Isolation and characterization of tissue and development-specific microRNAs from Drosophila, mouse and human

PhD Thesis

in partial fulfillment of the requirements

for the degree “Doctor of Philosophy (PhD)”

in the Molecular Biology Program

at the Georg August University Göttingen, Faculty of Biology

submitted by

Mariana Lagos-Quintana

born in

Mexico City

2005

Isolation and characterization of tissue and development-specific microRNAs from Drosophila, mouse and human

PhD Thesis

in partial fulfillment of the requirements

for the degree “Doctor of Philosophy (PhD)”

in the Molecular Biology Program

at the Georg August University Göttingen, Faculty of Biology

submitted by

Mariana Lagos-Quintana

born in

Mexico City

2005

Advisor, First doctoral committee member:	Dr. Thomas Tuschl
Second doctoral committee member:	Prof. Dr. Reinhard Lührmann
Third doctoral committe member:	Prof. Dr. Ralf Ficner

Date of submission of the PhD thesis: 15 April 2005

Acknowledgements

Dr. Thomas Tuschl, my supervisor and mentor for guiding me into the RNA world and giving encouragement and support.

Professors Dr. Reinhard Lührmann and Dr. Ralf Ficner, members of my thesis committee for their advice and guidance.

I would like to express my gratitude to the MS/PhD Program Molecular Biology of the Georg August University Göttingen and the International Max Planck Research School, in particular to Dr. Steffen Burkhardt.

I'd like to thank former and present and members of the Tuschl lab for discussions, advice, oligonucleotide synthesis and friendship.

Dr. Robert Sheridan for help in A20 library sequence analysis.

Dr. Stuart Shuman at the Memorial Sloan-Ketterin Cancer Center for providing the Rnl2(1-249) ligase.

Dr. Peter Unrau at the Simon Fraser University for providing the pRHT4 plasmid and advice on T4 RNA ligase purification.

Dr. James Russo and his group at the Columbia Genome Center, for sequencing small RNA libraries.

My family.

TABLE OF CONTENTS

Abstract

MicroRNAs (miRNAs) are short non-coding RNAs that regulate gene expression in plants, animals and viruses. These small RNAs provide specificity to protein effector complexes mediating posttranscriptional silencing. The *Caenorhabditis elegans lin-4* gene was the first miRNA characterized and for a long time was deemed to be unique to nematodes. Later, the *let-7* miRNA was discovered in *C. elegans*, and it was found to be evolutionary conserved in the genomes of *Drosophila melanogaster* and human. This work describes the cloning of small RNAs from Drosophila, mouse and human. Among the cloned RNAs was *let-7* and many novel invertebrate and mammalian miRNAs. Several miRNAs were found to be tissue-specific and developmentally regulated. Besides miRNAs, repeat-associated small interfering RNAs were cloned from *Drosophila*; this class of small regulatory RNAs derived from transposons and other repeat regions from fly. During the last few years miRNAs have been further characterized by many laboratories. Their mode of expression and biosynthesis pathway have been studied in depth, and proteins involved in the miRNA-guided silencing pathway have been identified. The current understanding of miRNA and related RNA silencing phenomena is presented. In addition, this work describes the optimization and validation of the cloning procedure used for the discovery of miRNAs. The improvements of the protocol are discussed in the context of other small RNA profiling methods.

General introduction

Discovery of small regulatory RNAs

In 1981 the *Caenorhabditis elegans* gene *lin-4* was discovered (Chalfie et al., 1981). Mutations in *lin-4* caused heterochronic defects in larval development. Heterochronic genes coordinate the succession of cell fates within lineages in *C. elegans* larval tissue. At each of the four larval stages, a subset of these regulatory genes acts as a switch of alternative cell fates (reviewed by Ambros, 2000). In *lin-4* mutants, particular cells undergo division patterns normally associated with their progenitors (Chalfie et al., 1981). It was later discovered that the *lin-4* gene does not code for a protein (Lee et al., 1993). Two *lin-4* transcripts of different length were identified, one of 22 nt and one of 61 nt. The 61 nt RNA was predicted to form a hairpin secondary structure and is the precursor of the 22 nt RNA. Lee and colleagues noticed that the *lin-4* transcripts were partially complementary to the 3' untranslated region (UTR) of the heterochronic gene *lin-14* (Lee et al., 1993). *lin-14* mRNA bears seven *lin-4* binding sites necessary for LIN-14 repression in vivo after the second larval stage. The base-pairing of the small RNA to the 3' UTR of *lin-14* confers regulation of protein synthesis (Ha et al., 1996; Olsen and Ambros, 1999; Wightman et al., 1993). *lin-4* also regulates LIN-28, a cold shock protein that initiates the transition between the larval stages 2 and 3 (Moss et al., 1997).

Almost seven years after describing the *lin-4* guided gene regulation, the Ruvkun laboratory identified *let-7*, the second heterochronic small RNA (Reinhart et al., 2000). Its expression is observed at the third larval stage triggering the transition from late larva to adult; loss of *let-7* caused reiteration of larval cell fates in the adult worm, whereas over-expression of *let-7* caused precocious expression of adult cell fates during larval stages. Like *lin-4*, *let-7* is complementary to the 3' UTRs of heterochronic genes, and it was shown to repress a reporter gene bearing *let-7* complementary sites in its 3' UTR (Reinhart et al., 2000).

For some time *lin-4* and *let-7* were the only naturally occurring antisense RNAs known to function as trans-acting regulatory elements in eukaryotes, and were deemed to be unique to worms. Later, homologues of *let-7* were found in the genomes of *D. melanogaster* and

human, and the *let-7* 21 nt transcript was detected in the three main clades of bilaterian animals. (Pasquinelli et al., 2000). The temporal regulation of *C. elegans let-7* was also conserved in other animals and it was therefore proposed to name this RNA type small temporal RNAs (stRNAs) (Pasquinelli et al., 2000).

Studies in plants (Napoli et al., 1990), fungi (Cogoni et al., 1996) and *Drosophila melanogaster* (Pal-Bhadra et al., 1997) were uncovering a phenomenon by which trans-genes inhibit the expression of homologous genes. It was first described as co-suppression or posttranscriptional gene silencing (PTGS). As a debate started about the trigger of PTGS in plants, the requirement for an RNA antisense to the target was recognized (Baulcombe, 1996), and later the it was shown that transcripts capable of forming double stranded RNA (dsRNA) were better than antisense RNA to induce PTGS (Waterhouse et al., 1998). In animals, the antisense RNA approach to inactivate a *C. elegans* gene found that injection of sense-strand RNA was actually as effective as the antisense RNA at inhibiting gene function (Guo and Kemphues, 1995). Further studies revealed that introduction of dsRNA into nematodes lead to silencing of cognate endogenous genes (Fire et al., 1998). It was surprising at the time that dsRNA was a better silencing trigger than antisense RNA. This discovery was fundamental for the identification of dsRNA as the trigger for PTGS. The term RNA interference (RNAi) was coined, and soon thereafter RNAi was recognized as a cellular mechanism conserved among most eukaryotes aiding in viral defense and transposon silencing (reviewed by Mello and Conte, 2004). Two major findings supported the connection between small regulatory RNAs and RNAi. The first discovery was the identification of ~23-25 nt RNA species in plants that showed PTGS induced by a transgene (Hamilton and Baulcombe, 1999; Hutvágner et al., 2000). It was later shown that dsRNA is specifically processed to 21 nt RNA in *Drosophila* embryonic extract. These RNAs were named small interfering RNAs (siRNAs) by the Tuschl laboratory (Elbashir et al., 2001a). Secondly, the observation that components of the RNAi machinery were involved in the processing or function of *let-7* and *lin-4* (Bernstein et al., 2001; Grishok et al., 2001; Hutvágner et al., 2001; Ketting et al., 2001).

The Tuschl laboratory had cloned siRNAs from *Drosophila* embryonic extract after the addition of long dsRNA (Elbashir et al., 2001a). Besides the sequences derived from the

exogenous dsRNA were small RNAs encoded in the *Drosophila* genome. I was prompted to look into small RNAs present in human cell lines and to conduct a more detailed analysis of small RNAs in fly (Lagos-Quintana et al., 2001). Interestingly, the presence of *let-7* in HeLa cells and *Drosophila* was confirmed, and other human and *Drosophila* RNAs with conserved stem-loop precursors were discovered, which were then collectively named microRNAs (miRNAs). At the same time the Bartel and Ambros laboratories identified new miRNAs from *C. elegans* (Lau et al., 2001; Lee and Ambros, 2001). Having discovered that the miRNAs were a diverse family of expressed RNAs, I initiated a major effort to identify more mammalian miRNA genes and their expression pattern. Northern blot analysis of the first group of identified sequences indicated that miRNAs are both developmentally regulated and tissue specific. I created libraries from different mouse tissues (Chapter I and II), and from developmental stages in *Drosophila* (Chapter III). Other laboratories followed these efforts and cloned miRNAs from mammals, fish, worms and flies (Ambros et al., 2003; Dostie et al., 2003; Houbaviy et al., 2003; Kim et al., 2004; Lim et al., 2003a; Lim et al., 2003b; Mourelatos et al., 2002; Pfeffer et al., 2004; Poy et al., 2004).

A distinct group of small RNAs derived from heterochromatic centromeric repeats was cloned from *Schizosaccharomyces pombe* (Reinhart and Bartel, 2002). Some cloned RNAs from *Drosophila* were derived from transposable elements, satellite, microsatellite and Suppressor of Stellate transcripts (Chapter III). These RNAs were named repeat-associated small interfering RNAs (rasiRNAs) and constitute a third class of small regulatory RNAs besides miRNAs and siRNAs. rasiRNAs are supposed to silence the transcription of repeat elements in the genome (see “other small regulatory RNAs and RNAi-related mechanisms”)

Besides extensive efforts to identify miRNAs by cloning, computational efforts were undertaken. Computational algorithms designed to predict miRNAs are based on the structural characteristics of the stem-loop precursors, the conservation of miRNA genes between related species and the proximity of stem-loops to previously identified miRNAs, (Bonnet et al., 2004; Grad et al., 2003; Jones-Rhoades and Bartel, 2004; Lai et al., 2003; Lim et al., 2003b; Ohler et al., 2004; Seitz et al., 2004; Wang et al., 2004). miRscan, an algorithm developed by the Burge and Bartel laboratories (Lim et al.,

2003c), finds hairpins by sliding a 110 nt window across a single genome and folding each segment as RNA to identify hairpins of sufficient free energy and length. In the next step hairpins are analyzed for conservation in other species by BLAST. Finally, the program creates a consensus structure and evaluates all possible 21 nt candidate miRNAs based on their distance to the terminal loop, extent of their conservation in the 5' end versus the 3' end of the miRNA (see Regulatory function of miRNAs) and a tendency of miRNAs to begin with a uridine (~80% of known miRNAs have a uridine on the 5' end). miRseeker, by Eric Lai and colleagues, on the other hand, first identifies conserved sequences through a genome alignment and extracts 100 nt windows within conserved non-coding regions (Lai et al., 2003). miRseeker then folds the fragments and evaluates the length, the quality and the free energy of a conserved hairpin structure. Finally, the distribution of divergent nucleotides across the hairpin is evaluated; the miRNA should be more conserved than the rest of the precursor.

Computational methods predicting miRNAs have estimated the total number of miRNA genes per genome, and have been important in identifying new miRNAs.

miRNA genes

The miRNA Registry (Rfam) (Griffiths-Jones, 2004) is a web-based database created to compile all validated miRNAs and to assign names. The database includes miRNAs from vertebrates, invertebrates, plants and viruses. By the time of submission of this thesis, the database (Rfam, release 5.1) listed 222 miRNAs in human, 224 in mouse, 78 in *D. melanogaster*, 116 in *C. elegans*, 121 in chicken, 30 in zebra fish, 112 in *Arabidopsis thaliana* and 5 in Epstein Barr Virus, a member of the herpes virus family. Most of the miRNA genes are conserved among mammalian species, and about one third of the insect and worm miRNAs have homologues in mammalian genomes (reviewed by Bartel, 2004).

The genomic location of a number of miRNAs was analyzed to understand their mode of transcription. miRNAs were located in intergenic regions, suggesting that they have their own promoter and function as independent transcription units; about33% of the miRNAs were present in introns of either protein coding genes or previously characterized non-coding RNAs (ncRNAs). Roughly 10% of them were located in introns, but in antisense

orientation (Chapter II). Recently a more comprehensive study surveyed the genomic location of all mammalian miRNAs known to date (Rodriguez et al., 2004). Close to 40% of miRNAs were found in introns of protein coding genes, 12% in introns of long ncRNA transcripts, 13% in exons of ncRNAs and a few miRNAs genes were found either in introns or exons, depending on alternative splicing. Only 30% of all miRNAs were found to be intergenic, suggesting that most of them are co-transcribed with other genes and processed from introns in the nucleus (4 of 5 miRNAs analyzed were co-expressed with the host mRNA). Some intronic miRNAs are of particular interest since their intronic locations are evolutionary conserved. The most striking case is miR-7, which is located in the last intron of the *heterogeneous nuclear ribonucleoprotein K* (*hnRNPK*) of *D. melanogaster*, *A. gambiae*, mouse and human (Chapter III).

Some miRNA genes are located in previously characterized ncRNAs. mir-155 was found in the BIC RNA and miRNA-15, -16 cluster is located in DLEU2 RNA (Chapters I and II). Rodriguez and colleagues (Rodriguez et al., 2004) identified additional miRNAs located in earlier described ncRNAs; for example, mir-206 is found in an exon of the synapse-specific 7H4 transcript, and mir-135a-2 in the intron of NCRMS. These ncRNAs are spliced and polyadenylated.

Many miRNAs are frequently clustered in the genome with similar expression patterns implying a multi-cistronic primary transcript (see (Lagos-Quintana et al., 2001) and Chapter II). The miRNAs within a cluster are often related in sequence. We showed examples of these clusters (Lagos-Quintana et al., 2001), which include different variants of a miRNA or many copies of the same miRNA with minor sequence changes in the precursor sequence. Some clusters have been duplicated; for example, in *Drosophila* mir-2 and mir-13 genes are related to each other in sequence; mir-2 variants are clustered on chromosome 2 while mir-13 variants are found on chromosome 3. Homologues of *lin-4* and *let-7* are clustered in insect and mammalian genomes (Chapter III). The mir-100, let-7 and mir-125 (homologue of *lin-4*) cluster is developmentally regulated in both flies and mammals. In *C. elegans lin-4* and *let-7* are not clustered and their expression during development differs as well.

miRNA biogenesis

Several miRNA primary transcripts (pri-miRNAs) have been cloned and characterized from human, *C. elegans* and plants (reviewed by Cullen, 2004). These transcripts are capped and polyadenylated. Based on these features, the current consensus is that most pri-miRNAs are transcribed by RNA polymerase II (pol II), and share structural and processing features with mRNAs. However, several recently discovered miRNAs from mouse gammaherpesvirus 68 are located immediately downstream of tRNA sequences. It was suggested that these miRNA transcripts are initiated from a tRNA-specific pol III promoter (Pfeffer et al., 2005).

Primary miRNA transcripts (pri-miRNAs) are processed into miRNA hairpin precursors (pre-miRNAs) and subsequently into mature miRNAs. Mature miRNAs are then loaded into an effector complex, responsible for regulating gene expression either by mRNA cleavage or translational repression. The processing of pri-miRNAs takes place in a stepwise manner by two RNase III enzymes: Drosha and Dicer (Figure 1). RNase III enzymes are conserved in organisms from bacteria to animals and regulate RNA metabolism; they are specific for dsRNA and their cleavage products have a 5'phosphate and 3' hydroxyl with a 2 nt 3' overhang (Robertson, 1982; Robertson et al., 1968). The *Escherichia coli* RNase III is involved in maturation of rRNAs, mRNAs and tRNAs. This nuclease has one RNase III domain and one dsRNA-binding domain (dsRBD). Drosha is predominantly nuclear and cleaves pri-miRNAs to pre-miRNAs of about 60 nt. The N-terminus of the mammalian Drosha is thought to be involved in protein-protein interactions; it contains a proline-rich region (PRR), an arginine-serine-rich domain (RS), and two dsRBDs. In cultured cells, Drosha knockdown results in accumulation of pri-miRNA transcripts (Lee et al., 2003). DGCR8, a protein containing two dsRBDs and a WW domain, is necessary for pri-miRNA processing and it interacts with Drosha in the microprocessor complex (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004a; Landthaler et al., 2004; Shiohama et al., 2003).

The transport of pre-miRNAs from the nucleus to the cytoplasm is mediated by Exportin 5 in a Ran-GTP dependent manner (Bohnsack et al., 2004; Lund et al., 2004; Park et al., 2005; Yi et al., 2003). In the cytoplasm Dicer clips off the loop sequence of the pre-miRNA producing a miRNA duplex.

Dicer is present in all organisms that are competent for RNAi. Plants have four Dicers, DCL-1 through 4; *Drosophila* has two Dicers, Dcr-1 and Dcr-2; and mammals, *C. elegans* and *S. pombe* encode only one. Dicer contains two RNase III domains, a

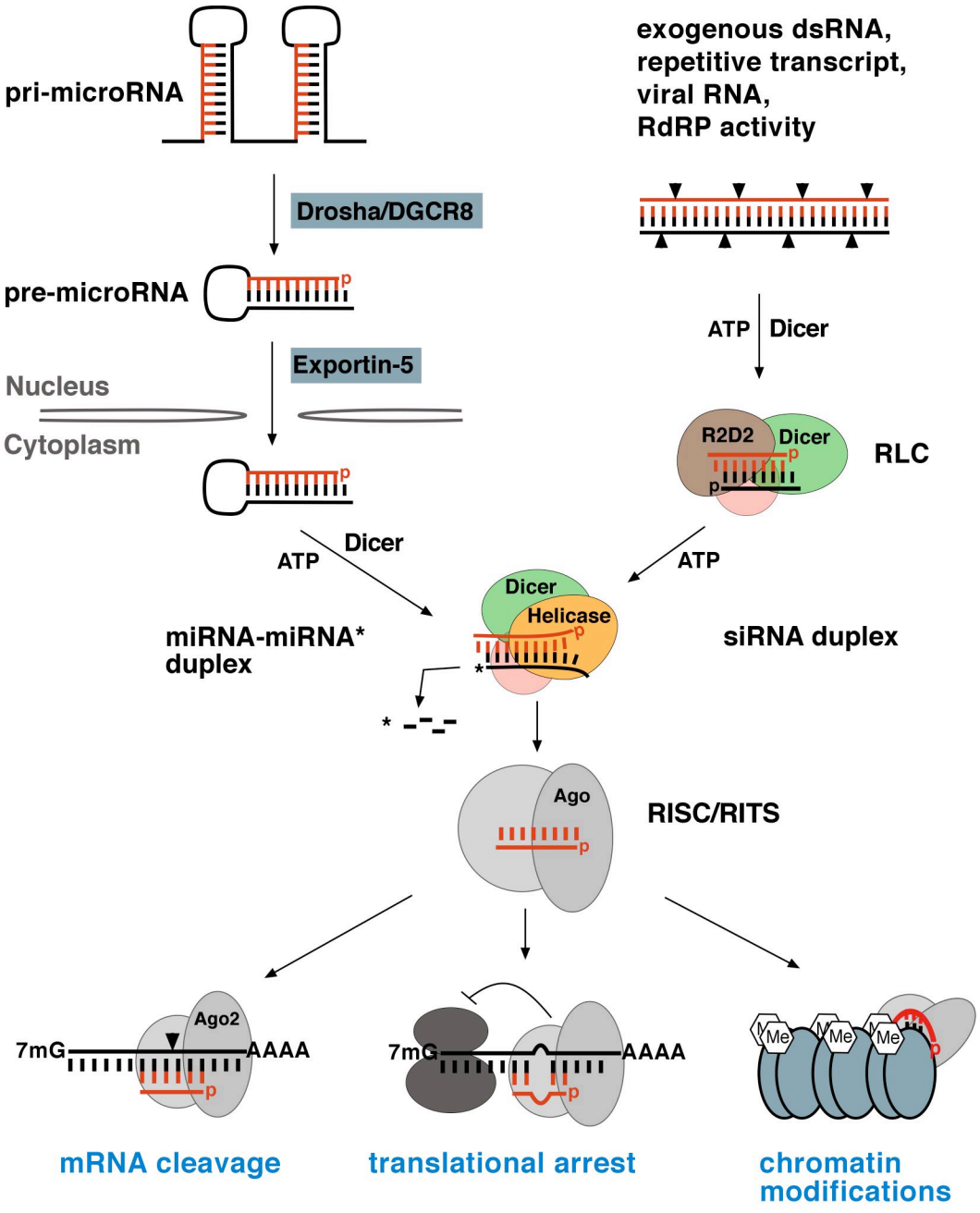


Figure 1. RNA-mediated gene silencing pathways

Biogenesis of miRNAs (left) and siRNAs (right) and incorporation of guide RNAs into different effector complexes. The pri-miRNA is processed by Drosha into pre-miRNA intermediates, which are exported by Exportin-5 into the cytoplasm, where they are processed by Dicer. siRNA incorporation into RISC requires the Dicer interactor R2D2 in *D. melanogaster*. The complex is termed RLC for RISC loading complex. Upon strand separation by a helicase, the miRNA* strand (black) is degraded and the miRNA (red) is incorporated into RISC. A monophosphate (P) marks the 5' terminus of each strand. Single stranded RNAs are present in effector complexes and guide silencing of complementary mRNAs (RISC) or DNA (RITS).

conserved DExH/DEAH box helicase domain, a PAZ domain, a dsRNA-binding motif and a DUF283 domain of unknown function.

In *Drosophila* the two Dicers have distinct functions; Dcr-1 produces predominantly miRNAs, while Dcr-2 cleaves exogenous or endogenous dsRNA to produce siRNAs (Lee et al., 2004b).

Arabidopsis Dicers have different cellular localizations as well as different functions.

The best-characterized Arabidopsis Dicer is DCL-1, the miRNA processing Dicer. DCL-1 mutations cause diverse developmental defects. DCL-1 is localized in the nucleus (Papp et al., 2003). It was recently shown that DCL-1 catalyses both the first and second processing steps in miRNA biogenesis in plants (Kurihara and Watanabe, 2004). Thus, the strategy for miRNA processing differs significantly between plants and animals. Unlike animal miRNAs, plant miRNAs are methylated on the ribose on the 3' end by HEN1 (Yu et al., 2005), which is necessary for both miRNA accumulation and siRNA production (Boutet et al., 2003; Park et al., 2002; Xie et al., 2003). HEN1 is associated with DCL-1 and with the nuclear dsRNA binding protein HYL1, also necessary for miRNA accumulation (Han et al., 2004b; Vazquez et al., 2004). Two of the remaining three Arabidopsis Dicers have defined functions; DCL-2 is implicated in processing viral dsRNA, (Xie et al., 2004), while DCL-3 processes rasiRNAs (Xie et al., 2004).

The PAZ domain

The PAZ domain is common to Dicer and proteins of the Argonaute family (Ago) that are key mediators of RNAi. The PAZ domain was thought to regulate protein-protein interactions until several groups resolved its structure and concluded that it binds to nucleic acids. The *Drosophila* Ago1 (Yan et al., 2003) and Ago2 (Lingel et al., 2003; Song et al., 2003), and the human Ago1 (Ma et al., 2004) PAZ domains are very similar to an oligonucleotide binding fold, previously characterized as a single stranded RNA-binding domain (ssRBD) (reviewed by Carmell and Hannon, 2004). The PAZ domain contains a cleft for nucleic acid binding with conserved aromatic residues. PAZ binding to nucleic acids is sequence unspecific (Lingel et al., 2003; Ma et al., 2004; Song et al., 2003; Yan et al., 2003). Ma and colleagues resolved a structure of a 9 nt duplex with 2 nt 3' overhangs bound to the PAZ domain. The overhang of one of the strands binds to a

conserved pocket and the rest of the phosphate backbone is bound in the cleft. The complementary strand only contacts the PAZ domain through the 5' terminal residue (Ma et al., 2004). The overhang contributes to the binding affinity of a duplex in agreement with the structure of miRNA and siRNA duplexes.

The RNA induced silencing complex

The mature single-stranded miRNAs are loaded into effector complexes termed miRNP (for miRNA ribonucleoprotein) and RISC (for RNA-induced silencing complex). In general, the first term refers to a miRNA-loaded complex and the second to a siRNA-loaded complex, thus, they have also been termed miRISC and siRISC, respectively. Initially these two complexes were seen as components of two different regulatory pathways. However, it was found in human cells that the endogenous let-7 miRNA could guide cleavage of an exogenously introduced mRNA with perfect complementarity to let-7 (Hutvágner and Zamore, 2002), and that siRNAs could repress mRNA translation, if they were partially complementary to sequence elements in the 3' UTR of mRNAs (Doench et al., 2003). Based on these observations it was proposed that the functional differences between RISC and miRNP depend on the degree of complementarity between the small RNA in the effector complex and the target mRNA. It was later shown that a miRNA can only guide cleavage of an mRNA, if it is loaded in a complex containing Ago2 (Meister et al., 2004b).

More recent evidence in *Drosophila* showed that miRNAs and siRNAs are incorporated into the effector complexes through different assembly pathways. The biological significance of these may be to distinguish between dsRNA triggers and load siRNAs and miRNAs into different complexes. Effector complexes contain a member of the Argonaute protein family as the core component, as indicated by genetic data from plants, nematodes and fungi (reviewed by Carmell et al., 2002), and more recently by biochemical data (reviewed by Meister and Tuschl, 2004). In addition, a number of proteins have been found to be associated with the Argonautes. The assembly process, purification and function of the effector complexes are described in the following paragraphs.

Duplex unwinding

Strand separation or unwinding of the siRNA/miRNA duplex in the silencing pathway must occur prior to formation of an active effector complex (Martinez et al., 2002; Tomari et al., 2004a).

The involvement of an RNA helicase is foreseeable since a stable RNA duplex needs to be unwound. Dicer has a helicase domain and it is not proven if Dicer is involved in duplex unwinding. Genetic and biochemical evidence support the requirement of a helicase in the RNAi pathway, although it is not known if any of these helicases act at the step of siRNA/miRNA duplex unwinding. A putative DEA(H/D) helicase is required in *Chlamydomonas reinhardtii* for PTGS (Wu-Scharf et al., 2000). In *C. elegans* DCR-1, RDE-4 and the Ago protein RDE-1 were co-purified with two helicases, DRH-1 and DRH–2 (Tabara et al., 2002). In *Drosophila*, mutations in the *spindle-E* gene, which encodes a DEAD-box helicase, cause de-repression of the *Stellete* (*Ste*) loci, which is normally repressed in male germ cells by a dsRNA-mediated mechanism resulting in expression of other normally silenced loci such as retrotransposons (Aravin et al., 2001). *Spindle-E* is also required for RNAi in *Drosophila* oocytes (Kennerdell et al., 2002). Likewise, the *Drosophila* gene *armitage* (*armi*) encodes a putative helicase. *armi* mutants fail to silence *Ste* in male germ cells, and *armi* mutant ovaries are defective for RNAi (Cook et al., 2004). The *armi* homolog in Arabidopsis, SDE3 is also required for PTGS (Dalmay et al., 2001). In *Drosophila* S2 cells, the putative helicase Dmp68 was co-purified with RISC activity (Ishizuka et al., 2002). In human cell extracts, a putative DEAD-box RNA helicase, Gemin 3, is associated with RISC (Hutvágner and Zamore, 2002; Mourelatos et al., 2002).

Duplex asymmetry and strand incorporation into RISC

The incorporation of one of the strands of the duplex into RISC is not random; the parameter defining which of the strands is preferably incorporated into RISC is the relative free energy between the two ends of the duplex. Whichever strand presents the weaker base-pairing on the 5’ end, is the one to preferably enter RISC (Khvorova et al., 2003; Schwarz et al., 2003). The other strand is thought to be degraded. Weak base-pairing at the 5’ end of the guide siRNA strand is now an important criteria for siRNA

design. Recent studies by the Zamore laboratory provide a model for selection of one of the strands from a duplex based on thermodynamic asymmetry (Tomari et al., 2004b). The model suggests that the Dcr-2-R2D2 protein complex binds asymmetrically to the siRNA duplex, in a manner that R2D2 binds to the thermodynamically more stable end and Dcr-2 to the less stable end. This asymmetry determines which strand enters RISC once the duplex is unwound (Tomari et al., 2004b).

RNA silencing is initiated by recognition of the dsRNA trigger and processing of the siRNA/miRNA duplex followed by assembly of the effector complex. Once loaded into RISC, a single standed small RNA guides distinct modes of silencing. Proteins of RNA silencing pathways can be placed either into the siRNA/miRNA biogenesis, the RISC loading complex or into the effector complex. However, these pathways must be linked since the guide RNA (siRNA or miRNA) has to be "handed" from one complex to the other.

Assembly of the effector complex

Many organisms encode multiple Dicer paralogues and/or Argonaute proteins with different roles in silencing pathways; the transition from the initiation to the effector complexes may involve specific factors for each case, and therefore multiple intermediate complexes. The best-studied assembly is the *Drosophila* siRISC formation from the Dcr-2-containing initiation complex. Dcr-2 is associated with R2D2. R2D2 does not affect Dcr-2 processing of siRNAs, but Dcr-2-R2D2 interaction is required for loading siRNAs into RISC (Liu et al., 2003).

The Zamore and Sontheimer groups have followed RISC assembly in *Drosophila* embryo extracts (Pham et al., 2004; Tomari et al., 2004a). Both groups found a complex of lower molecular weight than RISC, which contains Dcr-2 and R2D2. The complex defined as RISC loading complex (RLC) contains double stranded siRNA. The assembly of this complex requires ATP and it contains other protein factors in addition to Dcr-2 and R2D2 (Tomari et al., 2004a). A similar complex, defined by Pham and colleagues differs from the RLC in that its assembly does not require ATP and only consists of Dcr-2 and R2D2 (Pham et al., 2004). Embryo extracts lacking Armitage, the Ago protein aubergine

(Tomari et al., 2004a), or Ago2 (Okamura et al., 2004) fail to form any complex beyond RLC.

Assembly of miRISC is less well understood in any organism studied. As mentioned previously, *Drosophila* uses one Dicer for siRNA and a second one for miRNA production. The miRNA producing Dicer is independent of R2D2. Mammals and *C. elegans*, however, have only one Dicer that processes both types of small RNAs. It is not known how a single Dicer interacts with different Argonaute-containing effector complexes and how the specificity for different sources of dsRNA is controlled.

Effector complexes

The RISC assembly pathway culminates in a complex that has a single stranded siRNA or miRNA bound. If the Ago protein loaded with ssRNA is Ago2, it harbors siRNA-directed target-mRNA cleavage activity (Meister et al., 2004b). Several active forms of this complex have been reported by different groups. Purification of active RISC has yielded complexes falling in a wide range of molecular sizes from 160 kDa to 80S (reviewed by Meister and Tuschl, 2004; Sontheimer, 2005). The size of the complex depends on the species and purification conditions, and it is not generally clear if the association of certain components is specific. The one common component for all purified active RISCs is an Ago protein. In human cultured cells, Agos 1 through 4 have been shown to co-purify with miRNAs/siRNAs. However, only Ago2 is associated with cleavage activity (Hammond et al., 2001; Liu et al., 2004; Meister et al., 2004b). And it was recently shown that human Ago2 can combine with a siRNA to form "minimal RISC" that accurately cleaves substrate RNAs (Rivas et al., 2005). This is consistent with previous purifications of minimal RISC, purified from human cell extracts using affinity-tagged siRNAs under high salt conditions with a size of ~160 kDa and contained Ago2 as the major protein component (Martinez et al., 2002). A human complex of about 550 kDa also associates with cleavage activity (Hutvágner and Zamore, 2002; Mourelatos et al., 2002). In *Drosophila* S2 cells fractionation, RISC activity is pelleted with polysomes. When released by high salt wash and separated by gel filtration chromatography, the active complex appears to be about 500 kDa (Hammond et al., 2001). More stringent purification yielded an active fraction mostly composed of Ago2 (Rand et al., 2004).

Drosophila embryo lysates yielded a RISC peaking in the ~200 kDa fraction. Recently, a complex was purified from *Drosophila* embryo lysates, known as holo-RISC with an apparent size of about 80S (Pham et al., 2004).

Analyses of the different purified RISCs by mass spectrometry or Western analysis have revealed some of the protein components. A mass spectrometric analysis of the minimal RISC confirmed the presence of Ago1 and Ago2 (Martinez et al., 2002). An RNP complex containing the putative helicase Gemin3 and its partner Gemin4 was co-immunoprecipitated with Ago2 (Mourelatos et al., 2002). It was shown later that Gemin3 and Gemin4 immunoprecipitate miRNA-guided mRNA cleavage activity (Hutvágner and Zamore, 2002). Mammalian Ago2 and Dicer were co-immunoprecipitated with fragile X mental retardation protein (FMRP) in a complex that also contained miRNAs (Jin et al., 2004). Western analysis showed that *Drosophila* holo-RISC includes proteins previously linked with RNA silencing, such as the *D. melanogaster* homolog of human FMRP, Fmr1; the vasa intronic gene (Vig) protein, and Tsn, a protein with multiple repeats of the staphylococcal nuclease domain (Pham et al., 2004). Fmr1 also co-purifies with Ago2 and components of the large ribosomal subunit and the putative RNA helicase Dmp68 (Ishizuka et al., 2002).

mRNA cleavage

When miRISC or siRISC encounters a complementary target it cleaves the phosphodiester backbone yielding products that bear a 3'hydroxyl and a 5' phosphate in a reaction that requires magnesium ions (Martinez et al., 2002; Schwarz et al., 2004). mRNA cleavage can occur in an ATP-independent manner; the siRNA remains intact, which allows RISC to function as a multiple turnover enzyme (Hutvágner and Zamore, 2002). ATP, however, may aid target release since its presence accelerates turnover (Haley and Zamore, 2004). The mRNA backbone cleavage occurs at a defined site: only the phosphodiester linkage that joins the residues paired to siRNA nucleotides 10 and 11 (counting from the 5' end of the siRNA) is cleaved (Elbashir et al., 2001b). Mismatches and nucleotide modifications between the guide RNA and its target can only be tolerated near the ends of the siRNA strand (Chiu and Rana, 2003; Martinez and Tuschl, 2004).

5' mRNA fragments generated by RISC cleavage are rapidly degraded from their 3' ends by the exosome (Ski-complex) that requires ongoing translation, whereas the 3' fragments are degraded from their 5' ends by XRN1. Hence, mRNAs are degraded from the ends at the site of RISC cleavage, without undergoing decapping or deadenylation (Orban and Izaurralde, 2005).

Regulatory function of miRNAs

Computational prediction of targets

Our initial understanding of miRNA function has been primarily based on the early findings of *let-7* and *lin-4* guided silencing: they bind to the 3' UTR of their target mRNAs with imperfect complementarity leading to translational repression (Figure 1). Thus, the general consensus is that miRNAs are gene regulators that base-pair to their target mRNAs. In plants, the complementarity of miRNAs to mRNAs is perfect or near perfect and regulation by miRNAs results in mRNA cleavage. In animals, most of the miRNAs are partially complementary to their targets and their regulation is thought to result in translational repression, although they also reduce the levels of their target transcripts (Lim et al., 2005).

lin-4 and *let-7* were identified by loss of function mutations that caused aberrant phenotypes in larval development. After acknowledging them as noncoding small RNAs, their target genes were identified based on function and sequence complementarity to *lin-4* and *let-7* in their 3' UTRs. Since the first collection of animal miRNA sequences were published, target identification has been the major challenge in the field, and researchers have tried to develop both experimental and computational approaches. From the bioinformatics point of view, miRNA:target interactions present a complex problem. First, complementarity between miRNA and their targets is not extensive and includes bulges, loops and mismatches in most cases. Second, the stretches of continuous base-pairs between miRNAs and their targets are too short to provide statistically significant complementarity within a genome. Third, the number of known miRNA:target interactions is very limited, which hinders the establishment of rules for miRNA target prediction. A few known features of miRNA:target interactions were defined to develop

algorithms for target predictions creating output data sets with a low estimate of false positives. All research groups that have published computational screens for animal targets used the same assumptions: (1) miRNA binding sites are located in the 3' UTRs of mRNA targets, which is true for all experimentally validated animal miRNA targets. (2) base-pairing to the target is asymmetrically weighted along the miRNA, in some cases a higher score is given to the 5' segment of the miRNA, and in others, perfect base-pairing for this region is required. It was experimentally determined that perfect base-pairing on the miRNA 5' end is crucial for the ability of a miRNA to repress a target, and that G:U wobble base-pairing disrupts this regulation (Brennecke et al., 2005; Doench and Sharp, 2004; Kloosterman et al., 2004; Vella et al., 2004a). (3) miRNA binding sites are conserved among orthologous genes of different species: the conservation criteria significantly reduces the predicted number of false positive targets (Enright et al., 2003; Lewis et al., 2003; Rajewsky and Socci, 2004; Stark et al., 2003).

A subset of the 3' UTRs predicted to be miRNA targets were assayed in vivo for their regulatory capacity. The Cohen laboratory used sensor transgenes that are GFP constructs fused to the 3' UTR of predicted targets in *Drosophila* (Stark et al., 2003). The sensor transgene is down-regulated in cells expressing a miRNA hairpin predicted to target the 3' UTR. In mammalian cells, 3' UTRs were assessed by dual luciferase assays (Kiriakidou et al., 2004; Lewis et al., 2003).

Experimental identification of miRNA targets

Concurrent with computational efforts, the function of a small number of miRNAs has been elucidated experimentally (Table 1 summarizes validated miRNA targets). Like the founding members, other miRNAs have been identified in genetic screens in *Drosophila* and *C. elegans*. *Bantam* mutants, identified in *Drosophila*, were shown to stimulate cell proliferation. The cloned *bantam* locus was found to encode a miRNA, thus, apoptotic genes were screened for *bantam* binding sites that led to identification of cell death inducer *hid* as a target (Brennecke et al., 2003). Similarly, miR-14 was found in a locus implied in cell death and fat storage in the fly in a screen that looked for reaper-induced apoptosis in the fly eye (Xu et al., 2003). Unlike *bantam*, miR-14 does not have a

validated target, even though evidence suggests that it could regulate the apoptotic inducer *Drice*.

Table 1. Biological function of animal miRNAs

organism	miRNA	target gene	biological function	reference
C. elegans	lin-4	*lin-14, lin-28*	developmental timing	(Moss et al., 1997; Olsen and Ambros, 1999; Wightman et al., 1993)
	let-7	*lin-14, lin-28, lin-41, daf-12, hbl-1, pha-4, lss-4, die-1, let-60*	developmental timing	(Abrahante et al., 2003; Johnson et al., 2005; Lin et al., 2003; Reinhart et al., 2000; Slack et al., 2000)
	lsy-6	*cog-1*	neuronal cell fate	(Johnston and Hobert, 2003)
	miR-273	*die-1*	neuronal cell fate	(Chang et al., 2004)
Drosophila	miR-14	unknown	apoptosis, fat storage	(Xu et al., 2003)
	bantam	*hid*	Apoptosis	(Brennecke et al., 2003)
M. musculus	miR-181	unknown	hematopoietic lineage differentiation	(Chen et al., 2004; Chen and Lodish, 2005)
	miR-223	unknown	hematopoietic lineage differentiation	(Chen et al., 2004)
	miR-142	unknown	hematopoietic lineage differentiation	(Chen et al., 2004)
	miR-375	*mtpn*	insulin secretion	(Poy et al., 2004)
	miR-196a	*HOXB8*	development	(Yekta et al., 2004)
H. sapiens	miR-143	unknown	adipocite diferentiation	(Esau et al., 2004)
	let-7	RAS	downregulated in lung cancer	(Johnson et al., 2005)

A screen for genes involved in asymmetric neuronal specification led to the identification of two *C. elegans* miRNAs. The asymmetric chemosensory neurons (ASE) play a role in sensory discrimination, ASE left and ASE right (ASEL and ASER), develop their asymmetry through differential gene expression. The miRNAs *lsy-6* and miR-273 were identified in a screen for mutants with either two ASEL or two ASER cells (Chang et al., 2004). *lsy-6* was found to down-regulate the *cog-1* transcription factor in ASEL, whereas *cog-1* is expressed in ASER. The transcription factor *die-1* has a role in the same gene

regulatory cascade: *die-1* promotes *lsy-6* expression in ASEL, while it is down-regulated by the ASER specific miR-273. Since both *lsy-6* and miR-273 are expressed only in a single cell in *C. elegans*, they had never been cloned. miR-273 was predicted in a computational study (Grad et al., 2003), but its expression could not be confirmed by Northern blotting analysis.

Other studies have pursued tissue-specific or tissue-enriched miRNAs. The Bartel laboratory identified miRNAs specifically expressed in hematopoietic cells (Chen et al., 2004). Over-expression of miR-181 in bone marrow hematopoietic progenitors increased the number of B cells in vitro and in vivo, and resulted in decreased levels of CD8+ T cells.

Poy and colleagues (Poy et al., 2004) cloned miRNAs from the murine pancreatic β-cell line MIN6. The most abundantly cloned miRNA, miR-375, was also specific for pancreatic β-cells and found to reduce insulin secretion when over-expressed. Aided by bioinformatic target predictions, V-1 myotrophin and *Vti1a*, a yeast t-SNARE homolog were identified as mRNA targets. Both of these targets were implicated in secretion of neurotransmitters, but until now the function of these proteins in pancreatic islets and insulin secretion had not been studied before.

The location of a miRNA in the genome can also provide information of its function. The conserved genomic location of mir-10 in the Hox gene cluster suggested that it could regulate Hox genes or genes involved in development (discussed in Chapter II). The colinearity of temporal and spatial Hox gene expression with their genomic location is also expected to be true for other miRNAs encoded within the cluster. In addition to miR-10 variants, miR-196 is encoded in the Hox gene cluster. Mammals have four Hox gene clusters (HOX A to D). All of the clusters encode a mir-196 variant, and many Hox genes have miR-196 complementary elements in their 3'UTRs (Yekta et al., 2004). And *HOXB8* has a near complete sequence complementarity to miR-196 (except for one G:U wobble) and targets its cleavage. In fact, miRNA-guided mRNA cleavage is common in plants in vivo (Llave et al., 2002), but miR-196 presents the only known example of an animal miRNA to cleave its target. As expected by the high evolutionary conservation of the hox cluster, complementarity of miR-196a to *HOXB8* is also conserved in fish and frog.

Other small regulatory RNAs and related RNAi mechanisms

In addition to gene silencing by target mRNA cleavage and translational repression, the RNAi machinery is also involved in regulation of gene expression at the transcriptional level. dsRNA and proteins of the RNAi machinery can direct epigenetic alterations to homologous DNA sequences to induce transcriptional gene silencing or even DNA elimination in some organisms (reviewed by Matzke and Birchler, 2005). Effector complexes involved in these mechanisms are guided by small RNAs to complementary genomic regions generally within heterochromatin (Pal-Bhadra et al., 2004; Verdel et al., 2004; Volpe et al., 2002)

Heterochromatin contains degenerate transposon sequences and tandem arrays of repeat sequences. Regions of heterochromatin are often associated with centromeres and are low in gene expression. Silencing of chromatin requires that histone 3 is deacetylated and then methylated on lysine 9. Methylated histone 3 on lysine 9 (H3K9) binds the heterochromatin protein HP1 through its chromodomain (Hall et al., 2002; Pal-Bhadra et al., 2004). In *Drosophila* this methylation is mediated by SU(VAR)3-9, and a local high concentration of these two proteins is characteristic for heterochromatin (Schotta et al., 2003).

The first direct link between the RNAi machinery and heterochromatin formation was made in the fission yeast *Schizosaccharomyces pombe*. Deletion of Dicer (Dcr1), Ago1 or RNA-dependent RNA polymerase, Rdp1 (present in some organisms and with a function in RNAi signal amplification) leads to loss of gene silencing in centromeric repeats. RNA from both strands of these repeats accumulated in these mutants (Volpe et al., 2002). The silencing complex in this pathway was later defined as RNA-induced initiation of transcriptional gene silencing (RITS) complex (Verdel et al., 2004). It contains siRNAs originating from heterochromatin regions, Ago1, Chp1 (a centromere-associated chromodomain protein), and Tas3 (a serine-rich protein specific to *S. pombe*) (Verdel et al., 2004). RITS interacts with the RNA-dependent RNA polymerase complex (RDRC) containing the helicase Hrr1, the RNA-dependent RNA polymerase Rdp1 and Cid12, a member of the poly(A) polymerase family. The interaction between these two complexes is dependent on Dcr1 and the histone methyltransferase Crl4. The association of RDRC

to RITS may be necessary for regulating the RNA synthesis by RDRC in vivo, restricting this activity to transcripts derived from heterochromatic regions (Motamedi et al., 2004). RITS is tethered to loci with H3K9 methylation, and this interaction is essential for silencing and the production of siRNAs (Noma et al., 2004). Furthermore, de novo heterochromatin formation can be triggered in *S. pombe* by transcription of an artificial hairpin homologous to sequences present throughout the genome. This silencing is associated to the accumulation of Swi6 and Crl4 (yeast homologues of HP1 and SU(VAR)3-9, respectively) and dependent on the RNAi machinery (Schramke and Allshire, 2003).

It was recently shown in *Drosophila* that loss of function mutations of the Ago genes *piwi* and *aubergine* as well as the RNA helicase *spindle-E*, suppressed the silencing that occurs when a normally euchromatic gene is juxtaposed to a heterochromatic domain (Pal-Bhadra et al., 2004). These mutations also resulted in delocalization of heterochromatic proteins and reduction of H3K9 methylation. Our laboratory cloned *Drosophila* rasiRNAs, which are derived from transposons and heterchromatic regions of the genome, suggesting these small RNAs might be involved in the silencing of these genetic elements (Chapter III).

Recent evidence suggests that similar processes occur in vertebrates. Knockout of Dicer caused chromosome segregation defects and aberrant accumulation of transcripts from α-satellite sequences, which are derived from human centromeric repeats (Fukagawa et al., 2004; Kanellopoulou et al., 2005).

Another epigenetic modification linked to the RNAi machinery is RNA-dependent DNA methylation (RdDM). RdDM has been thoroughly studied in plants where dsRNAs, which contain sequences homologous to promoter regions, can trigger promoter methylation and transcriptional silencing. The process requires siRNAs (Mette et al., 2000) and proteins of the RNAi pathway. In Arabidopsis, a mutant allele for AGO4 activated silent alleles of the SUP gene and decreased DNA and histone methylation while accumulating of siRNAs from retroelements (Zilberman et al., 2003). RNAi components, including AGO4 and DCL3, are necessary for de novo DNA methylation of tandem repeats in the promoter of a transgene in *A. thaliana* (Chan et al., 2004). DCL3 is one of the four plant Dicers, which is localized in the nucleus and produces 24 nt RNAs

that have been implicated in chromatin modifications of repetitive sequences (Zilberman et al., 2003).

There is some evidence that RdDM also occurs in mammals: siRNAs were shown to target promoters of endogenous genes and induce GC methylation, transcriptional silencing and H3K9 methylation in human cells (Kawasaki and Taira, 2004; Morris et al., 2004).

Rationale

After addition of 500 base-pair long dsRNA to *Drosophila* embryonic extracts, it was observed that the RNA was processed into duplexes of 21 and 22 nt, which were shown to mediate RNAi. To characterize the products of the dsRNA processing, RNA was isolated from these extracts and small RNAs were cloned. In addition to dsRNA cleavage products derived from the exogenous dsRNA, endogenous small RNAs were identified.

The aim this study was to clone dsRNA-specific ribonuclease cleavage products from *Drosophila*, mouse and human and to identify small RNAs with a role in RNAi or RNAi-related phenomena. The sequence analysis of the clones resulted in the identification more than 100 new genes defining a large family of translational regulators similar to *lin-4* and *let-7* genes previously described in *C. elegans*. These genes were collectively named microRNAs. Their expression profiles were characterized as well as their genomic location and conservation in other organisms.

In the course of this study most of my results were published. Other members of the laboratory and collaborators contributed to this work. Reprints of the publications are included and constitute the first three chapters of this thesis. In the fourth chapter, I describe my work on the optimization of the small RNA cloning protocol.

Chapter I: Identification of tissue specific microRNAs from mouse

Mariana Lagos-Quintana, Reinhard Rauhut, Abdulah Yalcin, Jutta Meyer, Winfried Lendeckel, Thomas Tuschl, **Identification of tissue specific microRNAs from mouse**, *Current Biology*., 2002, 12, 735-739.

Statement of contribution

Small RNA cloning

Bioinformatic analysis

Writing

Current Biology, Vol. 12, 735–739, April 30, 2002, ©2002 Elsevier Science Ltd. All rights reserved. PII S0960-9822(02)00809-6

Identification of Tissue-Specific MicroRNAs from Mouse

Mariana Lagos-Quintana, Reinhard Rauhut, Abdullah Yalcin, Jutta Meyer, Winfried Lendeckel, and Thomas Tuschl[1]
Department of Cellular Biochemistry
Max-Planck-Institute for Biophysical Chemistry
Am Fassberg 11
D-37077 Göttingen
Germany

Summary

MicroRNAs (miRNAs) are a new class of noncoding RNAs, which are encoded as short inverted repeats in the genomes of invertebrates and vertebrates [1, 2]. It is believed that miRNAs are modulators of target mRNA translation and stability, although most target mRNAs remain to be identified. Here we describe the identification of 34 novel miRNAs by tissue-specific cloning of approximately 21-nucleotide RNAs from mouse. Almost all identified miRNAs are conserved in the human genome and are also frequently found in nonmammalian vertebrate genomes, such as pufferfish. In heart, liver, or brain, it is found that a single, tissue-specifically expressed miRNA dominates the population of expressed miRNAs and suggests a role for these miRNAs in tissue specification or cell lineage decisions. Finally, a miRNA was identified that appears to be the fruitfly and mammalian ortholog of *C. elegans* lin-4 stRNA.

Results and Discussion

MicroRNAs (miRNAs) represent a new class of gene products, which are believed to sequence-specifically control translation of target mRNAs by binding to sites of antisense complementarity in 3′ untranslated regions (UTRs) [1–5]. Several miRNAs, such as let-7 RNA, miR-1, miR-34, miR-60, and miR-87, are highly conserved between invertebrates and vertebrates, implicating that they may recognize multiple sites and/or multiple targets of presumably conserved function [3–6]. The small temporal RNAs (stRNAs) lin-4 and let-7 represent a subclass of miRNAs identified by genetic analysis in *Caenorhabditis elegans*, which are developmentally regulated and themselves control developmental programs, such as timing of neuronal rewiring, Dauer larva formation, vulva formation, and the terminal differentiation of hypodermal cells [7–11].

miRNAs are typically excised from 60- to 70-nucleotide foldback RNA precursor structures, which are sometimes detected at the onset of miRNA precursor expression [12–14] or during expression of very abundant miRNAs [3–5]. Generally, only one of the strands of the hairpin precursor molecule is excised and accumulates, presumably because it is protected by associated proteins from RNA degradation. These putative proteins may as well mediate the translational suppression. The miRNA precursor processing reaction requires Dicer RNase III and Argonaute family members [12–14]; Dicer and Argonaute proteins are also involved in RNAi [15–18].

To gain insights into the distribution and function of miRNAs in mammals, we investigated the tissue-specific distribution of miRNAs in adult mouse. Cloning of miRNAs from specific tissues was preferred over whole organism-based cloning because low-abundance miRNAs that normally go undetected by Northern blot analysis are identified clonally. Also, in situ hybridization techniques for detecting 21-nt RNAs have not yet been developed. Therefore, 19- to 25-nucleotide RNAs were cloned and sequenced from total RNA, which was isolated from 18.5-week-old BL6 mice (see the Supplementary Material available with this article online). Because RNA was prepared from combining tissues of several mice, minor sequence variations that were detected multiple times in multiple clones may reflect polymorphisms rather than RT/PCR mutations. Public database searching was used to identify the genomic sequences encoding the ~21-nt RNAs. The occurrence of a 20 to 30 basepair foldback structure involving the immediate upstream or downstream flanking sequences was used to assign miRNAs [1, 3–5].

We examined nine different mouse tissues and identified 34 novel miRNAs, some of which are highly tissue-specifically expressed (Table 1 and Figure 1). miR-1 was previously shown by Northern analysis to be strongly expressed in human adult heart but not in human brain, liver, kidney, lung, or colon [5]. Here we show that miR-1 accounts for 45% of all mouse miRNAs found in heart, yet miR-1 was still expressed at a low level in liver and midbrain, even though it remained undetectable by Northern analysis. Three copies or polymorphic alleles of miR-1 were found in mice. The conservation of tissue-specific miR-1 expression between mouse and human provides additional evidence for a conserved regulatory role of this miRNA. In liver, variants of miR-122 account for 72% of all cloned miRNAs, and miR-122 was undetected in all other tissues analyzed. In spleen, miR-143 appeared to be most abundant, at a frequency of ~30%. In colon, miR-142-as was cloned several times and also appeared at a frequency of 30%. In small intestine, too few miRNA sequences were obtained to permit statistical analysis. This was due to strong RNase activity in this tissue, which caused significant breakdown of abundant noncoding RNAs, e.g., rRNA, so that the fraction of miRNA in the cloned sequences was very low. For the same reason, no miRNA sequences were obtained from pancreas.

To gain insights in neural tissue miRNA distribution, we analyzed cortex, cerebellum, and midbrain. Similar to heart, liver, and small intestine, variants of a particular miRNA, miR-124, dominated and accounted for 25% to 48% of all brain miRNAs. miR-101, -127, -128, -131, and -132, also cloned from brain tissues, were further

[1]Correspondence: ttuschl@mpibpc.gwdg.de

Table 1. Mouse miRNA Sequences Identified by Cloning from Distinct Mouse Tissues

		Number of Clones[a]							
miRNA	Sequence (5′ to 3′)	ht	lv	sp	si	co	cx	cb	mb
let-7a	UGAGGUAGUAGGUUGUAUAGUU		3			1	1		7
let-7b	UGAGGUAGUAGGUUGUGUGGUU		1	1				2	5
let-7c	UGAGGUAGUAGGUUGUAUGGUU		2				2	5	19
let-7d	AGAGGUAGUAGGUUGCAUAGU	2				2	2		2
let-7e	UGAGGUAGGAGGUUGUAUAGU			1					2
let-7f	UGAGGUAGUAGAUUGUAUAGUU			2				3	3
let-7g	UGAGGUAGUAGUUUGUACAGUA						1	1	2
let-7h	UGAGGUAGUAGUGUGUACAGUU						1	1	
let-7i	UGAGGUAGUAGUUUGUGCU						1	1	
miR-1b	UGGAAUGUAAAGAAGUAUGUAA	4	2						1
miR-1c	UGGAAUGUAAAGAAGUAUGUAC	7							
miR-1d	UGGAAUGUAAAGAAGUAUGUAUU	16							1
miR-9	UCUUUGGUUAUCUAGCUGUAUGA						3	4	4
miR-15a	UAGCAGCACAUAAUGGUUUGUG	1							2
miR-15b	UAGCAGCACAUCAUGGUUUACA	1							
miR-16	UAGCAGCACGUAAAUAUUGGCG	1			1	2	1	2	3
miR-18	UAAGGUGCAUCUAGUGCAGAUA			1					
miR-19b	UGUGCAAAUCCAUGCAAAACUGA			1					
miR-20	UAAAGUGCUUAUAGUGCAGGUAG					1			
miR-21	UAGCUUAUCAGACUGAUGUUGA	1		1	2	1			
miR-22	AAGCUGCCAGUUGAAGAACUGU	2	1		1			1	2
miR-23a	AUCACAUUGCCAGGGAUUUCC	1							
miR-23b	AUCACAUUGCCAGGGAUUACCAC						1		
miR-24	UGGCUCAGUUCAGCAGGAACAG	1				1	1		1
miR-26a	UUCAAGUAAUCCAGGAUAGGCU							3	2
miR-26b	UUCAAGUAAUUCAGGAUAGGUU		2				4	1	
miR-27a	UUCACAGUGGCUAAGUUCCGCU	1		2		1	1	2	1
miR-27b	UUCACAGUGGCUAAGUUCUG								1
miR-29a	CUAGCACCAUCUGAAAUCGGUU	1				1		1	
miR-29b/miR-102	UAGCACCAUUUGAAAUCAGUGUU	1				1	5		3
miR-29c/	UAGCACCAUUUGAAAUCGGUUA	1					3		1
miR-30a-s/miR-97	UGUAAACAUCCUCGACUGGAAGC			1			1		1
miR-30a-as[b]	CUUUCAGUCGGAUGUUUGCAGC							1	
miR-30b	UGUAAACAUCCUACACUCAGC			1				2	
miR-30c	UGUAAACAUCCUACACUCUCAGC	2					1	1	
miR-30d	UGUAAACAUCCCCGACUGGAAG		1						
miR-99a/miR-99	ACCCGUAGAUCCGAUCUUGU						1		
miR-99b	CACCCGUAGAACCGACCUUGCG							1	
miR-101	UACAGUACUGUGAUAACUGA						2	1	1
miR-122a	UGGAGUGUGACAAUGGUGUUUGU		3						
miR-122b	UGGAGUGUGACAAUGGUGUUUGA		11						
miR-122a,b	UGGAGUGUGACAAUGGUGUUUG		23						
miR-123	CAUUAUUACUUUUGGUACGCG	1	2						
miR-124a[c]	UUAAGGCACGCGG-UGAAUGCCA				1		37	41	24
miR-124b	UUAAGGCACGCGGGUGAAUGC						1	3	
miR-125a	UCCCUGAGACCCUUUAACCUGUG						1	1	
miR-125b	UCCCUGAGACCCU--AACUUGUGA						1		
miR-126	UCGUACCGUGAGUAAUAAUGC	4						1	
miR-127	UCGGAUCCGUCUGAGCUUGGCU							1	
miR-128	UCACAGUGAACCGGUCUCUUUU						2	2	2
miR-129	CUUUUUUCGGUCUGGGCUUGC							1	
miR-130	CAGUGCAAUGUUAAAAGGGC							1	
miR-131	UAAAGCUAGAUAACCGAAAGU						1	1	1
miR-132	UAACAGUCUACAGCCAUGGUCGU							1	
miR-133	UUGGUCCCCUUCAACCAGCUGU	4					1		
miR-134	UGUGACUGGUUGACCAGAGGGA						1		
miR-135	UAUGGCUUUUUAUUCCUAUGUGAA						1		
miR-136	ACUCCAUUUGUUUUGAUGAUGGA						1		
miR-137	UAUUGCUUAAGAAUACGCGUAG						1		1
miR-138	AGCUGGUGUUGUGAAUC						1		
miR-139	UCUACAGUGCACGUGUCU					1	1		
miR-140	AGUGGUUUUACCCUAUGGUAG				1				
miR-141	AACACUGUCUGGUAAAGAUGG			1	1		1		
miR-142-s	CAUAAAGUAGAAAGCACUAC				1	1			
miR-142-as[c]	UGUAGUGUUUCCUACUUUAUGG			1	1	6			
miR-143	UGAGAUGAAGCACUGUAGCUCA	3		7			2		1

(*continued*)

Table 1. Continued

miRNA	Sequence (5′ to 3′)	Number of Clones[a]							
		ht	lv	sp	si	co	cx	cb	mb
miR-144	UACAGUAUAGAUGAUGUACUAG	2				1			
miR-145	GUCCAGUUUUCCCAGGAAUCCCUU	1							
miR-146	UGAGAACUGAAUUCCAUGGGUUU	1							
miR-147	GUGUGUGGAAAUGCUUCUGCC			1					
miR-148	UCAGUGCACUACAGAACUUUGU			1					
miR-149	UCUGGCUCCGUGUCUUCACUCC	1							
miR-150	UCUCCCAACCCUUGUACCAGUGU					1			
miR-151	CUAGACUGAGGCUCCUUGAGGU					1			
miR-152	UCAGUGCAUGACAGAACUUGG					1			
miR-153	UUGCAUAGUCACAAAAGUGA								1
miR-154	UAGGUUAUCCGUGUUGCCUUCG								1
miR-155	UUAAUGCUAAUUGUGAUAGGGG					1			

The sequences indicated represent the longest miRNA sequences identified by cloning. The 3′ terminus of miRNAs is often truncated by one or two nucleotides. miRNAs that are more than 85% identical in sequence (i.e., share 18 out of 21 nucleotides) or contain 1- or 2-nucleotide internal deletions are referred to by the same gene number followed by a lowercase letter. Minor sequence variations between related miRNAs are generally found near the ends of the miRNA sequence and are thought to not compromise target RNA recognition. Minor sequence variations may also represent A to G and C to U changes, which are accommodated as G-U wobble base pairs during target recognition. miRNAs with the suffix -s or -as indicate RNAs derived from either the 5′ half or the 3′ half of a miRNA precursor. Mouse brains were dissected into midbrain (mb), cortex (cx), and cerebellum (cb). The tissues analyzed were heart (ht), liver (lv), small intestine (si), colon (co), cortex (ct), cerebellum (cb), and midbrain (mb).

[a] The total number of clones, including breakdown products of noncoding RNAs and yet to be identified sequences, is listed in the Supplementary Material.

[b] The originally described miR-30 [3] was renamed to miR-30a-as in order to distinguish it from the miRNA derived from the opposite strand of the precursor encoded by the *mir-30a* gene. miR-30a-s is equivalent to miR-97 [22].

[c] A 1-nt length heterogeneity is found on both the 5′ and 3′ end. The 22-nt miR sequence is shown, but only 21-nt miRNAs were cloned.

analyzed by Northern blotting and shown to be predominantly brain specific (see Supplementary Material). Members of another class of noncoding RNAs, C/D-box small nucleolar RNAs (snoRNAs) and H/ACA-box snoRNA, in mouse and human have also shown brain-specific expression patterns [23].

miR-125a and miR-125b are very similar to the sequence of *C. elegans* lin-4 stRNA and may represent its orthologs (Figure 2A). This is of great interest because, unlike let-7 that was readily detected in other species, lin-4 has acquired a few mutations in the central region and thus escaped bioinformatic database searches. Using the mouse sequence miR-125b, we could readily identify its ortholog in the *D. melanogaster* genome (see Supplementary Material). miR-125a and miR-125b differ only by a central diuridine insertion and a U to C change. miR-125b is very similar to lin-4 stRNA with the differences located only in the central region, which is presumed to be bulged out during target mRNA recognition [11]. miR-125a and miR-125b were cloned from brain tissue, but expression was also detected by Northern analysis in other tissues, consistent with the role for lin-4 in regulating neuronal remodeling by controlling *lin-14* expression [19]. Unfortunately, orthologs to *C. elegans lin-14* have not been described, and miR-125 targets remain to be identified in *D. melanogaster* or mammals. Finally, miR-125b expression is also developmentally regulated and only detectable in pupae and adult but not in embryo or larvae of *D. melanogaster* (Figure 2B).

Sequence comparison of mouse miRNAs with previously described miRNA reveals that miR-99 variants are similar to *D. melanogaster*, mouse, and human miR-10 [3] as well as *C. elegans* miR-51 [4]; miR-141 is similar to *D. melanogaster* miR-8 [3]; miR-29 variants are similar to *C. elegans* miR-83 [4]; and miR-131 and miR-142-s are similar to *D. melanogaster* miR-4 [3] and *C. elegans* miR-79 [4]. miR-124a is conserved between invertebrates and vertebrates. In this respect, it should be noted that almost every miRNA cloned from mouse was also encoded in the human genome and frequently detected in other vertebrates, such as the pufferfish, *Fugu rubripes*, and the zebrafish, *Danio rerio*. Sequence conservation may point to conservation in function of these miRNAs. Comprehensive information about orthologous sequences is listed in Table S2 in the Supplementary Material).

In two cases, both strands of miRNA precursors were cloned (Table 1), which was previously observed once for a *C. elegans* miRNA [4]. It is thought that the most frequently cloned strand of a miRNA precursor represents the functional miRNA, which is miR-30c-s and miR-142-as, "s" and "as" indicating the 5′ or 3′ side of the foldback structure, respectively.

The *mir-142* gene is located on chromosome 17 but was also found at the breakpoint junction of a t(8;17) translocation, which causes an aggressive B cell leukemia due to strong upregulation of a translocated *MYC* gene [20]. The translocated *MYC* gene, which was also truncated at the first exon, was located only 4-nt downstream of the 3′ end of the miR-142 precursor. This suggests that translocated *MYC* was under the control of the upstream *mir-142* promoter. Alignment of mouse and human miR-142 containing EST sequences indicate an ~20 nt conserved sequence element downstream of the *mir-142* hairpin. This element was lost in the translocation. It is conceivable that the absence of the conserved downstream sequence element in the putative miR-142/mRNA fusion prevented the recognition of the

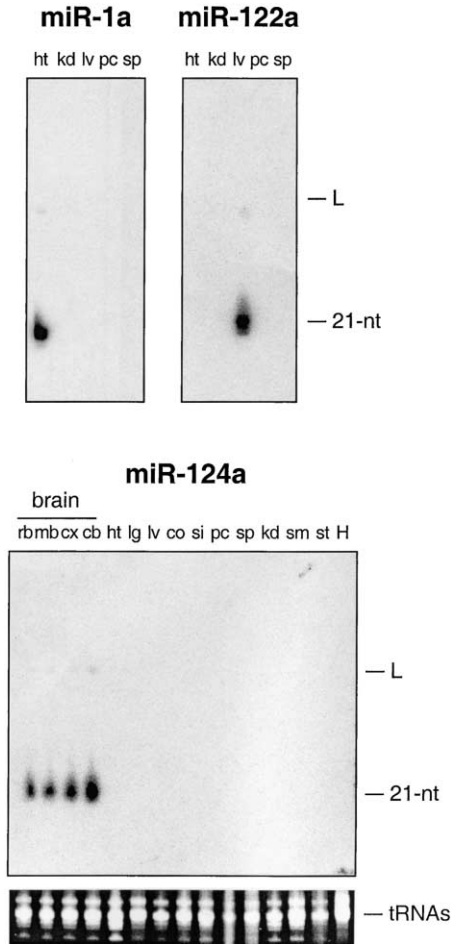


Figure 1. Northern Blot Analysis of Tissue-Specific miRNAs

Total RNA from different mouse tissues was blotted and probed with a 5′-radiolabeled oligodeoxynucleotide complementary to the indicated miRNA. Equal loading of total RNA on the gel was verified by ethidium bromide staining prior to transfer; the band representing tRNAs is shown. The foldback precursors are indicated with capital "L." Mouse brains were dissected into midbrain (mb), cortex (cx), and cerebellum (cb). The rest of the brain (rb) was also used. Other tissues were heart (ht), lung (lg), liver (lv), colon (co), small intestine (si), pancreas (pc), spleen (sp), kidney (kd), skeletal muscle (sm), stomach (st); H, human HeLa SS3 cells.

transcript as a miRNA precursor and therefore may have caused accumulation of fusion transcripts and overexpression of *MYC*.

miR-155, which was cloned from colon, is excised from the known noncoding BIC RNA [21]. *BIC* was originally identified as a gene transcriptionally activated by promoter insertion at a common retroviral integration site in B cell lymphomas induced by avian leukosis virus. Comparison of *BIC* cDNAs from human, mouse, and chicken revealed 78% identity over 138 nucleotides [21]. The identity region covers the miR-155 foldback precursor and a few conserved boxes downstream of the foldback sequence. The relatively high level of expression of *BIC* in lymphoid organs and cells in human, mouse, and chicken implies an evolutionary conserved function, but BIC RNA has also been detected at low levels in nonhematopoietic tissues [21].

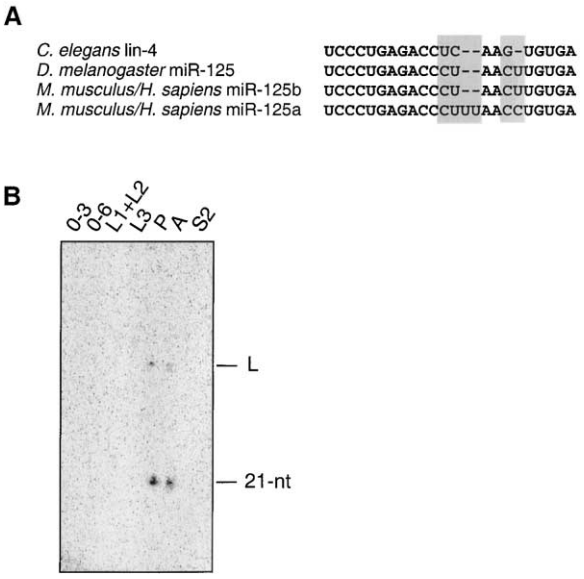


Figure 2. Potential Orthologs of lin-4 stRNA

(A) Sequence alignment of *C. elegans* lin-4 stRNA with mouse miR-125a and miR-125b and the *D. melanogaster* miR-125. Differences are highlighted by gray boxes.

(B) Northern blot of total RNA isolated from staged populations of *D. melanogaster*, probed for miR-125. E, embryo; L, larval stage; P, pupae; A, adult; S2, Schneider-2 cells.

Another interesting observation was that segments of perfect complementarity to miRNAs are not observed in mRNA sequences or in genomic sequences outside the miRNA inverted repeat. Although this could be fortuitous, based on the link between RNAi and miRNA processing [12–14], it may be speculated that miRNAs retain the potential to cleave perfectly complementary target RNAs. Because translational control without target degradation could provide more flexibility, it may be preferred over mRNA degradation.

In summary, 34 novel miRNAs were identified from mouse, which are conserved in human and often also in other nonmammalian vertebrates. A few of these miRNAs appear to be extremely tissue specific, suggesting a critical role for some miRNAs in tissue specification and cell lineage decisions. We may have also identified the fruitfly and mammalian ortholog of *C. elegans* lin-4 stRNA. The establishment of a comprehensive list of miRNA sequences will be instrumental for bioinformatic approaches that make use of completed genomes and the power of phylogenetic comparison in order to identify miRNA-regulated target mRNAs.

Supplementary Material

Supplementary Material including additional methodological details, figures, and tables can be found online at http://images.cellpress.com/supmat/supmatin.htm.

Acknowledgments

We are very grateful to S.M. Elbashir and C. Karschin for preparation of mouse tissues and sectioning of mouse brains; G. Dowe, G. Heyne, and M. Killian for sequencing; J. Aach for bioinformatic assistance; and R. Lührmann for support. This work was funded by a

Bundesministerium für Bildung und Forschung (BMBF) Biofuture grant, number 0311856.

Received: February 19, 2002
Revised: March 6, 2002
Accepted: March 6, 2002
Published: April 30, 2002

References

1. Ambros, V. (2001). microRNAs: tiny regulators with great potential. Cell *107*, 823–826.
2. Moss, E.G. (2002). MicroRNAs: hidden in the genome. Curr. Biol. *12*, R138–R140.
3. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–858.
4. 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.
5. Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. Science *294*, 862–864.
6. Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.D., Degnan, B., Müller, P., et al. (2000). Conservation of the sequence of *let-7* heterochronic regulatory RNA. Nature *408*, 86–89.
7. Ambros, V., and Horvitz, H.R. (1984). Heterochronic mutants of the nematode Caenorhabditis elegans. Science *226*, 409–416.
8. 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.
9. 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.
10. 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.
11. Rougvie, A.E. (2001). Control of developmental timing in animals. Nat. Rev. Genet. *2*, 690–701.
12. Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. Cell *106*, 23–34.
13. Hutvágner, G., McLachlan, J., Bálint, É., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA interference enzyme dicer in small temporal RNA maturation. Science *93*, 834–838.
14. Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. *15*, 2654–2659.
15. Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 363–366.
16. Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146–1150.
17. Billy, E., Brondani, V., Zhang, H., Muller, U., and Filipowicz, W. (2001). Specific interference with gene expression induced by long, double-stranded RNA in mouse embryonal teratocarcinoma cell lines. Proc. Natl. Acad. Sci. USA *98*, 14428–14433.
18. Knight, S.W., and Bass, B.L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in C. elegans. Science *2*, 2.
19. Hallam, S.J., and Jin, Y. (1998). *lin-14* regulates the timing of synaptic remodelling in Caenorhabditis elegans. Nature *395*, 78–82.
20. Gauwerky, C.E., Huebner, K., Isobe, M., Nowell, P.C., and Croce, C.M. (1989). Activation of *MYC* in a masked t(8;17) translocation results in an aggressive B-cell leukemia. Proc. Natl. Acad. Sci. USA *86*, 8867–8871.
21. Tam, W. (2001). Identification and characterization of human BIC, a gene on chromosome 21 that encodes a noncoding RNA. Gene *274*, 157–167.
22. Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. *16*, 720–728.
23. Cavaille, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J., and Hüttenhofer, A. (2000). Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. Proc. Natl. Acad. Sci. USA *97*, 14311–14316.
24. Lai, E.C. (2002). MicroRNAs are complementary to 3′ UTR sequence motifs that mediate negative post-transcriptional regulation. Nat. Genet. *30*, 363–364.

Note Added in Proof

It was recently noted that the 5′ ends of a subset of *Drosophila* microRNAs are perfectly complementary to 3′ UTR sequence motifs that mediate negative posttranscriptional regulation [24].

Chapter II: New microRNAs from mouse and human

Mariana Lagos-Quintana, Reinhard Rauhut, Jutta Meyer, Arndt Borkhardt, Thomas Tuschl. **New microRNAs from mouse and human**. *RNA*. 2003 Feb;9(2):175-9.

Statement of contribution

Small RNA cloning

Bioinformatic analysis

Writing

New microRNAs from mouse and human

MARIANA LAGOS-QUINTANA,[1] REINHARD RAUHUT,[1] JUTTA MEYER,[1] ARNDT BORKHARDT,[2] and THOMAS TUSCHL[1]

[1]Department of Cellular Biochemistry, Max-Planck-Institute for Biophysical Chemistry, D-37077 Göttingen, Germany
[2]Children's University Hospital, Pediatric Hematology & Oncology, D-35392 Gießen, Germany

ABSTRACT

MicroRNAs (miRNAs) represent a new class of noncoding RNAs encoded in the genomes of plants, invertebrates, and vertebrates. MicroRNAs regulate translation and stability of target mRNAs based on (partial) sequence complementarity. Although the number of newly identified miRNAs is still increasing, target mRNAs of animal miRNAs remain to be identified. Here we describe 31 novel miRNAs that were identified by cloning from mouse tissues and the human Saos-2 cell line. Fifty-three percent of all known mouse and human miRNAs have homologs in *Fugu rubripes* (pufferfish) or *Danio rerio* (zebrafish), of which almost half also have a homolog in *Caenorhabditis elegans* or *Drosophila melanogaster*. Because of the recurring identification of already known miRNAs and the unavoidable background of ribosomal RNA breakdown products, it is believed that not many more miRNAs may be identified by cloning. A comprehensive collection of miRNAs is important for assisting bioinformatics target mRNA identification and comprehensive genome annotation.

Keywords: microRNA; hairpin RNA

INTRODUCTION

MicroRNAs (miRNAs) represent a class of noncoding RNAs encoded in the genomes of plants and invertebrate and vertebrate animals. Mature miRNAs are ~21 nt long and excised from 60- to 80-nt double-stranded RNA foldbacks (dsRNA hairpins) by Dicer RNase III (Grishok et al. 2001; Hutvágner et al. 2001; Ketting et al. 2001; Park et al. 2002; Reinhart et al. 2002). The processing reaction is generally asymmetric, and only one of the strands from the hairpin is accumulating and is referred to as the mature miRNA (Lee et al. 1993; Reinhart et al. 2000; Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001). The hairpins themselves are excised from longer primary transcripts, some of which have intron/exon pre-mRNA structure, whereby the miRNA precursors are located in either exon or intron sequence (Lagos-Quintana et al. 2001, 2002; Lau et al. 2001; Lee et al. 2002). Clusters of miRNAs are also expressed as long primary transcripts (Lee et al. 2002). Trimming of the primary transcripts to the miRNA precursor hairpin takes place in the nucleus (Lee et al. 2002). Subsequently, the hairpin is exported to the cytoplasm, and Dicer excises the mature miRNA (Lee et al. 2002).

Reprint requests to: Thomas Tuschl, Department of Cellular Biochemistry, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany; e-mail: ttuschl@gwdg.de; fax: 49-551-201-1197.

Article and publication are at http://www.rnajournal.org/cgi/doi/10.1261/rna.2146903.

Our understanding of miRNA function originates from studies of the developmentally regulated miRNAs lin-4 and let-7 in *Caenorhabditis elegans* (for reviews, see Slack and Ruvkun 1997; Ambros 2000; Rougvie 2001; Pasquinelli and Ruvkun 2002). Biochemical characterization of the lin-4 miRNA of *C. elegans* revealed that miRNAs are able to block protein synthesis after transcription initiation or possibly regulate protein stability (Wightman et al. 1993; Olsen and Ambros 1999; Seggerson et al. 2002). The lin-4 miRNA recognizes complementary sequences within the 3′-UTR of its targets by forming an imperfect, bulged RNA duplex structure. The evolutionarily conserved let-7 miRNA of *C. elegans* also regulates the expression of downstream genes based on partial sequence complementarity between the miRNA and the target 3′-UTR (Reinhart et al. 2000). The targets for lin-4 and let-7 miRNAs were first identified genetically, and only after visual inspection of the sequences was the partial sequence complementarity detected. Intriguingly, the 5′-ends of some *Drosophila melanogaster* miRNAs were found to be complementary to 3′-UTR sequence motifs already known to mediate negative post-transcriptional regulation (Lai 2002). Furthermore, single nucleotide polymorphisms in the 3′-UTR of certain mRNAs have been associated with increased risk for certain types of cancer, indicating possible interference with miRNA regulation (Conne et al. 2000; Hayward 2000; Mendell and Dietz 2001; Kiyohara et al. 2002).

In contrast to animal miRNAs, almost all identified plant miRNAs show near-perfect complementarity to subsets of

mRNAs that are the likely targets for these miRNAs (Rhoades et al. 2002). *Arabidopsis thaliana* miR-171, which is also conserved in *Nicotiana benthamiana* and rice, was shown experimentally to specifically target *SCL6-III* and *SCL-IV* mRNAs for degradation (Llave et al. 2002b), indicating that plant miRNAs can act as small interfering RNAs (siRNAs) that guide target RNA cleavage (Elbashir et al. 2001) rather than mediating translational control. At the same time, it was observed that endogenous human let-7 miRNA is able to guide target RNA cleavage as long as a target RNA was introduced that carried a fully complementary segment of let-7 miRNA (Hutvágner and Zamore 2002). Argonaute proteins that are associated with single-stranded siRNAs in the target RNA-cleaving RNA-induced silencing complex (RISC; Hammond et al. 2001; Martinez et al. 2002) were found to be associated with mature miRNAs (Mourelatos et al. 2002; Schwarz et al. 2002). Together, these data indicate that miRNAs and siRNAs are present in similar RNP complexes, but that their function depends on the specific base-pairing structure formed between the small RNA and the target mRNA. Prediction of the targets of animal miRNAs remains difficult because these miRNAs show no more matches to mRNAs than random sequences, so that it must be assumed that few, if any, of the miRNAs will recognize their respective targets with near-perfect complementarity (Rhoades et al. 2002).

To date, nearly 200 miRNAs have been described from *C. elegans*, *D. melanogaster*, human, mouse, and *A. thaliana* (Lagos-Quintana et al. 2001, 2002; Lau et al. 2001; Lee and Ambros 2001; Llave et al. 2002a; Mourelatos et al. 2002; Park et al. 2002; Reinhart et al. 2002). However, not all tiny RNAs qualify as miRNAs, and care has to be taken not to annotate breakdown products of abundant RNAs like rRNAs or tRNAs as miRNAs. Important criteria for a small RNA to qualify as an miRNA are 20–30-bp hairpin precursor structures, phylogenetic conservation in multiple species, and experimental evidence for their existence by cDNA cloning or Northern blot analysis (see Annotation guide for miRNAs, V. Ambros, B. Bartel, D.P. Bartel, C.B. Burge, J.C. Carrington, X. Chen, G. Dreyfuss, S.R. Eddy, S. Griffiths-Jones, M. Matzke, et al., in prep.). The 21-nt sequences that have been cloned but do not match to the genome of the respective organism or a closely related species do not qualify as miRNAs.

Small cloned RNAs may originate from transcribed repetitive heterochromatic sequences, recently identified in the fission yeast (Reinhart and Bartel 2002). These small RNAs, also referred to as heterochromatic siRNAs (Reinhart and Bartel 2002), provide a link to chromatin silencing and DNA methylation, likely by nucleating histone H3 lysine 9 methylation (Hall et al. 2002; Volpe et al. 2002). Other small RNAs that were derived from retro-elements have been detected in *D. melanogaster* (Elbashir et al. 2001) and *A. thaliana* (Hamilton et al. 2002; Llave et al. 2002a).

MicroRNAs are expressed constitutively or in a developmentally regulated manner, and may accumulate in certain tissues while being excluded from others (Lee et al. 1993; Reinhart et al. 2000; Lagos-Quintana et al. 2001, 2002; Lau et al. 2001; Lee and Ambros 2001; Llave et al. 2002a; Park et al. 2002; Reinhart et al. 2002). To function, the miRNAs need to be coexpressed with their respective target mRNAs. We have generated a comprehensive collection of mammalian miRNAs by cloning them from various mouse tissues and human cell lines to provide a starting point for their functional characterization. This report describes another 31 novel miRNAs, more than half of which are conserved between other vertebrate animals, and discusses some features of the more interesting miRNAs. We have now reached what we believe are the limits of miRNA isolation from somatic tissues or cells by conventional RNA isolation and cloning methods.

RESULTS AND DISCUSSION

We have continued the characterization of mouse miRNAs by cloning the ~21-nt RNAs from lung, kidney, skin, testis, ovary, thymus, spinal cord, and eye of 18.5-week-old adult mice using the method described previously (Lagos-Quintana et al. 2001, 2002). Additionally, the miRNAs expressed in the human osteoblast sarcoma cell line Saos-2 were cloned. In total, ~600 clones of miRNAs were obtained, 91% of which represent previously identified miRNAs and 9% as-yet unidentified miRNAs. The newly identified miRNAs are listed in Table 1. The novel miRNAs were identified in the various genome databases by BLAST sequence analysis. It was then confirmed that the flanking sequences can base-pair to the cloned sequence, forming the typical hairpin precursor structures (data not shown). When human and mouse orthologous miRNAs differed slightly in sequence, the species name is indicated before the miRNA name. In all, 31 novel miRNAs were identified. From all the newly identified mouse miRNAs, only one could not be identified within the human genome. Vice versa, three miRNAs cloned from human could not be identified in the mouse genome sequence, although one of them was identified in zebrafish. Because the majority (91%) of sequenced miRNAs matched to previously identified miRNAs, we believe that we have reached near-complete identification of miRNA genes expressed in somatic mouse or human cells. Although we cannot exclude the possibility that certain rare cell types or tissues, such as stem cells, may express distinct sets of miRNAs, it remains impractical to identify these miRNAs by the conventional cloning protocols, because the amount or the purity of RNA that can be obtained is insufficient.

One of the features of miRNAs is that they can occur in clusters that are coexpressed (Lagos-Quintana et al. 2001, 2002; Lau et al. 2001; Lee et al. 2002). Seven of the newly identified miRNAs are found in clusters (Fig. 1), and the

TABLE 1. Novel mouse and human miRNA sequences

miRNA	Sequence (5′ to 3′)	Number of clones										
		ht	ln	lv	sp	kd	sk	ts	ov	thy	eye	S
miR-10b	CCCUGUAGAACCGAAUUUGUGU					1			3			
miR-129b	CUUUUUGCGGUCUGGGCUUGUU							1			1	1
miR-181	AACAUUCAACGCUGUCGGUGAGU	1					1				2	
miR-182	UUUGGCAAUGGUAGAACUCACA										1	
miR-183	UAUGGCACUGGUAGAAUUCACUG										1	
miR-184	UGGACGGAGAACUGAUAAGGGU										2	
miR-185	UGGAGAGAAAGGCAGUUC										1	
miR-186	CAAAGAAUUCUCCUUUUGGGCUU										1	1
miR-187	UCGUGUCUUGUGUUGCAGCCGG					1						
miR-188	CAUCCCUUGCAUGGUGGAGGGU					1						
miR-189	GUGCCUACUGAGCUGACAUCAGU					1						
miR-190	UGAUAUGUUUGAUAUAUUAGGU					2						
miR-191	CAACGGAAUCCCAAAAGCAGCU				2	1						
miR-192	CUGACCUAUGAAUUGACA			2		1						
miR-193	AACUGGCCUACAAAGUCCCAG					1						
miR-194	UGUAACAGCAACUCCAUGUGGA					1						
miR-195	UAGCAGCACAGAAAUAUUGGC		2			1	1					
miR-196	UAGGUAGUUUCAUGUUGUUGG								1			1
miR-197	UUCACCACCUUCUCCACCCAGC											3
miR-198	GGUCCAGAGGGGAGAUAGG											2
miR-199-s	CCCAGUGUUCAGACUACCUGUU											2
miR-199-as	UACAGUAGUCUGCACAUUGGUU						1					
miR-200a	UAACACUGUCUGGUAACGAUG					1						
miR-200b	UAAUACUGCCUGGUAAUGAUGAC		2					1				
miR-201	UACUCAGUAAGGCAUUGUUCU							1				
miR-202	AGAGGUAUAGCGCAUGGGAAGA							1				
miR-203	UGAAAUGUUUAGGACCACUAG						2	1				
miR-204	UUCCCUUUGUCAUCCUAUGCCUG										1	
miR-205	UCCUUCAUUCCACCGGAGUCUG						1					
miR-206	UGGAAUGUAAGGAAGUGUGUGG						2					
miR-207	GCUUCUCCUGGCUCUCCUCCCUC									1		
miR-208	AUAAGACGAGCAAAAAGCUUGU	1										

The number of clones identified from the indicated mouse tissues or the human osteoblast sarcoma cell line Saos-2 is presented. Abbreviations: ht, heart; ln, lung; lv, liver; sp, spleen; kd, kidney; sk, skin; ts, testis; ov, ovary; thy, thymus; S, Saos-2. The suffixes -s and -as indicate that the RNA is derived from either the 5′ half or the 3′ half of an miRNA precursor. Comprehensive tables for mouse and human miRNAs, miRNA precursors, and their respective tissue distribution are provided online at http://www.mpibpc.gwdg.de/abteilungen/100/105/mirna.html.

distance between the clustered miRNAs varies between 500 and 5000 bp.

It was previously noted that some miRNAs were strongly enriched in specific tissues, such as miR-1 variants in heart, miR-122 variants in liver, or miR-124 variants in brain (Lee and Ambros 2001; Lagos-Quintana et al. 2002). Here, we find that the miRNA profiles of eye and spinal cord were very similar to those observed from brain tissues, in which miR-124 variants are strongly enriched, presumably reflecting the high neuronal cell content. None of the other examined tissues or cell lines showed a strong dominance of a specific miRNA. Variants of let-7 miRNA were fairly abundant in all examined somatic tissues or cells. Similar to the ubiquitous let-7 distribution, we find miR-16, miR-26a, miR-27a, and miR-143a in almost all somatic cells or tissues. miR-21 is also ubiquitously expressed but excluded from neuronal tissue. The exclusion of miR-21 and the presence of neuron-specific miR-124 variants from neuronal tissue may contribute to neuronal cell specification.

To establish tissue culture systems for analyzing miRNA

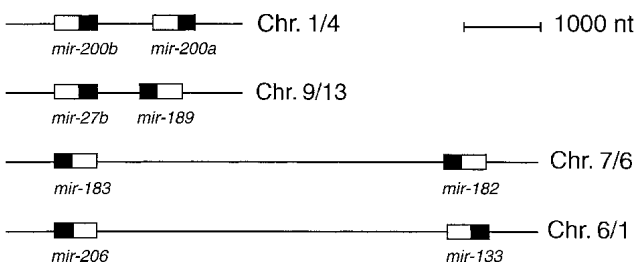


FIGURE 1. MicroRNA gene clusters. The precursor structure is indicated as a box, and the location of the miRNA within the precursor is shown in black. The clusters are transcribed from *left* to *right*. To the *right*, the chromosome location is indicated for human/mouse. The cluster of *mir-183* and *mir-182* is also conserved in zebrafish.

function, we have previously cloned the miRNAs from the cervical cancer cell line HeLa SS3 (Lagos-Quintana et al. 2001). To examine the miRNA profiles between different cell lines in culture, we cloned the miRNAs from the osteoblast sarcoma cell line Saos-2, another commonly used laboratory cell line. Six new miRNAs (miR-129b, miR-186, miR-196 to miR-199) were cloned from this cell line, albeit at such low frequency that it cannot be stated with certainty that these miRNAs are expressed cell-line-specifically. The expression pattern of the most abundant miRNAs (let-7a, miR-16, and miR-21) was very similar between Saos-2 and HeLa cells.

We have analyzed the genomic position of 54 mouse (60 human) miRNAs and find that 31 (33) are localized in intergenic regions. Of the remaining miRNAs, 11 (13) are found in sense orientation within introns of coding transcripts, 7 (7) in sense orientation within introns of noncoding genes, and 5 (7) in the reverse orientation within an intronic region. Therefore, miRNAs are either transcribed from their own promoters or derived from a pre-mRNA that frequently codes for an additional gene product. An example of an intronic miRNA is miR-186: Human and mouse miR-186 are both located in intron 8 of the pre-mRNA of the zinc finger protein 265; both introns are ~2 kb, but only the miRNA precursors and ~50 nt of flanking sequences are conserved between the two species. An example of intronic miRNAs located within a non-protein-coding transcript are miR-15a and miR-16: These miRNAs are clustered and located within the intron of a transcribed region referred to as *LEU2*. The *LEU2* gene lies within the deleted minimal region (DMR) of the B-cell chronic lymphocytic leukemia (B-CLL) tumor suppressor locus (Bullrich et al. 2001; Migliazza et al. 2001), for which the tumor suppressor gene yet has to be identified; miR-15a and miR-16 are strong candidates.

Some miRNAs are found in interesting genomic locations. For example, miR-10, which was originally identified by cloning from *D. melanogaster* embryos (Lagos-Quintana et al. 2001), is located in the Hox gene cluster between *Dfd* (*Hox4*) and *Scr* (*Hox5*). miR-10 also appears to be conserved in the Hox cluster of *Anopheles gambiae* (mosquito) and *Tribolium castaneum* (red flour beetle). miR-10 is also preserved in the Hox clusters of zebrafish, pufferfish, mouse, and human. In mouse and human the *mir-10* gene has been duplicated and is now present in the form of two variants. miR-10a is located between *Hox4b* and *Hox5b* in mouse and human, whereas miR-10b is found within intron 4 of *Hox4d* in mouse, and between *Hoxd4* and *Hoxd8* in human. Considering the spatial and temporal colinearity of Hox gene expression and the positional conservation of miR-10 within the Hox gene clusters, it is conceivable that miR-10 is important for regulating developmental events. The evolutionary conserved miRNAs, such as miR-10, are probably the most interesting starting points for functional analysis of this vast gene family (Table 2).

TABLE 2. miRNA and miRNA families conserved between vertebrates and invertebrates

miRNA family	*H. sap.*	*M. mus.*	*F. rub.*	*D. rerio*	*D. mel.*	*C. elegans*
let-7/miR-98	*p c n*	*p c n*	*p*	*p n*	*p c n*	*p c n*
lin-4/miR-125	*p*	*p c n*	*p*		*p n*	*p c n*
miR-1	*p c n*	*p c n*		*p*	*p c n*	*p c n*
miR-4/75/79/131	*p c*	*p c*	*p*	*p*	*p c n*	*p*
miR-7	*p*	*p*	*p*		*p c n*	
miR-8/141	*p*	*p c*			*p c n*	
miR-9	*p*	*p c*	*p*		*p c n*	
miR-10/99/100	*p c*	*p c*	*p*	*p*	*p c n*	
miR-29/83	*p c*	*p c*	*p*	*p*		*p c*
miR-31/73	*p c*	*p*			*p*	*p*
miR-34	*p*				*p*	*p c*
miR-25/92	*p c*	*p c*	*p*	*p*	*p c*	
miR-124	*p*	*p c n*	*p*	*p*	*p*	*p*
miR-133	*p*	*p c*	*p*		*p*	
miR-184	*p*	*p c*	*p*	*p*	*p*	

(*p*) A predicted miRNA identified by BLAST searching of a cloned miRNA in another species. The flanking sequences of the predicted miRNA have the capacity to fold with the miRNA into a hairpin precursor. (*c*) The miRNA was cloned from the respective organism. (*n*) The miRNA expression was confirmed by Northern blot analysis (Lagos-Quintana et al. 2001, 2002; Lee et al. 2002). Abbreviations: *H. sap., Homo sapiens; M. mus., Mus musculus; F. rub. Fugu rubripes; D. rerio, Danio rerio; D. mel., Drosophila melanogaster; C. elegans, Caenorhabditis elegans.*

In summary, 31 novel mammalian miRNAs were identified in an effort to provide a comprehensive list of these regulatory molecules. The identification of functional RNA genes is important for obtaining fully annotated genome sequences in order to assist researchers determined to identify genes linked to a particular disease and to understand posttranscriptional regulation of gene expression.

MATERIALS AND METHODS

Total RNA isolation, cloning, and bioinformatics analysis

RNA preparation and cloning of miRNAs was performed as last described (Lagos-Quintana et al. 2002). Total RNA from Saos-2 cells was isolated from a cell line that was stably transfected with an inducible *p53* gene. No significant differences in miRNA composition were observed between RNA preparations from p53-induced or p53-non-induced cells. For sequence searches and analysis, we used the Ensembl database (http://www.ensembl.org), NCBI online resources (http://www.ncbi.nlm.nih.gov), the DOE Joint Genome Institute Web site (http://www.jgi.doe.gov), the UCSC Genome Bioinformatics Web site (http://genome.ucsc.edu), and the mfold Web server (http://www.bioinfo.rpi.edu/applications/mfold/old/rna).

ACKNOWLEDGMENTS

We thank S.M. Elbashir and W. Lendeckel for the preparation of mouse tissues, G. Dowe for some of the sequencing, A. Aravin and

T. Fischer for bioinformatics assistance and program development, Y. Dorsett and J. Martinez for critical comments on the manuscript, and R. Lührmann for support. This work was funded by a Bundesministerium für Bildung und Forschung (BMBF) Biofuture grant, number 0311856.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Received September 23, 2002; accepted September 30, 2002.

REFERENCES

Ambros, V. 2000. Control of developmental timing in *Caenorhabditis elegans*. *Curr. Opin. Genet. Dev.* **10:** 428–433.

Bullrich, F., Fujii, H., Calin, G., Mabuchi, H., Negrini, M., Pekarsky, Y., Rassenti, L., Alder, H., Reed, J.C., Keating, M.J., et al. 2001. Characterization of the 13q14 tumor suppressor locus in CLL: Identification of ALT1, an alternative splice variant of the LEU2 gene. *Cancer Res.* **61:** 6640–6648.

Conne, B., Stutz, A., and Vassalli, J.D. 2000. The 3′ untranslated region of messenger RNA: A molecular 'hotspot' for pathology? *Nat. Med.* **6:** 637–641.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. 2001. RNA interference is mediated by 21 and 22 nt RNAs. *Genes & Dev.* **15:** 188–200.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106:** 23–34.

Hall, I.M., Shankaranarayana, G.D., Noma, K.I., Ayoub, N., Cohen, A., and Grewal, S.I. 2002. Establishment and maintenance of a heterochromatin domain. *Science* **297:** 2232–2237.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. 2002. Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21:** 4671–4679.

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. 2001. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293:** 1146–1150.

Hayward, N. 2000. New developments in melanoma genetics. *Curr. Oncol. Rep.* **2:** 300–306.

Hutvágner, G. and Zamore, P.D. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297:** 2056–2060.

Hutvágner, G., McLachlan, J., Bálint, É., Tuschl, T., and Zamore, P.D. 2001. A cellular function for the RNA interference enzyme Dicer in small temporal RNA maturation. *Science* **93:** 834–838.

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

Kiyohara, C., Otsu, A., Shirakawa, T., Fukuda, S., and Hopkin, J. 2002. Genetic polymorphisms and lung cancer susceptibility: A review. *Lung Cancer* **37:** 241.

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

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.

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

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.

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

Lee, R.C., Feinbaum, R.L., and Ambros, V. 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75:** 843–854.

Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. 2002. microRNA maturation: Stepwise processing and subcellular localization. *EMBO J.* **21:** 4663–4670.

Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. 2002a. Endogenous and silencing-associated small RNAs in plants. *Plant Cell* **14:** 1605–1619.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. 2002b. Cleavage of *Scarecrow*-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* **297:** 2053–2056.

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

Mendell, J.T. and Dietz, H.C. 2001. When the message goes awry: Disease-producing mutations that influence mRNA content and performance. *Cell* **107:** 411–414.

Migliazza, A., Bosch, F., Komatsu, H., Cayanis, E., Martinotti, S., Toniato, E., Guccione, E., Qu, X., Chien, M., Murty, V.V., et al. 2001. Nucleotide sequence, transcription map, and mutation analysis of the 13q14 chromosomal region deleted in B-cell chronic lymphocytic leukemia. *Blood* **97:** 2098–2104.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. 2002. miRNPs: A novel class of ribonucleoproteins containing numerous microRNAs. *Genes & Dev.* **16:** 720–728.

Olsen, P.H. and Ambros, V. 1999. The lin-4 regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216:** 671–680.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* **12:** 1484.

Pasquinelli, A.E. and Ruvkun, G. 2002. Control of developmental timing by microRNAs and their targets. *Annu. Rev. Cell Dev. Biol.* **18:** 495–513.

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

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403:** 901–906.

Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. 2002. MicroRNAs in plants. *Genes & Dev.* **16:** 1616–1626.

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.

Rougvie, A.E. 2001. Control of developmental timing in animals. *Nat. Rev. Genet.* **2:** 690–701.

Schwarz, D.S., Hutvágner, G., Haley, G., and Zamore, P.D. 2002. siRNAs function as guides, not primers, in the RNAi pathway in *Drosophila* and human cells. *Mol. Cell* (in press).

Seggerson, K., Tang, L., and Moss, E.G. 2002. Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene lin-28 after translation initiation. *Dev. Biol.* **243:** 215–225.

Slack, F. and Ruvkun, G. 1997. Temporal pattern formation by heterochronic genes. *Annu. Rev. Genet.* **31:** 611–634.

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

Wightman, B., 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.

Chapter III: The small RNA profile during *Drosophila melanogaster* development

Alexei A. Aravin, **Mariana Lagos-Quintana**, Abdlulah Yalcin, Mihaela Zavolan, Debora Marks, Ben Snyder, Terry Gaasterland, Jutta Meyer and Thomas Tuschl. The small RNA profile durind *Drosophila melanogaster* development. Dev Cell. 2003 Aug;5(2):337-50.

Statement of contribution

Bioinformatic analysis

Writing

Developmental Cell, Vol. 5, 337–350, August, 2003, Copyright ©2003 by Cell Press

The Small RNA Profile during *Drosophila melanogaster* Development

Alexei A. Aravin,[1] Mariana Lagos-Quintana,[2] Abdullah Yalcin,[3] Mihaela Zavolan,[4] Debora Marks,[5] Ben Snyder,[4] Terry Gaasterland,[4] Jutta Meyer,[2] and Thomas Tuschl[2,*]

[1]Department of Animal Molecular Genetics
Institute of Molecular Genetics
Kurchatov sq. 2
Moscow 123182
Russia

[2]Laboratory for RNA Molecular Biology
The Rockefeller University
1230 York Avenue, Box 186
New York, New York 10021

[3]Department for Cellular Biochemistry
Max Planck Institute for Biophysical Chemistry
Am Fassberg 11
37077 Göttingen
Germany

[4]Laboratory for Computational Genomics
The Rockefeller University
1230 York Avenue, Box 250
New York, New York 10021

[5]Center for Computational Biology and Bioinformatics
Columbia University Genome Center
Russ Berrie Pav., Room 520b
1150 St. Nicholas Avenue
New York, New York 10032

Summary

Small RNAs ranging in size between 20 and 30 nucleotides are involved in different types of regulation of gene expression including mRNA degradation, translational repression, and chromatin modification. Here we describe the small RNA profile of *Drosophila melanogaster* as a function of development. We have cloned and sequenced over 4000 small RNAs, 560 of which have the characteristics of RNase III cleavage products. A nonredundant set of 62 miRNAs was identified. We also isolated 178 repeat-associated small interfering RNAs (rasiRNAs), which are cognate to transposable elements, satellite and microsatellite DNA, and *Suppressor of Stellate* repeats, suggesting that small RNAs participate in defining chromatin structure. rasiRNAs are most abundant in testes and early embryos, where regulation of transposon activity is critical and dramatic changes in heterochromatin structure occur.

Introduction

One of the recent surprises in molecular and cellular biology was the identification of 20 to 30 nucleotide (nt) RNA molecules that guide transcriptional and posttranscriptional gene silencing (for reviews see Denli and Hannon, 2003; Grishok and Mello, 2002; Moss, 2002; Voinnet, 2002; Wassenegger, 2002; Zamore, 2002). The small RNAs are derived from double-stranded RNA (dsRNA) precursors that are processed by the ribonuclease type III enzyme Dicer (Bernstein et al., 2001; Elbashir et al., 2001b; Grishok et al., 2001; Hutvágner et al., 2001; Ketting et al., 2001; Park et al., 2002; Provost et al., 2002; Reinhart et al., 2002; Zhang et al., 2002). At least three types of small RNAs have been described: (1) Small interfering RNAs (siRNAs) (Caplen et al., 2001; Elbashir et al., 2001a, 2001b), (2) microRNAs (miRNAs) (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001), and (3) repeat-associated small interfering RNAs (rasiRNAs) (Djikeng et al., 2001; Elbashir et al., 2001b; Hamilton et al., 2002; Llave et al., 2002a; Mette et al., 2002; Reinhart and Bartel, 2002).

siRNAs and miRNAs are derived from different sources of dsRNA and act in common pathways interchangeably depending on the degree of complementarity with their target RNA sequence (Doench et al., 2003; Hutvágner and Zamore, 2002; Llave et al., 2002b; Rhoades et al., 2002; Tang et al., 2003). Sources for siRNAs are dsRNAs that are typically hundreds of base pairs (bp) long, e.g., replication intermediates of RNA viruses (for reviews see Plasterk, 2002; Waterhouse et al., 2001). miRNAs originate from genes that encode short 20 to 30 bp dsRNA hairpins and represent a growing class of noncoding RNAs that are found in plants and animals (Ambros et al., 2003b; Dostie et al., 2003; Grad et al., 2003; Lagos-Quintana et al., 2001, 2002, 2003; Lau et al., 2001; Lee and Ambros, 2001; Lee et al., 1993; Lim et al., 2003a, 2003b; Llave et al., 2002a; Mette et al., 2002; Mourelatos et al., 2002; Park et al., 2002; Reinhart et al., 2000, 2002). Many miRNAs are conserved in sequence between distantly related organisms, suggesting that the interactions between these miRNAs and their targets constitute essential processes (Pasquinelli et al., 2000).

miRNA expression is regulated in a spatial and temporal manner. The miRNA transcripts are trimmed in the nucleus to a dsRNA precursor and subsequently exported to the cytoplasm for Dicer processing (Lee et al., 2002). Most of our current knowledge of miRNA function derives from the two *Caenorhabditis elegans* miRNAs lin-4 and let-7, originally named small temporal RNAs (stRNAs) for their role in developmental timing (Lee et al., 1993; Reinhart et al., 2000). Investigation of *lin-4* and *let-7* genetic interactions identified genes containing sequences in their 3′ UTRs partially complementary to these miRNAs (Abrahante et al., 2003; Lin et al., 2003; Pasquinelli and Ruvkun, 2002). lin-4 and let-7 miRNAs form bulged RNA duplexes with their target mRNAs, which leads to repression of protein synthesis after translation initiation while the mRNA remains intact (Olsen and Ambros, 1999; Seggerson et al., 2002; Wightman et al., 1993). Recently, two *D. melanogaster* miRNAs, bantam (Brennecke et al., 2003) and miR-14 (Xu et al., 2003), were genetically characterized and some of the respective target mRNAs were identified. However, the precise mechanism of target repression by

*Correspondence: ttuschl@rockefeller.edu

these *D. melanogaster* miRNAs remains to be characterized.

rasiRNAs are presumably derived from long dsRNAs and match to repetitive sequence element in sense and antisense orientation (Djikeng et al., 2001; Llave et al., 2002a; Reinhart and Bartel, 2002). Repetitive sequences are often associated with regions of heterochromatin. Transposable elements (TEs), for example, may produce dsRNA upon random integration nearby transcriptionally active sequences. The high density of transposons in particular genomic regions makes it more likely that dsRNA is generated by transcription from adjacent opposing promoters of transposons rather than cellular genes. rasiRNAs are likely to function as guide RNAs during the establishment and/or maintenance of heterochromatin in plants (Hamilton et al., 2002; Llave et al., 2002a; Mette et al., 2002), *Trypanosoma brucei* (Djikeng et al., 2001), *Drosophila melanogaster* (Aravin et al., 2001; Pal-Bhadra et al., 2002), and fission yeast (Hall et al., 2002; Reinhart and Bartel, 2002; Volpe et al., 2002). Small RNAs and proteins related to RNA silencing have also been identified in *Tetrahymena thermophila* and are thought to participate in guiding programmed DNA elimination of dispersed sequence elements in order to form the transcriptionally active macronucleus after sexual conjugation (Mochizuki et al., 2002; Tautz et al., 1988; Taverna et al., 2002).

Two distinct classes of small RNAs were found in plants. These classes are 21 nt and 24 nt in size and mediate posttranscriptional and transcriptional gene silencing (Hamilton et al., 2002; Mallory et al., 2002; Tang et al., 2003). siRNA product inhibition experiments suggest that different homologs of Dicer are responsible for production of these two types of small RNAs (Tang et al., 2003). siRNAs and miRNAs are near 21 nt in size (Llave et al., 2002a; Reinhart et al., 2002), whereas rasiRNAs are 24 nt long (Hamilton et al., 2002; Llave et al., 2002a). Certain viral suppressors of RNA silencing in plants specifically affect the accumulation of 24 nt rasiRNAs (Hamilton et al., 2002). It was recently discovered that plants with mutations in the *Argonaute 4* gene were impaired in transcriptional silencing at certain loci (Zilberman et al., 2003). These mutants lacked accumulation of 24 nt siRNAs of the AtSN1 retroelement and derepressed a normally silent gene because of loss of DNA methylation and histone H3-lysine 9 methylation (Zilberman et al., 2003). Plant miRNAs are mostly 21 nt consistent with their role in posttranscriptional regulation (Llave et al., 2002a; Park et al., 2002; Reinhart et al., 2002).

A natural case in which repetitive DNA sequences function to silence a cellular gene was first described in *D. melanogaster* (Livak, 1984, 1990; Palumbo et al., 1994). Deletion of *Suppressor of Stellate* [*Su(Ste)*] repeats located on the Y chromosome results in derepression of *Stellate* located on the X chromosome leading to meiotic abnormalities and male infertility due to crystallization of overexpressed Stellate protein in sperm cells. Derepression of *Stellate* is also observed in mutants of the *spn-E* DE-H helicase (Aravin et al., 2001; Stapleton et al., 2001) and the Argonaute *aub* (Schmidt et al., 1999). Interestingly, *spn-E* mutants accumulate transcripts from retrotransposons in the germline (Aravin et al., 2001; Kogan et al., 2003) and both *spn-E* and *aub* mutants are defective for initiation of RNAi in the female germline (Kennerdell et al., 2002). *Stellate* silencing is correlated with the presence of 25–27 nt siRNAs deriving from both strands of *Su(Ste)* (Aravin et al., 2001). A study investigating multicopy transgene silencing in *D. melanogaster* implicated another Argonaute gene, *piwi*, in posttranscriptional and transcriptional gene silencing (Pal-Bhadra et al., 2002). Links between protein factors involved in RNAi and the silencing of endogenous transposable elements have also been made in other species like *C. elegans* (Ketting et al., 1999; Tabara et al., 1999) and *Chlamydomonas reinhardtii* (Wu-Scharf et al., 2000).

The Argonautes have been shown to be the largest conserved class of proteins that link the various pathways of RNA silencing. Various Argonaute family members function in both posttranscriptional and transcriptional gene silencing in protists, fungi, plants, and animals (for reviews see Carmell et al., 2002). Biochemical characterization of the RNAi machinery identified a ribonucleoprotein complex termed RISC (RNA induced silencing complex) that contains the siRNAs as guide RNAs (Elbashir et al., 2001b; Hammond et al., 2000; Zamore et al., 2000). In *D. melanogaster*, RISC was found to contain Ago2 (Hammond et al., 2001), and, in human cells, RISC contains the Argonaute proteins eIF2C1 and/or eIF2C2 (Martinez et al., 2002). Interestingly, eIF2C2 was also identified in a protein complex associated with miRNAs in humans (Mourelatos et al., 2002), and it was subsequently shown that this complex was able to cleave miRNA-complementary target mRNAs (Hutvágner and Zamore, 2002). The reciprocal case, that siRNAs can also function as miRNAs, was also shown (Doench et al., 2003). Therefore, small RNAs associated with Argonaute proteins can either act as siRNA or miRNAs depending on the degree of complementarity to the target mRNA.

Here we cloned and sequenced the short RNAs present in *D. melanogaster* at different stages of development as well as in adult testes. Several different classes of dsRNA-derived short RNAs were identified. We validated 62 unique miRNAs and annotated 178 rasiRNAs. The distinct features of the two classes of small RNAs suggest different regulatory mechanisms and function for the two classes. This study also indicated the fundamental importance of recording the small RNA profile for understanding gene regulation in eukaryotic cells.

Results

Cloning of *D. melanogaster* Small RNAs

Small RNAs in the size range of 16–29 nt were cloned from defined developmental stages of *D. melanogaster*. In addition, small RNAs were also cloned from microdissected adult testis. To obtain the small RNA fraction, total RNA was isolated using acid guanidinium thiocyanate-phenol-chloroform (Chomczynski and Sacchi, 1987) adapted to maximize recovery of small RNAs (Lagos-Quintana et al., 2001, 2002). The small RNAs were then size fractionated on denaturing polyacrylamide gels and excised from the gel directly below the 2S rRNA band that is detectable by UV shadowing. In contrast to many other species, *D. melanogaster* ribosomal RNA (rRNA) is composed of four individual RNAs, 28S, 18S, 5.8S, and 2S, the latter of which is 30 nt in

Table 1. Composition of Small RNA cDNA Libraries Prepared from Different Developmental Stages and Testes of *D. melanogaster*

	Embryo Stage (hr)					Larva Stages							
Type	0–2	2–4	4–6	6–12	12–24	L1	L1 + L2	L3	Pupa	Adult	Testes	Total (clones)	Total (%)
rRNA	161	119	160	85	23	193	111	293	91	438	451	2160	53.0
tRNA[a]	25	14	25	1	2	35	64	53	9	99	25	354	8.7
miRNA	112	42	25	7	3	12	8	8	—	121	41	382	9.4
rasiRNA	79	39	13		1	2	3	—	—	12	28	178	4.4
mRNA	3	7	11	2	1	26	30	32	1	16	13	143	3.5
snRNA/snoRNA[b]	4	—	4	2	3	1	2	2	—	16	1	35	0.9
Other ncRNA[c]	8	1	7	1	—	1	3	1	1	13	3	40	1.0
S. cerevisiae	45	101	49	22	19	53	61	26	1	141	7	525	12.9
Bacteria, plants	1	1	2	—	—	—	6	4	1	2	—	17	0.4
DCV	—	—	—	—	—	1	1	11	1	3	—	17	0.4
Unknown[d]	31	14	7	2	7	15	32	25	8	56	25	223	5.5
Total	469	338	303	122	59	339	321	455	113	920	594	4074	100

The number of sequenced clones is indicated according to matches to the annotation provided within the various public databases including the annotation of the *D. melanogaster* genome (version 3.1 from http://www.bdgp.org), a dataset of *D. melanogaster* sequences from GenBank (http://www.fruitfly.org/sequence/sequence_db/na_gb.dros), a database of transposable elements (http://www.bdgp.org/p_disrupt/datasets/VERSION3/ALL_SEQUENCES_dmel_RELEASE3.FASTA.ALL.v3) and canonical sequences (http://www.bdgp.org/p_disrupt/datasets/NATURAL_TRANSPOSABLE_ELEMENTS.fa), a database of *D. melanogaster* tRNA sequences (http://rna.wustl.edu/GtRDB/Dm/Dm-seqs.html), a database of small RNA sequences provided by A. Hüttenhofer, and a database of miRNAs (http://www.sanger.ac.uk/Software/Rfam/ftp.shtml) and predicted miRNA sequences (Lim et al., 2003). The assignment of annotation was performed in a hierarchical manner. Perfect matches of small RNA sequences to the genomes of *S. cerevisiae* were classified as *S. cerevisiae* sequences followed by matches to *D. melanogaster* rRNA, tRNA, and then snRNAs/snoRNAs and other ncRNAs. The remaining sequences were then checked against euchromatic and heterochromatic *D. melanogaster* genomic sequences as well as GenBank sequences of other organisms and classified as mRNAs, *Drosophila C Virus* (*DCV*), bacterial, and plant genomes.

[a] The annotation for small RNAs as tRNA breakdown products was performed using the tRNA sequences provided at the Genomic tRNA Database at http://rna.wustl.edu/tRNAdb/ as well as annotated GenBank sequences.

[b] snRNA and snoRNA hits were assigned by using the sequence sets provided in (Yuan et al., 2003).

[c] The noncoding RNAs (ncRNAs) were provided from A. Hüttenhofer and contained longer RNA sequences (>40 nt) without annotation or assigned function.

[d] Unknown describes a category for which most sequences do not match to *S. cerevisiae* or *D. melanogaster*. This category also contains some sequences with near perfect match to regions of *D. melanogaster* with no available annotation and no evidence for a repetitive character or a miRNA fold-back structure.

size (Tautz et al., 1988). The small RNAs were then cloned and sequenced (Pfeffer et al., 2003).

A total of 4074 clone sequences were obtained and current public databases were used to annotate 95.6% of these sequences; the residual sequences could not be annotated because they did not match to any of the sequenced genomes in the database (183 clones) or because they matched to a region of the *D. melanogaster* genome for which no functional or sequence homology assignment could be made (40 clones). The largest class of cloned RNAs represents breakdown products of abundant noncoding (or nonmessenger) RNAs (rRNA, tRNA, snRNA, snoRNA, and others) of *D. melanogaster* (63.5%), followed by breakdown product sequences from *Saccharomyces cerevisiae* (12.9%), which constitutes the preferred diet of *D. melanogaster*. A few bacterial and plant rRNA fragments from salmonella, cereal, and hops were also found, presumably because the baking yeast that was used for fly food was brewing yeast. A small fraction of *D. melanogaster* mRNA breakdown products (3.5%) was also identified (Table 1). The residual 577 sequences (14.1%) fell into the following three classes: miRNAs (9.3%), repeat-associated small RNAs derived from sense and antisense strands of repetitive elements (4.4%), and small RNAs from *Drosophila* virus C (0.4%).

The proportion of small RNAs with regulatory function relative to rRNA breakdown products varied drastically for the different developmental stages. Early embryos and adults showed the highest content of small regulatory RNAs (between 15% and 40%) while in late embryo and larvae the number was significantly lower. From pupae, almost all cloned small RNAs represented rRNA and tRNA breakdown products. Presumably, extensive apoptosis during metamorphic tissue reorganization caused extensive rRNA breakdown and made it impossible to clone small RNAs at high frequency. Also, we found that pupae- and adult-specific miRNAs, such as let-7, were still present in pupae RNA preparations and were readily detectable by Northern blotting (Figure 1B).

miRNAs and siRNAs are generated by Dicer RNase III cleavage from dsRNA precursors and therefore contain 5′ phosphates and 2′,3′-hydroxyl termini (Bernstein et al., 2001; Elbashir et al., 2001b; Grishok et al., 2001; Hutvágner et al., 2001; Nykänen et al., 2001). In contrast, breakdown products of longer RNAs that are generated by single-strand specific RNases or by simple hydrolysis should contain 5′-hydroxyl and 2′,3′ cyclic phosphate or 2′ or 3′ monophosphate termini. In theory, it should be possible to discriminate between small regulatory RNAs and RNA breakdown products by the presence or absence of a 5′ phosphate. The cloning method described by Lau et al. (2001), which takes advantage of the presence of a 5′ phosphate, should therefore discriminate against RNA breakdown products and enrich for RNase III processing products. However, when we compared the two cloning protocols using identical sources of starting material, we found that the content

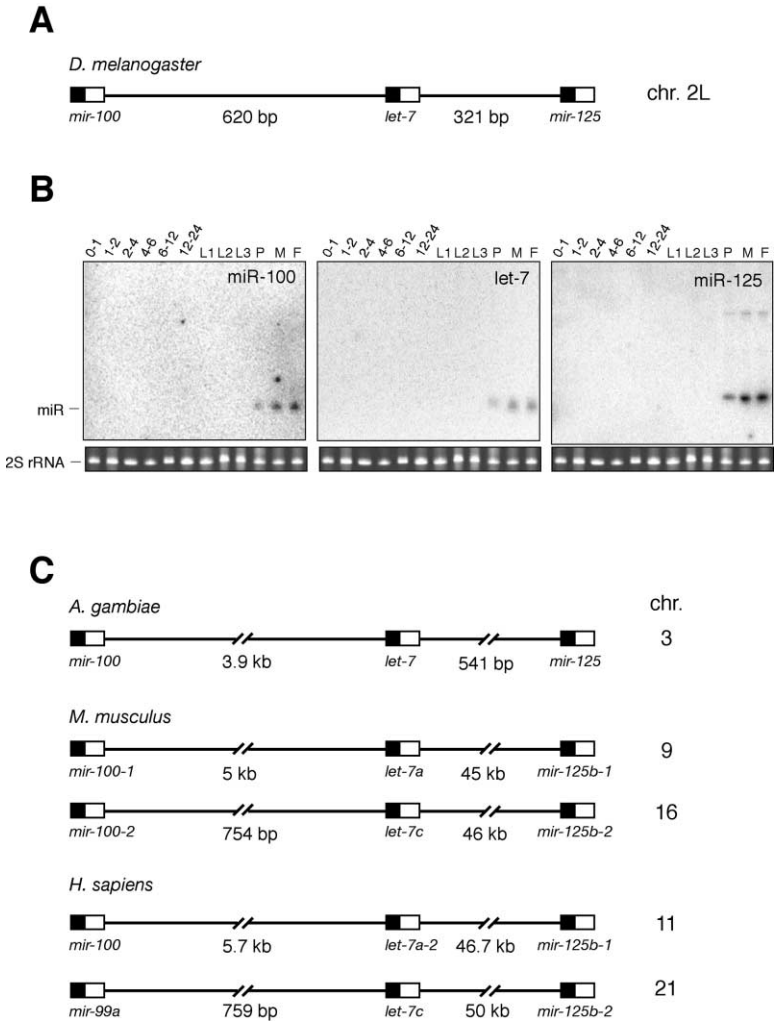


Figure 1. miRNA Genes *mir-100*, *let-7* and the *lin-4* Homolog *mir-125* Are Clustered

(A) Arrangement of the miRNA genes in *D. melanogaster*. The 70 nt fold-back precursor is indicated as box, the position of the miRNA in the precursor is shown in black. The chromosome location is indicated to the right.

(B) Northern blots confirming the coexpression of miR-100, let-7, and miR-125 in pupa and adult. As control for loading the 30 nt 2S rRNA band was visualized by ethidium bromide staining of the polyacrylamide gel before transfer of the RNA to the blotting membrane. Development stages of embryos are indicated in hours after egg laying. Larval stages are indicated as L1, L2, and L3. P indicates pupal stage; M, adult male; F, adult female.

(C) The miRNA clustering is conserved between invertebrates and vertebrates but gene duplication occurred in mammals and the spacing between miRNA precursors increased with increasing genome size.

of rRNA and tRNA breakdown RNAs was practically the same. Both methods yielded 15% of miRNA sequences using 0 to 2 hr embryo small RNAs and between 2% and 3% miRNAs from combined L1 and L2 stage small RNAs after sequence analysis of more than 100 clones for each RNA source and each cloning method. This suggests that small RNAs without a 5′ phosphate are rapidly phosphorylated in the cell and become indistinguishable from Dicer RNase III processing products. Rapid addition of 5′ phosphates to nonphosphorylated synthetic siRNA duplexes as well as rapid dephosphorylation of 3′ phosphates of synthetic siRNA duplexes was observed previously in *D. melanogaster* and human cell extracts (Martinez et al., 2002; Nykänen et al., 2001). However, it was not previously examined if other single- or double-stranded RNAs were also substrates for these kinases and phosphatases. We therefore continued using our original cloning protocol (Lagos-Quintana et al., 2001, 2002; Pfeffer et al., 2003).

Identification of Distinct Classes of Small RNAs

Cloning of small RNAs provides an overview of the RNAs expressed in a cell or organism. If a sufficiently large number of clones were examined, complete contigs of all nonmessenger RNAs could have been assembled from partly overlapping small RNA segments similar to shotgun genome sequencing. In contrast to dsRNA processing products, the breakdown products of noncoding RNAs and mRNAs show no significant length bias and are broadly distributed throughout the size window used for cloning. miRNAs and rasiRNAs show a much tighter size distribution (Supplemental Table S1 [http://www.developmentalcell.com/cgi/content/full/5/2/337/DC1]), which centers on distinct sizes for both classes. About 70% of the identified sequences of miRNAs are between 21 and 23 nt in size with an average of 22.0 nt. This size distribution is similar to the length distribution of siRNAs generated by in vitro processing of dsRNA in *D. melanogaster* embryo lysate (Elbashir et al., 2001a). About 65% of the rasiRNAs are between 23 and 26 nt in size with an average of about 23.6 nt (Supplemental Table S1).

miRNAs, siRNAs, and rasiRNAs have a strong preference for pyrimidine residues, uridine in particular, at their 5′-most position. The average nucleotide composition for the first, second, and last nucleotide of the various RNA classes is shown in Supplemental Table S2. Only the 5′-most position of RNase III processing products displays this sequence bias. Additional support

for the presence of dsRNA precursors for rasiRNAs is provided by the presence of sense as well as antisense oriented sequences cognate to repeat elements of LTR and non-LTR retrotransposons.

Absence of Evidence for Antisense mRNA Regulation

Bioinformatic analysis indicated that a significant number of genes are partially overlapping in opposite orientation in various organisms (Merino et al., 1994; Shendure and Church, 2002; Yelin et al., 2003). Partly complementary transcripts may participate in gene regulation by antisense or RNAi-related mechanisms. We cloned 143 small RNAs that matched to cDNAs or predicted cDNA sequences in the *D. melanogaster* genome. One hundred forty-one of these sequences were in sense (+) orientation to the mRNA without overlap to other genes, suggesting that these small RNAs are breakdown products of mRNAs. This is supported by the broad length distribution similar to rRNA and tRNA fragments and the absence of a 5′ pyrimidine sequence bias characteristic for dsRNA processing products (Supplemental Table S2). Only two sequences mapped to the coding region of predicted genes in antisense (−) orientation but without any evidence for the presence of overlapping transcripts. These findings are in strong contrast to observations in *C. elegans*, where 50%–90% of cloned small RNAs with a corresponding match to a protein-coding sequence were in antisense orientation to the open reading frame (Ambros et al., 2003b; Lim et al., 2003b).

Small RNAs Derived from *Drosophila C virus* RNA

Drosophila C virus (*DCV*) belongs to the class of *Dicistroviridae* and the genus *Cripavirus* (Cricket paralysis-like virus) and contains a positive-sense genomic RNA of 9264 nucleotides in length (Johnson and Christian, 1998). *DCV* is phylogenetically related to mammalian picornaviruses. *DCV* is commonly associated with *Drosophila* in nature and in laboratory culture. Infections are not usually associated with a noticeable disease state although they commonly reduce life expectancy of infected individuals. We isolated 17 small RNAs from *DCV*, 16 of which were in + orientation and one in − orientation. Considering the broad size distribution of the + orientation *DCV* small RNAs, which is similar to the sense mRNA fragments described above, it is likely that most of the + strand *DCV* small RNAs are not RNase III cleavage products. Although the possibility remains that the 21 nt *DCV* antisense RNA is a breakdown derived from the − strand of DCV, it is also conceivable that the 21 nt antisense *DCV* small RNA is a true siRNA that derived from the dsRNA replication intermediate of *DCV*.

microRNA Identification

We identified 382 clones that were derived from 60 miRNA genes (Table 2). Some of the cloned miRNAs, such as miR-1 through miR-14, let-7, and bantam were described previously (Brennecke et al., 2003; Lagos-Quintana et al., 2001; Pasquinelli et al., 2000). According to the convention for miRNA annotation (Ambros et al., 2003a), new miRNA gene names were assigned based on the evidence of cloning the small RNA from our cDNA libraries and their phylogenetic conservation as a fold-back precursor structure in other species (Supplemental Tables S3 and S4). miRNAs that closely resemble in sequence previously described miRNAs were miR-9b, 9c, 31a, 31b, 34, 79, 92a, 92b, 124, 184, and 210. These were named based on their evolutionary relationships to preidentified miRNAs. Based on the cloned sequences of bantam miRNAs, its previously deduced sequence (Brennecke et al., 2003) needs to be revised. All cloned bantam miRNAs begin with the same 5′ uridine residue encoded one nucleotide downstream of the proposed sequence.

For some miRNAs, we cloned the strand opposite to the accumulating and conserved miRNA. We refer to these sequences as the miR* sequence (Lau et al., 2001) (Table 2). In most cases, miR* is clearly less abundant, but for miR-10*, miR-13a*, and miR-281-2*, too few sequences were cloned to make any conclusion on the relative abundance. A similar situation was found for two small RNAs that are excised from a hairpin residing in the noncoding RNA transcript *iab-4* of the Bithorax complex. *iab-4* contributes to proper formation of abdominal segments (Mattick and Gagen, 2001).

miRNA genes are often found in close proximity to each other forming larger miRNA gene clusters (Lagos-Quintana et al., 2001; Lau et al., 2001). We found 11 gene clusters in the *D. melanogaster* genome containing on average three miRNAs, with the longest cluster containing eight miRNAs (Table 2). Some clusters are found within intergenic regions, while others are located within the intronic regions of protein-coding genes (Supplemental Table S3). Some clusters contain only highly homologous miRNAs, such as *mir-92a* and *mir-92b* or *mir-281-1* and *mir-281-2*, suggestive of rather recent gene duplication. For example, the *mir-2*/*mir-13* family is composed of eight genes that are encoded in four different genomic regions in *D. melanogaster*. There are two single copy genes and two clusters containing three gene copies each. In *Anopheles gambiae*, only five members of this gene family are found, and they are all within one cluster.

Because positional clustering of miRNAs is a common genomic feature of miRNAs, we examined the regions adjacent to miRNA genes for the presence of additional miRNAs that may have escaped the nonsaturating cloning and sequencing protocol. By identifying fold-back structures as well as sequence homologs of known miRNAs and only considering those candidates that were also conserved in other insect genomes, we identified eight more miRNAs. These include *mir-283* clustered with *mir-12* and *mir-304*, *mir-100* clustered with *let-7* and *mir-125*, *mir-313* and *mir-310* clustered with *mir-311* and *mir-312*, and *mir-2c* clustered with *mir-13a* and *mir-13b*. Nonclustered conserved fold-back structures were identified for *mir-87* and *mir-133*. The expression of miR-87, -100, -125, -133, -283, -310, and -312 was confirmed by Northern analysis (Figure 2). The validation of the predicted miR-2c was not attempted because of predictable problems of cross-hybridization of the Northern blotting probe to miR-2a and miR-2b. To date, including the predicted clustered and/or conserved miRNAs, a total of 62 unique miRNA sequences encoded by 68 genes have been identified in *D. melanogaster*.

Table 2. MicroRNA Profile During *D. melanogaster* Development

Gene	Mature miRNA and miRNA* Sequence	Size Range (nt)	Chr	Position, Orientation	Embryo (hr)					Larva			A	T	Total
					0–2	2–4	4–6	6–12	12–24	1	1 + 2	3			
mir-303	UUUAGGUUUCACAGGAAACUGGU	23	X	4106836…58,−										1	1
mir-31b	UGGCAAGAUGUCGGAAUAGCUG	22	X	8732382…403,−									1		1
mir-13b-2§	UAUCACAGCCAUUUUGACGAGU	17–24	X	8830259…80,+	3					1					4
mir-283[p,n]	UAAAUAUCAGCUGGUAAUUCU	21[p]	X	15238989…9009,+											
mir-304	UAAUCUCAAUUUGUAAAUGUGAG	21–23	X	15239984…40006,+	2								1	4	7
mir-12	UGAGUAUUACAUCAGGUACUGGU	21–24	X	15240493…515,+	5		3						2	3	13
mir-210	UUGUGCGUGUGACAGCGGCU	20–23	X	17859230…49,+									3		3
mir-275	UCAGGUACCUGAAGUAGCGCGCG	19–25	2L	7418084…106,+	4	1				1					6
mir-305	AUUGUACUUCAUCAGGUGCUCUG	19–23	2L	7418210…32,+	1								1		2
mir-2b-1§	UAUCACAGCCAGCUUUGAGGAGCG	19–24	2L	8250853…76,−	4	2		1					5		12
mir-87[p,n]	UUGAGCAAAAUUUCAGGUGUG	21	2L	9942846…65,−											
mir-9c	UCUUUGGUAUUCUAGCUGUAGA	22	2L	16676020…41,+	1	1			1						3
mir-306	UCAGGUACUUAGUGACUCUCAA	22	2L	16676502…22,+									1	1	2
	*GGGGGUCACUCUGUGCCUGUGC	22	2L	16676539…60,+		1									1
mir-79	UAAAGCUAGAUUACCAAAGCAU	22	2L	16676697…718,+	1										1
mir-9b	UCUUUGGUGAUUUUAGCUGUAUG	21–23	2L	16676842…64,+	1	1									2
mir-124[n]	UAAGGCACGCGGUGAAUGCCAAG	19–23	2L	17544509…31,+			2			2		1	1		6
mir-100[p,n]	AACCCGUAAAUCCGAACUUGUG	22	2L	18449528…49,+											
let-7[n]	UGAGGUAGUAGGUUGUAUAGU	20–21	2L	18450127…147,+									3	2	5
mir-125[p,n]	UCCCUGAGACCCUAACUUGUGA	22	2L	18450426…47,+											
mir-2a-2[n]	UAUCACAGCCAGCUUUGAUGAGCU	20–24	2L	19547578…601,−	3	3	2			1			4	2	15
	*CCUCAUCAAGUGGUUGUGAUA	21	2L	19547617…37,−	2								2		4
mir-2a-1[n]	UAUCACAGCCAGCUUUGAUGAGCU	20–24	2L	19547984…8007,−	3	3	2			1			4	2	15
mir-2b-2§	UAUCACAGCCAGCUUUGAGGAGCG	19–24	2L	19548271…94,−	4	2		1					5		12
mir-1[n]	UGGAAUGUAAAGAAGUAUGGAG	20–27	2L	20457238…59,+	23	1	4			4	2	3	34	3	74
mir-133[p,n]	UUGGUCCCCUUCAACCAGCUGU	22	2L	20586415…36,+											
mir-14	UCAGUCUUUUUCUCUCUCCUA	19–25	2R	4614434…54,+	2		2								4
mir-307	UCACAACCUCCUUGAGUGAG	20	2R	4681896…915,−									1		1
mir-281-2§	UGUCAUGGAAUUGCUCUCUUUGU	23	2R	7234876…98,−									1		1
	*AAGAGAGCUAUCCGUCGACAGU	22	2R	7234922…43,−									1		1
mir-281-1§	UGUCAUGGAAUUGCUCUCUUUGU	23	2R	7235096…118,−									1		1
mir-184	UGGACGGAGAACUGAUAAGGGC	19–23	2R	8394134…55,−	5	2	1		1	1			12	4	26
	*CCUUAUCAUUCUCUCGCCCCG	21	2R	8394174…94,−									1		1
mir-308	AAUCACAGGAUUAUACUGUGAG	18–22	2R	9279528…49,−	4	1							1		6
mir-278[n]	UCGGUGGGACUUUCGUCCGUUU	20–22	2R	10720847…68,+								1	1	6	8
mir-8[n]	*AUCUUACCGGGCAGCAUUAGA	21	2R	11895173…93,+									1		1
	UAAUACUGUCAGGUAAAGAUGUC	17–25	2R	11895209…31,+	2	1	1			1	1	2	5	1	14
mir-31a	UGGCAAGAUGUCGGCAUAGCUGA	22–23	2R	12847576…98,−	1						1				2
mir-6-3§[n]	UAUCACAGUGGCUGUUCUUUUU	21–22	2R	14724442…63,−	9	1									10
mir-6-2§[n]	UAUCACAGUGGCUGUUCUUUUU	21–22	2R	14724594…615,−	9	1									10
mir-6-1§[n]	UAUCACAGUGGCUGUUCUUUUU	21–22	2R	14724732…53,−	9	1									10

(*continued*)

Table 2. Continued

Gene	Mature miRNA and miRNA* Sequence	Size Range (nt)	Chr	Position, Orientation	Embryo (hr) 0–2	2–4	4–6	6–12	12–24	Larva 1	1 + 2	3	A	T	Total
mir-5	AAAGGAACGAUCGUUGUGAUAUG	22–23	2R	14724920…42,−	3	5	1								9
mir-4[n]	AUAAAGCUAGACAACCAUUGAA	21–22	2R	14725020…41,−	5	2									7
	*CUUUGGUCGUCCAGCCUUAGGUGA	24–25	2R	14725053…76,−		2									2
mir-286	UGACUAGACCGAACACUCGUGCU	22–24	2R	14725155…77,−	1	7	3	2					1		14
mir-3[n]	UCACUGGGCAAAGUGUGUCUCA	18–22	2R	14725320…41,−	7	2									9
mir-309	GCACUGGGUAAAGUUUGUCCUA	22	2R	14725430…51,−		2									2
mir-310[p,n]	UAUUGCACACUUCCCGGCCUUU	22	2R	15647477…99,−											
mir-311[n]	UAUUGCACAUUCACCGGCCUGA	22–23	2R	15647600…21,−	1	1									2
mir-312[n]	UAUUGCACUUGAGACGGCCUGA	22–23	2R	15647770…91,−	2	1									3
mir-313[p]	UAUUGCACUUUUCACAGCCCGA	22	2R	15647905…27,−											
mir-7[n]	UGGAAGACUAGUGAUUUUGUUGUU	21–24	2R	15669793…816,+	3	1	1								5
bantam	UGAGAUCAUUUUGAAAGCUGAUU	20–23	3L	622888…910,+			2			1	2		3		8
mir-276b	UAGGAACUUAAUACCGUGCUCU	22	3L	10277375…96,+									1		1
mir-276a	UAGGAACUUCAUACCGUGCUCU	22	3L	10322810…31,+				2					6		8
mir-314	UAUUCGAGCCAAUAAGUUCGG	21	3L	11730629…49,+	1								2	1	4
mir-285	UAGCACCAUUCGAAAUCAGUGC	22	3L	11903655…76,−									1		1
mir-263b	CUUGGCACUGGGAGAAUUCAC	21	3L	15792339…59,−									1		1
mir-315	UUUUGAUUGUUGCUCAGAAAGC	22	3L	18809861…82,+					1						1
mir-9a[n]	UCUUUGGUUAUCUAGCUGUAUGA	21–23	3L	19515097…119,+	5	1	2	2			1			1	12
mir-316	UGUCUUUUUCCGCUUACUGGCG	20–22	3L	21586044…65,−								1		1	2
mir-10[n]	*CAAAUUCGGUUCUAGAGAGGUUU	23	3R	2635235…57,−							1				1
	ACCCUGUAGAUCCGAAUUUGU	21	3R	2635277…97,−	1										1
mir-317	UGAACACAGCUGGUGGUAUCCAGU	20–24	3R	5916921…44,+									5	3	8
mir-277	UAAAUGCACUAUCUGGUACGACA	21–23	3R	5925820…42,+									13		13
mir-34	UGGCAGUGUGGUUAGCUGGUUGUG	17–25	3R	5926692…715,+									6	7	13
mir-318	UCACUGGGCUUUGUUUAUCUCA	22	3R	6234084…105,+									3		3
mir-13b-1§	UAUCACAGCCAUUUUGACGAGU	17–24	3R	11243152…73,−	3					1					4
mir-13a	UAUCACAGCCAUUUUGAUGAGU	22	3R	11243287…308,−											
	*CUCCUCAAAGGGUUGUGAAAUG	22	3R	11243328…49,−		1									1
mir-2c[p]	UAUCACAGCCAGCUUUGAUGGGC	22	3R	11243500…23,−											
mir-iab-4-5p	ACGUAUACUGAAUGUAUCCUGA	22	3R	12682018…39,+		1									1
mir-iab-4-3p	CGGUAUACCUUCAGUAUACGUAAC	24	3R	12682052…75,+			1								1
mir-11	AUCACAGUCUGAGUUCUUGC	20–23	3R	17439190…209,−	6	1									7
mir-92a	CAUUGCACUUGUCCCGGCCUAU	22–25	3R	21461647…68,+	2	1									3
mir-92b	AAUUGCACUAGUCCCGGCCUGC	22	3R	21466544…65,+	1										1
mir-279	UGACUAGAUCCACACUCAUUAA	22	3R	25030739…60,+	1										1

Clusters of miRNAs are represented by vertical bars between the columns that indicate the chromosome location. An asterisk is used to denote small RNAs that are derived from the strand opposite to the miRNA strand within the fold-back precursor. For the iab-4 derived miRNAs, 5p and 3p indicate 5′ and 3′ location within the conserved fold-back sequence of the iab-4 transcript.

§More copies of this miRNA are found in the genome and the clone numbers indicated cannot be assigned to a unique locus.

[p]Predicted miRNA based on phylogenetic conservation or vicinity to other clustered miRNAs.

[n]Expression of miRNA was also confirmed by Northern blotting (Figures 1 and 2).

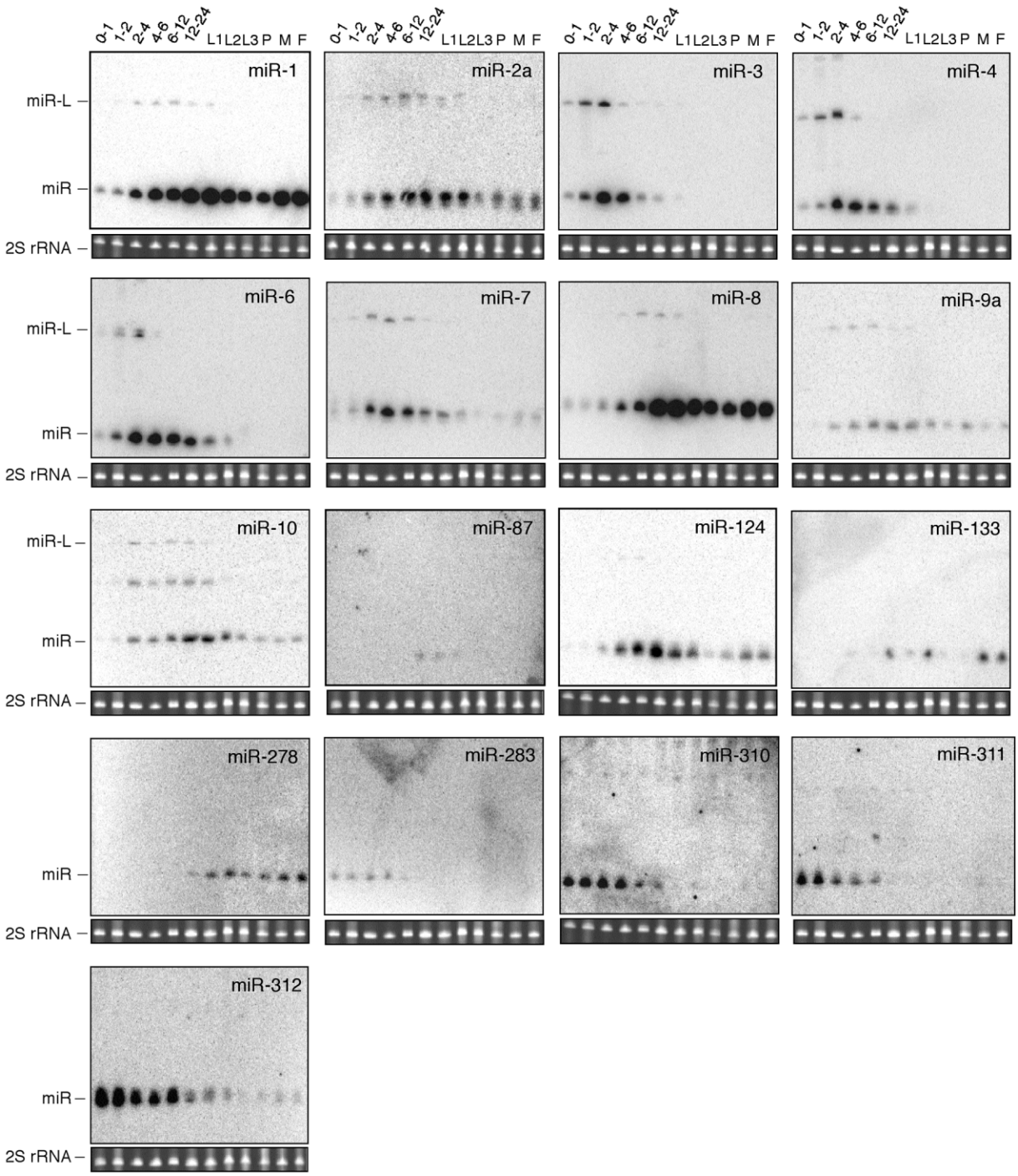


Figure 2. Expression Profiles of a Subset of Identified miRNA Genes

Northern blotting was performed as described in the legend to Figure 1B.

Although we did not clone miR-125, the lin-4 homolog of fly, we previously showed that *mir-125* was expressed in pupa and adult (Lagos-Quintana et al., 2002). As expected from their close proximity, the expression pattern of *mir-125* was identical to that of *D. melanogaster let-7* (Pasquinelli et al., 2000). A third miRNA gene that is homologous to mammalian *mir-100* was found in close proximity to *mir-125* and *let-7* (Figure 1A). We confirmed that *D. melanogaster* miR-100 is coexpressed with miR-125 and let-7 by Northern blotting (Figure 1B). Coexpression of miR-125 and let-7 and induction by the steroid hormone ecdysone at the onset of metamorphosis has been reported recently (Bashirullah et al., 2003; Sempere et al., 2003), as well as the coregulated expression of miR-100 (Sempere et al., 2003). The polycistronic expression of this miRNA cluster was also confirmed by

RT-PCR analysis, which detected a long primary transcript comprising all the hairpin precursor sequences (A.A., unpublished data). The clustering for *mir-100*, *mir-125*, and *let-7* is conserved in *A. gambiae*, although the distance between the genes has increased (Figure 1C). In mouse and human, the gene cluster underwent duplication and the distance between miRNA genes was increased further (Figure 1C). In *C. elegans*, neither *lin-4* nor its paralog *mir-237* are spatially linked with *let-7* family members (Lim et al., 2003b). The *lin-4* gene is expressed at an earlier stage than the *let-7* gene (Reinhart et al., 2000), but *mir-237* shows a very similar expression pattern to that of *let-7* (Lim et al., 2003b). Evidence further supporting the conservation of *lin-4* and *let-7* coregulation in nematodes comes from the identification of both lin-4 and let-7 RNA binding sites in the 3′ UTR of *lin-14*, *lin-28*, *lin-41*, and *lin-57* transcripts (Abrahante et al., 2003; Lin et al., 2003; Reinhart et al., 2000). Homologs to miR-100 were not identified in *C. elegans*.

The 5′ portion of the miRNAs shows the highest degree of conservation and miRNAs can be grouped in families based on this criterion (Ambros et al., 2003b; Lim et al., 2003b). It was also noticed that the 5′ ends of some miRNAs are complementary to sequence motifs in the 3′ UTR of some mRNA that are subject to posttranscriptional regulation (Lai, 2002). It was suggested that these motifs represent miRNA binding sites, which has yet to be confirmed. The 62 unique miRNA sequences can be grouped into 43 families, 14 of which are universally conserved (Supplemental Table S3). For three families, the conservation extends to mammals only, and for six families the conservation extends to nematodes only. This indicates that *D. melanogaster* is a good model system to bridge the gap for studying miRNA function in mammals and nematodes.

Repeat-Associated Small Interfering RNAs

A surprisingly large number of small RNAs (178 clones) map to repetitive sequence elements of the *D. melanogaster* genome. The relative ratio of miRNAs to rasiRNAs varies as a function of development. Whereas about equal proportions of miRNAs to rasiRNAs are found in the early embryo, the fraction of rasiRNAs drops significantly during the transition to adults, in which only about 10% of the small regulatory RNAs are represented as rasiRNAs. The content of rasiRNAs in microdissected testes is comparable to that of early embryos. Together, these findings suggest that the taming of transposable elements and the establishment of chromatin structure may be initiated in germline tissue and during early embryonic development in an RNA-dependent manner.

The rasiRNAs contain sequences from all known forms of repetitive sequence elements, such as retrotransposons, DNA transposons, satellite, and microsatellite DNA sequences, complex repeats such as the *Su(Ste)* locus, as well as vaguely characterized repetitive sequence motifs (Table 3). The transposable elements are categorized by their mode of transposition and fall into two major subgroups. The class I elements transpose via an RNA intermediate, while the class II elements transpose through a DNA intermediate. Class I elements are again subdivided into long-terminal repeat (LTR) retrotransposons and non-LTR transposons, also referred to as long-interspersed nuclear elements (LINE) or poly(A)-type retrotransposons. Class II elements are characterized by a terminal-inverted repeat (TIR). We have cloned small RNAs from 38 different transposable elements corresponding to 40% of all known transposable elements in *D. melanogaster* (Kaminker et al., 2002). The most frequently cloned rasiRNAs were derived from the LTR transposon *roo*, which is also the most abundant transposable element in the euchromatic portion of the *D. melanogaster* genome. This indicates a role for RNAi in controlling the mobility of transposable elements in *D. melanogaster*.

The production of small sense and antisense RNAs from the repetitive *Su(Ste)* locus sequences was shown previously (Aravin et al., 2001), indicating a role for dsRNA in the homology-dependent regulation of *Stellate*. In agreement with this, we cloned several siRNAs in sense and antisense orientation from the *Su(Ste)* locus in testes tissue. The identification of rasiRNAs from other not yet annotated repetitive sequences suggests that systems similar to that of *Ste/Su(Ste)* might exist in *D. melanogaster*. Therefore, small RNA cloning is a viable strategy for identification of such cases. A complete list of all cloned repeat-associated sequences is provided in Supplemental Table S5.

Some repeat sequences such as HeT-A, TART, and subterminal minisatellites are restricted to telomeres and required for telomere maintenance. These elements are absent from euchromatic regions (Kaminker et al., 2002) and therefore link small RNAs to regulation of heterochromatin. Small RNAs of the GAGAA microsatellite family were cloned from testes and originated probably from transcripts of the fully heterochromatic Y chromosome (Lohe et al., 1993).

Another group of 10 rasiRNAs could be assigned to a 200 kb region from chromosome 2R (band 42AB), which is composed of numerous different types of transposable elements, many of which are damaged and have diverged from canonical sequences. There are no predicted genes in this region, but several lethal mutations have been described that map to this region (Bender, 1996). The reasons for this lethality are unclear, but it is conceivable that this region is an important source for small RNAs involved in some aspect of heterochromatin regulation.

Discussion

Two Size Classes of Small RNAs in *D. melanogaster*

Here we describe the small RNA profile of the various stages of *D. melanogaster* development. Similar to plants (Hamilton et al., 2002; Llave et al., 2002a), *D. melanogaster* produces two distinct classes of small RNAs from dsRNA precursors. The two predominant classes of small RNAs are 21 and 22 nt miRNAs and 24–26 nt rasiRNAs. The strong bias for uridine at the 5′ termini and the match to repetitive elements in sense and antisense orientation is indicative of dsRNA-specific RNase III processing (Bernstein et al., 2001) and suggests a dsRNA origin for rasiRNAs. We also identified one antisense siRNA derived from the common *Drosophila C virus*. The propagation of the positive RNA strand *Flock House*

Table 3. Repeat-Associated Small Interfering RNA Profile During *D. melanogaster* Development

Repeat	Name	Size Range (nt)	Embryo (hr)				L1–L3	P	A	T	Total	Orientation
			0–2	2–4	4–6	6–24						
LINE	aurora-element	22–25			1		1		1	1	4	3(+)/1(−)
	baggins	24	1	1							2	2(+)/0(−)
	BS	22–25	1							1	2	1(+)/1(−)
	Cr1a	22	1								1	1(+)/0(−)
	Doc3-element	23	1								1	0(+)/1(−)
	F-element	24–27			1				1	1	3	0(+)/3(−)
	G-element	24								1	1	0(+)/1(−)
	HeT-A element	24–25	2	1							3	2(+)/1(−)
	Ivk	25		1							1	0(+)/1(−)
	R1-element	24–29	1	1						1	3	1(+)/2(−)
	Rt1a	23–27	4	2							6	0(+)/6(−)
	Rt1b	22–27	2	3	1		1				7	2(+)/5(−)
	Rt1c, diver	16		1							1	0(+)/1(−)
	TART-element	21–27	2	3					1		6	3(+)/3(−)
	X-element	21–24	2							1	3	0(+)/3(−)
LTR	17.6	22–24	2			1					3	0(+)/3(−)
	1731	19–26		1						1	2	1(+)/1(−)
	3S18	25		1							1	1(+)/0(−)
	412	27			1						1	0(+)/1(−)
	accord	28		1							1	0(+)/1(−)
	blood	26			1						1	0(+)/1(−)
	copia	24–25	1							2	3	2(+)/1(−)
	GATE	24–26	1		1						2	2(+)/0(−)
	gypsy3, springer	23		1							1	0(+)/1(−)
	HMS-Beagle	23–26		1					1		2	0(+)/2(−)
	invader1	20–23	2						1		3	2(+)/1(−)
	invader4	21	1								1	0(+)/1(−)
	Max-element	19–27	6	2							8	4(+)/4(−)
	McClintock	26		1							1	0(+)/1(−)
	mdg3	26	1								1	1(+)/0(−)
	opus	19–25	2						1		3	2(+)/1(−)
	Quasimodo	21	1								1	0(+)/1(−)
	roo	18–27	16	5	1				1		23	4(+)/19(−)
	rooA	24	1								1	0(+)/1(−)
	Stalker	16–24			1				1		2	0(+)/2(−)
TIR	1360	17–25	1							3	4	4(+)/0(−)
	hopper2	24–26	3								3	3(+)/0(−)
	transib4	26			1						1	0(+)/1(−)
Su(Ste)	Su(Ste)	23–26								5	5	1(+)/4(−)
Other Heterochromatic Repeats	1.688 satellite DNA	19–28	2				2				4	undef.
	microsatellite	17–24								7	7	undef.
	subterminal minisatellite at telomere of 2L	24		1							1	undef.
	repeat on chr. X	21	1								1	undef.
	42 AB region on chr. 2R	23–28	6	3	1						10	undef.
	unspecified	16–28	15	9	3		1	1	4	4	37	undef.

The developmental stages are larva 1, L1; larva 2, L2; larva 3, L3; pupa, P; adult, A. and testes, T. Transposable elements are abbreviated as LINE, long interspersed nuclear element, LTR, long terminal repeat retrotransposons, and TIR, terminal inverted repeat DNA transposons. The orientation in + indicates the same orientation as the open reading frame within the transposable element. The orientation in *Su(Ste)* is given with respect to the coding sequence of *stellate*.

virus (*FHV*) in *D. melanogaster S2* cells is associated with the formation of 21 and 22 nt siRNA (Li et al., 2002), further supporting that RNA viruses can elicit an RNAi response. The 21 and 22 nt class of small RNAs (siRNAs and miRNAs) is associated with mRNA degradation and translational regulation, and both small RNAs interact with common components of the gene-silencing machinery (Doench et al., 2003; Hutvágner and Zamore, 2002). In plants, the longer class of small RNAs has been suggested to mediate repression of retroelements via histone H3 methylation as well as asymmetric DNA methylation (Zilberman et al., 2003) and systemic silencing (Hamilton et al., 2002).

The identification of two size classes of small RNAs in plants and *D. melanogaster* suggests that at least two distinct dsRNA-processing enzymes or enzyme complexes are involved in their production. *Arabidopsis thaliana* encodes at least 4 Dicer-like (dcl) proteins (Schauer et al., 2002). Complete loss of *dcl-1* function is lethal (Golden et al., 2002). Plants homozygous for

the weak loss-of-function allele *dcl-1-9* are strongly impaired in miRNA precursor processing (Park et al., 2002; Reinhart et al., 2002), yet retain activity for mediating RNAi (Finnegan et al., 2003). This suggested that another member of the Dicer family generates siRNAs, although it cannot be excluded that dcl-1 may also be able to generate 21 nt siRNAs (Finnegan et al., 2003). Addition of long siRNAs into wheat germ extracts competitively inhibited the production of long siRNAs from dsRNA (Tang et al., 2003). However, the addition of short siRNAs did not inhibit the production of either long or short siRNAs from dsRNA, indicating the presence of distinct Dicer-like enzymes.

D. melanogaster encodes two members of Dicer, *dcr1* and *dcr2* (Bernstein et al., 2001). In vitro processing of dsRNA in cytoplasmic extracts from 0–2 hr embryos or late embryo-derived *Schneider 2* (*S2*) cells resulted only in the formation of the 21 and 22 nt siRNAs (Bernstein et al., 2001; Zamore et al., 2000), although EST data suggest that both Dicer proteins are expressed in *S2* cells. Similarly, nuclear extracts prepared from *S2* cells only generated 21 and 22 nt siRNAs (M.L.-Q., unpublished data). The presence of long siRNAs was however documented by Northern blotting for *Su(Ste)* derived siRNAs (Aravin et al., 2001). It is possible that cofactors needed for production of longer siRNAs were missing in the embryonic and *S2* cell extracts.

miRNAs

The number of miRNA genes identified in various organisms is still increasing. The miRNA profile of *C. elegans* is probably the most complete, with the number of validated miRNA genes being approximately 95 (Ambros et al., 2003b; Lim et al., 2003b). Independent estimates for the number of miRNA genes in *C. elegans* are 90–120 genes (Lim et al., 2003b) and 140–300 genes (Grad et al., 2003), although the lower estimate is based on more sensitive computational analyses and likely to be more accurate. For *D. melanogaster*, we have identified 62 unique miRNAs that are encoded by 68 genes. Cloning identified 55 of the unique miRNAs. The predicted miRNAs were validated by Northern blotting. Cloning of miRNAs from different developmental stages provided information on the developmental regulation of miRNAs. However, the cloning frequencies were generally not high enough to accurately quantify the level of expression. Therefore, to further our understanding of the developmental expression patterns, we performed Northern blots for a sample set consisting of 20 miRNAs (Figures 1 and 2). The spatial expression pattern of miRNAs was not examined because methods for in situ hybridization of miRNAs remain to be developed.

Intergenic miRNAs are under the control of their own promoter. miRNAs that are localized in introns might be expected, in some cases, to show the same expression pattern as the protein-coding gene, since all 18 of the identified intronic miRNAs are in the same orientation as the protein coding sequence. miR-11 is located in one of the introns of the transcription factor *E2f*, which is predominantly expressed during early development but not in adult (Dynlacht et al., 1994). Consistent with the expression pattern of *E2f*, miR-11 was cloned from early embryos, and Northern blot analysis confirmed that expression was absent from adults (Lagos-Quintana et al., 2001). The cluster including miRNAs 9c, P310, 79, and 9b is distributed over two introns of the protein serine/threonine kinase *grp* that is involved in the developmentally controlled DNA replication/damage checkpoint (Fogarty et al., 1997). It is possible that these miRNAs are also involved in regulating irradiation damage response because the *grp*/miRNA gene structure is conserved in *Drosophila* and *A. gambiae*. The most striking case of intronic conservation is presented by *mir-7*, which is found within the last intron of the *heterogeneous nuclear ribonucleoprotein K* (*hnRNP K*) gene of *Drosophila*, *A. gambiae*, mouse, and human. Overexpression of *hnRNP K* (also referred to as *transformation upregulated nuclear protein* [*TUNP*]) has been observed in cancer cells (Bomsztyk et al., 1997; Dejgaard et al., 1994), but so far the attention focused only on the encoded protein and the role of the encoded miR-7 was ignored. These examples highlight the importance of understanding miRNA gene structure and their genomic distribution and also indicate the need for comprehensive annotation of noncoding RNAs in sequenced genomes.

rasiRNAs

Our set of 178 rasiRNAs represents the largest collection of repeat-associated small RNAs (rasiRNAs) to date. It provides insights into the complexity of RNA-guided regulation of heterochromatin. A comprehensive annotation of rasiRNAs is difficult because methods for reliable assembly of contigs containing highly repetitive sequences are only currently being developed (Hoskins et al., 2002). Many rasiRNAs are derived from classical regions of heterochromatin, such as the telomeres, or they match segments of unassembled genomic sequences classified as heterochromatin. rasiRNAs are very rich in sequences cognate to all classes of transposable elements. By analogy to the plant world, it is presumed that this longer class of siRNAs is also mediating transcriptional regulation. A role for rasiRNAs in systemic silencing can be ruled out because systemic silencing does not exist in *D. melanogaster* (Roignant et al., 2003).

D. melanogaster contains five members of the Argonaute protein family, three of which (*Ago1*, *Ago2, piwi*) have been implicated in mechanisms involving siRNAs and miRNAs (Carmell et al., 2002; Hammond et al., 2001; Pal-Bhadra et al., 2002; Williams and Rubin, 2002). *Ago3*, *aub*, and *piwi* are predominantly expressed in germ cells and the early embryo (Williams and Rubin, 2002). The increased abundance of rasiRNAs in embryo and testes strongly suggests that these Argonaute members are specific binding partners for rasiRNAs. Mutation of *aub* as well as of the RNA helicase *spn-E* leads to a loss of the *Su(Ste)* function (Aravin et al., 2001; Schmidt et al., 1999; Stapleton et al., 2001). In addition, *spn-E* but not *aub* mutants show derepression of some transposable elements in the germline (Aravin et al., 2001). The transcriptional repression of certain multiple-copy transgenes in *D. melanogaster* requires *piwi* (Pal-Bhadra et al., 2002). In fission yeast only a single Argonaute protein has been identified, which is

required for establishment and maintenance of heterochromatin (Hall et al., 2002; Volpe et al., 2002). In addition, some small RNAs have been found in fission yeast that are homologous to these regions (Reinhart and Bartel, 2002). Together with the biochemical evidence for Argonaute proteins as binding partners of siRNAs (Hammond et al., 2001; Martinez et al., 2002), a picture for RNA-guided sequence-specific regulation of chromatin structure is gradually emerging.

In summary, small RNAs are not only involved in guiding RNA degradation and translational control but presumably also contribute to chromatin modification and transcriptional silencing. Recording the small RNA profile will become as important as recording mRNA expression profiles to understand how cells regulate and modulate gene expression. *D. melanogaster* is an excellent model organism for investigating the function of the small RNAs in these processes.

Experimental Procedures

RNA preparation was performed as previously described (Lagos-Quintana et al., 2002). Total RNA was isolated from embryos at different times after egg laying as well as from larval stages 1, 2, and 3 stages, from pupae, adults, and microdissected adult testes. Cloning of miRNAs was performed as described (Pfeffer et al., 2003). Northern blots were stripped and reprobed several times. Before reprobing, it was confirmed that the stripping was complete by phosphorimaging of the stripped membrane.

Acknowledgments

We thank G. Dowe and M. Killian for sequencing, W. Lendeckel for help on cloning, A. Hüttenhofer for providing *D. melanogaster* noncoding RNA sequences to facilitate sequence annotation, Y. Dorsett, M. Landthaler, S. Pfeffer, J. Silverman, and J. Martinez for comments on the manuscript, H. Jäckle, R. Rivera-Pomar, and R. Lührmann for support. This work was funded by a Bundesministerium und Forschung (BMBF) Biofuture grant number 0311856, Program Physico-Chemical Biology RAS, and an EMBO short term fellowship ASTF 9963 to A.A.

Received: June 16, 2003
Revised: June 23, 2003
Accepted: June 27, 2003
Published online: July 2, 2003

References

Abrahante, J.E., Daul, A.L., Li, M., Volk, M.L., Tennessen, J.M., Miller, E.A., and Rougvie, A.E. (2003). The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. Dev. Cell *4*, 625–637.

Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., et al. (2003a). A uniform system for microRNA annotation. RNA *9*, 277–279.

Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003b). MicroRNAs and Other Tiny Endogenous RNAs in C. elegans. Curr. Biol. *13*, 807–818.

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.

Bashirullah, A., Pasquinelli, A.E., Kiger, A.A., Perrimon, N., Ruvkun, G., and Thummel, C.S. (2003). Coordinate regulation of small temporal RNAs at the onset of Drosophila metamorphosis. Dev. Biol. *259*, 1–8.

Bender, M. (1996). New lethal loci in the 42AB region of the Drosophila melanogaster 2nd chromosome. Drosoph. Inf. Serv. *77*, 143.

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

Bomsztyk, K., Van Seuningen, I., Suzuki, H., Denisenko, O., and Ostrowski, J. (1997). Diverse molecular interactions of the hnRNP K protein. FEBS Lett. *403*, 113–115.

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.

Caplen, N.J., Parrish, S., Imani, F., Fire, A., and Morgan, R.A. (2001). Specific inhibition of gene expression by small double-stranded RNAs in invertebrate and vertebrate systems. Proc. Natl. Acad. Sci. USA *98*, 9742–9747.

Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev. *16*, 2733–2742.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. *162*, 156–159.

Dejgaard, K., Leffers, H., Rasmussen, H.H., Madsen, P., Kruse, T.A., Gesser, B., Nielsen, H., and Celis, J.E. (1994). Identification, molecular cloning, expression and chromosome mapping of a family of transformation upregulated hnRNP-K proteins derived by alternative splicing. J. Mol. Biol. *236*, 33–48.

Denli, A.M., and Hannon, G.J. (2003). RNAi: an ever-growing puzzle. Trends Biochem. Sci. *28*, 196–201.

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.

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

Dostie, J., Mourelatos, Z., Yang, M., Sharma, A., and Dreyfuss, G. (2003). Numerous microRNPs in neuronal cells containing novel microRNAs. RNA *9*, 180–186.

Dynlacht, B.D., Brook, A., Dembski, M., Yenush, L., and Dyson, N. (1994). DNA-binding and trans-activation properties of Drosophila E2F and DP proteins. Proc. Natl. Acad. Sci. USA *91*, 6359–6363.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001a). Duplexes of 21-nucleotide RNAs mediate RNA interference in mammalian cell culture. Nature *411*, 494–498.

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

Finnegan, E.J., Margis, R., and Waterhouse, P.M. (2003). Posttranscriptional gene silencing is not compromised in the Arabidopsis CARPEL FACTORY (DICER-LIKE1) mutant, a homolog of Dicer-1 from Drosophila. Curr. Biol. *13*, 236–240.

Fogarty, P., Campbell, S.D., Abu-Shumays, R., Phalle, B.S., Yu, K.R., Uy, G.L., Goldberg, M.L., and Sullivan, W. (1997). The Drosophila grapes gene is related to checkpoint gene chk1/rad27 and is required for late syncytial division fidelity. Curr. Biol. *7*, 418–426.

Golden, T.A., Schauer, S.E., Lang, J.D., Pien, S., Mushegian, A.R., Grossniklaus, U., Meinke, D.W., and Ray, A. (2002). SHORT INTEGUMENTS1/SUSPENSOR1/CARPEL FACTORY, a Dicer homolog, is a maternal effect gene required for embryo development in Arabidopsis. Plant Physiol. *130*, 808–822.

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.

Grishok, A., and Mello, C.C. (2002). RNAi (Nematodes: Caenorhabditis elegans). Adv. Genet. *46*, 339–360.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the

small temporal RNAs that control *C. elegans* developmental timing. Cell *106*, 23–34.

Hall, I.M., Shankaranarayana, G.D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S.I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232–2237.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J. *21*, 4671–4679.

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

Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146–1150.

Hoskins, R.A., Smith, C.D., Carlson, J.W., Carvalho, A.B., Halpern, A., Kaminker, J.S., Kennedy, C., Mungall, C.J., Sullivan, B.A., Sutton, G.G., et al. (2002). Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. Genome Biol. *3*, RESEARCH0085.1–0085.16.

Hutvágner, G., McLachlan, J., Bálint, É., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA interference enzyme Dicer in small temporal RNA maturation. Science *93*, 834–838.

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

Johnson, K.N., and Christian, P.D. (1998). The novel genome organization of the insect picorna-like virus Drosophila C virus suggests this virus belongs to a previously undescribed virus family. J. Gen. Virol. *79*, 191–203.

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., et al. (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol. *3*, RESEARCH0084.1–0084.20.

Kennerdell, J.R., Yamaguchi, S., and Carthew, R.W. (2002). RNAi is activated during Drosophila oocyte maturation in a manner dependent on aubergine and spindle-E. Genes Dev. *16*, 1884–1889.

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

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. Cell *99*, 133–141.

Kogan, G.L., Tulin, A.V., Aravin, A.A., Abramov, Y.A., Kalmykova, A.I., Maisonhaute, C., and Gvozdev, V.A. (2003). The GATEretrotransposon in Drosophila melanogaster: mobility in heterochromatin and aspects of its expression in germline tissues. Mol. Genet. Genomics *269*, 234–242.

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

Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. (2003). New microRNAs from mouse and human. RNA *9*, 175–179.

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.

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

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.

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

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

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

Li, H., Li, W.X., and Ding, S.W. (2002). Induction and suppression of RNA silencing by an animal virus. Science *296*, 1319–1321.

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

Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003b). The microRNAs of *Caenorhabditis elegans*. Genes Dev. *17*, 991–1008.

Lin, S.Y., Johnson, S.M., Abraham, M., Vella, M.C., Pasquinelli, A., Gamberi, C., Gottlieb, E., and Slack, F.J. (2003). The C. elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. Dev. Cell *4*, 639–650.

Livak, K.J. (1984). Organization and mapping of a sequence on the Drosophila melanogaster X and Y chromosomes that is transcribed during spermatogenesis. Genetics *107*, 611–634.

Livak, K.J. (1990). Detailed structure of the Drosophila melanogaster stellate genes and their transcripts. Genetics *124*, 303–316.

Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002a). Endogenous and silencing-associated small RNAs in plants. Plant Cell *14*, 1605–1619.

Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002b). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053–2056.

Lohe, A.R., Hilliker, A.J., and Roberts, P.A. (1993). Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. Genetics *134*, 1149–1174.

Mallory, A.C., Reinhart, B.J., Bartel, D., Vance, V.B., and Bowman, L.H. (2002). A viral suppressor of RNA silencing differentially regulates the accumulation of short interfering RNAs and micro-RNAs in tobacco. Proc. Natl. Acad. Sci. USA *99*, 15228–15233.

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

Mattick, J.S., and Gagen, M.J. (2001). The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. Mol. Biol. Evol. *18*, 1611–1630.

Merino, E., Balbas, P., Puente, J.L., and Bolivar, F. (1994). Antisense overlapping open reading frames in genes from bacteria to humans. Nucleic Acids Res. *22*, 1903–1908.

Mette, M.F., van der Winden, J., Matzke, M., and Matzke, A.J. (2002). Short RNAs can identify new candidate transposable element families in Arabidopsis. Plant Physiol. *130*, 6–9.

Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell *110*, 689–699.

Moss, E.G. (2002). MicroRNAs: hidden in the genome. Curr. Biol. *12*, R138–140.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. *16*, 720–728.

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

Olsen, P.H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Dev. Biol. *216*, 671–680.

Pal-Bhadra, M., Bhadra, U., and Birchler, J.A. (2002). RNAi related mechanism affect both transcriptional and posttranscriptional transgene silencing in *Drosophila*. Mol. Cell *9*, 315–327.

Palumbo, G., Bonaccorsi, S., Robbins, L.G., and Pimpinelli, S. (1994). Genetic analysis of Stellate elements of Drosophila melanogaster. Genetics *138*, 1181–1197.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL

FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Curr. Biol. *12*.

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

Pasquinelli, A.E., and Ruvkun, G. (2002). Control and developmental timing by microRNAs and their targets. Annu. Rev. Cell Dev. Biol. *18*, 495–513.

Pfeffer, S., Lagos-Quintana, M., and Tuschl, T. (2003). Cloning of small RNA molecules. In Current Protocols in Molecular Biology, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidmann, J.A. Smith, and K. Struhl, eds. (New York: Wiley and Sons), 26.24.21–26.24.16.

Plasterk, R.H. (2002). RNA silencing: the genome's immune system. Science *296*, 1263–1265.

Provost, P., Dishart, D., Doucet, J., Frendewey, D., Samuelsson, B., and Radmark, O. (2002). Ribonuclease activity and RNA binding of recombinant human Dicer. EMBO J. *21*, 5864–5874.

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

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901–906.

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

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

Roignant, J.Y., Carre, C., Mugat, B., Szymczak, D., Lepesant, J.A., and Antoniewski, C. (2003). Absence of transitive and systemic pathways allows cell-specific and isoform-specific RNAi in Drosophila. RNA *9*, 299–308.

Schauer, S.E., Jacobsen, S.E., Meinke, D.W., and Ray, A. (2002). DICER-LIKE1: blind men and elephants in Arabidopsis development. Trends Plant Sci. *7*, 487–491.

Schmidt, A., Palumbo, G., Bozzetti, M.P., Tritto, P., Pimpinelli, S., and Schafer, U. (1999). Genetic and molecular characterization of sting, a gene involved in crystal formation and meiotic drive in the male germ line of Drosophila melanogaster. Genetics *151*, 749–760.

Seggerson, K., Tang, L., and Moss, E.G. (2002). Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation. Dev. Biol. *243*, 215–225.

Sempere, L.F., Sokol, N.S., Dubrovsky, E.B., Berger, E.M., and Ambros, V. (2003). Temporal regulation of microRNA expression in Drosophila melanogaster mediated by hormonal signals and broad-Complex gene activity. Dev. Biol. *259*, 9–18.

Shendure, J., and Church, G.M. (2002). Computational discovery of sense-antisense transcription in the human and mouse genomes. Genome Biol. *3*, RESEARCH0044.1–0044.14.

Stapleton, W., Das, S., and McKee, B.D. (2001). A role of the Drosophila homeless gene in repression of Stellate in male meiosis. Chromosoma *110*, 228–240.

Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A., and Mello, C.C. (1999). The rde-1 gene, RNA interference, and transposon silencing in *C. elegans*. Cell *99*, 123–132.

Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev. *17*, 49–63.

Tautz, D., Hancock, J.M., Webb, D.A., Tautz, C., and Dover, G.A. (1988). Complete sequences of the rRNA genes of Drosophila melanogaster. Mol. Biol. Evol. *5*, 366–376.

Taverna, S.D., Coyne, R.S., and Allis, C.D. (2002). Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. Cell *110*, 701–711.

Voinnet, O. (2002). RNA silencing: small RNAs as ubiquitous regulators of gene expression. Curr. Opin. Plant Biol. *5*, 444–451.

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.

Wassenegger, M. (2002). Gene silencing-based disease resistance. Transgenic Res. *11*, 639–653.

Waterhouse, P.M., Wang, M.B., and Lough, T. (2001). Gene silencing as an adaptive defence against viruses. Nature *411*, 834–842.

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.

Williams, R.W., and Rubin, G.M. (2002). ARGONAUTE1 is required for efficient RNA interference in *Drosophila* embryos. Proc. Natl. Acad. Sci. USA *99*, 6889–6894.

Wu-Scharf, D., Jeong, B., Zhang, C., and Cerutti, H. (2000). Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-Box RNA helicase. Science *290*, 1159–1163.

Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. (2003). The Drosophila microRNA Mir-14 suppresses cell death and is required for normal fat metabolism. Curr. Biol. *13*, 790–795.

Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., et al. (2003). Widespread occurrence of antisense transcription in the human genome. Nat. Biotechnol. *21*, 379–386.

Yuan, G., Klambt, C., Bachellerie, J.P., Brosius, J., and Hüttenhofer, A. (2003). RNomics in Drosophila melanogaster: identification of 66 candidates for novel non-messenger RNAs. Nucleic Acids Res. *31*, 2495–2507.

Zamore, P.D. (2002). Ancient pathways programmed by small RNAs. Science *296*, 1265–1269.

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.

Zhang, H., Kolb, F.A., Brondani, V., Billy, E., and Filipowicz, W. (2002). Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. EMBO J. *21*, 5875–5885.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716–719.

Note Added in Proof

A recent study computationally identified numerous *D. melanogaster* miRNAs and used Northern blotting for validation (Lai, E.C., Tomancak, P., Williams, R.W., and Rubin, G.M. [2003]. Computational identification of *Drosophila* microRNA genes. Genome Biol. *4*, R42. Published online: June 30, 2003). Thirty-two previously unknown miRNA genes were identified in common, eight novel miRNA genes were not found in our study, and twelve novel miRNAs were absent from the prediction list. The 5′ ends of 13 commonly identified miRNAs were misidentified in the prediction between +5 and −2 nucleotides according to cloned sequences. The miRNAs derived from the *mir-281-1* and *mir-281-2* gene cluster were predicted antisense to cloned sequences, and miR-276a and miR-276b were predicted from the strand opposite to the cloned sequences. The curated set of 76 miRNA genes will become available at http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml.

Chapter IV: Optimization and validation of the small RNA cloning protocol

Background

Small RNAs are key components of dsRNA-dependent gene regulation. Cloning and sequencing them has been crucial for identifying miRNAs. Efficient methods for small RNA isolation, cDNA library preparation and high-throughput sequencing are needed in order to comprehensively characterize small RNA profiles. A previously developed method (Pfeffer et al., 2003) used directional ligation of 3' and 5' adapter oligonucleotides to small RNAs, followed by reverse transcription. One of the disadvantages of this protocol was the circularization of small RNAs by T4 RNA ligase, competing with transligation of small RNAs in the sample to the 3' adapter.

RNA ligases join 3' hydroxyl and 5' phosphate RNA termini via a series of three nucleotidyl transfer steps. First, the ligase reacts with ATP and forms a covalent ligase-(lysyl-N)-AMP intermediate releasing pyrophosphate. Second, AMP is transferred from the enzyme to the 5' phosphate end of the RNA to form an RNA-adenylate intermediate (AppRNA). Third, the AppRNA reacts with the 3' hydroxyl RNA end to form a 3'-5' phosphodiester bond and release AMP (Cranston et al., 1974; Krug and Uhlenbeck, 1982)

In vivo, T4 RNA 1 ligase (Rnl1) repairs a break in the anticodon loop of *Escherichia coli* $tRNA^{lys}$ triggered by phage activation of the host-encoded anticodon nuclease PrrC (Amitsur et al., 1987). The cleavage activity of PrrC depleats $tRNA^{lys}$ to halt phage protein synthesis and arrest infection. Bacteriophage T4 counteracts the cleavage of the RNA with the repair enzymes polynucleotide kinase, phosphatase and Rnl1 to seal the broken tRNA.

The efficiency of the overall ligation reaction by T4 RNA ligase is dependent on the nature of both the donor and the acceptor oligonucleotide (England and Uhlenbeck, 1978; Romaniuk et al., 1982). Secondary structure, as well as variations in sequence, have effects on the reaction efficiency (Bruce and Uhlenbeck, 1978; England and Uhlenbeck, 1978; Romaniuk et al., 1982)

Bacteriophage T4 also encodes T4 RNA ligase 2 (Rnl2), which belongs to a different family of ligases that includes the *Trypanosoma* and *Leshmania* RNA editing ligases (RELs) (Ho et al., 2004). Ho and colleagues characterized a variant of Rnl2 with a truncated C-terminus (Ho et al., 2004). They showed that the N-terminus carries an

adenyltransferase/AppRNA ligase domain, incapable of sealing a 3'-hydroxyl and 5'-phosphate termini. Therefore this polypeptide can only seal AppRNA and 5' phosphate ends, which takes place in the absence of ATP.

I have used Rnl2(1-249) for the ligation of Dicer cleavage products bearing a 5' phosphate to a chemically adenylated 3' adapter AppRNA, therefore minimizing circularization of input RNAs. The efficiency of the ligation by T4 RNA ligase depends on the donor adenylation (McLaughlin et al., 1985), thus a preadenylated donor is likely to ligate more efficiently. This was reflected in ligation efficiency assays for the 3' and 5' adapters, which was in general better for preadenylated 3' adapters than for 5' adapters.

My approach to optimize the cloning protocol was based on three modifications of our previous protocol: the use of the modified T4 RNA ligase 2, Rnl2(1-249), the design of new 3' and 5' adapter oligonucleotides, and new RNA size markers. Ligation assays for the new adapter and marker oligonucleotides indicated a bias for certain RNA sequences by RNA ligases. I therefore tested to which extent this could affect the cloning procedure.

Results

Ligase selection

In the original adapter ligation protocol, outlined in Figure 2A, small RNAs were dephosphorylated prior to ligation of a 3'-blocked 3' adapter to prevent their circularization. The ligation product was subsequently 5' phosphorylated and subjected to 5' adapter ligation. The dephosphorylation and phosphorylation steps can be bypassed by using a preadenylated 3' adapter in the first ligation step in the absence of ATP (Lau et al., 2001), (Figure 2A). However, T4 RNA ligases are not completely deadenylated, and the enzyme is capable of transferring its adenylate to 5' phosphorylated RNAs, which results in circularization. To avoid the problem of circularization of small RNAs, T4 RNA ligase was replaced in the first ligation step with Rnl2(1-249).

In addition, this approach allowed the enrichment of Dicer cleavage products bearing a 5' phosphate compared to random RNA degradation products with 5' hydroxyl termini. Cloning with adenylated adapters enriched the fraction of miRNAs obtained in a cDNA library compared to cloning with unadenylated adapters from 26% to 90% (Sebastien

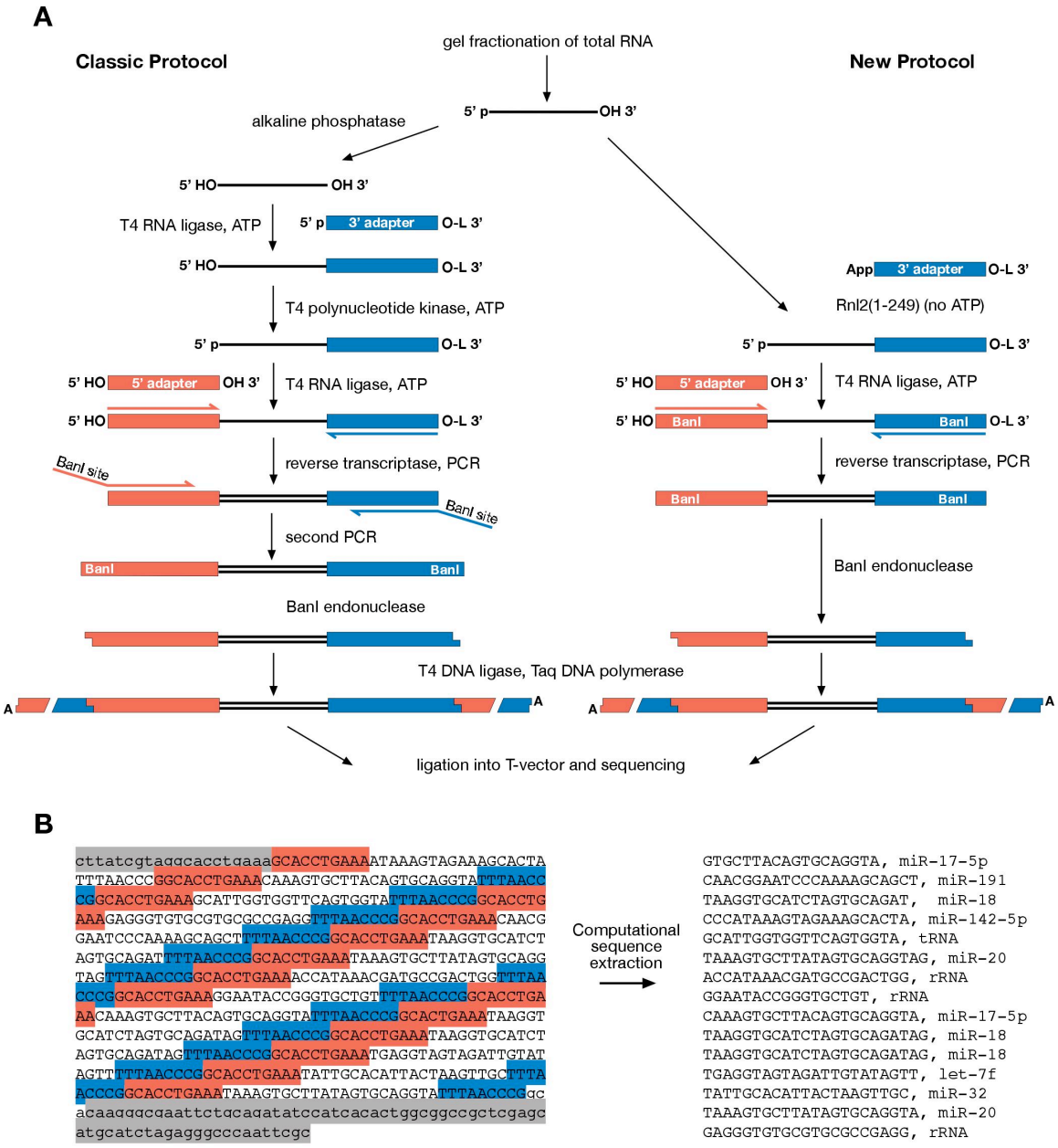


Figure 2. Small RNA cloning procedure

(A) Outline of the classic and new small RNA cloning procedure. P and OH indicate phosphate and hydroxyl at both ends of the RNA and oligonucleotides; L, 3' blocking group (see Table 2). (B) Example of a sequenced clone. On the left side the small RNAs are flanked by the adapter sequences highlighted in blue and pink. Vector sequences are highlighted in gray. The sequences on the right are extracted from the clone on the left.

Pfeffer, unpublished). To compare the efficiency of both ligases, preadenylated oligonucleotide *3A1* was ligated to radiolabeled oligonucleotide *M35* (Table 2) by two commercial T4 RNA ligases and by Rnl2(1-249). Near complete transligation was accomplished using Rnl2(1-249). In contrast, a significant fraction of the input RNA was

circularized by T4 RNA ligase even in the absence of ATP (Figure 3A). Therefore Rnl2(1-249) was used for all 3' adapter ligations.

His-tagged T4 RNA ligase purification

The enzyme for ligation of the 5' adapter to the small RNAs linked to 3' adapters is T4 RNA ligase, which seals 5' phosphate and 3' hydroxyl termini. I purified a His-tagged T4 RNA after recombinant expression in *E. coli* BL21(DE3) strain transformed with pRHT4 plasmid previously described (Wang and Unrau, 2002). The protein was purified over a nickel column and further purified over a sepharose Q column. 5.7 mg of purified protein were obtained from a bacterial culture of 1 liter. Ligase activity was assayed by circularization of a linear oligoribonucleotide (oligo *22*) with 5' phosphate and 3' hydroxyl termini as shown in Figure 3B

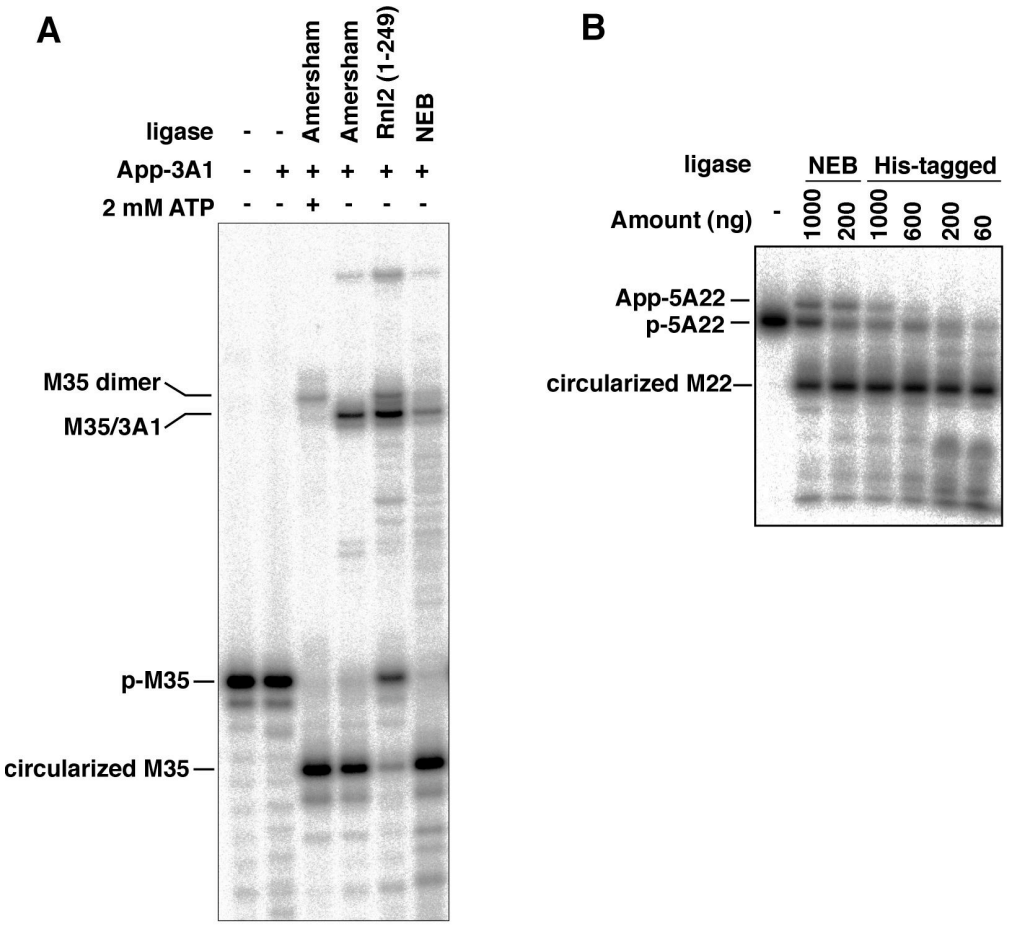


Figure 3. Ligation activity of RNA ligases

(A) Pre-adenylated DNA oligo *3A1* was tested for ligation efficiency to a ^{32}P-labeled 19 nt oligoribonucleotide (*M35*). Different RNA ligases were used: Rnl2(1-249), T4 RNA ligase from Amersham and from NEB. Desired ligation products and cirular side products are indicated. (B) Activity of purified His-tagged T4 RNA ligase (His-tagged) was compared to T4 RNA ligase from NEB. The amount of enzyme used in each reaction is indicated. App, and p indicate adenylate and phosphate on the 5' end of the oligonucleotide.

RNA size markers

In the original developed protocol for small RNA cloning (Pfeffer et al., 2003), RNA size markers are loaded separately on polyacrylamide gels. However, I occasionally encountered that the size of the cloned RNAs did not correspond to the desired range (19-24 nt), This was due to different migration behavior between the separately loaded size markers and the total RNA sample. To overcome this problem, RNA size markers were mixed to the samples prior to PAGE size fractionation. In addition, these radiolabeled size markers can be used as internal control for the ligation reactions. New RNA size

markers with an eight base-pair PmeI restriction site were designed (none of the currently known miRNAs contains a PmeI restriction site). After PCR amplification, the sample can be digested with PmeI to avoid cloning the markers. To select one marker pair that was efficiently ligated by T4 RNA ligase, I examined the products of four different marker pairs ligated to four different 3' adapters (Figure 4). Marker oligonucleotides *M27* and *M28* (19 and 24 nt long respectively) were the most efficiently ligated to all of the 3' adapter nucleotides tested, and therefore these marker oligonucleotides were selected for subsequent small RNA cDNA library preparation.

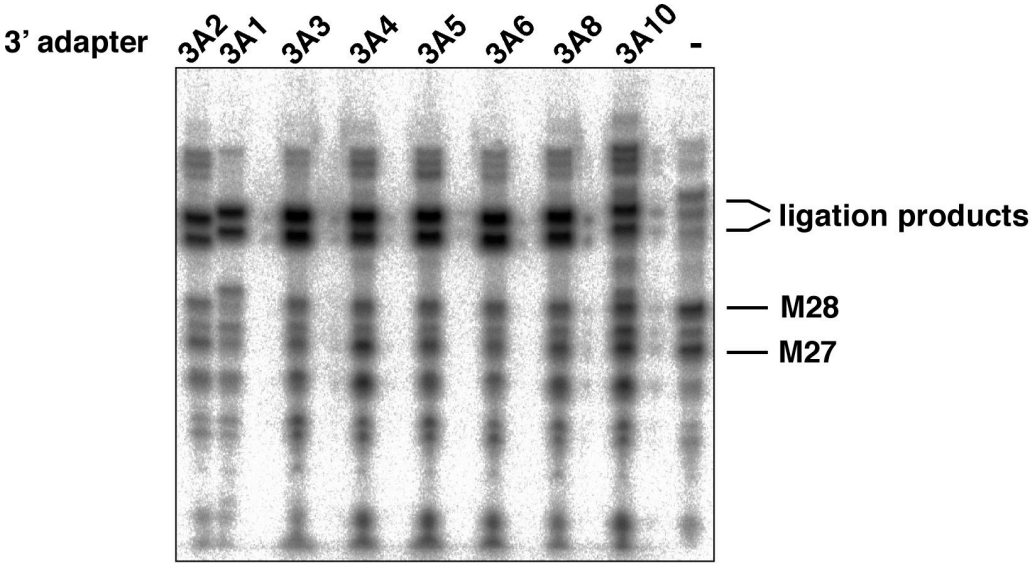


Figure 5. Ligation efficiency of 3' adapter oligonucleotides

3' adapter oligonucleotides were ligated to two ^{32}P-labeled RNA size markers 19 and 24 nt long (*M27* and *M28*, respectively) using Rnl2(1-249). Ligation reactions were analyzed by PAGE. Desired ligation products and unligated size markers are indicated. See Table 2 for adapter and size marker sequences.

3' and 5' adapters

Following the ligation of the two adapters, cDNA synthesis and PCR, BanI restriction sites were introduced in our original protocol by PCR amplification using primers with additional BanI restriction sites. After PCR, BanI digestion and concatamerization, constant fragments of 36 nt in length flank the small RNA sequences (Figure 2A). I designed new 3' and 5' adapters that already contained a BanI restriction site to bypass

the second PCR and shorten the nucleotide stretches between the small RNAs. After PCR and digestion the units for concatamerization are almost 30% shorter using the new adapter sets (42 versus 59 nt). Therefore, when concatamerized and ligated into a plasmid, more sequence information per sequenced clone can be obtained (Figure 2B).

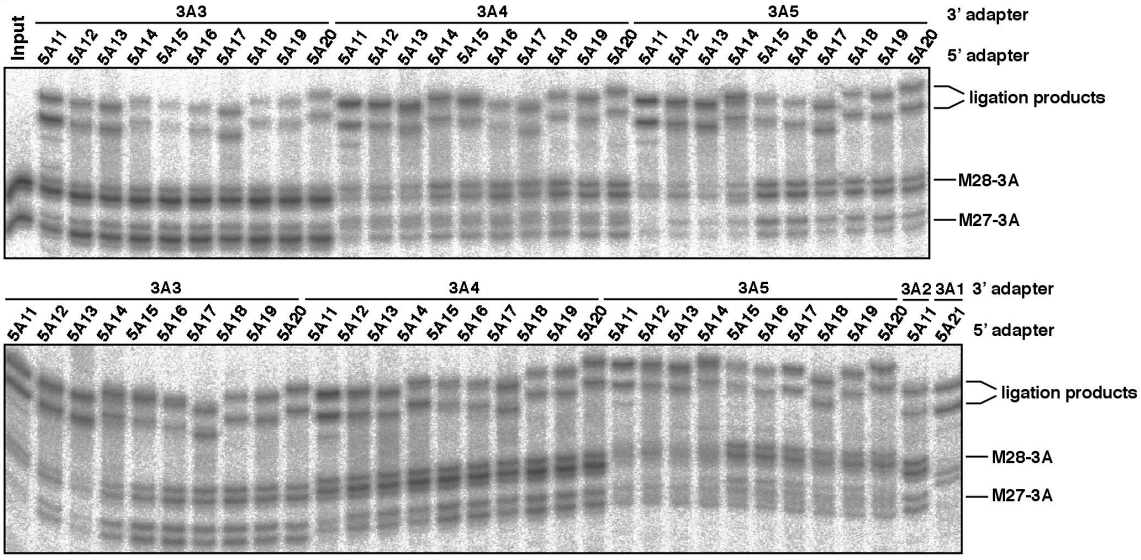
3' adapter oligonucleotides were synthesized and adenylated as described (Pfeffer et al., 2003). Oligonucleotides *3A1*, *3A2*, *3A3*, *3A4*, *3A5*, *3A6*, *3A7*, *3A8*, and *3A10* were ligated to RNA size markers *M27* and *M28*, 19 and 24 nt in length, respectively. All 3' adapters ligated to both RNA markers with efficiencies greater than 80% (Figure 5). After gel purification of the two ligation products derived from each adapter ligation (adapter+19mer and adapter+24mer), the isolated products were subsequently ligated to 5' adapters *5A11* through *5A20*. The ligation efficiencies are summarized in Figure 6. 5' adapters *5A11*, *5A12*, and *5A13* ligated more efficiently to each marker-adapter ligation product than any of the other 5'adapters. Surprisingly, the efficiency of the 5' adapter ligation reaction was dependent on the 3' adapters to which the RNA size marker was initially ligated. RNA size markers joined to the 3' adapter oligonucleotides *3A5* and *5A6* ligated more efficiently than markers linked to oligonucleotides *3A3*, *3A4*, *3A8* and *3A10*. My data suggest that the ligation reaction catalyzed by T4 RNA ligase was dependent on the nucleotides at the site of bond formation as well as on the larger sequence context, which likely influences the RNA secondary structure.

Synthetic miRNA pools

To assess any bias introduced by T4 RNA ligases, oligoribonucleotide pools were created containing only synthetic miRNAs (see Materials and methods). The pools consisted of six oligoribonucleotides. Pool I contained the oligos in concentrations that differed up to about 32 fold, and in pool II oligos differed up to 1.6 fold (see materials and methods). Small RNAs were cloned from both pools using two different adapter pairs: adapter pair B, 3' adapter *3A6* and 5' adapter *5A11*; adapter pair C, 3' adapter *3A5* and 5' adapter *5A17* (Table 2). More than 100 miRNA sequences were obtained from each pool. The sequencing output of pool I, cloned with adapter pair B shows a very good correlation with the relative abundances of oligonucleotides in the pool (Figure 7). The rest of the

cloning sets show underrepresented oligonucleotides in the sequencing outputs. In this experiment miR-16 and miR-29a were underrepresented when cloning with either

A



B

Adapter pair	Percentage ligated M28	M27
3A3-5A11	25	66
3A3-5A12	14	19
3A3-5A13	22	31
3A3-5A14	06	08
3A3-5A15	03	03
3A3-5A16	06	07
3A3-5A17	15	23
3A3-5A18	03	06
3A3-5A19	06	09
3A3-5A20	10	14
3A4-5A11	74	64
3A4-5A12	69	50
3A4-5A13	75	58
3A4-5A14	45	48
3A4-5A15	56	44
3A4-5A16	19	22
3A4-5A17	37	43
3A4-5A18	20	30
3A4-5A19	37	41
3A4-5A20	26	35

Adapter pair	Percentage ligated M28	M27
3A5-5A11	82	91
3A5-5A12	72	82
3A5-5A13	85	89
3A5-5A14	71	77
3A5-5A15	16	32
3A5-5A16	14	36
3A5-5A17	42	69
3A5-5A18	20	49
3A5-5A19	29	51
3A5-5A20	36	64
3A6-5A11	73	92
3A6-5A12	75	80
3A6-5A13	87	91
3A6-5A14	70	61
3A6-5A15	64	29
3A6-5A16	37	17
3A6-5A17	36	45
3A6-5A18	19	31
3A6-5A19	36	44
3A6-5A20	41	49

Adapter pair	Percentage ligated M28	M27
3A7-5A11	57	85
3A7-5A12	35	59
3A7-5A13	29	49
3A7-5A14	18	63
3A7-5A15	12	24
3A7-5A16	12	27
3A7-5A17	29	51
3A7-5A18	09	28
3A7-5A19	17	34
3A7-5A20	23	42
3A8-5A11	69	66
3A8-5A12	59	54
3A8-5A13	69	62
3A8-5A14	63	52
3A8-5A15	16	25
3A8-5A16	17	32
3A8-5A17	35	38
3A8-5A18	41	58
3A8-5A19	31	47
3A8-5A20	37	54
3A1-5A11	21	33
3A2-5A21	60	89

Figure 6. Ligation efficiency of 5' adapter oligoribonucleotides

(A) 5' adapters *5A11* through *5A20* were ligated to size markers *M27* and *M28* previously linked to 3' adapters (obtained after elution of ligation products shown in figure 5). Ligation reactions were analyzed by PAGE. Ligation products and unligated RNA size markers-3' adapters are indicated. (B) Ligation efficiencies were determined by phosphorimager analysis. Ligations with efficiency of 50 to 69% are highlighted in yellow, and with efficiency greater than 70% are highlighted in red. See Table 2 for oligonucleotide sequences.

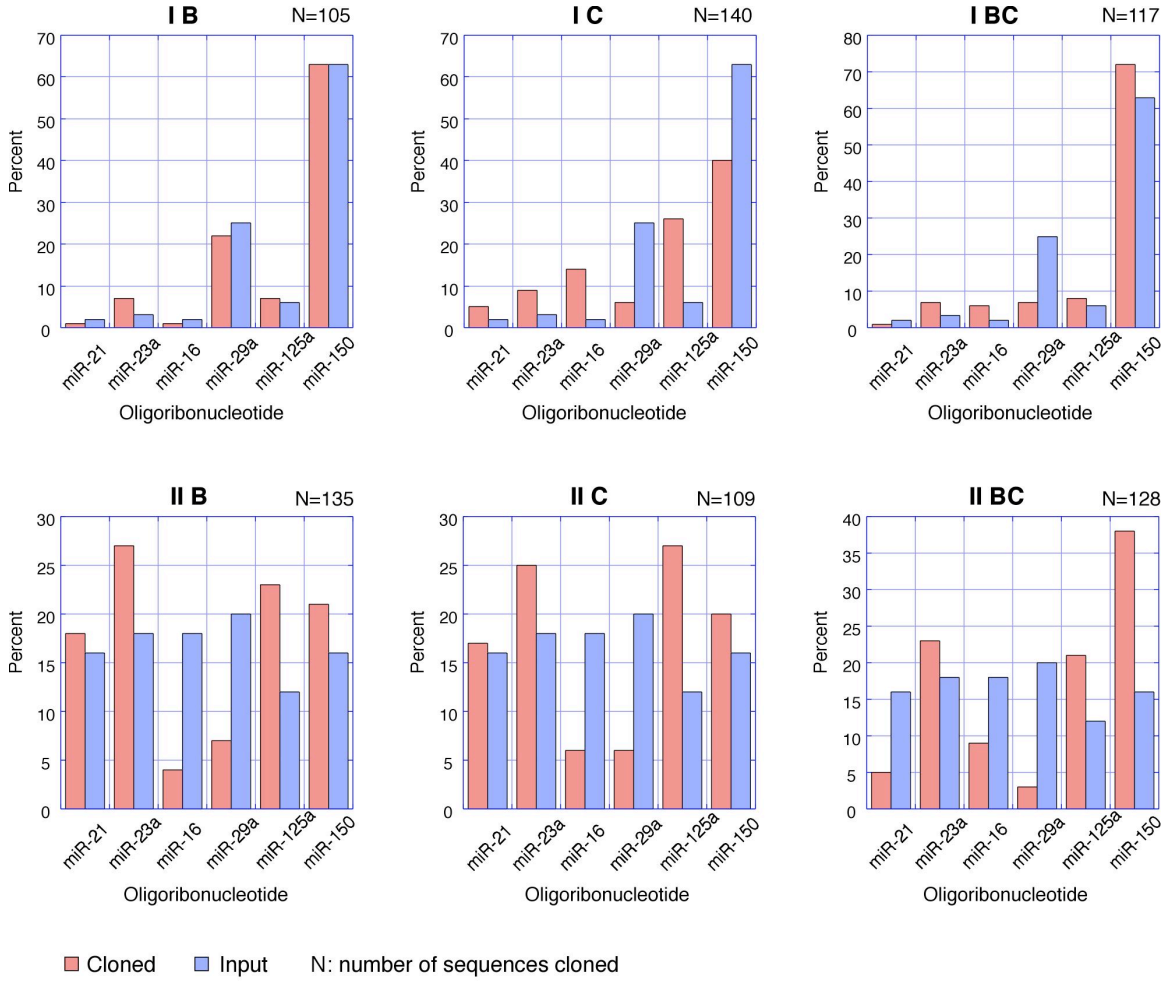


Figure 7. Cloning synthetic miRNAs

Six different synthetic miRNAs were mixed in different concentrations to create miRNA pools I and II. The two different pools were cloned using adapter pair B, C and both adapter pairs in a single experiment. Blue bars indicate the relative abundance of each miRNA in a pool; pink bars indicate the cloning output for each miRNA as indicated. N, number of cloned sequences for each pool.

adapter pair, and miR-125a was in most of the cases overrepresented. In these miRNA pools the combination of two 3' adapters and two 5' adapters in a single cloning procedure was not an advantage to eliminate or decrease the sequence bias (Figure 7). For example, miR-21 shows a similar distribution between input and output when cloned with either pair of adapters from pool II, however, when both adapter pairs were used at the same time, miR-21 was underrepresented in the output. Both miR-150 and miR-125a were either well represented or overrepresented in all libraries. These two RNAs share the same terminal nucleotides in the 5' and 3' ends (miR-125a: 5'-UC(N20)UG-3'; miR-150: 5'-UC(N19)UG-3'). This similarity may suggest that certain nucleotide arrangements are preferentially ligated. On the other hand, the case of miR-16, which is always underrepresented in cloning outputs from pool II but well represented or overrepresented in pool I outputs, cannot be explained by a ligase bias only.

Small RNA library of the mouse B cell line A20

To further asses any bias introduced by adapter sequences in an authentic small RNA sample, I cloned small RNAs from the mouse B cell line A20. Cloning was carried out with two different adapter pairs in parallel: pair B, described above; and pair A (the adapter pair of the original protocol), composed of adenylated 3' adapter *3A1* and 5' adapter *5A21* (Table 2). Cloning with each adapter pair was performed using the same RNA preparation with 100 μg RNA as starting material. Although a second PCR is not needed to introduce the restriction sites for marker pair B, which contains, BanI restriction sites, this sample went through the same number of PCR cycles as the sample which required a second PCR (see Figure 2 for protocol outline). 845 sequences were obtained by cloning with the original adapters (adapter pair A) and 664 sequences with adapter pair B. The proportion of miRNAs in the sample cloned with adapter pair A was higher than with adapter pair B, 84% and 74% respectively, also the diversity of the miRNAs cloned was higher with adapter pair A than with adapter pair B, 61 and 52 different miRNAs, respectively (Figure 8A). The frequency of cloning differed significantly between libraries for miR-15, miR-17-5p, miR-19a, miR-20, miR-21 and miR-142-5p (Figure 8). miRNAs cloned one or two times do not provide significant data to assess cloning bias or diversity. miR-16 was the only miRNA cloned with high

frequency with both adapter pairs. Relative expression levels were tested for miR-16, miR-17, miR-17*, miR-19a, miR-20, miR-142-5p and miR-142-3p by Northern blot analysis (Figure 8B). If one assumes that all probes hybridize with the same efficiency to their complementary miRNAs and without cross hybridization, Northern blots showed that miR-20 followed by miR-19 were the most abundant among the ones probed for A20 RNA. Nevertheless, miR-16 was cloned with both adapters 7 times more frequently than miR-19a. The Northern signal for miR-16 was 2.3 times stronger than for miR-17. However, when cloned with adapter pair B, miR-17 appears to be more abundant than miR-16. In contrast, adapter pair A selected against miR-17, and therefore cloning with adapter pair A would have misrepresented the abundance of this miRNA. This is consistent with (Lagos-Quintana et al., 2001), where miR-17* was cloned with adapter pair A, but not the higher expressed miR-17. Higher expression of miR-17 compared to miR-17* was confirmed by Northern hybridization in A20 cells (Figure 8B) and HeLa cells (data not shown). Similarily, adapter pair A, selected for miR-142-5p. This miRNA was more frequently cloned than miR-16 with this adapter pair, but Northern data indicates that miR-142-5p is 6 times less abundant than miR-16. The Northern data was better reflected when cloning with adapter pair B.

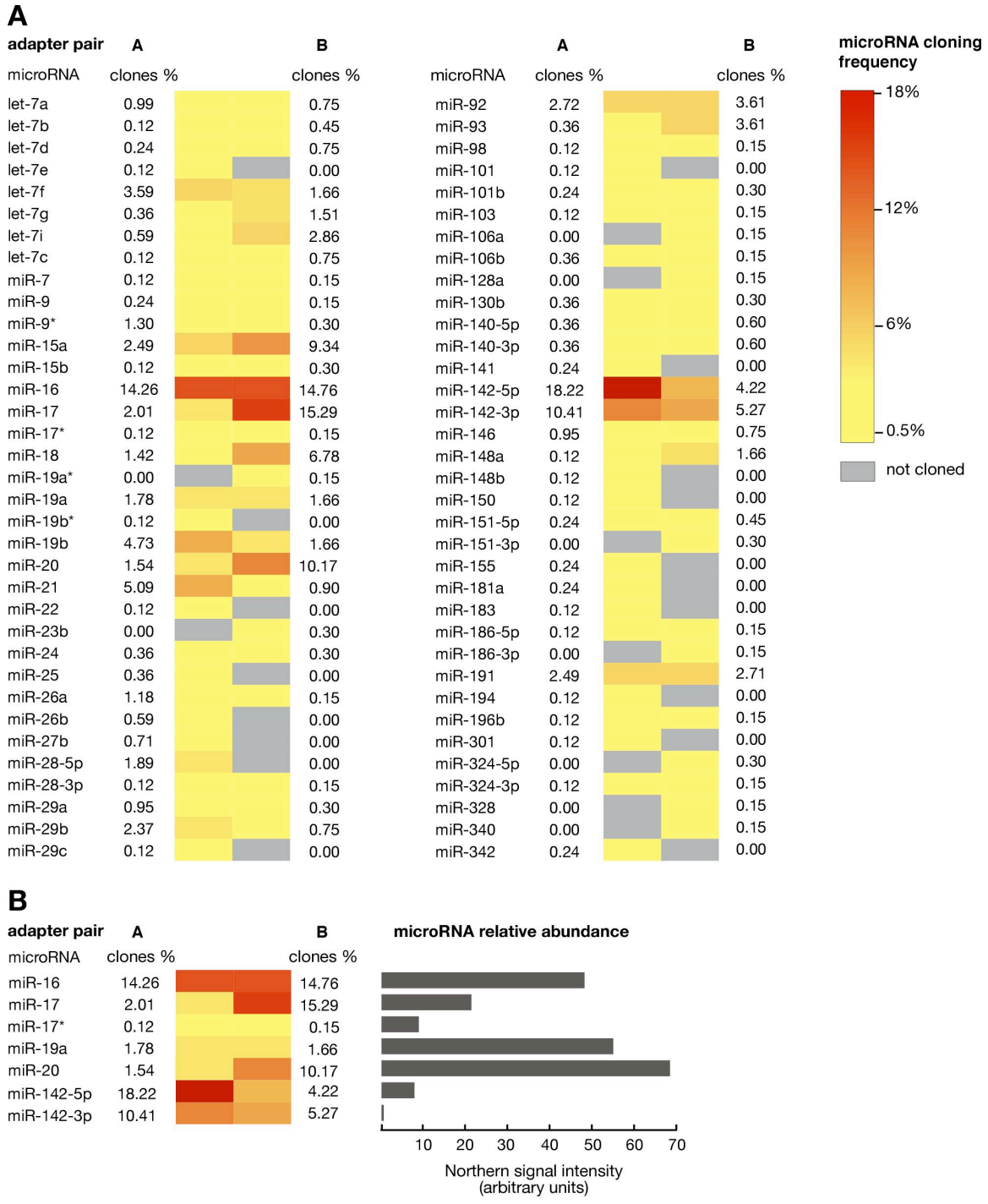


Figure 8. A20 small RNA libraries cloned with two different adapter pairs

(A) Frequency of cloned miRNAs with adapter oligonucleotide pairs A and B. (B) A subset of miRNAs present in the A20 libraries were analyzed by Northern blotting. Cloning frequency as shown in (A) is compared to normalized Northern signal intensity.

Discussion

Small RNA cloning has been widely used for miRNA identification and expression profiling. It has been an essential tool for identifiying miRNAs, and other RNAs like rasiRNAs (Chapter III; Reinhart and Bartel, 2002). With an increasing number of known miRNAs, other methods for small RNA profiling have become available. Microarrays and Northern blot analyses are now of general use to determine the expression profile of known miRNAs.

I have used miRNA cloning as a technique to identify new miRNAs and to profile miRNA expression in different tissues (Chapters I, II and III). Expression of a sequence is supported by Northern analysis. The results presented above address the accuracy of such a protocol to portrait the actual miRNA population in cell culture or tissue. The synthetic oligoribonucleotide pools and the A20 cell miRNA profiles obtained with different adapter pairs, support the idea that T4 RNA ligase introduces somedegree of bias in the cloning procedure. Abundant miRNAs showed up to 7.6 fold difference in cloning outputs obtained with different adapter pairs. The frequency in which a small RNA will ligate to an adapter is not only dependent on its abundance but also on its sequence and it is not clear which sequence elements determine the ligation efficiency of two RNAs. In addition the cDNA synthesis by reverse transcription and PCR amplification steps could also contribute to the observed variation. For example, for different miRNAs the cDNA synthesis starts with different deoxynucleotides since the reverse transcription primer is the full-length complement of the 3' adapter (Figure 2). Thus, interpreting cloning frequency as abundance of a certain miRNA in an RNA sample can be misleading and such interpretations should be made carefully, and should be supported by other means of quantification.

In most cases frequently cloned miRNAs are abundant in the cell and Northern analyses can be performed for those miRNAs to confirm their relative expression. Cloning studies and microarray or Northern blot profiling have coincided to detect the most abundant miRNAs from a certain tissue. For example, cloning from adult mouse tissues suggest that tissue specific miRNAs, miR-1 in heart, miR-122 in liver, and miR-124 in neuronal tissue, represent one of the most abundant miRNAs in each tissue (Lagos-Quintana et al., 2001) and Chapter II). These findings were confirmed by Northern analysis (Chapter I),

and microarray data from other groups have supported these data (Babak et al., 2004; Baskerville and Bartel, 2005; Krichevsky et al., 2003; Thomson et al., 2004)

However, cloning data and Northern analyses frequently appear to contradict each other. Figure 8B shows such a discrepancy, in which miR-142-3p was abundantly cloned but was not detected by Northern analysis. Another example is provided in Chapter III where we failed to clone miR-125 from *Drosophila* but detected a strong signal by Northern analysis. Similar cases appear in the literature when expression of predicted miRNAs is confirmed by Northern hybridization where cloning profiling failed to detect those miRNAs (Lai et al., 2003; Lim et al., 2003b). It should be noted that cloning reflects the relative levels of miRNAs, and in many cases if the RNA input is low, a non-abundant miRNA can be cloned frequently due to amplification of the sequence by PCR.

Most of the currently known miRNAs have been identified by cloning techniques, and their expression confirmed by Northern blot analysis. Lately, bioinformatics has predicted many miRNA genes that need to be validated by hybridization-based methods or by cloning. miRNA microarrays have emerged as a tool for miRNA profiling. Microarrays provide a high throughput method to determine relative expression levels of miRNAs. This technique is based on hybridization of a fluorescently labeled miRNA library to antisense miRNA sequences immobilized on a glass slide. The measured fluorescence indicates the level of expression of each miRNA relative to the others in the sample. Hybridization methods seem to overcome the problem of sequence bias that I encounter in the cloning protocol. However thermodynamic stability and secondary structure of a small RNA will have a significant impact on hybridization efficiency. Moreover, most microarray protocols label the RNA samples by directly attaching the fluorescent tag to the 3' end of the miRNA using T4 RNA ligase (Miska et al., 2004; Thomson et al., 2004), and another protocol ligates two adapters with T4 RNA ligase followed by reverse transcription and PCR (Baskerville and Bartel, 2005). Nelson and colleagues eliminated the use of T4 RNA ligase as well as reverse transcription and PCR in a procedure called RNA-primed, array-based Klenow enzyme (RAKE) assay (Nelson et al., 2004). The RAKE assay uses a Klenow reaction to extend the miRNA, which functions as a primer, only when a miRNA binds with complete complementarity to an immobilized probe. This method is very sensitive to distinguish nucleotide mismatches at

the 3' end, where miRNA homologues commonly share the greatest sequence disparity. It remains to be determined whether the RAKE assay is the miRNA profiling method with the least bias and the most specific hybridization-based method.

Sequence similarity between miRNAs of the same family presents a disadvantage to hybridization-based methods. Complementary sequences with a few mismatches can hybridize on a Northern blot or microarray and give an unspecific signal depending on the position and nature of the mismatch. However, it is debatable if closely related miRNAs recognize the same targets and therefore are functionally equivalent. Specificity remains a desirable feature of profiling assays, and microarrays seem to be better in discriminating between paralogs than Northern analysis (Nelson et al., 2004). In addition, microarrays are suitable for small samples of RNA. It is possible to detect abundant miRNAs with as little as 50 ng of total RNAs and non-abundant miRNAs and 50% of miRNAs with 5 μg of total RNA (Thomson et al., 2004). The smallest amount of total RNA tested to create a cDNA library was 20 μg. However the results produced with such an amount varied from cloning results obtained from the same RNA sample using 200 μg (Pablo Landgraf, unpublished). These results suggest the variation from experiment to experiment exist by changing experimental conditions other than the adapter pairs.

Cloning efforts continue to identify new miRNA sequences and to reveal the existence of other small RNAs like rasiRNAs. The latter class of RNAs cannot be predicted by computational methods, since they do not share structural features. Thus, small RNA cloning remains a very powerful technique to identify and detect miRNAs in specific tissues, subcellular compartments and organisms where these RNAs most likely play a role in gene silencing.

In summary, I have improved our previous protocol for small RNA cloning by using a modified RNA ligase and optimizing the adapter oligonucleotides. The use of pre-adenylated adapters significantly enriched for Dicer cleavage products, the small RNAs involved in gene silencing mechanisms. Despite the evidence of some degree of selectivity, it remains the most reliable method to discover new small RNAs. However, hybridization-based methods where ligation, reverse transcription and PCR are bypassed might be more accurate to asses relative abundances of miRNAs.

Table 2. adapter and size marker oligonucleotides.

function	oligonucleotide	sequence (5'→3')
3' adapter	*3A1*	uuuAACCGCGAATTCCAG-L
	3A2	CTGTAGGCACCATCAA-L
	3A3	TTTAACCGCGGCACCAG-L
	3A4	TTTAGCTGGGCACCTCA-L
	3A5	TTTTACGAGGCACCCAG-L
	3A6	TTTAACCCGGCACCCTC-L
	3A7	TTTAACCTGGCACCGGA-L
	3A8	TTTGTCGCGGACCTAAA-L
	3A9	uuuAACGACGGCACCCAG-L
	3A10	uuuAACGAGGCACCCAGAG-L
5' adapter	*5A11*	ATCGTaggcaccugaaa†
	5A12	agggaggcaccgaugcgg
	5A13	agggauggcaccgaugc
	5A14	acggaauuggcaccacuaaa
	5A15	aucgaaggcaccucacuaa
	5A16	acatgaggcaccggauua
	5A17	aggguggcaccacgaaa
	5A18	ACGGTTAAGGCACCuaa
	5A19	AGCCATAGGCACCGaau
	5A20	TTGGCAGGCACCAGGauu
	5A21	ACGGAATTCCTCACTaaa
	5A22	AGAGGCACCGTCTCTaaa
	5A23	CAATGGCACCGACTCAaaa
	5A24	ACGGCACCCTCACTaaa
RNA size markers	*M25*	cguacgcggaauacuucga
	M26	cguacgcggaaua**guuuaaac**ugu
	M27	cguacgcgg**guuuaaac**ga
	M28	cguacgcggaaua**guuuaaac**ugu
	M29	c**guuuaaac**aauacuucga
	M30	c**guuuaaac**aauacuucgaaaugu
	M31	cgua**guuuaaac**uaaacga
	M32	cgua**guuuaaac**uaaacgaacugu
	M33	uuuaac**guuuaaac**uccag
	M34	uuuaaccgc**guuuaaac**auuccag
	M35	cguacgcggaauacuucga
	M36	cguacgcggaauacuucgaaaugu

Lowercase, RNA; uppercase, DNA; L, C6-3' aminolinker, (Dharmacon). 3' adapters were preadenylated according to (Pfeffer et al., 2003). Underlined, BanI restriction site; bold, PmeI restrictions site. † DNA-RNA chimeric 5' adapter described by the Bartel lab (Lau et al., 2001).

Materials and methods

Labeling of RNA oligonucleotides

RNA oligonucleotides 1 μM were phosphorylated in a 10 μl reaction using T4 polynucleotide kinase (New England Biolabs) in 1x PNK buffer, and 1 μl [γ^{32}P]-ATP (~3000 Ci/mmol, Amersham Biosciences) for 15 min at 37°C. The labeled RNA was gel-purified in a 15% acrylamide gel.

T4 RNA ligase purification

His-tagged T4 RNA ligase was expressed and purified as previously described (Wang and Unrau, 2002) with the following modifications. One liter of LB medium was inoculated with 30 ml overnight culture in a 4 liter bottle with flat bottom to avoid excessive aeration. Cell culture was grown at 37ºC. At OD_{600} of 0.7 the culture was induced with 1 mM IPTG and grown for 3 h at 37ºC shaking at 260 rpm. Cells were harvested at room temperature, resuspended in 10 ml lysis buffer, frozen in liquid nitrogen, and stored at -80ºC for later use. Cells were disrupted by sonication in 60 ml lysis buffer and debris was removed at 10,000 x g for 20 min. 2 ml His-Bind resin (Novagen) was charged with 100 mM Nickel Sulfate as recommended by the manufacturer and equilibrated in lysis (50 mM KH_2PO_4, 10 mM $Na_4P_2O_7$, 300 mM NaCl, and 5 mM imidazole). $Na_4P_2O_7$ was used to deadenylate the enzyme. The resin was added to the lysate and rotated for 1 h at 4ºC without generating air bubbles. The resin was washed in a glass column with 20 ml of wash buffer (50 mM KH_2PO_4, 300 mM NaCl, 10 mM imidazole, and 10 mM $Na_4P_2O_7$), and 20 ml of wash buffer without $Na_4P_2O_7$. Bound protein was eluted with 5 ml elution buffer (50 mM KH_2PO_4, 300 mM NaCl, 100 mM imidazole). The eluate was dialyzed against buffer C (40 mM sodium phosphate [pH 7.7], 2 mM EDTA, 2 mM DTT, and 10% glycerol). The dialyzed protein was further purified over a sepharose Q column (Hi Trap Q FF, 1 ml, Amersham Biosciences) using an ÄKTA prime FPLC and eluted with a 0.05 to 1.0 M NaCl gradient over 20 ml. Fractions of 0.5 ml were collected. The protein eluted around 0.5 M NaCl and was dialyzed against storage buffer (25 mM HEPES [pH 7.5], 1 mM DTT, 0.1 mM EDTA, and 50 % glycerol)

and stored at –20ºC. The purity and deadenylation state of the protein were analyzed by SDS-PAGE. The enzyme was diluted in storage buffer to a working concentration of 1 μg/μl (comparable to commercial T4 RNA ligases).

Recombinant T4 RNA ligase activity assay

Purified T4 RNA ligase was assayed for activity by circularization of a radiolabled RNA oligonucleotide (oligo *5A22*) 5' phosphate and 3' hydroxyl termini. Ligation was performed in 50 mM Tris-HCl [pH 7.6], 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0.1 mg/ml acetylated bovine serum albumin (Sigma), 15% v/v DMSO, 2 mM ATP, and 100 nM oligonucleotide in a reaction volume of 10 μl. Different reactions were performed using 60, 200, 600, and 1000 ng of enzyme. As positive control two reactions were performed using the same conditions with 200 and 1000 ng T4 RNA ligase (New England Biolabs).

3' adapter ligation to radiolabeled size marker oligoribonucleotides

Preadenylated oligonucleotides with a blocked 3' end were ligated to 5' ^{32}P-labeled oligoribonucleotides. Reactions were performed with Rnl2(1-249) ligase (Rnl2), or with T4 RNA ligase when indicated. Ligation was performed in 50 mM Tris-HCl [pH 7.6], 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0.1 mg/ml acetylated bovine serum albumin (Sigma), and 15% v/v DMSO. The concentration of radiolabeled acceptor oligoribonucleotide was 100 nM, the concentration of adenylated donor oligonucleotides was 1 μM. The final concentration of enzyme used was 100 ng/μl.

Ligation efficiency assays

Each adapter was assayed for ligation under the reaction conditions described above. The products were separated by 15% PAGE and visualized and quantified using a phosphorimager (Fuji).

Synthetic miRNA pools

Different oligoribonucleotide pools were cloned using different 3' and 5' adapter pairs (pair A and pair B) and combing both.

Adapter pair A 3' adapter: 5'-uuuAACCGCGAATTCCAG-L (oligo *3A1*)
5' adapter: 5'-ACGGAATTCCTCACTaaa (oligo *5A21*)

Adapter pair B 3' adapter: 5'-TTTAACCCGGCACCCTC-L (oligo *3A6*)
5' adapter: 5'-ATCGTaggcaccugaaa (oligo *5A11*)

Adapter pair C 3' adapter: 5'-TTTTACGAGGCACCCAG-L (oligo *3A5*)
5' adapter: 5'-aggguggcaccacgaaa (oligo *5A17*)

(Lower case, RNA; uppercase, DNA; L, C7-3' aminolinker (Chemgenes). 3' adapters were pre-adenylated as described (Pfeffer et al., 2003)).

The pools consisted of six synthetic RNA oligonucleotides containing 3' hydroxyl and 5' phosphate termini. The oligonucleotides were mixed in a 10 μl 3' adapter ligation as follows

miRNA	Sequence	Concentration pool I (nM)	Concentration pool II (nM)
miR-21	UAGCUUAUCAGACUGAUGUUGA	44	382
miR-23a	AUCACAUUGCCAGGGAUUUCC	72	445
miR-16	UAGCAGCACGUAAAUAUUGGCG	50	445
miR-29a	CUAGCACCAUCUGAAAUCGGUU	555	492
miR-125a	UCCCUGAGACCCUUUAACCUGUG	136	287
miR-150	UCUCCCAACCCUUGUACCAGUG	1400	380

The reaction contained 3 different oligoribonucleotides containing two 3' terminal deoxynucleotides that are not the preferred substrate for T4 RNA ligase (Nandakumar et al., 2004).

miRNA	Sequence	Concentrarion pool I (nM)	Concentrarion pool II (nM)
miR-30a-3p	GCUGCAAACAUCCGACUGAAAAdTdT	36	287
miR-124a	GGCAUUCACCGCGUGCCUUAUdTdT	146	195

let-7a	UGAGGUAGUAGGUUGUAUAGdTdT	97	195

Ligation was performed in 50 mM Tris-HCl [pH 7.6], 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0.1 mg/ml acetylated bovine serum albumin (Sigma), 15% v/v DMSO in a reaction volume of 10 μl. The concentration of 3' adapter oligonucleotide was 10 μM when only one adapter was used, and 5 μM of each adapter when adapters A and B were used in the same reaction. The final concentration of Rnl2(1-249) ligase used was 100 ng/μl. The ligation reactions were stopped by addition of urea loading dye and the products were analyzed by PAGE. The eluted products of the 3' ligation were dissolved in water and the 5' adapter ligation and product purification was performed in the same reaction conditions as above with the exception of addition of 0.2 mM ATP for the ligation reaction and the use of the recombinant His-tagged T4 RNA ligase.

A20 library

Mouse cell line A20 RNA isolation and miRNA cloning were performed as described (Pfeffer et al., 2003). Adapter pair A as described by Pfeffer and colleagues (oligonucleotides *3A1* and *5A21* in Table 2). Adapter pair B is described above. 3' adapter ligations were performed with Rnl2(1-249) and 5' with His-tagged T4 RNA ligase (see above).

Northern Blots

Northern blot analysis was performed as described in (Lagos-Quintana et al., 2001) Oligodeoxynucleotides used as probes were miR-16, CGCCAATATTTACGTGCTGCTA; miR-17, ACTACCTGCACTGTAAGCACTTTG; miR-17*, TACAAGTGCCCTCACTGCAGT; miR-19a, TCAGTTTTGCATAGATTTGCACA; miR-20, CTACCTGCACTATAAGCACTTTA; miR-142-5p, GTAGTGCTTTCTACTTTATG; miR-142-3p, CCATAAAGTAGGAAACACTACA; val-tRNA TGGTGTTTCCGCCCGGTTT. tRNA was used as loading control. miRNA signals were quantified using a phosphorimager (Fuji) and normalized to tRNA signals.

General Discussion

miRNA identification

Four years ago the Ambros, Bartel and Tuschl laboratories discovered a novel family of small regulatory RNA genes (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001). These landmark papers on the identification of miRNAs collectively described 102 of these small RNAs, 27 from human, 17 from *Drosophila* and 58 from *C. elegans*. One year later we cloned 82 additional miRNAs derived from 122 genes, from human SAOS-2 osteosarcoma cells and mouse, describing differential expression of miRNAs in various tissues (Chapters II and III). Subsequently, a more exhaustive study in *D. melanogaster* identified 46 additional fly miRNAs (Chapter III). At the time of writing, 199 miRNAs (miRNA* sequences excluded), encoded by 224 genes have been characterized in humans as listed in the miRNA registry (Rfam, release 5.1), (Griffiths-Jones, 2004). In the last four years miRNAs were identified from several organisms and their expression in different tissues and developmental stages was described. Several techniques now aid the initial cloning approach for the identification of new miRNAs and analysis of their expression patterns.

Computational methods to predict miRNAs have estimated the number of miRNA genes per genome, and they have identified new miRNAs (Chapter III; Berezikov et al., 2005; Lai et al., 2003; Lim et al., 2003b; Ohler et al., 2004; Wang et al., 2004). However the expression of predicted miRNAs must be validated experimentally by hybridization-based methods or by small RNA cloning. Computational predictions identify evolutionarily conserved genomic sequences that are able to fold like a characteristic miRNA precursor. Therefore it is not possible to know the exact miRNA sequence derived from a predicted precursor, and in many cases it is not known which strand is transcribed. Recent work in *C. elegans* has identified an 8 nt motif 200 nt upstream of most intergenic miRNAs. Identification of such upstream element has helped identifying *C. elegans* miRNAs by computational methods (Ohler et al., 2004).

I believe that most ubiquitously expressed and highly abundant miRNAs have already been identified. Cloning efforts should now focus on specific cell types rather than whole

tissues containing mixed cell types. miRNAs may have specific roles in specialized cells as it has been shown in miR-375 in murine pancreatic β-cells (Poy et al., 2004), and the *C. elegans lsy-6*, which expressed in a single cell, contributing to neuronal asymmetry (Johnston and Hobert, 2003). miR-375 was discovered by cloning, whereas *lsy-6* was identified in a genetic screen for neuronal defects.

How many miRNA genes?

Based on cross-species conservation and structural characteristics of miRNAs, several groups have estimated the total number of miRNAs in a genome. Early estimates concluded that mammals had at most 255 conserved miRNA genes (Lim et al., 2003b). Considering that currently 224 mouse miRNA genes appear in Rfam, only about 30 genes would remain to be identified. More recent publications agree in that early estimates of the number of miRNAs were low (Berezikov et al., 2005; Xie et al., 2005). This number has been raised by more than a factor of two. The Plasterk laboratory (Berezikov et al., 2005) has a conservative estimate of 200 to 300 novel miRNA genes, but states that this number may also be an under-estimate, and that there could be as many as 1,000 miRNAs in vertebrate genomes. As a consequence, the estimate by Lewis and colleagues (Lewis et al., 2003) that 10% of all protein-coding genes are regulated by miRNAs, has also gone up substantially to 20-30% (Lewis et al., 2005; Xie et al., 2005). Not surprisingly, an extensive cloning effort to identify new miRNAs from different human and mouse cell lines and tissues by the Tuschl laboratory, has identified more than 100 new miRNAs (Alexei Aravin, Pablo Landgraf and Sebastien Pfeffer, unpublished data).

Genomics

The present lack of knowledge concerning the function of miRNA genes has directed the focus on the mode of miRNA expression. As the numbers of identified miRNAs increased, their mode of expression became clearer. In fact, most of our current knowledge on miRNAs derived from their genomic distribution and sequence context. The proximity of some miRNAs to each other has been indicative of their transcription as multicistronic units. Coexpression of closely clustered miRNAs supports this idea

(Baskerville and Bartel, 2005; Lau et al., 2001; Sempere et al., 2004). Some polycistronic transcripts have been confirmed by RT-PCR (Lee et al., 2002; Lee et al., 2004a; Luciano et al., 2004). Intergenic miRNAs constitute 30% of all miRNA genes. Regarding their transcription, the consensus is that pri-miRNAs are transcribed by RNA polymerase II (pol II) with exception of a small number of viral miRNAs that are located downstream of tRNAs and transcribed by pol III (Pfeffer et al., 2005). The few miRNA transcripts fully characterized share features with pol II transcripts like a cap, polyadenylation and splicing but remain and are processed in the nucleus. A few mammalian miRNA pol II promoters have been described in the literature, as in the case of *mir-23a-27a-24-2* cluster (Lee et al., 2004a) and mir-21 (Cai et al., 2004). The tight temporal and spatial regulation of miRNA expression, in some cases an extremely tissue-specific and cell-type expression, raises the question of what functional cis-elements and promoters regulate the transcription of miRNA genes. The first step in understanding the upstream regions of miRNAs would be defining their transcriptional start site using a computational approach searching EST databases or conventional techniques like 5' rapid amplification of cDNA ends (5'RACE). Since miRNA precursors might be rapidly processed, the identification of the start site could be accomplished by isolation of primary miRNA transcripts in cell lines or tissues depleted of Drosha or DGCR8, which are essential in the processing of the primary transcript (Denli et al., 2004; Gregory et al., 2004; Han et al., 2004a; Landthaler et al., 2004). Defining tissue and cell-type specific pol II promoters could provide a valuable tool for gene therapy approaches for localized expression of proteins and non-coding transcripts.

The genomic context of a miRNA provides information about its expression pattern and function. miRNAs encoded in introns, accounting for almost half of the miRNA genes, are expected to be co-transcribed with their host gene (Baskerville and Bartel, 2005; Rodriguez et al., 2004). Thus the expression pattern of the host indicates in which cellular processes the miRNA might be involved. On the other hand, genomic context of an intergenic miRNA can also be related to its expression pattern. For example, miR-196, encoded in the Hox gene cluster, is transcribed during vertebrate development and posttranscriptionally restricts *HOXB8*, a gene in this cluster (Yekta et al., 2004). Such

assumptions about miRNA expression based on genomic location have not yet been considered for genome-wide target predictions.

miRNA function

Over the last few years the miRNA field has drawn a lot of attention. In a short period of time hundreds of miRNA genes with unknown function were discovered. It rapidly became evident that they were widespread among higher eukaryotes. The rare examples of miRNAs with known functions helped to define rules for the interaction between miRNAs and their targets. The availability of a continuously growing list of miRNAs and of an equally expanding list of animal genomes prompted the creation of different computational tools for target identification. The first list of target predictions became available three years after the discovery of miRNAs (Enright et al., 2003; Lewis et al., 2003; Stark et al., 2003). With a limited complementarity between miRNAs and their targets, and scarce examples of validated targets in the literature, it has been a major challenge to produce reliable target predictions.

Computational predictions were described in *Drosophila* (Enright et al., 2003; Rajewsky and Socci, ; Stark et al., 2003) and in mammals (John et al., 2004; Kiriakidou et al., 2004; Lewis et al., 2005; Lewis et al., 2003). If data sets of the same species are compared, the overlap of predicted targets for a certain miRNA is very rare. This inconsistency can be attributed to the basis of their algorithms even though they all parted from similar principles. For instance, the emphasis on the 5' end of the miRNA versus the 3' end is determined differently. Some groups allow G:U wobbles or even mismatches, when located in a high pairing energy context, while others would only consider sites with perfect complementarity from nucleotides 2-8 (5'-3') of the miRNA as putative targets. In some predictions the existence of more than one miRNA binding site adds to the score of a candidate (Enright et al., 2003; Stark et al., 2003), while others do not require more than one targeting site per 3' UTR (Kiriakidou et al., 2004). Despite the evidence for an additive effect of repression of multiple miRNA binding sites (Stark et al., 2003; Vella et al., 2004b), single sites are known to confer regulation in vivo and should not be disregarded (Moss et al., 1997; Yekta et al., 2004). In general, although the assumptions are similar in all computational approaches, the way they are incorporated into the

algorithms is arbitrary. And finally, all predictions use different 3'UTR datasets, which are not well defined in particular for species with less characterized genomes that are used in the homology filters. Therefore in these studies many 3' UTRs were defined as a segment of certain length downstream of the stop codon.

Recently, the Cohen laboratory conducted a systematic study on miRNA target interaction with an in vivo reporter assay (Brennecke et al., 2005). Their findings supported many of the assumptions made by target-finding algorithms. However, this study explored in detail the sequence requirements for miRNA-guided target regulation and expanded the possibilities of miRNA:mRNA binding. Apart from supporting the widely accepted importance of the interaction of the 5' end (seed) of the miRNA with the mRNA, "3' compensatory sites" were defined as sites with a weak seed binding, yet with more extensive complementarity towards the 3' end of the miRNA. The analysis also showed that multiple miRNA binding sites are not necessary for regulation in all cases. Detailed and systematic studies like this one will be the key to the establishment of less ambiguous rules of miRNA:target interactions.

With hundreds of miRNA sequences available, high-throughput methods for miRNA target validation are desirable using a list of miRNA target predictions as a starting point. In order to be more reliable, these predictions need to incorporate expression data of both miRNAs and mRNAs, as well as more detailed information on miRNA:mRNA interactions. But even with better-defined rules and expression information, candidate targets need to be validated experimentally.

For target validation on a genomic scale, a feasible strategy is the use of reporter assays in tissue culture, where a reporter gene fused to the putative target 3'UTR is transfected into the cells. The expression level of the reporter gene bearing the target 3'UTR should be lower than that of a control. If the reporter is regulated, a more detailed set of analyses should follow. Mutations in target sites disrupting miRNA binding is expected to increase the reporter activity (Doench and Sharp, 2004; Lewis et al., 2003). In addition, the miRNA or a miRNA family, regulating the target mRNA, can be blocked in the cells by co-transfection of antisense 2'-O-methyl oligoribonucleotides (Hutvágner et al., 2004; Meister et al., 2004a). Finally, the miRNA can be ectopically expressed in a cell type, where it is normally absent, and one can see if reporter expression levels are reduced.

Several groups have implemented reporter assays in tissue culture. The preferred approach is the dual luciferace assay (Kiriakidou et al., 2004; Lewis et al., 2003; Poy et al., 2004). However, the most comprehensive studies on animal miRNA function have emerged from genetic screens in *C. elegans* and *Drosophila*. In nematodes, defects in developmental timing led to the characterization of *lin-4* more than10 years ago, long before the concepts of RNAi or miRNAs existed. Moreover, more thorough studies on phenotypes of *C elegans let-7* mutants have recently characterized *let-60*/RAS as a novel target for let-7. Interestingly, let-7 expression is lower in lung tumors than in normal lung tissue, while RAS protein is significantly higher in lung tumors, providing a possible role for let-7 in cancer (Johnson et al., 2005). Other screens for specific nervous system asymmetry defects in nematodes lead to the functional characterization of two miRNAs, *lsy-6* and miR-273. Similarly, in *Drosophila*, screens for cell death and cell proliferation resulted in the functional characterization of *bantam* and miR-14 (Brennecke et al., 2003; Xu et al., 2003).

While forward genetic screens as described above are productive, a more straightforward approach would be to determine the phenotype that results from mutating miRNA genes. Reverse genetic approaches, like the creation of knockout animals, in which miRNA genes are deleted, might provide insights into the function of miRNAs in vivo. However, such a time-consuming and costly venture might be complicated due to the presence of several copies of a given miRNA gene in a genome and the potential functional redundancy of sequence related miRNAs.

Model organisms like fly, worm, and mouse, in which genetic analyses are feasible, do not constitute the best choice for high-throughput miRNA target validation. In addition, it may be of great interest to focus on miRNA target genes specific to humans. However, valuable information might be extracted from insights in model organisms as it is illustrated by the finding of the conserved *let-7* target RAS (Johnson et al., 2005), which provides a link between miRNAs and disease and potentially offers new approaches to therapeutic interventions.

On the other hand tissue culture models favor the biochemical characterization of protein-RNA complexes. Such methods may be exploited to study directly the components of human miRNP complexes and associated mRNAs. Small RNA-guided effector silencing

complexes have an Ago protein as a core component (Martinez et al., 2002; Meister et al., 2004b; Mourelatos et al., 2002). Ago proteins likely interact directly with miRNAs, which guide the binding of the effector complexes to target mRNAs. These complexes can be isolated by immunoprecipitation (IP) with antibodies against the Ago proteins. Our laboratory has been successful in stably expressing epitope-tagged Ago proteins in HeLa cells, which allowed IPs with epitope-specific antibodies to characterize the composition of the Ago protein complexes. In addition mRNAs associated with the isolated effector complexes could be identified by microarray analysis. The enriched mRNA population in the immunoprecipitated fraction is very likely under miRNA regulation. A similar approach was used to identify mRNAs associated with FMRP in mouse brain lysates (Brown et al., 2001).

An alternate approach is the CLIP assay (Ule et al., 2003), which consists in UV-irradiating a cell lysate to covalently link RNAs to their associated proteins prior to IPs with antibodies specific to the protein of interest. This procedure allows stringent washes of the isolated protein-RNA complex, increasing the specificity of the assay. In addition, covalent cross-linking allows partial digestion of the RNA, while retaining the core element involved in protein binding. These short RNA tags can be isolated, possibly allowing the identification of the bound target mRNA and the location of the miRNA binding site.

miRNAs and human disease

miRNAs are a large family of gene regulators; therefore, it is foreseeable that some of them have important functions in development and cell physiology. The only known cases of miRNA mutants with conspicuous phenotypes have been found in *C. elegans* and *Drosophila*. miRNAs have been implicated in human diseases based on their expression patterns or on their chromosomal location. miR-155 has been shown to be overexpressed in both, Burkitt lymphoma and Hodgkin lymphoma derived cell lines (Metzler et al., 2004; van den Berg et al., 2003). miR-143 and miR-145 are reduced in precancerous and neoplastic colorectal tissue as well as in several cancer-derived cell lines (Michael et al., 2003). A role in cancer is speculated for miR-15 and -16, since they are located in chromosomal region 13q14, which is deleted in more than 50% of B cell

chronic lymphocytic leukemias (B-CLL), and to a significant extent in other types of cancer. Expression analysis shows that in B-CLLs both miRNAs are down-regulated (Calin et al., 2002) (discussed in Chapter I). An analysis of genomic locations of miRNAs revealed that nearly 20% of the miRNAs studied were located in fragile genomic sites (FRAs) (Calin et al., 2004). FRA-associated genes are frequently deleted or rearranged in cancer cells highlighting their importance in tumorigenesis. Some of the miRNA-containing FRAs do not contain identified tumor suppressor genes, and miRNAs may be good candidates. Recently, a more direct involvement of miRNAs in cancer has been discovered by Johnson and colleagues (Johnson et al., 2005). Reduced levels of let-7 were observed in lung cancer, which inversely correlated with expression of the RAS oncogene. Independently, Karube and colleagues implicated Dicer in lung cancer by showing that Dicer expression levels were reduced in a fraction of lung cancer patients (Karube et al., 2005).

Proteins in the miRNA pathway have also been implicated in human pathology. The human Argonaute proteins Ago1, Ago3, and Ago4, reside in tandem on chromosome 1 (1p34-35), a region frequently lost in Wilms' tumors (Dome and Coppes, 2002). Another human Ago protein, Hiwi, is located on chromosome 12q24.33, a locus linked to the development of testicular germ cell tumors (Skotheim et al., 2001; Summersgill et al., 2001). The protein responsible for fragile X mental retardation syndrome (FMRP), and its *Drosophila* homolog Fmr1 interact with components of the RNA silencing pathway. Mammalian FMRP was co-immunoprecipitated with Dicer and Ago2 (Jin et al., 2004), and *Drosophila* Fmr1 was found to be associated to holo-RISC (Caudy et al., 2002; Ishizuka et al., 2002).

Outlook

Besides miRNA identification, profiling and target identification, several questions remain unsolved that are of great interest and biological significance. Once the function is elucidated for a number of miRNAs, many questions regarding miRNA-dependent regulation can be addressed. The mechanism of translational repression guided by miRNAs remains to be elucidated. It is not known whether this translational arrest is reversible and if the translation process is regulated at the step of initiation or elongation.

Analyses of miRNA-mediated silencing detected a decrease on the protein level, but the mRNA level remained almost unchanged (Doench et al., 2003; Olsen and Ambros, 1999; Zeng et al., 2002). This also raises the question of stability of a miRISC and if it undergoes multiple turnover interactions with its target mRNA. Experimental evidence for siRNAs suggests that once in siRISC, siRNAs are very stable and do not dissociate (Martinez et al., 2002). It remains to be clarified if the same is true for miRNA-containing complexes and whether target availability has any effect on the stability of the complex.

Knowledge about the factors regulating miRNA expression, such as the identity of their promoters and the changes in miRNA expression levels as a response to different cellular inputs will also shed light on the regulatory network in which miRNAs are involved in. Some steps in the miRNA processing pathway remain to be identified. Though Drosha and Dicer, the key enzymes in maturation of miRNAs, have been identified, the nuclear processes by which the Drosha-containing microprocessor identifies a miRNA primary transcript, are unknown.

Finally, one of the most challenging tasks is the identification of the biochemical components and the function of the different Argonaute-containing complexes. It is now known that fly and human Ago2 are the RISC “slicer”, but the purpose of the remaining members of this protein family is unknown. The same is true for the molecular composition and the in vivo function of each effector complex. Clearly, an understanding of the function of the Ago-containing effector complexes in transcriptional and posttranscriptional gene silencing will provide insight into the complexity of small RNA-mediated gene regulation and chromosome structure.

Summary

This work describes the identification of miRNAs from *Drosophila*, mouse and human, and the analysis of their expression patterns. miRNAs were cloned from 15 mouse tissues and tissue-specific miRNAs were identified in liver, heart and neuronal tissue. Developmentally regulated miRNAs were described in *Drosophila*. Bioinformatic analysis of the miRNAs led to identification of their genomic location, which are insights into miRNA biogenesis and function. Some miRNAs were found to be linked to disease-associated genomic loci. Features of miRNA expression, such as clustering of miRNA genes and their location within introns of protein coding genes or intergenic regions emerged. Furthermore the evolutionary conservation of miRNAs in vertebrate and invertebrate genomes was assessed.

The identification and profiling of miRNAs in human, mouse and *Drosophila* along with the findings by others in *C. elegans* provided the groundwork for the study of small RNA mediated gene regulation.

As the impact of this discovery became apparent, the interest for identifying miRNAs and other regulatory small RNAs, like rasiRNAs, increased. Our laboratory started a project for large-scale cloning of small RNAs. I therefore improved the cloning procedure, making it more efficient and shortening its steps. I also evaluated its accuracy in reflecting the small RNA population in a cell or tissue.

Bibliography

Abrahante, J. E., Daul, A. L., Li, M., Volk, M. L., Tennessen, J. M., Miller, E. A., and Rougvie, A. E. (2003). The *Caenorhabditis elegans* hunchback-like gene *lin-57/hbl-1* controls developmental time and is regulated by microRNAs. Dev Cell *4*, 625-637.

Ambros, V. (2000). Control of developmental timing in *Caenorhabditis elegans*. Curr Opin Genet Dev *10*, 428-433.

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.

Amitsur, M., Levitz, R., and Kaufmann, G. (1987). Bacteriophage T4 anticodon nuclease, polynucleotide kinase and RNA ligase reprocess the host lysine tRNA. EMBO J *6*, 2499-2503.

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.

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.

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

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

Baulcombe, D. C. (1996). RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. Plant Mol Biol *32*, 79-88.

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.

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

Bohnsack, M. T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA *10*, 185-191.

Bonnet, E., Wuyts, J., Rouze, P., and Van de Peer, Y. (2004). Detection of 91 potential conserved plant microRNAs in Arabidopsis thaliana and Oryza sativa identifies important target genes. Proc Natl Acad Sci U S A *101*, 11511-11516.

Boutet, S., Vazquez, F., Liu, J., Béclin, C., Fagard, M., Gratias, A., Morel, J. B., Crete, P., Chen, X., and Vaucheret, H. (2003). Arabidopsis HEN1. A Genetic Link between Endogenous miRNA Controlling Development and siRNA Controlling Transgene Silencing and Virus Resistance. Curr Biol *13*, 843-848.

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.

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

Brown, V., Jin, P., Ceman, S., Darnell, J. C., O'Donnell, W. T., Tenenbaum, S. A., Jin, X., Feng, Y., Wilkinson, K. D., Keene, J. D., et al. (2001). Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. Cell *107*, 477-487.

Bruce, A. G., and Uhlenbeck, O. C. (1978). Reactions at the termini of tRNA with T4 RNA ligase. Nucleic Acids Res *5*, 3665-3677.

Cai, X., Hagedorn, C. H., and Cullen, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA *10*, 1957-1966.

Calin, G. A., Dumitru, C. D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., et al. (2002). Frequent deletions and down-regulation of microRNA genes miR-15 and miR-16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A *99*, 15524-15529.

Calin, G. A., Sevignani, C., Dumitru, C. D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M., and Croce, C. M. (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A *101*, 2999-3004.

Carmell, M. A., and Hannon, G. J. (2004). RNase III enzymes and the initiation of gene silencing. Nat Struct Mol Biol *11*, 214-218.

Carmell, M. A., Xuan, Z., Zhang, M. Q., and Hannon, G. J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev *16*, 2733-2742.

Caudy, A. A., Myers, M., Hannon, G. J., and Hammond, S. M. (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev *16*, 2491-2496.

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.

Chan, S. W., Zilberman, D., Xie, Z., Johansen, L. K., Carrington, J. C., and Jacobsen, S. E. (2004). RNA silencing genes control de novo DNA methylation. Science *303*, 1336.

Chang, S., Johnston, R. J., Jr., Frokjaer-Jensen, C., Lockery, S., and Hobert, O. (2004). MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. Nature *430*, 785-789.

Chen, C. Z., Li, L., Lodish, H. F., and Bartel, D. P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. Science *303*, 83-86.

Chen, C. Z., and Lodish, H. F. (2005). MicroRNAs as regulators of mammalian hematopoiesis. Semin Immunol *17*, 155-165.

Chiu, Y. L., and Rana, T. M. (2003). siRNA function in RNAi: a chemical modification analysis. RNA *9*, 1034-1048.

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

Cook, H. A., Koppetsch, B. S., Wu, J., and Theurkauf, W. E. (2004). The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. Cell *116*, 817-829.

Cranston, J. W., Silber, R., Malathi, V. G., and Hurwitz, J. (1974). Studies on ribonucleic acid ligase. Characterization of an adenosine triphosphate-inorganic pyrophosphate exchange reaction and demonstration of an enzyme-adenylate complex with T4 bacteriophage-induced enzyme. J Biol Chem *249*, 7447-7456.

Cullen, B. R. (2004). Transcription and Processing of Human microRNA Precursors. Mol Cell *16*, 861-865.

Dalmay, T., Horsefield, R., Braunstein, T. H., and Baulcombe, D. C. (2001). SDE3 encodes an RNA helicase required for post-transcriptional gene silencing in Arabidopsis. EMBO J *20*, 2069-2078.

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

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

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

Dome, J. S., and Coppes, M. J. (2002). Recent advances in Wilms tumor genetics. Curr Opin Pediatr *14*, 5-11.

Dostie, J., Mourelatos, Z., Yang, M., Sharma, A., and Dreyfuss, G. (2003). Numerous microRNPs in neuronal cells containing novel microRNAs. RNA *9*, 180-186.

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

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

England, T. E., and Uhlenbeck, O. C. (1978). Enzymatic oligoribonucleotide synthesis with T4 RNA ligase. Biochemistry *17*, 2069-2076.

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. (2003). MicroRNA targets in *Drosophila*. Genome Biol *5:RI*, 1-14.

Esau, C., Kang, X., Peralta, E., Hanson, E., Marcusson, E. G., Ravichandran, L. V., Sun, Y., Koo, S., Perera, R. J., Jain, R., et al. (2004). MicroRNA-143 regulates adipocyte differentiation. J Biol Chem *279*, 52361-52365.

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.

Fukagawa, T., Nogami, M., Yoshikawa, M., Ikeno, M., Okazaki, T., Takami, Y., Nakayama, T., and Oshimura, M. (2004). Dicer is essential for formation of the heterochromatin structure in vertebrate cells. Nat Cell Biol *6*, 784-791.

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.

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

Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res *32*, D109-111.

Grishok, A., Pasquinelli, A. E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D. L., Fire, A., Ruvkun, G., and Mello, C. C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. Cell *106*, 23-34.

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

Ha, I., Wightman, B., and Ruvkun, G. (1996). A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation. Genes Dev *10*, 3041-3050.

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

Hall, I. M., Shankaranarayana, G. D., Noma, K., Ayoub, N., Cohen, A., and Grewal, S. I. (2002). Establishment and maintenance of a heterochromatin domain. Science *297*, 2232-2237.

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

Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R., and Hannon, G. J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146-1150.

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

Han, M. H., Goud, S., Song, L., and Fedoroff, N. (2004b). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proc Natl Acad Sci U S A *101*, 1093-1098.

Ho, C. K., Wang, L. K., Lima, C. D., and Shuman, S. (2004). Structure and mechanism of RNA ligase. Structure (Camb) *12*, 327-339.

Houbaviy, H. B., Murray, M. F., and Sharp, P. A. (2003). Embryonic stem cell-specific MicroRNAs. Dev Cell *5*, 351-358.

Hutvágner, G., McLachlan, J., Bálint, É., Tuschl, T., and Zamore, P. D. (2001). A cellular function for the RNA interference enzyme Dicer in small temporal RNA maturation. Science *93*, 834-838.

Hutvágner, G., Mlynarova, L., and Nap, J. P. (2000). Detailed characterization of the posttranscriptional gene-silencing-related small RNA in a GUS gene-silenced tobacco. RNA *6*, 1445-1454.

Hutvágner, G., Simard, M. J., Mello, C. C., and Zamore, P. D. (2004). Sequence-specific inhibition of small RNA function. PLoS Biol *2*, E98.

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

Ishizuka, A., Siomi, M. C., and Siomi, H. (2002). A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. Genes Dev *16*, 2497-2508.

Jin, P., Zarnescu, D. C., Ceman, S., Nakamoto, M., Mowrey, J., Jongens, T. A., Nelson, D. L., Moses, K., and Warren, S. T. (2004). Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. Nat Neurosci *7*, 113-117.

John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C., and Marks, D. (2004). Human miRNA targets. PLoS Biol *2*, e363.

Johnson, S. M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K. L., Brown, D., and Slack, F. J. (2005). RAS Is Regulated by the let-7 MicroRNA Family. Cell *120*, 635-647.

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

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.

Kanellopoulou, C., Muljo, S. A., Kung, A. L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D. M., and Rajewsky, K. (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev *19*, 489-501.

Karube, Y., Tanaka, H., Osada, H., Tomida, S., Tatematsu, Y., Yanagisawa, K., Yatabe, Y., Takamizawa, J., Miyoshi, S., Mitsudomi, T., and Takahashi, T. (2005). Reduced expression of Dicer associated with poor prognosis in lung cancer patients. Cancer Sci *96*, 111-115.

Kawasaki, H., and Taira, K. (2004). Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. Nature *431*, 211-217.

Kennerdell, J. R., Yamaguchi, S., and Carthew, R. W. (2002). RNAi is activated during Drosophila oocyte maturation in a manner dependent on aubergine and spindle-E. Genes Dev *16*, 1884-1889.

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

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

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 U S A *101*, 360-365.

Kiriakidou, M., Nelson, P. T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., and Hatzigeorgiou, A. (2004). A combined computational-experimental approach predicts human microRNA targets. Genes Dev *18*, 1165-1178.

Kloosterman, W. P., Wienholds, E., Ketting, R. F., and Plasterk, R. H. (2004). Substrate requirements for let-7 function in the developing zebrafish embryo. Nucleic Acids Res *32*, 6284-6291.

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.

Krug, M., and Uhlenbeck, O. C. (1982). Reversal of T4 RNA ligase. Biochemistry *21*, 1858-1864.

Kurihara, Y., and Watanabe, Y. (2004). Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc Natl Acad Sci U S A *101*, 12753-12758.

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

Lai, E. C., Tomancak, P., Williams, R. W., and Rubin, G. M. (2003). Computational identification of Drosophila microRNA genes. Genome Biol *4*, R42.

Landthaler, M., Yalcin, A., and Tuschl, T. (2004). The Human DiGeorge Syndrome Critical Region Gene 8 and Its D. melanogaster Homolog Are Required for miRNA Biogenesis. Curr Biol *14*, 2162-2167.

Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science *294*, 858-862.

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

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843-854.

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

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

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H., and Kim, V. N. (2004a). MicroRNA genes are transcribed by RNA polymerase II. EMBO J *23*, 4051-4060.

Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J., and Carthew, R. W. (2004b). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell *117*, 69-81.

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15-20.

Lewis, B. P., Shih, I., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

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

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S., and Johnson, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769-773.

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

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

Lin, S. Y., Johnson, S. M., Abraham, M., Vella, M. C., Pasquinelli, A., Gamberi, C., Gottlieb, E., and Slack, F. J. (2003). The *C. elegans* hunchback homolog, *hbl-1*, controls temporal patterning and is a probable microRNA target. Dev Cell *4*, 639-650.

Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2003). Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature *426*, 465-469.

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

Liu, Q., Rand, T. A., Kalidas, S., Du, F., Kim, H. E., Smith, D. P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921-1925.

Llave, C., Xie, Z., Kasschau, K. D., and Carrington, J. C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053-2056.

Luciano, D. J., Mirsky, H., Vendetti, N. J., and Maas, S. (2004). RNA editing of a miRNA precursor. RNA *10*, 1174-1177.

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

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

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

Martinez, J., and Tuschl, T. (2004). RISC is a 5' phosphomonoester-producing RNA endonuclease. Genes Dev *18*.

Matzke, M. A., and Birchler, J. A. (2005). RNAi-mediated pathways in the nucleus. Nat Rev Genet *6*, 24-35.

McLaughlin, L. W., Piel, N., and Graeser, E. (1985). Donor activation in the T4 RNA ligase reaction. Biochemistry *24*, 267-273.

Meister, G., Landthaler, M., Dorsett, Y., and Tuschl, T. (2004a). Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. RNA *10*, 544-550.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004b). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol Cell *15*, 185-197.

Meister, G., and Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. Nature *431*, 343-349.

Mello, C. C., and Conte, D., Jr. (2004). Revealing the world of RNA interference. Nature *431*, 338-342.

Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A., and Matzke, A. J. M. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J *19*, 5194-5201.

Metzler, M., Wilda, M., Busch, K., Viehmann, S., and Borkhardt, A. (2004). High expression of precursor microRNA-155/BIC RNA in children with Burkitt lymphoma. Genes Chromosomes Cancer *39*, 167-169.

Michael, M. Z., O' Connor, S. M., van Holst Pellekaan, N. G., Young, G. P., and James, R. J. (2003). Reduced accumulation of specific microRNAs in colorectal neoplasia. Mol Cancer Res *1*, 882-891.

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.

Morris, K. V., Chan, S. W., Jacobsen, S. E., and Looney, D. J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. Science *305*, 1289-1292.

Moss, E. G., Lee, R. C., and Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the lin-4 RNA. Cell *88*, 637-646.

Motamedi, M. R., Verdel, A., Colmenares, S. U., Gerber, S. A., Gygi, S. P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. Cell *119*, 789-802.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev *16*, 720-728.

Nandakumar, J., Ho, C. K., Lima, C. D., and Shuman, S. (2004). RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2. J Biol Chem *279*, 31337-31347.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a chimeric chalcone synthase gene into Petunia results in reversible co-suppression of homologous genes in trans. Plant Cell *2*, 279-289.

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 Methods *1*, 155-161.

Noma, K., Sugiyama, T., Cam, H., Verdel, A., Zofall, M., Jia, S., Moazed, D., and Grewal, S. I. (2004). RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. Nat Genet *36*, 1174-1180.

Ohler, U., Yekta, S., Lim, L. P., Bartel, D. P., and Burge, C. B. (2004). Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. RNA *10*, 1309-1322.

Okamura, K., Ishizuka, A., Siomi, H., and Siomi, M. C. (2004). Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev *18*, 1655-1666.

Olsen, P. H., and Ambros, V. (1999). The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. Dev Biol *216*, 671-680.

Orban, T. I., and Izaurralde, E. (2005). Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. Rna *11*, 459-469.

Pal-Bhadra, M., Bhadra, U., and Birchler, J. A. (1997). Cosuppression in Drosophila: gene silencing of *Alcohol dehydrogenase* by *white-Adh* transgenes is *Polycomb* dependent. Cell *90*, 479-490.

Pal-Bhadra, M., Leibovitch, B. A., Gandhi, S. G., Rao, M., Bhadra, U., Birchler, J. A., and Elgin, S. C. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science *303*, 669-672.

Papp, I., Mette, M. F., Aufsatz, W., Daxinger, L., Schauer, S. E., Ray, A., van der Winden, J., Matzke, M., and Matzke, A. J. (2003). Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. Plant Physiol *132*, 1382-1390.

Park, M. Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., and Poethig, R. S. (2005). Nuclear processing and export of microRNAs in Arabidopsis. Proc Natl Acad Sci U S A.

Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Curr Biol *12*, 1484-1495.

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

Pfeffer, S., Alain , S., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F. A., van Dyk, L. F., Ho, C. K., Shuman, S., Chien, M., et al. (2005). Identification of microRNAs of the herpesvirus family. Nat Methods *2*, 269-276.

Pfeffer, S., Lagos-Quintana, M., and Tuschl, T. (2003). Cloning of small RNA molecules. In Current Protocols in Molecular Biology, F. M. Ausubel, R. Brent, R. E. Kingston , D.

D. Moore, J. G. Seidmann, J. A. Smith, and K. Struhl, eds. (New York, John Wiley and Sons), pp. 26.24.21-26.24.16.

Pfeffer, S., Zavolan, M., Grasser, F. A., Chien, M., Russo, J. J., Ju, J., John, B., Enright, A. J., Marks, D., Sander, C., and Tuschl, T. (2004). Identification of virus-encoded microRNAs. Science *304*, 734-736.

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.

Poy, M. N., Eliasson, L., Krutzfeldt, J., Kuwajima, S., Ma, X., Macdonald, P. E., Pfeffer, S., Tuschl, T., Rajewsky, N., Rorsman, P., and Stoffel, M. (2004). A pancreatic islet-specific microRNA regulates insulin secretion. Nature *432*, 226-230.

Rajewsky, N., and Socci, N. D. (2004). Computational identification of microRNA targets. Dev Biol *267*, 529-535.

Rand, T. A., Ginalski, K., Grishin, N. V., and Wang, X. (2004). Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. Proc Natl Acad Sci U S A *101*, 14385-14389.

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

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., Horvitz, H. R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature *403*, 901-906.

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.

Robertson, H. D. (1982). *Escherichia coli* ribonuclease III cleavage sites. Cell *30*, 669-672.

Robertson, H. D., Webster, R. E., and Zinder, N. D. (1968). Purification and properties of ribonuclease III from Escherichia coli. J Biol Chem *243*, 82-91.

Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L., and Bradley, A. (2004). Identification of Mammalian microRNA Host Genes and Transcription Units. Genome Res *14*, 1902-1910.

Romaniuk, E., McLaughlin, L. W., Neilson, T., and Romaniuk, P. J. (1982). The effect of acceptor oligoribonucleotide sequence on the T4 RNA ligase reaction. EMBO J *125*, 639-643.

Schotta, G., Ebert, A., and Reuter, G. (2003). SU(VAR)3-9 is a conserved key function in heterochromatic gene silencing. Genetica *117*, 149-158.

Schramke, V., and Allshire, R. (2003). Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. Science *301*, 1069-1074.

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

Schwarz, D. S., Tomari, Y., and Zamore, P. D. (2004). The RNA-induced silencing complex Is a Mg(2+)-dependent endonuclease. Curr Biol *14*, 787-791.

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.

Sempere, L. F., Freemantle, S., Pitha-Rowe, I., Moss, E., Dmitrovsky, E., and Ambros, V. (2004). Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biol *5*, R13.

Shiohama, A., Sasaki, T., Noda, S., Minoshima, S., and Shimizu, N. (2003). Molecular cloning and expression analysis of a novel gene DGCR8 located in the DiGeorge syndrome chromosomal region. BBRC *304*, 184-190.

Skotheim, R. I., Kraggerud, S. M., Fossa, S. D., Stenwig, A. E., Gedde-Dahl, T., Jr., Danielsen, H. E., Jakobsen, K. S., and Lothe, R. A. (2001). Familial/bilateral and sporadic testicular germ cell tumors show frequent genetic changes at loci with suggestive linkage evidence. Neoplasia *3*, 196-203.

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

Song, J. J., Liu, J., Tolia, N. H., Schneiderman, J., Smith, S. K., Martienssen, R. A., Hannon, G. J., and Joshua-Tor, L. (2003). The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. Nat Struct Biol *10*, 1026-1032.

Sontheimer, E. J. (2005). Assembly and function of RNA silencing complexes. Nat Rev Mol Cell Biol.

Stark, A., Brennecke, J., Russell, R. B., and Cohen, S. M. (2003). Identification of *Drosophila* microRNA targets. PLoS Biol *1*, 397-409.

Summersgill, B., Osin, P., Lu, Y. J., Huddart, R., and Shipley, J. (2001). Chromosomal imbalances associated with carcinoma in situ and associated testicular germ cell tumours of adolescents and adults. Br J Cancer *85*, 213-220.

Tabara, H., Yigit, E., Siomi, H., and Mello, C. C. (2002). The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans. Cell *109*, 861-871.

Thomson, J. M., Parker, J., Perou, C. M., and Hammond, S. M. (2004). A custom microarray platform for analysis of microRNA gene expression. Nat Methods *1*, 47-53.

Tomari, Y., Du, T., Haley, B., Schwarz, D. S., Bennett, R., Cook, H. A., Koppetsch, B. S., Theurkauf, W. E., and Zamore, P. D. (2004a). RISC assembly defects in the Drosophila RNAi mutant armitage. Cell *116*, 831-841.

Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P. D. (2004b). A protein sensor for siRNA asymmetry. Science *306*, 1377-1380.

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.

van den Berg, A., Kroesen, B. J., Kooistra, K., de Jong, D., Briggs, J., Blokzijl, T., Jacobs, S., Kluiver, J., Diepstra, A., Maggio, E., and Poppema, S. (2003). High expression of B-cell receptor inducible gene BIC in all subtypes of Hodgkin lymphoma. Genes Chromosomes Cancer *37*, 20-28.

Vazquez, F., Gasciolli, V., Crete, P., and Vaucheret, H. (2004). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr Biol *14*, 346-351.

Vella, M. C., Choi, E. Y., Lin, S. Y., Reinert, K., and Slack, F. J. (2004a). The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. Genes Dev *18*, 132-137.

Vella, M. C., Reinert, K., and Slack, F. J. (2004b). Architecture of a Validated MicroRNA::Target Interaction. Chem Biol *11*, 1619-1623.

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

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

Wang, Q. S., and Unrau, P. J. (2002). Purification of histidine-tagged T4 RNA ligase from E. coli. Biotechniques *33*, 1256-1260.

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.

Waterhouse, P. M., Graham, M. W., and Wang, M. B. (1998). Virus resistance and gene silencing in plants can be induced by simultaneous expression of sense and antisense RNA. Proc Natl Acad Sci U S A *95*, 13959-13964.

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.

Wu-Scharf, D., Jeong, B., Zhang, C., and Cerutti, H. (2000). Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-Box RNA helicase. Science *290*, 1159-1163.

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.

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.

Xie, Z., Kasschau, K. D., and Carrington, J. C. (2003). Negative feedback regulation of Dicer-Like1 in Arabidopsis by microRNA-guided mRNA degradation. Curr Biol *13*, 784-789.

Xu, P., Vernooy, S. Y., Guo, M., and Hay, B. A. (2003). The *Drosophila* microRNA mir-14 suppresses cell death and is required for normal fat metabolism. Curr Biol *13*, 790-795.

Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M. M. (2003). Structure and conserved RNA binding of the PAZ domain. Nature *426*, 468-474.

Yekta, S., Shih, I. H., and Bartel, D. P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science *304*, 594-596.

Yi, R., Qin, Y., Macara, I. G., and Cullen, B. R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev *17*, 3011-3016.

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.

Zeng, Y., Wagner, E. J., and Cullen, B. R. (2002). Both natural and designed microRNAs can inhibit the expression of cognate mRNAs when expressed in human cells. Mol Cell *9*, 1327-1333.

Zilberman, D., Cao, X., and Jacobsen, S. E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716-719.

Abbreviations

ASEL	asymmetric chemosensory neuron left
ASER	asymetric chemosensory neuron right
ATP	adenosine triphosphate
B-CLL	B cell chronic lymphocytic leukemia
C. elegans	*Caenorhabditis elegans*
cDNA	complementary DNA
DNA	deoxyribonucleic acid
dsRBD	double stranded RNA binding domain
dsRNA	double stranded RNA
E. coli	*Escherichia coli*
FMRP	fragile X mental retardation protein
FPLC	fast protein liquid chromatography
GFP	green fluorescent protein
H. sapiens	*Homo sapiens*
IPTG	isopropyl-β-D-thiogalactopyranoside
M. musculus	*Mus musculus*
miRNA	microRNA
mRNA	messenger RNA
ncRNA	non-coding RNA
nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pre-miRNA	precursor miRNA
pri-miRNA	primary miRNA
PTGS	posttranscriptional gene silencing
rasiRNA	repeat-associated small interfering RNA
RDRC	RNA dependent RNA polymerase complex
RISC	RNA induced silencing complex
RITS	RNA-induced initiation of transcriptional gene silencing
RLC	RISC loading complex
RNA	ribonucleic acid
rRNA	ribosomal RNA
RT-PCR	reverse transcription-PCR
S. pombe	*Schizosacharomyces pombe*
SDS-PAGE	sodium dodecyl sulfate-PAGE
siRNA	small interfering RNA
ssRNA	single stranded RNA
tRNA	transfer RNA
UTR	untranslated region

Appendix - Updated miRNA names -

Some miRNA sequences presented in published tables were later found to have mistakes. Cloned sequences published that do not have a genomic match are likely due PCR errors, so with evidence lacking for their existence they were removed from the miRNA Registry. Sequences with one or two base changes are assigned suffixes of the form miR-181a and miR-181b, thus, some of the published sequences listed above have been renamed when a close related sequence appeared in the literature. In the case of miRNAs encoded by the same precursor, the forms miR-142–s and miR-142–as are no longer used. Nomenclature changed in favor of using names of the form miR-142-5p and miR-142-3p to designate miRNAs from the 5' and 3' arms, respectively. Where cloning studies allow to determine which arm of the precursor gives rise to the predominantly expressed miRNA, an asterisk denotes the less predominant form, as in miR-199a and miR-199a*

Published name	Rfam entry[a]	Chapter	Notes
let-7h	no entry	I	PCR error. Cloned twice. No precursor shown in Paper II, Table S2 (Chapter I)
miR-1b	no entry	I	PCR error. Chapter I miR-1b precursor shown corresponds to hsa-mir-1-1
miR-1c	no entry	I	PCR error. Cloned 7 times from mouse heart where miR-1 is very abundant
miR-1d	miR-1		The onfly mammalian miR-1 in Rfam
miR-30a-s	miR-30a-5p	I	Based in recent data[b] this miRNA is the predomingnly expressed miRNA.
miR-30a-as	miR-30a-as	I	Based in recent data[b] this is the miRNA*
miR-101	miR-101a	I	
miR-122b	no entry	I	PCR error
miR-124b	no entry	I	PCR error. Chapter I table S2 shows precursor that corresponds to hsa-miR-124a-3
miR-128	miR-128a	I	
miR-130	miR-130a	I	
miR-131	miR-9*	I	This miRNA* is encoded by the 3' arm of miR-9
miR-133	miR-133a	I	
miR-135	miR-135a	I	Table 1 (Chapter I) miR-135 sequence has a 5' terminal A which is not encoded by any of the two mir-135 gene copies
miR-142-s	miR-142-5p	I	Figure 8B suggests that this miRNA is the predominant miRNA
miR-142-as	miR-142-3p	I	Figure R1 B suggests that this miRNA is the miRNA*

miR-147	no mouse entry	I	The sequence cloned in paper II from mouse spleen is the human miR-147. There is no match found in the mouse genome. Figure S2 (Chapter I) shows the human precursor
miR-148	miR-148a	I	
miR-181	miR-181a	I	
miR-189	no entry	II	This miRNA is miR-24*. It derives from the 5' arm of the mir-24 precursor. Rfam kept the original name for the mature miR-189.
miR-196	miR-196a	II	
miR-199-s	miR-199a	II	
miR-199-as	miR-199a*	II	

[a] Rfam release 5.1 (Griffiths-Jones, 2004)

[b] Pablo Landgraf and Sebastien Pfeffer, unpublished data.

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.
Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

New York, den 14. April, 2005

_____

Curriculum Vitae

Mariana Lagos Quintana

RNA Molecular Biology Laboratory
The Rockefeller University
1230 York Ave Box 186
10021, New York, NY

Tel: 212 327 7697
Fax: 212 327 7652
email: lagosm@rockefeller.edu

Education/Positions

1996-2000	B.S. Biology, Biotechnology, Universidad de las Américas-Puebla, Puebla, Mexico.
2000	B.S. thesis: Caracterization of a novel mutation causing coagulation factorXI deficiency, Laboratorios Clínicos de Puebla, Puebla, Mexico
2000-date	PhD Student in Molecular Biology, University of Göttingen, Göttingen, Germany. Visiting graduate student, Rockefeller University, New York.

Awards

- 2002 AAAS Newcomb Cleveland Prize
- 1997-2000 Scholarship Support to Sciences, Jenkins Foundation, UDLA-P, Mexico

Publications

Pfeffer, S., Alain , S., **Lagos-Quintana, M.**, Sheridan, R., Sander, C., Grasser, F. A., van Dyk, L. F., Ho, C. K., Shuman, S., Chien, M., et al. (2005). Identification of microRNAs of the herpesvirus family. *Nat. Methods* **2**, 269-276.

Pfeffer, S., **Lagos-Quintana, M.**, and Tuschl, T. (2003). Cloning of small RNA molecules. In Current Protocols in Molecular Biology, F. M. Ausubel, R. Brent, R. E. Kingston , D. D. Moore, J. G. Seidmann, J. A. Smith, and K. Struhl, eds. (New York, John Wiley and Sons), pp. 26.24.21-26.24.16.

Aravin, A. A., **Lagos-Quintana, M**., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. *Dev. Cell* **5**, 337-350.

Bechert, K., **Lagos-Quintana, M.**, Harborth, J., Weber, K., Osborn, M. (2003) Effects of expressing lamin A mutant protein causing Emery-Dreifuss muscular dystrophy and familial partial lipodystrophy in HeLa cells. *Exp. Cell. Res.* **286**, 75-86.

Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. (2003) New microRNAs from mouse and human. *RNA* **9**, 175-179.

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.

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

Patent

T. Tuschl, **M. Lagos-Quintana**, W. Lendeckel, J. Meyer, R. Rauhut, Max-Planck Gesellschaft zur Förderung der Wissenschaften e. V. (2002) Identification of novel genes coding for small temporal RNAs.

Recent Presentations

Poster presentation "Optimization of small RNA cloning" Keystone Symposia, Keystone CO, April 14th-19th, 2004.

Short talk, "Small RNA profile during *Drosophila melanogaster* development" Epigenetics Annual Meeting of the German Genetics Society, Kassel, Germany, Sept. 26th-29th, 2003