nature genetics

Approaches to microRNA discovery

Eugene Berezikov, Edwin Cuppen & Ronald H A Plasterk

MicroRNAs (miRNAs) are noncoding RNAs that can regulate gene expression. Several hundred genes encoding miRNAs have been experimentally identified in animals, and many more are predicted by computational methods. How can new miRNAs be discovered and distinguished from other types of small RNA? Here we summarize current methods for identifying and validating miRNAs and discuss criteria used to define an miRNA.

MicroRNAs are RNA molecules of approximately 22 nucleotides (nt), derived from genome-encoded stem-loop precursors, that recognize target mRNAs by base-pairing and thereby regulate their expression. Other reviews in this issue focus on details of the biosynthesis of these small RNAs, their mode of action and their function. Such details are also relevant to the subject of this review: namely, the discovery of miRNAs.

After the initial wave of miRNA identification[1–3], about 250 miRNAs were estimated to be encoded in the human genome[4], but it was subsequently recognized that this estimate could be low[5]. Later studies, based on combinations of computational and experimental techniques, support a substantially larger number of miRNAs[6–8]. It is still difficult to estimate the upper limit of miRNAs in humans and other mammals, in part because of the difficulty in defining what a 'true' miRNA is. Advances in technology and methodology may ultimately lead to the description of thousands of candidate miRNA genes, and the issue then will be which of these encode real, biologically functional miRNAs. Currently accepted standards for miRNA classification[9] are largely based on evidence obtained from initial studies on the small RNA repertoire of a eukaryotic cell.

Here, we first provide an overview of these initial studies and review recent advances in the methodology used to discover miRNAs. We then discuss the standards used for miRNA classification in connection with recent progress in our understanding of miRNA biology. As a focus for the review, we concentrate on work done in animal systems.

Classic definition of an miRNA

Previously, miRNAs were defined as noncoding RNAs that fulfill the following combination of expression and biogenesis criteria[9]. First, mature miRNA should be expressed as a distinct transcript of ~22 nt that is detectable by RNA (northern) blot analysis or other experimental means such as cloning from size-fractionated small RNA libraries. Second, mature miRNA should originate from a precursor with a characteristic secondary structure, such as a hairpin or fold-back, that does not contain large internal loops or bulges. Mature miRNA should occupy the stem part of the hairpin. Third, mature miRNA should be processed by Dicer, as determined by an increase in accumulation of the precursor in Dicer-deficient mutants.

In addition, an optional but commonly used criterion is that mature miRNA sequence and predicted hairpin structure should be conserved in different species. An 'ideal' miRNA would meet all the above criteria. In practice, variations are possible, but at the very minimum expression of a ~22-nt form and the presence of a hairpin precursor need to be demonstrated to classify a sequence as an miRNA.

All approaches to discovering miRNAs, as we review below, are based on these definitions of an miRNA gene and can be split into two groups. In experiment-driven methods, the expression of small RNAs is first established, and bioinformatics is then used to identify RNAs that meet structural requirements. In computation-driven approaches, candidate miRNA are first predicted in (whole) genome sequences on the basis of structural features, and experimental techniques are then used to validate these predictions by demonstrating expression of the corresponding sequences.

Identification of miRNAs by forward genetics

As is often found in biology, forward genetics methods were instrumental in identifying the first miRNA genes, *lin-4* and *let-7*. Analysis of a *Caenorhabditis elegans* mutant with a defective cell lineage indicated that a mutation in a small noncoding RNA, *lin-4*, was responsible for the phenotype[10]. The *lin-4* RNA showed several intriguing features. It had two forms: a large ~60-nt RNA, which could fold into a stem-loop structure (a hairpin); and a small ~22-nt RNA, which was part of the stem of the larger form and repressed expression of the *lin-14* gene by imperfect pairing with the 3′ UTR of *lin-14* mRNA[11].

Several years later, another RNA with similar characteristics, *let-7*, was identified in *C. elegans*[12]; this discovery jump-started the miRNA field because *let-7*, in contrast to *lin-4*, was found to be conserved among a wide range of phylogenetic taxa[13,14], indicating that miRNA-mediated gene regulation might be more ancient and common than was previously thought. Since then, forward genetics approaches have yielded only four additional miRNAs, *bantam*, *miR-14* and *miR-278* in *Drosophila melanogaster*[15–17] and *lsy-6* in *C. elegans*[18].

Several explanations can be given for the relative inefficiency of miRNA gene discovery by forward genetics methods. The small size of miRNAs and their potential tolerance to mutations that do not affect

Eugene Berezikov, Edwin Cuppen and Ronald H.A. Plasterk are in the Hubrecht Laboratory, Uppsalalaan 8, 3584CT Utrecht, The Netherlands.
e-mail: plasterk@niob.knaw.nl
Published online 30 May 2006; doi:10.1038/ng1794

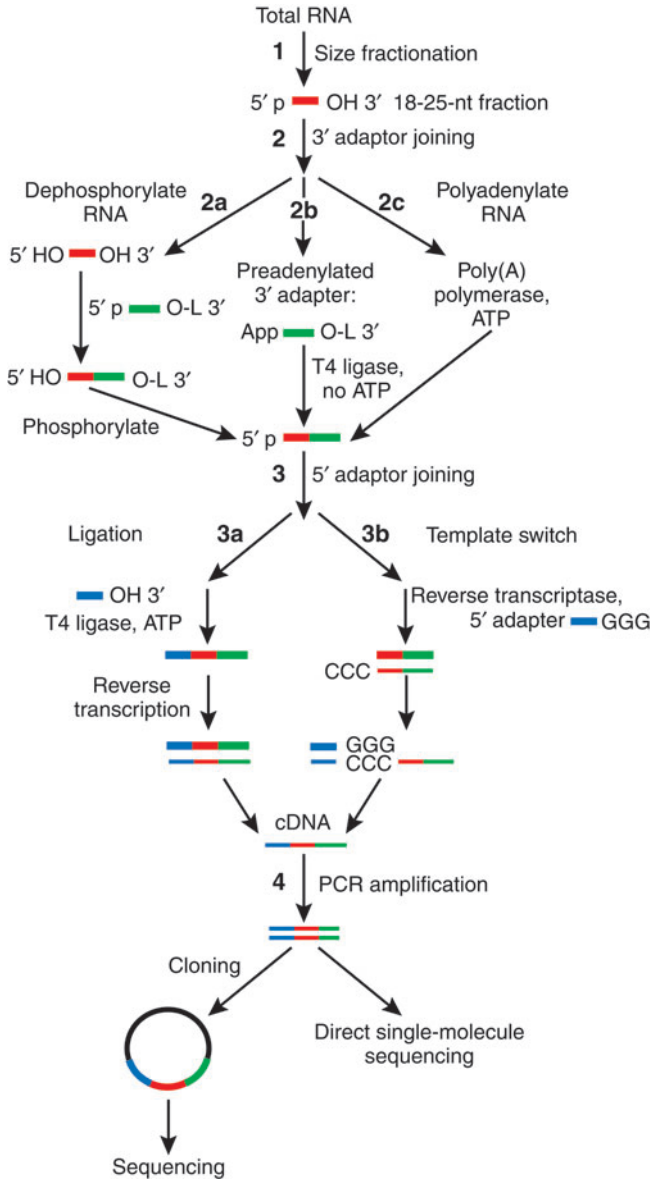

Figure 1 Methods for cloning miRNA. In **1**, total RNA is separated on a polyacrylamide gel and the fraction corresponding to RNAs of 18–25 nt is recovered. In **2**, a 3′ adapter can be introduced in different ways: the adapter can be ligated to a dephosphorylated RNA, which is then phosphorylated (**2a**); a preadenylated adapter can be ligated to RNA without free ATP in the reaction (**2b**); or the RNA can be polyadenylated by poly(A) polymerase (**2c**). In **3**, a 5′ adapter is introduced either by ligation (**3a**), or by template switching during reverse transcription (**3b**). In **4**, cDNA is amplified by PCR and cloned into a vector to create a library. Alternatively, PCR products can be sequenced directly by single-molecule sequencing methods (massive parallel sequencing).

the so-called 'seed sequence' make miRNAs genes 'difficult-to-hit' targets in spontaneous or induced mutagenesis. Even if an miRNA is hit and knocked out, researchers may still miss it because efforts to map a mutation are usually concentrated on protein-coding regions. Finally, many miRNA mutants may not be recognized in a phenotype-driven screen because of redundancy[19]. Although more miRNA genes might be identified in the future by means of forward genetics, this approach is unlikely to be a main contributor to the ever-growing list of biologically functional miRNAs.

Identification of miRNAs in small RNA libraries

The preferred approach to *de novo* identification of miRNAs is to sequence size-fractionated cDNA libraries. Initially, a protocol, originally devised for cloning ~22-nt small interfering RNA molecules[20], seemed to be also useful for identifying endogenous small RNAs, many of which turned out to be miRNAs[1]. Variations of protocols for cloning small RNAs have been developed independently[2,3], and these different protocols have been successfully applied to identification of most of the currently known miRNAs. All protocols follow the same principle but differ in their details[21] (**Fig. 1**). In brief, an RNA sample is separated in a denaturing polyacrylamide gel and the size fraction corresponding to 20–25 nt is recovered. Next, 5′ and 3′ adapters are attached to the RNAs, RT-PCR is carried out and the fragments are cloned into vectors to create a cDNA library. Individual clones are then sequenced and analyzed to determine the genomic origin of the small RNA.

For the first-strand cDNA synthesis, a 3′ adapter needs to be ligated to the mature miRNA to introduce a site at which to anneal the primer used by reverse transcriptase. To prevent self-circularization of the mature miRNAs and the adapter, small RNAs are usually dephosphorylated before ligation and the 3′-hydroxyl terminus of the 3′ adapter is blocked by incorporating a non-nucleotide group during chemical synthesis of the oligonucleotide[22]. In another popular variation of the protocol, the 3′ adapter is preadenylated, removing the need to dephosphorylate the small RNA[2,23]. Alternatively, ligation of the 3′ adapter can be replaced by addition of a poly(A) tail to the small RNAs using poly(A) polymerase[24], in which case oligo(dT) is used as a primer for reverse transcription.

Before the reverse transcription reaction, a 5′ adapter is ligated to the gel-purified and, if necessary, phosphorylated product of 3′ adapter ligation. Ligation of the 5′ adapter can be omitted in protocols using cDNA cloning by SMART technology (Clontech)[3], which relies on the property of some reverse transcriptases to add several nontemplated nucleotides (predominantly deoxycytidine) to the 3′ ends of synthesized cDNAs. These overhanging nucleotides can be subsequently used in switching the template from miRNA to the 5′ adapter.

Finally, PCR-amplified cDNAs are often concatamerized into large fragments before being cloned into vectors to increase the length of informative sequence obtained from each sequenced clone[1,2,22]. Recently, a serial analysis of gene expression (SAGE)-like variation of the concatamerization step has been developed[25] that increases the average number of small RNA tags per clone from 5 to 35, thereby boosting the throughput and cost-efficiency of sequencing small RNA libraries.

New approaches that increase sequencing depth may replace conventional sequencing[26]. In a first example, Lu *et al.*[27] applied massive parallel signature sequencing to elucidate of the small RNA component of the *Arabidopsis thaliana* transcriptome. This technology enables hundreds of thousands of short (17-nt) sequencing tags to be generated in one run. An alternative technology that is emerging can produce a similar number of longer (100–150-nt) sequence reads in a single analysis run[28]. Although no small RNA libraries sequenced by this technology have been published so far, preliminary results from different research groups, including ours, indicate that this sequencing method is likely to have a big impact on the discovery of miRNAs and other small RNAs.

A common limitation in the discovery of miRNAs by cloning is that it is difficult to find miRNAs that are expressed at a low level, at very specific stages or in rare cell types. In principle, this limitation can be overcome by deep sequencing of small RNA libraries from a broad range of samples. In addition, a more difficult problem to solve is that some miRNAs may be hard to clone owing to physical properties, including sequence composition, or to post-transcriptional modifications, such as editing or methylation[29–31].

Once a small RNA is cloned from a cDNA library, bioinformatics is required to identify its origin in the genome. It may seem a trivial task to determine the genomic location (or locations) of a 22-nt sequence and to check whether a hairpin precursor is encoded in the genomic region and whether the hairpin structure is conserved in other species. This analysis is complicated, however, by the fact that hairpin structures are common in eukaryotic genomes and are not a unique feature of miRNAs. Additional care should be taken to distinguish miRNAs from other types of endogenous small RNA[21,32] and from degradation products of mRNAs or structural RNAs. Unfortunately, there is no publicly available software for processing sequencing results from small RNA libraries at present, and research groups hunting for miRNAs by sequencing will need to implement in-house bioinformatics analysis programs.

Computational prediction of miRNA genes

Surveying genomic sequences to predict miRNAs became popular after initial cloning efforts generated sufficient information about miRNA properties to recognize a set of distinctive miRNA features[33,34]. Numerous approaches to miRNA prediction have been developed and can be categorized on the basis of the particular miRNA features that they use for prediction. First, all approaches use secondary structure information, because the presence of a fold-back structure is an essential characteristic of miRNA. Second, many rely on phylogenetic conservation of both sequence and structure to distinguish miRNA candidates from irrelevant genomic hairpins. Last, other methods assess the thermodynamic stability of hairpins and sequence and structure similarity to known miRNAs, or use information on genomic location relative to known miRNAs.

The first methods of miRNA prediction relied heavily on conservation criteria. MirScan software identified and ranked conserved hairpins on the basis of their similarity to experimentally confirmed miRNAs and predicted 35 new miRNA candidates in *C. elegans*[35] and 107 in human[4], many of which were experimentally confirmed. The predictions in *C. elegans* were subsequently refined by incorporating into the algorithm the conservation of a characteristic motif upstream of the hairpin structures[36]. Another conservation-based software, snarloop, has been used to predict 214 candidate miRNAs in *C. elegans*[37] and has provided a basis for estimates of between 140 and 300, or more, miRNA genes in the *C. elegans* genome. In *D. melanogaster*, 48 candidate miRNAs have been predicted by miRSeeker[38], which does not simply use conservation but recognizes conservation patterns specific to miRNAs (such as a more diverged loop sequence and a more conserved hairpin stem). A similar approach, based on the shapes of conservation patterns of known miRNAs, has been used to predict more than 800 new miRNA candidates that are conserved between human and rodents[6].

Conservation of potential target sequences rather than hairpins can be used as an alternative starting point in miRNA prediction. Xie *et al.*[7] analyzed conserved motifs that are overrepresented in the 3′ UTRs of genes and found that many of them correspond to complements of seed sequences of known miRNAs. The seed sequence is formed by seven or eight nucleotides of the mature miRNA, starting from the first or second nucleotide, and is most crucial for interaction between the miRNA and its target[39–45]. Using motifs that did not match to known miRNAs, Xie *et al.*[7] predicted 129 new miRNA candidates in human. Similar 'target-driven' approaches have been recently applied to the prediction of miRNAs in *A. thaliana*[46], flies and worms[47].

Thermodynamic stability of secondary structure is another characteristic that can be used to distinguish miRNAs from other hairpins. Bonnet *et al.*[48] demonstrated that miRNAs, in contrast to tRNAs and rRNAs, have free energies of folding that are significantly lower than those of shuffled sequences. RNAz software combines thermodynamic stability and conservation of secondary structure to predict noncoding RNAs[49,50], and has been successfully used to predict miRNAs in various organisms[51–53].

Recently, several alignment-type methods for identifying homologs of known miRNAs have been developed[54–56] that search for genomic sequences that can be 'aligned' with original miRNAs at both the sequence and structural level. Importantly, not only close but also distant homologs can be identified in this way[55].

Obviously, methods that rely on phylogenetic conservation of the structure and sequence of an miRNA cannot predict nonconserved genes. To overcome this problem, several groups have developed *ab initio* approaches to miRNA prediction[8,57,58] that use only intrinsic structural features of miRNAs and not external information. Each of these methods builds classifiers that can measure how a candidate miRNA is similar to known miRNAs on the basis of several features (for example, Sewer *et al.*[57] distinguish 40 features, such as free energy of folding, length of the perfect longest stem, average size of symmetrical loops, and proportion of different nucleotides in the stem, among others). Once a set of features is defined, a popular machine learning approach called 'support vector machines' is used to build a model, based on positive and negative training sets, that assigns weights to different features such that their contribution to an overall score results in the optimal separation of positives and negatives. With these *ab initio* prediction methods, many nonconserved miRNAs have been discovered and experimentally verified in viruses[59] and human[8].

Another productive way to discover miRNAs is to explore genomic sequences surrounding known miRNAs, because many miRNAs are clustered or located close to one other[2]. Numerous human and mouse miRNAs have been identified in this way[60,61] and the indications are that more will be found[57].

Experimental validation of candidate miRNAs

Computationally predicted candidate miRNAs need experimental validation. In principle, an miRNA can be considered to be validated when expression of its mature ~22-nt form is demonstrated. Validation approaches can be split in two categories: those that determine the exact ends of the mature RNA; and those that demonstrate expression but do not identify the exact ends (**Fig. 2**). It is important to realize that miRNA prediction algorithms often cannot predict the location of the mature miRNA in a precursor with nucleotide precision. However, establishing the mature miRNA ends, especially the 5′ end, is essential for downstream applications such as miRNA target prediction. For this reason, validation approaches based on the cloning and sequencing of small RNAs are the most informative.

A combination of random cloning from small RNA libraries and miRNA prediction is a viable, albeit not directed, approach in which predictions are not used at the experimental stage but they simplify analysis of the cloned sequences[21]. In a PCR-based directed cloning approach, one of the primers is universal and corresponds to the 5′ adapter, whereas the other overlaps with the 3′ region of the miRNA, facilitating amplification of specific cDNA clones from a small RNA library[4,35]. With this approach, only one of the miRNA ends (the 5′ end) can be determined. In another directed cloning method, biotinylated oligonucleotides corresponding to predicted miRNAs are used to enrich for specific cDNAs before library construction[8]. The advantage of this method is that the complete sequence of the mature miRNA can be deduced.

Different hybridization-based methods can be used to demonstrate the expression of predicted miRNAs. RNA (northern) blot analysis is a robust technique that can provide information on the size and expression of predicted miRNAs. It is also frequently used to confirm expression

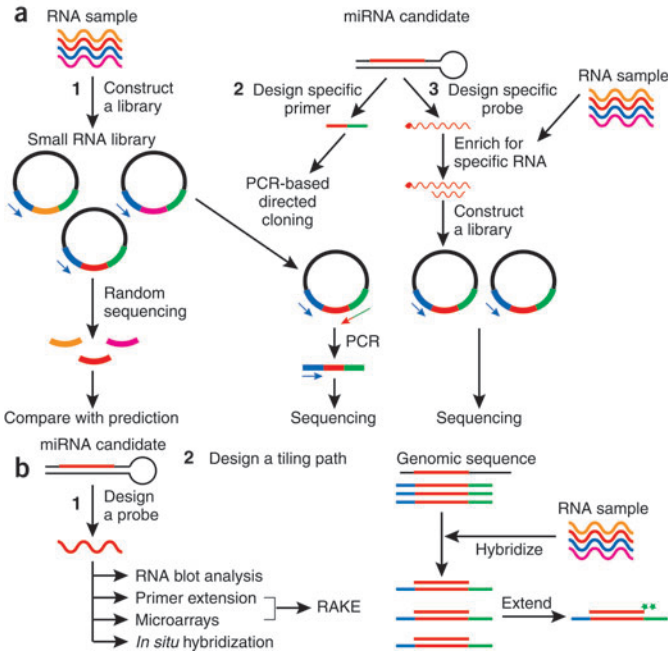

Figure 2 Validation of miRNA candidates. (**a**) Cloning-based approaches. In **1**, miRNA can be validated indirectly by random sequencing from small RNA libraries. In **2**, primers overlapping the predicted miRNA and the adapter can be designed to amplify specific candidate miRNAs from a library. In **3**, a biotinylated probe can be used to enrich the RNA sample before library construction. (**b**) Hybridization-based methods. In **1**, candidate-specific probes can be used in RNA blot analysis, primer extension, microarray analysis or *in situ* hybridization. In **2**, a tiling path of probes overlapping the predicted 3′ end of a mature miRNA can be designed and used in an RNA-primed, array-based Klenow enzyme (RAKE) assay to establish the exact 3′ end of the miRNA.

of miRNAs cloned from size-fractionated cDNA libraries[1–3]. The disadvantages of RNA hybridization are its low throughput and limited sensitivity for detecting rare miRNAs.

Another hybridization-based method for miRNA validation is primer extension. In this approach, a primer that is several nucleotides shorter than the predicted miRNA is hybridized to an RNA sample and extended by reverse transciptase using RNA as template; gel electrophoresis is then used to detect extended products[62]. Only the 5′ end of an miRNA can be identified in this way. An inversed version of this approach is used in the RNA-primed array-based Klenow extension (RAKE) assay, in which miRNAs are hybridized to probes on a microarray and are used by Klenow enzyme as primers in an extension reaction[63]. RAKE was originally developed for expression profiling of known miRNAs, but it can be adapted to map the 3′ ends of predicted miRNAs in a high-throughput fashion. We have designed RAKE tiling path probes with single-nucleotide resolution that cover the potential ends of a set of previously predicted miRNAs[6], and have used high-density 44K Agilent custom microarrays to confirm the expression of several hundred miRNA candidates (E.B., E.C. and R.H.A.P., unpublished data). Although conventional miRNA microarrays can be also used for high-throughput testing of candidate miRNA expression[21], RAKE provides the additional advantage of establishing the dominant 3′ end of an miRNA.

Finally, *in situ* hybridization methods for miRNA detection have been developed recently[64–66] and can be used to determine the spatio-temporal expression patterns of candidate miRNAs. *In situ* hybridization does not provide information on the size or the ends of the RNAs detected, and thus has limited value for validating predicted miRNAs.

What is an miRNA?

In the literature, miRNAs are sometimes described as ~22-nt RNA molecules that originate from fold-back precursors and can regulate the expression of genes. This biologically intuitive definition implies that miRNAs should have a demonstrated function; however, biological function has been so far elucidated for only a few miRNAs, and the criteria established for miRNA classification[9] deliberately do not include the requirement that a small RNA must have a demonstrated function to be annotated as an miRNA. Instead, phylogenetic conservation—an indirect indication of a possible function—is proposed as supporting evidence for annotation as an miRNA. Strictly speaking, and in keeping with the general guidelines for annotating noncoding RNAs[67], the term 'candidate miRNA' should be used as long as the function of the miRNA is unknown. This may, however, not always be practical, and once expression and biogenesis evidence is obtained for reliable annotation of a gene as an miRNA, the prefix 'candidate' can be dropped without a specific function assigned to the gene. For border-line cases—for example, when the only criteria satisfied are expression by sequencing of a single clone and the presence of a nonconserved hairpin—the 'candidate miRNA' terminology may be justified.

Function aside, the main objective of the original guidelines was to establish a uniform system for miRNA annotation and to prevent the misclassification of other types of small RNA as miRNA. The guidelines have successfully fulfilled their role so far, but how will recently generated data on small RNA sequencing and other approaches to miRNA identification affect them in the future? Recent work on extensive sequencing of small RNAs from human colorectal cells[25] has provided a first glimpse of some of the issues that will undoubtedly arise. In this study, Velculescu and coworkers[25] identified 200 known and 133 unknown miRNAs by sequencing more than 270,000 cDNA tags. All previously unknown miRNAs met the expression (=cloned) and biogenesis (=hairpin) criteria required for miRNA annotation, and 89 had additional supporting evidence such as conservation, multiple observations of expression, genomic clustering, cloning of the star sequences, or homology to known miRNAs. Yet the set of newly discovered miRNAs was fundamentally different from the set of known miRNAs.

First, the set was supported by only 2,000 tags as compared with 70,000 tags for known miRNAs, indicating that the newly identified miRNAs are expressed at substantially lower levels. Second, only six of the new miRNAs were differentially expressed in a cell line in which Dicer expression was knocked down, as compared with 55 of the 97 known miRNAs. Third, only 32 of the 133 new miRNAs are conserved. Last, our analysis (unpublished data) indicates that 25 of the new miRNAs overlap with repeat annotations (as provided in Ensembl v.36), including L1 elements, 2 overlap with tRNA annotations (*hsa-mir-565* and *hsa-mir-594*) and 1 overlaps with *Alu* repeats (*hsa-mir-566*). Does this mean that some RNAs identified in this study are erroneously annotated as miRNAs?

This question is difficult to answer because for every argument against miRNA annotation, a counterargument can be provided. Low expression, as judged by a few observed clones, would be expected, because abundant miRNAs are easily cloned and were identified long ago; however, in this study 20 known miRNAs were also observed only once[25]. Thus, the number of times that an RNA sequence is observed in the experiment is not always a good filter. Neither is conservation, because this is not a general feature of functional noncoding RNAs[68], and several nonconserved miRNAs have been identified[8,69]. In addition, overlap with repeat annotations does not immediately taint a candidate, because examples of repeat- and pseudogene-derived miRNAs are known[70,71]. No differential expression in a Dicer mutant is perhaps more worrying, but can be explained by the technical difficulties of detecting miRNAs

expressed at low levels. This leaves us with only two operational criteria: expression of a ~22-nt RNA, and the presence of a potential fold-back structure. Although these criteria are indeed not sufficient to declare an RNA as a *bona fide* miRNA[9], they do warrant its annotation as a candidate miRNA. From the practical side, these candidate miRNAs should still be deposited in miRBase[72] with appropriate remarks and under distinctive names, thereby providing a consistent basis for their further investigation.

What evidence can be considered to generate a definition of an miRNA that is precise, generally applicable, and based on our biological meaning of an miRNA? Given the known functions of miRNAs, one might require that a real miRNA regulates target mRNAs. We would not, however, want to exclude miRNAs that enhance, rather than repress, a target mRNA (perhaps by stabilizing it). We also would not want to exclude an miRNA that can be clearly shown to bind targets, but without affecting their expression. After all, recent reports show that external signals may function to release a mRNA from miRNA binding[73,74]; thus, a miRNA may affect the expression of a target only under specific *in vivo* conditions and/or in some tissues or cell lines, and regulatory effects might not be observed under the experimental conditions used.

The criterion that an miRNA needs to interact with a target is also probably not appropriate: at least in animal cells the current thought is that many, if not all, miRNAs that share their 7-bp seed sequence with an mRNA in the cell will bind to that mRNA[43–45]. Because other sequences within the 22 nt may also affect binding (there is selective pressure on the conservation of such residues), because the structure of the target mRNA may have a role, and because there may even be hindrance by other bound miRNAs, we probably need to move away from the clean discrete picture that is seen in the early miRNA literature: namely, one mRNA with a fixed set of miRNAs cleanly bound to its 3′ UTR. There may be many miRNAs that jump off and on, there may be partially occupied sites, and thus there may be different copies of the same mRNA in the same cell that are occupied by a partially different set of miRNAs.

Certainly, a useful criterion is that miRNAs are made by the Dicer pathway. Other small regulatory RNAs are very likely to exist, but if they are made by other nucleases, then they are not included in the definition of miRNAs. Indeed, this criterion is precise, it fits with the intuitive biological definition, and it is probably robust (although the Dicer pathway may turn out to be a member of a larger family of related small RNA pathways and then even this defining criterion may need to be reconsidered). Demonstrating that a small RNA is made in a Dicer pathway can be technically challenging, however, because Dicer homozygous mutants are unviable[75–77] and 'tricks' are required to circumvent this problem[25,78]. An additional general feature of miRNAs, which can be useful in their discrimination, is that they are 'handed over' by Dicer to the RNA-induced silencing complex or RISC protein complex. This criterion, however, is also technically challenging to assess.

Outlook

Technologies such as massively parallel sequencing will boost the discovery of many expressed small RNAs and will undoubtedly result in the identification of more candidate miRNAs. The currently accepted standards for classifying *bona fide* miRNAs will remain the basis of the miRNA definition, even though the limitations of some of the criteria are becoming obvious. It is, however, premature and perhaps impossible to propose an absolute set of standards that will apply to all systems, and considerable validation assays that establish functionality will be needed to advance our understanding of miRNAs and to provide a realistic estimate of the total number of miRNAs encoded by a human or mammalian genome.

ACKNOWLEDGMENTS

We thank R. Ketting, M. Tijsterman, W. Kloosterman and other colleagues for discussion and critical comments on the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

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

1. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. & Tuschl, T. Identification of novel genes coding for small expressed RNAs. *Science* **294**, 853–858 (2001).
2. Lau, N.C., Lim, L.P., Weinstein, E.G. & Bartel, D.P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**, 858–862 (2001).
3. Lee, R.C. & Ambros, V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**, 862–864 (2001).
4. Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B. & Bartel, D.P. Vertebrate microRNA genes. *Science* **299**, 1540 (2003).
5. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
6. Berezikov, E. *et al.* Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* **120**, 21–24 (2005).
7. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434**, 338–345 (2005).
8. Bentwich, I. *et al.* Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* **37**, 766–770 (2005).
9. Ambros, V. *et al.* A uniform system for microRNA annotation. *RNA* **9**, 277–279 (2003).
10. 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).
11. 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).
12. Reinhart, B.J. *et al.* The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**, 901–906 (2000).
13. Pasquinelli, A.E. *et al.* Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* **408**, 86–89 (2000).
14. Slack, F.J. *et al.* The lin-41 RBCC gene acts in the *C. elegans* heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* **5**, 659–669 (2000).
15. Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B. & Cohen, S.M. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. *Cell* **113**, 25–36 (2003).
16. Xu, P., Vernooy, S.Y., Guo, M. & Hay, B.A. The *Drosophila* microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr. Biol.* **13**, 790–795 (2003).
17. Teleman, A.A. & Cohen, S.M. *Drosophila* lacking microRNA miR-278 are defective in energy homeostasis. *Genes Dev.* **20**, 417–422 (2006).
18. Johnston, R.J. & Hobert, O. A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. *Nature* **426**, 845–849 (2003).
19. Abbott, A.L. *et al.* The let-7 MicroRNA family members mir-48, mir-84, and mir-241 function together to regulate developmental timing in *Caenorhabditis elegans*. *Dev. Cell* **9**, 403–414 (2005).
20. Elbashir, S.M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* **15**, 188–200 (2001).
21. Aravin, A. & Tuschl, T. Identification and characterization of small RNAs involved in RNA silencing. *FEBS Lett.* **579**, 5830–5840 (2005).
22. Pfeffer, S., Lagos-Quintana, M. & Tuschl, T. Cloning of small RNA molecules. In *Current Protocols in Molecular Biology* Vol. 4 (eds. Ausubel F. *et al.*) 26.4.1–26.4.18 (2003).
23. Pfeffer, S. *et al.* Identification of microRNAs of the herpesvirus family. *Nat. Methods* **2**, 269–276 (2005).
24. Fu, H. *et al.* Identification of human fetal liver miRNAs by a novel method. *FEBS Lett.* **579**, 3849–3854 (2005).
25. Cummins, J.M. *et al.* The colorectal microRNAome. *Proc. Natl. Acad. Sci. USA* **103**, 3687–3692 (2006).
26. Meyers, B.C., Souret, F.F., Lu, C. & Green, P.J. Sweating the small stuff: microRNA discovery in plants. *Curr. Opin. Biotechnol.* **17**, 139–146 (2006).
27. Lu, C. *et al.* Elucidation of the small RNA component of the transcriptome. *Science* **309**, 1567–1569 (2005).
28. Margulies, M. *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
29. Luciano, D.J., Mirsky, H., Vendetti, N.J. & Maas, S. RNA editing of a miRNA precursor. *RNA* **10**, 1174–1177 (2004).
30. Yang, W. *et al.* Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat. Struct. Mol. Biol.* **13**, 13–21 (2006).
31. Yang, Z., Ebright, Y.W., Yu, B. & Chen, X. HEN1 recognizes 21–24 nt small RNA duplexes and deposits a methyl group onto the 2′ OH of the 3′ terminal nucleotide. *Nucleic Acids Res.* **34**, 667–675 (2006).
32. Kim, V.N. & Nam, J.W. Genomics of microRNA. *Trends Genet.* **22**, 165–173 (2006).

33. Berezikov, E. & Plasterk, R.H.A. Camels and zebrafish, viruses and cancer: a microRNA update. *Hum. Mol. Genet.* **14**, R183–R190 (2005).
34. Bentwich, I. Prediction and validation of microRNAs and their targets. *FEBS Lett.* **579**, 5904–5910 (2005).
35. Lim, L.P. *et al.* The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* **17**, 991–1008 (2003).
36. Ohler, U., Yekta, S., Lim, L.P., Bartel, D.P. & Burge, C.B. Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA* **10**, 1309–1322 (2004).
37. Grad, Y. *et al.* Computational and experimental identification of *C. elegans* microRNAs. *Mol. Cell* **11**, 1253–1263 (2003).
38. Lai, E.C., Tomancak, P., Williams, R.W. & Rubin, G.M. Computational identification of *Drosophila* microRNA genes. *Genome Biol.* **4**, R42 (2003).
39. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. & Burge, C.B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
40. Doench, J.G. & Sharp, P.A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **18**, 504–511 (2004).
41. Kloosterman, W.P., Wienholds, E., Ketting, R.F. & Plasterk, R.H. Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Res.* **32**, 6284–6291 (2004).
42. 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).
43. Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. Principles of microRNA-target recognition. *PLoS Biol.* **3**, e85 (2005).
44. Farh, K.K. *et al.* The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* **310**, 1817–1821 (2005).
45. Stark, A., Brennecke, J., Bushati, N., Russell, R.B. & Cohen, S.M. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′ UTR evolution. *Cell* **123**, 1133–1146 (2005).
46. Adai, A. *et al.* Computational prediction of miRNAs in *Arabidopsis* thaliana. *Genome Res.* **15**, 78–91 (2005).
47. Chan, C.S., Elemento, O. & Tavazoie, S. Revealing posttranscriptional regulatory elements through network-level conservation. *PLoS Comput. Biol.* **1**, e69 (2005).
48. Bonnet, E., Wuyts, J., Rouze, P. & Van de Peer, Y. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* **20**, 2911–2917 (2004).
49. Washietl, S., Hofacker, I.L. & Stadler, P.F. Fast and reliable prediction of noncoding RNAs. *Proc. Natl. Acad. Sci. USA* **102**, 2454–2459 (2005).
50. Washietl, S., Hofacker, I.L., Lukasser, M., Huttenhofer, A. & Stadler, P.F. Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotechnol.* **23**, 1383–1390 (2005).
51. Missal, K., Rose, D. & Stadler, P.F. Non-coding RNAs in *Ciona intestinalis*. *Bioinformatics* **21** (Suppl.), ii77–ii78 (2005).
52. Hsu, P.W. *et al.* miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucleic Acids Res.* **34**, D135–D139 (2006).
53. Missal, K. *et al.* Prediction of structured non-coding RNAs in the genomes of the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *J. Exp. Zoolog. B Mol. Dev. Evol.* published online 19 January 2006 (10.1002/jez.b.21086).
54. Legendre, M., Lambert, A. & Gautheret, D. Profile-based detection of microRNA precursors in animal genomes. *Bioinformatics* **21**, 841–845 (2005).
55. Nam, J.W. *et al.* Human microRNA prediction through a probabilistic co-learning model of sequence and structure. *Nucleic Acids Res.* **33**, 3570–3581 (2005).
56. Wang, X. *et al.* MicroRNA identification based on sequence and structure alignment. *Bioinformatics* **21**, 3610–3614 (2005).
57. Sewer, A. *et al.* Identification of clustered microRNAs using an ab initio prediction method. *BMC Bioinformatics* **6**, 267 (2005).
58. Xue, C. *et al.* Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *BMC Bioinformatics* **6**, 310 (2005).
59. Pfeffer, S. *et al.* Identification of microRNAs of the herpesvirus family. *Nat. Methods* **2**, 269–276 (2005).
60. Seitz, H. *et al.* A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res.* **14**, 1741–1748 (2004).
61. Altuvia, Y. *et al.* Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res.* **33**, 2697–2706 (2005).
62. Seitz, H. *et al.* A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res.* **14**, 1741–1748 (2004).
63. Nelson, P.T. *et al.* Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat. Methods* **1**, 155–161 (2004).
64. Wienholds, E. *et al.* MicroRNA expression in zebrafish embryonic development. *Science* **309**, 310–311 (2005).
65. Kloosterman, W.P., Wienholds, E., de Bruijn, E., Kauppinen, S. & Plasterk, R.H.A. *In situ* detection of miRNAs in animal embryos using LNA-modified oligonucleotide probes. *Nat. Methods* **3**, 27–29 (2006).
66. Nelson, P.T. *et al.* RAKE and LNA-ISH reveal microRNA expression and localization in archival human brain. *RNA* **12**, 187–191 (2006).
67. Huttenhofer, A. & Vogel, J. Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res.* **34**, 635–646 (2006).
68. Pang, K.C., Frith, M.C. & Mattick, J.S. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet.* **22**, 1–5 (2006).
69. Chen, P.Y. *et al.* The developmental miRNA profiles of zebrafish as determined by small RNA cloning. *Genes Dev.* **19**, 1288–1293 (2005).
70. Smalheiser, N.R. & Torvik, V.I. Mammalian microRNAs derived from genomic repeats. *Trends Genet.* **21**, 322–326 (2005).
71. Devor, E.J. Primate microRNAs miR-220 and miR-492 lie within processed pseudogenes. *J. Hered.* **97**, 186–190 (2006).
72. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **34**, D140–D144 (2006).
73. Schratt, G.M. *et al.* A brain-specific microRNA regulates dendritic spine development. *Nature* **439**, 283–289 (2006).
74. Ashraf, S.I., McLoon, A.L., Sclarsic, S.M. & Kunes, S. Synaptic protein synthesis associated with memory is regulated by the RISC pathway in *Drosophila*. *Cell* **124**, 191–205 (2006).
75. Wienholds, E., Koudijs, M.J., van Eeden, F.J.M., Cuppen, E. & Plasterk, R.H.A. The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nat. Genet.* **35**, 217–218 (2003).
76. Bernstein, E. *et al.* Dicer is essential for mouse development. *Nat. Genet.* **35**, 215–217 (2003).
77. Fukagawa, T. *et al.* Dicer is essential for formation of the heterochromatin structure in vertebrate cells. *Nat. Cell Biol.* **6**, 784–791 (2004).
78. Giraldez, A.J. *et al.* MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**, 833–838 (2005).