The Importance of RNA Pairing Stability and Target Concentration for Regulation by MicroRNAs

by

David M. Garcia

B.S., Biochemistry and Molecular Biology (2004)
University of California, Santa Cruz

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2012

Signature of Author……
David M. Garcia
Department of Biology
May 25, 2012

Certified by……
David P. Bartel
Professor of Biology
Thesis Supervisor

Accepted by…….....
Robert T. Sauer
Salvador E. Luria Professor of Biology
Chair, Biology Graduate Committee

The Importance of RNA Pairing Stability and Target Concentration
for Regulation by MicroRNAs

by

David M. Garcia

Submitted to the Department of Biology on May 25, 2012
In Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

ABSTRACT

Regulation of gene expression in eukaryotes is highly precise and complex. Changes in expression can define the fate of each cell, convert healthy tissues to diseased ones, and even lead to speciation. Regulation occurs at the steps of transcription, mRNA processing and stability, and translation. In the last decade, the scope of post-transcriptional regulation has been dramatically widened through uncovering widespread small RNAs as critical regulators of gene expression in eukaryotes.

MicroRNAs (miRNAs) compose a major class of small regulatory RNAs. They are ~22-nt in length and bind to complementary sites in messenger RNAs to direct their degradation and translational repression. A central question for uncovering the biological roles of miRNAs is to understand how they find their target mRNAs, and a decade of work has highlighted one feature as most critical: base pairing between the 5′ end of the miRNA and a complementary site usually located in the 3′ UTR. One particular miRNA from the model organism *Caenorhabditis elegans*, called *lsy-6,* had in earlier studies not followed this principal, as most complementary sites were not repressed, which both intrigued and confounded the field.

This thesis presents studies of *lsy-6* targeting, conducted in human cell lines using heterologous reporter assays, which uncovered the reasons for this miRNA's generally poor targeting proficiency. These reasons are the weak pairing stability between *lsy-6* and a target site in an mRNA, as well as the high number of endogenous mRNAs *lsy-6* can bind to. Through a collaboration, the importance of RNA pairing stability and target concentration for miRNA targeting was extended to other miRNAs and siRNAs. Besides reconciling the unusual targeting behavior of *lsy-6* with the widely accepted model of miRNA targeting, these results also further suggest a mechanism of repression of its *in vivo* targets that is more complex than for most other miRNAs.

Thesis Advisor: David P. Bartel
Title: Professor

Acknowledgements

I thank Dave for giving me the wonderful opportunity to work in his lab, and for all that he taught me. His dedication, intellect, and fearlessness represent a model I will strive to follow in my career.

I thank my thesis committee members Phil Sharp and Steve Bell for their valuable input to my project development, and my training experience in general. I have mountains of respect for them and their work; they've provided me with equal great example to follow.

I thank my collaborators Daehyun Baek, Andrew Grimson, and Chanseok Shin for their valuable discussions and essential contributions.

I thank many people in the lab for providing an extremely fun and stimulating environment in which to work: Alex, Andrew, Anna, Calvin, DK, Graham, Huili, Igor, Jinkuk, Jin-Wu, Kathy, Laura, Lena, Lori, Michael, MLam, Muhammed**,** Noah, Robin, Sheq, Sue-Jean, Vikram, Vincent, Weinberg, and Wendy. Among them I've established several close friendships, and I hope to maintain personal and professional relationships with these people for years to come.

Elsewhere at MIT, I thank my cycling buddies—Calvin Jan, James Partridge, and Vincent Auyeung. I've also had a blast drumming in the MIT Jazz Combos for the last five years. Coach Keala Kaumeheiwa and dozens of fellow musicians have taught me as much about playing Jazz as I've learned about biology at MIT. Biograds2006 provided fun and some sanity in my first couple years here, and I'm fortunate for the friends I made as a result.

A big thanks to my previous scientific mentors, who, one by one, lead me on a path to MIT—Steve Beckendorf, Russell Sanchez, Bill Scott, and Corey Largman. Also thanks to the CAMP and SHURP programs which made my undergraduate research experiences further possible.

I lastly thank my parents for their love and support.

Table of Contents

Chapter 1

Introduction

Cis-regulatory elements: important regulators of gene expression

Over billions of years, evolution has tinkered with the composition of life, adjusting not only which genes are included in a cell or organism, but how much each gene is expressed, and where and when this happens. A large enough ensemble of small changes in the expression of different genes can lead to different biological outcomes, such as the divergence of one species into two, or the conversion of a healthy tissue to a diseased one. Regulation of gene expression is thus characterized by a high degree of precision and complexity.

Sampling genetic variation in *cis*-regulatory elements—relatively short tracts of DNA or RNA sequence that influence how genes are expressed—enables the testing of different levels of gene activity in an organism. This process can potentially yield new phenotypes that could be subject to selection, but yet are subtler than those resulting from changes in coding sequence that often directly impact the molecular structure and function of the encoded gene product. Comparative genome analyses of different model organisms has shown that *cis*-regulatory complexity is correlated with organismal complexity (Levine and Tjian, 2003); modulation of *cis*-regulatory sequences may also be a principal driver of morphological diversity during evolution (Carroll, 2000).

DNA *cis*-elements are central to transcription regulation, where they serve as binding sites for regulatory proteins (*trans*-factors). In all the post-transcriptional steps of gene expression in eukaryotes—including mRNA processing, translation and RNA localization—

RNA *cis*-elements can serve as binding sites for regulatory proteins or ribonucleoprotein complexes. RNA *cis*-elements can also fold into defined structures that modulate gene activity. In the cases where *cis*-elements bind *trans*-factors, interplay between the evolution of *cis*-sequences and the sequence specificities of *trans*-factors can add further layers of complexity to gene regulation.

One classic example of gene regulation—oft repeated in biology textbooks and Ph.D. theses—is that proposed by Jacob and Monod, a broadly descriptive model developed from studies of the lactose system in *E. coli* (Jacob and Monod, 1961). In essence, they developed the idea that repressors could interrogate genes to regulate their corresponding messenger RNA synthesis or protein synthesis, thereby controlling protein output. Their prescient notion of a regulatory molecule binding to a target gene to control its expression remains, 50 years later, the paradigm in post-transcriptional control of gene expression.

A perfectly emblematic modern-day example of this, which has garnered significant interest in the last decade, is the repression of mRNAs by tiny ~22-nt RNAs called "microRNAs." MicroRNAs (miRNAs) bind complementary sites in the mRNAs of protein coding genes to inhibit their expression through degradation and translational repression (Ambros, 2004; Bartel, 2004). Since miRNAs and their binding sites are ubiquitous in the mRNAs of plants and animals, this pathway represents a compelling example of post-transcriptional regulation of gene expression, one detail of which is the subject of this thesis.

RNAi

It all started with a purple petunia. Working at a biotech company in California in the late 1980's, Rich Jorgensen and others sought to impress investors with their skill in genetically modifying plants, so they added extra copies of a gene conferring purple pigment to petunias, to intensify the flowers' purple color. Unexpectedly, they produced flowers that appeared variegated, or even completely white, and a mystery was born.

Jorgensen and colleagues termed the phenomenon "co-suppression," referring to the fact that the activity of both the transgene and the homologous endogenous pigment gene was suppressed (Napoli et al., 1990), and in parallel another group in the Netherlands arrived at identical results (van der Krol et al., 1990). Post-transcriptional gene silencing (PTGS) induced by transgenes or viruses was soon found in other plants, as well as in the fungus *Neurospora crassa* where it was termed "quelling" (Cogoni et al., 1996; Romano and Macino, 1992). Through these studies it was found that introduction of transgenes or viruses could lead to their rapid degradation, as well as that of any endogenous homologs, but what triggered the process remained unknown.

Studies in the nematode *C. elegans* would find a similar role for post-transcriptional gene silencing, and importantly, define the trigger: double-stranded RNA (dsRNA). One early hint that RNA was involved came when researchers injected an RNA antisense to the *par-1* mRNA in order to block its expression through presumed interference with translation, a common technique at the time. It worked, but injection of a sense RNA (the negative control) worked as well (Guo and Kemphues, 1995). A few years later, Andrew Fire and Craig Mello and colleagues discovered that injecting dsRNA corresponding to an endogenous gene was much more efficient

at gene silencing, in some cases even spreading to the offspring (Fire et al., 1998). The silencing observed from sense or antisense RNAs to *par-1* in the Guo and Kemphues paper probably resulted from contaminating amounts of dsRNA.

"RNA interference," or "RNAi" as it was known, thus represented the potent and specific silencing by dsRNA (Fire et al., 1998). The following year it was reported that 25-nt RNAs were associated with PTGS in plants (Hamilton and Baulcombe, 1999). It was confirmed that small RNAs, processed from long dsRNA precursors, were directly responsible for gene silencing in studies using *Drosophila* cell or embryo extracts in which 21- to 23-nt RNAs processed from dsRNA precursors lead to cleavage of homologous mRNAs (Hammond et al., 2000; Zamore et al., 2000).

The nature of dsRNA that enters the RNAi pathway depends on its origins. It can be a bimolecular duplex, as is the case for siRNAs, or it can be a unimolecular RNA that folds into a hairpin with extended double-stranded character, as is the case for miRNAs. Bimolecular duplexes can result from either the pairing of sense and antisense transcripts, or the activity of RNA-dependant RNA polymerases (RdRPs) that are present in plants, fungi, and nematodes, but absent in flies and mammals (Ahlquist, 2002). Viral derived dsRNAs, such as a replication intermediate for an RNA virus, will also readily enter the pathway in plants, as well as in animals with more rudimentary immune systems like nematodes and flies, highlighting what is likely to be one of the original purposes of RNAi—defense against viruses (Shabalina and Koonin, 2008).

An RNase III enzyme called Dicer processes long dsRNAs into ~22-nt siRNA duplexes in a phased manner, leaving each strand with a 5′ phosphate and 2-nt 3′ hydroxyl overhangs (Bernstein et al., 2001; Elbashir et al., 2001a; Ketting et al., 2001; Zamore et al., 2000). After

loading of the duplex into the RNA induced silencing complex, or "RISC" (Hammond et al., 2000; Martinez et al., 2002; Nykanen et al., 2001), one strand is stably incorporated. Selection of the loaded strand is dictated by the relative thermodynamic stabilities of each end of the duplex—the strand whose 5′ end lies at the end with lower stability (resulting from mismatches, bulges, or more A:U or G:U pairs) is the one that stays (Khvorova et al., 2003; Schwarz et al., 2003), known as the guide strand. The opposite passenger strand is cleaved by RISC, and then the cleavage fragments are released (Matranga et al., 2005; Miyoshi et al., 2005; Rand et al., 2005). The sequence of the loaded siRNA then provides specificity for finding a target that is perfectly or nearly perfectly complementary (Bernstein et al., 2001; Elbashir et al., 2001a; Zamore et al., 2000). Once the siRNA pairs to a target within RISC, the central effector protein of complex, the endonuclease Argonaute, directs cleavage of the target between nucleotides opposite positions 10 and 11 of the siRNA (Elbashir et al., 2001a; Elbashir et al., 2001b; Liu et al., 2004; Song et al., 2004). Upon target release, the siRNA loaded RISC can go on to mark other target RNAs for cleavage (Haley and Zamore, 2004; Hutvagner and Zamore, 2002).

Huge numbers of small RNAs

miRNAs

While the molecular details of RNAi were being worked out, researchers were also discovering hundreds of endogenous small RNAs in plants, worms, flies, and mammals. Years earlier, researchers had unlocked a regulatory role for the non-coding 22-nt RNA *lin-4* in controlling larval development in *C. elegans* (Lee et al., 1993). Another RNA of the same size was later discovered, and it was also shown to have a role in the *C. elegans* heterochronic pathway, at a

later stage than *lin-4* (Reinhart et al., 2000; Slack et al., 2000). Homologs of this RNA, *let-7*, were soon found in the human and *Drosophila* genomes, and could be detected by Northern blots of samples from more than a dozen bilateral animals (Pasquinelli et al., 2000). As had been suggested for *lin-4* earlier, the mature *let-7* RNA was also predicted to originate from an unstable precursor transcript that formed a conserved stem-loop structure. Within a year this class of small RNAs would grow from two to more than one hundred when three labs cloned dozens of new genes encoding small RNAs in worm, fly, and human (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). These additional small non-coding RNAs were found to be conserved, have varied expression profiles, and joining founding members *lin-4* and *let-7*, became known as microRNAs, or miRNAs.

In the last ten years, the number of newly described miRNAs has skyrocketed, and current estimates number well over 500 in humans, about 200 in the model plant *Arabidopsis thaliana*, and close to 150 each in *Drosophila* and *C. elegans*. MicroRNAs have even been found in distantly branching simple animals like the marine sponge *Amphimedon queenlandica,* where 8 were cloned (Grimson et al., 2008). An Internet registry called miRBase contains a comprehensive list of published miRNA sequences (albeit containing a number of false positives resulting from sequencing and computational predictions that fail further experimental confirmation)(Kozomara and Griffiths-Jones, 2011).

Endogenous siRNAs

Deciphering the trigger for RNAi depended in part on the robustness of *C. elegans* processing exogenously sourced long dsRNA molecules into short siRNA duplexes capable of gene

silencing. Worms also have two major classes of endogenous siRNAs, 22- or 26-nt in length (Claycomb et al., 2009; Gu et al., 2009b; Han et al., 2009; Ruby et al., 2006). They map to a fraction of protein coding genes (and also intergenic regions), including many germline-associated and stage-specific genes, in both sense and antisense orientations with respect to coding sequences. *C. elegans* encodes an RdRP that is thought to amplify initial "primary" siRNAs into a greatly increased number of "secondary" siRNAs, using transcripts complementary to the primary siRNAs as templates (Pak and Fire, 2007; Sijen et al., 2007). The details of how RdRP is recruited by primary siRNA-loaded Argonautes to templates are unknown. The amplification is also not Dicer-dependent, so the mechanism by which the long dsRNA precursors are processed into small secondary siRNA duplexes also remains a mystery. Since the secondary siRNAs heavily outnumber the primary siRNAs, distinguishing the two populations by high-throughput sequencing methods remains an unsolved challenge.

Flies have endogenous siRNAs originating from a variety of sources—transposons, mRNAs, and long hairpins precursors (Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008). Like in worms, the canonical RNAi pathway in flies also readily processes exogenously derived dsRNA substrates into small siRNA duplexes that can silence genes.

Even yeast have siRNAs. In contrast to their post-transcriptional silencing roles in plants and animals, siRNAs in the fission yeast *Schizosaccharomyces pombe* direct transcriptional silencing. After being loaded into the RNA-induced initiation of transcriptional gene silencing (RITS) complex, they direct histone methylation on their target loci, leading to heterochromatin formation (Reinhart and Bartel, 2002; Verdel et al., 2004; Volpe et al., 2002). The budding yeast

Saccharomyces castellii has siRNAs that direct post-transcriptional silencing through an RNAi pathway in which a single Dicer and Argonaute protein are sufficient for post-transcriptional silencing (Drinnenberg et al., 2009).

In *Arabidopsis thaliana*, there are several classes of siRNAs that can be distinguished by their lengths, modes of biogenesis, and targets (Ghildiyal and Zamore, 2009; Mallory and Vaucheret, 2006). Natural antisense siRNAs (natsiRNAs) and *trans*-acting siRNAs (tasiRNAs) are endogenous siRNAs that silence near perfectly matched mRNAs through cleavage, while *cis*-acting siRNAs (casiRNAs) direct DNA methylation and histone modification at homologous loci.

piRNAs

Another class of metazoan small regulatory RNAs is the PIWI-interacting RNAs (piRNAs)(Aravin et al., 2006; Girard et al., 2006; Lau et al., 2006). They arise from genomic clusters that are expressed and amplified by a still to be worked out mechanism and silence transposons and repeat elements (Aravin et al., 2007). Their activity seems to be especially important in the germline, where transposons can be very active during early stages of development. The *C. elegans* version of piRNAs appears to be the 21U-RNAs which were originally named for their size and first nucleotide preference (Ruby et al., 2006); 21U-RNAs were later found to share some similarities to piRNAs (Batista et al., 2008; Das et al., 2008).

Others

The lack of similarity between animal and plant miRNA identity (and structure—plant miRNAs

tend to be processed from stem loop precursor transcripts that are generally larger and more variable in size) implies that these pathways evolved convergently. An ancestral RNAi pathway probably consisted of siRNA- and piRNA-like small RNAs for defense against viruses and transposons. It is likely that still other mechanisms exist by which siRNAs can be generated or function in other less studied organisms, yet to be discovered. Recently bacteria and archaea have been found to have their own cohort of small RNAs that can silence genes using a distinct set of proteins in the CRISPR pathway (Marraffini and Sontheimer, 2010). Some bacterial and archaeal species also contain Argonaute related proteins (Makarova et al., 2009). CRISPR differs in many ways from RNAi—the small RNAs tend to be slightly longer, the effector proteins are different and not conserved in any metazoans examined thus far, and silencing occurs at the DNA level (to destroy plasmids and phage) in all but one example characterized (Hale et al., 2009). Nevertheless, it's interesting to note that both CRISPR and RNAi appear to have evolved originally to silence parasitic elements like phage, viruses, and transposons, often by cleaving the nucleic acid portions of these elements directly and using the resulting products to target additional copies of these elements. In plants and animals, the RNAi machinery was adapted to serve the miRNA pathway, in which endogenous small RNAs repress genes in *trans*, targeting mRNAs arising from loci apart from which the miRNAs arise (Bartel, 2004).

miRNA biogenesis and origins

The distinction between endogenous miRNAs and siRNAs in metazoa arises not from their functional differences, but from their biogenesis and origins. Since both classes are similar in length and have the same chemical moieties on their ends, miRNAs can direct target cleavage on

sites with high complementarity (Hutvagner and Zamore, 2002; Song et al., 2004; Yekta et al., 2004), and siRNAs can function as miRNAs on sites with only partial complementarity, like those typical of miRNA targets (Doench et al., 2003).

The types of genomic loci miRNAs arise from are diverse: independent transcription units, polycistronic transcripts encoding multiple miRNAs, and within the introns or exons of protein coding genes (Kim et al., 2009). MicroRNAs are transcribed by RNA Polymerase II into primary miRNA transcripts (pri-miRNAs) of ~120-nt that fold into stem-loop structures (Figure 1). The pri-miRNA is composed of a hairpin containing the miRNA and a partially complementary sequence that are connected by a loop at one end, flanked by long single stranded regions at the other end. The first processing step removes the single-stranded flanks through the nuclear RNase III enzyme Drosha with its cofactor DGCR8/Pasha (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004; Landthaler et al., 2004; Lee et al., 2003). Cleavage occurs at ~11-nt, or roughly one helical turn, from the base of the stem, which determines one end of the eventual mature miRNA duplex (Han et al., 2006). The resulting pre-miRNA is then exported to the cytoplasm via Exportin-5 (Lund et al., 2004). In the cytoplasm, Dicer, the same RNase III enzyme responsible for processing long dsRNA into siRNA duplexes, processes pre-miRNAs. After recognizing the free end containing the 2-nt 3′ overhang left by Drosha, Dicer cleaves off the terminal loop at the other end by measuring roughly two helical turns away (Lee et al., 2002; Macrae et al., 2006), generating the final ~22-nt miRNA duplex containing the miRNA and miRNA* (miRNA "star") with 2-nt, 3′ overhangs on each strand. This short duplex maintains the imperfect pairing characteristic of the stem of the original pri-miRNA. Another protein known as TRBP/Loquacious assists Dicer in miRNA processing and subsequent

recruitment and loading into Argonaute (Ago), the core protein component of the miRNA RISC (miRISC)(Chendrimada et al., 2005; Forstemann et al., 2005). If the miRNA duplex is loaded into an Ago protein with cleavage activity (e.g. Ago2 in mammals), the miRNA* can be cleaved and removed (Matranga et al., 2005; Miyoshi et al., 2005; Rand et al., 2005). If the duplex is loaded into a non-cleaving Ago (e.g. Ago1, Ago3, and Ago4 in mammals, all of which can repress target mRNAs without cleaving them), the miRNA* is removed from miRISC by an unknown mechanism. Occasionally the miRNA* is stably loaded over the miRNA, and regulation of which strand is loaded can even be dynamic during development (Chiang et al., 2010). Deep sequencing of many organisms, however, has generally shown strong preference for loading one strand of the duplex—by definition this strand is usually referred to as the miRNA.

There are also a few non-canonical miRNA biogenesis pathways. In one special case, miRNAs arise from introns where the termini of the pre-miRNA species is defined by splice sites processed by the spliceosome. These molecules therefore bypass Drosha processing, and are called mirtrons (Ruby et al., 2007a). Mirtrons were first found in worms and flies which both have an average intron size close to that of pre-miRNAs, but they've also been found in mammals (Babiarz et al., 2008; Berezikov et al., 2007). They suggest a mechanism for the generation of new miRNAs before Drosha emerged with its dominant role in the first step of biogenesis. There is also one case of a miRNA that bypasses Dicer processing, the conserved miR-451. It has an unusual stem loop structure which is first processed by Drosha and then enters directly into Ago which cleaves the passenger arm and trims the 3′ end of the miRNA to generate mature miR-451, now primed for targeting (Cheloufi et al., 2010; Cifuentes et al., 2010).

Loaded miRNAs are considered to be very stable, with estimated half-lives sometime exceeding days (van Rooij et al., 2007). Once loaded though, miRNAs are not totally refractory to modification, as they can be trimmed and tailed at their 3′ end, with tailing adding untemplated adenosines or uridines (Ameres et al., 2010; Cazalla et al., 2010). Such modifications are suggested to affect the stabilities of miRNAs.

Some miRNAs can be grouped into families based on sharing an identical (or nearly identical) sequence at their 5′ end, know as the "seed" sequence (Lewis et al., 2003). Since the seed sequence is the most influential factor in determining targets (more in the following section), miRNAs that share a seed are also predicted to share the same targets. Of course, not all family members will be co-expressed similarly, and differences in sequence at their 3′ ends can further distinguish targets that have the potential for additional pairing with this region, as well as affecting their loading or stability (Bartel, 2009).

While most miRNA genes arise from individual transcripts, some come from polycistronic transcripts containing multiple miRNAs often related in sequence or function. As could be expected, these clustered miRNAs tend to have coordinated expression patterns, and intronic miRNAs also tend to have coordinated expression with their host genes (Baskerville and Bartel, 2005).

Theories about the origins of new miRNAs implicate gene duplication followed by promotion of the miRNA* as the dominant species of the duplicated copy (subfunctionalization); a duplicated copy acquiring mutations conferring novel function in the miRNA or miRNA* (neofunctionalization); or *de novo* emergence from a portion of a pre-existing RNA transcript

(including introns) capable of folding into a hairpin structure capable of entering the miRNA biogenesis pathway (Ruby et al., 2007b).

miRNA targeting

The seed rules

The first clues to how miRNAs find their targets came through the first studies of the *C. elegans* miRNA *lin-4* when it was noted that this RNA had sequence complementarity to several conserved sites in the 3′ UTR of the *lin-14* mRNA (Lee et al., 1993; Wightman et al., 1993). These sites had previously been shown to be required for regulation of *lin-14* by *lin-4*, even before it was known that *lin-4* was a non-coding RNA (Wightman et al., 1991). Genetic analyses uncovered these first examples of miRNA–target relationships, but once larger scale efforts in cloning and computational discovery revealed hundreds of miRNAs, predicting all of their targets became a much larger and more complex puzzle.

The predicted binding sites for *lin-4* in the *lin-14* mRNA were not fully complementary to the miRNA. In contrast, for plant miRNAs it was noted that targets could be confidently predicted by searching for near perfect matches, suggesting they would lead to mRNA target cleavage, as is the case for siRNAs (Rhoades et al., 2002). For animals however, targets with extensive complementarity subject to cleavage have been found only in rare cases (Davis et al., 2005; Yekta et al., 2004). There are also examples of so called “center sites,” in which a target site pairs with the center of the miRNA, which can lead to cleavage or mRNA repression (Shin et al., 2010). It is now clear, however, that the dominant class of miRNA target site in animals has complementarity with the 5′ end of the miRNA, known as the seed sequence (Figure 2).

Quantitative analyses of microarray, proteomics, and ribosome profiling datasets monitoring the response of gene expression to transfection of a miRNA into cell lines has revealed a hierarchy in the types of sites that mediate repression (Figure 3)(Bartel, 2009; Guo et al., 2010). Basal sites have perfect Watson-Crick pairing to positions 2–7 of the miRNA—defined as the seed—and these sites are called 6mers, which tend to confer only weak repression. In terms of the amount of repression conferred, the next most effective site has perfect seed pairing and an Adenine (A) across from the 1 position of the miRNA—these are called 7mer-A1 sites. The A in the target site across from the first nucleotide of the miRNA benefits site efficacy regardless of the identity of the first nucleotide of the miRNA, as these bases do not pair within the silencing complex (Lewis et al., 2005; Ma et al., 2005; Parker et al., 2005). Thus this A contributes to some aspect of targeting other than seed pairing. The next most effective site has perfect pairing at positions 2–8, known as a 7mer-m8 site. The most effective sites on average are 8mers, which combine aspects of the 7mer-A1 and 7mer-m8 sites—they have perfect pairing at positions 2–8 and an A in the site at position 1. Conservation analyses has also uncovered the offset 6mer site, which has pairing with positions 3–8 of the miRNA (Friedman et al., 2009), but these sites are repressed even less than canonical 6mers.

Computational studies provide strong support for the importance of seed pairing for target recognition. Requiring conserved Watson-Crick base pairing with the miRNA seed can significantly reduce the rate of false-positive miRNA target predictions when searching through a genome, significantly more than matches to any other region of the miRNA (Lewis et al., 2003). After subtracting the number of sites expected to be conserved by chance (e.g. shuffled sequences with similar properties to seed sequences), the number of conserved matches to the

seed region of a miRNA can be in the hundreds in coding sequences, with greater numbers of predicted targets for highly conserved miRNAs (Brennecke et al., 2005; Krek et al., 2005; Lewis et al., 2005; Xie et al., 2005). More than half of human protein-coding genes appear to be selectively maintaining sites complementary to the seed of one or more miRNAs based on conservation analysis (Friedman et al., 2009). In this most recent conservation analysis, the ability to predict the number of selectively conserved targets improved through the use of better control sequence cohorts that minimize the inclusion of sites conserved for reasons other than miRNA regulation. For example, these analyses utilize control shuffled sequences that correct for factors that can affect the conservation level of short sequences, such as GC content, dinucleotide content, and local conservation rates. Transcriptome-level experimental data, monitoring the effects of overexpressing a miRNA for example, has greatly improved our understanding of those types of miRNA target sites are functional and those that are not. This has in turn lead to the development of miRNA target prediction algorithms that rank the repression levels of all mRNAs predicted to be targeted by a given miRNA, in a quantitative fashion (Bartel, 2009; Grimson et al., 2007).

Numerous types of experimental studies also support the importance of seed pairing for target recognition. One early report noted that short sequence elements previously shown to mediate post-transcriptional repression of mRNAs were complementary to only the 5′ region of *Drosophila* miRNAs (Lai, 2002). Subsequent global approaches helped further cement the model. Transfecting exogenous miRNA duplexes into HeLa cells resulted in the downregulation of hundreds of mRNAs containing seed matches, enough to shift the HeLa expression profile toward a cell type in which the transfected miRNA is normally highly expressed (Lim et al.,

2005). Complementing the transfection based experiments, sequestering miR-122 (a miRNA highly expressed in liver) in mice using novel chemically engineered oligonucleotides lead to the observation that hundreds of upregulated messages in liver were enriched for sites matching the seed of miR-122 (Krutzfeldt et al., 2005). In another *in vivo* example, compared to mRNA levels in wild-type zebrafish embryos, the mRNA levels in embryos in which Dicer was knocked out showed significant enrichment for seed matches to miR-430—a miRNA normally highly expressed in embryos—in hundreds of upregulated messages (Giraldez et al., 2006). Analogous results were found in microarray analyses of murine immune cells lacking miR-155 (Rodriguez et al., 2007). Proteomics based approaches have also shown seed matches as being the sequence motif most highly associated with downregulation of protein levels upon miRNA transfection into cell lines, or derepression of protein levels in the absence of an endogenous miRNA (Baek et al., 2008; Selbach et al., 2008). Studies utilizing co-immunoprecipitation of Argonaute protein cross-linked to physically associated mRNAs, followed by deep sequencing of bound mRNA fragments, showed strong enrichment for seed matches to transfected or highly expressed endogenous miRNAs (Chi et al., 2009; Hafner et al., 2010; Leung et al., 2011). Other studies have shown directly the dependence on seed-matched sites in mRNAs for co-immunoprecipitation with miRISC (Karginov et al., 2007).

Regulation of sites lacking perfect seed pairing, such as those with mismatches or bulges, and those containing GU wobble pairs, has been suggested in a few cases, most notably for *let-7* sites in the *lin-41* 3′ UTR in *C. elegans* (Vella et al., 2004). Evaluating the general effectiveness of these non-canonical sites on a larger scale, however, is hampered by the numerous possible pairing schemes, and in the cases of co-immunoprecipitation of miRISC complexes, the fact that

it's difficult to confidently assign target sites pulled down to the action of a specific miRNA, since each cell expresses dozens to hundreds of them.

One noteworthy study questioned the overall predictive value of seed matches, despite much evidence to the contrary. In this study the authors found that when they monitored the response of GFP reporter constructs bearing the 3′ UTRs of 14 predicted targets of the *C. elegans* miRNA *lsy-6*, only one was repressed (Didiano and Hobert, 2006). Each 3′ UTR contained one or two sites, most of them conserved, and the reporter transgenes were expressed in two neurons, one of which endogenously expresses *lsy-6* and one that does not. In comparing GFP levels between the two cells, it could be evaluated which targets respond to endogenous *lsy-6*, and the only UTR that responded was from *cog-1*. This UTR contains two well conserved *lsy-6* sites, and was a known target of the miRNA from earlier genetic studies (Johnston and Hobert, 2003). Because only 1 of 14 targets was repressed, and repression of *lsy-6* sites placed ectopically in other UTRs was not always observed, the authors concluded that "perfect seed pairing is not a generally reliable predictor for miRNA target interactions" (Didiano and Hobert, 2006). Even in the shadow of a large body of evidence in support of a seed based targeting model, the intriguing data of this study, which profiled *in vivo* targeting interactions, warranted further study.

Beyond the seed: other determinants

While often predictive of repression, a seed match site in an mRNA is not always sufficient. Other determinants include pairing with the 3′ end of the miRNA, but more importantly, the sequence context of the site in the UTR (Bartel, 2009).

Supplementary pairing to nucleotides 13–16 of the miRNA can boost repression modestly (Grimson et al., 2007), and these site types are preferentially conserved (Friedman et al., 2009). In a few cases, more extensive pairing to the 3′ end of the miRNA is believed to compensate for poor pairing with the seed (Brennecke et al., 2005; Friedman et al., 2009). For example, the two *let-7* sites in the *lin-41* 3′ UTR of *C. elegans* both have imperfect seed pairing, with either a GU wobble or a single nucleotide asymmetric bulge, and extensive pairing with the 3′ end of one of the *let-7* miRNA family members (probably making these sites specific for this single family member (Bartel, 2009)). While there are other examples of conserved sites with this type of compensatory pairing, they still make up only a very small minority of preferentially conserved sites, and it has not been determined if these sites consistently mediate repression (Friedman et al., 2009).

Several features of site context are important. The first is that to be effective, sites in 3′ UTRs must be positioned at least 15-nt from the stop codon. This feature can be explained by the footprint of the ribosome as it arrives at the stop codon, covering the first ~15-nt of the UTR (Grimson et al., 2007). This "ribosome shadow" is predicted to interfere with binding of a miRISC silencing complex. Consistent with ribosome interference making certain areas refractory to miRISC binding, 5′ UTRs don't generally have any effective sites, and sites in the ORF tend to be less effective than those in 3′ UTRs (Bartel, 2009). Effective ORF sites may rely on a context of reduced translation (Gu et al., 2009a).

The second context feature correlated with target site efficacy is positioning of sites away from the middle of long 3′ UTRs (Grimson et al., 2007). The reasons for this aren't totally clear, but may involve proximity to translation machinery in the context of a circularized mRNA,

leading to more efficient repression since the UTR ends would be closer to either the ORF or the poly-A tail that interacts with the 5′ cap. There may also be a higher chance for the center of long UTRs to form secondary structure than the ends, since ORFs and poly-A tailed ends of UTRs are less likely to form structure with neighboring UTR sequence (Grimson et al., 2007).

A third context feature correlated with target site efficacy is multiple sites in close proximity to each other. Although multiple sites in a UTR generally act independently (Doench et al., 2003; Grimson et al., 2007), sites spaced in close proximity can act cooperatively (Grimson et al., 2007; Saetrom et al., 2007). This results in more repression from a pair or combination of sites than expected if the sites functioned independently. These studies estimated a distance requirement of ~10–40-nt between the sites to observe cooperativity, but the spacing could potentially be larger if two sites distantly apart in UTR sequence were brought physically close together through secondary structure in the UTR. The mechanism of cooperativity in miRNA targeting remains unknown, but possibilities include: cooperative binding of Ago (pre-associated or non-pre-associated) in which one Ago binding event strongly favors a second binding event nearby; after the binding of two Agos to two closely spaced sites, increased retention of one Ago by its interactions with the second Ago; cooperative recruitment of downstream silencing factors. One recent study reported that in mammalian cells only specific Argonautes could mediate cooperativity (Broderick et al., 2011).

The fourth context feature is AU-rich local sequence context, which is the additional determinant most correlated with target site efficacy (Bartel, 2009). Context is most important in the immediate vicinity of the site, ~30-nt upstream or downstream, after which the benefit declines (Grimson et al., 2007). Local AU-rich content is also observed to be elevated around

conserved sites compared to non-conserved sites. Some of the benefit of local AU content may be to minimize secondary structure and make sites more accessible to a silencing complex. Using computational methods to predict UTR secondary structure, several studies have proposed that reduced local secondary structure is correlated with miRNA target efficacy (Hammell et al., 2008; Kertesz et al., 2007; Robins et al., 2005; Zhao et al., 2005). One study used their structure-based model to explain that inhibitory structure could explain why so few *lsy-6* targets were repressed in the worm reporter assays from Didiano and Hobert (Long et al., 2007). Despite the correspondence between AU content and secondary structure, the AU content feature alone is still capturing effects missing from structural models. Including the AU content feature leads to a greater increase in performance of a target prediction algorithm than including secondary structure as a feature (Baek et al., 2008; Grimson et al., 2007).

Conservation and expression context considerations

MicroRNAs that have been conserved over longer evolutionary time scales, such as those found in both nematodes and mammals, tend to be expressed at higher levels and in broader cell contexts than poorly conserved miRNAs found only in closely related species, like those limited to the nematode clade (Bartel, 2009; Chen and Rajewsky, 2007). More broadly conserved miRNAs also tend to have a greater number of conserved predicted targets than do "nonconserved" miRNAs (Lewis et al., 2005; Lewis et al., 2003), demonstrating that miRNAs retained through evolution help shape the conservation of their cognate seed matches in the transcriptome. Conserved miRNAs with broader tissue expression patterns also tend to have more predicted targets than those with more restricted expression (Ruby et al., 2007b). As a

result, it is more difficult to predict a significant number of functional sites above background for poorly conserved miRNAs or those with lower or more restricted expression. For example, when evaluating 53 miRNA families conserved in mammals but not broadly in other vertebrates, the predicted number of functional sites above background levels was considerably lower than for target predictions for miRNAs conserved across vertebrates (Friedman et al., 2009).

Since the recognition motif in a miRNA is only 6–7-nt, nonconserved target sites outnumber conserved ones by 10:1 (Farh et al., 2005). Do all these nonconserved sites work? Could they just be titrating the miRNA away from its real sites? Reporter assays monitoring the response of nonconserved sites to miRNA transfection showed that many of these sites do indeed function, albeit less frequently than conserved ones (Farh et al., 2005). This supports the idea that the presence of a seed match alone can be a good predictor of which mRNAs will respond to a co-expressed miRNA, and implies that the presence of additional site context features might make conserved sites more effective. Since a target site alone has the potential to confer repression, highly expressed messages avoid harboring sites for co-expressed miRNAs (Farh et al., 2005; Stark et al., 2005). This phenomenon is known as selective avoidance, and thus expression of a miRNA can directly influence 3′ UTR evolution. Selective avoidance isn't complete however, as several studies observed derepression of many messages containing nonconserved sites upon loss of a miRNA (Baek et al., 2008; Giraldez et al., 2006; Krutzfeldt et al., 2005; Rodriguez et al., 2007).

That highly expressed mRNAs specifically avoid sites for co-expressed miRNAs is suggestive of a concentration dependence. In heterologous reporter assays, repression of conserved and nonconserved target sites can be lowered steadily by titrating down the amount of

transfected miRNA (Farh et al., 2005). Overexpression of transgenes containing target sites for endogenous miRNAs causes derepression of their endogenous targets (Ebert et al., 2007). Similarly, overexpression of exogenous miRNAs causes derepression of the endogenous targets of highly expressed endogenous miRNAs (Khan et al., 2009). These latter data suggest that the abundance of Ago/miRISC can be limiting.

Role of Argonaute

Although the context of pairing between a miRNA and mRNA is essential for promoting targeting, the Ago protein is what holds everything together and mediates repression. Three domains compose the portion of Argonaute that contacts a loaded miRNA. The MID domain has specificity for a 5′ monphosphate and uracil that is the predominant first base observed in miRNAs (Boland et al., 2010; Frank et al., 2010). The PIWI domain contacts the central part of the miRNA and contains the endonuclease DDH motif required for target cleavage in some Agos (Parker et al., 2004; Song et al., 2004). The 3′ end of the miRNA makes contacts with the PAZ domain, which enforces a steric block that limits the length of miRNAs to ~22-nt (Ma et al., 2004).

The importance of seed pairing for miRNAs (and also for nucleating perfectly matched sites for siRNAs) is reconciled structurally with the observation that within Ago, the miRNA seed region has greater binding affinity for target sites and is more structurally constrained than other regions of the miRNA (Haley and Zamore, 2004; Lambert et al., 2011; Parker et al., 2009). Sites with a mismatch or wobble to the seed are not conserved above background (Brennecke et al., 2005; Lewis et al., 2005; Lewis et al., 2003), and repression is highly sensitive to base

mismatches and GU wobbles (Brennecke et al., 2005; Doench and Sharp, 2004; Kloosterman et al., 2004; Lai et al., 2005). As such, sites that introduce mismatches are used as negative controls in heterologous reporter assays, and all messages lacking a site with perfect seed pairing are bundled together as a control set in genome-wide analyses ("no site"). Despite contributing positively to RNA duplex stability *in vitro* in the absence of protein, GU wobbles are apparently disfavored in Ago and therefore generally not included as criteria for making target predictions (Bartel, 2009).

The number of Argonaute proteins per organism varies widely. Worms have the most known of any organism, with 27 Agos, split among 3 general clades that associate with specific classes of small RNAs with differing functions. Mammals have a total of 8, split between two classes: 4 PIWI-type Agos and 4 regular Agos. Not all four human regular Agos are represented equally in cells. Ago2, which can cleave extensively paired targets, is the most highly expressed family member in at least two human cell lines, constituting ~60% of all Ago protein, followed by Ago1 and Ago3 (Petri et al., 2011). In this study Ago4 protein was barely detectable, despite reasonable mRNA levels. These four Agos do not exhibit loading preferences for certain miRNAs since, with a few exceptions, all miRNAs are distributed between them at similar frequencies (Burroughs et al., 2011; Meister et al., 2004).

Several lines of evidence suggest that Ago protein is limiting *in vivo*. Overexpression of any of the four human Ago proteins increases levels of mature miRNA from ectopically expressed constructs (Diederichs and Haber, 2007). This study also found that in Ago2 knockout murine cell lines, lower endogenous miRNA expression is observed, which is rescued by reintroduction of Ago2 in these cells. Overexpression of Ago2 can also enhance siRNA-directed

cleavage of perfect matching sites (Diederichs et al., 2008). Results mentioned in the previous section that demonstrated changes to endogenous silencing upon miRNA/target site mimic overexpression, lend further support to Ago levels being an essential component of small RNA directed silencing capacities. Silencing should also depend on rates of miRNA loading or turnover in Ago, which could vary for different miRNAs. For example, some miRNAs could reach loading saturation in Ago at lower expression levels than other miRNAs. Post-translational modifications of Ago proteins should also influence their function (Johnston and Hutvagner, 2011).

Mechanisms of miRNA repression

Upon pairing stably to a target site, miRNAs direct a combination of mRNA degradation and translational repression, mediated by the silencing complex and associated factors.

For years, many in the field were influenced by reports that translational repression was the dominant mode of repression, although there wasn't complete agreement as to whether this occurred at translation initiation or at a later step. Techniques including polysome profiling, reporter assays, and *in vitro* reconstitution systems were used, measuring where miRNAs co-sediment in polysome gradients or final protein output as a result of miRNA repression. Meanwhile, other studies pointed toward mRNA degradation as an important aspect of miRNA repression. One study monitored the response of messages to miRNA transfection in cell lines using microarrays, seeing a significant response from messages with seed matches which demonstrated that mRNA levels are significantly impacted (Lim et al., 2005). A complementary study reported derepression of many messages containing sites for miR-122 when the miRNA's

activity was blocked in liver cells where it is highly expressed (Krutzfeldt et al., 2005). Reexamination of the *lin-4*:*lin-14* interaction—that started the miRNA field and first observed translational repression without mRNA degradation—demonstrated that the *lin-14* mRNA was indeed reduced (Bagga et al., 2005).

The most common routes of mRNA degradation *in vivo* are removal of the protective 5′ caps or 3′ poly-A tails, and miRNA directed repression leads to both outcomes. Widespread deadenylation was first reported in the clearance of maternal mRNAs by miR-430 during zebrafish embryogenesis (Giraldez et al., 2006). Other studies connected canonical mRNA degradation machinery—the CCR4-NOT and PAN2-PAN3 deadenylase complexes, and the DCP1:DCP2 decapping complex—to miRNA directed mRNA degradation (Behm-Ansmant et al., 2006; Rehwinkel et al., 2005). The GW182/TNRC6 class of proteins are essential for repression (Jakymiw et al., 2005; Liu et al., 2005; Meister et al., 2005; Rehwinkel et al., 2005), and physically link Ago with Poly-A Binding Protein (PABP)(Fabian et al., 2009; Zekri et al., 2009) and deadenylation complexes (Chekulaeva et al., 2011; Fabian et al., 2011).

To investigate the relative contributions of mRNA degradation and translational repression by miRNAs, one study performed microarray analyses and high-throughput proteomics on miRNA transfected cell lines and miRNA knockout cells. This study found that most of the repression (or derepression for the knockout) at the protein level could be explained by changes at the mRNA level (Baek et al., 2008). A pair of subsequent studies looked at translation directly using global polysome profiling or ribosome profiling and put a number on the contributions, with translation repression accounting for ~20% of the signal and mRNA degradation the rest (Guo et al., 2010; Hendrickson et al., 2009). It's important to note that these

studies have mostly looked at late time points after introduction/deletion of a miRNA, and closer examination of the kinetics of repression will be necessary to get a complete view of the mechanisms of repression. Future studies will hopefully also determine whether the exact mode and magnitude of repression depends on the nature of the miRNA–target interaction, or the cellular context.

Relatively little is known about where in the cell repression occurs exactly. Ago and GW182/TNRC6 proteins and repressed mRNAs have been observed in P bodies and stress granules, sometimes in a miRNA-dependant way (Leung et al., 2006; Liu et al., 2005; Pillai et al., 2005; Sen and Blau, 2005). Since these two cellular foci are known to contain mRNA degradation machinery and translationally repressed mRNAs, respectively, they are good candidates for sites of repression, but it is not yet clear all the locations in the cell where miRNAs are active.

miRNAs–target relationships and biological function

While miRNA annotation and our understanding of targeting mechanisms has grown quickly in the last decade, efforts to assign specific regulatory roles to each miRNA has developed at a slower pace. Unbiased genetic screens in model organisms like worms paved the way toward understanding roles for the first characterized animal miRNAs, and these continued efforts, along with several miRNAs knockouts in mammals, and other pathway studies, have deciphered more.

lin-4 and *let-7*, the first two miRNAs to be characterized in detail, both emerged from screens for genes involved in developmental timing in *C. elegans* larva (Ambros, 2004). Loss of these miRNAs results in ectopic expression of their target mRNAs, which leads to omission, or

reiteration of cell fate decision events during development (Lee et al., 1993; Reinhart et al., 2000). The roles of these miRNAs in these developmental events can also be connected to single targets. For example, animals with a gain of function mutation in *lin-14*, the target of *lin-4*, phenocopy a *lin-4* null (Lee et al., 1993; Wightman et al., 1993). Since *lin-14* contains three sites for *lin-4*, this targeting interaction is strong enough to be classified as a developmental switch.

More than two-dozen miRNAs have been knocked out in mice, some with broad expression, and others with more tissue specific expression in hematopoietic cells or heart and skeletal muscle tissue (Park et al., 2010). Some knockouts produce no obvious phenotype, while others produce phenotypes in the tissues where the miRNA is specifically expressed. The task of assigning specific targets of the miRNA to a specific phenotype is inherently difficult, because the only sure way to know is to mutate the predicted target sites for the miRNA in individual candidate target genes and compare the phenotype to the miRNA knockout. Nearly all of the known *C. elegans* miRNAs have been knocked out, and only a few had obvious phenotypes (Miska et al., 2007). This implies that many could be functionally redundant, which is supported by the observation that while knocking out an entire miRNA family was sometimes sufficient to produce a phenotype, knocking out individual miRNA family members was not. Another option is that some of the miRNAs are not essential for the worm's fitness under laboratory conditions. A mutant phenotype therefore may only result when testing a specific behavior or subjecting the animal to a specific stress not normally present in the lab.

When trying to predict targets strongly repressed by a miRNA, UTRs with multiple sites to the seed are often the best starting candidates. For example, *Hmga2* is one of the top predicted targets of the *let-7* family in mammals because it has seven sites for *let-7*, and this UTR is

strongly repressed by the miRNA in cell culture (Mayr et al., 2007). A large portion of this UTR is lost in a type of chromosomal rearrangement that results in overexpression of HMGA2 protein, and this event is correlated with cell transformation and cancer. Therefore in this relationship, *let-7* has a tumor suppressor role. As one of the most broadly and highly expressed miRNAs in animals, it surely has numerous other roles as well.

Most miRNA–target interactions do not fall into the category of switches. Instead they are more likely to be "fine tuners," either actively dampening target gene expression in a modest but still physiologically important way, or simply dampening expression of targets with neutrally evolving sites with no immediate consequence for the cell (including if the interactions are lost)(Bartel and Chen, 2004). Individual miRNAs may act as switches on some targets, fine tuners of others, and have neutral interactions with others, roles that could change between cell types and developmental states. Like switch interactions, fine tuning interactions are under selective pressure to be maintained over evolution by conserving the site, to be contrasted with neutral interactions where the sites are not conserved above neutrally evolving background sequence. The ability of fine tuning interactions to dampen protein output appears to benefit cells by optimizing target protein levels or dampening noise in gene expression, otherwise they would not be conserved above background levels. Neutral interactions, on the other hand, appear to confer repression that is tolerated or offset in the cell, but is not relevant enough to have been maintained over evolution (Bartel, 2009). Among the conserved targeting interactions, our understanding of which are switches and which are fine tuners remains limited by the pace at which they are characterized individually in model systems.

One of the better studied miRNA–target relationships is that between the *C. elegans lsy-6* miRNA and its target *cog-1*. Based on promoter tagging experiments, *lsy-6* is expressed in no greater than ten cells in an entire worm (~1000 somatic cells total). The miRNA was discovered in a forward genetic screen for genes involved in neuronal patterning (Johnston and Hobert, 2003). Using the model of two morphologically bilateral taste receptor neurons, called ASE left (ASEL) and ASE right (ASER), a mutagenesis screen was performed in search of animals in which expression of GFP driven by the promoters of terminal cell fate markers distinct in ASER and ASEL cells was aberrant. Wild-type ASER and ASEL cells differ in their ability to discriminate different ions by expressing distinct sets of chemoreceptor genes, which is important for worms' perception of bacterial food sources. A gene that was important for establishing the fate of one of the ASE neurons, when mutated, would often result in symmetric GFP expression (representing the terminal cell fate markers), and these mutants were assigned to the class "laterally symmetric," or *lsy*. The screen produced dozens of known or predicted protein-coding genes, but there was one, *lsy-6*, which was found to encode a miRNA that is normally only expressed in the ASEL cell and not in the ASER cell. *lsy-6* mutants display a phenotype in which both ASE cells adopt an ASER fate. *lsy-6* loss-of-function animals do not exhibit an obvious phenotype morphologically or behaviorally, but do show impaired chemotaxis when assayed specifically for this trait. The miRNA emerged in this assay designed to find genes involved in a neuronal patterning years before it was cloned by deep sequencing methods, given its low abundance in the animal. The miRNA acts like a switch because ectopic expression of *lsy-6* in the neuron destined for an ASER fate causes it to adopt an ASEL cell fate instead.

lsy-6 was found to target two closely spaced sites in the 3′ UTR of the *cog-1* mRNA that encodes a transcription factor with another essential role in ASE cell fate specification (Figure 4). *cog-1* loss-of-function mutants display the opposite phenotype of *lsy-6* loss of function mutants—they adopt a dual ASEL cell fate. Promoter tagging experiments suggest that *lsy-6* and *cog-1* are only co-expressed in one cell in the entire worm, the ASEL cell (Hobert, 2006). Transient *lsy-6* expression before the embryonic comma stage alone is also sufficient to direct the ASEL fate that is established at a later point in development (Zheng et al., 2011). Thus *lsy-6* can act as a switch on *cog-1*, possibly its only target in the entire animal since 13 other candidate targets were not repressed by *lsy-6*, as previously mentioned (Didiano and Hobert, 2006). A later study noted that regions outside of the *lsy-6* sites in the *cog-1* 3′ UTR are also required for repression (Didiano and Hobert, 2008). The layers of specificity in the *lsy-6:cog-1* interaction that continue to emerge with further study highlight that not all miRNA–target interactions will be as simple as just seed pairing.

References

Ahlquist, P. (2002). RNA-dependent RNA polymerases, viruses, and RNA silencing. Science *296*, 1270-1273.

Ambros, V. (2004). The functions of animal microRNAs. Nature *431*, 350-355.

Ameres, S.L., Horwich, M.D., Hung, J.H., Xu, J., Ghildiyal, M., Weng, Z., and Zamore, P.D. (2010). Target RNA-directed trimming and tailing of small silencing RNAs. Science *328*, 1534-1539.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T.*, et al.* (2006). A novel class of small RNAs bind to MILI protein in mouse testes. Nature *442*, 203-207.

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science *318*, 761-764.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev *22*, 2773-2785.

Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature *455*, 64-71.

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.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell *136*, 215-233.

Bartel, D.P., and Chen, C.Z. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet *5*, 396-400.

Baskerville, S., and Bartel, D.P. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA *11*, 241-247.

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S.*, et al.* (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in C. elegans. Mol Cell *31*, 67-78.

Behm-Ansmant, I., Rehwinkel, J., Doerks, T., Stark, A., Bork, P., and Izaurralde, E. (2006). mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. Genes Dev *20*, 1885-1898.

Berezikov, E., Chung, W.J., Willis, J., Cuppen, E., and Lai, E.C. (2007). Mammalian mirtron genes. Mol Cell *28*, 328-336.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363-366.

Boland, A., Tritschler, F., Heimstadt, S., Izaurralde, E., and Weichenrieder, O. (2010). Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. EMBO Rep *11*, 522-527.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol *3*, e85.

Broderick, J.A., Salomon, W.E., Ryder, S.P., Aronin, N., and Zamore, P.D. (2011). Argonaute protein identity and pairing geometry determine cooperativity in mammalian RNA silencing. RNA *17*, 1858-1869.

Burroughs, A.M., Ando, Y., de Hoon, M.J., Tomaru, Y., Suzuki, H., Hayashizaki, Y., and Daub, C.O. (2011). Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. RNA Biol *8*, 158-177.

Carroll, S.B. (2000). Endless forms: the evolution of gene regulation and morphological diversity. Cell *101*, 577-580.

Cazalla, D., Yario, T., and Steitz, J.A. (2010). Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. Science *328*, 1563-1566.

Chekulaeva, M., Mathys, H., Zipprich, J.T., Attig, J., Colic, M., Parker, R., and Filipowicz, W. (2011). miRNA repression involves GW182-mediated recruitment of CCR4-NOT through conserved W-containing motifs. Nat Struct Mol Biol *18*, 1218-1226.

Cheloufi, S., Dos Santos, C.O., Chong, M.M., and Hannon, G.J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature *465*, 584-589.

Chen, K., and Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. Nat Rev Genet *8*, 93-103.

Chendrimada, T.P., Gregory, R.I., Kumaraswamy, E., Norman, J., Cooch, N., Nishikura, K., and Shiekhattar, R. (2005). TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature *436*, 740-744.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479-486.

Chiang, H.R., Schoenfeld, L.W., Ruby, J.G., Auyeung, V.C., Spies, N., Baek, D., Johnston, W.K., Russ, C., Luo, S., Babiarz, J.E*., et al.* (2010). Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev *24*, 992-1009.

Cifuentes, D., Xue, H., Taylor, D.W., Patnode, H., Mishima, Y., Cheloufi, S., Ma, E., Mane, S., Hannon, G.J., Lawson, N.D*., et al.* (2010). A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science *328*, 1694-1698.

Claycomb, J.M., Batista, P.J., Pang, K.M., Gu, W., Vasale, J.J., van Wolfswinkel, J.C., Chaves, D.A., Shirayama, M., Mitani, S., Ketting, R.F*., et al.* (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell *139*, 123-134.

Cogoni, C., Irelan, J.T., Schumacher, M., Schmidhauser, T.J., Selker, E.U., and Macino, G. (1996). Transgene silencing of the al-1 gene in vegetative cells of Neurospora is mediated by a cytoplasmic effector and does not depend on DNA-DNA interactions or DNA methylation. EMBO J *15*, 3153-3163.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R*., et al.* (2008). An endogenous small interfering RNA pathway in Drosophila. Nature *453*, 798-802.

Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R*., et al.* (2008). Piwi and piRNAs act upstream of an endogenous siRNA pathway to suppress Tc3 transposon mobility in the Caenorhabditis elegans germline. Mol Cell *31*, 79-90.

Davis, E., Caiment, F., Tordoir, X., Cavaille, J., Ferguson-Smith, A., Cockett, N., Georges, M., and Charlier, C. (2005). RNAi-mediated allelic trans-interaction at the imprinted *Rtl1/Peg11* locus. Curr Biol *15*, 743-749.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature *432*, 231-235.

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.

Didiano, D., and Hobert, O. (2008). Molecular architecture of a miRNA-regulated 3' UTR. RNA *14*, 1297-1317.

Diederichs, S., and Haber, D.A. (2007). Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. Cell *131*, 1097-1108.

Diederichs, S., Jung, S., Rothenberg, S.M., Smolen, G.A., Mlody, B.G., and Haber, D.A. (2008). Coexpression of Argonaute-2 enhances RNA interference toward perfect match binding sites. Proc Natl Acad Sci U S A *105*, 9284-9289.

Doench, J.G., Petersen, C.P., and Sharp, P.A. (2003). siRNAs can function as miRNAs. Genes Dev *17*, 438-442.

Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. Genes Dev *18*, 504-511.

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R., and Bartel, D.P. (2009). RNAi in budding yeast. Science *326*, 544-550.

Ebert, M.S., Neilson, J.R., and Sharp, P.A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. Nat Methods *4*, 721-726.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001a). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev *15*, 188-200.

Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001b). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. Embo J *20*, 6877-6888.

Fabian, M.R., Cieplak, M.K., Frank, F., Morita, M., Green, J., Srikumar, T., Nagar, B., Yamamoto, T., Raught, B., Duchaine, T.F.*, et al.* (2011). miRNA-mediated deadenylation is orchestrated by GW182 through two conserved motifs that interact with CCR4-NOT. Nat Struct Mol Biol *18*, 1211-1217.

Fabian, M.R., Mathonnet, G., Sundermeier, T., Mathys, H., Zipprich, J.T., Svitkin, Y.V., Rivas, F., Jinek, M., Wohlschlegel, J., Doudna, J.A.*, et al.* (2009). Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. Mol Cell *35*, 868-880.

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.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.

Forstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., and Zamore, P.D. (2005). Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. PLoS Biol *3*, e236.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. Nature *465*, 818-822.

Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res *19*, 92-105.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z*., et al.* (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science *320*, 1077-1081.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat Rev Genet *10*, 94-108.

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.

Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature *442*, 199-202.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. (2004). The Microprocessor complex mediates the genesis of microRNAs. Nature *432*, 235-240.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91-105.

Grimson, A., Srivastava, M., Fahey, B., Woodcroft, B.J., Chiang, H.R., King, N., Degnan, B.M., Rokhsar, D.S., and Bartel, D.P. (2008). Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature *published on-line*.

Gu, S., Jin, L., Zhang, F., Sarnow, P., and Kay, M.A. (2009a). Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. Nat Struct Mol Biol *16*, 144-150.

Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J*., et al.* (2009b). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. Mol Cell *36*, 231-244.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835-840.

Guo, S., and Kemphues, K.J. (1995). par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. Cell *81*, 611-620.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M.*, et al.* (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129-141.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945-956.

Haley, B., and Zamore, P.D. (2004). Kinetic analysis of the RNAi enzyme complex. Nat Struct Mol Biol *11*, 599-606.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science *286*, 950-952.

Hammell, M., Long, D., Zhang, L., Lee, A., Carmack, C.S., Han, M., Ding, Y., and Ambros, V. (2008). mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. Nat Methods.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature *404*, 293-296.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev *18*, 3016-3027.

Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887-901.

Han, T., Manoharan, A.P., Harkins, T.T., Bouffard, P., Fitzpatrick, C., Chu, D.S., Thierry-Mieg, D., Thierry-Mieg, J., and Kim, J.K. (2009). 26G endo-siRNAs regulate spermatogenic and zygotic gene expression in Caenorhabditis elegans. Proc Natl Acad Sci U S A *106*, 18674-18679.

Hendrickson, D.G., Hogan, D.J., McCullough, H.L., Myers, J.W., Herschlag, D., Ferrell, J.E., and Brown, P.O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. PLoS Biol *7*, e1000238.

Hobert, O. (2006). Architecture of a microRNA-controlled gene regulatory network that diversifies neuronal cell fates. Cold Spring Harb Symp Quant Biol *71*, 181-188.

Hutvagner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056-2060.

Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol *3*, 318-356.

Jakymiw, A., Lian, S., Eystathioy, T., Li, S., Satoh, M., Hamel, J.C., Fritzler, M.J., and Chan, E.K. (2005). Disruption of GW bodies impairs mammalian RNA interference. Nat Cell Biol *7*, 1267-1274.

Johnston, M., and Hutvagner, G. (2011). Posttranslational modification of Argonautes and their role in small RNA-mediated gene regulation. Silence *2*, 5.

Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. Nature *426*, 845-849.

Karginov, F.V., Conaco, C., Xuan, Z., Schmidt, B.H., Parker, J.S., Mandel, G., and Hannon, G.J. (2007). A biochemical approach to identifying microRNA targets. Proc Natl Acad Sci U S A *104*, 19291-19296.

Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., Okada, T.N., Siomi, M.C., and Siomi, H. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature *453*, 793-797.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. Nat Genet *39*, 1278-1284.

Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev *15*, 2654-2659.

Khan, A.A., Betel, D., Miller, M.L., Sander, C., Leslie, C.S., and Marks, D.S. (2009). Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. Nat Biotechnol *27*, 549-555.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol *10*, 126-139.

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.

Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res *39*, D152-157.

Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M.*, et al.* (2005). Combinatorial microRNA target predictions. Nat Genet *37*, 495-500.

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.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853-858.

Lai, E.C. (2002). Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. Nat Genet *30*, 363-364.

Lai, E.C., Tam, B., and Rubin, G.M. (2005). Pervasive regulation of *Drosophila* Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. Genes Dev *19*, 1067-1080.

Lambert, N.J., Gu, S.G., and Zahler, A.M. (2011). The conformation of microRNA seed regions in native microRNPs is prearranged for presentation to mRNA targets. Nucleic Acids Res *39*, 4827-4835.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. Curr Biol *14*, 2162-2167.

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.

Lau, N.C., Seto, A.G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D.P., and Kingston, R.E. (2006). Characterization of the piRNA complex from rat testes. Science *313*, 363-367.

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science *294*, 862-864.

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.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S.*, et al.* (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.

Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J *21*, 4663-4670.

Leung, A.K., Calabrese, J.M., and Sharp, P.A. (2006). Quantitative analysis of Argonaute protein reveals microRNA-dependent localization to stress granules. Proc Natl Acad Sci U S A *103*, 18125-18130.

Leung, A.K., Young, A.G., Bhutkar, A., Zheng, G.X., Bosson, A.D., Nielsen, C.B., and Sharp, P.A. (2011). Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. Nat Struct Mol Biol *18*, 237-244.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. Nature *424*, 147-151.

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.

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.

Lim, L.P., Lau, N.C., Garrett-Engele, 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.

Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441.

Liu, J., Valencia-Sanchez, M.A., Hannon, G.J., and Parker, R. (2005). MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. Nat Cell Biol *7*, 719-723.

Long, D., Lee, R., Williams, P., Chan, C.Y., Ambros, V., and Ding, Y. (2007). Potent effect of target structure on microRNA function. Nat Struct Mol Biol *14*, 287-294.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95-98.

Ma, J.B., Ye, K., and Patel, D.J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature *429*, 318-322.

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.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science *311*, 195-198.

Makarova, K.S., Wolf, Y.I., van der Oost, J., and Koonin, E.V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct *4*, 29.

Mallory, A.C., and Vaucheret, H. (2006). Functions of microRNAs and related small RNAs in plants. Nat Genet *38 Suppl*, S31-36.

Marraffini, L.A., and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet *11*, 181-190.

Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563-574.

Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., and Zamore, P.D. (2005). Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell *123*, 607-620.

Mayr, C., Hemann, M.T., and Bartel, D.P. (2007). Disrupting the pairing between *let-7* and *Hmga2* enhances oncogenic transformation. Science *315*, 1576-1579.

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.

Meister, G., Landthaler, M., Peters, L., Chen, P.Y., Urlaub, H., Luhrmann, R., and Tuschl, T. (2005). Identification of novel argonaute-associated proteins. Curr Biol *15*, 2149-2155.

Miska, E.A., Alvarez-Saavedra, E., Abbott, A.L., Lau, N.C., Hellman, A.B., McGonagle, S.M., Bartel, D.P., Ambros, V.R., and Horvitz, H.R. (2007). Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. PLoS Genet *3*, e215.

Miyoshi, K., Tsukumo, H., Nagami, T., Siomi, H., and Siomi, M.C. (2005). Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev *19*, 2837-2848.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. Plant Cell *2*, 279-289.

Nykanen, A., Haley, B., and Zamore, P.D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell *107*, 309-321.

Okamura, K., Chung, W.J., Ruby, J.G., Guo, H., Bartel, D.P., and Lai, E.C. (2008). The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature *453*, 803-806.

Pak, J., and Fire, A. (2007). Distinct populations of primary and secondary effectors during RNAi in C. elegans. Science *315*, 241-244.

Park, C.Y., Choi, Y.S., and McManus, M.T. (2010). Analysis of microRNA knockouts in mice. Hum Mol Genet *19*, R169-175.

Parker, J.S., Parizotto, E.A., Wang, M., Roe, S.M., and Barford, D. (2009). Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. Mol Cell *33*, 204-214.

Parker, J.S., Roe, S.M., and Barford, D. (2004). Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J *23*, 4727-4737.

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.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P.*, et al.* (2000). Conservation of the sequence and temporal expression of *let-7* heterochronic regulatory RNA. Nature *408*, 86-89.

Petri, S., Dueck, A., Lehmann, G., Putz, N., Rudel, S., Kremmer, E., and Meister, G. (2011). Increased siRNA duplex stability correlates with reduced off-target and elevated on-target effects. RNA *17*, 737-749.

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.

Rand, T.A., Petersen, S., Du, F., and Wang, X. (2005). Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell *123*, 621-629.

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.

Reinhart, B.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.

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.

Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell *110*, 513-520.

Robins, H., Li, Y., and Padgett, R.W. (2005). Incorporating structure to predict microRNA targets. Proc Natl Acad Sci U S A *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.

Romano, N., and Macino, G. (1992). Quelling: transient inactivation of gene expression in Neurospora crassa by transformation with homologous sequences. Mol Microbiol *6*, 3343-3353.

Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. Cell *127*, 1193-1207.

Ruby, J.G., Jan, C.H., and Bartel, D.P. (2007a). Intronic microRNA precursors that bypass Drosha processing. Nature *448*, 83-86.

Ruby, J.G., Stark, A., Johnston, W.K., Kellis, M., Bartel, D.P., and Lai, E.C. (2007b). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. Genome Res *17*, 1850-1864.

Saetrom, P., Heale, B.S., Snove, O., Jr., 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.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199-208.

Selbach, M., Schwanhausser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. Nature *455*, 58-63.

Sen, G.L., and Blau, H.M. (2005). Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. Nat Cell Biol *7*, 633-636.

Shabalina, S.A., and Koonin, E.V. (2008). Origins and evolution of eukaryotic RNA interference. Trends Ecol Evol *23*, 578-587.

Shin, C., Nam, J.W., Farh, K.K., Chiang, H.R., Shkumatava, A., and Bartel, D.P. (2010). Expanding the microRNA targeting code: functional sites with centered pairing. Mol Cell *38*, 789-802.

Sijen, T., Steiner, F.A., Thijssen, K.L., and Plasterk, R.H. (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. Science *315*, 244-247.

Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. Mol Cell *5*, 659-669.

Song, J.J., Smith, S.K., Hannon, G.J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. Science *305*, 1434-1437.

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.

van der Krol, A.R., Mur, L.A., Beld, M., Mol, J.N., and Stuitje, A.R. (1990). Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. Plant Cell *2*, 291-299.

van Rooij, E., Sutherland, L.B., Qi, X., Richardson, J.A., Hill, J., and Olson, E.N. (2007). Control of stress-dependent cardiac growth and gene expression by a microRNA. Science *316*, 575-579.

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.

Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I., and Moazed, D. (2004). RNAi-mediated targeting of heterochromatin by the RITS complex. Science *303*, 672-676.

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I., and Martienssen, R.A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science *297*, 1833-1837.

Wightman, B., Burglin, T.R., Gatto, J., Arasu, P., and Ruvkun, G. (1991). Negative regulatory sequences in the *lin-14* 3'-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. Genes Dev *5*, 1813-1824.

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.

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.

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.

Zekri, L., Huntzinger, E., Heimstadt, S., and Izaurralde, E. (2009). The silencing domain of GW182 interacts with PABPC1 to promote translational repression and degradation of microRNA targets and is required for target release. Mol Cell Biol *29*, 6220-6231.

Zhao, Y., Samal, E., and Srivastava, D. (2005). Serum response factor regulates a muscle-specific microRNA that targets *Hand2* during cardiogenesis. Nature *436*, 214-220.

Zheng, G., Cochella, L., Liu, J., Hobert, O., and Li, W.H. (2011). Temporal and Spatial Regulation of MicroRNA Activity with Photoactivatable Cantimirs. ACS Chem Biol.

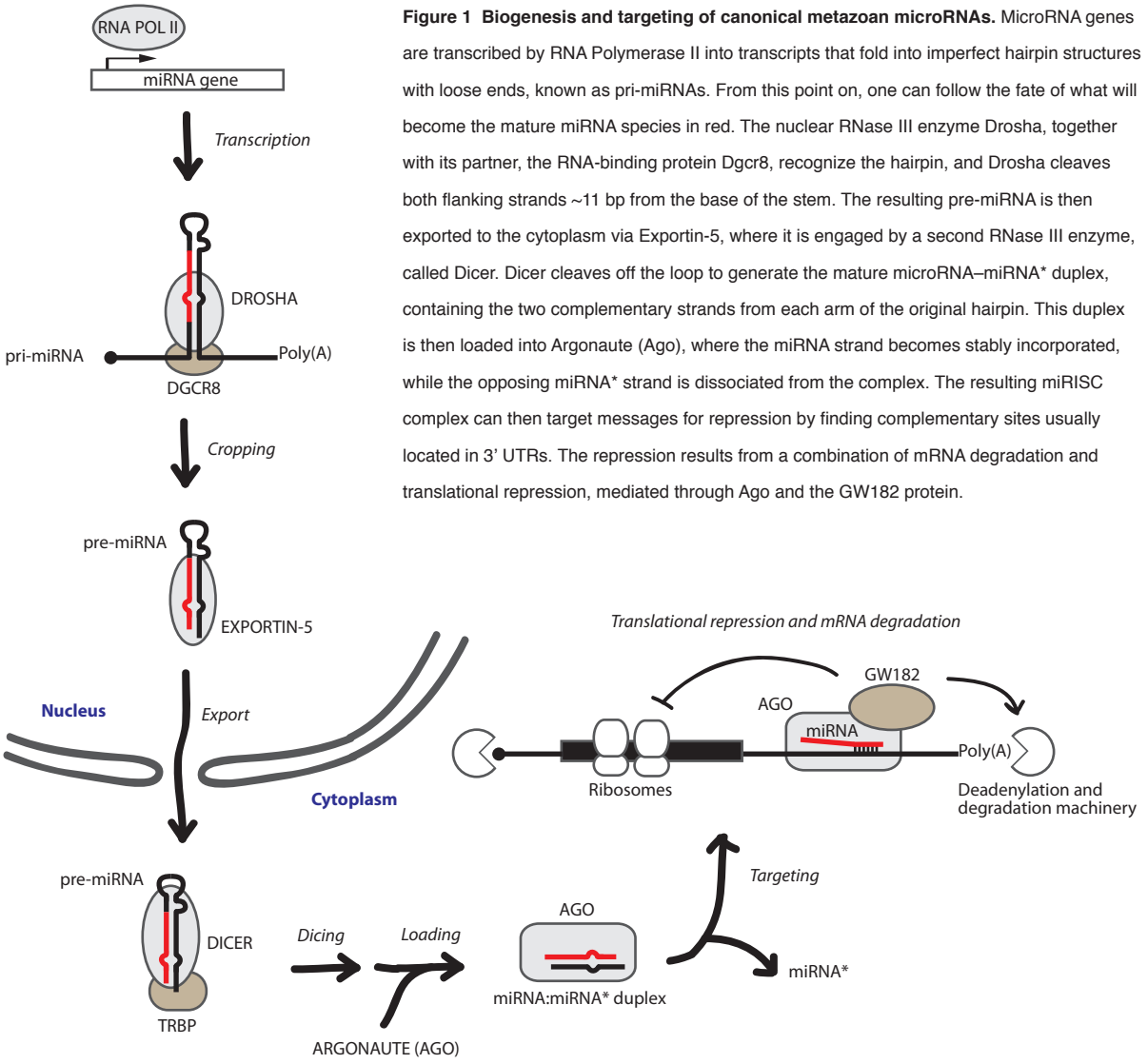


Figure 1 Biogenesis and targeting of canonical metazoan microRNAs. MicroRNA genes are transcribed by RNA Polymerase II into transcripts that fold into imperfect hairpin structures with loose ends, known as pri-miRNAs. From this point on, one can follow the fate of what will become the mature miRNA species in red. The nuclear RNase III enzyme Drosha, together with its partner, the RNA-binding protein Dgcr8, recognize the hairpin, and Drosha cleaves both flanking strands ~11 bp from the base of the stem. The resulting pre-miRNA is then exported to the cytoplasm via Exportin-5, where it is engaged by a second RNase III enzyme, called Dicer. Dicer cleaves off the loop to generate the mature microRNA–miRNA* duplex, containing the two complementary strands from each arm of the original hairpin. This duplex is then loaded into Argonaute (Ago), where the miRNA strand becomes stably incorporated, while the opposing miRNA* strand is dissociated from the complex. The resulting miRISC complex can then target messages for repression by finding complementary sites usually located in 3' UTRs. The repression results from a combination of mRNA degradation and translational repression, mediated through Ago and the GW182 protein.

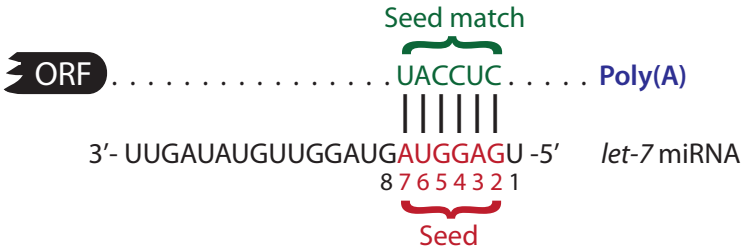


Figure 2 Canonical seed-matched site. Seed pairing between the *C. elegans let-7* miRNA and a complementary site in the 3' UTR of an mRNA target. The seed sequence spans nucleotides 2–7 of the miRNA.

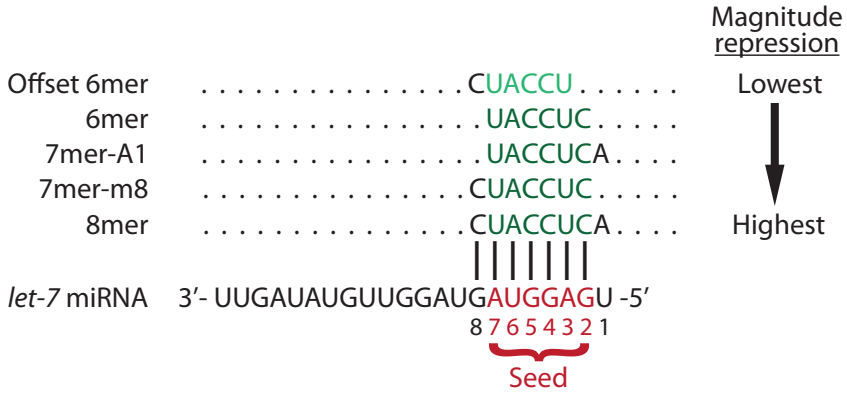


Figure 3 Types of miRNA target sites. Different degrees of pairing with the seed region yield different average levels of repression. Pairing with positions 2–7 (or positions 3–8) of the miRNA alone imparts only marginal repression (6mer, offset 6mer sites). Seed pairing plus an adenosine across from nucleotide 1 of the miRNA increases site efficacy (7mer-A1 site). Seed pairing plus an additional base pair with miRNA nucleotide 8 yields still greater repression (7mer-m8 site). Combining the features of the two 7mer site types confers the most repression (8mer site).

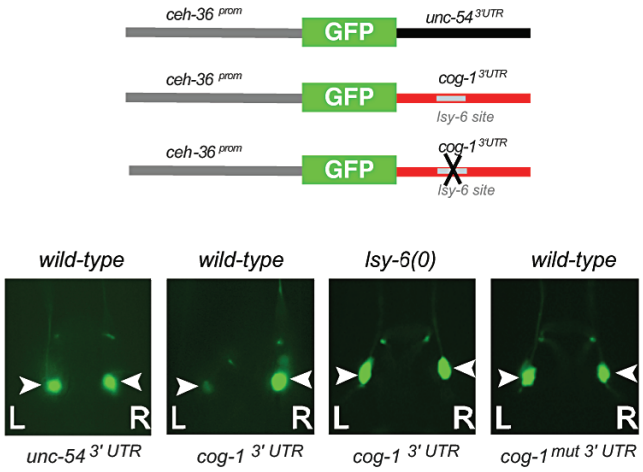


Figure 4 Down-regulation of the *cog-1* 3' UTR by *lsy-6* as monitored using a GFP sensor strategy (Johnston and Hobert, 2004; Didiano and Hobert, 2006). Expressing a GFP sensor containing the 3' UTR of *cog-1* in the ASE neurons of *C. elegans* demonstrates regulation by *lsy-6* specifically in the ASEL cell where the miRNA is expressed. Expressing a sensor containing the *cog-1* 3' UTR in which the *lsy-6* sites are mutated abolishes regulation, as does expressing a control UTR (*unc-54*). Down-regulation of the *cog-1* 3' UTR is dependent on *lsy-6* expression, as regulation is lost in a *lsy-6* null animal. Figure from (Hobert, 2006).

Chapter 2

Weak seed-pairing stability and high target-site abundance decrease the proficiency of *lsy-6* and other microRNAs

David M. Garcia[1–3,8], Daehyun Baek[1–5,8], Chanseok Shin[1–3,6], George W. Bell[1], Andrew Grimson[1–3,7] & David P. Bartel[1–3]

[1]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, USA. [2]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]School of Biological Sciences, Seoul National University, Seoul, Republic of Korea. [5]Bioinformatics Institute, Seoul National University, Seoul, Republic of Korea. [6]Department of Agricultural Biotechnology, Seoul National University, Seoul, Republic of Korea. [7]Present address: Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York, USA. [8]These authors contributed equally to this work.

D.M.G. performed most reporter assays and associated experiments and analyses. D.B. performed all the computational analyses except for reporter analyses. G.W.B. implemented revisions to the TargetScan site. C.S and A.G performed assays and analyses involving miR-23. D.M.G, D.B. and D.P.B wrote the paper.

Published as: Garcia, D.M., Baek, D., Shin, C., Bell, G.W., Grimson, A., and Bartel, D.P. *Nature Structural and Molecular Biology*, Volume 18, Number 10, pages 1139–1146, October 2011.

Abstract

Most metazoan microRNAs (miRNAs) target many genes for repression, but the nematode *lsy-6* miRNA is much less proficient. Here we show that the low proficiency of *lsy-6* can be recapitulated in HeLa cells and that miR-23, a mammalian miRNA, also has low proficiency in these cells. Reporter results and array data indicate two properties of these miRNAs that impart low proficiency: their weak predicted seed-pairing stability (SPS) and their high target-site abundance (TA). These two properties also explain differential propensities of small interfering RNAs (siRNAs) to repress unintended targets. Using these insights, we expand the TargetScan tool for quantitatively predicting miRNA regulation (and siRNA off-targeting) to model differential miRNA (and siRNA) proficiencies, thereby improving prediction performance. We propose that siRNAs designed to have both weaker SPS and higher TA will have fewer off-targets without compromised on-target activity.

Introduction

MicroRNAs are ~22-nucleotide (nt) RNAs that pair with the messages of protein-coding genes to direct post-transcriptional repression of these target mRNAs[1,2]. In animals, many studies using a wide range of methods, including comparative sequence analysis, site-directed mutagenesis, genetics, mRNA profiling, coimmunoprecipitation and proteomics, have shown that perfect pairing with miRNA nucleotides 2–7, known as the miRNA seed, is important for the recognition of many miRNA targets[3]. To impart more than marginal repression of mammalian targets, this seed pairing is usually augmented by either a match with miRNA nucleotide 8 (7-mer-m8 site)[4–7], an A across from nucleotide 1 (7-mer-A1 site)[4,7] or both (8-mer site)[4,7]. In rare

instances, targeting also occurs through 3′-compensatory sites[4,5,8] and centered sites[9], for which substantial pairing outside the seed region compensates for imperfect seed pairing.

A single miRNA can target hundreds of distinct mRNAs through seed-matched sites[10]. Indeed, most human mRNAs are conserved regulatory targets[8], and many additional regulatory interactions occur through nonconserved sites[11–13]. However, not every site is effective; 8-nt sites are effective more often than 7-nt sites, which are effective more often than 6-nt sites[7,14]. Another factor is site context. For example, sites in the 3′ untranslated regions (3′ UTRs) are effective more often than those in the path of the ribosome[7]. Among 3′ UTR sites, those away from the centers of long UTRs and those within high local A-U sequence context are effective more often[7], consistent with reports that sites predicted to be within more accessible secondary structure tend to be more effective[15–19]. Site efficacy is also influenced by proximity to other miRNA-binding sites[7,20], to protein-binding sites[21] and to sequences that can pair with the 3′ region of the miRNA, particularly nucleotides 13–17 (ref. 7).

Studies of site efficacy have focused primarily on different sites for the same miRNA, without systematic investigation of whether some miRNA sequences are more proficient at targeting than others. Broadly conserved miRNAs typically have many more conserved targeting interactions than do other miRNAs[4,8], and highly or broadly expressed miRNAs seem to target more mRNAs than do others[22], but these phenomena reflect evolutionary happenstance more than intrinsic targeting proficiency.

Our interest in targeting proficiency was spurred by results regarding the *lsy-6* miRNA. When tested in *Caenorhabditis elegans*, only 1 of 14 predicted targets with 7- to 8-nt seed matched sites responds to *lsy-6*, which was interpreted to show that perfect seed pairing is not a

reliable predictor for miRNA-target interactions[23]. Alternatively, and in keeping with findings for many other miRNAs3, the results for *lsy-6* might not apply to other miRNAs because *lsy-6* could have unusually high targeting specificity owing to unusually low targeting proficiency. A similar rationale might explain results for mammalian miR-23, another miRNA that confers unusually weak responses from most reporters designed to test predicted targets.

When considering properties that might confer a low targeting proficiency, we noted that both *lsy-6* and miR-23 have unusually (A+U)-rich seed regions, which could lower the stability of seed-pairing interactions. Perhaps a threshold of SPS is required for the miRNA to remain associated with targets long enough to achieve widespread seed-based targeting. Indeed, predicted SPS is correlated with the propensity of siRNAs to repress unintended targets[24], a process called "off-targeting," which occurs through the same seed-based recognition as that for endogenous miRNA targeting[10]. Potentially confounding this interpretation, however, miRNAs with (A+U)-rich seed regions have more 3′ UTR–binding sites, a consequence of the (A+U)-rich nucleotide composition of 3′ UTRs, which could dilute the effect on each target message. Indeed, TA can be manipulated to titrate miRNAs away from their normal targets[25,26], and natural TA has been proposed to influence miRNA targeting and siRNA off-targeting[27,28], although these reported TA effects have not been fully disentangled from potential SPS effects. Here, we find that both SPS and TA have a substantial impact on targeting proficiency, and apply these insights to improve miRNA target predictions.

Results

lsy-6 targeting specificity is recapitulated in HeLa cells

lsy-6 targeting was originally examined in a *C. elegans* neuron[23], whereas more proficient targeting by other miRNAs has been experimentally demonstrated in other systems, sometimes in vertebrate tissues or primary cells[11,13,29,30] but more often in cell lines[3]. To test whether differences in targeting proficiency could be attributed to the different biological contexts in which the miRNAs had been examined, we ported the 14 3′ UTRs tested in *C. elegans* into a luciferase reporter system typically used in mammalian cell lines and introduced the *lsy-6* miRNA by co-transfecting an imperfect RNA duplex representing the miRNA (Fig. 1a) and the short RNA from the other arm of the hairpin, known as the miRNA* (Supplementary Fig. 1a). As has been observed in worms[23], only the *cog-1* 3′ UTR responded in HeLa cells (Fig. 1b). Repression was lost when a control miRNA (miR-1) replaced *lsy-6* or when the two *cog-1* sites were mutated, introducing either mismatches (Fig. 1b) or G•U wobbles (Supplementary Fig. 1b,c).

Each of the 14 3′ UTRs had at least one canonical 7- to 8-nt *lsy-6* site, and 11 UTRs had a site conserved in three sequenced nematodes (Supplementary Table 1). When evaluated using the context-score model, some sites had scores comparable to those of sites that mediate repression in this assay[7] (Supplementary Table 1). Moreover, the C27H6.9 3′ UTR had two 8-mer sites with scores matching those of the two *cog-1* sites. The close match between the results in our heterologous reporter assay and previous results in *C. elegans* neurons indicated that the specificity for targeting the *cog-1* 3′ UTR did not require the endogenous cellular context of *lsy-6* repression; it was operable in HeLa cell culture and thereby attributable to the intrinsic properties

of *lsy-6* and its targets. This result also indicated that these properties could be investigated in mammalian cell culture, which is easier than using stable reporter lines in worms.

Modifying both SPS and TA elevates targeting proficiency

As expected for a miRNA with sequence UUUGUAU at nucleotides 2–8, the calculated free energy ($\Delta G°$) of the predicted SPS for the *lsy-6* 8-mer or 7-mer-m8 sites (both 7 base pairs, bp) was weak (-3.65 kcal mol^{-1}), which was weaker than that of all but one conserved nematode miRNA (Fig. 1c). The SPS predicted for *lsy-6* was also weaker than that of the weakest of 87 broadly conserved vertebrate miRNAs (Fig. 1d). The predicted $\Delta G°$ of an 8-mer or 7-mer-m8 seed match for miR-23 was -5.85 kcal mol^{-1}, in the bottom quintile for broadly conserved vertebrate miRNAs (Fig. 1d). We observed similar results for 7-mer-A1 or 6-mer sites (both 6 bp) for both miRNAs (Supplementary Fig. 1d,e).

lsy-6 is also at the extreme end of the distribution of TA for miRNAs in nematodes and human (Fig. 1e,f). To predict the TA in a genome, we counted the number of sites in a curated set of distinct 3′ UTRs. When considering a particular cell type, we converted the genome TA to a transcriptome TA by considering the relative levels of each mRNA bearing a site, although in practice the genome and transcriptome TA levels were highly correlated. For example, the transcriptome TA for HeLa cells (TA_{HeLa}) was correlated nearly exactly with the genome TA ($R^2 = 0.98$, $P < 10^{-100}$, Spearman's correlation test, Supplementary Fig. 1f). For 8-mer and 7-mer-m8 sites (which both pair with nucleotides 2–8), *lsy-6* had a genome TA that ranked second among 60 *C. elegans* miRNA families and a TA_{HeLa} near that of miR-23, which ranks fifth among the 87 vertebrate families (Fig. 1e and Supplementary Fig. 1g).

To test the hypothesis that either the weak SPS or high TA of *lsy-6* influences its targeting proficiency, we made three substitutions in the *lsy-6* seed that changed both properties. The three substitutions converted the *lsy-6* seed to that of miR-142-3p (Fig. 1a and Supplementary Fig. 1a), which changed the predicted SPS to −7.70 kcal mol^{-1}, which was 4.05 kcal mol^{-1} stronger than that of *lsy-6* and near the median values for conserved nematode and vertebrate miRNAs (Fig. 1c,d). The substitutions also changed the predicted TA to $10^{2.957}$ sites in *C. elegans* and $10^{3.207}$ sites in human, values below the median of conserved miRNAs in both genomes (Fig. 1e,f). We co-transfected this miR-142*lsy-6* chimeric miRNA and assayed it using reporters with compensatory substitutions in their seed matches, and found it repressed 9 of 14 reporters, a fraction within the range expected in this system using reporters with the site types and contexts assayed (Fig. 1g). We repeated the experiment using the full-length miR-142-3p sequence (Fig. 1a and Supplementary Fig. 1a) and found similar results, indicating that miRNA sequence outside the seed region was irrelevant for repression of both the *cog-1* 3′ UTR and the other *C. elegans* 3′ UTRs (Fig. 1h).

Like *lsy-6*, miR-23 also had low targeting proficiency in our system. We surveyed 17 human 3′ UTR fragments, randomly chosen from a set with two 7- to 8-nt miR-23 sites (conserved or nonconserved) spaced within 700 nt of each other, and found that only one fragment was repressed by miR-23 endogenous to either HeLa or HepG2 cells (data not shown). In subsequent experiments focusing on the six UTRs with the most favorable context scores (Supplementary Table 1), we found that co-transfecting additional miR-23a imparted marginal or no repression (Fig. 1i).

To test whether strengthening SPS while decreasing TA could increase the targeting proficiency of miR-23a, we converted two A:U seed pairs into two G:C pairs (Fig. 1a and Supplementary Fig. 1a); this strengthened the predicted SPS from −5.85 kcal mol^{-1} to −8.67 kcal mol^{-1} while reducing the TA from the fifth highest of the 87 vertebrate families to below the lowest. We assayed this miRNA, called miR-CGCG, using reporters with compensatory substitutions in their seed matches, and found that the sporadic and marginal repression observed with the wild-type UTRs became much more robust (Fig. 1j). These results indicate that miR-23a had low targeting proficiency because of its weak SPS, its high TA, or both, thereby extending our findings to a mammalian miRNA and mammalian 3′ UTRs.

Separating the effects of SPS and TA on miRNA targeting

To differentiate the potential effects of SPS from those of TA, we considered the relationship between these two properties for all 16,384 possible heptamers. In the *C. elegans* 3′ UTRs, these properties were highly anticorrelated (Fig. 2a, $R^2 = 0.680$, $P < 10^{-100}$, Spearman's correlation test). In mammalian 3′ UTRs the relationship was still highly significant, but the substantial depletion of CG dinucleotides in the vertebrate transcriptome[31] created more spread in TA, which led to lower correlation coefficients for both human (Fig. 2b, $R^2 = 0.121$, $P < 10^{-100}$) and mouse (Supplementary Fig. 2a, $R^2 = 0.081$, $P < 10^{-100}$). In general, each additional CG dinucleotide imparted an additional $\log_{10}$ reduction in TA.

To test the influence of TA on *lsy-6* targeting proficiency, we designed the low-TA (LTA) version of *lsy-6*, which had two point substitutions in the *lsy-6* seed (Fig. 2c and Supplementary Fig. 1a). Substituting U4 with a C (substitution U4C) introduced a CG dinucle-

otide, whereas the other substitution, U2A, facilitated later investigation of SPS. Because of the CG dinucleotide, LTA-*lsy-6* had a predicted TA_{HeLa} 95% lower than that of *lsy-6*, a value that would be third lowest among the conserved vertebrate miRNA families. Although the substitutions also led to stronger SPS, the predicted SPS of −5.49 kcal mol^{-1} was still slightly weaker than that of miR-23 and well below the median for both nematode and vertebrate conserved miRNAs (Fig. 1c,d). When assayed using reporters with compensatory substitutions in their seed matches, LTA-*lsy-6* repressed the *cog-1* reporters and only three others (Fig. 2d). Two reporters (F55G1.12 and C27H6.9) were repressed only marginally (<1.3 fold), reminiscent of the marginal repression imparted by miR-23 when using its cognate sites. For the third reporter, T20G5.9, we attributed much of the apparent repression to normalization to the miR-1 results, which in the case of this UTR were unusual (Supplementary Fig. 2d). Taken together, the LTA-*lsy-6* results indicate that lowering TA was not sufficient alone to confer robust targeting proficiency.

To strengthen SPS without changing TA, we replaced each of the two seed adenines of LTA-*lsy-6* with 2,6-di-aminopurine (DAP or D). DAP is an adenine analog with an exocyclic amino group at position 2, enabling it to pair with uracil with geometry and thermodynamic stability resembling that of a G:C pair (Fig. 2e). Because nearest-neighbor parameters had not been determined for model duplexes containing D:U pairs, we estimated SPS by using the values for A:U pairs and adding −0.9 kcal mol^{-1} for each D:U pair, as this is the value of an additional hydrogen bond in model duplexes[32]. With this approximation, the D-LTA-*lsy-6* miRNA had a predicted SPS of −7.29 kcal mol^{-1}, which approached −7.87 kcal mol^{-1}, the median predicted SPS of the conserved vertebrate miRNAs. When assayed using the same reporters as used for

LTA-*lsy-6*, D-LTA-*lsy-6* repressed 7 of 14 reporters (Fig. 2f). Although this repression was weaker than that observed with the miR-142 seed (Fig. 1g,h), it was greater than that observed for LTA-*lsy-6* and on par with that expected for mammalian miRNAs in this system using reporters with the site types and site contexts assayed.

We next tested D-miR-23, which also had two seed adenines replaced by DAP, thereby strengthening the predicted SPS from −5.85 kcal mol^{-1} to −7.65 kcal mol^{-1}. Five of the six reporters with miR-23 sites showed significantly greater repression by D-miR-23a than by wild-type miR-23a (Fig. 2g), demonstrating a favorable effect for increasing SPS in the context of very high TA (93rd percentile). However, repression was still considerably lower than that conferred by miR-CGCG, presumably because miR-CGCG had lower TA and somewhat stronger SPS (−8.67 kcal mol^{-1}), although we cannot exclude the possibility that the non-natural DAP in the miRNA compromised activity.

The results for DAP-substituted miRNAs show that for miRNAs with weak SPS, strengthening SPS can enhance targeting proficiency, regardless of whether these miRNAs have high or low TA. Because DAP substitution changed the predicted SPS without changing the sites in the UTRs, these results indicate that the low proficiency was due to weak SPS rather than occlusion of the sites by RNA-binding proteins that recognized the miRNA seed matches. Taken together, our reporter results also suggest that lowering TA can further enhance targeting proficiency, particularly for miRNAs with moderate to strong SPS.

Global impact of TA and SPS on targeting proficiency

To examine the global impact of TA and SPS on targeting, we collected 175 published

microarray data sets that monitored the response of transfecting miRNAs or siRNAs (together referred to as sRNAs) into HeLa cells (Supplementary Data 1). Data sets reporting the effects of sRNAs with the same seed region were combined, yielding results for 102 distinct seeds that covered a broad spectrum of TA and predicted SPS (Fig. 3a). For each of these 102 data sets, we determined the mean repression of mRNAs with a single 3′ UTR 8-mer site and no other sites in the message, and plotted these values with respect to both the TA_{HeLa} and predicted SPS of the transfected sRNA (Fig. 3b, top). sRNAs with lower TA_{HeLa} were more effective than those with higher TA_{HeLa}, and those with stronger predicted SPS were more effective than those with weaker predicted SPS ($P = 0.0006$ and 0.0054 for TA_{HeLa} and SPS, respectively, Pearson's correlation test; Table 1). We used multiple linear regression to account for the cross-correlation between TA_{HeLa} and SPS and found that correlations were at least marginally significant for the individual features ($P = 0.005$ and 0.05, *t*-test; Table 1), indicating that both properties were independently associated with the proficiency of targeting 3′ UTR sites. We observed similar results for targeting 7-mer-m8, 7-mer-A1 and 6-mer sites (Fig. 3b and Table 1).

Although both TA and SPS each significantly influenced targeting proficiency, together they explained only a minority of the variability (Table 1). Most of the variability could be from factors unrelated to targeting, such as array noise, differential transfection efficiencies or differential sRNA loading or stability. To reduce variability from these sources, we focused on 74 data sets for which responsive messages were significantly enriched in 3′ UTR sites to the transfected sRNA (Fig. 3a, red squares; Supplementary Data 1). In these filtered data sets, correlations between proficiency and both TA_{HeLa} and SPS were stronger and observed with similar significance, even though the filtering reduced the quantity of data analyzed and might

have preferentially discarded data sets for which high TA or weak SPS prevented detectable repression (Supplementary Fig. 3a,b and Supplementary Table 2).

Studies monitoring global effects of miRNAs on target repression have concluded that sites in open reading frames (ORFs) can mediate repression but that the efficacy of these sites is generally less than that of sites in 3′ UTRs[7,30,33,34]. To examine the impact of TA and SPS on targeting in ORFs, we considered expressed messages that had a single ORF site but no additional sites in the rest of the message. For 7-mer-m8 and 6-mer sites, mean repression was significantly correlated with both TA_{HeLa} and predicted SPS, and for the other two sites in ORFs, mean repression was significantly correlated with TA_{HeLa} (Fig. 3c and Table 1). The response of sites in 5′ UTRs was not significantly correlated with either TA or predicted SPS (Table 1), consistent with the idea that 5′ UTRs harbor few effective sites[3].

We next examined the quantitative impact of TA and SPS on targeting proficiency. We considered the same sets of mRNAs with single sites to the cognate sRNAs, and for each site type and each mRNA region, we binned mRNAs into quartiles ranked by either low TA or strong predicted SPS. For each site type, messages in the top quartile responded more strongly than those in the bottom (Fig. 3d). The differences usually were substantial. For example, repression of the top quartile of mRNAs with 7-mer-A1 sites matched the mean repression of mRNAs with 7-mer-m8 sites, whereas repression of the bottom quartile resembled the mean repression of mRNAs with 6-mer sites.

Improved miRNA target prediction

An effective tool for mammalian miRNA target prediction is the context score[30]. Context scores

are used to rank mammalian miRNA target predictions by modeling the relative contributions of previously identified targeting features, including site type, site number, site location, local A+U content and 3′-supplementary pairing, to predict the relative repression of mRNAs with 3′ UTR sites[7]. However, the context-score model was not designed to consider differences between sRNAs, such as TA or SPS, which can cause sites of one miRNA to be more robustly targeted than those of another (assuming equal expression of the two miRNAs).

To build a model appropriate for predicting the relative response of targets of different miRNAs, we considered TA and SPS as two independent variables when carrying out multiple linear regression on the 11 microarray data sets used previously for the initial development and training of the context-score model[7]. The other parameters were local A+U content, the location of the site within the 3′ UTR, and 3′-supplementary pairing[7]. For each site type, TA and/or SPS robustly contributed (Supplementary Table 3). The scores generated by these models were called context+ scores, because they consider site type and context plus sRNA proficiency. We then generated the total context+ score for each mRNA with 3′ UTR sites, relying on the observation that multiple sites typically act independently with respect to each other[7].

We tested the predictive value of the new model using data from array data sets not used to train the model, and comparing the performance of the predicted targets ranked using the total context+ scores to those ranked using scores of the original model. To examine whether any improvement over the original model was due to training the model with multiple linear regression rather than simple linear regression, we also used multiple linear regression to build a model that considered only the three parameters used to build the original model (context-only scores, Supplementary Table 4). For each model, we ranked predicted targets with 7- to 8-nt sites

by score and assigned them to ten bins. The context+ scores performed better than the old context scores at predicting the response to the sRNAs (Fig. 4a), yielding significantly stronger mean repression for the top two bins ($P = 5 \times 10^{-56}$ and 3×10^{-8} for bins 1 and 2, respectively) and significantly weaker repression in the bottom four bins ($P = 6 \times 10^{-10}$, 1.5×10^{-5}, 1×10^{-7} and 3×10^{-4} for bins 7–10, respectively, Wilcoxon's rank sum test). Improved specificity was also demonstrated in receiver operating characteristic (ROC) curves (Supplementary Fig. 4a).

Because most 6-mer sites and ORF sites are either nonresponsive or only marginally responsive to the miRNA, algorithms that achieve useful prediction specificity do so at the expense of ignoring these sites[3]. As low TA and strong SPS were correlated with substantially greater efficacy of these marginal sites (Fig. 3c,d), we extended the context+ scores to 6-mer sites. For the context+ model, the top bin of mRNAs with 6-mer 3′ UTR sites but no larger sites (Fig. 4b) had average repression resembling that of the third bin of mRNAs with 7- to 8-nt 3′ UTR sites (Fig. 4a; ROC curves, Supplementary Fig. 4b). We also generated context-only and context+ scores for ORF sites by changing only the parameter of site location; this was not applicable for ORF sites because it accounts for the lower efficacy of sites near the middle of long 3′ UTRs[7]. In ORFs, we found that sites farther from the stop codon tended to be less effective, and thus we included the distance from the stop codon (linearly scaled distance of 0 to ≥1,500 nt) as a parameter. Although this context+ model was not substantially more predictive than the context-only model for ORF sites (perhaps because data from only 11 miRNAs were used in the regression), both models had predictive value. We compared mRNAs with at least one 8-mer ORF site (Fig. 4c) and found that those ranked in the top bin had average repression resembling that of the second or third bins of mRNAs with 7- to 8-nt 3′ UTR sites (Fig. 4a).

Overall, our findings show that taking TA and SPS into account can significantly improve miRNA target prediction when pooling results from multiple sRNAs. Training on the 11 miRNA transfection data sets used for the original context scores was appropriate for demonstrating the improvement that could be achieved by taking TA and SPS into account. We reasoned, however, that training on the 74 filtered data sets could generate a more precise context+ model to be used to quantitatively predict repression. As we expected, correlations for all four parameters had even greater significance when we trained the model on more data (Supplementary Table 5). Although a support vector machine (SVM) approach should in principle yield even greater specificity by capturing effects lost in multiple linear regression due to multicollinearity, we did not observe enhanced performance with SVM (Supplementary Fig. 4c–e). Therefore, we used multiple linear regression because it enabled more convenient calculation of context+ scores (Supplementary Fig. 5a). We will use these new scores in version 6.0 of TargetScan (http://www.targetscan.org/).

Additional considerations

A caveat of the reporter experiments was that miRNA sequence changes designed to alter TA or SPS could have influenced other factors, such as miRNA stability or its loading into the silencing complex. However, our computational analyses of 102 array data sets also showed that TA and SPS each independently influence targeting efficacy. Therefore, if differences in sRNA stability or loading confounded interpretation of our results, these differences would be correlated with either predicted SPS or TA. Analysis of published miRNA overexpression data countered this possibility, showing no correlation between miRNA accumulation and predicted SPS or TA

(Supplementary Fig. 3c,d). Furthermore, experiments examining the RNAs co-purifying with AGO2 indicated that the difference in proficiency between *lsy-6* and miR-142*lsy-6* was not merely attributable to less accumulation of *lsy-6* in the silencing complex (Supplementary Fig. 1m–s).

Discussion

The correlation between strong SPS and low TA has confounded earlier efforts to examine the influence of these parameters on targeting efficacy, with one study implicating SPS and not TA[24] and others implicating TA and not SPS[27,28]. Our results indicate that both parameters influence efficacy and solve one of the mysteries in miRNA targeting, the failure of *lsy-6* to repress all but one of the 14 examined seed-matched mRNAs. Previous studies have hypothesized that the seed-based targeting model is unreliable[23] or that sites of the 13 nonresponsive mRNAs fall in inaccessible UTR structure[18]. Our work shows that the solution is the unusually weak SPS and high TA of the *lsy-6* miRNA. Changing these parameters to resemble those of more typical miRNAs imparted typical seed-based targeting proficiency, even though the sites were in their original UTR contexts, thereby demonstrating that neither the reliability of seed-based targeting nor the accessibility of the sites were at issue.

MicroRNAs with unusually weak predicted SPS and unusually high TA, such as miR-23 and *lsy-6*, seem to have few targets. Indeed, *lsy-6* might have only a single biological target, the *cog-1* mRNA—an extreme exception to the finding that metazoan miRNAs generally have dozens if not hundreds of preferentially conserved targets[4,8,35,36]. Determining why so few mRNAs respond to *lsy-6* brings to the fore a second mystery, still unsolved: how is the *cog-1* 3′

UTR so efficiently recognized and repressed by a miRNA with such weak targeting proficiency? This UTR has two 8-mer sites, which by virtue of their conservation make *cog-1* the top predicted target of *lsy-6* (ref. 3), but this is only part of the answer[37]. Improving the context-score model to take into account the differential SPS and TA of different miRNAs may help focus attention on the predicted targets of miRNAs with more typical proficiencies, but leaves unsolved the problem of how to predict the few biological sites of the less proficient miRNAs without considering site conservation.

MicroRNAs with very high TA, such as *lsy-6* or miR-23, and those with very low TA, such as miR-100 or miR-126, two broadly conserved vertebrate miRNAs containing CG dinucleotides in their seeds (Supplementary Data 2), seem to represent two strategies for targeting very few genes, accomplished at opposite ends of the TA spectrum. For miRNAs with very high TA, other UTR features flanking the seed sites are required for regulation, as has been shown for *lsy-6* regulation of *cog-1* (ref. 37), whereas miRNAs with very low TA have far fewer potential target sites to begin with.

Our results also have implications for how siRNA could be designed to reduce off-targets. Earlier studies have proposed that off-targets could be reduced by designing siRNAs with low TA[27] or weak SPS[24], and our results suggest that off-targets could be largely eliminated by designing siRNAs with both high TA and weak SPS. However, such siRNAs might also be ineffective at recognizing the desired mRNA target because pairing with this target would nucleate on a match with weak SPS and might be titrated by the many other mRNAs with seed matches. To investigate this concern, we examined a published data set of high-throughput luciferase assays reporting the response to 2,431 different siRNAs[38]. siRNAs with weak

predicted SPS knocked down the desired target more effectively than did those with strong predicted SPS (Fig. 4d; $P < 10^{-100}$, *t*-test), presumably because of preferential loading into the silencing complex[39,40]. Moreover, high TA did not compromise the desired targeting efficacy, even after we corrected for the cross-correlation between TA and SPS ($P = 0.16$, *t*-test). Therefore, designing siRNAs with high TA and weak SPS should minimize off-target effects without compromising knockdown of the desired target.

Highly expressed mRNAs tend to be evolutionarily depleted in sites for coexpressed miRNAs, a phenomenon partly attributed to the possibility that these mRNAs might otherwise titrate the miRNAs from their intended targets[12,41,42]. Titration can also provide a useful mechanism for cells to regulate miRNA activity, as has been shown by *IPS1* titration of miR-399 in *Arabidopsis thaliana*[25]. Beneficial titration has even been proposed to explain why so many miRNA sites are conserved[43]. However, because most preferentially conserved sites are in lowly to moderately expressed mRNAs, and because these sites each comprise only a tiny fraction of the TA, each could impart at most a correspondingly tiny effect on the effective miRNA concentration—much less than that required to selectively retain the site. Although titration functions cannot explain most site conservation, TA could be dynamic during development, with notable consequences. For example, the increase of a miRNA during development is often accompanied by a decrease in its transcriptome TA, a consequence of the evolutionary depletion of sites in mRNAs coexpressed at high levels with the miRNA[12,42]. This accompanying TA decrease would sharpen the transition between the nonrepressed and repressed states of targets.

When predicting SPS, we used parameters derived from model RNA duplexes, which presumably underestimated the affinity of RNA segments pairing with Argonaute-bound seed

regions[2,3,44,45]. The extent to which Argonaute enhances affinity might vary for different seed sequences. These potential differences, however, did not obscure our detection of an influence of SPS on targeting proficiency. Thus, our study provides a lower bound on the influence of SPS, and an approach for determining its full magnitude once accurate SPSs of Argonaute-bound complexes are known.

Methods

Reporter assays

For *lsy-6* reporter assays, HeLa cells were plated in 24-well plates at 5 x 10^4 cells per well. After 24 h, each well was transfected with 20 ng TK-*Renilla*-luciferase reporter (pIS1)[46], 20 ng firefly-luciferase control reporter (pIS0)[46] and 25 nM miRNA duplex (Dharmacon; Supplementary Fig. 1a), using Lipofectamine 2000 (Invitrogen). For miR-23 reporter assays, conditions were the same except for transfected DNA: 10 ng SV40-*Renilla*-luciferase reporter (pIS2)[46], 25 ng firefly-luciferase control reporter (pIS0) and 1.25 μg pUC19 carrier DNA. Luciferase activities were measured 24 h after transfection with the Dual-Luciferase Assay (Promega) and a Veritas microplate luminometer (Turner BioSystems). For every construct assayed, four independent experiments, each with three biological replicates, were done. To control for transfection efficiency, firefly activity was divided by *Renilla* activity. Values for constructs with sites matching the cognate miRNA were then normalized to the geometric mean of values for otherwise identical constructs in which the sites were mutated. To control for differences not attributable to the cognate miRNA, the ratios were further normalized to ratios for the same

constructs tested with a noncognate miRNA, miR-1. These double-normalized results are in figures; singly normalized results are in Supplementary Figures 1h–l and 2d–f.

Constructs

3′ UTRs of *lsy-6* predicted targets23 were subcloned into XbaI and EagI sites in pIS1, and 3′ UTRs of miR-23 predicated targets were cloned into SacI and SpeI sites in pIS2 after amplification (UTR sequences, Supplementary Table 1). Mutations were introduced using Quikchange (Stratagene) and confirmed by sequencing.

Predicted SPS

SPS was predicted using nearest-neighbor thermodynamic parameters, including the penalty for terminal A:U pairs[32]. The contribution of the A at position 1 of 8-mer and 7-mer-A1 sites was not included because this A does not pair with the miRNA[4] and thus its contribution is not expected to differ predictably for different miRNAs. For linear regression analyses, the predicted SPS of positions 2–8 was used for 8-mer and 7-mer-m8 sites, and the predicted SPS of positions 2–7 was used for 7-mer-A1 and 6-mer sites. To assign a single value for 7- to 8-nt sites (7-mer-A1, 7-mer-m8 and 8-mer), we used a mean weighted value of the three site types. This mean SPS was calculated as [(6-mer SPS)(7-mer-A1 TA) + (7-mer-m8 SPS)(7-mer-m8 TA + 8-mer TA)] / (7-mer-A1 TA + 7-mer-m8 TA + 8-mer TA).

Reference mRNAs

To generate a list of unique mRNAs, human full-length mRNAs obtained from RefSeq[47] and H-Invitational[48] databases were aligned to the human genome[49] (hg18) using BLAT50 software and

processed as described to represent each gene by the mRNA isoform with the longest UTR[30]. These unique full-length mRNAs, which were each represented by the genomic sequence of their exons (as the genomic sequence was of higher quality than the mRNA sequence), were the reference mRNAs (Supplementary Data 3). Mouse full-length mRNAs were obtained from RefSeq[47] and FANTOM DB[51] databases, aligned against the mouse genome[52] (mm9) and processed similarly. For *C. elegans* and *Drosophila melanogaster*, we obtained 3′ UTR sequences from TargetScan (targetscan.org)[22,53]. Mature miRNA sequences were downloaded from the miRBase web site[54].

Microarray processing and mapping to reference mRNAs

We collected published data sets reporting the response of HeLa mRNAs 24 h after 100 nM sRNA transfection using Agilent arrays (two-color platform), excluding data sets for which either multiple sRNAs were simultaneously transfected or the transfected RNAs contained chemically modified nucleotides (Supplementary Data 1). If probe sequences for an array platform were available, they were mapped to genomic locations in the human genome using BLAT[50] software. For some arrays (for example, GSE8501), probe sequences were unavailable, but associated cDNA or EST sequence IDs were available. In such cases, genomic coordinates of cDNAs and ESTs obtained from the UCSC Genome Browser[55] were used as if they were coordinates of array probes. Each probe and its associated mRNA fold-change value were mapped to the reference mRNA sharing the greatest overlap with the probe's genomic coordinates, ≥15 bases. When multiple probes were mapped to a single reference mRNA, the median fold change was used. To avoid analysis of mRNAs not expressed in HeLa cells, only mRNAs with signal greater than the median in the mock-transfection samples were considered.

For each array, the median fold change of reference mRNAs without any 6- to 8-nt site was used to normalize the fold changes of all reference mRNAs. To correct for the global association between mRNA fold change and A+U content of the mRNA transcript, the LOWESS filtering was applied by using the malowess function within MATLAB (Mathworks) (Supplementary Data 4). For some arrays, the transfected sRNA was designed to target nearly perfectly matching (≥18 nt) mRNAs, in which case these intended targets were excluded from analysis.

Motif-enrichment analysis for array filtering

To evaluate array data sets, we carried out motif-enrichment analysis using the Fisher's exact test for a 2 x 2 contingency table, populated based on whether the reference mRNA had a 7-mer motif for the cognate sRNA in its 3′ UTR and whether it was among the top 5% most downregulated mRNAs. If multiple arrays examined the effects of transfecting sRNAs with identical seed regions (positions 2–8), the *P* value of the Fisher's exact test for site enrichment (considering either of the two 7-mer sites and picking the one with the lower *P* value) was assessed for each array, and the array with the median *P* value was chosen to represent that seed region, yielding 102 representative arrays (Supplementary Data 1). To obtain a filtered data set, this test was repeated for the 16,384 heptamers, and arrays were retained if the motif most significantly associated with downregulation was the 7-mer-m8 or 7-mer-A1 site of the transfected sRNA; 74 arrays passed this filter (Supplementary Data 1). Results of multiple linear regression and other analyses were robust to cutoff choice (other cutoffs tested were 10, 15 and 20%; data not shown).

Target site abundance

TA in the human transcriptome was calculated as the number of nonoverlapping 3′ UTR 8-mer, 7-mer-m8 and 7-mer-A1 sites in the reference mRNAs. An analogous process was used to calculate TA in mouse, *C. elegans* and *D. melanogaster*. To calculate TA_{HeLa}, each site was weighted based on mRNA-Seq data[33]. Predicted SPS and TA values for all heptamers in *C. elegans*, human and HeLa, mouse and *D. melanogaster* are in Supplementary Data 5.

miRNA target prediction and analysis of siRNA efficacy

Context scores were calculated for the cognate sites of the reference mRNAs using the simple linear regression parameters reported earlier[7]. Before fitting, scores for each parameter were scaled from 0 to 1 (Supplementary Fig. 5b). To account for site type without the complication of multiple sites, we developed models for each type individually, using mRNAs with only a single site to the cognate miRNA (Supplementary Fig. 5c). The multiple linear regression models for context-only and context+ were computed by using the lm function in the R package version 2.11.1.

Acknowledgments

We thank D. Didiano and O. Hobert (Columbia University) for *lsy-6* target constructs and V. Auyeung, R. Friedman, C. Jan and H. Guo for helpful discussions and for sharing data sets before publication. This work was supported by US National Institutes of Health grant GM067031 (D.P.B.) and a Research Settlement Fund for the new faculty of SNU (D.B.). D.P.B. is an investigator of the Howard Hughes Medical Institute.

Figure Legends

Figure 1 Strengthening SPS while decreasing TA imparted typical targeting proficiency to *lsy-6* and miR-23 miRNAs. (**a**) Sequences of miRNAs and target sites tested in reporter assays. Each miRNA was co-transfected with reporter plasmids as a duplex designed to represent the miRNA paired with its miRNA* strand (Supplementary Fig. 1a). (**b**) Response of reporters with 3′ UTRs of predicted *lsy-6* targets after co-transfection with *lsy-6*. As a specificity control, the experiment was also done using a noncognate miRNA, miR-1 (gray bars). Geometric means are plotted relative to those of reporters in which the predicted target sites were mutated after also normalizing for the repression observed for miR-1 (gray bars). Mutant sites of this experiment were the cognate sites of Figure 2d. Error bars, third largest and third smallest values among 12 replicates from 4 independent experiments. Significant differences in repression by cognate miRNA compared to that by noncognate miRNA are indicated. (**c**) Distribution of predicted SPSs for 7-mer-m8 sites of 60 conserved nematode miRNA families[36] (Supplementary Data 2). Values were rounded down to the next half-integer unit. (**d**) SPS distribution for 7-mer-m8 sites of 87 conserved vertebrate miRNA families[8] (Supplementary Data 2). (**e**) Distributions of predicted genome TA for 7-mer-m8 3′ UTR sites of 60 conserved nematode miRNA families (Supplementary Data 2). Values were rounded up to the next tenth of a unit. (**f**) Distributions of predicted genome TA for 7-mer-m8 3′ UTR sites of 87 conserved vertebrate miRNA families (Supplementary Data 2). (**g**) Response of reporters mutated such that their sites matched the miR-142 seed. The cognate miRNA was the miR-142*lsy-6* chimera; noncognate sites were *lsy-6* sites. Otherwise, as in **b**. (**h**) As in **g**, except showing the response to miR-142 transfection. (**i**) Response of reporters with 3′ UTRs of predicted miR-23 targets after co-transfection with miR-

23a. Noncognate sites were for miR-CGCG. Otherwise, as in **b**. (**j**) Response of reporters mutated such that their sites matched the seed of miR-CGCG, which was co-transfected as the cognate miRNA. Noncognate sites were for miR-23. Otherwise, as in **i**. **P* < 0.01, ***P* < 0.001, Wilcoxon rank-sum test.

Figure 2 Separating the effects of SPS and TA on miRNA targeting proficiency. (**a**) Relationship between predicted SPS and genomic TA for *lsy-6* and the 59 other conserved nematode miRNAs (red squares), and all other heptamers (light blue, blue, dark blue or purple squares indicating 0, 1, 2 or 3 CpG dinucleotides within the heptamer, respectively). TA was defined as total number of canonical 7- to 8-nt sites (8-mer, 7-mer-m8 and 7-mer-A1) in annotated 3′ UTRs. SPS values were predicted using the respective 7-mer-m8 sites. (**b**) Relationship between predicted SPS and TA in human 3′ UTRs for miR-23 and the 86 other broadly conserved vertebrate miRNA families (red squares). Otherwise, as in **a**. (**c**) Sequences of miRNAs and target sites tested in reporter assays of this figure. (**d**) Response of reporters with 3′ UTRs of predicted *lsy-6* targets mutated such that their sites matched the seed of LTA-*lsy-6*, which was co-transfected as the cognate miRNA. Noncognate sites were for *lsy-6*. Otherwise, as in Figure 1b. (**e**) 2,6-di-aminopurine (DAP or D)-uracil base pair. (**f**) Response of reporters used in **d** after co-transfecting D-LTA-*lsy-6* as the cognate miRNA. Otherwise, as in **d**. (**g**) Response of reporters used in Figure 1i after co-transfecting D-miR-23a as the cognate miRNA, alongside results for miR-23a that was repeated in parallel. Otherwise, as in Figure 1i. **P* < 0.01, ***P* < 0.001, Wilcoxon rank-sum test.

Figure 3 Impact of TA and SPS on sRNA targeting proficiency, as determined using array data. (**a**) Distribution of TA_{HeLa} and predicted SPS for the sRNAs from the 102 array data sets analyzed in this study (orange squares) and sRNAs from data sets that passed the motif-enrichment analysis (red squares). Otherwise, plotted as in Figure 2b. (**b**) Response of expressed mRNAs with a single 3′ UTR site to the cognate sRNA, with respect to TA_{HeLa} and predicted SPS. Fold-change values are plotted according the key to the right of each plot, comparing mRNAs with a single site of the type indicated (and no additional sites to the cognate sRNA elsewhere in the mRNA) to those with no site to the cognate sRNA; note different scales for different plots. In areas of overlap, mean values are plotted. Correlation coefficients and *P* values are in Table 1. (**c**) Response of expressed mRNAs with a single ORF site to the cognate sRNA, with respect to TA_{HeLa} and predicted SPS. Otherwise, as in **b**. (**d**) Response of mRNAs with indicated single sites when binning cognate sRNA by TA_{HeLa} (top) or predicted SPS (bottom). The key indicates the data considered, with the first quartiles at top comprising data for sRNAs with the lowest TA_{HeLa} and those at bottom comprising data for sRNAs with the strongest predicted SPS. Error bars, 95% confidence intervals.

Figure 4 Predictive performance of the context+ model, which considers miRNA or siRNA proficiency in addition to site context. (**a**) Improved predictions for mRNAs with canonical 7- to 8-nt 3′ UTR sites. Predicted interactions between mRNAs and cognate sRNA were distributed into ten equally populated bins based on total context scores generated using the model indicated (key), with the first bin comprising interactions with most favorable scores. Plotted for each bin is the mean mRNA change on the arrays (error bars, 95% confidence intervals). (**b**) Prediction of

responsive interactions involving mRNAs with only 3′ UTR 6-mer sites. Otherwise, as in **a**. (**c**) Prediction of responsive interactions involving mRNAs with at least one 8-mer ORF site but no 3′ UTR sites. Otherwise, as in **a**. (**d**) Impact of TA and SPS on siRNA-directed knockdown of the desired target. Efficacy in luciferase activity knockdown for 2,431 siRNAs transfected into H1299 cells[38]. Efficacy is linearly scaled (key), with positive and negative controls having values of 0.900 and 0.354, respectively[38].

Figure 1

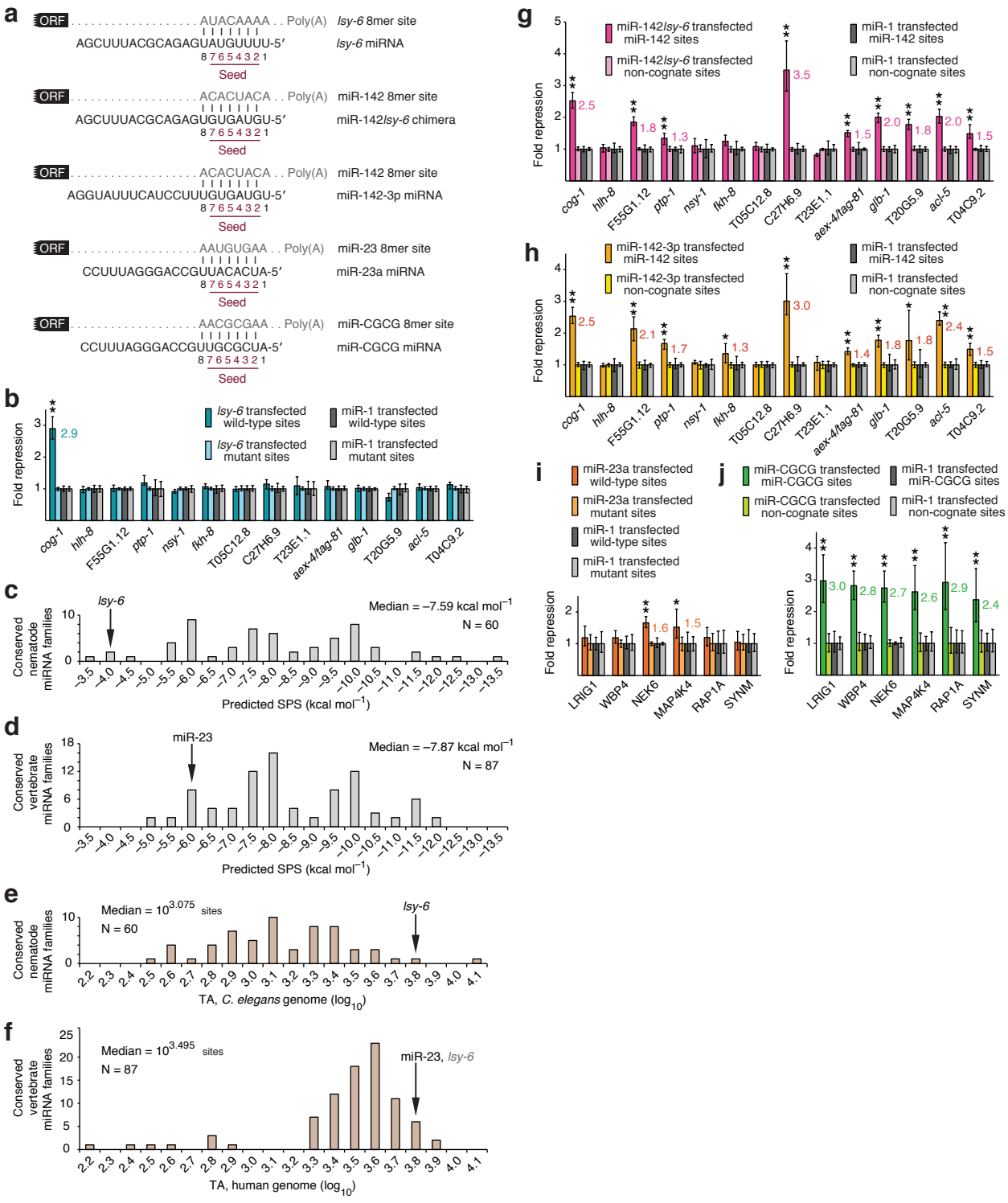

a
ORF . . . . . AUACAAAA . . Poly(A) lsy-6 8mer site
AGCUUUACGCAGAGUAUGUUUU-5′ lsy-6 miRNA
8 7 6 5 4 3 2 1
Seed
ORF . . . . . ACACUACA . . Poly(A) miR-142 8mer site
AGCUUUACGCAGAGUGUGAUGU-5′ miR-142/lsy-6 chimera
8 7 6 5 4 3 2 1
Seed
ORF . . . . . ACACUACA . . Poly(A) miR-142 8mer site
AGGUAUUUCAUCCUUUGUGAUGU-5′ miR-142-3p miRNA
8 7 6 5 4 3 2 1
Seed
ORF . . . . . AAUGUGAA . . Poly(A) miR-23 8mer site
CCUUUAGGGACCGUUACACUA-5′ miR-23a miRNA
8 7 6 5 4 3 2 1
Seed
ORF . . . . . AACGCGAA . . Poly(A) miR-CGCG 8mer site
CCUUUAGGGACCGUUGCGCUA-5′ miR-CGCG miRNA
8 7 6 5 4 3 2 1
Seed
b
Fold repression
lsy-6 transfected wild-type sites
miR-1 transfected wild-type sites
lsy-6 transfected mutant sites
miR-1 transfected mutant sites
2.9
cog-1 hlh-8 F55G1.12 ptp-1 nsy-1 fkh-8 T05C12.8 C27H6.9 T23E1.1 aex-4/tag-81 glb-1 T20G5.9 acl-5 T04C9.2
c
Conserved nematode miRNA families
lsy-6
Median = –7.59 kcal mol–1
N = 60
Predicted SPS (kcal mol–1)
d
Conserved vertebrate miRNA families
miR-23
Median = –7.87 kcal mol–1
N = 87
Predicted SPS (kcal mol–1)
e
Conserved nematode miRNA families
Median = 10^3.075 sites
N = 60
lsy-6
TA, C. elegans genome (log10)
f
Conserved vertebrate miRNA families
Median = 10^3.495 sites
N = 87
miR-23, lsy-6
TA, human genome (log10)
g
Fold repression
miR-142/lsy-6 transfected miR-142 sites
miR-142/lsy-6 transfected non-cognate sites
miR-1 transfected miR-142 sites
miR-1 transfected non-cognate sites
2.5 1.8 1.3 3.5 1.5 2.0 1.8 2.0 1.5
cog-1 hlh-8 F55G1.12 ptp-1 nsy-1 fkh-8 T05C12.8 C27H6.9 T23E1.1 aex-4/tag-81 glb-1 T20G5.9 acl-5 T04C9.2
h
Fold repression
miR-142-3p transfected miR-142 sites
miR-142-3p transfected non-cognate sites
miR-1 transfected miR-142 sites
miR-1 transfected non-cognate sites
2.5 2.1 1.7 1.3 3.0 1.4 1.8 1.8 2.4 1.5
cog-1 hlh-8 F55G1.12 ptp-1 nsy-1 fkh-8 T05C12.8 C27H6.9 T23E1.1 aex-4/tag-81 glb-1 T20G5.9 acl-5 T04C9.2
i
miR-23a transfected wild-type sites
miR-23a transfected mutant sites
miR-1 transfected wild-type sites
miR-1 transfected mutant sites
Fold repression
1.6 1.5
LRIG1 WBP4 NEK6 MAP4K4 RAP1A SYNM
j
miR-CGCG transfected miR-CGCG sites
miR-CGCG transfected non-cognate sites
miR-1 transfected miR-CGCG sites
miR-1 transfected non-cognate sites
Fold repression
3.0 2.8 2.7 2.6 2.9 2.4
LRIG1 WBP4 NEK6 MAP4K4 RAP1A SYNM

Figure 2

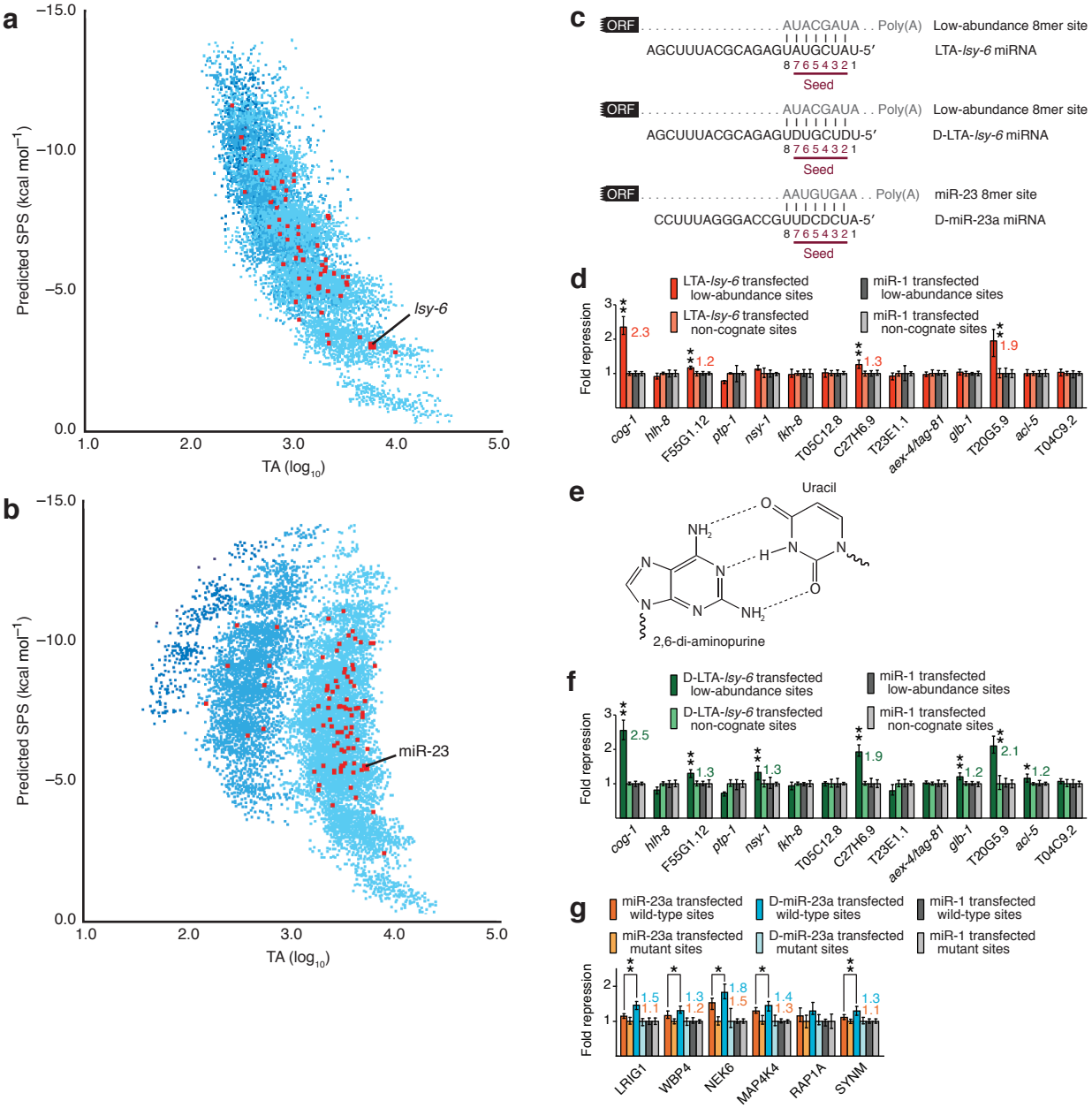

a
Predicted SPS (kcal mol−1)
−15.0
−10.0
−5.0
0.0
1.0
2.0
3.0
4.0
5.0
TA (log10)
lsy-6
b
miR-23
c
ORF
AUACGAUA . . Poly(A)
Low-abundance 8mer site
AGCUUUACGCAGAGUAUGCUAU-5′
LTA-lsy-6 miRNA
8 7 6 5 4 3 2 1
Seed
AUACGAUA . . Poly(A)
Low-abundance 8mer site
AGCUUUACGCAGAGUDUGCUDU-5′
D-LTA-lsy-6 miRNA
AAUGUGAA . . Poly(A)
miR-23 8mer site
CCUUUAGGGACCGUUDCDCUA-5′
D-miR-23a miRNA
d
LTA-lsy-6 transfected low-abundance sites
LTA-lsy-6 transfected non-cognate sites
miR-1 transfected low-abundance sites
miR-1 transfected non-cognate sites
Fold repression
cog-1
hlh-8
F55G1.12
ptp-1
nsy-1
fkh-8
T05C12.8
C27H6.9
T23E1.1
aex-4/tag-81
glb-1
T20G5.9
acl-5
T04C9.2
e
Uracil
2,6-di-aminopurine
f
D-LTA-lsy-6 transfected low-abundance sites
D-LTA-lsy-6 transfected non-cognate sites
miR-1 transfected low-abundance sites
miR-1 transfected non-cognate sites
g
miR-23a transfected wild-type sites
D-miR-23a transfected wild-type sites
miR-1 transfected wild-type sites
miR-23a transfected mutant sites
D-miR-23a transfected mutant sites
miR-1 transfected mutant sites
LRIG1
WBP4
NEK6
MAP4K4
RAP1A
SYNM

Figure 3

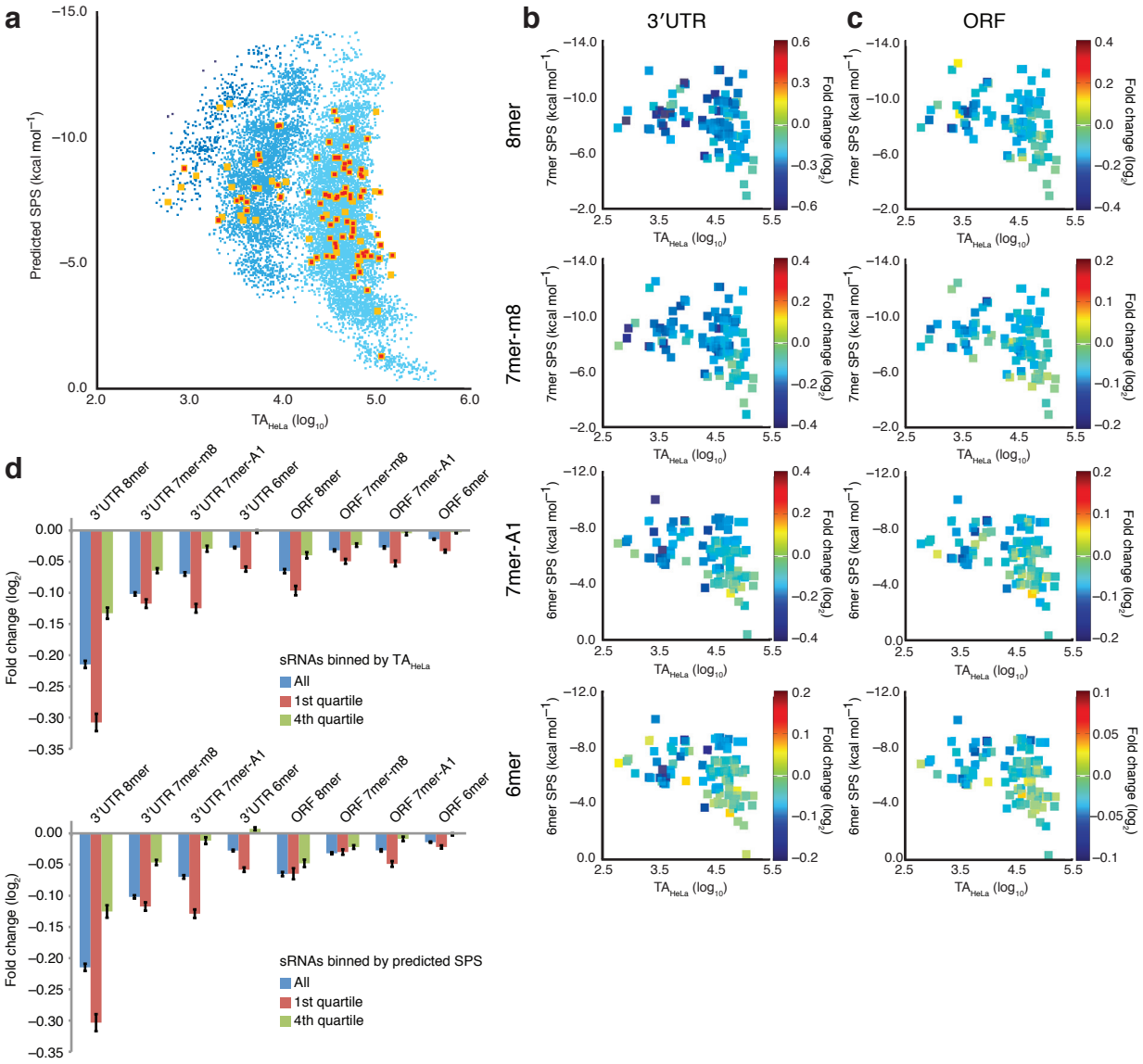

a
Predicted SPS (kcal mol−1)
TAHeLa (log10)
b
3′UTR
c
ORF
8mer
7mer-m8
7mer-A1
6mer
7mer SPS (kcal mol−1)
6mer SPS (kcal mol−1)
Fold change (log2)
d
3′UTR 8mer
3′UTR 7mer-m8
3′UTR 7mer-A1
3′UTR 6mer
ORF 8mer
ORF 7mer-m8
ORF 7mer-A1
ORF 6mer
Fold change (log2)
sRNAs binned by TAHeLa
All
1st quartile
4th quartile
sRNAs binned by predicted SPS

Figure 4

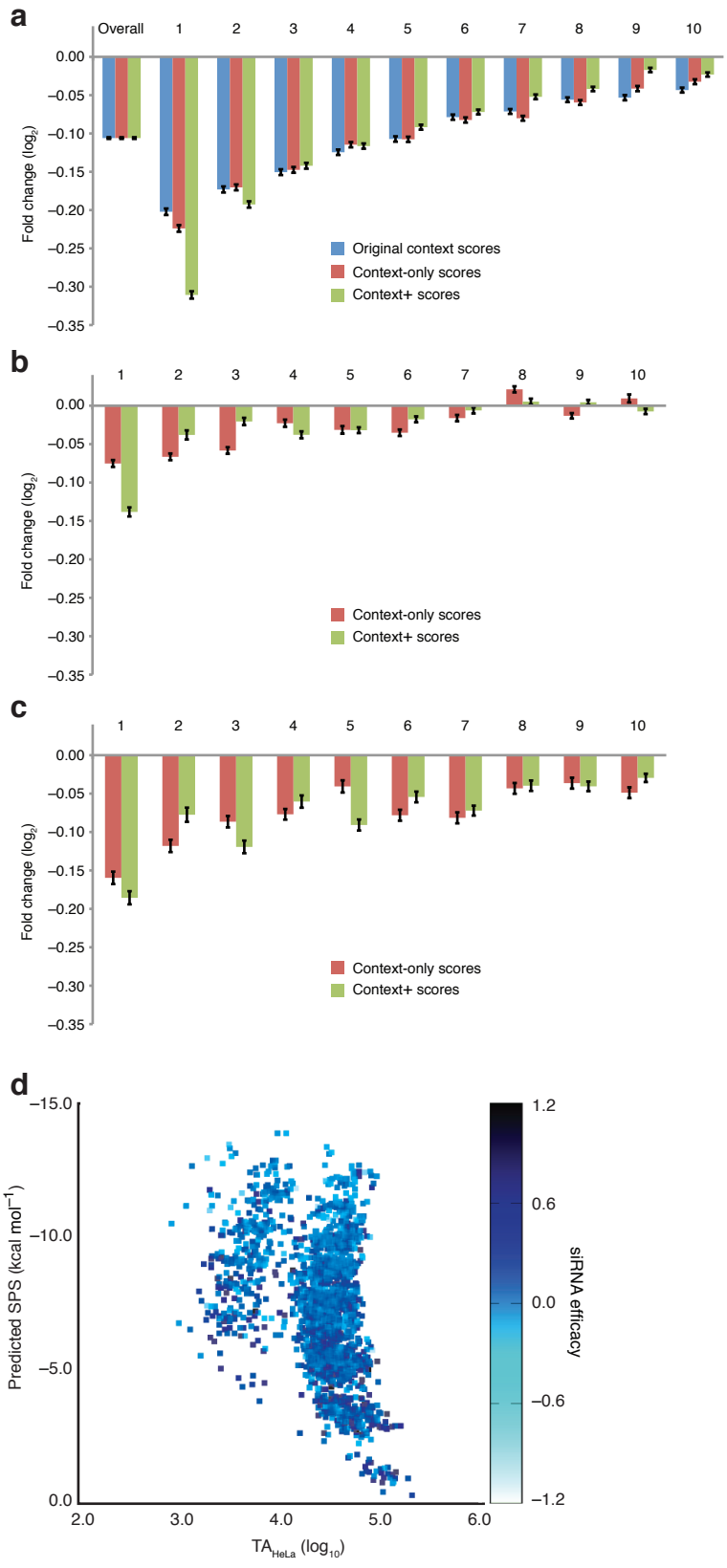
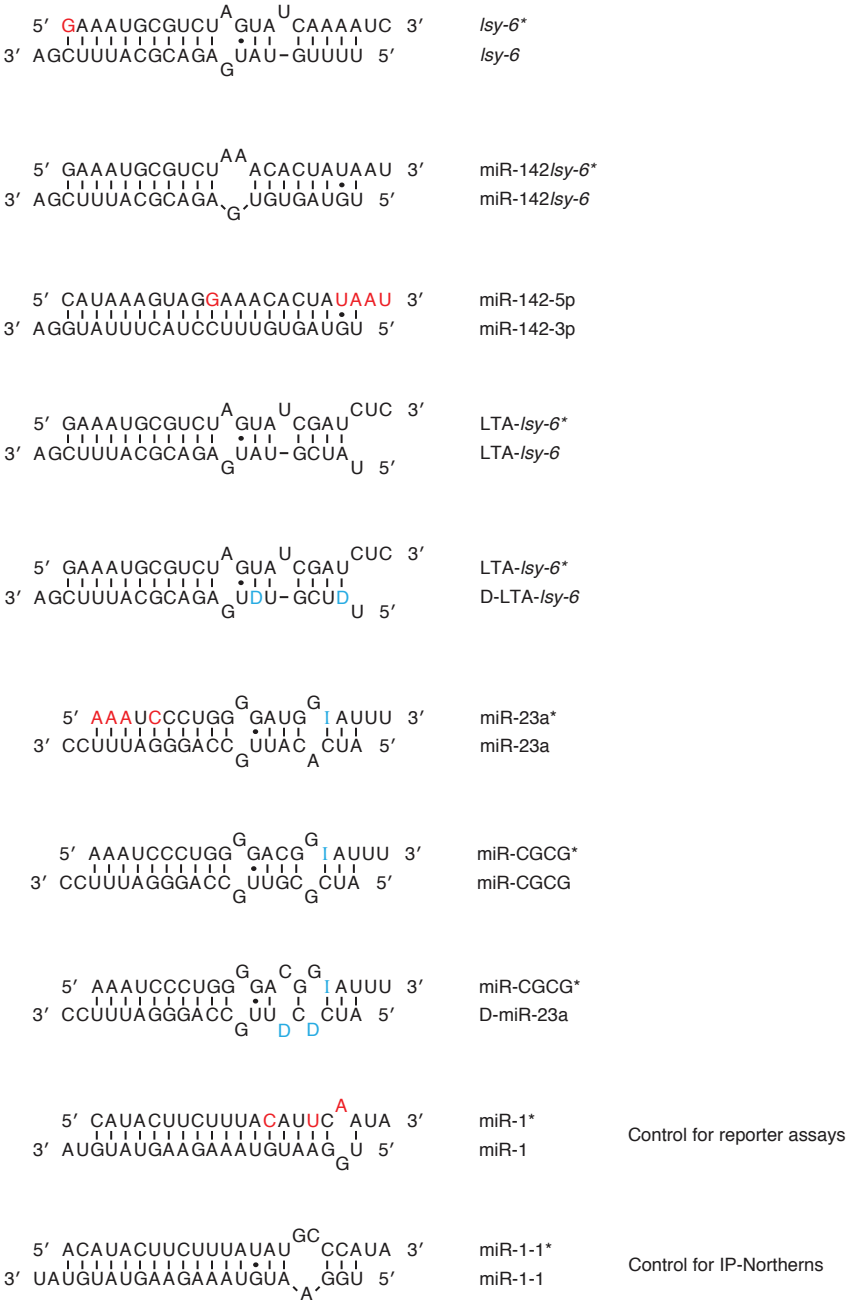

a
Overall
1
2
3
4
5
6
7
8
9
10
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
−0.30
−0.35
Fold change (log2)
Original context scores
Context-only scores
Context+ scores
b
Context-only scores
Context+ scores
c
Context-only scores
Context+ scores
d
−15.0
−10.0
−5.0
0.0
Predicted SPS (kcal mol−1)
2.0
3.0
4.0
5.0
6.0
TAHeLa (log10)
1.2
0.6
0.0
−0.6
−1.2
siRNA efficacy

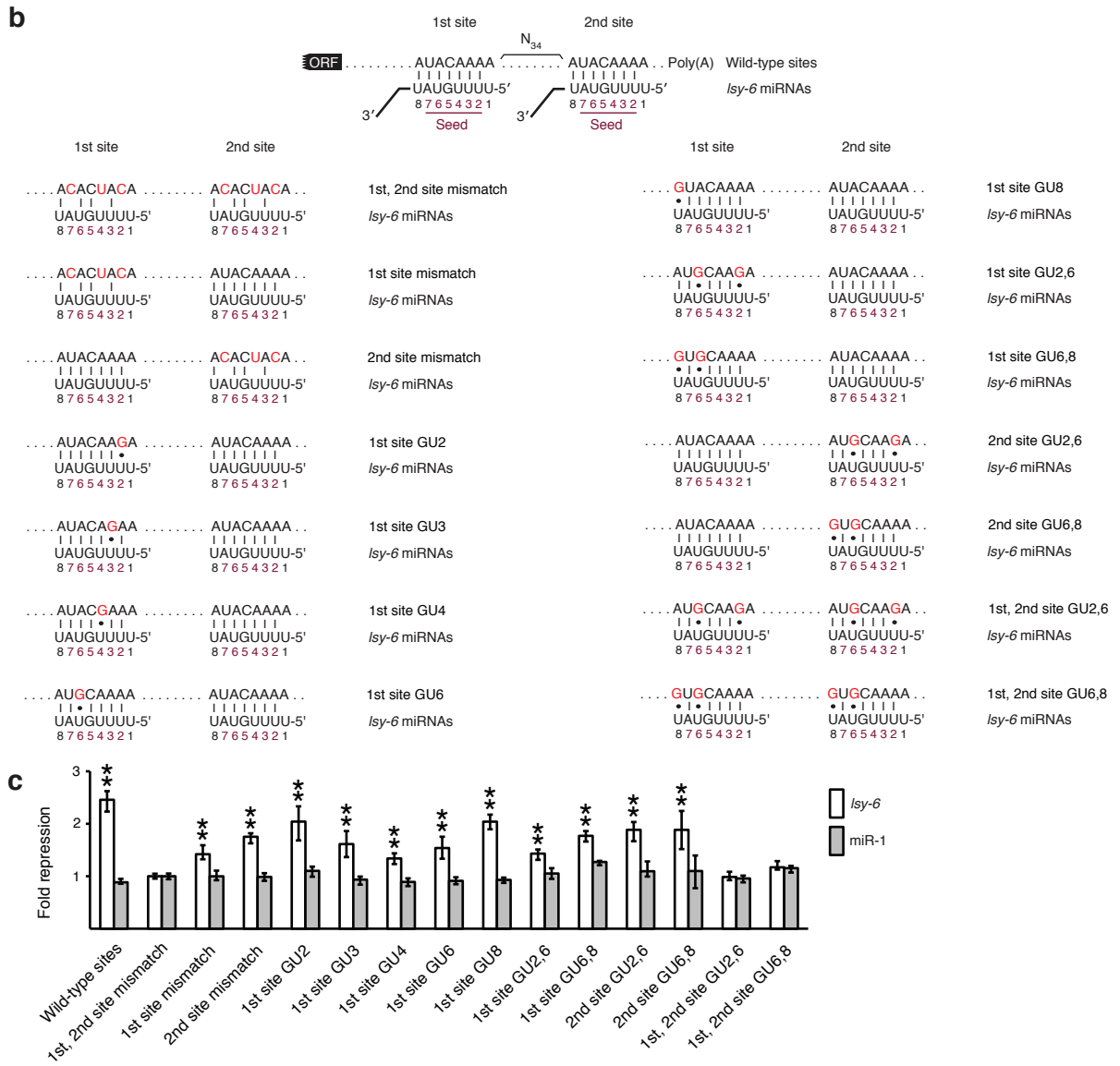
Table 1 Relationship between mean mRNA repression and either TA or predicted SPS for the indicated site types, as determined from microarray data (Fig. 3b,c).

Site location and type	Multiple linear regression			Simple linear regression			
	Multiple R^2	*P* value		TA_{HeLa}		SPS	
		TA_{HeLa}	SPS	R^2	*P* value	R^2	*P* value
3′UTR 8mer	0.149	0.0049	0.051	0.115	0.0006	0.076	0.0054
3′UTR 7mer-m8	0.190	0.0081	0.0047	0.122	0.0003	0.131	0.0002
3′UTR 7mer-A1	0.335	0.0009	2×10^{-5}	0.196	3×10^{-6}	0.256	6×10^{-8}
3′UTR 6mer	0.177	0.039	0.0025	0.097	0.0014	0.141	0.0001
ORF 8mer	0.104	0.018	0.14	0.085	0.0030	0.052	0.021
ORF 7mer-m8	0.171	0.019	0.0054	0.103	0.0010	0.123	0.0003
ORF 7mer-A1	0.135	0.010	0.073	0.106	0.0008	0.076	0.0052
ORF 6mer	0.228	0.010	0.0008	0.133	0.0002	0.174	1×10^{-5}
5′UTR 8mer	0.004	0.75	0.68	0.002	0.64	0.003	0.59
5′UTR 7mer-m8	0.003	0.63	0.72	0.002	0.70	0.000	0.84
5′UTR 7mer-A1	0.012	0.60	0.49	0.007	0.41	0.009	0.35
5′UTR 6mer	0.011	0.97	0.32	0.001	0.74	0.011	0.29

a

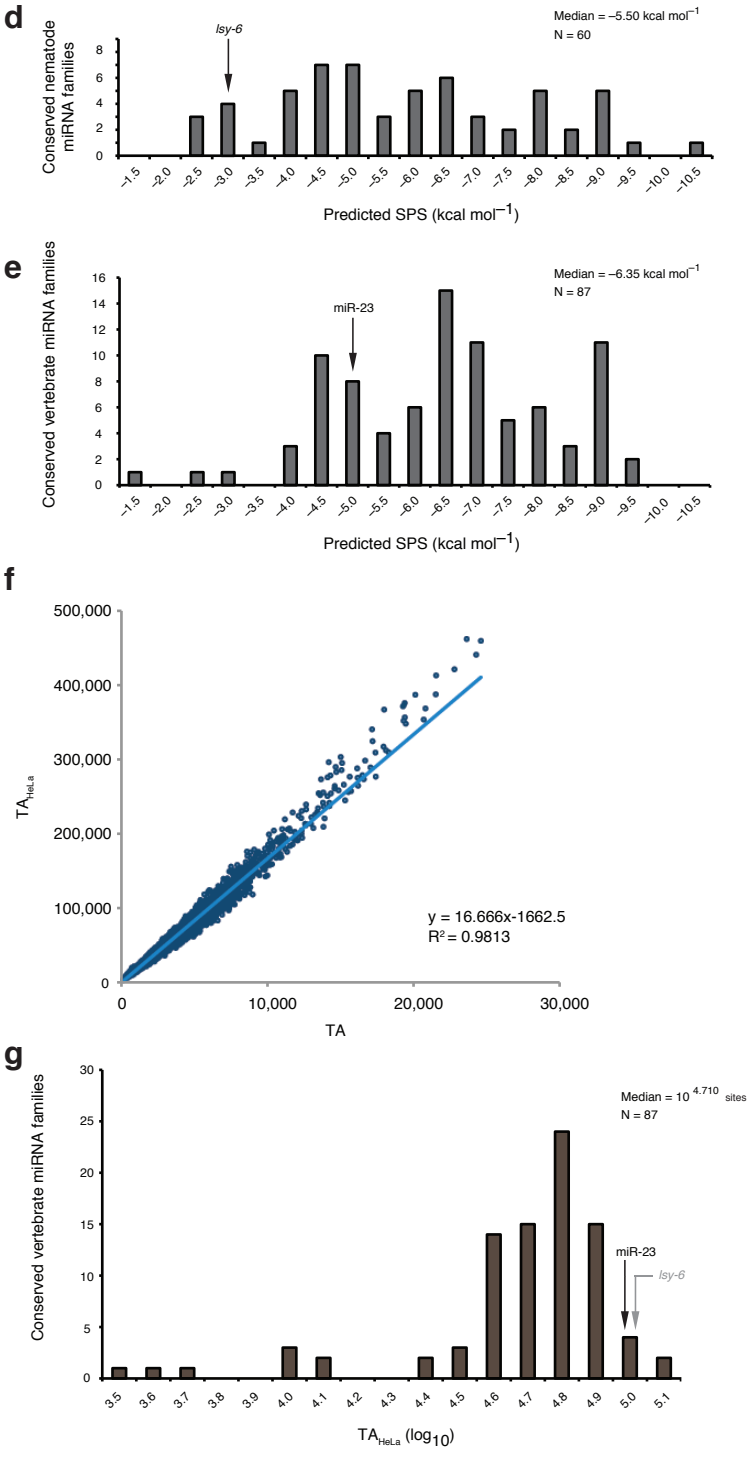


Supplementary Figure 1. Information and analyses related to **Figure 1**. (**a**) Predicted structures for miRNA duplexes transfected in this study. For miRNA mimics of endogenous sequences (*lsy-6*, miR-142-3p, miR-23a, miR-1, miR-1-1), miRNA* nucleotides that differed from their endogenous identities[36,56] are highlighted in red. These changes were designed to facilitate loading of the miRNA. Additionally, a guanine present within endogenous miR-142-5p was deleted (not shown). Non-canonical nucleotides used to either increase SPS (D = 2,6-di-aminopurine), or facilitate loading (I = Inosine), are highlighted in cyan.

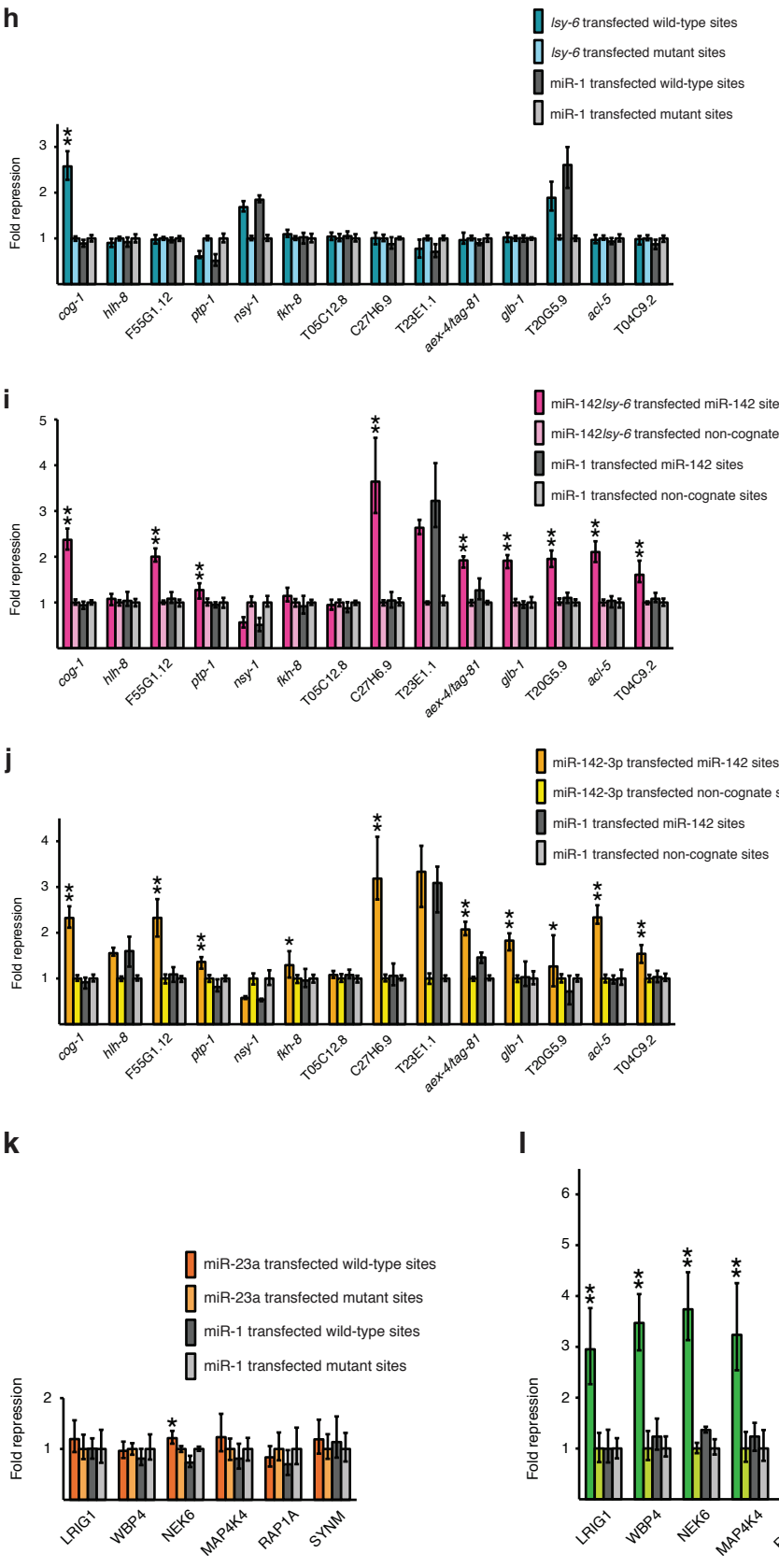


Supplementary Figure 1 continued. (**b**) *cog-1* 3′UTR wild-type and mutant sites containing mismatches or G:U wobbles to the indicated nucleotide(s) of *lsy-6*. Illustrations of mutant sites, with mutated positions shown in red, are simplified from the wild-type sites at top. (**c**) Response of the *cog-1* 3′UTR reporter to mutations in the *lsy-6* sites. Repression of each construct by *lsy-6* was normalized to a construct with two mutated *lsy-6* sites, each containing two mismatches (1st, 2nd site mismatch). In parallel, activity was measured using a non-cognate miRNA, miR-1 (grey bars). Normalization was as panels **h–l** of this figure. Error bars and statistical significance is as in **Figure 1b**.

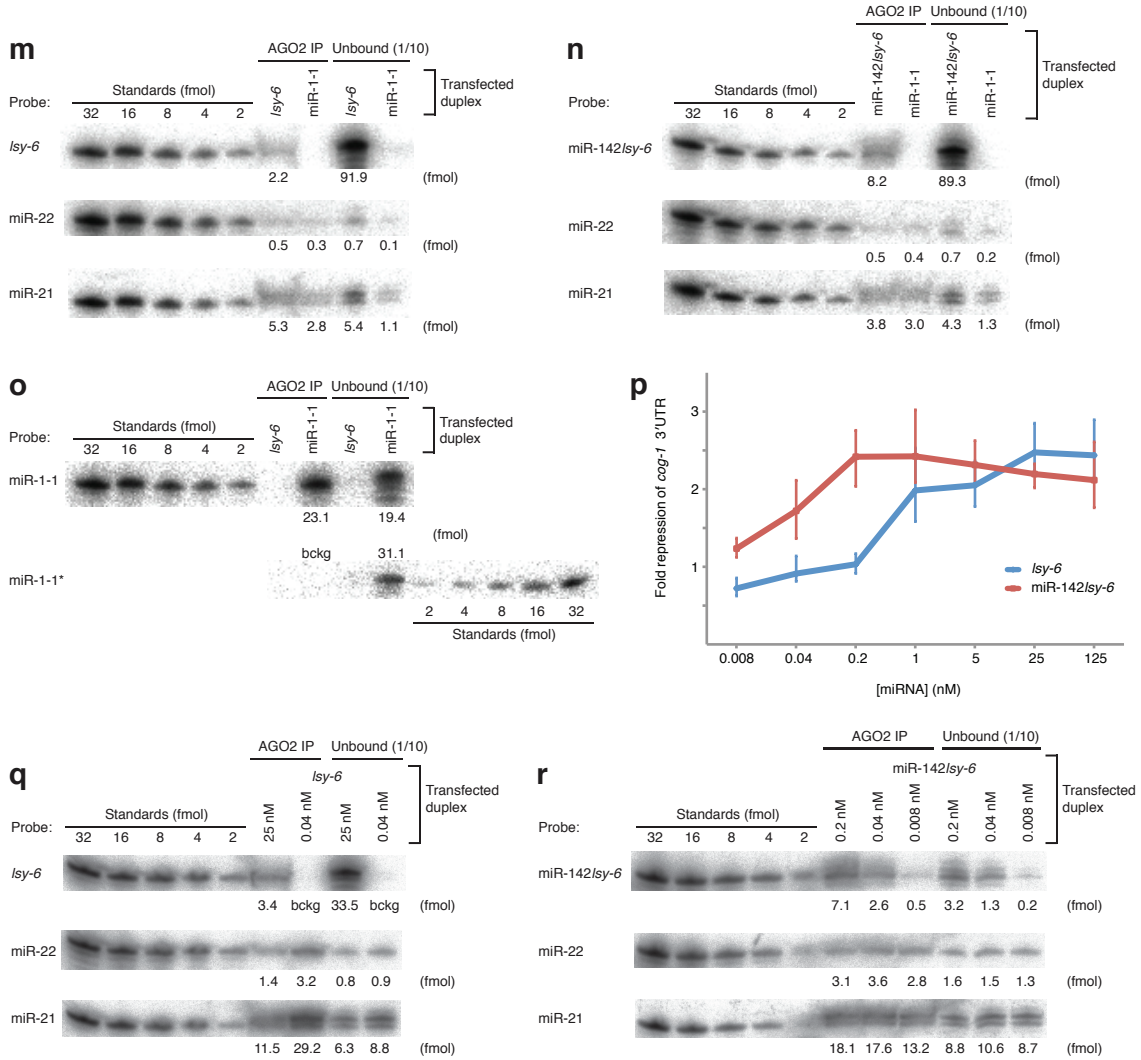
The original study using *in vivo* reporter assays in *C. elegans* concludes that repression of *cog-1* by *lsy-6* is not strongly diminished by the introduction of G:U wobbles into the seed match[23], which contrasts with conclusions from studies using reporters in mammalian cells[57] and *D. melanogaster*[5] as well as many other studies using comparative sequence analysis and large-scale experimental datasets[3]. A second study of the *lsy-6*:*cog-1* interaction concludes that some G:U wobble combinations diminish repression of *cog-1* by *lsy-6* in the *in vivo* reporter assay[37]. We used luciferase reporter assays in HeLa cells to examine the same G:U wobble changes as those examined in worms, as well as some additional changes (**Supplementary Table 1**). Introducing G:U wobbles into the upstream *lsy-6* site in *cog-1* was detrimental in all cases. G:U wobbles in the downstream *lsy-6* site also reduced repression, although the effect was less pronounced than for wobbles in the upstream site. Introducing two wobbles into both sites abolished repression.



Supplementary Figure 1 continued. (**d**) Distribution of predicted SPSs for 6mer miRNA sites to 60 conserved nematode miRNA families (**Supplementary Table 7**), as in **Figure 1c**. (**e**) Distribution of predicted SPSs for 6mer miRNA sites to 87 conserved vertebrate miRNA families (**Supplementary Table 7**), as in **Figure 1d**. (**f**) Relationship between human TA and TA_{HeLa} for all heptamers. The least-squares linear fit to the data is shown, with the equation for the line and its Spearman's R^2. (**g**) Distribution of TA_{HeLa}, counting 7mer-m8 3′UTR sites for 87 conserved vertebrate miRNA families, plotted as in **Figure 1f**. TA_{HeLa} values for all 16,384 heptamers are provided in **Supplementary Table 10**.



Supplementary Figure 1 continued. (**h–l**) Reporter results presented in **Figure 1** before normalizing to ratios obtained for the non-cognate miRNA, miR-1. In the main figures, cognate miRNA repression values are normalized to repression values by miR-1. This normalization method was useful because expression differences between the test and control constructs were sometimes observed in the absence of the cognate miRNA (e.g., *nsy-6* or *ptp-1* in **h**).



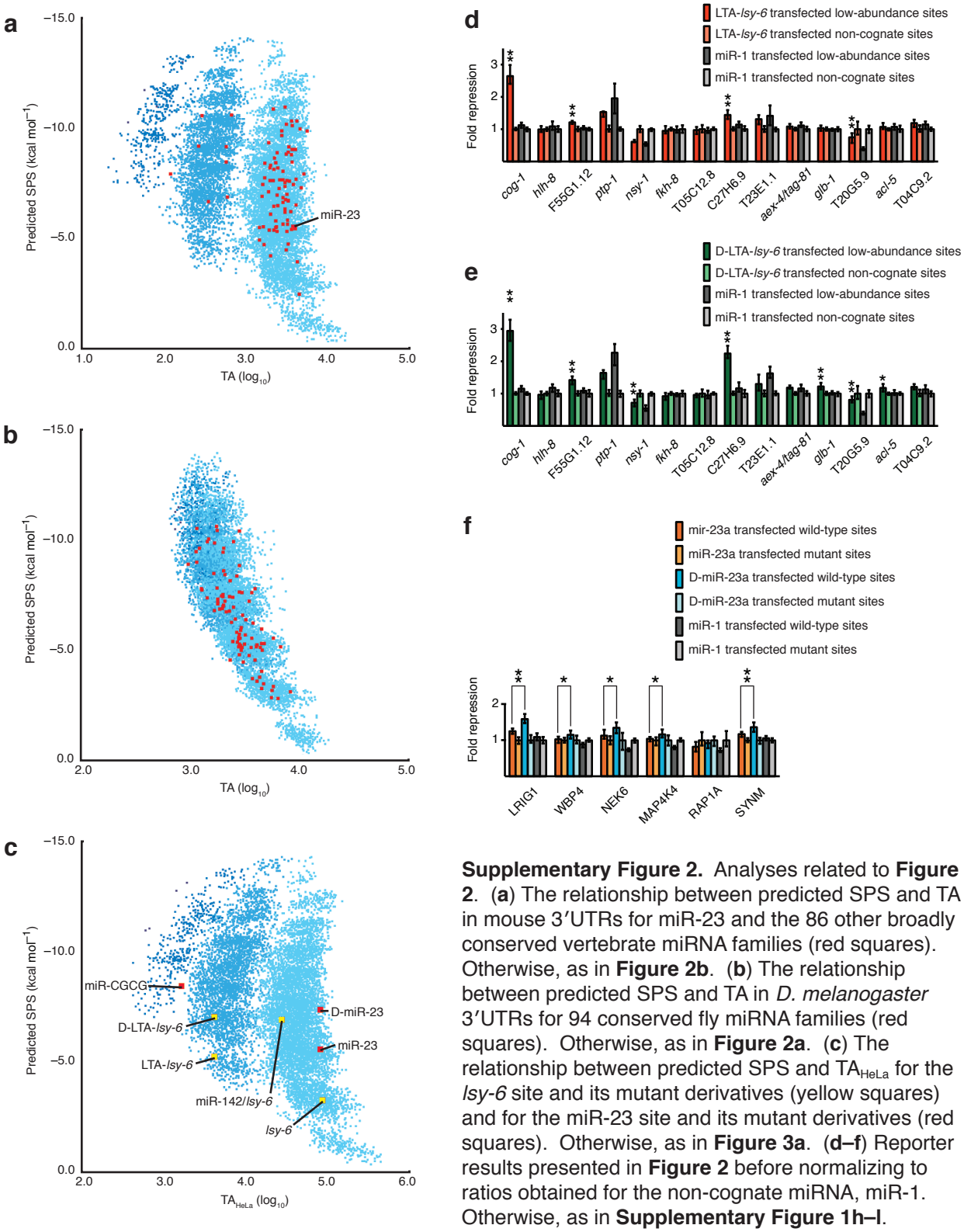
Supplementary Figure 1 continued. Accumulation of transfected miRNAs within the AGO2 silencing complex. (**m**) Quantitative RNA blot probing for *lsy-6* and endogenous controls (miR-21 and miR-22), comparing samples with synthetic RNA standards to samples with material that co-purified with AGO2 (AGO2 IP) and material that did not co-purify (unbound) after transfecting the indicated miRNA duplexes. miR-1-1 samples contained half of the material present in *lsy-6* samples. Because a large fraction of the transfected miRNA did not co-purify with AGO2, only one-tenth of unbound material corresponding to bound material was loaded on the gel. (**n**) Quantitative RNA blot probing for miR-142/*lsy-6* chimera and endogenous controls, otherwise as in **m**. (**o**) Control blot probing for miR-1-1 and miR-1-1*, which demonstrated the specificity of the co-purification for loaded miRNA. Otherwise, as in **m**. (**p**) Repression of *cog-1* reporters containing cognate sites for either *lsy-6* or

Supplementary Figure 1m–s continued.
miR-142*lsy-6* chimera measured across a range of transfected miRNA concentrations. Data is plotted as in **Figure 1**, except error bars represent the second largest and second smallest values among 9 replicates from 3 independent experiments. For normalization, a non-cognate miRNA (miR-1) was co-transfected in parallel at the same concentrations as the cognate miRNAs. (**q,r**) Repeat of the experiment in panels **m** and **n**, transfecting less miR-142*lsy-6* chimera to account for its more efficient accumulation in the AGO2 silencing complex. (**s**) Response of reporters to transfection of miR-142*lsy-6* chimera at 0.2 nM, otherwise as in **Figure 1g**.

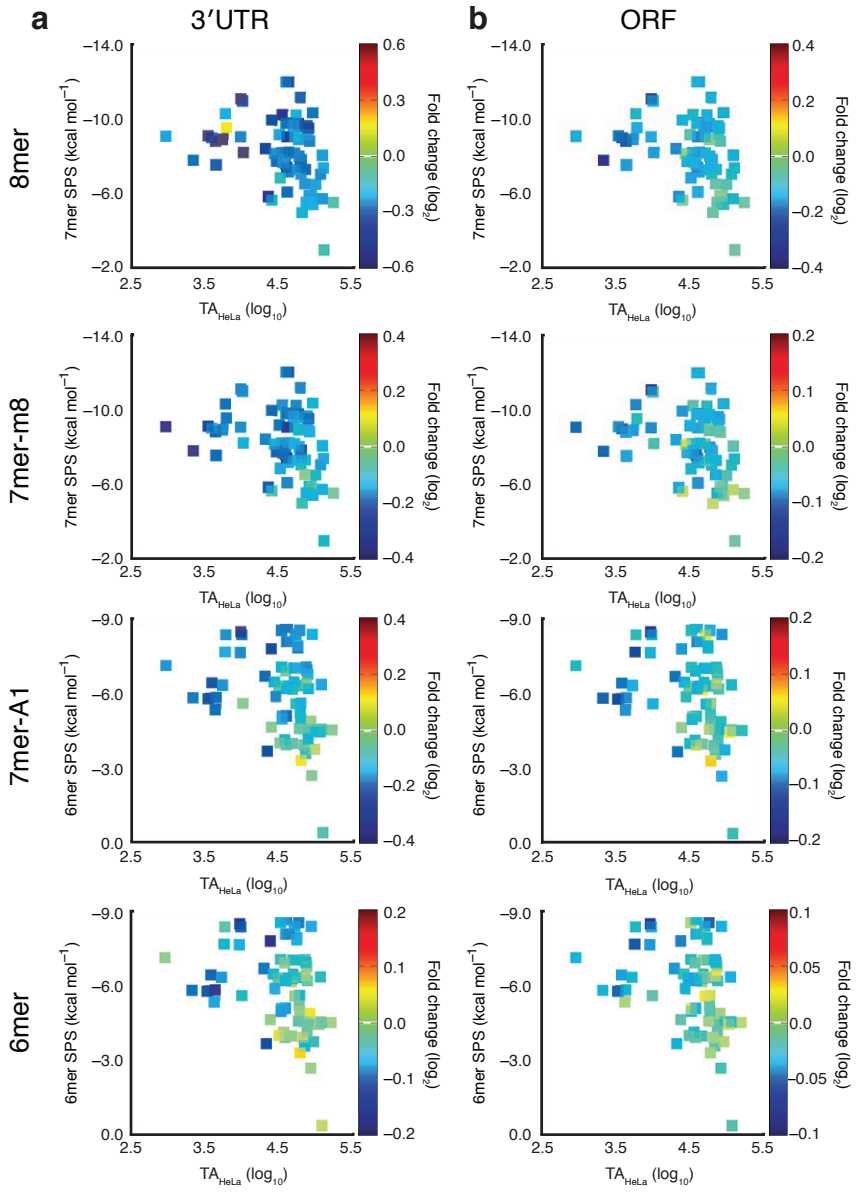
Analyses of the co-purfication results in panels **m** and **n** (geometric mean of ratios normalized to the endogenous internal controls, miR-22 and miR-21) indicated that miR-142*lsy-6* chimera accumulated in AGO2 at a level 4.4-fold higher than did *lsy-6*. This difference represented an estimate of relative accumulation in the silencing complex because levels in AGO1, AGO3, and AGO4 were not determined and because loaded miRNAs might have different degradation rates over the 24 hours after transfection. Because eight targets in **Figure 1** were not significantly repressed by *lsy-6* but were repressed between 1.3- and 3.5-fold by miR-142*lsy-6* chimera, an accumulation difference of less than 5-fold could not explain the difference in proficiency. Consistent with this interpretation were miRNA titration results (**p**), which indicated a rather shallow relationship between miRNA tranfection concentration and fold repression, such that 5-fold differences in miRNA concentration would not be expected to result in the binary differences observed between *lsy-6* and miR-142*lsy-6* chimera, particularly near the concentration used (25 nM).

To find transfection concentrations yielding equal the levels of AGO2-bound *lsy-6* and miR-142*lsy-6* chimera, AGO2 immunopurification was repeated after transfecting miR-142*lsy-6* chimera at concentrations matching those tested in panel **p**. Analyses of these results (panels **q** and **r**) suggested that transfection of miR-142*lsy-6* chimera at 0.2 nM resulted in accumulation of AGO2-bound miRNA to a level similar to that of *lsy-6* transfected at 25 nM. At even lower transfection concentrations, miR-142*lsy-6* chimera levels in AGO2 decreased further, consistent with the reduced repression of *cog-1* at these concentrations (panel **p**). Transfection of miR-142*lsy-6* chimera at 0.2 nM yielded greater reporter repression than that observed in **Figure 1b**, but less than that observed in **Figure 1g** (panel **s**). These results indicate that the relative level of miRNA in the silencing complex (presumed functions of miRNA turnover and loading efficiencies) was not the only factor contributing to proficiency, thereby supporting our conclusion that properties of the seed also played a role. Additional experiments will be needed to learn whether the less efficient accumulation of AGO2-bound *lsy-6* is attributable to poorer loading or faster turnover. If faster turnover of loaded *lsy-6* were a factor, then comparing the results of panel **s** with **Figure 1b** would underestimate the effects of SPS and TA, because the luciferase reporter assay results represented cumulative effects of the miRNA on targets since transfection, and at earlier times the levels of loaded *lsy-6* in **Figure 1b** would have been relatively higher than levels of loaded miR-142*lsy-6* in panel **s**.

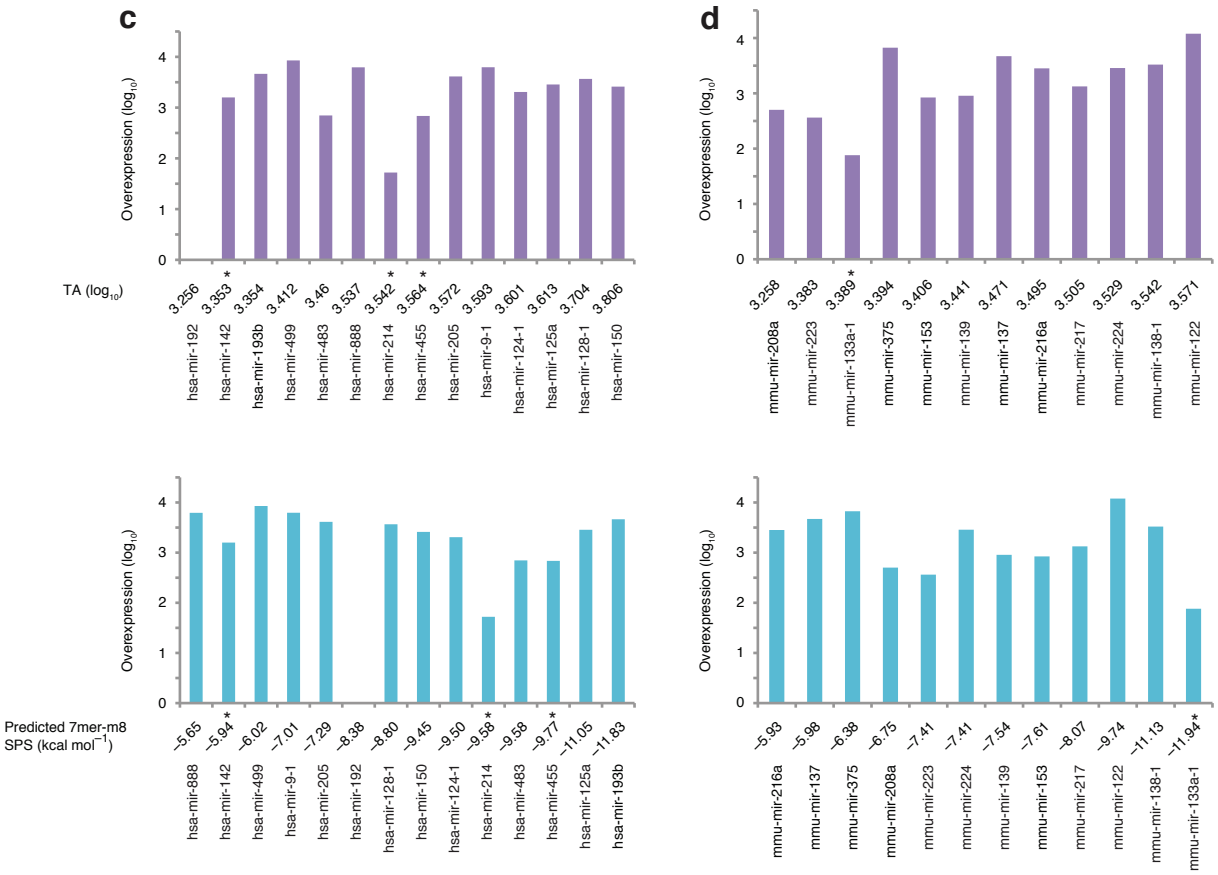
The methods for the immunopurification experiment were as follows: For each miRNA duplex, four (**m–o**) or three (**q,r**) 24-well plates of HeLa cells were transfected as described for the reporter assays (at 25 nM unless otherwise labeled). Half of the wells were co-transfected with pIS0 and pIS1 containing wild-type *lsy-6* sites, the other half with pIS0 and pIS1 containing mutated *lsy-6* sites, and cells were mixed during harvesting. After 24 hours, cells were washed once with 1X PBS and trypsinized, after which all remaining steps were carried out either at 4° C or on ice. Cells were harvested by resuspension in growth media, pelleted (200 x g for 5 minutes), washed with 1X PBS and re-pelleted, then lysed with 4.8 mL or 3.6 mL (50 μL per well) Ago Lysis Buffer (ALB)(25 mM Tris-Cl pH 7.4, 150 mM KCl, 0.5 mM EDTA, 0.5% NP-40, 0.5 mM DTT, one Roche EDTA-free Protease Inhibitor Cocktail tablet per 10 mL) for 1 hour. Cellular debris was spun out (200 x g for 5 minutes), and for each sample, supernatant was mixed with 15 μL of Anti-Human AGO2 antibody (Wako, clone 4G8). After 1 hour, 80 μL EZview Red Protein G Affinity Gel (Sigma) was added, and the mixture was incubated another 4 hours with rocking. Beads were spun down and supernatant ("Unbound") was set aside for later RNA isolation. Beads were washed two times in ALB and then two times in Minimal Cleavage Buffer (MCB) (400 mM KCl, 1mM $MgCl_2$, 10 mM Tris-Cl pH 7.4, 20% w/v Glycerol, 0.5mM DTT). Yeast total RNA was added to IP samples to a concentration of 200 ng per μL, and RNA from IP and Unbound samples was isolated using TRI reagent (Ambion). Small RNA blots were generated and probed as described (http://web.wi.mit.edu/bartel/pub/protocols.html). To enable quantification of RNA levels in the IP and unbound samples, dilution series of synthetic standards for the relevant RNAs were also loaded and used to generate a standard curve–AAGCUGCCAGUUGAAGAACUGU (miR-22); UAGCUUAUCAGACUGAUGUUGA (miR-21); *lsy-6*, miR-142*lsy-6*, miR-1-1, and miR-1-1* sequences are shown in **Supplementary Figure 1a**. Probe sequences: TCGAAATGCGTCTCATACAAAA (*lsy-6*); TCGAAATGCGTCTCACACTACA (miR-142*lsy-6*); TACATACTTCTTTACATTCCA (miR-1-1); TATGGGCATATAAAGAAGTATGT (miR-1-1*); ACAGTTCTTCAACTGGCAGCTT (miR-22); TCAACATCAGTCTGATAAGCTA (miR-21).



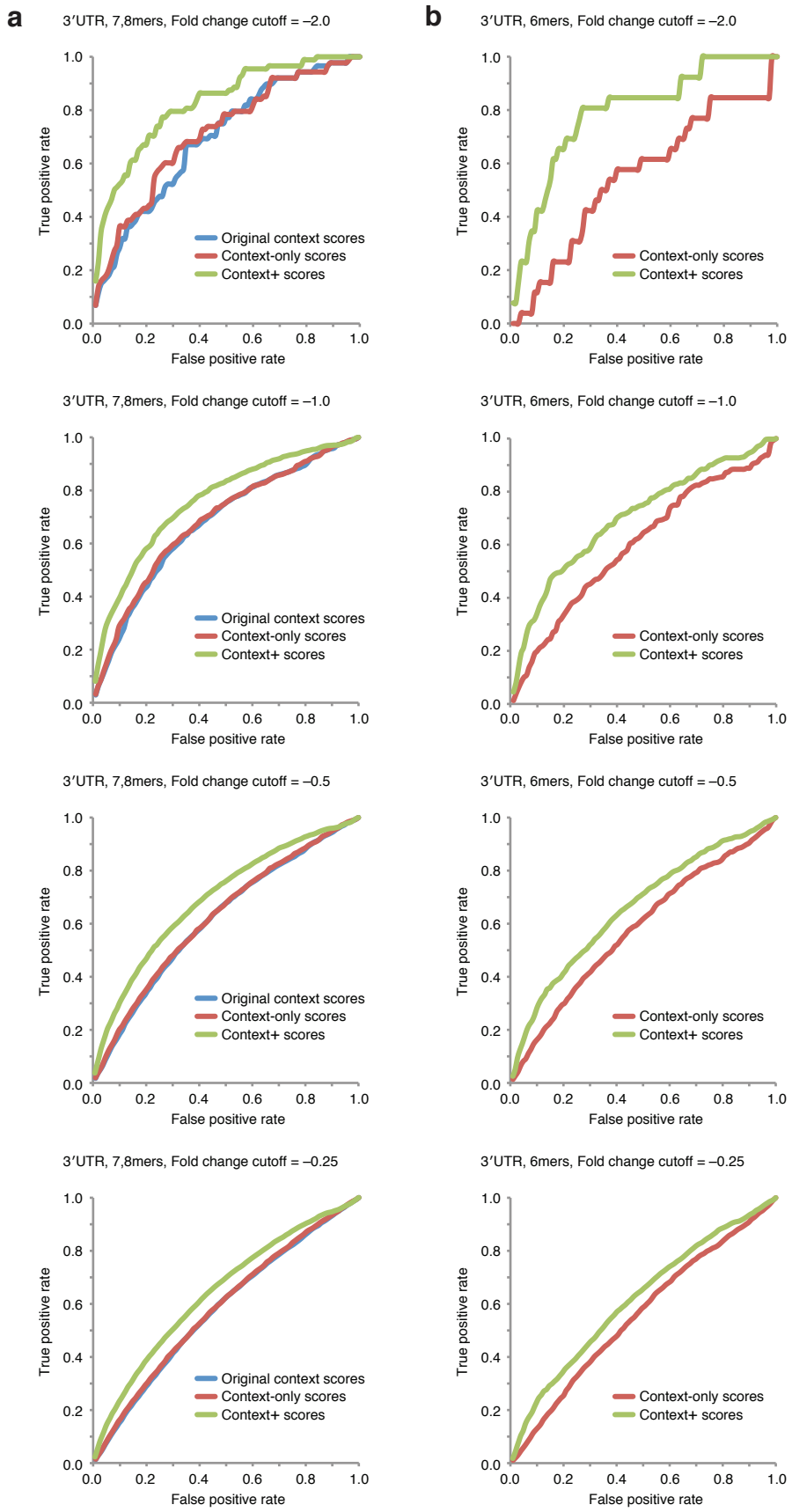
Supplementary Figure 2. Analyses related to **Figure 2**. (**a**) The relationship between predicted SPS and TA in mouse 3′UTRs for miR-23 and the 86 other broadly conserved vertebrate miRNA families (red squares). Otherwise, as in **Figure 2b**. (**b**) The relationship between predicted SPS and TA in *D. melanogaster* 3′UTRs for 94 conserved fly miRNA families (red squares). Otherwise, as in **Figure 2a**. (**c**) The relationship between predicted SPS and TA_{HeLa} for the *lsy-6* site and its mutant derivatives (yellow squares) and for the miR-23 site and its mutant derivatives (red squares). Otherwise, as in **Figure 3a**. (**d–f**) Reporter results presented in **Figure 2** before normalizing to ratios obtained for the non-cognate miRNA, miR-1. Otherwise, as in **Supplementary Figure 1h–l**.



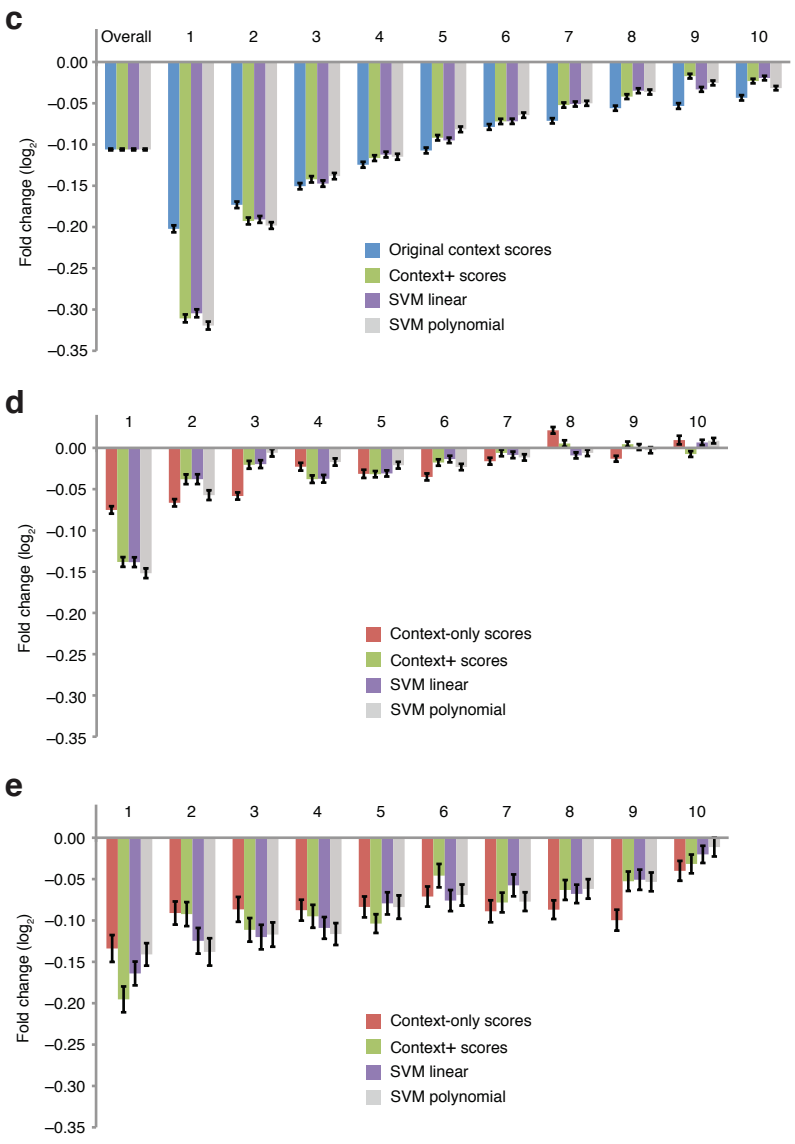
Supplementary Figure 3. Analyses related to **Figure 3**. Impact of TA and SPS on sRNA targeting proficiency of single 3′UTR sites (**a**) and single ORF sites (**b**) to the cognate sRNA, as measured using array data from 74 datasets that passed the motif-enrichment analysis (**Figure 3a**, red squares). Otherwise, as in **Figure 3b,c**.



Supplementary Figure 3 continued. Plots showing the relationship between predicted SPS or TA and the accumulation of mature miRNA after over-expressing the miRNAs from DNA vectors in HEK 293 cells[59]. (**c**) Results from analyses of 14 human miRNAs. Overexpression was calculated as the number of sequencing reads from the most dominant mature miRNA species minus the number of reads found in the mock-transfection control, after normalizing to the reads of endogenous miRNAs that were not overexpressed[59]. For miRNAs marked with asterisks, the most dominant mature miRNA sequence was offset by 1–2 nucleotides with respect to the miRBase annotations, and therefore the predicted SPS and TA values shown differed from those found in **Supplementary Table 7**. These plots show that miRNA accumulation does not decrease with weaker SPS or higher TA. (**d**) Results from analysis of 12 mouse miRNAs. Otherwise, as in **c**.

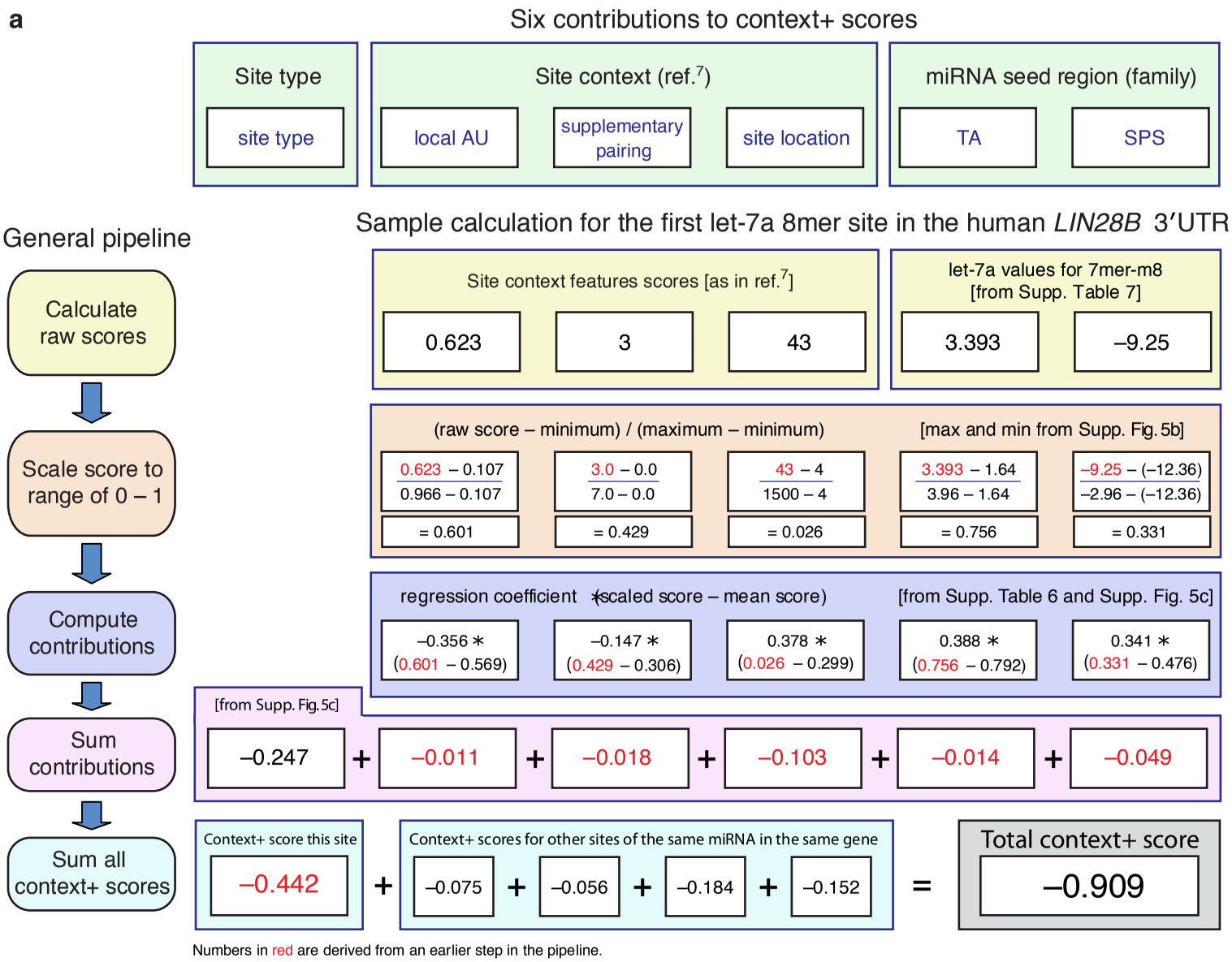


Supplementary Figure 4. Analyses related to **Figure 4**. This page shows ROC curves demonstrating improvements in sRNA target prediction after integrating TA and predicted SPS as features in context+ scores. (**a**) Analyses of mRNAs with 7-8-nucleotide sites in 3′UTRs, performed at four different fold-change cutoffs. (**b**) Analyses of mRNAs with 6mer 3′UTR sites but no larger sites, performed at four different fold-change cutoffs.



Supplementary Figure 4 continued. Performance of the context+ model and SVM regression models with either linear or polynomial kernel. (**c**) Predictions for mRNAs with canonical 7–8-nucleotide 3′UTR sites. Predicted interactions between mRNAs and cognate sRNA were distributed into 10 equally populated bins based on scores generated using the indicated models (key), with the first bin comprising interactions with the most favorable scores. Plotted for each bin is the mean mRNA change on the arrays (error bars, 95% confidence intervals). To perform SVM regression, SVM*light* version 6.02 was used with default parameters[58]. Performance of other SVM kernels (radial basis function and sigmoid tanh) was similar or worse (data not shown). (**d**) Prediction of responsive interactions involving mRNAs with only 3′UTR 6mers sites. Otherwise, as in **c**. (**e**) Prediction of responsive interactions involving mRNAs with at least one 8mer ORF site but no 3′UTR sites. Otherwise, as in **c**.

Supplementary Figure 5. Calculation of context+ scores for TargetScan 6.



Supplementary Figure 5 continued.

b Minimum and maximum values used to scale each parameter.

Site location and type	Local AU content		3′-Supplementary pairing		Site location		TA		SPS	
	Min	Max	Min	Max	Min*	Max	Min	Max	Min	Max
3′UTR 8mer	0.107	0.966	0.0	7.0	4	1500	1.64	3.96	–12.36	–2.96
3′UTR 7mer-m8	0.093	0.990	0.0	7.5	3	1500	1.64	3.96	–12.36	–2.96
3′UTR 7mer-A1	0.122	0.984	0.5	7.5	3	1500	1.64	3.96	–10.00	–0.40
3′UTR 6mer	0.071	0.989	0.0	7.0	3	1500	1.64	3.96	–10.00	–0.40
ORF 8mer	0.033	0.893	0.0	6.5	4	1000	1.64	3.96	–12.36	–2.96
ORF 7mer-m8	0.024	0.914	0.0	7.5	3	1000	1.64	3.96	–12.36	–2.96
ORF 7mer-A1	0.045	0.891	0.0	7.5	3	1000	1.64	3.96	–10.00	–0.40
ORF 6mer	0.024	0.918	0.0	7.5	3	1000	1.64	3.96	–10.00	–0.40

*Although sites within 15 nt of the stop codon were not included as UTR sites because they are in the path of the ribosome as it approaches the stop codon, 3′UTR sites could nonetheless be within 3–4 nucleotides of the polyadenylation site.

c The mean parameters to be used to compute the individual contribution of each determinant in TargetScan6, from analysis of 74 microarrays chosen after motif-enrichment analysis (see main text).

Site location and type	The mean parameter values					
	Fold change	Local AU content	3′-Supplementary pairing	Site location	TA	SPS
3′UTR 8mer	–0.247	0.569	0.306	0.299	0.792	0.476
3′UTR 7mer-m8	–0.120	0.509	0.285	0.289	0.796	0.457
3′UTR 7mer-A1	–0.074	0.555	0.236	0.303	0.794	0.450
3′UTR 6mer	–0.019	0.524	0.306	0.293	0.792	0.437
ORF 8mer	–0.078	0.554	0.334	0.640	0.761	0.438
ORF 7mer-m8	–0.035	0.499	0.288	0.641	0.751	0.422
ORF 7mer-A1	–0.027	0.554	0.290	0.635	0.767	0.415
ORF 6mer	–0.007	0.501	0.289	0.641	0.757	0.404

Supplementary Table 1. Predicted target genes investigated in this study.

Predicted *lsy-6* target genes investigated in this study. Conservation indicates sites present in orthologous UTRs of *C. elegans*, *C. briggsae,* and *C. remanei*. More negative context scores indicate sites predicted to be in more favorable contexts for miRNA recognition[7]. A new tool that precisely maps the 3′ ends of transcripts was applied to *C. elegans*[36], enabling us to check the 3′UTR annotations of these targets. These data indicated that for some of the predicted targets the UTRs end before reaching the *lsy-6* sites. However, this information did not change our conclusions regarding the targeting proficiency of the *lsy-6* miRNA because many of the predicted sites not retained in the worm UTRs must have been retained in those UTRs in HeLa cells — otherwise, repression would not have been observed in **Figure 1g,h**.

Target gene	Site	Sequence name	*C. elegans* site type	Conserved	Context score
cog-1	1	R03C1.3(A)	8mer	Yes	–0.43
cog-1	2	R03C1.3(A)	8mer	Yes	–0.46
hlh-8		C02B8.4	7mer-m8	Yes	–0.26
F55G1.12		F55G1.12	8mer	As a 7mer-A1	–0.51
ptp-1	1	C48D5.2A	7mer-A1	No	–0.14
ptp-1	2	C48D5.2A	8mer	No	–0.52
nsy-1		F59A6.1	7mer-A1	No	–0.12
fkh-8		F40H3.4	7mer-A1	Yes	–0.21
T05C12.8		T05C12.8	7mer-m8	Yes	–0.30
C27H6.9*	1	C27H6.9	8mer	As a 7mer-A1	–0.50
C27H6.9*	2	C27H6.9	8mer	No	–0.43
T23E1.1		T23E1.1	7mer-m8	No	–0.19
aex-4/tag-81		T14G12.2	7mer-m8	Yes	–0.27
glb-1		ZK637.13	7mer-A1	Yes	–0.15
T20G5.9		T20G5.9	7mer-A1	Yes	–0.15
acl-5		R07E3.5	7mer-A1	Yes	+0.01
T04C9.2		T04C9.2	7mer-m8	Yes	–0.23

*Listed as C27H6.3 in ref 23.

Predicted miR-23 target genes investigated in this study. Conservation status and site context scores (calculated for miR-23a) from TargetScan 5.1[7]. More negative scores indicate sites predicted to be in more favorable contexts for miRNA recognition[7].

Target gene	Site	Human site type	Conserved	Context score
LRIG1	1	7mer-A1	Yes	–0.19
LRIG1	2	8mer	Yes	–0.29
WBP4	1	7mer-A1	Yes	–0.28
WBP4	2	7mer-A1	Yes	–0.26
NEK6	1	8mer	Yes	–0.36
NEK6	2	8mer	Yes	–0.43
MAP4K4	1	7mer-m8	Yes	–0.20
MAP4K4	2	8mer	Yes	–0.44
RAP1A	1	7mer-A1	Yes	–0.17
RAP1A	2	7mer-A1	Yes	–0.18
DMN	1	7mer-A1	No	–0.11
DMN	2	8mer	No	–0.32

Supplementary Table 1 continued. Sequences of UTR fragments assayed. Listed are the plasmid name in brackets, gene name, RefSeq accession number (where applicable, longest isoform is cited), and UTR sequence tested with miRNA seed sites underlined. For *lsy-6* and miR-23 targets, the full-length sequence shown is wild-type. For mutant constructs, only the miRNA site (underlined, mutations in uppercase) is shown; the remainder of the UTR sequence was identical to wild-type. For all *cog-1* UTRs assayed in **Supplementary Figure 1b,c**, full-length sequences are shown.

lsy-6 target UTRs:

[pDMG1a] *cog-1*; NM_001027093
cttttaagcgttctacctctccccctcccttcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtactttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttccaaatcatcgtcacttatacaaaaaccaaactccctttttaccgttaaaccatgcccaaatacaaaaaaatttcccatttaattgtacgttttttttctcttcaaattgg
attctaatgacataaatttattagattaa
[pDMG1b] miR-142 sites: aCacTaCa, aCacTaCa; otherwise as for pDMG1a
[pDMG1c] LTA sites: atacGaTa, atacGaTa; otherwise as for pDMG1a

[pDMG2a] *hlh-8*; NM_076966
tttcgaatatggaaaaaactggccagctcctaatttatttgataatgtatgcttctcaatcaacatagtcctcatgatatagtgccttattctcatttttgatgtatcaaatctgtctaa
taatatctactcgatttcaattttcgtttgctcaaaacttaaaaatttgcttgagaattttgcaagagactattagtaagcccattgttaaatgaaacaagtttatcccgcaaacaga
aactatgtgtgaatatgatcaaactataatacaaacacgtaaaaaaaaaattttgaatcataattatcatttgaccactaagccatgcaatgatgaaccaatttcaacttgacattc
acacccagtagtggtatcaattgactcttttacccagtcatcgccattctgtctcatcacatcgatcgtcctcattattggcttgcattctccgaatcctaaaaaaagtgtgggtc
agcggcgtgatggatgggcgtctatgaaaaaaaacgagcccatcggagcccaaatg
[pDMG2b] miR-142 site: aCacTaC; otherwise as for pDMG2a
[pDMG2c] LTA site: atacGaT; otherwise as for pDMG2a

[pDMG3a] F55G1.12; NM_068805
attgattttatttattttaatttaatgaatctcgccggaatgtctgatttgttgcttggtttggtttgaaaattatatacaaaaatagtggattaaatgat
[pDMG3b] miR-142 site: aCacTaCa; otherwise as for pDMG3a
[pDMG3c] LTA site: atacGaTa; otherwise as for pDMG3a

[pDMG4a] *ptp-1*; NM_065331.2
gcttttatccaaaaaatacatatatcgtttttgtttcttcaaattcttcttcccatcgaactctcatgaatcacggatcccgcgaggtgctagctatttttgccttttttctttcttcttttttt
ttattgcatagttaattagctattgttttcctacacaaaactagtcaatgttttaagtaattaaatatcatcatttaatatttcaacaaaaaatctatctcaatgggtcacccgatgtgat
tttcgtaccaattgttccccatcactacgtcataattgtaccacccccccccatctttcatgtacgaaaaatcgcccaaacttgtatgtaaaaaaaacaaaaaagtcctctcaaa
tcatcacaaactttccttctttttcatataaaatgttacagtctgtgtttccattgtacaaaaaaaaagtgtgatcggggaaaaagaaacggggctacatgatcgggaaagtgtg
aacagtttgcttgatttcggaaatcaccaggtttcaaaatttctaaataaaattggaagggaagggaagggaagggaaaagagaaatatatataaatatatagaacccaaga
aaaaatggaacaaaaaaaacggagaatgaggtgtttagatgaacgaaaaaatgcaaatttttagagtttngtcgatccagcgaataaaatcgaaacttngaaaaaaaaaa
agagacactgcctattagaaaacaaaaaaaaacatttcacaaaaaaaattagatgggggtgagggacgaagagtagatcagaaaattgtgaaaaagaaaatatttttacat
cggttttccataacaaaacggtacataaaatgatggagagaatcgaggggaatgatcaaggaatgggacatggcctgccatgagacaaaagacgnttcncaatacaca
antagagcgggggatagaagantagaagaagcttattccagtgnttatggggctatttatatgatgtagaaaaatacaaaaatgtatttttatacatnttccccc
[pDMG4b] miR-142 sites: CacTaCa, aCacTaCa; otherwise as for pDMG4a
[pDMG4c] LTA sites: tacGaTa, atacGaTa; otherwise as for pDMG4a

[pDMG5a] *nsy-1*; NM_062524.6
ttttctgtagtttctctgttctctctctctctctctctctggtcatttttctctcttctagtttttctgtctctttctctcatttttattgtgatatctttttctctctctctctctctctctctctaat
cctctgtttcgtgtacaaagttttcagtttcttaatttgttctctgaaaatcatcaaaccccctccaaaatttgcttgcgtgtagaacttttcattacaaacaaaaaccaaattgctagt
gtctttcctatccaactacattatagagatttctaatctcatgtcaattgtttcatgtatcattctacaaaactcacacacatccgaatcatcataactatcataataataagttttatta
accaaaaataaataaatatatatttcatgatcgttattctccgagcacttccgcatcatcttcggcaacccataagtaaatcatttgtcctgctgcagtcatgagtcgttctttttttc
ggatgagccgtcagagaatgaatcacttttgaaggatgttcgagtttcgacacgatgtgtgaatcgaggagactgtaatttattctttctttgaaactaatattgttacctaccta
tacacataaacgaatccgtcttctgagccacttgccacgtgctcgatggattgtaaaactcggcagtctaatttgtattccgtgttttgatgacctttatagctgaaaatttataaa
tttatagctccaattttcatacaatattcaaaaatatgtaccttgctagcaattttccactagatttatcaataagccgaacaattccacccattactccagccaacaggcaattact
gtctggggtaaagctgacactgtttactgaatctcccatgtagtcgactgtcatctganaattagaaatttctaataacttgaaaattcttttttttttttcaattacatttccatctctaa

tactatagactctgtaatttccatcagcacttccagcaacaatctcatggccatta
[pDMG5b] miR-142 site: CacTaCa; otherwise as for pDMG5a
[pDMG5c] LTA site: tacGaTa; otherwise as for pDMG5a

[pDMG6a] *fkh-8*; NM_062834.2
attgttaatctaaaggttcaaaaactcacatatttttcacacagtgtccaaatttcatttgtacaaaaatacattgttagttagttttcattttcatattcatttttcgtaaacattcaa
[pDMG6b] miR-142 site: CacTaCa; otherwise as for pDMG6a
[pDMG6c] LTA site: tacGaTa; otherwise as for pDMG6a

[pDMG7a] T05C12.8; NM_063322.2
tacaactaaatgtggcaaagcttcttcattgtttgaataatgaaaaacgaatacaaacaactttgaaatcaaaacattaaaacttacaacatttcgttcaataaagtatcatcaaa
agaagaaaacaaaagctgaacatgagaatttgggataaggagcagcagatcggaattatgtgagaagcacgcggaaaacagggatatataaacggggtaaacggga
aaat
[pDMG7b] miR-142 site: aCacTaC; otherwise as for pDMG7a
[pDMG7c] LTA site: atacGaT; otherwise as for pDMG7a

[pDMG8a] C27H6.9; NM_001129395.1
ttggaaaatgtgatgtttttctataaataaatattctcacaactctttttcatgttttatataatacaaaatgcacatcaagcagaaaaatttcaacataaagtttacaccagaagtga
atttagggatgaagaggaaccaaattacgtaaatacaaaagtatcgaaacatgatagat
[pDMG8b] miR-142 sites: aCacTaCa, aCacTaCa; otherwise as for pDMG8a
[pDMG8c] LTA sites: atacGaTa, atacGaTa; otherwise as for pDMG8a

[pDMG9a] T23E1.1; NM_067895.3
aattgagatcaaattgttctttttatatgtatgtactgaaaacaataaagaatttttttgaaattaaaaatttaaagtcttcactcacacccgcctgggaacccccctcttctagccctga
aaacgccttaaattgcacacggagcaagtaaggagtggatgccttgtaggcttaggctcggacttaggcctaggctcaggattaggtttaggcttaggcttagactgggc
gggggaagagagcaaaaataagttccagaaaattcaagaattaaaaaaaggaaataagcctcctaattaggcgaggaggctggcgagaggcgagttttcaatccataat
atccgtgttaagctatttttttttaataaactcttcgaaaatatctactttccctgcaccagtttttctcttccaaaatgttccaaatatgtattgttgagtggcgtaagcaaaacaaag
tcaagtctctagtgaatacaaacacacgctcttcatttttt
[pDMG9b] miR-142 site: aCacTaC; otherwise as for pDMG9a
[pDMG9c] LTA site: atacGaT; otherwise as for pDMG9a

[pDMG10a] *aex-4/tag-81*; NM_076240.5
cttcacaaaaagtgtggtgcgcgcattccacgggctacgaacacatgggtaaactgtacattttcaaatattgttgaaaacttttaatttttcaattttaaattcaaactttatgtttt
aagcaataaaatgatgatttaatccgttatacaaagacatggaaaagttacagttagttttttttttttaagcggtcgttatttatagggggttcgtttaatggtgtcacatactgctttg
cgt
[pDMG10b] miR-142 site: aCacTaC; otherwise as for pDMG10a
[pDMG10c] LTA site: atacGaT; otherwise as for pDMG10a

[pDMG11a] *glb-1*; NM_066573.5
ttgagcctttatattgtatttgaatgagctttgagtattataatgattatctctcttggaaacgtttttgtacaaaataaacaaag
[pDMG11b] miR-142 site: CacTaCa; otherwise as for pDMG11a
[pDMG11c] LTA site: tacGaTa; otherwise as for pDMG11a

[pDMG12a] T20G5.9; NM_066860.2
gcaacgattaaatatagattctacctctctgtttcatttcatgtgcgatagtttcagataattatttatttttatattttgtattttatgaacgggttcgatacttgtctttttttcggttggaat
gtacaaaaatacacagaatacacgaattga
[pDMG12b] miR-142 site: CacTaCa; otherwise as for pDMG12a
[pDMG12c] LTA site: tacGaTa; otherwise as for pDMG12a

[pDMG13a] *acl-5*; NM_001047817.1
agtttttgatgtacaaaactagccaatttttttgtatcagatcttttattgattgtttacgtttgaacggttccatttgccaaa
[pDMG13b] miR-142 site: CacTaCa; otherwise as for pDMG13a
[pDMG13c] LTA site: tacGaTa; otherwise as for pDMG13a

[pDMG14a] T04C9.2; NM_065904.1
cgcgataactttgtttcggctcctatacaaatttggttatttttgttggtcgctccaacatttttttcgtcctcatccgagccatgacttctcttctccttttcctctatttcgtctcaaact
tccgttctttttacctaatcatcattattagccccatccttatcatcttctggaacccacatcgtcatcttcggtttctttcttttgaggcacaagcaacaactacttttctcgcatctt
ctctcctccagcttctcttctttatgagccgggttaggggctcttcgaaaattgtttccactcggctgccttcgtgttttcgacgtgcccgaacttgctcaaaaccgaagctcac
gcatcgttaggtaacgaaagaattacacgtaggagggacgcactgccgtttgattcttatctgtcatcgtcaggattgttgcaacctcttcaatcctccggatgtgcgttgatt
cccgcacgattagacaatttgttgtg
[pDMG14b] miR-142 site: aCacTaC; otherwise as for pDMG14a
[pDMG14c] LTA site: atacGaT; otherwise as for pDMG14a

miR-23 target UTRs

[pAG247] LRIG1; NM_015541.2
gataaaagcaaatgtggccttctcagtatcattcgattgctatttgagacttttaaattaaggtaaaggctgctggtgttggtacctgtggatttttctatactgatgttttcgttttg
ccaatataatgagtattacattggccttgggggacagaaaggaggaagttctgacttttcagggctaccttatttctactaaggacccagagcaggcctgtccatgccattcc
ttcgcacagatgaaactgagctgggactggaaaggacagcccttgacctgggttctgggtataatttgcacttttgagactggtagctaaccatcttatgagtgccaatgtgt
catttagtaaaacttaaatagaaacaaggtccttcaaatgttcctttggccaaaagctgaagggagttactgagaaaatagttaacaattactgtcaggtgtcatcactgttcaa
aaggtaagcacatttagaattttgttcttgacagttaactgactaatcttacttccacaaaatatgtgaatttgctgcttctgagaggcaatgtgaaagagggagtattacttttat
gtacaaagttatttatttatagaaattttggtacagtgtacattgaaaaccatgtaaaatattgaag
[pCS247] miR-CGCG sites: aCgCgaa, aaCgCgaa; otherwise as for pAG247

[pAG249] WBP4; NM_007187.3
catgcttttaggacagaatggagacttatacacccaaagtttatctgtgtttgtttgtaagtattatgatgctaaaaatttagatttattctaaatgtatttgatgtgaattaaaataaa
tatttttttcatgtgaaatttatttttggttcctaaaatggaagcctaccacattgcattgtaatacagtgtattatgttcagtgtctaaaaactgctaattaagtcataatttaagatgcta
tgtatctgttatttaaaacatggagaaacagggccttttattccattcatattcataagagcatatttatcctgcattgaaaatgcattacttttgcacattgatattaactgttgtccaa
caaataagtatcggagtacgtgagaatattccc
[pCS249] miR-CGCG sites: aCgCgaa, aCgCgaa; otherwise as for pAG249

[pAG250] NEK6; NM_001145001.2
aacagctaagaccacagggttcagcaggttccccaaaaggctgcccagccttacagcagatgctgaaggcagagcagctgagggaggggcgctggccacatgtcac
tgatggtcagattccaaagtcctttctttatactgttgtggacaatctcagctgggtcaataagggcaggtggttcagcgagccacggcagccccctgtatctggattgtaat
gtgaatctttagggtaattcctccagtgacctgtcaaggcttatgctaacaggagacttgcaggagaccgtgtgatttgtgtagtgagcctttgaaaatggttagtaccgggtt
cagtttagttcttggtatcttttcaatcaagctgtgtgcttaatttactctgttgtaaagggataaagtggaaatcatttttttccgtggagtggtgattctgctaacatttttatctacg
ttttataacttggtgagtgacgatgagagccctgcacctggccagagtgtcacaggcaaaaggcatcgggaagcaggagcatcttcttggcagccaggctgggccatctt
ctcctggacacctgctgtgtaccaggaacttcgtcacctccttgaatgctggcggttcatttcatgatcagtgttaagcattttcctccatgggaaggaagcatgggatataga
aaagcgaagggctgtcctttacaaattctggttctgcaacttcctagcgtgacttttgggcttgggcaagtttcttagccgttctgagccttcatttcctcatctgtacaatgagatt
aatagtacctatcatctaccttcaggattgctgacagacagaatttgaaataaaatatgcaagttagctaatacaaaaagtagatgatccaaaaatggtagccactcaccctc
acaaactgaagtccatggaccacggaagtcgagaattaatgtacacctgtatcatgtgtaggaaaccagaaatgtgttccttatttcttgttcccaaacaggattaactgtgaa
gactaatttataatgtgaacctaagaaaactccacctctgaaggaaatcatttgaattttgtttttgtacgtaaagttaaccttccaattgtctgagctgtcgtcactgacttcatg
acagtctggccctccagacaagagcagcgctggcatcgggcaggtgattcctgacacct
[pCS250] miR-CGCG sites: aaCgCgaa, aaCgCgaa; otherwise as for pAG250

[pAG252] MAP4K4; NM_145686.2
tttgggattgagcatcatactggaaagcaaacacctttcctccagctccagaattccttgtctctgaatgactctgtcttgtgggtgtctgacagtggcgacgatgaacatgcc
gttggttttattggcagtgggcacaaggaggtgagaagtggtggtaaaaggagcggagtgctgaagcagagagcagatttaatatagtaacattaacagtgtatttaattg
acatttctttttttgtaatgtgacgatatgtggacaaagaagaagatgcaggtttaagaagttaatatttataaaatgtgaaagacacagttactaggataactttttttgtgggtgg
ggcttgggagatggggtggggtgggttaaggggtcccattttgtttctttggatttggggtgggggtcctggccaagaactcagtcatttttctgtgtaccaggttgcctaaat
ca
[pCS252] miR-CGCG sites: aaCgCga, aaCgCgaa; otherwise as for pAG252

[pAG253] RAP1A; NM_001010935.1
gccagattacaggaatgaagaactgttgcctaattggaaagtgccagcattccagacttcaaaaataaaaaatctgaagaggcttctcctgttttatatattatgtgaagaatttt
agatcttatattggtttgcacaagttccctggagaaaaaaattgctctgtgtatatctcttggaaaataagacaatagtatttctcctttgcaatagcagttataacagatgtgaaa

atatacttgactctaatatgattatacaaaagagcatggatgcatttcaaatgttagatattgctactataatcaaatgatttcatattgatctttttatcatgatcctccctatcaagc
actaaaaagttgaaccattatactttatatctgtaatgatactgattatgaaatgtcccctgaa
[pCS253] miR-CGCG sites: aCgCgaa, aCgCgaa; otherwise as for pAG253

[pAG260] SYNM; NM_145728.2
cagacagagatgtgctgattttgttttagctgtaacaggtaatggtttttggatagatgattgactggtgagaatttggtcaaggtgacagcctcctgtctgatgacaggacag
actggtggtgaggagtctaagtgggctcagtttgatgtcagtgtctgggctcatgacttgtaaatggaagctgatgtgaacaggtaattaatattatgacccacttctatttactt
tgggaaatatcttggatcttaattatcatctgcaagtttcaagaagtattctgccaaaagtatttacaagtatggactcatgagctattgttggttgctaaatgtgaatcacgcgg
gagtgagtgtgcccttcacactgtgacattgtgacattgtgacaagctccatgtcctttaaaatcagtcactctgcacacaagagaaatcaacttcgtggttggatggggccg
gaacacaaccagtctttttgtatttattgttactgagacaaaacagtactcactgagtgtttttcagtttcctactggtggttttga
[pCS260] miR-CGCG sites: aCgCgaa, aaCgCgaa; otherwise as for pAG260

cog-1 UTR sequences assayed in **Supplementary Figure 1b,c**.

[pDMG1a] wild-type sites
cttttaagcgttctacctctccccctccctcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttttccaaatcatcgtcacttatacaaaaaccaaactcccttttaccgttaaaccatgcccaaatacaaaaaaatttcccatttaattgtacgtttttttctcttcaaattgg
attctaatgacataaatttattagattaa

[pDMG1b] 1st, 2nd site mismatch
cttttaagcgttctacctctccccctccctcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttttccaaatcatcgtcacttaCacTaCaaccaaactcccttttaccgttaaaccatgcccaaaCacTaCaaatttcccatttaattgtacgtttttttctcttcaaat
tggattctaatgacataaatttattagattaa

[pDMG1d] 1st site mismatch
cttttaagcgttctacctctccccctccctcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttttccaaatcatcgtcacttaCacTaCaaccaaactcccttttaccgttaaaccatgcccaaatacaaaaaaatttcccatttaattgtacgtttttttctcttcaaattg
gattctaatgacataaatttattagattaa

[pDMG1e] 2nd site mismatch
cttttaagcgttctacctctccccctccctcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttttccaaatcatcgtcacttatacaaaaaccaaactcccttttaccgttaaaccatgcccaaaCacTaCaaatttcccatttaattgtacgtttttttctcttcaaattg
gattctaatgacataaatttattagattaa

[pDMG1f] 1st site GU2
cttttaagcgtctacctctccccctccctcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttccttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttttccaaatcatcgtcacttatacaaGaaccaaactcccttttaccgttaaaccatgcccaaatacaaaaaaatttcccatttaattgtacgtttttttctcttcaaattg
gattctaatgacataaatttattagattaa

[pDMG1g] 1st site GU3
cttttaagcgtctacctctccccctccctcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttttccaaatcatcgtcacttatacaGaaaccaaactcccttttaccgttaaaccatgcccaaatacaaaaaaatttcccatttaattgtacgtttttttctcttcaaattgg
attctaatgacataaatttattagattaa

[pDMG1h] 1st site GU4
ctttaagcgttctacctctccccctcccttcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttccaaatcatcgtcacttatacGaaaaaccaaactccctttaccgttaaaccatgcccaaatacaaaaaaatttcccatttaattgtacgttttttctcttcaaattgg
attctaatgacataaatttattagattaa

[pDMG1i] 1st site GU6
ctttaagcgttctacctctccccctcccttcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttccaaatcatcgtcacttatGcaaaaaccaaactccctttaccgttaaaccatgcccaaatacaaaaaaatttcccatttaattgtacgttttttctcttcaaattgg
attctaatgacataaatttattagattaa

[pDMG1j] 1st site GU8
ctttaagcgttctacctctccccctcccttcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttccaaatcatcgtcacttGtacaaaaaccaaactccctttaccgttaaaccatgcccaaatacaaaaaaatttcccatttaattgtacgttttttctcttcaaattgg
attctaatgacataaatttattagattaa

[pDMG1k] 1st site GU2,6
ctttaagcgttctacctctccccctcccttcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttccaaatcatcgtcacttatGcaaGaaccaaactccctttaccgttaaaccatgcccaaatacaaaaaaatttcccatttaattgtacgttttttctcttcaaattg
gattctaatgacataaatttattagattaa

[pDMG1l] 1st site GU6,8
ctttaagcgttctacctctccccctcccttcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttccaaatcatcgtcacttGtGcaaaaaccaaactccctttaccgttaaaccatgcccaaatacaaaaaaatttcccatttaattgtacgttttttctcttcaaattgg
attctaatgacataaatttattagattaa

[pDMG1m] 2nd site GU2,6
ctttaagcgttctacctctccccctcccttcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttccaaatcatcgtcacttatacaaaaaccaaactccctttaccgttaaaccatgcccaaatGcaaGaaatttcccatttaattgtacgttttttctcttcaaattg
gattctaatgacataaatttattagattaa

[pDMG1n] 2nd site GU6,8
ctttaagcgttctacctctccccctcccttcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttccaaatcatcgtcacttatacaaaaaccaaactccctttaccgttaaaccatgcccaaGtGcaaaaaatttcccatttaattgtacgttttttctcttcaaattg
gattctaatgacataaatttattagattaa

[pDMG1o] 1st, 2nd site GU2,6
ctttaagcgttctacctctccccctcccttcaaccgagtgtattattcccccaatttgtttgcaatttttttcctgaagccctctaagaaaatccaaaatcatgacctacttccgtcttt
acacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatttt
ttttctttttttttccaaatcatcgtcacttatGcaaGaaccaaactccctttaccgttaaaccatgcccaaatGcaaGaaatttcccatttaattgtacgttttttctcttcaaatt
ggattctaatgacataaatttattagattaa

[pDMG1p] 1st, 2nd site GU6,8
ctttaagcgttctacctctccccctcccttcaaccgagnngtattattcccccaatttgtttgcaatttttttcctgaagccctctaanaaaatccaaaatcatgacctacttccgtc
tttacacctgattacctgaataccaacaccccacacagatgccatgatctctcgtcttttctcgtacttttgtataatttttttttcttaatttttttgcatgttttcccatagttatagccatt
ttttttttctttttttttccaaatcatcgtcacttGtGcaaaaaccaaactccctttaccgttaaaccatgcccaaGtGcaaaaaatttcccatttaattgtacgttttttctcttcaaat
tggattctaatgacataaatttattagattaa

Supplementary Table 2. Relationship between mean mRNA repression and either TA or predicted SPS for the indicated site types, from analysis of microarrays chosen after motif-enrichment analysis.

Site location and type	Multiple linear regression			Simple linear regression			
	Multiple R^2	*P* value		TA_{HeLa}		SPS	
		TA_{HeLa}	SPS	R^2	*P* value	R^2	*P* value
3′UTR 8mer	0.189	0.032	0.012	0.113	0.0034	0.134	0.0013
3′UTR 7mer-m8	0.320	9.3×10^{-5}	0.013	0.258	3.8×10^{-6}	0.156	5.0×10^{-4}
3′UTR 7mer-A1	0.442	4.6×10^{-5}	2.2×10^{-5}	0.280	1.3×10^{-6}	0.294	6.0×10^{-7}
3′UTR 6mer	0.345	2.3×10^{-4}	0.0013	0.241	8.9×10^{-6}	0.206	4.8×10^{-5}
ORF 8mer	0.350	2.7×10^{-6}	0.087	0.323	1.3×10^{-7}	0.112	0.0036
ORF 7mer-m8	0.306	7.4×10^{-5}	0.032	0.259	3.7×10^{-6}	0.132	0.0014
ORF 7mer-A1	0.298	1.8×10^{-5}	0.14	0.276	1.5×10^{-6}	0.089	0.0099
ORF 6mer	0.287	0.0031	0.0017	0.179	1.7×10^{-4}	0.193	9.1×10^{-5}
5′UTR 8mer	0.006	0.52	0.81	0.006	0.54	0.000	0.97
5′UTR 7mer-m8	0.000	0.91	0.97	0.000	0.91	0.000	0.99
5′UTR 7mer-A1	0.022	0.42	0.49	0.016	0.29	0.013	0.33
5′UTR 6mer	0.016	0.33	0.47	0.009	0.42	0.003	0.65

Supplementary Table 3. Multiple linear regression statistics for miRNA target prediction for context+ scores, using 11 microarray datasets previously used to build the TargetScan context score model.

Site location and type	Multiple linear regression intercept and coefficients (*P* value)					
	Intercept	Local AU content	3′-Supplementary pairing	Site location	TA_{HeLa}	SPS
3′UTR 8mer	–0.674 (0.003)	–0.447 (2×10^{-7})	–0.006 (1)	0.312 (1×10^{-7})	0.431 (0.1)	0.416 (1×10^{-5})
3′UTR 7mer-m8	–0.309 (0.02)	–0.443 (2×10^{-22})	–0.186 (0.01)	0.213 (4×10^{-13})	0.300 (0.06)	0.310 (2×10^{-12})
3′UTR 7mer-A1	–0.596 (1×10^{-7})	–0.226 (3×10^{-8})	–0.111 (0.07)	0.119 (3×10^{-6})	0.681 (6×10^{-7})	0.163 (0.002)
3′UTR 6mer	–0.350 (7×10^{-10})	–0.164 (5×10^{-16})	–0.023 (0.4)	0.084 (2×10^{-12})	0.431 (2×10^{-10})	0.106 (7×10^{-6})
ORF 8mer	–0.317 (0.02)	–0.191 (2×10^{-4})	–0.048 (0.5)	0.117 (2×10^{-7})	0.289 (0.07)	0.134 (0.007)
ORF 7mer-m8	–0.110 (0.2)	–0.139 (1×10^{-5})	–0.042 (0.4)	0.052 (9×10^{-5})	0.149 (0.1)	0.019 (0.5)
ORF 7mer-A1	–0.077 (0.2)	–0.077 (0.01)	–0.050 (0.2)	0.052 (3×10^{-5})	0.089 (0.3)	0.042 (0.2)
ORF 6mer	–0.104 (0.01)	–0.059 (0.002)	–0.016 (0.6)	0.025 (8×10^{-4})	0.144 (.004)	0.007 (0.7)

Supplementary Table 4. Multiple linear regression statistics for miRNA target prediction for context-only scores, using 11 microarrays previously used to build the TargetScan context score.

Site location and type	Multiple linear regression intercept and coefficients (*P* value)			
	Intercept	Local AU content	3′-Supplementary pairing	Site location
3′UTR 8mer	–0.150 (0.03)	–0.376 (1×10^{-5})	–0.076 (0.6)	0.290 (1×10^{-6})
3′UTR 7mer-m8	0.061 (0.08)	–0.395 (7×10^{-18})	–0.230 (0.002)	0.198 (3×10^{-11})
3′UTR 7mer-A1	0.019 (0.5)	–0.188 (3×10^{-6})	–0.163 (0.008)	0.100 (9×10^{-5})
3′UTR 6mer	0.045 (0.002)	–0.143 (9×10^{-13})	–0.043 (0.1)	0.074 (8×10^{-10})
ORF 8mer	–0.022 (0.6)	–0.189 (2×10^{-4})	–0.057 (0.4)	0.113 (6×10^{-7})
ORF 7mer-m8	0.019 (0.4)	–0.138 (1×10^{-5})	–0.043 (0.3)	0.052 (8×10^{-5})
ORF 7mer-A1	0.009 (0.7)	–0.073 (0.01)	–0.054 (0.2)	0.051 (3×10^{-5})
ORF 6mer	0.017 (0.4)	–0.054 (0.005)	–0.014 (0.6)	0.025 (8×10^{-4})

Supplementary Table 5. Context+ parameters to be used for improved target predictions in TargetScan 6. Analysis is with 74 filtered representative array datasets (**Supplementary Data 1**).

Site location and type	Multiple linear regression intercept and coefficients (*P* value)					
	Intercept	Local AU content	3′-Supplementary pairing	Site location	TA	SPS
3′UTR 8mer	–0.583 (7×10^{-25})	–0.356 (1×10^{-16})	–0.147 (0.03)	0.378 (2×10^{-45})	0.388 (1×10^{-10})	0.341 (6×10^{-17})
3′UTR 7mer-m8	–0.243 (6×10^{-23})	–0.366 (1×10^{-74})	–0.139 (2×10^{-5})	0.212 (4×10^{-63})	0.243 (4×10^{-20})	0.207 (3×10^{-28})
3′UTR 7mer-A1	–0.298 (2×10^{-28})	–0.187 (1×10^{-17})	–0.048 (0.1)	0.164 (6×10^{-39})	0.239 (5×10^{-16})	0.220 (2×10^{-26})
3′UTR 6mer	–0.114 (1×10^{-19})	–0.084 (7×10^{-15})	–0.048 (0.002)	0.094 (3×10^{-51})	0.106 (7×10^{-15})	0.098 (1×10^{-22})
ORF 8mer	–0.260 (1×10^{-18})	–0.147 (5×10^{-8})	–0.035 (0.3)	0.122 (1×10^{-24})	0.203 (2×10^{-11})	0.095 (1×10^{-4})
ORF 7mer-m8	–0.095 (1×10^{-11})	–0.074 (6×10^{-7})	–0.033 (0.1)	0.056 (5×10^{-19})	0.071 (2×10^{-7})	0.043 (8×10^{-4})
ORF 7mer-A1	–0.164 (2×10^{-21})	–0.014 (0.4)	–0.041 (0.07)	0.063 (2×10^{-21})	0.130 (1×10^{-14})	0.040 (0.007)
ORF 6mer	–0.054 (3×10^{-10})	0.004 (0.7)	–0.035 (0.005)	0.028 (2×10^{-14})	0.037 (7×10^{-6})	0.023 (0.004)

References

1. Ambros, V. The functions of animal microRNAs. *Nature* 431, 350–355 (2004).
2. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297 (2004).
3. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233 (2009).
4. Lewis, B.P., Burge, C.B. & Bartel, D.P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20 (2005).
5. Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. Principles of microRNA-target recognition. *PLoS Biol.* 3, e85 (2005).
6. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* 37, 495–500 (2005).
7. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* 27, 91–105 (2007).
8. Friedman, R.C., Farh, K.K., Burge, C.B. & Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105 (2009).
9. Shin, C. *et al.* Expanding the microRNA targeting code: functional sites with centered pairing. *Mol. Cell* 38, 789–802 (2010).
10. Lim, L.P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433, 769–773 (2005).
11. Krützfeldt, J. *et al.* Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* 438, 685–689 (2005).
12. Farh, K.K. *et al.* The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science* 310, 1817–1821 (2005).
13. Giraldez, A.J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312, 75–79 (2006).
14. Nielsen, C.B. *et al.* Determinants of targeting by endogenous and exogenous microRNAs and siRNAs. *RNA* 13, 1894–1910 (2007).
15. Robins, H., Li, Y. & Padgett, R.W. Incorporating structure to predict microRNA targets. *Proc. Natl. Acad. Sci. USA* 102, 4006–4009 (2005).
16. Zhao, Y., Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets *Hand2* during cardiogenesis. *Nature* 436, 214–220 (2005).
17. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nat. Genet.* 39, 1278–1284 (2007).
18. Long, D. *et al.* Potent effect of target structure on microRNA function. *Nat. Struct. Mol. Biol.* 14, 287–294 (2007).
19. Hammell, M. *et al.* mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat. Methods* 5, 813–819 (2008).
20. Saetrom, P. *et al.* Distance constraints between microRNA target sites dictate efficacy and cooperativity. *Nucleic Acids Res.* 35, 2333–2342 (2007).
21. Kedde, M. *et al.* RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* 131, 1273–1286 (2007).
22. Ruby, J.G. *et al.* Evolution, biogenesis, expression, and target predictions of a substantially expanded set of *Drosophila* microRNAs. *Genome Res.* 17, 1850–1864 (2007).

23. Didiano, D. & Hobert, O. Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat. Struct. Mol. Biol.* 13, 849–851 (2006).
24. Ui-Tei, K., Naito, Y., Nishi, K., Juni, A. & Saigo, K. Thermodynamic stability and Watson-Crick base pairing in the seed duplex are major determinants of the efficiency of the siRNA-based off-target effect. *Nucleic Acids Res.* 36, 7100–7109 (2008).
25. Franco-Zorrilla, J.M. *et al.* Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* 39, 1033–1037 (2007).
26. Ebert, M.S., Neilson, J.R. & Sharp, P.A. MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nat. Methods* 4, 721–726 (2007).
27. Anderson, E.M. *et al.* Experimental validation of the importance of seed complement frequency to siRNA specificity. *RNA* 14, 853–861 (2008).
28. Arvey, A., Larsson, E., Sander, C., Leslie, C.S. & Marks, D.S. Target mRNA abundance dilutes microRNA and siRNA activity. *Mol. Syst. Biol.* 6, 363 (2010).
29. Rodriguez, A. *et al.* Requirement of bic/microRNA-155 for normal immune function. *Science* 316, 608–611 (2007).
30. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* 455, 64–71 (2008).
31. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21 (2002).
32. Xia, T. *et al.* Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* 37, 14719–14735 (1998).
33. Guo, H., Ingolia, N.T., Weissman, J.S. & Bartel, D.P. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835–840 (2010).
34. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* 455, 58–63 (2008).
35. Lall, S. *et al.* A genome-wide map of conserved microRNA targets in *C. elegans*. *Curr. Biol.* 16, 460–471 (2006).
36. Jan, C.H., Friedman, R.C., Ruby, J.G. & Bartel, D.P. Formation, regulation and evolution of *Caenorhabditis elegans* 3′ UTRs. *Nature* 469, 97–101 (2011).
37. Didiano, D. & Hobert, O. Molecular architecture of a miRNA-regulated 3′ UTR. *RNA* 14, 1297–1317 (2008).
38. Huesken, D. *et al.* Design of a genome-wide siRNA library using an artificial neural network. *Nat. Biotechnol.* 23, 995–1001 (2005).
39. Schwarz, D.S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115, 199–208 (2003).
40. Khvorova, A., Reynolds, A. & Jayasena, S.D. Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115, 209–216 (2003).
41. Bartel, D.P. & Chen, C.Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat. Rev. Genet.* 5, 396–400 (2004).
42. Stark, A., Brennecke, J., Bushati, N., Russell, R.B. & Cohen, S.M. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′UTR evolution. *Cell* 123, 1133–1146 (2005).
43. Seitz, H. Redefining microRNA targets. *Curr. Biol.* 19, 870–873 (2009).
44. Ameres, S.L., Martinez, J. & Schroeder, R. Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* 130, 101–112 (2007).

45. Parker, J.S., Parizotto, E.A., Wang, M., Roe, S.M. & Barford, D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol. Cell* 33, 204–214 (2009).
46. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. & Burge, C.B. Prediction of mammalian microRNA targets. *Cell* 115, 787–798 (2003).
47. Pruitt, K.D., Katz, K.S., Sicotte, H. & Maglott, D.R. Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. *Trends Genet.* 16, 44–47 (2000).
48. Imanishi, T. *et al.* Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol.* 2, e162 (2004).
49. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* 409, |860–921 (2001).
50. Kent, W.J. BLAT–the BLAST-like alignment tool. *Genome Res.* 12, 656–664 (2002).
51. Okazaki, Y. *et al.* Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420, 563–573 (2002).
52. Waterston, R.H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562 (2002).
53. Ruby, J.G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* 127, 1193–1207 (2006).
54. Griffiths-Jones, S., Saini, H.K., van Dongen, S. & Enright, A.J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36, D154–D158 (2008).
55. Rhead, B. *et al.* The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.* 38, D613–D619 (2010).
56. Landgraf, P. et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129, 1401-14 (2007).
57. Doench, J.G. & Sharp, P.A. Specificity of microRNA target selection in translational repression. *Genes Dev* 18, 504-11 (2004).
58. Joachims, T. Optimizing Search Engines Using Clickthrough Data. *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)* (2002).
59. Chiang, H.R. et al. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. *Genes Dev* 24, 992-1009 (2010).

Chapter 3

Discussion

When reports in the literature reach conclusions that contradict widely accepted models, quickly dismissing them can preclude appropriate follow-up investigation. The paper highlighting the poor proficiency of *lsy-6* could have easily fallen into this category, but in addressing this discrepancy, the work in this thesis lead to discovering the influence of SPS and TA on miRNA targeting. From analyses of the 74 array datasets, SPS and TA had correlation coefficients near or above those of local AU content for three site types in 3′ UTRs (Supplementary Table 5, Chapter 2). Future miRNA target prediction could uncover additional features of similar importance, though in even the most up-to-date analyses searching for new features, SPS and TA still rank near the top of features for modeling targeting in human cells (V. Agarwal, personal communication). Starting with the intriguing worm data on *lsy-6* targeting, it was fortunate to find new features with broader relevance to miRNAs and siRNAs, and hopefully other such examples will emerge. Even after this study and others on *lsy-6*, we're still far from understanding how *cog-1* is so robustly repressed. An explanation could lie in its 3′ UTR structure, or in some unknown *trans* factor that is conserved between worms and HeLa cells, since the repression response of *cog-1* compared to other predicted targets is so similar in these two cell types. It might be interesting to test if *cog-1* is still repressed by *lsy-6* in reporter assays in *Drosophila* S2 cells, for further comparison. Bashing the *cog-1* 3′ UTR and testing these mutants in reporter assays in human cells, together with secondary structure prediction of these mutant UTRs could reveal structural motifs important for repression.

Integrating SPS and TA into target prediction

One of the more practical benefits of considering SPS and TA is the improvement in miRNA target prediction. In applying these features to a target prediction program like TargetScan, the predicted target rankings—from estimates of their repression levels by a given a miRNA—will be readjusted. For each predicted target site, TargetScan scores different features like site type, local AU content, and now, SPS and TA, which are each scaled using different coefficients. These scores are added to generate a context score that quantitatively predicts how much repression a site would confer in HeLa cells with transfected miRNA, which are the conditions under which these features were modeled (see Chapter 2). These rankings are also believed to predict relative repression for targets under different experimental or biological conditions.

Ranking predicted targets of a single miRNA or miRNA seed family, which has the same TA and SPS for all its sites, will not change (or change only slightly due to rescaling coefficients for other features after integrating SPS and TA). There are two practical cases, however, in which the new rankings will significantly impact target prediction. The first is in predicting which sites for different miRNAs in a single mRNA of interest are the most likely to be effective. The sites to miRNAs with more favorable TA and SPS values will tend to impact target expression to a greater extent. The second case is when considering all the miRNA targets in a particular cell type, those containing sites for the higher expressed miRNAs with favorable SPS and TA scores most likely represent the most robust and relevant targeting interactions in the cell.

TA and SPS in endogenous targeting

To see if these improvements in target prediction modeled on miRNA overexpression in human cell lines could apply to datasets monitoring the effects of endogenous targeting, the context+ model was compared to context-only in publically available datasets from mouse embryonic stem cells (mESCs) and *Drosophila* S2 cells. For mESCs, when microarray data from wild-type cells was compared to DGCR8 knockout cells (which should block expression of miRNAs that depend on the canonical biogenesis pathway), the context+ model demonstrated improvements over the context-only model through ROC curve analysis, at three log fold-change cutoff values (cutoffs of derepression of target mRNAs in knockouts vs. wild-type)(Figure 1a)(J.-W. Nam, unpublished data). This analysis computes the number of true positive target events vs. false positive target events at indicated expression fold-change cutoffs, with larger area under curve (AUC) values indicating better predictions. A similar trend is seen in *Drosophila* S2 cells in which endogenous miRNA pools are depleted through Drosha siRNA knock-down (Figure 1b)(J.-W. Nam, unpublished data). In contrast, no improvement is seen in comparing data from a maternal-zygotic Dicer knockout from zebrafish to wild-type, 24 hours post-fertilization (Figure 1c)(J-W. Nam, unpublished data). This is due to the fact that targeting in zebrafish embryos is dominated by miR-430 (Giraldez et al., 2005), so the majority of productive seed matches have the same SPS and TA and consideration of these features cannot improve the predictions beyond context-only.

These improvements in target predictions in mouse and *Drosophila* lend further support to SPS and TA being important features in targeting by endogenous miRNAs. One reason for the baseline AUC values being fairly low (a value of 0.5 means an equal chance of predicting a true

positive target or a false positive target), and the gains from context+ being small include the fact that these datasets profile effects by endogenous miRNAs expressed at lower levels than in the transfections, leading to less repression. Another reason is that these datasets profiling cells where nearly the whole miRNA pathway has been perturbed (by knocking down/out central processing components) are inherently noisy. A more comprehensive endogenous picture of the importance of SPS and TA will emerge once more large-scale datasets from individual miRNA knockouts are available for comparison, especially in cell types in which the miRNA is normally highly expressed.

RNA pairing stability in miRNA targeting

The concept of RNA pairing stability influencing gene output is an old one in biology. To give one illustrative example from bacteria, intrinsic (or non-factor mediated) transcription termination is controlled in part by the production of a GC-rich hairpin toward the end of the nascent RNA transcript followed by a tract of uracils. The strong hairpin structure causes RNA polymerase to pause, and then the weak RNA pairing between the U-tract in the transcript and complementary A's in the template DNA promotes release. Thus strong RNA–RNA pairing in one structural motif, and weak DNA:RNA pairing in another control gene expression of these intrinsic terminators.

What are ways in which SPS could increase miRNA targeting proficiency? It has been shown that Argonaute can greatly increase the affinity of a guide ~22-nt RNA for target RNA (compared to the RNA–RNA pairing alone), up to ~300-fold (Parker et al., 2009). This is believed to result from a reduction in entropy from Ago contacts with the backbone of the guide

strand, pre-organizing the seed region for binding to an incoming target RNA. In contrast, Ago makes interactions between the 3′ end of the guide and the target disfavored (Parker et al., 2009), consistent with this part of the small RNA being disordered in crystal structures (Wang et al., 2008), and less structurally constrained than the seed region in solution (Lambert et al., 2011). Therefore, besides providing target specificity, seed pairing is probably accomplishing two major things: (1) serving as a nucleation site for pairing between the miRNA and a target, and (2) maintaining favorable pairing long enough to lead to mRNA repression. Since a seed match is only comprised of 6–7 base pairs, which would likely melt in isolation under physiological conditions, Ago's job to make this interaction stable could presumably be aided by even a few additional kcals of thermostability. In isolation, RNA duplex formation is a very fast process after nucleation of just a couple base pairs, with second order rate constants of stable complex formation on the order of ~10^6 M^{-1} sec^{-1} (Craig et al., 1971). These rate constants do not vary significantly for the free pairing segments from a host of RNAs in widely different secondary structure contexts (Zeiler and Simons, 1998). Therefore one might not expect SPS to significantly affect the association rate constant (k_{on}) of Ago onto a target. But once on, the stability of miRISC on a target could be expected to vary depending on the sequence of the interacting bases, and so SPS could indeed affect the off rate (k_{off}) of the complex (Figure 2). A more stable miRISC would promote commitment to mRNA repression, and potentially help counteract outside forces that lead to displacement of the complex (e.g. impeding secondary structure, competing RNA binding proteins).

SPS and TA are highly correlated due to their sequence dependencies, and there is at least one case where SPS could influence the true TA. In the study presented in Chapter 2, TA was

calculated based on the number of sites counted in mRNAs in a genome or transcriptome. The effective TA, however, will also depend on which seed matches in mRNAs are accessible. Seed matches to miRNAs with high SPS will be GC-rich, and therefore more likely to be directly paired with neighboring UTR sequence, occluding access to Ago. One would predict this to be more common for non-conserved sites, not under selection pressure to maintain interactions with the miRNA, than conserved ones that have presumably made the site contexts more accessible for Ago binding to counteract this effect. This could make the effective TA for these miRNAs even lower. Since non-conserved sites far outnumber conserved ones (Friedman et al., 2009), however small, this effect could still be quantifiable.

Contrasting TA and endogenous miRNA sponges

While TA can dilute the activity of miRNAs that have many potential targets in a cell, this phenomenon should be distinguished from the concept of an endogenous "miRNA sponge." Exogenous miRNA sponges were originally developed as an experimental tool to study the effects of depleting miRNAs from cell, and in this capacity they were engineered to be highly expressed transcripts containing multiple binding sites to sequester a miRNA and cause derepression of its endogenous targets (Ebert et al., 2007). Reasons for the existence of endogenous miRNA sponges have been discussed (Ebert and Sharp, 2010), and there is one striking example from plants (see Chapter 2 Discussion)(Franco-Zorrilla et al., 2007). Several research groups have recently proposed that many types of endogenous RNAs, including individual mRNAs, RNAs transcribed from pseudogenes, and lincRNAs, can act as miRNA sponges in human cells and in turn co-regulate each other by competing for miRNAs (Cesana et

al., 2011; Karreth et al., 2011; Salmena et al., 2011; Sumazin et al., 2011; Tay et al., 2011). In these reports, many endogenous RNAs were considered sponge candidates on the basis of them harboring sites to multiple miRNAs that are shared with other transcripts in the cell, although their endogenous expression levels were not quantified systematically. The function of a endogenous sponge, however, should be rather specific to a particular miRNA, be very highly expressed at a specific point to diminish the activity of a miRNA, and is likely to be conserved if this function is important, all which were shown for the convincing example in plants (Franco-Zorrilla et al., 2007). Given that the median TA among conserved vertebrate miRNA families in HeLa cells is 51,286 sites (Supplementary Figure 1f, Chapter 2), a true endogenous miRNA sponge would probably need to be in thousands to tens-of-thousands of copies per cell in order for it to exert a meaningful effect on the repression levels of other targets of the same miRNA. The prospect of such high expression levels seems even less likely for individual pseudogenes and lincRNAs, both of which tend to be expressed at much lower levels than protein coding genes (Cabili et al., 2011; Zou et al., 2009). It remains to be seen whether the RNAs classified as endogenous sponges in these recent reports are expressed highly enough, and perhaps in a dynamic way during some cellular process. Such conditions seem necessary to implicate them as having bona fide regulatory roles, and not just simply being individual RNAs among so many in the cell that contain miRNA target sites.

TA is not a gene regulatory mechanism in of itself, rather it's a manifestation of the genomic environment within which miRNAs emerge and function. TAs should be quite predictable within a given cell type at specific developmental points. And once established, TAs should be robust to change even over relatively long evolutionary times because a large shift in

gene expression would be required to drastically change them for many miRNAs. The phenomenon known as selective avoidance (introduced in Chapter 1) in which highly expressed messages lose sites to co-expressed miRNAs, could be considered one exception. For selective avoidance, the TA for a miRNA would be higher before it emerged than after, but this TA change would have to do more with messages that don't want to be repressed losing their target sites than an active modulation of the miRNA's activity on its targets. The events that lead to the underrepresentation of CG dinucleotides in mammalian genomes was another event that could have redistributed TAs for ancestral miRNAs maintained in mammalian lineages, but again this would have represented a major change in the genomic environment, not an active regulatory mechanism for miRNAs. In order for TA to have a major impact on a network of miRNAs, the expression of all targets or their miRNA binding sites would have to change simultaneously.

lsy-6 targeting

The results presented in Chapter 2 demonstrated that different seed sequences can have different proficiencies, and this implies that there has been selection for some miRNAs to have weak, average, or strong seeds. Once a miRNA emerges with a particular seed sequence and establishes a productive relationship with a particular set of targets, it is unlikely this interaction can be fine tuned by changing the seed proficiency, because it would necessitate the unlikely double event of a mutation in the seed and compensatory change in the site (or vice versa). (The miRNA–target interaction can, however, be tuned by transcriptional regulation of each gene.)

One can hypothesize that in spite of its weak SPS and high TA, *lsy-6* has succeeded as a miRNA because these qualities strongly limit its potential to repress targets other than *cog-1*.

Such target interference could come from present-day competitor mRNAs in ASEL cells, or those arising at any time during evolution. This has effectively made the *lsy-6* targeting very specific for *cog-1*, contingent of course on *cog-1* developing features that allow it to be repressed by a weak seed. Additional support for this idea comes from experiments that showed that *lsy-6* seed matches placed ectopically in other UTRs—including those from *lin-14*, *lin-28*, and *lin-41* that are subject to regulation by other miRNAs—aren't usually sufficient to confer repression of reporters in the ASEL cell (Didiano and Hobert, 2008). That study also began to uncover other sequence elements in the *cog-1* 3′ UTR outside of the seed matches that are also essential for regulation.

If transcriptional control of *lsy-6* became aberrant and it was expressed ectopically in other cells or during other developmental periods, the miRNA would be unlikely to have negative effects on messages with seed matches due to its low proficiency. Thus, once the *lsy-6:cog-1* gene switch was established, it could be refractive to perturbation by the emergence of new sites, and have minimal off-target effects if the miRNA were ectopically expressed. *cog-1* and *lsy-6* have been shown to have overlapping expression only in the ASEL cell, but they are each expressed individually in other cell types (*lsy-6* in a few head and tail neurons; *cog-1* in head and tail neurons and the uterus and vulva)(Hobert, 2006). In these other cell types, *cog-1* is known to play a role in reproductive system development, but it remains unclear what other roles *lsy-6* is serving, since no other UTRs with sites have responded to the miRNA in reporter assays. It's possible *lsy-6* could be acting as a failsafe mechanism in these other cells if ectopic *cog-1* expression had detrimental effects in these cells.

Could *cog-1* be the only target of *lsy-6*? The data thus far support this possibility, along with additional evidence. A new tool that precisely maps the 3′ ends of transcripts was applied to *C. elegans* (Jan et al., 2011), enabling us to check the 3′ UTR annotations of the 14 previously predicted *lsy-6* targets. These new data indicated that for half the predicted targets, the UTRs end before reaching the predicted *lsy-6* sites. Because many of the predicted sites must have been retained in the UTRs expressed in HeLa cells (otherwise repression would not have been observed in Figure 1g,h in Chapter 2), this information did not change our conclusions regarding the targeting proficiency of the *lsy-6* miRNA. This confirmed that at least half of the predicted targets are not authentic targets of *lsy-6 in vivo*. The conserved sites that are expressed in worm could be conserved by chance or for reasons other than miRNA regulation. Since expression profiling of these other targets has not been compared to *lsy-6* expression, it has yet to be determined if they are ever co-expressed *in vivo*.

If weak SPS together with high TA (or strong SPS together with low TA) can limit off-target effects, why should more miRNAs not take advantage of this? There must be a balance between a miRNA being able to establish high target-specificity and its availability for broader targeting, and *lsy-6* falls at one extreme end of this spectrum. The nematode clade is richly diverse, with tens of thousands of described species living in environments as diverse as 3 km below the surface of the earth (Borgonie et al., 2011) and inside the human body. *lsy-6* is found in most sequenced nematodes, and has even been cloned from *Brugia malayi*, a filarial nematode that spends the early part of its larval life cycle (when *lsy-6* is presumed to repress *cog-1*) in a human host (Larry McReynolds, personal communication). Thus this regulatory circuit seems to be robust in diverse environments. Hopefully more interesting examples will emerge from

studies of other miRNAs with uncommon SPS and TA values that are important for their targeting fidelity.

lsy-6 sites in cog-1 are not cooperative in HeLa cells

Despite having two closely spaced *lsy-6* sites (separated by 34-nt) that confer strong repression in reporter assays in worms and in HeLa cells, the *cog-1* 3′ UTR tested negatively for cooperativity in HeLa cells (Supplementary Figure 1b,c, Chapter 2). The amount of repression conferred by the two sites together was equal to the product of the repression levels from the two sites functioning individually. Earlier studies had shown that a pair of sites separated by ~10–40-nt could act cooperatively in reporter assays (Grimson et al., 2007; Saetrom et al., 2007). It's possible that some element needed for cooperativity is present in the ASEL cell but missing in HeLa cells, but since presently cooperativity in miRNA targeting is not well understood, it would be premature to offer additional conjecture on this case.

Although the *lsy-6* sites in *cog-1* do not appear to be cooperative, one could imagine cooperativity between the site for a weak seed and one for a more typical seed, leading to robust repression only in the presence of both miRNAs. If only the miRNA containing the weak seed were expressed, one might expect minimal or no repression. If only the miRNA containing the average seed were expressed, one could expect some repression. With both miRNAs expressed and acting on cooperative sites, repression could be more substantial.

Speculation about the independence of seed match identity for cog-1 repression

One interesting result from the reporter data on the *cog-1* 3′ UTR is that it was repressed by about the same amount (~2.3–3-fold) regardless of seed-match type. In the titration data, lowering the amount of each transfected miRNA below a certain level would lead to less repression, and the concentration below which this would happen was different for each miRNA (Supplementary Figure 1p, Chapter 2). But at transfection levels in which the miRNAs appear to have reached saturation in the silencing complex, as assayed by small RNA Northerns on Ago immunoprecipitations, the identity of the seed match did not matter for repression. This is in contrast to the other targets that responded differentially to seeds based on SPS and TA.

In the model proposed earlier, higher SPS could lower k_{off} of the silencing complex by increasing binding affinity. In this model *cog-1* would have a way, independent of SPS, to increase its affinity for miRISCs even when it contains the weak seed match of *lsy-6*, and which doesn't require cooperativity. If *cog-1* did this by making secondary structure maximally free, one might suspect there wouldn't be enough pairing energy to keep Ago on, and that all of the other targets that tested negative for repression by *lsy-6* had inhibitory secondary structure. One solution could be that *cog-1* forms restrictive structure around the Agos after they have bound, enforcing stable complex binding even with a weak seed match.

It also appears that *cog-1* has a way to neutralize the effect of stronger SPS on repression, because the miR-142*lsy-6* and D-LTA-*lsy-6* seed matches did not repress more than wild-type. If SPS helped dictate the input strength into this negative regulatory circuit in which the output is repression, then *cog-1* can fix the output regardless of input strength. It remains an intriguing

mystery as to how *cog-1* is able to so precisely tune its repression, but perhaps one somewhat similar phenomenon in bacterial translation, while not directly comparable, can be illustrative.

In prokaryotes, direct base pairing between a short motif in the 3′ end of the 16S rRNA and the Shine-Dalgarno (SD) sequence (also known as the ribosome binding site, or RBS) in an mRNA is central to establishing translation initiation at a downstream start codon. In studies of the control of translational efficiency for the bacteriophage MS2 coat gene, it was observed that when secondary structure was not inhibitory at the RBS, translational efficiency operated independently of the binding strength of the SD sequence for the ribosome (de Smit and van Duin, 1990, 1994). The RBS is normally occluded within the stem of a hairpin, and when the strength of the hairpin was made stronger, this predictably reduced translation. A thermodynamically weaker SD sequence (one that is predicted to pair less stably to the 16S rRNA) was more sensitive to stronger hairpin structure, and expectantly resulted in even less translation. But when the hairpin was mutated, making it unstable and thus making secondary structure around the RBS non-inhibitory, a weak SD sequence conferred just as much translation as a strong SD sequence. Therefore an open RBS is operating at maximal efficiency, regardless of the binding strength of the SD sequence. This implies the affinity of the 30S subunit for the SD sequence helps overcome the normally inhibitory effect of secondary structure, probably by reducing the k_{off} of the 30S before it shifts several codons downstream to the start codon. This is just meant to show another instance where if local secondary structure is non-inhibitory, gene expression that is regulated by a *trans* factor can be at maximal levels independent of the predicted binding strength of that factor to a cognate *cis* element. Whether this is relevant to *cog-1* is uncertain, but one could speculate that the sites in *cog-1* are initially found in very open

secondary structure to facilitate binding of Ago with a weak seed match, and then after binding, local structure becomes more restrictive to retain the silencing complexes, leading to repression.

Further questions arising from this study

After learning the importance of SPS and TA on targeting by *lsy-6*, miR-23, and other miRNAs, additional questions arise, listed below:

Do organisms with more or less AU-rich genomes adjust their SPS and TA distributions correspondingly?

MicroRNA expression levels have a significant influence on their ability to repress messages; do miRNAs with high TA have to be expressed at even higher levels to function effectively?

Do the targets of other miRNAs with weak SPS and high TA use mechanisms similar to *cog-1*, and if so may these features or variations thereof also be used more generally in targeting, not just for weak seed matches?

Are there limits to how much repression a stronger SPS can yield? Beyond a required threshold, does stronger SPS yield stronger repression, and if so why would some miRNAs rely on this versus having multiple sites in target?

Since a strong 7mer site can repress as well as a weak 8mer site, has SPS and TA at all influenced site type dynamics during evolution?

What sequence or structural features in the mature duplexes of miRNAs with CG rich seeds ensure that they get loaded efficiently if the loading rules would predict the passenger strand to be loaded instead?

As mentioned in Chapter 2, miRNAs with both weak SPS and high TA and those with both strong SPS and low TA could be two ways to regulate few targets (contingent on the former having targets tuned to be repressed by less proficient seeds). Do these types of miRNAs tend to have more switch-like targeting relationships—are they less likely to follow the fine tuner paradigm than miRNAs with more targets?

The influence of TA implies that Ago is limiting in small RNA targeting, which has implications for Ago specialization. For example, in organisms like *C. elegans* and *Drosophila* that have strong RNAi activities on exogenous dsRNA substrates (e.g. dsRNA viral elements), is one reason to have Argonautes dedicated to these silencing pathways that this avoids inference with endogenous miRNA targeting?

Why should RISC be limiting for endogenous miRNAs targeting? Does this facilitate miRNA turnover during developmental transitions, when some miRNAs suddenly elevate their expression, helping focus the cell's energy on the targets of the newly expressed miRNAs?

Can Ago expression be dynamic during stress or cell state transitions to magnify TA effects?

In Supplementary Figure 1m–s in Chapter 2, *lsy-6* duplex transfected at 25 nM yielded miRNA association in AGO2 equal to miR-142*lsy-6* transfected at 0.2 nM. This means 125-times more *lsy-6* than miR-142*lsy-6* was required in the transfection for them to associate with AGO2 at equal levels, despite these miRNAs varying in only a few nucleotides (and a few compensatory changes in the miRNA*). What sequence or structural determinants in these duplexes are responsible for such a large differences in loading or stability of the miRNAs in AGO2?

Acknowledgements

I thank Jin-Wu Nam for sharing unpublished data and associated discussions, and Vikram Agarwal and Dave Bartel for discussions that influenced how I thought about and presented ideas in this chapter.

References

Borgonie, G., Garcia-Moyano, A., Litthauer, D., Bert, W., Bester, A., van Heerden, E., Moller, C., Erasmus, M., and Onstott, T.C. (2011). Nematoda from the terrestrial deep subsurface of South Africa. Nature *474*, 79-82.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev *25*, 1915-1927.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell *147*, 358-369.

Craig, M.E., Crothers, D.M., and Doty, P. (1971). Relaxation kinetics of dimer formation by self complementary oligonucleotides. J Mol Biol *62*, 383-401.

de Smit, M.H., and van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc Natl Acad Sci U S A *87*, 7668-7672.

de Smit, M.H., and van Duin, J. (1994). Translational initiation on structured messengers. Another role for the Shine-Dalgarno interaction. J Mol Biol *235*, 173-184.

Didiano, D., and Hobert, O. (2008). Molecular architecture of a miRNA-regulated 3' UTR. RNA *14*, 1297-1317.

Ebert, M.S., Neilson, J.R., and Sharp, P.A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. Nat Methods *4*, 721-726.

Ebert, M.S., and Sharp, P.A. (2010). Emerging roles for natural microRNA sponges. Curr Biol *20*, R858-861.

Franco-Zorrilla, J.M., Valli, A., Todesco, M., Mateos, I., Puga, M.I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J.A., and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. Nat Genet *39*, 1033-1037.

Friedman, R.C., Farh, K.K., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res *19*, 92-105.

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.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. (2007). MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell *27*, 91-105.

Hobert, O. (2006). Architecture of a microRNA-controlled gene regulatory network that diversifies neuronal cell fates. Cold Spring Harb Symp Quant Biol *71*, 181-188.

Jan, C.H., Friedman, R.C., Ruby, J.G., and Bartel, D.P. (2011). Formation, regulation and evolution of Caenorhabditis elegans 3'UTRs. Nature *469*, 97-101.

Karreth, F.A., Tay, Y., Perna, D., Ala, U., Tan, S.M., Rust, A.G., DeNicola, G., Webster, K.A., Weiss, D., Perez-Mancera, P.A.*, et al.* (2011). In vivo identification of tumor- suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. Cell *147*, 382-395.

Lambert, N.J., Gu, S.G., and Zahler, A.M. (2011). The conformation of microRNA seed regions in native microRNPs is prearranged for presentation to mRNA targets. Nucleic Acids Res *39*, 4827-4835.

Parker, J.S., Parizotto, E.A., Wang, M., Roe, S.M., and Barford, D. (2009). Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. Mol Cell *33*, 204-214.

Saetrom, P., Heale, B.S., Snove, O., Jr., 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.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell *146*, 353-358.

Sumazin, P., Yang, X., Chiu, H.S., Chung, W.J., Iyer, A., Llobet-Navas, D., Rajbhandari, P., Bansal, M., Guarnieri, P., Silva, J.*, et al.* (2011). An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. Cell *147*, 370-381.

Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S.M., Ala, U., Karreth, F., Poliseno, L., Provero, P., Di Cunto, F.*, et al.* (2011). Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell *147*, 344-357.

Wang, Y., Sheng, G., Juranek, S., Tuschl, T., and Patel, D.J. (2008). Structure of the guide-strand-containing argonaute silencing complex. Nature *456*, 209-213.

Zeiler, B.N., and Simons, R.W. (1998). Antisense RNA structure and function. In RNA structure and function, R.W.a.G.-M. Simons, M., ed. (Cold Spring Harbor, Cold Spring Harbor Laboratory Press), pp. 437–464.

Zou, C., Lehti-Shiu, M.D., Thibaud-Nissen, F., Prakash, T., Buell, C.R., and Shiu, S.H. (2009). Evolutionary and expression signatures of pseudogenes in Arabidopsis and rice. Plant Physiol *151*, 3-15.

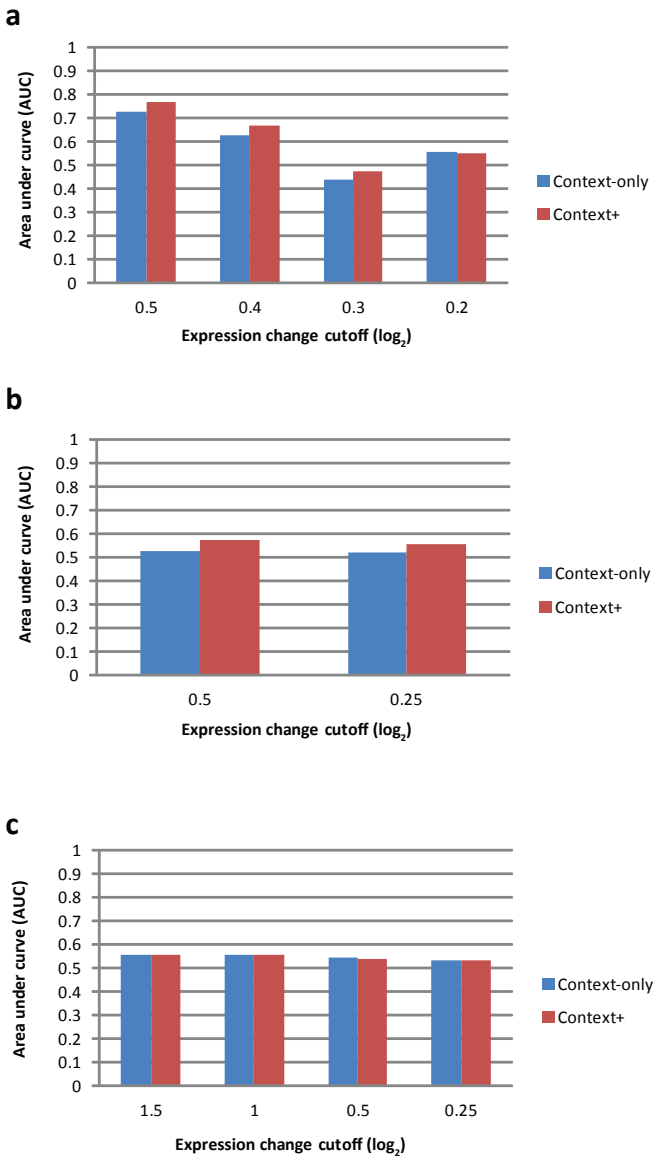


Figure 1 (**a**) ROC curve derived analysis of all messages containing single 7mer-m8 sites to the top 5 highly expressed miRNAs (representing 3 different seed families) in mouse ESCs. As a group, these 5 miRNAs yielded the largest improvements in predictions compared to other combinations tested, presumably because those miRNAs expressed below the top five were at levels too low to generate any impactful targeting. Cutoffs indicate all messages on the array that changed in expression at the indicated level or above, between WT and DGCR8 knockout cells. (**b**) Analysis of messages with sites to the top 20 expressed miRNAs (representing 14 different seed families), comparing expression changes between *Drosophila* S2 WT and Drosha knockdown cells. Otherwise as in **a**. Note that the AUC for context-only is close to 0.5, indicating an equal chance of predicting a non-target or a target, indicating that the dataset is quite noisy. (**c**) Analysis of messages with sites to the top 20 expressed miRNAs (representing 12 different seed families), comparing expression changes between WT and maternal-zygotic Dicer knockout cells at 24-hours post-fertilization in Zebrafish. Otherwise as in **a**.

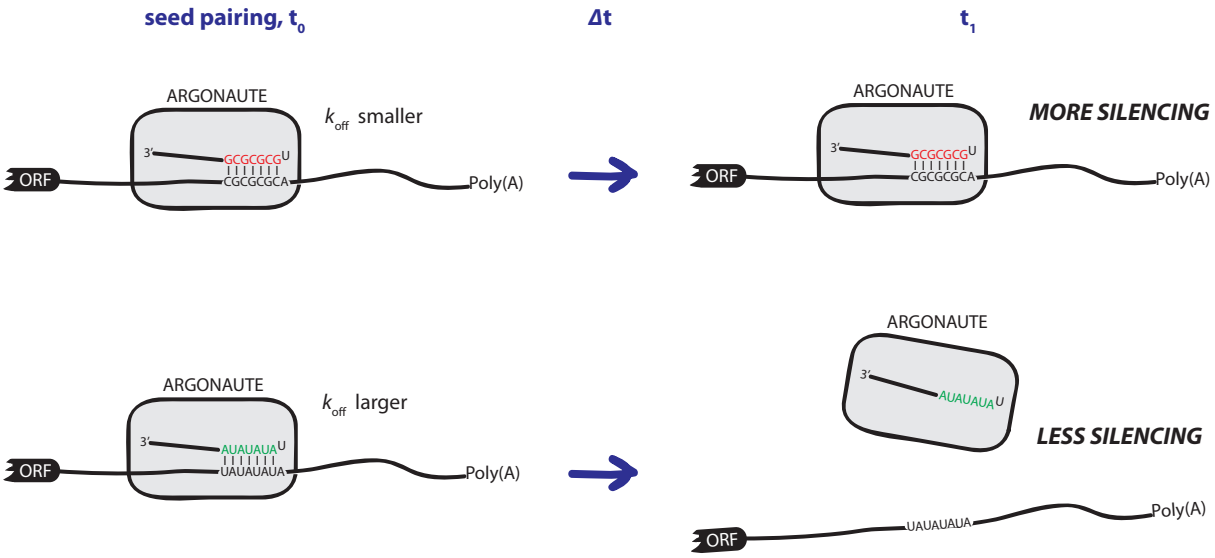


Figure 2 Proposed model for the influence of SPS on miRNA targeting. After binding of silencing complexes to a target and passage of time (t), miRNAs with strong SPS have reduced off rates of Argonaute (Ago), leading to more silencing (top). In contrast, for miRNAs with weak SPS, the off rate of Ago is larger, resulting in less silencing. Rates of intial seed pairing are expected to be similar for both types of seeds. Ago binding or retention could be inhibited by thermal effects, an inclination of the mRNA to form secondary structure, or other competing RNA binding proteins.

Curriculum vitae

David M. Garcia

Education:
Massachusetts Institute of Technology, Cambridge, MA
Ph.D. Biology, June 2012
Thesis: The Importance of RNA Pairing Stability and Target Concentration for Regulation by MicroRNAs

University of California, Santa Cruz, Santa Cruz, CA
B.S. Biochemistry and Molecular Biology, College Honors, 2004
Senior Thesis: Crystallization of a Spliceosomal RNA

Pontificia Universidad Católica de Chile, Santiago, Chile
Study abroad student (January–December 2003)

Research Experience:
MIT/Whitehead Institute, Cambridge, MA, 2007–2012
Advisor: David P. Bartel
Regulation by microRNAs

Veterans Affairs Medical Center/UCSF, San Francisco, CA, 2004–2006
Advisor: Corey Largman
Treating a mouse model of AML

UC Santa Cruz, Santa Cruz, CA, 2000–2002
Advisor: William G. Scott
RNA crystallography

Children's Hospital/Harvard Medical School, Boston, MA, Summer 2000
Advisors: Russell Sanchez and Frances Jensen
AMPA receptor function

UC Berkeley, Berkeley, CA, Summer 1999
Advisor: Steven Beckendorf
Drosophila development

Teaching Experience:
MIT 7.17 (Biotechnology III Laboratory), Teaching Assistant, Spring 2010
MIT 7.05 (General Biochemistry), Teaching Assistant, Spring 2008
MIT 5.111 (Principles of Chemical Science), Group Facilitator/Tutor, Fall 2007

Awards:
MIT S. Klein Prize for Science Writing, First Place, 2012
MIT Ragnar and Margaret Naess Certificate of Distinction, Jazz Performance, 2012
Phi Beta Kappa, 2004
UC Santa Cruz Alumni Association Scholarship, 2001
Michele Guard Memorial Scholarship, 1998

Publications:
Garcia, D.M.*, Baek, D.*, Shin, C., Bell, G.W., Grimson, A., Bartel, D.P. Weak seed-pairing stability and high target-site abundance decrease the proficiency of *lsy-6* and other microRNAs. *Nature Structural and Molecular Biology,* 18, 1139–1146, October 2011. *equal contribution

Additional Science Communication:
You'd Prefer An Argonaute, 2009–2012
http://youdpreferanargonaute.com