New microRNAs from mouse and human

MARIANA LAGOS-QUINTANA,1 REINHARD RAUHUT,1 JUTTA MEYER,1 ARNDT BORKHARDT,2 and THOMAS TUSCHL1

1Department of Cellular Biochemistry, Max-Planck-Institute for Biophysical Chemistry, D-37077 Göttingen, Germany
2Children’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 foldback (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).

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).

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

Article and publication are at http://www.rnajournal.org/cgi/doi/10.1261/rna.2146903.
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
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-15a and miR-16) 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 non-coding 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-15a and miR-16: These miRNAs have the capacity to fold with the miRNA into a hairpin precursor. (a) 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.

### 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. Dow for some of the sequencing, A. Aravin and

---

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>let-7/miR-98</td>
<td>p c n</td>
<td>p c n</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>lin-4/miR-125</td>
<td>p c n</td>
<td>p c n</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-1</td>
<td>p c n</td>
<td>p c n</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-4/75/79/131</td>
<td>p c</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p</td>
<td></td>
</tr>
<tr>
<td>miR-7</td>
<td>p p</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-8/141</td>
<td>p c</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-9</td>
<td>p c</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-10/100</td>
<td>p c</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-29/83</td>
<td>p c</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-31/73</td>
<td>p c</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-34</td>
<td>p p</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-25/92</td>
<td>p c</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-124</td>
<td>p c</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-133</td>
<td>p c</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
<tr>
<td>miR-184</td>
<td>p c</td>
<td>p c</td>
<td>p n</td>
<td>p c n</td>
<td>p c n</td>
<td>p c n</td>
</tr>
</tbody>
</table>

(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.
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


