FEBS Letters 579 (2005) 5830–5840 FEBS 29898

Minireview

Identification and characterization of small RNAs involved in RNA silencing

Alexei Aravin, Thomas Tuschl*

Laboratory of RNA Molecular Biology, The Rockefeller University, New York, NY 10021, USA

Received 25 July 2005; accepted 11 August 2005

Available online 18 August 2005

Edited by Shou-Wei Ding

Abstract Double-stranded RNA (dsRNA) is a potent trigger of sequence-specific gene silencing mechanisms known as RNA silencing or RNA interference. The recognition of the target sequences is mediated by ribonucleoprotein complexes that contain 21- to 28-nucleotide (nt) guide RNAs derived from processing of the trigger dsRNA. Here, we review the experimental and bioinformatic approaches that were used to identify and characterize these small RNAs isolated from cells and tissues. The identification and characterization of small RNAs and their expression patterns is important for elucidating gene regulatory networks.

Keywords: MicroRNA; miRNA; Small interfering RNA; siRNA

1. Introduction

1.1. Discovery of small RNAs guiding RNA silencing processes

The first gene silencing process guided by small RNAs was discovered in 1993 in *Caenorhabditis elegans* [1,2]. It was shown that a 21-nucleotide (nt) processing product of the non-coding RNA transcript of the *lin-4* gene mediated repression of *lin-14* mRNA. The repression was dependent on partial sequence complementarity between the *lin-4* 21-nt RNA and the 3′ untranslated region (UTR) of *lin-14* mRNA. The *lin-4* mutant was initially discovered in 1981 in a screen for genes controlling the invariant cell lineages of *C. elegans* [3]. Several years later, another *C. elegans* 21-nt small RNA derived from the non-coding RNA transcript of *let-7* was shown to regulate *lin-41* mRNA [4]. lin-4 and let-7 small RNAs are unrelated in sequence, yet their primary transcripts share a common structural feature: the lin-4 and let-7 21-nt sequence is embedded in the stem of a 30-bp double-stranded RNA (dsRNA) hairpin. A double-stranded precursor suggested that lin-4 and let-7 feed into general gene silencing mechanism triggered by dsRNA in *C. elegans*. Because let-7 is evolutionary conserved in all bilateral symmetrical animals it was further suggested that this general silencing mechanism was also conserved [5]. More recent genetic screens have identified additional small RNA coding genes, such as *lsy-6* in *C. elegans* [6] and *bantam* in *Drosophila melanogaster* [7], controlling neuronal asymmetry and apoptosis, respectively. While genetic approaches provide not only insight into the nature of the gene product but also its function, other experimental and computational approaches have led to the identification of now hundreds of such small RNA coding genes, referred to as microRNAs (miRNAs) [8–10].

While small RNAs were studied in *C. elegans*, studies in plants on the mechanism of co-suppression, transcriptional and post-transcriptional gene silencing (TGS and PTGS, respectively), and virus-induced gene silencing were tracking down "aberrant" RNA as the trigger of these processes [11,12]. Filamentous fungi also showed similar mechanisms known as quelling, repeat-induced point mutation and meiotic silencing [13]. Related processes contributing to transposon and transgene silencing were also described in *D. melanogaster* [14,15]. Similar to miRNA-guided gene silencing, these silencing processes are evolutionary conserved and homology-dependent, i.e., require extensive sequence similarity between the trigger sequence (e.g., the transgene) and the targeted genes or gene products.

The molecular nature of the trigger of homology-dependent gene silencing was only revealed upon the discovery of the process of RNAi in *C. elegans* in 1998 [16]. Injected long dsRNA (hundreds of bps) induced degradation of mRNAs cognate to the dsRNA trigger that sometimes lasted for several generations. This method was shown to be effective in other invertebrate animals as well as in mouse oocytes and early embryos [17]. However, long dsRNA applied to differentiated somatic mammalian cells not only triggers RNAi but also a strong sequence-unspecific response causing global translational arrest and apoptosis [18]. The first report linking the production of small RNAs to homology-induced gene silencing in plants appeared in 1999 [19]. Small 25-nt sense and antisense RNAs were only found in transgenic plants that showed PTGS, suggesting that these small RNAs might function as specificity determinants of silencing. The initially reported 25-nt RNAs were later shown to represent two distinctly sized small RNA populations of 21- and 24-nt [20]. Analysis of the biochemical mechanism of RNAi in *D. melanogaster* embryo lysate revealed that long dsRNA is processed to 21-nt small RNAs [21]. After processing, the 21-nt small RNAs are base-paired with a 2-nt 3′ overhang and carry a

*Corresponding author.
E-mail addresses: aravina@rockefeller.edu (A. Aravin), ttuschl@mail.rockefeller.edu (T. Tuschl).

0014-5793/$30.00
doi:10.1016/j.febslet.2005.08.009

5′ phosphate and free 3′ hydroxyl [22]. Because such 21-nt base-paired processing intermediates were effectively targeting mRNAs for degradation, they were named small interfering RNAs (siRNAs). In order to characterize siRNAs that are processed from long dsRNA, the first small RNA cloning protocol was developed [22]. Cloning of the siRNAs obtained by processing of exogenously added dsRNA to *D. melanogaster* embryo lysate identified at the same time endogenous small RNAs, including novel miRNAs, and small RNAs cognate to transposable elements [8,22]. These initial observations triggered new experiments aimed at elucidating cellular sources of dsRNA-derived small RNAs and their role in gene regulation. Small RNAs of different origin and function were subsequently identified in different organisms raging from fission yeast to human (Table 1).

1.2. Different types of small RNAs

RNA silencing is an evolutionary ancient regulatory mechanism. Although its core protein machinery is widely conserved in eukaryotes, excluding the yeast *Saccharomyces cerevisiae*, different sources of dsRNA and functionally distinct RNA silencing effector complexes are encountered in different organisms. Effector complexes are composed of a single-stranded small RNA [23] directly bound to a member of the Argonaute (Ago) protein family [24–26]. Ago proteins contain an endoribonuclease domain [27]. However, only a subset of Ago proteins is catalytically active and responsible for mediating RNA cleavage [28,29]; cleavage-incompetent Ago members likely mediate other sequence-specific regulatory processes, such as transcriptional and translational silencing. The core effector complex may also be associated with auxiliary proteins [30].

1.2.1. miRNAs. miRNAs are the most abundant type of small RNAs in plants and animals (for recent review see [31,32]). miRNAs are excised from primary transcripts (pri-miRNAs) by two rounds of endoribonuclease III processing involving first Drosha, producing a hairpin-shaped pre-miRNA, and then Dicer (reviewed in [33]). After Dicer processing, miRNAs emerge as siRNA-duplex-like structures, but only one strand, the mature miRNA, is then predominantly incorporated into the Ago effector complexes. The discarded RNA strand is frequently referred to as miRNA* and is degraded. In plants, miRNAs guide mRNA cleavage and the highly complementary target sites are readily identified using bioinformatic tools [34]. In contrast, animal miRNAs preferentially target mRNAs at partially complementary yet evolutionary conserved sites, which are predominantly located within the 3′ UTR [35–37].

Table 1
Different types of endogenous small RNA

Class of small RNA	Size of mature form (nt)	Structure of precursor	Biogenesis	Mechanism of action	Organism
miRNA	20–23	Imperfect hairpin	Successive cleavage by Drosha and Dicer resulting in a mature form with defined sequence	Translational repression,[a] mRNA cleavage[b]	*C. elegans* [6,9,10,53,56,97] *D. melanogaster* [7,8,40,93] *X. laevis* [101] *D. rerio* [61,80,96] Mammals [8,37,59,61,94,95,100,102–109] Plants [110–116] Viruses [47,117,118]
rasiRNA	23–28	Long dsRNA	Processing of long dsRNA by Dicer resulting in multiple short RNAs[d]	Regulation of chromatin structure, transcriptional silencing[c]	*S. pombe* [38] *T. brucei* [44] *C. elegans* [58] *D. melanogaster* [15,40] *D. rerio* [96] *A. thaliana* [20,41]
Endogenous siRNA	20–23	Long dsRNA	Processing of long dsRNA by Dicer.[e] Biogenesis requires RdRP activity	mRNA cleavage	*C. elegans* [53] *A. thaliana* [51]
tncRNA	19–23	Unknown	Unknown. Mature forms have defined sequence[f]	Unknown	*C. elegans* [53]

[a]Majority of animal miRNAs.
[b]Majority of plant miRNAs.
[c]Shown in fission yeast and plants; indirect evidences for *D. melanogaster* and *C. elegans*.
[d]rasiRNA in *S. pombe* and plants processed from long dsRNA by Dicer. Size of rasiRNA in *D. melanogaster* (24-27 nt) does not correspond to the size of standard Dicer products (21-23 nt).
[e]In plants miRNA-directed cleavage of transcript produce register for subsequent Dicer processing resulting in defined sequences of siRNAs spaced 21 nt apart from each other.
[f]Processing of tiny non-coding RNA depends on Dicer but no hairpin precursor can be identified.

1.2.2. rasiRNAs. The source of repeat-associated small interfering RNAs (rasiRNAs) is presumably dsRNA produced by annealing of sense and antisense transcripts that contain repeat sequences frequently related to transposable elements [38–40]. rasiRNAs are generally less abundant than miRNAs.

One of the most intriguing features of rasiRNAs is their distinct length distribution (Table 1). Their size varies between species and they are similar or a few nt longer than Dicer-processed siRNAs or miRNAs. In plants, multiple Dicer-like (Dcl) proteins have been identified; 21-nt miRNAs and siRNAs are generated by Dcl1, while 24-nt rasiRNAs are produced by Dcl3 [41]. In *D. melanogaster*, which encodes two Dicer paralogs, Dcr-1 matures 21-nt miRNAs and Dcr-2 produces 21-nt siRNAs [42,43]. The distinct size of 24–27 nt rasiRNAs and their mechanism of processing remains to be addressed. In the unicellular eukaryotes *Trypanosoma brucei* [44] and *Schizosaccharomyces pombe* [38], siRNAs and rasiRNAs were identified, but not miRNAs. rasiRNAs are involved in establishing and maintaining heterochromatin structure and in controlling transcripts emerging from repeat sequences [45,46]. In mammals, the existence of functional rasiRNAs is debated. Small RNA cloning studies, although they detect repeat-derived small RNAs at low abundance do not reveal a sharp size distribution characteristic for rasiRNAs in other species [47].

1.2.3. siRNAs. siRNAs are produced upon exogenous delivery of dsRNA or transgenic expression of long dsRNA, but endogenous sources for siRNAs are rare or non-existing. Although it can be envisioned that endogenous mRNA transcripts with segments of sequence complementarity may hybridize to each other or itself to yield dsRNA and subsequently siRNAs, there is no evidence of such mechanism. In animals, deamination of dsRNA, which destabilizes its structure, is a likely mechanism to balance unwanted dsRNA formation [48–50]. In plants, an interesting type of endogenous siRNAs was recently described and named *trans*-acting siRNAs (tasiRNAs) [51,52]. tasiRNAs were initially confounded with miRNAs because the same sequences were cloned multiple times, suggesting a defined register for tasiRNA excision. It was revealed later that specific miRNAs in plants target single-stranded non-coding RNA transcripts for cleavage, which then are used as template for RNA-dependent RNA polymerase (RdRP). Dicer processes the resulting dsRNA in 21-nt increments producing mature tasiRNAs, which then subject mRNAs that have one or a few tasiRNA complementary sites for degradation.

1.2.4. tncRNAs. Tiny non-coding RNAs (tncRNAs) are a specialty of *C. elegans* and are less abundant than miRNAs [53]. They are not evolutionary conserved, yet are matured to a defined sequence. Although the presence of tncRNAs is Dicer-dependent, double-stranded precursor molecules of tncRNAs were not yet identified experimentally or computationally. The functions of tncRNAs are unknown, and it remains unclear if tncRNAs guide RNA silencing or if their primary transcripts are involved in unrelated processes.

2. Methods for detection and identification of small RNAs

There are two possible experimental angles from which to approach small RNA characterization. Some researchers want to characterize expression of already known small RNAs, in particular miRNAs, while others wish to discover novel small RNAs. The first group of researchers would likely prefer nucleic acid hybridization-based technologies (Northern hybridization, RNase protection, primer extension, real-time RT/PCR, microarray hybridization or even more sophisticated technologies), while the second group wants to obtain direct sequence information by cloning and sequencing the population of small RNAs present in their sample. The characteristics and limitations of the different methods are summarized in Table 2. Finally, bioinformatic methods that rely on the characteristic miRNA precursor hairpin structure and its evolutionary sequence conservation were used to predict candidate miRNA genes, most of which remain to be validated.

The starting material for detection of small RNAs can either be total RNA or size fractionated total RNA, which is best isolated using standard Trizol (guanidinium isothiocyanate/acidic phenol) method. Size fractionation is needed for application of microarrays or for small RNA cloning, and is not required for Northern hybridization, RNase protection, primer extension or real-time RT/PCR. Size fractionation of nucleic acids is generally carried out using denaturing polyacrylamide gel electrophoresis, although some specialized methods have been developed to enrich for the fraction of small RNAs. Large RNAs, including the abundant rRNAs, can be removed by precipitation in the presence of polyethylene glycol (PEG-8000) and NaCl [19,54]. Alternative methods depend on the use of silica-based column materials (mirVana and PureLink miRNA Isolation Kits from Ambion and Invitrogen, respectively). It is important to keep in mind that RNA samples may contain different types of small RNAs that vary in size (Table 1). Furthermore, some species contain extremely abundant small RNA, such as 30 nt 2S ribosomal RNA in *D. melanogaster*, which one might want to remove by denaturing gel purification.

2.1. Hybridization-based detection methods

2.1.1. Northern hybridization. Polyacrylamide gel separation followed by Northern hybridization provides a simple and reliable method for small RNA detection. Northern hybridization provides information about the size of the detected molecules, which is important for discrimination of different classes of small RNAs [20,55]. Small RNAs that are excised from a longer precursor in a precise manner, like miRNAs or tncRNAs, are generally detected using a 5′ terminally labeled antisense oligodeoxynucleotide of the same length as the small RNA [1,8,53]. Small RNAs that originate from random processing of long dsRNA precursors, like siRNAs and rasiRNAs, are preferably detected using a RNA probe, which is prepared by partial hydrolysis of internally labeled long antisense RNA transcripts [19]. The amount of input total RNA needed for small RNA detection can be significant (5–50 μg); however, abundant miRNAs can be detected using as less as 5 μg of total RNA.

Northern hybridization is also a quantitative technique. In order to determine the absolute amount of a small RNA in a sample, a dilution series of a synthetic oligoribonucleotide of known concentration is blotted in parallel with the sample. The hybridization signal for the serial diluted standard samples is directly proportional to the applied amount and a standard curve can be obtained from which the concentration of small

Table 2
Comparison of small RNA characterization methods

	Small RNA cloning	Northern hybridization	RNase protection	Primer extension	Array hybridization
Identification of new small RNAs	+	–	–	–	–
Detection of endogenous siRNA/rasiRNA	+	+	+	–	–
Simultaneous detection of many miRNAs	+	–	–	–	+
Determination of small RNA size	+	+	Not precise	–	–
Identification of 5′ and 3′ ends of short RNA	+	–	Not precise	Only 5′ end	–
Discrimination between closely related miRNAs	Complete	Dependent on the position of the mismatch[a]	Dependent on the position of the mismatch[a]	Dependent on the position of the mismatch[a,b]	Dependent on the position of the mismatch[a]

[a]Centrally located sequence changes or insertions/deletions found between miRNA family members are readily discriminated in hybridization-based assays, but those that are located towards the termini of the molecule cannot be discriminated.
[b]Mismatches near the 3′ end of the RT primer will drastically affect extension of the primer.

RNA in the sample can be calculated. This method has been used to determine the number of miRNA molecules per cell, reaching up to 50 000 for abundant miRNAs [56].

2.1.2. RNase protection assay (RPA). Rather than using blotting, small RNAs can also be detected using solution-based hybridization of internally radiolabeled antisense transcripts. The excess of unpaired radiolabeled probe and of unpaired single-stranded segments of the probe/small RNA duplex are subsequently digested by RNase treatment. The protected segment of the probe is resolved on a polyacrylamide gel and quantified. To detect miRNAs or siRNAs after therapeutic dosing, radioactive probes are used that are just a few nucleotides longer than the siRNA/miRNA [57]. To detect siRNAs that are produced from long dsRNAs, transcripts that cover a few hundred nucleotides have been used [19,58]. Compared to Northern hybridization, the size estimates of small RNAs by RPA are not as precise because the RNase treatment yields a size distribution of protected cleavage products. The sensitivity of RPA is generally comparable to Northern blotting, although larger amounts of input total RNA can be used.

2.1.3. Primer extension. Primer extension is another solution-based RNA detection method, but it is only applicable to characterize defined small RNA sequences such as miRNAs or siRNAs derived from short hairpins. The small RNA is detected by hybridizing a radiolabeled DNA primer to the 3′ portion of the small RNA, followed by template-directed incorporation of nucleotides by reverse transcriptase. The primer should be a few nucleotides shorter than its small RNA target. The primer extension product is then separated from the non-extended primer using polyacrylamide gel electrophoreses. This method has been applied to detect predicted miRNAs [59].

2.1.4. Reverse transcription/PCR (RT-PCR). Real-time RT-PCR has been used to quantify individual miRNA precursors [60]. This method has also been adapted for mature miRNA quantification in a commercial setting (Applied Biosystems TaqMan MicroRNA Assays) using a looped primer design to add a universal primer sequence to the miRNA during cDNA synthesis. This universal primer sequence is then used for PCR amplification. The amount of starting material can be in the nanogram range of total RNA.

PCR-based methods have also been used to validate predicted miRNAs [61]. For this purpose, a 3′ adapter sequence has to be ligated to the isolated small RNAs. After cDNA synthesis using a RT primer complementary to the 3′ adapter sequence, the specific miRNA is PCR-amplified using the RT primer and a 5′ PCR primer that matches about 15 nt at the miRNA 5′ end. The PCR product is then cloned and sequenced. The sequence of the central segment of the PCR product (6–8 nt) that is not covered by the primer pair is examined to confirm its identity to the predicted miRNA.

2.1.5. miRNA microarrays. Another conventional though more technology-demanding approach to detect multiple or all known miRNAs is the use of microarray hybridization. In its first application, nylon membrane (dot-blot) arrays spotted with oligodeoxynucleotides antisense to mature miRNAs were used for hybridization with 5′ radiolabeled short RNAs [62]. The spot signal intensity on the membrane was analyzed using autoradiography or phosphorimaging.

Most (commercial) solutions for array analysis, however, use fluorescence for detection. The small RNA sample is either directly conjugated to biotin or a chromophore (direct labeling), or the label is introduced during cDNA synthesis or PCR-amplification of the cDNA (indirect labeling). If biotin label is introduced, chromophore-conjugated streptavidin is subsequently used for detection.

Direct labeling of small RNAs can be accomplished chemically or enzymatically. A fluorescent *cis*-platinum compound, which preferably couples to the N7 of guanosine, has been reacted with small RNAs followed by array hybridization [63]. This labeling method is also known as universal linkage system (ULS) [64] and is commercially available under the trademark Micromax (Perkin–Elmer). Human miRNAs contain between 1 and 10 guanosines, indicating that labeling efficiencies of different miRNAs will vary. The 2′,3′-diol functionality of the RNA 3′ end has also been used to introduce a label [65]. This method requires periodate oxidation of the diol to the dialdehyde, followed conversion with a labeled hydrazine derivative. Enzymatically, small RNA has been labeled by T4 RNA ligase (Rnl1) using a fluorescently modified dinucleotide and ATP [54]. However, it should be noted that Rnl1 rapidly circularizes small RNAs in the presence of ATP. This circularization reaction effectively competes with the 3′ ligation reaction of the labeled dinucleotide. The problem could be partly remedied by using pre-adenylated fluorescent or biotin compounds that are substrates for Rnl1 in the absence of ATP [66]. An alterna-

tive 3′ end labeling method relies on *Escherichia coli* poly(A) polymerase and the incorporation of amino-modified nucleotides, followed by coupling of the amino-modified small RNAs with the fluorophore (mirVana Labeling Kit, Ambion).

In contrast to above methods that label the small RNAs prior to array hybridization, an alternative enzymatic approach has been reported [67]. miRNAs were first hybridized to 5′ immobilized oligodeoxynucleotide probes, which contained a spacer/oligo-dT sequence 5′ to the miRNA complement. After hybridization, biotinylated ATP was incorporated opposite the oligo-dT sequence segment by a Klenow fragment. The biotinylated hybridization products were then visualized after incubation of the array with labeled streptavidin.

Indirect labeling of miRNAs during cDNA synthesis using biotinylated random octamer primers has been described [68,69]. In another application, unlabeled cDNA was generated first using unmodified random hexamers [70]. After alkaline hydrolysis of the template RNA, the single-stranded cDNA was 3′ labeled using terminal deoxynucleotidyl transferase. Other methods introduce the label during PCR-amplification of the cDNA. To allow for PCR amplification, 5′ and 3′ adapter oligonucleotides need to be joined to the small RNA sequence pool first (see below). In one case, a T7 RNA polymerase promoter was introduced by PCR and labeled single-stranded RNA was generated by T7 RNA polymerase transcription from the PCR template using modified nucleoside triphosphates [71]. In other cases, the fluorophore was introduced by PCR using a labeled primer [72–74]. While one group took great care to remove the unlabeled complementary strand of the PCR product prior to array hybridization [73,74], the other group did not [72].

In summary, direct labeling methods, if effective, are simple and time-saving and should not be limited by the amount of miRNA available, given their large copy number, 100 to 100 000 molecules per cell. Currently, it is difficult to evaluate the quality of the array data generated by different groups, because the use of a universal reference miRNA standard has not yet become a standard practice. Introduction of a miRNA standard for microarray experiments would allow different laboratories with different microarray platforms to compare and share data. So far, only two groups have used synthetic DNA oligonucleotides reference standards covering the miRNA sets represented on their arrays [54,73,74]. These reference standards, unfortunately, are distinct in design. Therefore, it is not yet possible to discuss the strength or weaknesses of the design of probes spotted on the array as well as the detection technology.

2.1.6. In situ hybridization. In situ hybridization is widely used to detect expression of mRNAs in various tissues, however, because of the size of the small RNAs, development of reliable protocols has been challenging. Such protocols need to take care of fixing small RNAs, which diffuse faster than longer RNAs and could be lost during hybridization or washing steps. At the same time, too much fixation or crosslinking would interfere with probe hybridization.

The first report claiming in situ detection of mature miR172 of *Arabidopsis thaliana* [75] used a probe, which was generated by incorporation of digoxigenin-modified ribonucleotides by in vitro transcription from a plasmid template that contained four repeats of the complement of the mature miRNA sequence. It is unclear if the detected signal was due to hybridization with mature or pri- and pre-miRNA, because the sense transcript that was generated as a control was not fully complementary to the pre-miRNA. *D. melanogaster* miR-10 expression was detected in the embryo aiming at the detection of pri- or pre-miRNA with DIG-labeled in vitro transcribed probes targeting the entire pre-miRNA or longer transcripts including the pre-miRNA [76].

To analyze miRNA expression in animals, modified synthetic oligonucleotides with increased duplex stability were applied. These oligonucleotides contain locked-nucleic-acid residues interspersed between 2′-deoxynucleotides [77] and miRNA probes are commercially available (Exiqon, Denmark). These reagents were used to examine the expression patterns of more than hundred miRNAs during zebrafish development [78]. A comparison of in situ signals between Dicer knockout and wild-type animals shows that the signals are predominantly derived from the recognition of the mature miRNAs. Though it appears from these experiments that small RNA fixation problems are overcome in zebrafish, adjustment of protocols may be required for detection of miRNAs in other samples, such as sliced tissues.

2.1.7. miRNA reporter transgenes. An indirect approach to determine miRNA expression pattern has first been developed in *D. melanogaster* [7] and was later also applied in mouse [79]. A reporter transgene (GFP or lacZ) under the control of a ubiquitous promoter carrying a 3′ UTR with a miRNA complementary sequence is introduced and its expression pattern is compared to a transgene with a mutant or deleted miRNA complementary site. Co-expression of this "sensor" mRNA with the miRNA leads to cleavage of the transgene transcript and loss of reporter expression. Rather than using transgenic expression, in zebrafish, injection of in vitro transcribed sensor mRNAs can also been used [80].

Transgene sensor expression was also used to validate miRNA target interactions in vivo [81–83]. The sensor encodes a reporter (e.g., GFP) gene fused with the 3′ UTR of the predicted miRNA target mRNA, while the miRNA precursor is expressed from the 3′ UTR of another reporter transgene (e.g., red fluorescent protein). The miRNA carrying transgene transcripts are only partially processed so that some transcripts release the miRNA, while others yield reporter protein. The miRNA carrying transgene and a transgenic sensor for its function are co-expressed in spatially distinct though partially overlapping patterns. If the miRNA recognizes the target sequence, the sensor is repressed in the overlapping regions of reporter expression.

2.2. Cloning of small RNAs

2.2.1. Preparation and sequencing of small RNA cDNA libraries. The majority of currently known miRNAs was determined by cloning and it remains the most powerful approach to discover new small RNAs. Small RNAs that are generated by RNase III processing have 5′ phosphate and 3′ hydroxyl termini, in contrast to most RNase hydrolysis products that have 5′ hydroxyl and 2′,3′ cyclic phosphate or 2′ or 3′ monophosphate termini. Although it may appear obvious that these chemical differences should be used to discriminate between silencing related small RNAs and RNA turnover products, it is important to realize that other nucleic acid modifying enzymes (kinases, phosphatases, methylases) can act on these products before incorporation into silencing complexes or

before degradation. In plants, the methyltransferase HEN1 modifies the 3′ terminus of miRNAs [84]. In animals, duplexes of siRNAs that carry 5′ hydroxyl termini are rapidly 5′ phosphorylated [23,85] and in human cell extracts, duplexes with 3′ phosphates are rapidly dephosphorylated [23]. The fate of single-stranded small RNAs is less clear because they are generally degrade rapidly unless they are associated with stabilizing proteins, such as the Ago proteins. Cell-type and organism-dependent variations of RNA silencing/turnover are anticipated, and may have an impact on the small RNA cloning result, depending on the experimental strategy.

Small RNA cloning is a multi-step process, for which different protocols have been developed independently, details of which are found in the literature [86–88], http://web.wi.mit.edu/bartel/pub/protocols_reagents.htm, and at http://banjo.dartmouth.edu/lab/MicroRNAs/mir.html. The preferred approach requires the presence of 5′ phosphate and free 3′ hydroxyl on the small RNAs for sequential ligation of 3′ and 5′ adapter oligonucleotides [10,47,87]. After reverse transcription, the cDNA is PCR-amplified using primers cognate to the ligated adapters. The PCR products, or concatamers thereof, are cloned and sequenced. The amount of starting material can be as low as 10 μg of total RNA though 200 μg is typically recommended. The following sections will discuss in more detail the technical challenges faced during small RNA cloning and how they have been addressed.

T4 RNA ligase 1, Rnl1, and a truncated form of T4 RNA ligase 2, Rnl2(1-249), have been used for adapters ligation. RNA ligases catalyze the joining of a 5′-phosphate (p) modified donor oligoribonucleotide (pDonor) to a 3′ hydroxyl containing acceptor oligoribonucleotide (Acceptor). Ligation is a three-step enzymatic process that requires ATP [89]. First, the ligase reacts with ATP to form a covalent enzyme-AMP intermediate and releases pyrophosphate. The adenylate (Ap) residue is then transferred to pDonor yielding the stable intermediate AppDonor. Finally, the ligase catalyzes the joining of Acceptor to AppDonor releasing AMP and forming Acceptor-pDonor. Oligoribonucleotides that contain both, a 5′ phosphate and 3′ hydroxyl will be rapidly circularized if no preventive measures are taken.

In order to effectively join the 3′ adapter to small RNA molecules, several countermeasures have been taken. (1) The 3′ adapter oligonucleotide was modified to prevent circularization. Typically, the 3′ hydroxyl was blocked by chemical synthesis of an oligonucleotide containing a 3′ non-nucleotidic moiety. An excess of adapter over small RNAs was used. (2) The small RNAs were dephosphorylated to avoid circularization (Fig. 1, step 1). This, however, posed the problem that small turnover products of large RNAs (without 5′ phosphate) and RNase III processed small RNAs (with 5′ phosphate) become chemically identical [87]. (3) Pre-adenylated 3′ adapter oligonucleotides were joined to 5′ phosphate containing small RNAs in the absence of ATP in order to eliminate the need for dephosphorylation of small RNAs [10,47,87] (Fig. 1, step 4). However, there are additional concerns. RNA ligases are partially adenylated after purification and therefore promote circularization independent of addition of ATP. Furthermore, since all enzymatic reactions are reversible, Rnl1 (the standard commercially available T4 RNA ligase) deadenylates the pre-adenylated adapter. If adenylate is transferred to the 5′ phosphorylated small RNAs, it is rapidly circularized. Therefore, reaction conditions that use a large amount of enzyme and an excess of adenylated adapter over enzyme and small RNA favor this side reaction. More recently, Rnl2(1-249) has been used to join pre-adenylated adapters to 5′ phosphorylated small RNAs to avoid small RNA circularization due to adenylate transfer [47,90]. Rnl2(1-249) is compromised for circularization of adenylated oligoribonucleotides. The use of pre-adenylated adapters can boost the representation of miRNAs in samples that contain partly degraded rRNAs and tRNAs [47]. (4) If pre-adenylated 3′ adapters are used, all nucleotides can be 2′-deoxynucleotides using Rnl1 or Rnl2(1-249). If non-adenylated 3′ adapters are used, the first few nucleotides need to be 2′ ribonucleotides to allow for enzymatic adenylation.

The 5′ adapter was ligated after gel-purification of the 3′ adapter ligation product using Rnl1 in the presence of ATP (Fig. 1, step 5) [10,47,87]. If small RNAs were dephosphorylated prior to 3′ adapter ligation, the 3′-adapter ligation product must be 5′ phosphorylated prior to 5′ adapter ligation (Fig. 1, step 3). The 5′ adapter oligonucleotide, or at least several nucleotides of its 3′ end need to be composed of 2′ ribose residues. The efficiency of joining the 5′ adapter is not only critically dependent on the adapter sequence itself but also on the sequence of the selected 3′ adapter. Different investigators selected different adapter pairs, which are not necessarily interchangeable.

The gel-purified second adapter ligation product was reverse transcribed priming with an oligodeoxynucleotide complementary to the 3′ adapter. One protocol for small RNA cloning did not use a 5′ adapter ligation step [9]. Instead, the 3′ adapter ligation product was reverse transcribed in the presence of a switch template (Fig. 1, step 7; SMART technology, Invitrogen), which introduced a constant sequence to the cDNA similar to a 5′ adapter. Because template switching is associated with the untemplated addition of nucleotides, the precise 5′ terminus of the small RNAs cannot be determined with certainty.

The cDNA library obtained by the various approaches was amplified by PCR using primers matching the 5′ and 3′ adapter sequences. Although the PCR product could be cloned and sequenced, it is not economic because of the small size of the PCR-generated insert. Concatamerization approaches have therefore been developed to produce longer inserts. Concatamerization requires directional ligation of non-palindromic restriction enzyme digested PCR products. These restriction sites were either placed into the adapter sequences [10] or were introduced by PCR after cDNA synthesis [87]. The concatamerized PCR products of 500–800 bp were ligated into a plasmid suitable for sequencing. A good overview of miRNA expression in a given sample is obtained by sequencing about 200 clones of concatamers.

Not all experimentalists followed the library construction with sequencing; instead some used the library as template for the synthesis of labeled probes to hybridize on miRNA microarrays (see above). Alternatively to array hybridization a labeled bead-based sorting technology has also been applied to examine the library composition for known miRNAs [91]. In order to test for the expression of a limited set of computationally predicted miRNAs, small RNA cDNA libraries also serve well as template for PCR-based miRNA detection. One primer of the primer pair was specific to segments of the predicted miRNA, while the other primer was matching the adapter sequence [37,61]. The PCR products were then cloned and

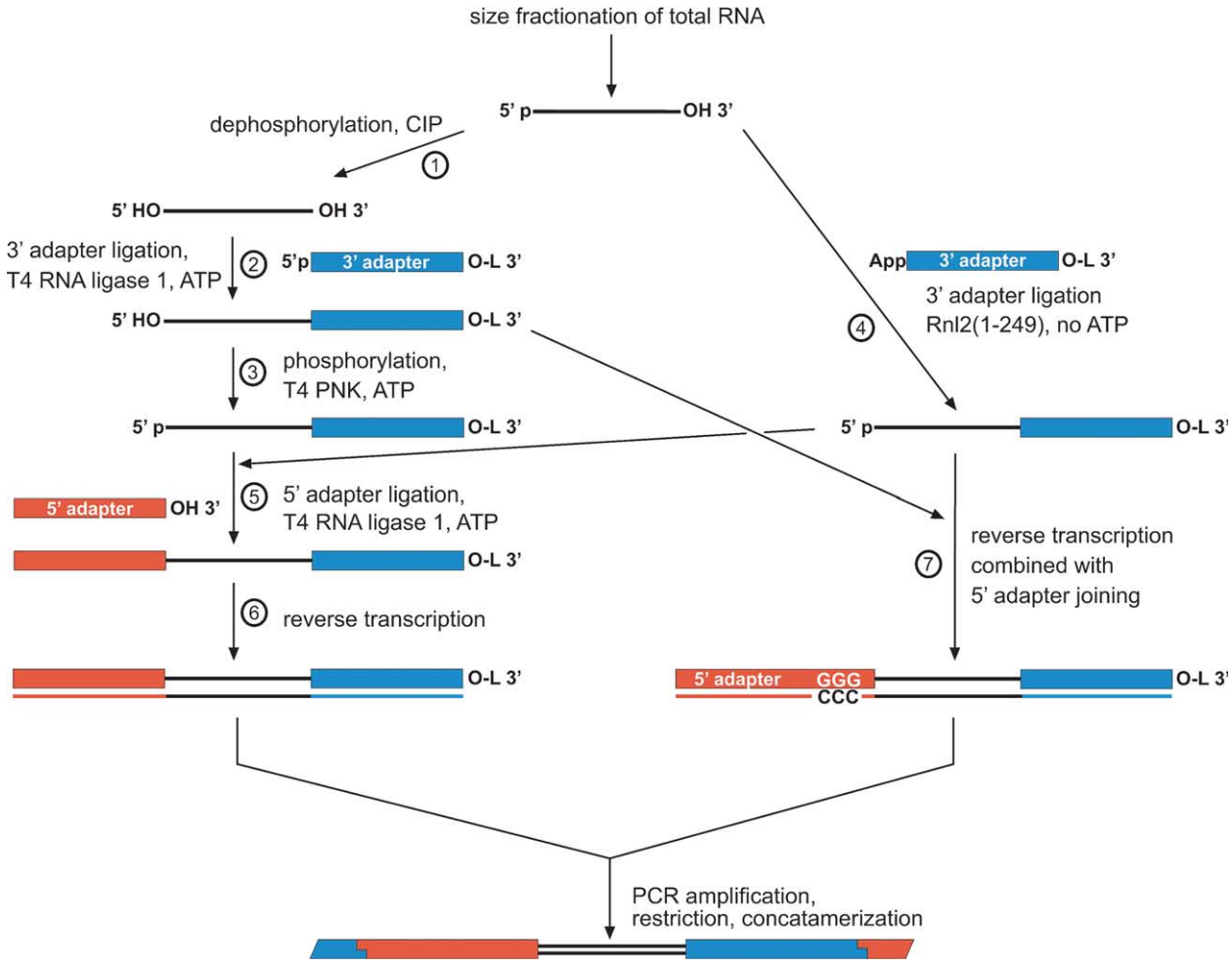

Fig. 1. Small RNA cloning procedure. Outline of the small RNA cloning procedure. RNA is dephosphorylated (step 1) for joining the 3′ adapter by T4 RNA ligase 1 in the presence of ATP (step 2). The use of a chemically adenylated adapter and truncated form of T4 RNA ligase 2 (Rnl2) allows eliminating the dephosphorylation step (step 4). If the RNA was dephosphorylated, it is re-phosphorylated (step 3) prior to 5′ adapter ligation with T4 RNA ligase 1 and ATP (step 5). After 5′ adapter ligation, a standard reverse transcription is performed (step 6). Alternatively, after 3′ adapter ligation, the RNA is used directly for reverse transcription simultaneously with 5′ adaptor joining (step 7). In this case, the property of reverse transcriptase to add non-templated cytidine residues at the 5′ end of synthesized DNA is used to facilitate template switch of the reverse transcriptase to the 3′ guanosine residues of the 5′ adapter (SMART technology, Invitrogen). Abbreviations: P and OH indicate phosphate and hydroxyl ends of the RNA; App indicates 5′ chemically adenylated adapter; L, 3′ blocking group; CIP, calf alkaline phosphatase and PNK, polynucleotide kinase.

sequenced to confirm the identity of the predicted miRNA by inspection of the non-primed region.

2.2.2. Analysis and interpretation of small RNA sequence data. In order to assign a putative function to cloned small RNAs, their sequences need first to be mapped to the genome. On average, about half of the clones represent RNA turnover products rather than siRNA or miRNAs. Depending on the degree of genome annotation, silencing-related small RNAs are readily distinguished from RNA turnover products of the abundant rRNAs, tRNAs, snRNAs, snoRNAs, or mRNAs fragments.

miRNAs meet the following features [31,47]: (1) They are typically cloned multiple times and their clone length distribution is narrow and peaks between 21 and 23 nt. (2) While the 5′ end of cloned RNAs is mostly invariant, and frequently uridine, their 3′ end is variable. At low frequency, the 3′ end is also post-transcriptionally modified by addition of predominantly adenosine or uridine. It is useful to note that post-transcriptional modifications of miRNAs, such as A-to-I editing [47,92], or disease-relevant mutations in miRNAs may only be identified using a cloning approach. (3) The genomic sequence flanking a miRNA contains a highly complementary, 20- to 30-nt segment, which is required to form the characteristic pre-miRNA hairpin structure. (4) pre-miRNA sequences are generally conserved in closely related species, though a small number of miRNAs are universally conserved in all animals. The loop sequence and sequences flanking the stem are much more variable than the miRNA and its complementary miRNA* sequence, so that the conservation profile resemble a saddle-like structure [93–95].

The cloning frequency of individual miRNAs is generally proportional to their relative concentration in the sample, given that a large enough number of small RNA clones ($\geqslant 1000$) were sequenced. To obtain absolute numbers of miRNA molecules in a sample, quantitative Northern blotting can be performed on one (or more) of the cloned miRNAs to convert relative miRNA cloning frequencies into absolute numbers of molecules.

Individual siRNAs or rasiRNAs are generally identified with a cloning frequency of one, likely because long dsRNA precursors can be variable in size and because the register of long dsRNA processing is not stable. The Dicer products from long dsRNA therefore appear in a more or less random positional distribution. Depending on the species, rasiRNAs show a narrow size distribution distinct from miRNAs. Similar to miRNAs, rasiRNAs may demonstrate 5′ terminal sequence bias towards uridine [40,96]. Genomic regions that yield long dsRNA transcripts can be identified if a significant number of small RNA clones cluster together. If these clustered sequences were found in sense and antisense orientation, it would suggest that the precursor dsRNA formed by annealing of sense and antisense transcripts. If the clustered sequences appeared only in a single orientation, it would either be indicative of internal fold-back structures due to long inverted repeat sequences or of Dicer-independent turnover of an abundant transcript.

2.3. Bioinformatic prediction of miRNA genes

The characteristic features of miRNAs, described above, can be used to predict new miRNA genes. The majority of analysis is strongly dependent on evolutionary conservation of miRNA genes [37,56,61,93,95,97–99]. In one approach, conserved candidate miRNAs were predicted after identification of conserved sequence motifs within 3′ UTRs of mRNAs [37]. These motifs were about 7-nt in length, many of which complementary to the 5′ end of known miRNA sequences as well as candidate sequences with fold-back structure.

Depending on the requirements for defined pre-miRNA secondary structure and the primary sequence conservation, different numbers of candidate miRNA genes were obtained, few of which are then actually experimentally confirmed. There are multiple explanations for not being able to experimentally confirm a predicted miRNAs. (1) They represent false positive candidates. (2) Experimental evidence was lacking because the expression pattern of the predicted gene was unknown and the gene was not expressed in the examined RNA sample. Other uncertainties associated with miRNA predictions include: (1) the orientation of the transcript (plus or minus strand) for a genomic location encoding hairpin sequence, (2) the position of the processing sites within the hairpin structure, and (3) the determination of which of the paired segments of the hairpin will constitute the mature miRNA.

Cloning of small RNAs from virus-infected cells demonstrated that viral miRNAs are rarely evolutionary conserved. This triggered the development of a prediction algorithm that did not utilize miRNA primary sequence conservation [47]. The prediction of miRNA relied on a supervised learning algorithm that was trained on known miRNA and pre-miRNA sequences (positives) as well as known non-coding RNAs unrelated to miRNAs (negatives). This approach is best applied to search for miRNAs in small genomes or to identify new miRNA genes in the closer vicinity of already known miRNAs. The application of analysis to large entire genomes, however, would produce so many candidates that would be difficult to be tested experimentally. The examination of the vicinity of known miRNAs, using structural and/or evolutionary conservation filters was useful to revealed new miRNAs [94,100].

The main difficulty associated with computational predictions is the absence of information about miRNA expression. Given the many stages of development and large number of cell and tissue types and environmental conditions, it is almost impossible to conclusively proof if a prediction is a false positive or if miRNA expression is exquisitely specific and highly regulated. Some attempts have been made to examine a large number of candidates leading to the identification of a large cluster of miRNAs specifically expressed in human placenta [100]. Considering the time and effort that is needed to follow up miRNA predictions, direct cloning and sequencing remains the most effective miRNA discovery route, though databases filled with conserved putative miRNA precursors significantly facilitate the identification of new miRNAs within the sets of cloned sequences.

3. Outlook

Currently, about 500 conserved miRNA genes have been identified in mammals. However, because many miRNAs underwent gene duplication only about 300 sequence-distinct mature miRNAs are expressed. The number of reasonably conserved predicted miRNAs is yet about twofold larger. Revealing of the actual number of miRNAs will largely depend on continued efforts to clone and sequence small RNA libraries.

miRNAs act as regulators of gene expression on many protein-coding genes, though the precise number of their targets remains debatable. The analysis and interpretation of global gene expression patterns is therefore tightly linked to understanding small RNA expression patterns. In fact, miRNA profiles seem to have an advantage over mRNA profiles when characterizing histologically uncertain types of cancer [91]. The miRNA microarray technology will undergo several updates with respect to new miRNA sequences, and will then likely become the preferred platform for whole genome miRNA expression analysis.

With respect to endogenous siRNAs and rasiRNAs, efforts to increase their cloning frequency should facilitate their annotation and functional characterization. Increasing the throughput of library sequencing using modern technology could provide sufficient clone numbers to detect rare endogenous sources of dsRNA. Procedures to isolate intact small RNA-loaded ribonucleoprotein complexes might be used to enrich for functional small RNAs.

Acknowledgments: We thank members of Tuschl laboratory, M. Zavolan, E. Bernstein, and A. Bosio for critical reading and stimulating discussions.

References

[1] Lee, R.C., Feinbaum, R.L. and Ambros, V. (1993) The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843–854.
[2] Wightman, B., Ha, I. and Ruvkun, G. (1993) Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. Cell 75, 855–862.
[3] Chalfie, M., Horvitz, H.R. and Sulston, J.E. (1981) Mutations that lead to reiterations in the cell lineages of *C. elegans*. Cell 24, 59–69.
[4] Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R. and Ruvkun, G.

(2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403, 901–906.
[5] Pasquinelli, A.E. et al. (2000) Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature 408, 86–89.
[6] Johnston, R.J. and Hobert, O. (2003) A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*. Nature 426, 845–849.
[7] Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B. and Cohen, S.M. (2003) Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in *Drosophila*. Cell 113, 25–36.
[8] Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001) Identification of novel genes coding for small expressed RNAs. Science 294, 853–858.
[9] Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. Science 294, 862–864.
[10] Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science 294, 858–862.
[11] Waterhouse, P.M., Wang, M.B. and Lough, T. (2001) Gene silencing as an adaptive defence against viruses. Nature 411, 834–842.
[12] Baulcombe, D. (2004) RNA silencing in plants. Nature 431, 356–363.
[13] Faugeron, G. (2000) Diversity of homology-dependent gene silencing strategies in fungi. Curr. Opin. Microbiol. 3, 144–148.
[14] Birchler, J.A., Bhadra, M.P. and Bhadra, U. (2000) Making noise about silence: repression of repeated genes in animals. Curr. Opin. Genet. Dev. 10, 211–216.
[15] Aravin, A.A., Naumova, N.M., Tulin, A.V., Vagin, V.V., Rozovsky, Y.M. and Gvozdev, V.A. (2001) Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. Curr. Biol. 11, 1017–1027.
[16] Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391, 806–811.
[17] Svoboda, P. (2004) Long dsRNA and silent genes strike back: RNAi in mouse oocytes and early embryos. Cytogenet. Genome Res. 105, 422–434.
[18] Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. and Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411, 494–498.
[19] Hamilton, A.J. and Baulcombe, D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286, 950–952.
[20] Hamilton, A., Voinnet, O., Chappell, L. and Baulcombe, D. (2002) Two classes of short interfering RNA in RNA silencing. EMBO J. 21, 4671–4679.
[21] Zamore, P.D., Tuschl, T., Sharp, P.A. and Bartel, D.P. (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell 101, 25–33.
[22] Elbashir, S.M., Lendeckel, W. and Tuschl, T. (2001) RNA interference is mediated 1- and 22-nucleotide RNAs. Genes Dev. 15, 188–200.
[23] Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. and Tuschl, T. (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell 110, 563.
[24] Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T. and Patel, D.J. (2005) Structural basis for 5′-end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. Nature 434, 666–670.
[25] Parker, J.S., Roe, S.M. and Barford, D. (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature 434, 663–666.
[26] Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J. and Joshua-Tor, L. (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. Nat. Struct. Mol. Biol. 12, 340–349.
[27] Song, J.J., Smith, S.K., Hannon, G.J. and Joshua-Tor, L. (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305, 1434–1437.
[28] Liu, J. et al. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. Science 305, 1437–1441.
[29] Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G. and Tuschl, T. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol. Cell 15, 185–197.
[30] Meister, G. and Tuschl, T. (2004) Mechanisms of gene silencing by double-stranded RNA. Nature 431, 343–349.
[31] Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116, 281–297.
[32] Ambros, V. (2004) The functions of animal microRNAs. Nature 431, 350–355.
[33] Cullen, B.R. (2004) Transcription and processing of human microRNA precursors. Mol. Cell 16, 861–865.
[34] Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B. and Bartel, D.P. (2002) Prediction of plant microRNA targets. Cell 110, 513–520.
[35] Lewis, B.P., Burge, C.B. and Bartel, D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15–20.
[36] Krek, A. et al. (2005) Combinatorial microRNA target predictions. Nat. Genet. 37, 495–500.
[37] Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S. and Kellis, M. (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. Nature 434, 338–345.
[38] Reinhart, B.J. and Bartel, D.P. (2002) Small RNAs correspond to centromere heterochromatic repeats. Science 297, 1831.
[39] Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I. and Martienssen, R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297, 1833–1837.
[40] Aravin, A.A. et al. (2003) The small RNA profile during *Drosophila melanogaster* development. Dev. Cell 5, 337–350.
[41] Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E. and Carrington, J.C. (2004) Genetic and functional diversification of small RNA pathways in plants. PLoS Biol. 2, E104.
[42] Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J. and Carthew, R.W. (2004) Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell 117, 69–81.
[43] Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W. and Sontheimer, E.J. (2004) A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in *Drosophila*. Cell 117, 83–94.
[44] Djikeng, A., Shi, H., Tschudi, C. and Ullu, E. (2001) RNA interference in *Trypanosoma brucei*: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. RNA 7, 1522–1530.
[45] Matzke, M., Aufsatz, W., Kanno, T., Daxinger, L., Papp, I., Mette, M.F. and Matzke, A.J. (2004) Genetic analysis of RNA-mediated transcriptional gene silencing. Biochim. Biophys. Acta 1677, 129–141.
[46] Lippman, Z. and Martienssen, R. (2004) The role of RNA interference in heterochromatic silencing. Nature 431, 364–370.
[47] Pfeffer, S. et al. (2005) Identification of microRNAs of the herpesvirus family. Nat. Meth. 2, 269–276.
[48] Tonkin, L.A. and Bass, B.L. (2003) Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. Science 302, 1725.
[49] Athanasiadis, A., Rich, A. and Maas, S. (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. 2, e391.
[50] Levanon, E.Y. et al. (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat. Biotech. 22, 1001–1005.
[51] Vazquez, F. et al. (2004) Endogenous trans-acting siRNAs regulate the accumulation of *Arabidopsis* mRNAs. Mol. Cell 16, 69–79.
[52] Allen, E., Xie, Z., Gustafson, A.M. and Carrington, J.C. (2005) MicroRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. Cell 121, 207–221.

[53] Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T. and Jewell, D. (2003) MicroRNAs and other tiny endogenous RNAs in *C. elegans*. Curr. Biol. 13, 807–818.
[54] Thomson, J.M., Parker, J., Perou, C.M. and Hammond, S.M. (2004) A custom microarray platform for analysis of microRNA gene expression. Nat. Meth. 1, 47–53.
[55] Aravin, A.A., Klenov, M.S., Vagin, V.V., Bantignies, F., Cavalli, G. and Gvozdev, V.A. (2004) Dissection of a natural RNA silencing process in the *Drosophila melanogaster* germ line. Mol. Cell. Biol. 24, 6742–6750.
[56] Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B. and Bartel, D.P. (2003) The microRNAs of *Caenorhabditis elegans*. Genes Dev. 17, 991–1008.
[57] Soutschek, J. et al. (2004) Therapeutic silencing of an endogenous gene by systemic administration of modified siRNAs. Nature 432, 173–178.
[58] Sijen, T. and Plasterk, R.H. (2003) Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. Nature 426, 310–314.
[59] Seitz, H., Royo, H., Bortolin, M.L., Lin, S.P., Ferguson-Smith, A.C. and Cavaille, J. (2004) A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. Genome Res. 14, 1741–1748.
[60] Schmittgen, T.D., Jiang, J., Liu, Q. and Yang, L. (2004) A high-throughput method to monitor the expression of microRNA precursors. Nucleic Acids Res. 32, e43.
[61] Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B. and Bartel, D.P. (2003) Vertebrate microRNA genes. Science 299, 1540.
[62] Krichevsky, A.M., King, K.S., Donahue, C.P., Khrapko, K. and Kosik, K.S. (2003) A microRNA array reveals extensive regulation of microRNAs during brain development. RNA 9, 1274–1281.
[63] Babak, T., Zhang, W., Morris, Q., Blencowe, B.J. and Hughes, T.R. (2004) Probing microRNAs with microarrays: tissue specificity and functional inference. RNA 10, 1813–1819.
[64] Wiegant, J.C., van Gijlswijk, R.P., Heetebrij, R.J., Bezrookove, V., Raap, A.K. and Tanke, H.J. (1999) ULS: a versatile method of labeling nucleic acids for FISH based on a monofunctional reaction of cisplatin derivatives with guanine moieties. Cytogenet. Cell Genet. 87, 47–52.
[65] Liang, R.Q., Li, W., Li, Y., Tan, C.Y., Li, J.X., Jin, Y.X. and Ruan, K.C. (2005) An oligonucleotide microarray for microRNA expression analysis based on labeling RNA with quantum dot and nanogold probe. Nucleic Acids Res. 33, e17.
[66] Richardson, R.W. and Gumport, R.I. (1983) Biotin and fluorescent labeling of RNA using T4 RNA ligase. Nucleic Acids Res. 11, 6167–6184.
[67] Nelson, P.T., Baldwin, D.A., Scearce, L.M., Oberholtzer, J.C., Tobias, J.W. and Mourelatos, Z. (2004) Microarray-based, high-throughput gene expression profiling of microRNAs. Nat. Meth. 1, 155–161.
[68] Liu, C.G. et al. (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. Proc. Natl. Acad. Sci. USA 101, 9740–9744.
[69] Calin, G.A. et al. (2004) MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. Proc. Natl. Acad. Sci. USA 101, 11755–11760.
[70] Sun, Y., Koo, S., White, N., Peralta, E., Esau, C., Dean, N.M. and Perera, R.J. (2004) Development of a micro-array to detect human and mouse microRNAs and characterization of expression in human organs. Nucleic Acids Res. 32, e188.
[71] Barad, O. et al. (2004) MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. Genome Res. 14, 2486–2494.
[72] Miska, E.A., Alvarez-Saavedra, E., Townsend, M., Yoshii, A., Sestan, N., Rakic, P., Constantine-Paton, M. and Horvitz, H.R. (2004) Microarray analysis of microRNA expression in the developing mammalian brain. Genome Biol. 5, R68.
[73] Baskerville, S. and Bartel, D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. RNA 11, 241–247.
[74] Axtell, M.J. and Bartel, D.P. (2005) Antiquity of microRNAs and their targets in land plants. Plant Cell 17, 1658–1673.
[75] Chen, X. (2004) A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. Science 303, 2022–2025.
[76] Kosman, D., Mizutani, C.M., Lemons, D., Cox, W.G., McGinnis, W. and Bier, E. (2004) Multiplex detection of RNA expression in *Drosophila* embryos. Science 305, 846.
[77] Valoczi, A., Hornyik, C., Varga, N., Burgyan, J., Kauppinen, S. and Havelda, Z. (2004) Sensitive and specific detection of microRNAs by Northern blot analysis using LNA-modified oligonucleotide probes. Nucleic Acids Res. 32, e175.
[78] Wienholds, E. et al. (2005) MicroRNA expression in zebrafish embryonic development. Science.
[79] Mansfield, J.H. et al. (2004) MicroRNA-responsive 'sensor' transgenes uncover Hoxlike and other developmentally regulated patterns of vertebrate microRNA expression. Nat. Genet. 36, 1079–1083.
[80] Giraldez, A.J. et al. (2005) MicroRNAs regulate brain morphogenesis in zebrafish. Science.
[81] Stark, A., Brennecke, J., Russell, R.B. and Cohen, S.M. (2003) Identification of *Drosophila* microRNA targets. PLoS Biol. 1, E60.
[82] Brennecke, J., Stark, A., Russell, R.B. and Cohen, S.M. (2005) Principles of microRNA-target recognition. PLoS Biol. 3, e85.
[83] Lai, E.C., Tam, B. and Rubin, G.M. (2005) Pervasive regulation of *Drosophila* Notch target genes by GY-box-, Brd-box-, and K-box-class microRNAs. Genes Dev..
[84] Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R.W., Steward, R. and Chen, X. (2005) Methylation as a crucial step in plant microRNA biogenesis. Science 307, 932–935.
[85] Nykanen, A., Haley, B. and Zamore, P.D. (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell 107, 309–321.
[86] Ambros, V. and Lee, R.C. (2004) Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. Methods Mol. Biol. 265, 131–158.
[87] Pfeffer, S., Lagos-Quintana, M. and Tuschl, T. (2003) in: Current Protocls in Molecular Biology (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidmann, J.G., Smith, J.A. and Struhl, K., Eds.), pp. 26.24.16–26.24.21, Wiley, New York.
[88] Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A. and Darnell, R.B. (2003) CLIP identifies Nova-regulated RNA networks in the brain. Science 302, 1212–1215.
[89] Wood, Z.A., Sabatini, R.S. and Hajduk, S.L. (2004) RNA ligase; picking up the pieces. Mol. Cell 13, 455–456.
[90] Ho, C.K., Wang, L.K., Lima, C.D. and Shuman, S. (2004) Structure and mechanism NA ligase. Structure (Camb) 12, 327–339.
[91] Lu, J. et al. (2005) MicroRNA expression profiles classify human cancers. Nature 435, 834–838.
[92] Luciano, D.J., Mirsky, H., Vendetti, N.J. and Maas, S. (2004) RNA editing of a miRNA precursor. RNA 10, 1174–1177.
[93] Lai, E.C., Tomancak, P., Williams, R.W. and Rubin, G.M. (2003) Computational identification of *Drosophila* microRNA genes. Genome Biol. 4, R42.
[94] Altuvia, Y. et al. (2005) Clustering and conservation patterns of human microRNAs. Nucleic Acids Res. 33, 2697–2706.
[95] Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R.H. and Cuppen, E. (2005) Phylogenetic shadowing and computational identification of human microRNA genes. Cell 120, 21–24.
[96] Chen, P.Y. et al. (2005) The developmental miRNA profiles of zebrafish as determined by small RNA cloning. Genes Dev. 19, 1288–1293.
[97] Grad, Y., Aach, J., Hayes, G.D., Reinhart, B.J., Church, G.M., Ruvkun, G. and Kim, J. (2003) Computational and experimental identification of *C. elegans* microRNAs. Mol. Cell 11, 1253–1263.
[98] Legendre, M., Lambert, A. and Gautheret, D. (2005) Profile-based detection of microRNA precursors in animal genomes. Bioinformatics 21, 841–845.
[99] Nam, J.W., Shin, K.R., Han, J., Lee, Y., Kim, V.N. and Zhang, B.T. (2005) Human microRNA prediction through a probabilistic co-learning model of sequence and structure. Nucleic Acids Res. 33, 3570–3581.

[100] Bentwich, I. et al. (2005) Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet.
[101] Watanabe, T., Takeda, A., Mise, K., Okuno, T., Suzuki, T., Minami, N. and Imai, H. (2005) Stage-specific expression of microRNAs during *Xenopus* development. FEBS Lett. 579, 318–324.
[102] Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W. and Tuschl, T. (2002) Identification of tissue-specific microRNAs from mouse. Curr. Biol. 12, 735–739.
[103] Mourelatos, Z. et al. (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. 16, 720–728.
[104] Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A. and Tuschl, T. (2003) New microRNAs from mouse and human. RNA 9, 175–179.
[105] Dostie, J., Mourelatos, Z., Yang, M., Sharma, A. and Dreyfuss, G. (2003) Numerous microRNPs in neuronal cells containing novel microRNAs. RNA 9, 180–186.
[106] Houbaviy, H.B., Murray, M.F. and Sharp, P.A. (2003) Embryonic stem cell-specific microRNAs. Dev. Cell 5, 351–358.
[107] Kim, J., Krichevsky, A., Grad, Y., Hayes, G.D., Kosik, K.S., Church, G.M. and Ruvkun, G. (2004) Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. Proc. Natl. Acad. Sci. USA 101, 360–365.
[108] Suh, M.R. et al. (2004) Human embryonic stem cells express a unique set of microRNAs. Dev. Biol. 270, 488–498.
[109] Poy, M.N. et al. (2004) A pancreatic islet-specific microRNA regulates insulin secretion. Nature 432, 226–230.
[110] Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B. and Bartel, D.P. (2002) MicroRNAs in plants. Genes Dev. 16, 1616–1626.
[111] Jones-Rhoades, M.W. and Bartel, D.P. (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol. Cell 14, 787–799.
[112] Sunkar, R., Girke, T., Jain, P.K. and Zhu, J.K. (2005) Cloning and characterization of microRNAs from rice. Plant Cell 17, 1397–1411.
[113] Sunkar, R. and Zhu, J.K. (2004) Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. Plant Cell 16, 2001–2019.
[114] Wang, X.J., Reyes, J.L., Chua, N.H. and Gaasterland, T. (2004) Prediction and identification of *Arabidopsis thaliana* microRNAs and their mRNA targets. Genome Biol. 5, R65.
[115] Wang, J.F., Zhou, H., Chen, Y.Q., Luo, Q.J. and Qu, L.H. (2004) Identification of 20 microRNAs from *Oryza sativa*. Nucleic Acids Res. 32, 1688–1695.
[116] Adai, A., Johnson, C., Mlotshwa, S., Archer-Evans, S., Manocha, V., Vance, V. and Sundaresan, V. (2005) Computational prediction of miRNAs in *Arabidopsis thaliana*. Genome Res. 15, 78–91.
[117] Pfeffer, S. et al. (2004) Identification of virus-encoded microRNAs. Science 304, 734–736.
[118] Cai, X., Lu, S., Zhang, Z., Gonzalez, C.M., Damania, B. and Cullen, B.R. (2005) Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. Proc. Natl. Acad. Sci. USA 102, 5570–5575.