Identification and Analysis of MicroRNAs

Shveta Bagga and Amy E. Pasquinelli

Molecular Biology Section

Division of Biology 0349

Bonner Hall, Room 4118

9500 Gilman Drive

University of California, San Diego

La Jolla, CA 92093-0349

Phone: 858-822-3006

FAX: 858-822-3021

E-mail: apasquin@ucsd.edu

sbagga@ucsd.edu

INTRODUCTION

The first microRNA (miRNA) gene was uncovered in 1993. After languishing in near obscurity for almost a decade, this gene is now recognized as the founding member of a new class of regulatory RNAs that control gene expression in all multicellular organisms. MicroRNA genes express ~22 nucleotide (nt) RNAs that regulate the expression of protein coding genes containing sequences of antisense complementarity. The intense interest in understanding the role of miRNAs in regulating gene expression has fueled the development of new methods to study how these tiny RNA genes are expressed and function. In this chapter, we present a brief history outlining the discovery of miRNAs and the current model for their biogenesis and mode of action. We then describe experimental approaches used to analyze miRNA expression patterns and regulatory functions.

EXPRESSION AND FUNCTION OF miRNAS

Discovery of miRNAs through nematode genetics

Forty years ago Sydney Brenner proposed adoption of the microscopic nematode *Caenorhabditis elegans* for studying the genetic basis of animal development and behavior. Not only has the worm proven to be a model experimental system for identifying the genes responsible for controlling cell fate and function, but it also enabled the discovery of an entirely unexpected class of genes and a novel regulatory mechanism. The first microRNA gene was uncovered through classical genetic methods to identify a

mutation responsible for abnormal development of certain worm cells. The Ambros lab found that the developmental defects resulted from mutation of the *lin-4* (lin = lineage) gene, which encoded a 21 nucleotide (nt) regulatory RNA (1). This type of gene was unprecedented but opportune work from the Ruvkun lab on another developmental gene, *lin-14*, provided the necessary clues for predicting how a tiny RNA product might control gene expression (1,2). The *lin-4* RNA recognizes sites of imperfect complementarity in the 3' untranslated region (UTR) of the *lin-14* messenger RNA (mRNA) and halts protein expression (Fig. 1). Insufficient *lin-4* RNA or deletion of the target sites in the *lin-14* 3' UTR leads to failed down-regulation of LIN-14 protein expression at the appropriate time and, thus, abnormal development (1-4). Although the LIN-14 protein disappears in response to the *lin-4* RNA, the *lin-14* mRNA remains and continues to associate with polysomes, indicating that translational inhibition is the mechanism at work (2,5).

Another target of *lin-4* regulation, the *lin-28* protein-coding gene, was also discovered by the Ambros lab, providing another example of a developmental gene under post-transcriptional control by the tiny RNA (6). Yet, the question of whether this novel mode of gene regulation was restricted to nematodes persisted until the turn of the century (7). The identification of the *let-7* (let=lethal) gene in *C. elegans* as another 22nt RNA that regulates the expression of protein-coding genes containing 3' UTR target sites raised the possibility that tiny RNA genes might abound in worms and beyond. Indeed, the remarkable conservation of the *let-7* RNA sequence enabled the Ruvkun lab to establish that this gene is expressed in diverse animals, including fruitflies, molluscs, sea urchins, zebrafish and humans (8). Moreover, temporally regulated expression of the *let-7* RNA

and potential target sites in *lin-41* homologs in all species assayed implied that this RNA gene may be essential for development of many animal species (8).

Around the time *let-7* RNA was discovered, another type of tiny RNA was gaining fame. In 1998, Fire and Mello reported that injection of double-stranded RNA (dsRNA) into *C. elegans* could elicit the degradation of homologous mRNA and, thus, potently inhibit gene expression in a process termed RNA interference (RNAi) (9). Shortly thereafter, several groups found that the dsRNA is cleaved to ~22nt small interfering RNAs (siRNAs) that serve as the guides to target complementary mRNA sequences for destruction (10-14). It was clear that tiny RNAs could be powerful regulators of gene expression and soon hundreds of ~22nt RNA genes were uncovered in animal and plant genomes (15-20). The small size of these regulatory RNAs inspired the name microRNA (21), and genes encoding these RNAs appear to be present in all multicellular organisms (22).

Transcription of miRNAs

Despite our relatively brief awareness of their existence, impressive progress has been made in understanding how miRNAs are expressed and function (22). The ~22nt, mature forms of miRNAs arise from multiple processing steps of longer substrate RNAs. So far, little is known about the composition of the initial miRNA transcripts, called primary miRNAs (pri-miRNAs). Both Pol II and Pol III promoters have been used to drive ectopic expression of pri-miRNAs (23-25). Recently, the Kim lab presented direct evidence that Pol II transcribes several mammalian miRNAs (26). Additionally, a few

complete pri-miRNAs that have been characterized show hallmarks of Pol II transcription - they apparently undergo 5'-end capping, 3'-end polyadenylation and splicing (26-28). These first examples of pri-miRNAs are over one thousand nucleotides long - remarkably lengthy transcripts to serve as substrates for ~22nt RNA products!

Many miRNAs are restricted to specific developmental periods or tissue types. At least in some cases, regulated expression of miRNAs is attributable to transcriptional control. Predicted promoter sequences of the *C. elegans let-7* gene can confine expression of green fluorescent protein (GFP) to late larval and adult stages - the same time in development when mature *let-7* is present (29). The transcriptional control sequences for *lys-6* miRNA, which regulates neuronal asymmetry in *C. elegans*, restricts GFP expression to a subset of neurons (30). Although these examples of predicted miRNA promoters directing protein expression support the likelihood that they recruit Pol II, the specialized transcription factors that afford temporal or spatial control are yet to be identified for any miRNA gene.

Processing of miRNAs

Generation of the mature miRNA form requires multiple processing and cellular transportation events (Fig. 2). In animals, the nuclear localized ribonuclease (RNase) Drosha clips the ~65nt hairpin miRNA precursor (pre-miRNA) from the primary transcript (31). The pre-miRNA is shuttled by Exportin-5 to the cytoplasm for final processing by the RNase Dicer (32-38). This enzyme appears to measure ~22nt from the 5' and 3' ends of the hairpin to position a staggered cut through both strands of the stem

(25,31,39-41). Typically, only one half of the resulting duplex is retained. Thermodynamic arguments have been made to explain the choice for which strand persists. The 5' end that is more easily peeled away from its antisense is favored for incorporation into a stable complex and, by default, the other half is unprotected and degraded (42,43).

Pri-miRNAs contain sequences and structures important for processing and generation of the functional ~22nt form. However, truncated pri-miRNA substrates, even the hairpin precursors, can suffice as substrates to produce mature miRNAs when over-expressed from heterologous constructs (23-25,31,35). In the endogenous situation, processing may be a critical control point in miRNA biogenesis. Deletion of cis-acting sequences in pri-miRNA transcripts or depletion of trans-acting processing factors can inhibit miRNA maturation (27,31-34). In some cases, the miRNA substrates accumulate *in vivo*, indicating that transcription of a miRNA gene and production of the mature form are not necessarily coupled.

Function of miRNAs

Mature miRNAs inhibit expression of genes containing sequences of antisense complementarity. In animals, the primary mechanism of gene regulation concurs with the original model proposed for the *lin-4* miRNA and *lin-14* mRNA in *C. elegans* (1,2). Imperfect base-pairing between the miRNA and sequences in the 3' UTR of the target mRNA results in inhibited protein expression (2,5,44). It remains to be determined how

partial base-pairing between miRNAs and target sequences results in blocked protein production.

In plants, many miRNAs exhibit perfect, or nearly complete, base-pair complementarity to their target mRNAs (45,46). Not only does this feature of plant miRNAs make it simpler to predict specific targets, but it also results in target degradation (46-50). Animal miRNAs can also direct mRNA destabilization if they share near perfect complementarity with target sequences (24,51-54). In fact the vertebrate miRNA, miR-196, can form a complete duplex with sequences in HOXB8 mRNA and direct degradation of this target (55). The general model holds that miRNAs can regulate gene expression by either translational inhibition or mRNA destabilization, depending on the nature of the duplex formed with the target sequences.

EXPERIMENTAL IDENTIFICATION OF miRNAS

The non-coding nature and the extraordinarily small size of miRNAs make their detection challenging. For a long time, the conventional cloning and identification techniques and the gene prediction databases were clearly biased for long protein coding sequences. The discovery of tiny RNAs in *C. elegans* and elucidation of the RNAi mechanism led several groups to adopt novel or modified conventional methods to detect miRNAs.

Genetic screening

The discovery of pioneer members of the miRNA family- *lin-4* and *let-7,* demonstrated the potential of classical genetic screens in detection of miRNAs (1,2,56). Although time consuming and labor intensive, identification of a miRNA through a genetic screen can readily give important clues about its function and gene targets. Also, rare and non-conserved miRNAs, which usually evade cloning and computational detection, can be identified by genetic methods. A loss-of-function screen led to the identification of *C. elegans* miRNA *lys-6,* which controls left/right neuronal asymmetry (30). Gain-of-function genetic screens based on mutations in negatively regulated targets or forcing altered expression of miRNAs have also contributed to the detection of new miRNAs (48,51,57,58). These studies emphasized an important distinction of present day genetic mapping- to look for non-coding, short stem-loop structures in addition to conventional open reading frames (ORFs). Increasing efforts towards developing full genome databases will facilitate the identification of more miRNAs through genetic screenings. Taking into account the abundance of miRNAs, it would not be surprising if many of the previously uncharacterized loci in genetic screens could be ascribed to miRNAs.

Biochemical cloning

Direct cloning of expressed miRNAs by the Ambros, Bartel and Tuschl labs led to the identification of the first populations of miRNAs in worms, flies and humans (15,16,20). Several unique as well as highly conserved miRNAs, like *let-7*, were detected in these initial cloning efforts. Northern blot analyses of cloned miRNAs revealed both tissue-

specific and stage-specific miRNAs, emphasizing their role in developmental timing and tissue specifications. The phylogenetic distribution of miRNAs was further expanded by the cloning of plant miRNAs (17-20,59). Cloning of tiny RNAs from specific ribonucleoprotein complexes also identified several novel miRNAs (60). To date, biochemical cloning has led to the identification of hundreds of distinct miRNAs (15-17,19,20,60-64).

An important characteristic that emerged from biochemical cloning and complied with *lin-4* and *let-7* sequences is the existence of animal miRNAs as a part of ~70 nucleotide stem-loop precursors (1,8,56). Processing of mature miRNAs from hairpin precursors is now considered a signature of animal miRNA genes (21). Although plant miRNAs also derive from precursors, composition of these substrates is not as well defined (19). The miRNA sequence can reside on either arm of the stem-loop structure, and hence the location on the precursor is not a determinant of its excision by Dicer. Cloning of miRNAs that are clustered in the genome and identification of some in expressed sequence tag (EST) databases hinted that miRNA precursors might derive from longer primary transcripts (15,20,28,35,61,65). After the discovery of Drosha, it was speculated that specific cleavage of primary transcripts determines the correct register of Dicer action and hence the mature ends of miRNAs are determined at the level of primary transcripts (31).

As the result of being RNase III Dicer products, miRNAs are cloned based on their three distinguishing features: a length of about 22nt, a 5’-terminal monophosphate and a 3’-

terminal hydroxyl group (10,33,66). The general protocol for miRNA cloning involves size fractionation of an RNA population followed by ligation with adapter molecules (Fig. 3) (10,15,16,20). The chimeric RNA is then subjected to RT-PCR, cloned and sequenced. One of the advantages of biochemical cloning of miRNAs is that the expressed miRNA population from any tissue or at any stage of development can be readily detected. Cloning of mouse brain-tissue miRNAs revealed probable orthologs of *C. elegans lin-4* RNA, and the mouse sequences revealed probable Drosophila orthologs as well (61). Homologs of the *lin-4* gene had, thus far, not surfaced from informatic searches of other organisms. Although powerful in terms of revealing expressed miRNAs directly, detection by cloning has an inevitable drawback of selecting clones of breakdown products of abundant cellular RNAs. Hence to qualify as a miRNA, a small cloned RNA should be able to form a stem-loop precursor structure with its flanking sequences and show conservation in related species (21). Endogenous siRNAs are usually distinguished from miRNAs by extended dsRNA structure of their precursors and by displaying less sequence conservation (21,67).

Interestingly, cloning efforts in *C. elegans* and Drosophila led to the identification of new categories of non-coding RNAs designated as “tiny noncoding RNAs” (tncRNAs) and “repeat associated small interfering RNAs” (rasiRNAs) (63,64). The 24-26 nt rasiRNAs apparently derive from various repetitive sequence elements including retrotransposons, DNA transposons, satellite and microsatellite sequences, complex as well as vaguely characterized repetitive sequence motifs (64). The tncRNAs are similar in size to miRNAs but are not processed from stem-loop precursors and do not have orthologs in

other species (63). Although some of tncRNAs exhibit temporal expression patterns, their exact role and significance awaits further experimentation.

Informatics

Although biochemical cloning led to the identification of several hundreds of new miRNAs, it is limited for identifying rare miRNAs or those that are triggered by specific environmental conditions. The availability of full genome databases of several organisms enabled the development of informatics approaches for identification of new miRNAs.

The fortuitous discovery of the first conserved miRNA, *let-7,* demonstrated the potential of simple homology searches using BLASTN (8). Homology searches with cloned miRNAs also revealed orthologs and paralogs in various organisms (15,16,20). A simple homology-based strategy originally involved the analysis of intergenic sequences among related organisms using the RNA folding program "mfold" (16,68). The output was scanned by eye for miRNA characteristic stem-loop structures and the expression was confirmed by Northern blotting. The proximal location of several miRNA genes prompted the search for new miRNAs adjacent to the previously identified ones (20,64,69,70). This approach is most suitable for identification of rapidly evolving miRNA genes, which are proximal to each other but are too divergent in sequence to be detected by general methods (22).

An important advance in detection of miRNA genes has been achieved by development of new computational approaches (63,67,71-76). All the programs primarily utilize

sequence conservation, presence of stem-loop structures, and intergenic location of miRNAs as basic criteria. One of the more sensitive programs, “MiRscan” has been applied to vertebrate and nematode genomes to identify new miRNA genes (67,74). The MiRscan program was developed by using the 50 cloned miRNAs from *C. elegans* as the training set (16,20). Based on its similarity to the training set a score is assigned to each putative genomic candidate that is identified by conserved stem-loop structures. The evaluation is based on seven features: base pairing of the miRNA portion of the fold-back, base-pairing of the rest of the fold-back, stringent sequence conservation in the 5’ half of the miRNA, slightly less stringent sequence conservation in the 3’ half of the miRNA, sequence biases in the first five bases of the miRNA, a tendency toward having symmetric internal loops and bulges in the miRNA region, and the presence of 2-9 consensus base pairs between the miRNA and the terminal loop region with a preference for 4-6 base pairs. The accuracy of MiRscan predictions has been further improved by the inclusion of conserved elements upstream of miRNA precursors (69). The successful application of this program, although developed using nematode miRNAs, to vertebrates demonstrated its universal application. It also emphasized that, despite sequence variations of miRNAs among diverse animals, their generic features are broadly conserved.

Using a reference set of Drosophila pre-miRNA sequences, another program called “miRseeker” identified novel miRNA genes (73). The miRseeker algorithm detects insect miRNA genes using a three-step filter strategy. The first step involves extraction of candidate genes using conserved and non-genic regions of *D. melanogaster* and *D.*

pseudoobscura genomes. The next step identifies and ranks the stem-loop structured regions based on the helical length and free energy values. Finally, high scoring regions averaged for two genomes are evaluated for divergence using the determinants of a reference set. In principle, miRseeker should be applicable to analysis of other sets of sequenced genomes of related organisms.

Recent informatics approaches specifically designed to detect plant miRNAs identified several new candidates (71,76). These strategies are similar to MiRscan and miRseeker in terms of using homologous fold-back sequences conserved between *Arabidopsis* and *Oryza sativa*. However, the parameters constraining the selection of fold-back structures were specifically designed for plant miRNAs. The MIRcheck algorithm utilizes the sequences and structures of putative miRNA hairpins and 20mers within them (71). MIRcheck selects the candidates by restricting the number of unpaired, bulged or asymmetrically unpaired, consecutive unpaired nucleotides, and the length of the hairpin. Unlike other programs, MIRcheck does not restrict based on pattern or extent of base pairing outside the 20mer sequence, a feature typical to plant miRNAs. Several of the plant miRNAs identified by this approach were confirmed by expression and target mRNA degradation (71).

ANALYSIS OF miRNAs

Many miRNAs exhibit diverse temporal and spatial expression patterns. Additionally, the relative levels of a particular miRNA can vary several orders of magnitude among

different cell types. Adaptations of traditional molecular techniques as well as novel methods have been developed to analyze when, where and how much of a specific miRNA exists and what is its biological function.

Expression patterns

Northern blot and RNase protection assays yielded the first molecular evidence for the existence of a ~22nt RNA product. A specific tiny RNA product was present in wildtype but not *lin-4* mutant worms, and this RNA reappeared upon rescue of the mutant with a transgene containing just 693nt of *lin-4* genomic sequence (1). Typical analyses for miRNA expression by Northern blots utilize high percentage (10-15%) polyacrylamide gel electrophoresis (PAGE), which enables detection of the mature and precursor forms of the miRNA (Fig. 4) (1,77). The relative level of a mature miRNA can be readily assessed by sampling total RNA from particular tissues, developmental time points or experimental conditions. However, Northern blotting to detect specific miRNAs can be labor intensive and insensitive to low level miRNAs.

Computational prediction of miRNA genes avoids the cloning bias of detecting the more abundant species. Confirmation of a predicted miRNA can be experimentally challenging, though, if the gene is weakly expressed or only activated under particular conditions. PCR-based approaches were developed to help validate the expression of miRNAs identified by informatics. Strong evidence for the existence of several elusive miRNAs was provided by a PCR protocol, which involves amplifying miRNA sequences from bulk miRNA cDNA libraries via the common adaptor sequences (67,75). Real-time

PCR assays have been employed for relatively high throughput analysis of miRNA precursor expression (78). In at least some cases, the level of precursor accurately reflected that of mature, as indicated by Northern analyses. More recently an exceptionally sensitive and quantitative method was reported for detecting precursor or mature miRNAs (79). The Invader miRNA assay can detect as little as 20,000 molecules of a specific miRNA and has been used to show that the amounts of human let-7a miRNA vary over several orders of magnitude among different tissues (79).

Microarray technology offers an efficient and sensitive method to assess global changes in miRNA expression patterns. Microchips containing oligonucleotides corresponding to miRNA sequences have been used to screen various cell types to uncover the miRNA profile (80,81). Additionally, this type of miRNA profiling was used to identify distinctions between normal human B cells and those derived from chronic lymphocytic leukemia cells (82). Since their discovery, miRNA genes have been considered possible disease candidates (83). High throughput profiling of miRNA expression patterns offers a powerful tool for correlating specific miRNAs with altered cell biological states.

Detection of miRNAs *in vivo* is particularly challenging considering the small size and potentially low abundance. Nonetheless, *in situ* hybridization results have indicated localized expression for a few miRNA transcripts. In plants, there is an inverse correlation between expression of specific miRNAs and proposed targets in specific tissues, supporting the model that these RNAs negatively regulate protein expression to control development (84-86). Localized expression of miR-10 in Drosophila embryos

indicates a role for this miRNA in regulating genes in the thoracic and abdominal primoridia, although specific targets of miR-10 are yet to be identified (87).

An indirect method for analyzing temporal and spatial expression of miRNA genes is to fuse predicted miRNA promoter sequences to a reporter gene, such as green fluorescent protein (GFP). This technique revealed tissue and developmental regulation of specific miRNA promoters in *C. elegans* that agreed with predictions about the function of the miRNAs (29,30,88). The *lys-6* miRNA was discovered as a gene that controls neuronal asymmetry in *C. elegans* by repressing expression of a transcription factor in a left taste neuron (30). Consistent with the proposed function of *lys-6*, a GFP reporter fused to the promoter for this miRNA gene is expressed in the left, but not right, neuron (30). These types of reporter experiments are very useful for predicting when and where a miRNA promoter functions as well as for studying its transcriptional control (29,30,88). However, the promoters and functions of most miRNAs are yet to be identified and, thus, caution is warranted for interpreting expression patterns based on fusions to miRNAs for which little is known about the natural biological role.

The *in situ* and reporter experiments described above can be used to indicate when and where a miRNA gene is active, but they do not demonstrate the production of functional miRNAs. Regulated processing and stabilization of some miRNAs may also influence their ability to control gene expression. An ingenious method to detect functional miRNAs *in vivo* was developed by the Cohen lab to show spatial and temporal expression of the Drosophila *bantam* miRNA (51). The "sensor" strategy is based on the

demonstration that miRNAs will direct degradation of target mRNAs containing sites of perfect antisense complementarity (Fig. 5) (24,48,52-54). A GFP-reporter gene containing *bantam* miRNA complementary sites was down-regulated in response to *bantam* expression. Thus, the presence of a functional miRNA can be assayed *in vivo* without the knowledge of its natural targets. Identification of specific miRNA expression patterns will greatly facilitate determination of biological functions.

Functional roles

The first miRNAs, *lin-4* and *let-7*, were initially discovered as genes essential for regulating developmental timing in *C. elegans* (1,56). Since the vast majority of RNAs to join the miRNA family were isolated by biochemical or computational means, biological functions are yet to be assigned. Considering their abundance, it is not surprising that miRNA genes are now being uncovered in mutant screens. Perhaps the lack of traditional gene structure allowed miRNAs to escape previous detection, but now mutations in miRNA genes account for broad-ranging phenotypes, including disrupted neuronal asymmetry, misregulated cell death, abnormal fat metabolism and cellular patterning defects (18,28,30,51,58,84,88,89). Isolation of genetic mutations in specific miRNA genes not only aids in determining biological function but also is valuable for identifying direct targets of regulation. A genetic suppressor screen of the *let-7* mutant revealed *lin-41* as a target of negative regulation, which then led to the recognition of *let-7* complementary sites in the 3'UTR of *lin-41* mRNA (56,90).

In many systems, targeted disruption or isolation of mutations in specific miRNAs is prohibitively laborious. Furthermore, homology among several groups of miRNAs suggests that redundancy may obscure phenotypes resulting from mutation of just one member. Overexpression or ectopic expression is an efficient alternative to study the function of particular miRNAs. The validity of this approach was established by introducing high copies of the *lin-4* gene to worms and observing developmental defects opposite of the *lin-4* loss of function phenotypes (4). More recently, ectopic expression of miR-181 in mouse hematopoietic stem cells biased their differentiation into B-lineage cells (23). Thus direct targets of miR-181 may be predicted by focusing on distinct changes in gene expression in the B-lineage pathway.

The biological function of specific miRNAs can also be revealed by inhibition with antisense oligonucleotides. Injection of antisense DNA oligonucleotides corresponding to specific miRNAs into Drosophila embryos resulted in developmental defects (91). More recently, 2'-O-methyl oligonucleotides were shown to potently block the function of targeted miRNAs in Drosophila, human cell and *C. elegans* systems (92,93). The 2'-O-methyl modification protects the oligonucleotide against cellular RNases (94). Base-pairing of the oligonucleotide to a miRNA titrates the miRNA from its endogenous targets, thus, revealing the loss of function phenotype. Although delivery of the antisense oligonucleotide can be technically prohibitive (92), this method of miRNA inhibition offers an efficient means to uncover the biological roles of miRNAs for which only the mature sequence is known.

PAIRING OF miRNAS WITH TARGETS

The combination of cloning and computational approaches has likely enabled identification of the majority of miRNAs (22). However, as of yet, only a few miRNAs have been paired with their *bona fide* targets. Identification of direct miRNA targets is essential for understanding their diverse functions.

In addition to the discovery of pioneer miRNAs, the credit for the discovery of the first miRNA targets also goes to classical genetics. Long before the broad significance of tiny regulatory RNAs was appreciated, the functional pairing of *lin-4* RNA with its target *lin-14* mRNA was proposed (1,2). The 3'UTR of *lin-14* mRNA had partial complementarity to *lin-4* RNA and was sufficient for temporal regulation of a reporter gene. The *let-7* target, *lin-41*, also supports the model, both in terms of partial complementarity and reporter gene regulation with the 3'UTR (56,90,95). The opposite phenotypes of *lin-4* and *lin-14* mutants helped pinpoint *lin-14* as a direct target of *lin-4* mediated negative regulation (96,97). A handful of other *bona fide* miRNA target genes were identified through genetic screens (28,30,51,57,88,89,98,99). However, for the majority of other miRNAs either mutants are not known or their mutant phenotypes are not apparent. Also, the small size and imperfect nature of base pairing, particularly in case of animal miRNAs, hampers straightforward prediction of miRNA targets.

Target identification and validation for plant miRNAs

Exact complementarity between miR171 and an mRNA target in *Arabidopsis* indicated that target prediction might be less complicated for plant compared to animal miRNAs (17,19,48). Indeed, "near perfect complementarity" appears to be a general rule for plant miRNA targets (46,71,76). Initially, targets were identified by searching annotated *Arabidopsis* mRNAs for 0-4 mismatches to specific miRNAs (19,46). Conservation of the predicted mRNA target sequences in rice and low hits with a random cohort of tiny RNA sequences strengthened the validity of these proposed plant targets (46,71). Plant miRNA targets show a clear bias towards transcription factors involved in cell differentiation and developmental patterning (46,71). In comparison to sequences regulated by animal miRNAs, most plant miRNA-target interactions exhibit two general distinctions: (1) Plant miRNA target sites are primarily found within open reading frames, (2) Multiple target sites within the same target mRNA are not detected in plants. These features may have significant functional implications for plant miRNAs - they favor an RNAi-like mechanism, as opposed to translational control, to inhibit gene expression (45).

A sensitive computational approach identified several novel plant miRNA targets belonging to families of transcription factors as well as other genes like ATP sulfurylase, laccase and superoxide dismutase (71). This approach allowed for gaps and mismatches between mRNA:miRNA duplexes but constrained the candidate targets to conservation between *Arabidopsis* and *Oryza*. Validation of predicted target sequences is facilitated by the fact that many plant miRNAs direct cleavage of their complementary mRNA

targets. The 3' cleavage product of the target, which maps to the tenth nucleotide of the miRNA and has a characteristic phosphate at its 5' end, can be cloned and sequenced (47,48,50,71,76). Although absence of a 3' cleavage product may suggest a false or alternatively regulated target, its presence is a convincing confirmation of regulation by a specific miRNA.

Informatics approaches for target identification in animals

New computational methods have matched animal miRNAs with numerous target genes, although many of the pairings still await experimental confirmation (100-104). The small number of validated miRNA targets in animals makes the development of reliable algorithms particularly challenging. As a starting point, most computational methods rely on conserved complementary sites within 3'UTRs of potential target genes.

Identification of the *hid* gene as a *bantam* miRNA target exhibited the potential of computational approaches for identifying targets in Drosophila (51,101). This approach was based on the presence of miRNA target sites in 3'UTRs of target mRNAs and their relatively better complementarity to the 5' end of miRNAs. The first step for identifying genome wide Drosophila miRNA targets involved generation of a conserved database comparing 3'UTRs of *D. melanogaster* and *D. pseudoobscura* (101). The candidate target genes were then scored based on their free energy of base pairing with the miRNAs, as determined by mFold (68,101). The combination of sensitive sequence databases with that of the RNA folding algorithm confirmed the known targets and identified several new ones (101). A striking feature of predicted targets was the

presence of clusters of functionally related targets regulated by specific miRNAs. This included *Notch* target genes for *mir-7*, proapoptotic genes for *mir-2* family and metabolic pathway enzymes for *mir-277* (101). Another computational method for target identification, miRanda, relies on evolutionary relationships between miRNAs and their targets using three insect genomes (104). The miRanda approach is a three phase method involving sequence matching of miRNA:mRNA pairs, estimating the energetics of the physical interaction and using evolutionary conservation as an informational filter. This method suggested both multiplicity (one miRNA targets several genes) and cooperativity (one gene targeted by several miRNAs) as general features of miRNA regulated gene expression (104).

The TargetScan algorithm predicted more than 400 target genes for mammalian miRNAs (102). TargetScan also combines thermodynamics-based modeling of RNA:RNA duplex interactions with comparative sequence analysis to predict miRNA targets conserved across multiple genomes. One of the criteria for filtering miRNA:mRNA pairs using this algorithm is exact complementarity between 2-8 bases of miRNAs counted from the 5'end of miRNA. The folding energy of each pair is calculated using RNAeval (105), after extending the pairing as far as possible. Each 3' UTR is then scored based on the number of miRNA:mRNA matches, free energy of interaction, and number and affinity of complementary sites. Comparative ranking of UTRs among different organisms sorted on the basis of this score then predicts the target. TargetScan revealed that in contrast to plant miRNA targets, only a small fraction of predicted mammalian targets participate in developmental control; they seem to regulate broadly diverse biological processes (102).

Another computational program, DIANA-microT, was developed to study the rules of single miRNA:MRE (target mRNA) pairing and to predict targets containing a single complementary site (103). Similar to other computational programs, DIANA-microT identifies the putative targets by estimating the binding energies between conserved miRNA:MRE pairs. A difference from other programs is that it also takes into account the G-U wobble dinucleotide pairs for calculating binding energies (103).

Computational identification, based on favorable energy statistics and evolutionary relationships, corroborated by experimental evidence provides reasonable substantiation of miRNA target validity. An important consideration in computational target prediction and confirmation is the use of correctly annotated genes. Already, ambiguity in annotated genes misguided attempts to validate a miRNA target (106,107). Absent or incomplete annotations of 3' UTRs also hinder the comprehensive analysis of miRNA targets.

Heterologous reporter assays are most commonly used for validation of miRNA targets (Fig. 6). Typically, a reporter gene, such as luciferase or β-galactosidase, is fused to sequences containing the miRNA complementary region from a putative target. Expression of the reporter is observed in the presence or absence of the proposed regulatory miRNA (6,56,90,98,99,101-103). Downregulation of a reporter gene in the presence of the miRNA indicates the presence of regulatory sites in the fragment used for fusion. However, concerns of extraneous effects due to multiple copies of complementary sites or very long UTR regions should be kept in mind. Also, failure to demonstrate

regulation of a heterologous reporter may reflect factors independent of the miRNA: (1) the cell system might not express additional co-factors or adequate levels of the miRNA to appreciably affect reporter expression, (2) additional mRNA elements are required but not included in the UTR segment of the reporter, and (3) steric hindrance imposed by the fusion of reporter on the putative sites blocks interaction with the miRNA.

SUMMARY and OUTLOOK

Genetics introduced us to the existence of miRNAs. Biochemical and molecular methods were essential for establishing the existence of vast numbers of miRNA genes in diverse organisms. Bioinformatic approaches contributed to the identification of additional, elusive, miRNAs as well as to the prediction of miRNA target genes. Combined experimental and computational methods will be required to advance our rudimentary understanding of miRNA expression and function. Central questions remain: How are transcription and processing of miRNAs regulated? How do miRNAs find their appropriate targets? What is the mechanism by which miRNAs regulate expression of their targets? The discovery of miRNAs established a new paradigm for gene regulation and understanding the biological roles of these abundant RNA genes undoubtedly will be a challenging endeavor.

ACKNOWLEDGMENTS

We are grateful to members of the Pasquinelli lab for comments and inspiration.

Research support is provided by the National Institutes of Health (GM071654-01) and the Searle Scholar, Peter Gruber and V Foundations.

FIGURE LEGENDS

Figure 1. Conserved sites in the 3'UTR of the *C. elegans lin-14* mRNA are complementary to the *lin-4* miRNA (2). Shaded blocks in the *lin-14* 3'UTR indicate sequences of high homology (at least 10 nucleotides of exact conservation) between the related nematodes *C. elegans* and *C. briggsae*. The striped blocks 1-7, represent regions of partial complementarity to the *lin-4* miRNA. The duplexes are shown with the *lin-14* top strand reading 5' to 3' base-paired with the bottom strand *lin-4* miRNA.

Figure 2. A model of miRNA biogenesis and function. The relatively long primary transcripts, called pri-miRNAs, are initially transcribed from miRNA genes (35). The pri-miRNAs are processed by the RNase Drosha to hairpin precursors (25,31,35). The precursors are recognized by Exportin-5 and delivered to the cytoplasm for maturation to ~22nt RNAs by Dicer (25,32-38). It has been proposed that a helicase activity separates the duplex (42,43), and typically only one half is retained and incorporated into a multi-factor RNA induced silencing complex (RISC) (12,52-54). The degree of complementarity between a miRNA and its target site determines the regulatory mechanism: near perfect base-pairing directs RNA degradation and bulged duplexes mediate translational repression (22). This model is based primarily on work in animal systems and note that organismal differences exist for the protein factors and subcellular location of processing events (22).

Figure 3. MiRNA cloning strategy. Typically, total RNA is fractionated to ~22nt size forms and miRNAs containing 5' phosphate and 3' hydroxyl groups are substrates for ligation to adaptor oligonucleotides (10,15,16,20). The chimeric RNA is subjected to RT-PCR, cloning and sequencing. Legitimate miRNAs match genomic sequences that support formation of a hairpin precursors (21).

Figure 4. Northern analysis of miRNA expression. Total RNA from wildtype worms or worms depleted of dicer was isolated, separated by 11% polyacrylamide gel electrophoresis and subjected to Northern hybridization analysis to detect *let-7* RNA. The ~22nt nucleotide mature form is predominant in wild type worms, whereas the 65nt precursor accumulates in dicer(-) worms (32).

Figure 5. The "sensor" approach to analyze miRNA expression *in vivo* (51). In this example, constructs expressing a reporter protein, such as GFP, fused to 3'UTR sequences +/- miRNA complementary sites are introduced into worms. A ubiquitous promoter drives reporter expression. If the miRNA is absent, such as in early larval development (middle panels), GFP will be detected. However, if the miRNA turns on later in development or in particular cell types, the reporter mRNA will be specifically degraded and GFP will disappear (last panel, top row). This example shows a predicted pattern for *let-7* expression in *C. elegans*: early in development *let-7* is absent and thus GFP is expressed ubiquitously (gray shading of entire worm), including in the 10 hypodermal seam cells, but later in development *let-7* miRNA is produced and shuts off reporter expression, perhaps specifically in the 16 seam cells of adult worms (absence of

gray shading) (29). Importantly, expression of a control reporter lacking the miRNA complementary sites is unaffected by miRNA expression (bottom panels).

Figure 6. Validation of miRNA target predictions. Multiple targets with sites of partial complementarity are often identified for a single miRNA. To test the function of such predictions, sequences containing the complementary elements are fused to reporter genes, such as luciferase. The reporter is assayed for expression in the presence and absence of the candidate miRNA partner. In the depicted example, a miRNA is predicted to recognize sites in the 3'UTR's of 3 different genes. Each UTR is fused to the reporter and the constructs are introduced to cells expressing the miRNA of interest. Only, the site from gene "1" mediates reporter repression via interaction with the miRNA.

REFERENCES

1. Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) *Cell*, **75,** 843-854.

2. Wightman, B., Ha, I. and Ruvkun, G. (1993) *Cell*, **75,** 855-862.

3. Wightman, B., Burglin, T.R., Gatto, J., Arasu, P. and Ruvkun, G. (1991) *Genes Dev*, **5,** 1813-1824.

4. Feinbaum, R. and Ambros, V. (1999) *Dev Biol*, **210,** 87-95.

5. Olsen, P.H. and Ambros, V. (1999) *Dev Biol*, **216,** 671-680.

6. Moss, E.G., Lee, R.C. and Ambros, V. (1997) *Cell*, **88,** 637-646.

7. Wickens, M. and Takayama, K. (1994) *Nature*, **367,** 17-18.

8. 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) *Nature*, **408,** 86-89.

9. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) *Nature*, **391,** 806-811.

10. Elbashir, S.M., Lendeckel, W. and Tuschl, T. (2001) *Genes Dev*, **15,** 188-200.

11. Hamilton, A.J. and Baulcombe, D.C. (1999) *Science*, **286,** 950-952.

12. Hammond, S.M., Bernstein, E., Beach, D. and Hannon, G.J. (2000) *Nature*, **404,** 293-296.

13. Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P. and Sharp, P.A. (1999) *Genes Dev*, **13,** 3191-3197.

14. Zamore, P.D., Tuschl, T., Sharp, P.A. and Bartel, D.P. (2000) *Cell*, **101,** 25-33.

15. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) *Science*, **294,** 853-858.

16. Lee, R.C. and Ambros, V. (2001) *Science*, **294,** 862-864.

17. Llave, C., Kasschau, K.D., Rector, M.A. and Carrington, J.C. (2002) *Plant Cell*, **14,** 1605-1619.

18. Park, W., Li, J., Song, R., Messing, J. and Chen, X. (2002) *Curr Biol*, **12,** 1484-1495.

19. Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. and Bartel, D.P. (2002) *Genes Dev*, **16,** 1616-1626.

20. Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001) *Science*, **294,** 858-862.

21. Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M. *et al*. (2003) *Rna*, **9,** 277-279.

22. Bartel, D.P. (2004) *Cell*, **116,** 281-297.

23. Chen, C.Z., Li, L., Lodish, H.F. and Bartel, D.P. (2004) *Science*, **303,** 83-86.

24. Zeng, Y., Wagner, E.J. and Cullen, B.R. (2002) *Mol Cell*, **9,** 1327-1333.

25. Zeng, Y. and Cullen, B.R. (2003) *Rna*, **9,** 112-123.

26. Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S.H. and Kim, V.N. (2004) *Embo J*, **in press**.

27. Bracht, J., Hunter, S., Eachus, R., Weeks, P. and Pasquinelli, A.E. (2004) *Rna*, **10**.

28. Aukerman, M.J. and Sakai, H. (2003) *Plant Cell*, **15,** 2730-2741.

29. Johnson, S.M., Lin, S.Y. and Slack, F.J. (2003) *Dev Biol*, **259,** 364-379.

30. Johnston, R.J. and Hobert, O. (2003) *Nature*, **426,** 845-849.

31. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. *et al*. (2003) *Nature*, **425,** 415-419.

32. Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G. and Mello, C.C. (2001) *Cell*, **106,** 23-34.

33. Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T. and Zamore, P.D. (2001) *Science*, **293,** 834-838.

34. Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J. and Plasterk, R.H. (2001) *Genes Dev*, **15,** 2654-2659.

35. Lee, Y., Jeon, K., Lee, J.T., Kim, S. and Kim, V.N. (2002) *Embo J*, **21,** 4663-4670.

36. Yi, R., Qin, Y., Macara, I.G. and Cullen, B.R. (2003) *Genes Dev*, **17,** 3011-3016.

37. Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E. and Kutay, U. (2004) *Science*, **303,** 95-98.

38. Bohnsack, M.T., Czaplinski, K. and Gorlich, D. (2004) *Rna*, **10,** 185-191.

39. Zhang, H., Kolb, F.A., Jaskiewicz, L., Westhof, E. and Filipowicz, W. (2004) *Cell*, **118,** 57-68.

40. Zhang, H., Kolb, F.A., Brondani, V., Billy, E. and Filipowicz, W. (2002) *Embo J*, **21,** 5875-5885.

41. Basyuk, E., Suavet, F., Doglio, A., Bordonne, R. and Bertrand, E. (2003) *Nucleic Acids Res*, **31,** 6593-6597.

42. Khvorova, A., Reynolds, A. and Jayasena, S.D. (2003) *Cell*, **115,** 209-216.

43. Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N. and Zamore, P.D. (2003) *Cell*, **115,** 199-208.

44. Seggerson, K., Tang, L. and Moss, E.G. (2002) *Dev Biol*, **243,** 215-225.

45. Bartel, B. and Bartel, D.P. (2003) *Plant Physiol*, **132,** 709-717.

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

47. Kasschau, K.D., Xie, Z., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A. and Carrington, J.C. (2003) *Dev Cell*, **4,** 205-217.

48. Llave, C., Xie, Z., Kasschau, K.D. and Carrington, J.C. (2002) *Science*, **297,** 2053-2056.

49. Xie, Z., Kasschau, K.D. and Carrington, J.C. (2003) *Curr Biol*, **13,** 784-789.

50. Tang, G., Reinhart, B.J., Bartel, D.P. and Zamore, P.D. (2003) *Genes Dev*, **17,** 49-63.

51. Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B. and Cohen, S.M. (2003) *Cell*, **113,** 25-36.

52. Doench, J.G., Petersen, C.P. and Sharp, P.A. (2003) *Genes Dev*, **17,** 438-442.

53. Hutvagner, G. and Zamore, P.D. (2002) *Science*, **297,** 2056-2060.

54. Zeng, Y., Yi, R. and Cullen, B.R. (2003) *Proc Natl Acad Sci U S A*.

55. Yekta, S., Shih, I.H. and Bartel, D.P. (2004) *Science*, **304,** 594-596.

56. Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. and Ruvkun, G. (2000) *Nature*, **403,** 901-906.

57. Hipfner, D.R., Weigmann, K. and Cohen, S.M. (2002) *Genetics*, **161,** 1527-1537.

58. Xu, P., Vernooy, S.Y., Guo, M. and Hay, B.A. (2003) *Curr Biol*, **13,** 790-795.

59. Mette, M.F., van der Winden, J., Matzke, M. and Matzke, A.J. (2002) *Plant Physiol*, **130,** 6-9.

60. Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M. and Dreyfuss, G. (2002) *Genes Dev*, **16,** 720-728.

61. Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. and Tuschl, T. (2002) *Curr Biol*, **12,** 735-739.

62. Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A. and Tuschl, T. (2003) *Rna*, **9,** 175-179.

63. Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T. and Jewell, D. (2003) *Curr Biol*, **13,** 807-818.

64. Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J. and Tuschl, T. (2003) *Dev Cell*, **5,** 337-350.

65. Smalheiser, N.R. (2003) *Genome Biol*, **4,** 403.

66. Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001) *Nature*, **409,** 363-366.

67. Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B. and Bartel, D.P. (2003) *Genes Dev*, **17,** 991-1008.

68. Mathews, D.H., Sabina, J., Zuker, M. and Turner, D.H. (1999) *J Mol Biol*, **288,** 911-940.

69. Ohler, U., Yekta, S., Lim, L.P., Bartel, D.P. and Burge, C.B. (2004) *Rna*, **10,** 1309-1322.

70. Seitz, H., Youngson, N., Lin, S.P., Dalbert, S., Paulsen, M., Bachellerie, J.P., Ferguson-Smith, A.C. and Cavaille, J. (2003) *Nat Genet*, **34,** 261-262.

71. Jones-Rhoades, M.W. and Bartel, D.P. (2004) *Mol Cell*, **14,** 787-799.

72. Bonnet, E., Wuyts, J., Rouze, P. and Van de Peer, Y. (2004) *Proc Natl Acad Sci U S A*, **101,** 11511-11516.

73. Lai, E.C., Tomancak, P., Williams, R.W. and Rubin, G.M. (2003) *Genome Biol*, **4,** R42.

74. Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B. and Bartel, D.P. (2003) *Science*, **299,** 1540.

75. Grad, Y., Aach, J., Hayes, G.D., Reinhart, B.J., Church, G.M., Ruvkun, G. and Kim, J. (2003) *Mol Cell*, **11,** 1253-1263.

76. Wang, X.J., Reyes, J.L., Chua, N.H. and Gaasterland, T. (2004) *Genome Biol*, **5,** R65.

77. Pasquinelli, A.E., McCoy, A., Jimenez, E., Salo, E., Ruvkun, G., Martindale, M.Q. and Baguna, J. (2003) *Evol Dev*, **5,** 372-378.

78. Schmittgen, T.D., Jiang, J., Liu, Q. and Yang, L. (2004) *Nucleic Acids Res*, **32,** e43.

79. Allawi, H.T., Dahlberg, J.E., Olson, S., Lund, E., Olson, M., Ma, W.P., Takova, T., Neri, B.P. and Lyamichev, V.I. (2004) *Rna*, **10,** 1153-1161.

80. Liu, C.G., Calin, G.A., Meloon, B., Gamliel, N., Sevignani, C., Ferracin, M., Dumitru, C.D., Shimizu, M., Zupo, S., Dono, M. *et al*. (2004) *Proc Natl Acad Sci U S A*, **101,** 9740-9744.

81. Krichevsky, A.M., King, K.S., Donahue, C.P., Khrapko, K. and Kosik, K.S. (2003) *Rna*, **9,** 1274-1281.

82. Calin, G.A., Liu, C.G., Sevignani, C., Ferracin, M., Felli, N., Dumitru, C.D., Shimizu, M., Cimmino, A., Zupo, S., Dono, M. *et al*. (2004) *Proc Natl Acad Sci U S A*, **101,** 11755-11760.

83. McManus, M.T. (2003) *Semin Cancer Biol*, **13,** 253-258.

84. Chen, X. (2004) *Science*, **303,** 2022-2025.

85. Juarez, M.T., Kui, J.S., Thomas, J., Heller, B.A. and Timmermans, M.C. (2004) *Nature*, **428,** 84-88.

86. Kidner, C.A. and Martienssen, R.A. (2004) *Nature*, **428,** 81-84.

87. Kosman, D., Mizutani, C.M., Lemons, D., Cox, W.G., McGinnis, W. and Bier, E. (2004) *Science*, **305,** 846.

88. Chang, S., Johnston, R.J., Jr., Frokjaer-Jensen, C., Lockery, S. and Hobert, O. (2004) *Nature*, **430,** 785-789.

89. Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C. and Weigel, D. (2003) *Nature*.

90. Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R. and Ruvkun, G. (2000) *Mol Cell*, **5,** 659-669.

91. Boutla, A., Delidakis, C. and Tabler, M. (2003) *Nucleic Acids Res*, **31,** 4973-4980.

92. Hutvagner, G., Simard, M.J., Mello, C.C. and Zamore, P.D. (2004) *PLoS Biol*, **2,** E98.

93. Meister, G., Landthaler, M., Dorsett, Y. and Tuschl, T. (2004) *Rna*, **10,** 544-550.

94. Inoue, H., Hayase, Y., Imura, A., Iwai, S., Miura, K. and Ohtsuka, E. (1987) *Nucleic Acids Res*, **15,** 6131-6148.

95. Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K. and Slack, F.J. (2004) *Genes Dev*.

96. Ambros, V. and Horvitz, H.R. (1984) *Science*, **226,** 409-416.

97. Ambros, V. (1989) *Cell*, **57,** 49-57.

98. Abrahante, J.E., Daul, A.L., Li, M., Volk, M.L., Tennessen, J.M., Miller, E.A. and Rougvie, A.E. (2003) *Dev Cell*, **4,** 625-637.

99. Lin, S.Y., Johnson, S.M., Abraham, M., Vella, M.C., Pasquinelli, A., Gamberi, C., Gottlieb, E. and Slack, F.J. (2003) *Dev Cell*, **4,** 639-650.

100. Rajewsky, N. and Socci, N.D. (2004) *Dev Biol*, **267,** 529-535.

101. Stark, A., Brennecke, J., Russell, R.B. and Cohen, S.M. (2003) *PLoS Biol*, **1,** E60.

102. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) *Cell*, **115,** 787-798.

103. Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z. and Hatzigeorgiou, A. (2004) *Genes Dev*, **18,** 1165-1178.

104. Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C. and Marks, D.S. (2003) *Genome Biol*, **5,** R1.

105. Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, S., Tacker, M. and Schuster, P. (1984) *Monatshefte fur Chemie*, **125,** 167-188.

106. Kawasaki, H. and Taira, K. (2003) *Nature*, **423,** 838-842.

107. Kawasaki, H. and Taira, K. (2003) *Nature*, **426,** 100.