PERSPECTIVE

nature structural & molecular biology

Desperately seeking microRNA targets

Marshall Thomas[1], Judy Lieberman[1] & Ashish Lal[1,2]

MicroRNAs (miRNAs) suppress gene expression by inhibiting translation, promoting mRNA decay or both. Each miRNA may regulate hundreds of genes to control the cell's response to developmental and other environmental cues. The best way to understand the function of a miRNA is to identify the genes that it regulates. Target gene identification is challenging because miRNAs bind to their target mRNAs by partial complementarity over a short sequence, suppression of an individual target gene is often small, and the rules of targeting are not completely understood. Here we review computational and experimental approaches to the identification of miRNA-regulated genes. The examination of changes in gene expression that occur when miRNA expression is altered and biochemical isolation of miRNA-associated transcripts complement target prediction algorithms. Bioinformatic analysis of over-represented pathways and nodes in protein-DNA interactomes formed from experimental candidate miRNA gene target lists can focus attention on biologically significant target genes.

MicroRNAs regulate virtually every aspect of biology, including developmental timing, differentiation, proliferation, antiviral defense and metabolism. MicroRNAs are ~22-nucleotide-long RNAs that are generated by sequential processing from longer transcripts that contain a stem-loop[1–4]. One strand is loaded into the miRNA-induced silencing complex (miRISC), which contains the proteins argonaute (Ago) and Tnrc6 (trinucleotide repeat–containing 6; GW182). The other strand is usually degraded. The mature miRNA guides the miRISC to partially complementary sequences, termed miRNA recognition elements (MREs), in target mRNAs to repress mRNA translation, promote transcript decay or both[1,5–7]. MicroRNAs probably regulate the expression of most coding genes[8].

Most metazoan miRNAs pair imperfectly with their cognate mRNAs, and it is difficult to identify their biologically important targets. Bioinformatic analysis of the first known miRNA-regulated genes showed that pairing of miRNA nucleotides 2–8, called the seed region, to the 3′ untranslated region (UTR) of the target mRNA is often important[5]. Algorithms based on seed pairing and evolutionary conservation became a powerful tool for identifying miRNA-regulated genes[5,8–14]. However, these algorithms typically predict hundreds to thousands of target genes for each miRNA, and most predicted genes are not *bona fide* targets[15]. Moreover, the algorithms sometimes fail to predict the most biologically important miRNA targets, such as the oncogenes *KRAS* and *HRAS* for the miRNA let-7 (ref. 16) or the transcription factor *E2F2* and the oncogene *MYC* for miR-24 (ref. 17). Recent studies provide examples of MREs located outside of the 3′ UTR (especially in the coding sequence (CDS)) or that lack exact seed pairing, but compensate by downstream complementarity[17–20].

Experimental methods for identifying miRNA targets identify mRNAs or proteins that are downregulated when a miRNA is overexpressed or that are upregulated when a miRNA is antagonized[17,21–25], or mRNAs that precipitate with miRISC-associated proteins[26–31]. High-throughput sequencing of Ago-immunoprecipitated RNAs after crosslinking now provides a way to identify miRNA-MRE pairings[32–34]. These methods confirm that current target prediction algorithms miss many genes. Both prediction algorithms and experimental methods generate large lists of candidate miRNA targets. However, choosing the important targets from these long lists is daunting. Gene ontology and interactome analyses of candidate target gene lists can be useful tools for this task[17]. Here, we review methods for identifying biologically relevant miRNA targets.

Computational prediction of miRNA targets

Bioinformatics captures the sequence and location characteristics of MREs to predict miRNA targets[5,14,35]. Commonly used algorithms place variable weight on: (i) complementarity to the miRNA seed region; (ii) evolutionary conservation of the MRE; (iii) free energy of the miRNA-mRNA heteroduplex; and (iv) mRNA sequence features outside the target site[17,36–38] (**Table 1**). Early algorithms, such as TargetScan and PicTar[8,9], focus on the seed region in miRNA targeting (**Table 1**). TargetScan[8,10,11] requires an exact match to ≥7 bases of the seed sequence, but PicTar[9] doesn't, instead imposing a stringent free energy cutoff for imperfect matches. (TargetScan includes a special class of seed matches with a hexamer match in positions 2–7, plus an adenosine at position 1). Both TargetScan and PicTar improve their predictions by taking into account evolutionary conservation. TargetScan also adds a 'context score', which considers features in the surrounding mRNA, including local A-U content and location (near either end of the 3′ UTR is preferred) and improves predictions for nonconserved sequences[39]. Messenger RNAs that have a high context score or multiple predicted MREs are more likely to be true targets.

The miRanda algorithm[12] aligns a miRNA to the target mRNA to identify highly complementary sequences. Seed pairing is weighed more strongly than pairing elsewhere[40], but seed G•U wobbles and mismatches (**Table 1**) are allowed. High-scoring targets are then filtered on a secondary criterion of heteroduplex free energy (ΔG). Finally, only conserved predictions are considered. Because miRanda does not

[1]Immune Disease Institute and Program in Cellular and Molecular Medicine, Children's Hospital Boston, Harvard Medical School, Boston, Massachusetts, USA. [2]Present address: Genetics Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. Correspondence should be addressed to J.L. (lieberman@idi.harvard.edu) or A.L. (lalas@mail.nih.gov).

Published online 6 October 2010; doi:10.1038/nsmb.1921

© 2010 Nature America, Inc. All rights reserved.

Table 1 miRNA target prediction algorithms differ in their ability to identify unconventional miRNA recognition elements (MREs)

MRE feature	Example		Target prediction algorithm[a]				
			TargetScan	PicTar	miRanda	PITA	rna22
Perfect seed	`HMGA2    5′ CCGACAUUCAAUUUCUACCUCA 3′` ` :	:| ||||||||` `let-7a 3′ UUGAUAUGUUGGAUGAUGGAGU 5′`	Ref. 36					
G:U wobble seed	`NF2      5′ UACAAGAGAUUCUCCUGCCUCA 3′` ` ||			:|| ||:|||||` `let-7a 3′ UUGAUAUGUUGGAUGAUGGAGU 5′`	Ref. 37	–				
Imperfect seed/seedless	`E2F2     5′ GUGGGUGCU-CUGGGCUGAACCA  3′` ` ||	|| |||::|||| |||` `miR-24 3′ GACA-AGGACGACUUGACUCGGU 5′`	Ref. 17	–				

Outside the 3′ UTR	`DNMT3B   5′ UGGCAAAGAAGAUGUUUUGUGGUGCACUGAG 3′`
	`               :||||       |||:|||:|||||||||`
	`miR-148  3′  -UGUUU......CAAGACAUCACGUGACU- 5′`
		Ref. 38	–	–	–	+	+

[a]The algorithms are listed in the rough order by which they weigh perfect seed pairing and location within the 3′ UTR. The algorithms that allow identification of noncanonical recognition sites generally are more inclusive, but have much higher false prediction rates. +, algorithm predicts targets of this class; –, algorithm does not predict targets of this class.

require exact seed pairing, it predicts sites such as the two *let-7* sites in the *Caenorhabditis elegans* gene *lin-41,* which contain either a bulge or a G•U wobble in the seed region[40,41].

Other algorithms account for target site accessibility[42,43]. Instead of considering only the ΔG of the miRNA-mRNA duplex[12,13], these algorithms estimate the free energy cost to unfold the mRNA secondary structure that surrounds the target site. The PITA algorithm[42] combines this free energy cost with the ΔG of miRNA-target pairing to measure $\Delta\Delta G$. The algorithm finds seed matches (allowing a user-specified number of mismatches), and calculates $\Delta\Delta G$ for each match, irrespective of conservation. Although designed and tested for 3′ UTRs, PITA can predict MREs outside the 3′ UTR. The algorithm mirWIP[43] is based on sequence features of mRNAs isolated by immunoprecipitation of the *C. elegans* miRISC[26]. The mirWIP program scores targets on the basis of: (i) perfect and imperfect seed matches; (ii) site accessibility from predicted mRNA secondary structure; and (iii) ΔG_{total}, an analog of $\Delta\Delta G$. The sensitivity and specificity of mirWIP are high relative to those of other algorithms for the miRISC dataset that was used to develop it. However, it was not tested against other high-throughput datasets.

The rna22 algorithm differs radically from other algorithms[13]. It relies on recognition of shared miRNA sequence patterns. Briefly, the training dataset considered all known mature miRNA sequences to identify statistically over-represented sequence patterns. Potential MREs were identified as complementary patterns. Predictions are made by 'folding' each known miRNA *in silico* into a heteroduplex with the predicted target sequence and identifying miRNAs with the lowest ΔG. The rna22 algorithm does not consider conservation, its predictions are not restricted to the 3′ UTR, and the user can allow seed mismatches. This algorithm has been used to locate 'seedless' and CDS MREs (**Table 1**). The developers of rna22 reported a low false-positive rate[13]. However, rna22 did not perform as well as seed-biased algorithms on high-throughput proteomic data[21,22], and only some rna22-predicted MREs were validated in low-throughput studies[17,44].

Robust comparisons of prediction algorithms are lacking. However, a few reviews have compared predictions against a 'benchmark' of experimentally validated or refuted miRNA targets[14,45]. In these comparisons seed-based predictions that require stringent pairing have the highest specificity and sensitivity[45]. Proteomics studies by the developers of TargetScan and PicTar suggest that their seed-based algorithms have superior predictive power[15,21,22]. Many experimental studies have found that miRNA targets are enriched for exact miRNA seed matches. Moreover, many experimentally validated targets were chosen on the basis of seed-based algorithms. However, recent high-throughput experimental analyses of Ago-bound miRNA-mRNA pairings suggest that around 25%–45% of MREs lack a perfect seed match to expressed miRNAs[32,33]. An increasing list of validated MREs lack a canonical seed[17,18,24,41]. Therefore, filtering gene lists by requiring an exact seed match will eliminate true targets.

All bioinformatic analyses need to consider variations in the annotated sequence databases used to generate predictions. The same algorithm run on two databases (UCSC and Ensembl) generated target predictions that overlapped by less than 50%[46]. Current databases do not take into account cell type–specific mRNA isoforms. Some mRNAs are alternatively polyadenylated in cellular proliferation[47], cancer[48] and differentiation[49], generating transcripts with unique 3′ UTRs. It is important to know whether the expressed isoforms in a cell under study contain the predicted MREs. High-throughput mRNA sequencing of tissue-, development- or disease-specific gene expression[50] can be combined with miRNA profiling to identify candidate miRNA targets[32].

Genetic approaches to miRNA target identification

Forward genetics involves screening for mutants with interesting phenotypes, whereas reverse genetics is used to examine the functional effect of manipulating a specific gene[51]. The founding miRNA lin-4 and its target *lin-14*, which were identified in forward genetic screens for larval defects in *C. elegans,* led to the discovery of miRNAs[51–53]. Other miRNAs and a few of their targets were identified by screens for mutant phenotypes in *C. elegans* and *Drosophila melanogaster*[51]. These targets were generally deduced from epistatic relationships and the complementary phenotypes that sometimes occur when a miRNA and a key target are mutated[51,53].

Reverse genetics has been especially useful for identifying the biological function of miRNAs, and a few examples are described here. Conditional overexpression of each of the six miRNAs in the oncogenic miR-17–92 cluster pinpointed a prominent role for miR-19, which largely recapitulated the oncogenic properties of the cluster, in hematological cancers *in vitro* and *in vivo*[54–56]. Furthermore, when combined with computational, expression or functional analyses, these studies all identified the tumor suppressor *Pten* as an important target of miR-19. The function of the miR-143–145 cluster, which is highly expressed in smooth muscle cells, was elucidated by knocking it out in mice. Knockout mice lacked gross developmental defects, indicating that the cluster is not essential for development[57,58]. However, smooth muscle cells from the knockout mice were less differentiated than those from wild-type animals, and were impaired in responding to vascular injury. Targets of the cluster were then identified by manually selecting interesting candidate genes from computational predictions[57]. Combining reverse genetics with computational and other molecular methods is a powerful method for dissecting miRNA function[21,27,32].

© 2010 Nature America, Inc. All rights reserved.

© 2010 Nature America, Inc. All rights reserved.

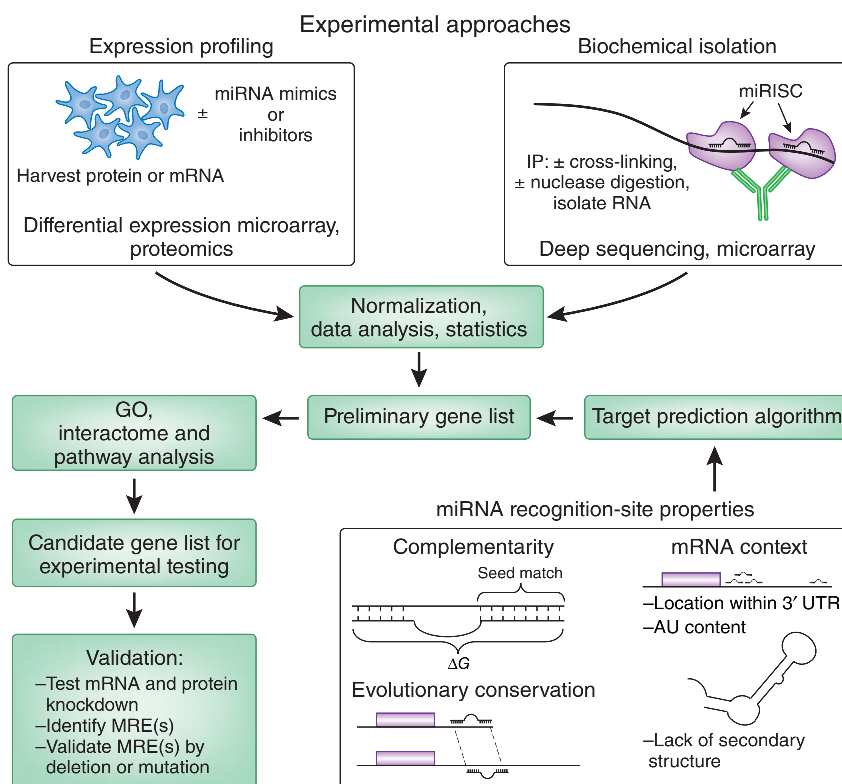

Figure 1 Methods for identifying miRNA targets. Putative target genes can be identified by expression profiling of cells in which the miRNA is overexpressed or antagonized, by biochemical isolation of the miRISC or by target prediction algorithms. These methods generally identify hundreds of candidate genes or more. Bioinformatic analysis of these large candidate gene lists for over-represented Gene Ontology (GO) terms, enriched biological pathways or gene interaction networks can then help researchers to select candidate genes to evaluate experimentally.

Forward genetics can also be used to identify miRNA targets without bias. In one study, candidate miR-19 targets were identified by a genome-wide short hairpin RNA screen to discover genes whose knockdown 'phenocopied' overexpression of miR-19 (ref. 54). The screen identified eight candidate miR-19–regulated genes, of which four were validated. An important benefit of forward genetic methods is that they might identify the most biologically meaningful targets.

Microarray analysis after miRNA overexpression or knockdown

Because miRNAs reduce the steady-state transcript levels of most target genes[6,21–23], identifying mRNAs that are downregulated after ectopic miRNA expression or upregulated after miRNA antagonism provides a useful way to identify putative miRNA-regulated genes (**Fig. 1**). Although this method cannot distinguish direct from indirect targets, harvesting cells soon after transfection might enhance the proportion of direct targets[24,59]. This strategy can also pinpoint miRNA function by bioinformatic analysis of the effect of miRNA manipulation on global gene expression[17,23,59] (**Fig. 1**). This approach was first used to investigate changes in mRNA after overexpression of muscle-specific miR-1 and brain-specific miR-124 (ref. 23). Remarkably, transfection of miR-1 into HeLa cells shifted their gene expression profile toward that of muscle cells, whereas transfection of miR-124 shifted the profile toward that of brain cells. For miR-1, 88% of downregulated genes contained a 3′ UTR hexamer seed match, and 76% of genes that were downregulated by miR-124 had a similar match. CDS seed matches were also somewhat enriched. When miR-124 with mutations in seed region bases 5 and 6 was transfected into cells, the downregulated genes no longer overlapped with genes downregulated by wild-type miR-124. However, when bases 9 and 10 were altered, 89% of downregulated genes overlapped. Recently, similar transcriptome analysis identified a class of functional 'seedless' targets that pair with 11 central bases of the miRNA[18].

Overexpression of miRNAs has also been used to study the biological effects and targets of cancer-related miRNAs. Overexpression of let-7 decreased the proliferation of liver and lung cancer cells, and more than 200 transcripts were downregulated in both cell types[24]. Gene ontology analysis of the downregulated genes indicated that they were enriched in genes for DNA replication and cell cycle pathways. Similarly, transfection of miR-15a and miR-16 into colon cancer cell lines increased the proportion of cells in cell cycle phase G0/G1 and genes that were downregulated were enriched in cell cycle–related genes[59]. Paradoxically, these downregulated cell cycle genes were not enriched in miR-15a or miR-16 seed matches, suggesting that much of the effect of these miRNAs might be indirect.

Loss of function has also been used to identify miRNA targets (**Fig. 1**). *In vivo* knockdown of the liver-specific miR-122 modestly upregulated hundreds of mRNAs[60]. The most enriched 3′ UTR hexamer motif of these genes matched the miR-122 seed and the upregulated genes were enriched in genes for cholesterol biosynthesis pathways. An analysis of changes in gene expression after knockdown of miR-15 and -16 was performed at the same time as the overexpression study described above[59]. The increase in mRNA expression after knockdown was small compared to the downregulation that occurred after ectopic overexpression, suggesting that ectopic miRNA expression might be more useful than miRNA antagonism for identifying targets. However, the relative benefits of overexpression and knockdown depend on biological context. Overexpression may result in supraphysiological miRNA levels, leading to artifacts. In particular, transfection of miRNA mimics can increase the expression of endogenous miRNA targets, probably because it limits the available miRISC[61]. Conversely, repression of an underexpressed miRNA is unlikely to have a substantial effect on target genes. It can also be difficult to suppress a highly expressed endogenous miRNA.

Proteomics

Proteomic analysis can also identify miRNA targets. Stable isotope labeling with amino acids in cell culture (SILAC) followed by mass spectrometry can assess the effect of the loss or overexpression of miRNAs on global protein expression[21,22] (**Fig. 1**). In one study, the nuclear proteome was analyzed after overexpression of miR-124, miR-1 and miR-181 in HeLa cells[21]. The results support the importance of the seed region. For miR-124 and miR-1, the most enriched heptanucleotide motif in the transcripts of the most downregulated proteins matched the miRNA seed sequence, and for miR-181, the second most enriched motif matched the seed. SILAC was also used to compare the expression of cytoplasmic and nuclear proteins in neutrophils from wild-type and miR-223 knockout mice. The mRNAs of proteins that were overexpressed in miR-223 knockout neutrophils were most highly enriched for miR-223 seed matches. Although algorithms that

require stringent seed pairing had the greatest power to predict changes in protein levels, only 33% of the targets predicted by such algorithms showed a change in protein expression, suggesting that the algorithms have a high false-positive rate[21].

SILAC was also used to analyze the proteome of HeLa cells that overexpressed one of five different miRNAs[22]. The 3′ UTR hexamer motifs that most correlated with changes in protein expression were complementary to the miRNA seed for each miRNA. To assess whether depletion and overexpression of miRNAs influenced the expression of the same proteins, the authors knocked down let-7b. The protein changes in cells that overexpressed let-7b and cells in which it was knocked down were negatively correlated. However, the effect of knockdown was only about one-third as great as that of overexpression. Therefore, miRNA overexpression has more of an effect than miRNA antagonism on both protein and mRNA expression.

Changes in mRNA expression correlated well with protein changes in both studies. Proteomics, like mRNA expression studies, cannot distinguish between direct and indirect miRNA targets. SILAC experiments are time-consuming, expensive and not accessible to most laboratories. Moreover, current methods can resolve only a fraction of the proteome at a time. Given these practical challenges, and because mRNA and protein expression changes are highly correlated, microarray analysis might be preferred for its simplicity.

Identification of miRISC-associated miRNA targets

The mammalian miRISC contains a mature miRNA and several proteins, including an Ago protein and Tnrc6 (refs. 62,63). Several studies have identified miRNA targets by their association with miRISC proteins, using immunoprecipitation of epitope-tagged miRISC components[27–30] or native miRISC[31], often while manipulating a specific miRNA to identify its targets[27,29,30] (**Fig. 1**). The Cohen laboratory used microarrays to identify mRNAs that immunoprecipitated with hemagglutinin (HA)-tagged Ago1 in *D. melanogaster* S2 cells[27]. The 89 reproducibly isolated mRNAs were enriched for 3′ UTR seed matches, and some contained CDS seed matches. When introduced into the luciferase open reading frame, CDS MREs were moderately active in reporter assays. The HA-Ago1 immunoprecipitation was performed in transgenic flies with mutant or wild-type miR-1, and 108 transcripts were selectively depleted in embryos with mutated miR-1. The 11 transcripts that contained miR-1 heptamer seed matches all had some activity by luciferase assay. This study highlighted the potential utility of biochemical pulldown to identify miRNA targets. However, it captured only 10% of the expected number of miR-1 target genes[28]. A more sensitive protocol was later developed by optimizing immunoprecipitation conditions and increasing amplification of the mRNAs that were pulled down[28].

Several other studies immunoprecipitated tagged miRISC components to identify targets. The Hannon laboratory identified 294 mRNAs that bound specifically to myc-tagged Ago2 in cells that overexpressed miR-124a (ref. 29). Of these mRNAs, 67% contained 3′ UTR heptamer miR-124a seed matches. Notably, many mRNAs that were enriched in the immunoprecipitation were not downregulated in the cells that overexpressed miR-124a. However, when their 3′ UTRs were tested by luciferase assay, 21 out of 30 were suppressed. A similar Ago2 pulldown to identify miR-124a and miR-1 targets[30] isolated 49% of the putative miR-124a targets identified by the Hannon laboratory. Many putative targets had seed matches in the CDS, but not the 3′ UTR. Expression of mRNAs with CDS seed matches decreased after transfection with miRNA, but genes with 3′ UTR matches were more strongly suppressed, consistent with earlier studies[27,30]. Immunoprecipitation with Tnrc6 has also been used to isolate targets in mammalian cells[62] and *C. elegans*[26]. Another approach to target identification has been to clone and sequence mRNAs that immunoprecipitated with endogenous Ago1 and Ago2 (ref. 31). Five out of six randomly chosen genes that were sequenced more than once were experimentally validated as miRNA targets. However, most of the clones were recovered only once, suggesting that the depth of sequencing may have been a limiting factor.

The biochemical approach to target identification has mostly been used to describe the general features of MREs, rather than to identify specific targets. Pulldown of Ago-associated mRNAs usually enriches transcripts containing seed matches. Experimental validation of putative miRNA targets identified by immunoprecipitation is high, supporting the usefulness of the method[29,30]. Immunoprecipitation studies have identified targets whose mRNAs do not decrease after overexpression of miRNAs[29] as well as unanticipated CDS MREs[27,30]. However, some questions remain about the utility of Ago immunoprecipitation for identifying miRNA targets. The miRNAs and mRNAs that are pulled down with Ago1 or Ago2 might not be identical to those found in miRISCs that contain other Agos[64]. Moreover, epitope-tagged Ago has some drawbacks. Epitope-tagged Ago proteins can associate with transfer RNAs[65], potentially introducing experimental artifacts[27]. Overexpression of Ago globally increases endogenous miRNA production[66,67], which could skew the profile of endogenous miRNA-mRNA interactions. As Ago proteins in cell extracts can associate with cognate mRNAs after lysis[29] some Ago-immunoprecipitated mRNAs might not be true targets. Finally, the Ago pulldowns have the limitation that they identify mRNAs rather than specific MREs and are usually not specific for an individual miRNA.

Pulldowns using miRISC components enrich for all miRNA targets, but don't directly identify mRNAs associated with a specific miRNA. To circumvent this problem, a recent study sought to identify mRNAs directly bound to transfected biotinylated miR-10a (refs. 68,69). However, the results are controversial because the pulled down mRNAs were not enriched for known targets or for miR-10a seed matches. Moreover, they were mostly abundant ribosomal mRNAs, suggesting they might have associated with biotinylated miR-10a nonspecifically. Most identified genes were translationally upregulated, rather than being downregulated, by miR-10a. The authors attributed this unexpected result to the presence of weak miR-10a binding sites in the 5′ UTR.

Identifying miRISC-bound MREs

An important recent development, called Ago HITS-CLIP (high throughput sequencing by crosslinking and immunoprecipitation), directly identifies miRNA-bound MREs[32–34]. Nucleic acids are crosslinked by ultraviolet radiation to miRISC proteins and then immunoprecipitated with an antibody to a miRISC component. Unbound RNA is digested to leave miRISC-protected RNA fragments, which are then analyzed by high-throughput RNA sequencing to identify both Ago-associated miRNAs and their target MREs. In the first study, protected transcript fragments associated with native Ago complexes in mouse brain or HeLa cells were most enriched for motifs that matched the seeds of the most abundant miRNAs. Fragments of mRNAs were also enriched for sequences just after the stop codon or before the polyadenylation site, consistent with studies indicating that MREs in the middle of 3′ UTRs contribute less to silencing than those at the ends[39]. Both the false-positive and false-negative rates of the MRE predictions were low in this study. Notably, detection of previously validated miRNA targets correlated with their expression; highly expressed targets were more readily identified. Of the enriched sequences, 25% mapped to the CDS and 27% lacked a perfect seed match to the 20 most highly expressed miRNAs. These results support recent findings that miRNAs regulate many genes by 'seedless' or CDS interactions[17,18,20,38,41,70,71].

A HITS-CLIP study in *C. elegans*[32] also found enrichment for 3′ UTR seed matches to isolated miRNAs. However, 37% of sequences lacked

© 2010 Nature America, Inc. All rights reserved.

a conserved seed match, even after allowing one G•U wobble. CDS sequences were significantly enriched for matches to the central region of the miRNA, but not for seed matches. The identified MREs were found in more accessible regions (lacking secondary structure).

An improved method for isolating protein-associated RNAs, termed PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation), has also been used to identify miRNA targets[34]. Cells are cultured with photoreactive 4-thiouridine, which can substitute for uridine during transcription, before cross-linking and immunoprecipitation. Protein-bound miRNAs and mRNAs are then analyzed by high-throughput sequencing. The incorporation of 4-thiouridine substantially improves RNA yields. Because reverse transcription of 4-thiouridine leads to T-C transitions, the MRE-miRISC interaction site can be accurately mapped. PAR-CLIP of Flag- or HA-tagged AGO1–4 identified almost 20,000 enriched sequences. Nine out of ten of the most enriched heptamers matched the seed sequence of highly expressed miRNAs. Fifty percent of the crosslinked candidate MREs mapped to the CDS, whereas 46% matched a 3′ UTR. Target site identification was tested by antisense inhibition of the 25 most highly expressed miRNAs followed by expression profiling. Targets with seed matches were more frequently upregulated than targets without seed matches. Similarly, mRNAs that were sequenced exclusively in the 3′ UTR were more frequently upregulated after the miRNA was antagonized than those with CDS-only sequence clusters.

Although these methods provide an elegant way to identify miRNA targets, they involve a technically challenging, multistep protocol. The challenges of the procedure need to be weighed against the benefit of genome-wide direct identification of miRNA target sites.

Gene network analysis of miRNA targets

Given the potential of a miRNA to regulate a large number of genes, it can be challenging to identify key miRNA targets and functions from the long lists of putative target genes generated by the methods described above. Gene ontology and interactome analysis can be used to probe the features of lists of candidate genes[17] (**Fig. 1**). We applied this approach to investigating the biological function of miR-24, a miRNA that is consistently upregulated during cellular differentiation[72,73]. Candidate miR-24 targets were identified by their downregulation after overexpression of miR-24. We found that 248 mRNAs had significantly reduced expression, of which 40% were predicted by TargetScan and 51% had a 3′ UTR miR-24 hexamer seed, suggesting that a significant proportion of the downregulated mRNAs were direct targets of miR-24. The cellular pathways that were enriched in the downregulated genes suggested that miR-24 regulates cell cycle progression and DNA repair. Consistent with this hypothesis, overexpression of miR-24 inhibits cellular proliferation and sensitizes cells to DNA damage[17,72].

We used the Ingenuity Pathways software to examine the interactome formed by the downregulated genes. The network of genes that were downregulated by miR-24 was enriched for proteins that have key roles in cell cycle progression. E2F2 and MYC formed highly connected nodes of this interactome, and the effects of miR-24 on cell cycle progression could be reproduced by manipulating E2F2 (a direct target that was not predicted by any algorithm and that lacks a 3′ UTR miR-24 seed match). We suggested, on the basis of these results, that the genes at nodes of a miRNA target interactome might be biologically important targets. This idea needs to be tested in other experimental systems. Many of the genes in the interactome that are downregulated by miR-24 are key elements of cell cycle regulation, including genes that are transcriptionally regulated by E2F2 and MYC. These were also validated as direct targets. Most of those also lacked a seed match, which is consistent with the Ago HITS-CLIP findings noted above. The importance of seed pairing for target mRNA selection could vary amongst miRNAs. This example of a miRNA that suppresses key transcription factors and also directly suppresses their transcriptional target genes suggests that some miRNAs function as master regulators by downregulating a dense network of genes in the same pathway. Applying unbiased systems biology approaches to experimental datasets may help to define miRNA function and pinpoint important target genes.

Validation of putative target genes

The identification of putative miRNA targets is only the first step. To validate candidate genes, the effect of manipulating the miRNA (by overexpression, knockdown or genetic ablation) on protein and mRNA levels of the candidate gene needs to be assessed. An inverse relationship between the gene product (at least the protein) and the miRNA is expected. Direct regulation of gene expression by an miRNA is then tested by reporter assays that use expression plasmids incorporating the entire 3′ UTR or CDS in cells that have been transfected to overexpress the miRNA mimic. Showing that the entire 3′ UTR of a gene is regulated by a miRNA adds confidence that the transcript is recognized in its native context. Further confirmation is provided by identifying the MRE. Candidate MREs can be identified using seed-based algorithms or algorithms that do not require an exact seed match, such as rna22 or PITA[13,42]. Direct regulation by the miRNA is then confirmed by mutating or deleting binding residues in the reporter and testing for restored expression. Most studies test direct regulation by miRNA overexpression using reporters driven by a strong promoter. Showing that a gene can be regulated by physiologically relevant levels of the miRNA provides additional evidence that regulation of the gene by the miRNA is biologically important.

Some genes that have been identified by experimental approaches have candidate CDS MREs. As CDS MREs have a weaker effect on gene expression than 3′ UTR MREs, it is important to determine whether these potential CDS MREs are functional. To do this, CDS MREs and (potentially) their surrounding sequences should be cloned in-frame into the CDS of an appropriate reporter[27,71]. The presence of rare codons upstream of a CDS MRE may slow down translation, permitting miRNAs to bind to CDS MREs and effectively to repress gene expression[74]. This fact needs to be considered in designing CDS MRE reporters.

Concluding remarks

Genome-wide miRNA target identification methods are constantly improving. Ectopic miRNA expression followed by microarray analysis may be the simplest way to identify the biological function and candidate targets of a miRNA. This approach is supported by proteomics studies, which have provided evidence that changes in the expression of mRNA and proteins correlate well[21,22]. However, this strategy has limitations. First, targets that are translationally repressed will be missed. Second, overexpression of a miRNA does not distinguish direct from indirect targets and supraphysiological overexpression can introduce artifacts. Knockdown of a miRNA may therefore be better than overexpression for identifying miRNA targets in a physiological context, although it is less sensitive. Biochemical pulldown methods are continuing to improve and will ultimately provide a more specific and sensitive method for identifying miRNA targets.

Although the importance of the miRNA seed region in miRNA target recognition is clear, a subset of biologically relevant MREs lack a canonical seed or are located in the CDS. Filtering miRNA overexpression or pulldown gene lists by requiring an exact 3′ UTR seed match will eliminate potentially important target genes from consideration. Therefore, we favor an unbiased selection of candidate targets. Gene ontology and interactome analysis or forward genetics can help to pinpoint biologically interesting targets from the hundreds of genes identified by any method. The identification of biologically important targets will be crucial for understanding miRNA function.

© 2010 Nature America, Inc. All rights reserved.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Ghildiyal, M. & Zamore, P.D. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**, 94–108 (2009).
2. Davis, B. & Hata, A. Regulation of microRNA biogenesis: a miRiad of mechanisms. *Cell Commun. Signal.* **7**, 18 (2009).
3. Winter, J., Jung, S., Keller, S., Gregory, R.I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* **11**, 228–234 (2009).
4. Kim, V.N., Han, J. & Siomi, M.C. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **10**, 126–139 (2009).
5. Bartel, D.P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).
6. Hendrickson, D.G. *et al.* Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biol.* **7**, e1000238 (2009).
7. Brodersen, P. & Voinnet, O. Revisiting the principles of microRNA target recognition and mode of action. *Nat. Rev. Mol. Cell Biol.* **10**, 141–148 (2009).
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. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nat. Genet.* **37**, 495–500 (2005).
10. 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).
11. Lewis, B.P., Shih, I., Jones-Rhoades, M.W., Bartel, D.P. & Burge, C.B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
12. John, B. *et al.* Human microRNA targets. *PLoS Biol.* **2**, e363 (2004).
13. Miranda, K.C. *et al.* A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell* **126**, 1203–1217 (2006).
14. Rajewsky, N. microRNA target predictions in animals. *Nat. Genet.* **38** (suppl.), S8–S13 (2006).
15. Alexiou, P., Maragkakis, M., Papadopoulos, G.L., Reczko, M. & Hatzigeorgiou, A.G. Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics* **25**, 3049–3055 (2009).
16. Johnson, S.M. *et al.* RAS is regulated by the let-7 microRNA family. *Cell* **120**, 635–647 (2005).
17. Lal, A. *et al.* miR-24 inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3′UTR microRNA recognition elements. *Mol. Cell* **35**, 610–625 (2009).
18. Shin, C. *et al.* Expanding the microRNA targeting code: functional sites with centered pairing. *Mol. Cell* **38**, 789–802 (2010).
19. Tay, Y.M. *et al.* MicroRNA-134 modulates the differentiation of mouse embryonic stem cells, where it causes post-transcriptional attenuation of Nanog and LRH1. *Stem Cells* **26**, 17–29 (2008).
20. Tay, Y., Zhang, J., Thomson, A.M., Lim, B. & Rigoutsos, I. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* **455**, 1124–1128 (2008).
21. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).
22. Selbach, M. *et al.* Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**, 58–63 (2008).
23. Lim, L.P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769–773 (2005).
24. Johnson, C.D. *et al.* The let-7 MicroRNA represses cell proliferation pathways in human cells. *Cancer Res.* **67**, 7713–7722 (2007).
25. Chang, T. *et al.* Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol. Cell* **26**, 745–752 (2007).
26. Zhang, L. *et al.* Systematic identification of *C. elegans* miRISC proteins, miRNAs, and mRNA targets by their interactions with GW182 proteins AIN-1 and AIN-2. *Mol. Cell* **28**, 598–613 (2007).
27. Easow, G., Teleman, A.A. & Cohen, S.M. Isolation of microRNA targets by miRNP immunopurification. *RNA* **13**, 1198–1204 (2007).
28. Hong, X., Hammell, M., Ambros, V. & Cohen, S.M. Immunopurification of Ago1 miRNPs selects for a distinct class of microRNA targets. *Proc. Natl. Acad. Sci. USA* **106**, 15085–15090 (2009).
29. Karginov, F.V. *et al.* A biochemical approach to identifying microRNA targets. *Proc. Natl. Acad. Sci. USA* **104**, 19291–19296 (2007).
30. Hendrickson, D.G., Hogan, D.J., Herschlag, D., Ferrell, J.E. & Brown, P.O. Systematic identification of mRNAs recruited to argonaute 2 by specific microRNAs and corresponding changes in transcript abundance. *PLoS ONE* **3**, e2126 (2008).
31. Beitzinger, M., Peters, L., Zhu, J.Y., Kremmer, E. & Meister, G. Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol.* **4**, 76–84 (2007).
32. Zisoulis, D.G. *et al.* Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans. Nat. Struct. Mol. Biol.* **17**, 173–179 (2010).
33. Chi, S.W., Zang, J.B., Mele, A. & Darnell, R.B. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* **460**, 479–486 (2009).
34. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
35. Hammell, M. Computational methods to identify miRNA targets. *Semin. Cell Dev. Biol.* **21**, 738–744 (2010).
36. Mayr, C., Hemann, M.T. & Bartel, D.P. Disrupting the pairing between let-7 and Hmga2 enhances oncogenic transformation. *Science* **315**, 1576–1579 (2007).
37. Meng, F. *et al.* The microRNA let-7a modulates interleukin-6-dependent STAT-3 survival signaling in malignant human cholangiocytes. *J. Biol. Chem.* **282**, 8256–8264 (2007).
38. Duursma, A.M., Kedde, M., Schrier, M., le Sage, C. & Agami, R. miR-148 targets human DNMT3b protein coding region. *RNA* **14**, 872–877 (2008).
39. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).
40. Betel, D., Wilson, M., Gabow, A., Marks, D.S. & Sander, C. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* **36**, D149–D153 (2008).
41. Vella, M.C., Choi, E., Lin, S., Reinert, K. & Slack, F.J. The *C. elegans* microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3′UTR. *Genes Dev.* **18**, 132–137 (2004).
42. 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).
43. Hammell, M. *et al.* mirWIP: microRNA target prediction based on microRNA-containing ribonucleoprotein-enriched transcripts. *Nat. Methods* **5**, 813–819 (2008).
44. Navarro, F. *et al.* miR-34a contributes to megakaryocytic differentiation of K562 cells independently of p53. *Blood* **114**, 2181–2192 (2009).
45. Sethupathy, P., Megraw, M. & Hatzigeorgiou, A.G. A guide through present computational approaches for the identification of mammalian microRNA targets. *Nat. Methods* **3**, 881–886 (2006).
46. Ritchie, W., Flamant, S. & Rasko, J.E.J. Predicting microRNA targets and functions: traps for the unwary. *Nat. Methods* **6**, 397–398 (2009).
47. Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A. & Burge, C.B. Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science* **320**, 1643–1647 (2008).
48. Mayr, C. & Bartel, D.P. Widespread shortening of 3′UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**, 673–684 (2009).
49. Ji, Z., Lee, J.Y., Pan, Z., Jiang, B. & Tian, B. Progressive lengthening of 3′ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl. Acad. Sci. USA* **106**, 7028–7033 (2009).
50. Wang, E.T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
51. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350–355 (2004).
52. Wightman, B., Ha, I. & Ruvkun, G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell* **75**, 855–862 (1993).
53. Lee, R.C., Feinbaum, R.L. & Ambros, V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843–854 (1993).
54. Mavrakis, K.J. *et al.* Genome-wide RNA-mediated interference screen identifies miR-19 targets in Notch-induced T-cell acute lymphoblastic leukaemia. *Nat. Cell Biol.* **12**, 372–379 (2010).
55. Mu, P. *et al.* Genetic dissection of the miR-17:92 cluster of microRNAs in Myc-induced B-cell lymphomas. *Genes Dev.* **23**, 2806–2811 (2009).
56. Olive, V. *et al.* miR-19 is a key oncogenic component of mir-17–92. *Genes Dev.* **23**, 2839–2849 (2009).
57. Xin, M. *et al.* MicroRNAs miR-143 and miR-145 modulate cytoskeletal dynamics and responsiveness of smooth muscle cells to injury. *Genes Dev.* **23**, 2166–2178 (2009).
58. Elia, L. *et al.* The knockout of miR-143 and -145 alters smooth muscle cell maintenance and vascular homeostasis in mice: correlates with human disease. *Cell Death Differ.* **16**, 1590–1598 (2009).
59. Linsley, P.S. *et al.* Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression. *Mol. Cell. Biol.* **27**, 2240–2252 (2007).
60. Krützfeldt, J. *et al.* Silencing of microRNAs *in vivo* with 'antagomirs'. *Nature* **438**, 685–689 (2005).
61. Khan, A.A. *et al.* Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat. Biotechnol.* **27**, 549–555 (2009).
62. Landthaler, M. *et al.* Molecular characterization of human Argonaute-containing ribonucleoprotein complexes and their bound target mRNAs. *RNA* **14**, 2580–2596 (2008).
63. Ding, L. & Han, M. GW182 family proteins are crucial for microRNA-mediated gene silencing. *Trends Cell Biol.* **17**, 411–416 (2007).
64. Su, H., Trombly, M.I., Chen, J. & Wang, X. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. *Genes Dev.* **23**, 304–317 (2009).
65. Maniataki, E. & Mourelatos, Z. Human mitochondrial tRNAMet is exported to the cytoplasm and associates with the Argonaute 2 protein. *RNA* **11**, 849–852 (2005).
66. Zhang, X., Graves, P.R. & Zeng, Y. Stable Argonaute2 overexpression differentially regulates microRNA production. *Biochim. Biophys. Acta* **1789**, 153–159 (2009).
67. Diederichs, S. & Haber, D.A. Dual role for Argonautes in microRNA processing and post-transcriptional regulation of microRNA expression. *Cell* **131**, 1097–1108 (2007).
68. Ørom, U.A. & Lund, A.H. Isolation of microRNA targets using biotinylated synthetic microRNAs. *Methods* **43**, 162–165 (2007).
69. Ørom, U.A., Nielsen, F.C. & Lund, A.H. MicroRNA-10a binds the 5′UTR of ribosomal protein mRNAs and enhances their translation. *Mol. Cell* **30**, 460–471 (2008).
70. 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).
71. Lal, A. *et al.* p16INK4a translation suppressed by miR-24. *PLoS ONE* **3**, e1864 (2008).
72. Lal, A. *et al.* miR-24-mediated downregulation of H2AX suppresses DNA repair in terminally differentiated blood cells. *Nat. Struct. Mol. Biol.* **16**, 492–498 (2009).
73. Neilson, J.R., Zheng, G.X., Burge, C.B. & Sharp, P.A. Dynamic regulation of miRNA expression in ordered stages of cellular development. *Genes Dev.* **21**, 578–589 (2007).
74. Gu, S., Jin, L., Zhang, F., Sarnow, P. & Kay, M.A. Biological basis for restriction of microRNA targets to the 3′ untranslated region in mammalian mRNAs. *Nat. Struct. Mol. Biol.* **16**, 144–150 (2009).

© 2010 Nature America, Inc. All rights reserved.