Physiol Genomics 29: 35–43, 2007.
First published November 14, 2006; doi:10.1152/physiolgenomics.00081.2006.

Discovery and profiling of bovine microRNAs from immune-related and embryonic tissues

Luiz L. Coutinho,[2]* Lakshmi K. Matukumalli,[1,3]* Tad S. Sonstegard,[1] Curtis P. Van Tassell,[1] Louis C. Gasbarre,[1] Anthony V. Capuco,[1] and Timothy P. L. Smith[4]

[1]*United States Department of Agriculture, Agricultural Research Center, Beltsville Area Research Center, Beltsville, Maryland;* [2]*University of Sao Paulo-ESALQ, Piracicaba, SP, Brazil;* [3]*Bioinformatics and Computational Biology, George Mason University, Manassas, Virginia; and* [4]*United States Department of Agriculture, Agricultural Research Center, U.S. Meat Animal Research Center, Clay Center, Nebraska*

Submitted 4 May 2006; accepted in final form 8 November 2006

Coutinho LL, Matukumalli LK, Sonstegard TS, Van Tassell CP, Gasbarre LC, Capuco AV, Smith TP. Discovery and profiling of bovine microRNAs from immune-related and embryonic tissues. *Physiol Genomics* 29: 35–43, 2007. First published November 14, 2006; doi:10.1152/physiolgenomics.00081.2006.—MicroRNAs are small ~22 nucleotide-long noncoding RNAs capable of controlling gene expression by inhibiting translation. Alignment of human microRNA stem-loop sequences (mir) against a recent draft sequence assembly of the bovine genome resulted in identification of 334 predicted bovine mir. We sequenced five tissue-specific cDNA libraries derived from the small RNA fractions of bovine embryo, thymus, small intestine, and lymph node to validate these predictions and identify new mir. This strategy combined with comparative sequence analysis identified 129 sequences that corresponded to mature microRNAs (miR). A total of 107 sequences aligned to known human mir, and 100 of these matched expressed miR. The other seven sequences represented novel miR expressed from the complementary strand of previously characterized human mir. The 22 sequences without matches displayed characteristic mir secondary structures when folded in silico, and 10 of these retained sequence conservation with other vertebrate species. Expression analysis based on sequence identity counts revealed that some miR were preferentially expressed in certain tissues, while bta-miR-26a and bta-miR-103 were prevalent in all tissues examined. These results support the premise that species differences in regulation of gene expression by miR occur primarily at the level of expression and processing.

small RNA; microRNA; embryo; immune

MICRORNAS ARE SMALL MOLECULES (~22 nucleotides in length) that influence the expression of hundreds of genes (24) and have a role in regulation of gene expression for numerous biological processes including brain morphogenesis (17, 24), cardiomyocyte proliferation and differentiation (41), insulin secretion (32), tumorogenesis, viral defense (23), and hematopoietic lineage differentiation (12). Mature microRNAs (miR) in animals interact mostly with the 3′-untranslated region (UTR) of targeted mRNA and modulate gene expression (3, 4, 7, 19, 20, 30, 31, 40).

miR have been identified by sequence and expression analyses (5). Genome sequence analysis algorithms based on phylogenetic conservation (10) and RNA folding (27) have identified potential stem-loop structures containing microRNAs (mir) in many species with available genome sequence. However, validation of these predictions has required detection of miR transcripts by Northern blot or by sequencing of cDNA libraries derived from size-fractioned RNA (5). The latter approach has also identified miR not predicted by in silico methods.

There are currently 4,039 mir described for primates, rodents, birds, fish, worms, flies, plants, and viruses. A total of 462 mir have been described for humans (18) (release 8.2). Genomic scans and cloning results have indicated that the actual total number of human mir may be closer to 800 (9). Some mir maintain conservation across vertebrate species, while others have a more limited species distribution (9). Probably the most interesting observations pertain to variation in miR abundance and expression among different tissues (6, 15, 24, 25, 35). Elucidation of the differences in miR and target mRNA expression between species and tissues will continue to be valuable in understanding the gene expression regulatory networks underlying biological differences between organisms.

Cattle have tremendous importance not only for food production but as a mammalian model organism for comparative genomics and biological studies (16). Despite the recognized importance of miR in regulating gene expression during development and other biological processes, there has been little information about miR expression in cattle. Thus, the main objective in this study was to identify conserved and novel miR present in cattle and to evaluate specific expression patterns in embryo and tissues that are important for immune responses.

MATERIALS AND METHODS

In silico mir predictions. Three data sets were used for mir identification in the bovine genome (Btau_2.0, July 2005). Basic local alignment search tool (BLAST) alignments (1) requiring at least 90% of mir length were generated by comparison to: *1*) a data set obtained from mirBase (version 8.2) (*set 1*; Ref. 18) contained confirmed sequences from different species, including both the mature and stem-loop sequences, *2*) a second data set contained 975 mir predicted from human/mouse phylogeny and secondary structure analyses (10), and *3*) a third data set contained the miRNAMap database (21) with 2,681 human mir predicted using RNAz software (38) that considers both sequence and RNA secondary structure conservation.

miR cloning. Day 30 (d30) bovine embryos (gestation period 280 days) were snap-frozen in liquid nitrogen immediately after removal from the reproductive tract of slaughtered cows. Immune and gut tissue samples were obtained from 8-mo-old Holstein steers raised in

* L. L. Coutinho and L. K. Matukumalli contributed equally to this work.

Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org).

Address for reprint requests and other correspondence: T. S. Sonstegard, USDA, ARS Bovine Functional Genomics Laboratory, Bldg. 200 Rm. 2A, BARC-East, Beltsville, MD 20705 (e-mail: tads@anri.barc.usda.gov).

concrete stalls. Approximately 200–300 mg of tissue from mesenteric (MLN) and abomasal (ALN) lymph nodes, thymus (THY), small intestine (SI), and whole embryos (EMB) were processed with TRIzol (Invitrogen, Carlsbad, CA) for RNA extraction according to manufacturer's instruction. Single insert cDNA libraries corresponding to expressed miR were constructed as described by Lu et al. (26) with the following modifications. In brief, RNA at each stage was separated by size on denaturing acrylamide gels, stained with Syber Gold (Molecular Probes, Eugene, OR), and eluted by FlashPAGE electrophoresis (Ambion, Austin, TX). PCR-amplified cDNA was cloned using Topo TA cloning (Invitrogen). Individual clones from the transformation were transferred into 384-well plates and sequenced with DYEnamic ET terminator (Amersham, Piscataway, NJ) on an ABI 3730 instrument (Applied Biosystems, Foster City, CA). All animal care and protocols were reviewed and approved by the Beltsville Agricultural Research Center's Animal Care and Use Committee (protocol number 05-013).

Sequence, quantitative RT-PCR and statistical analyses. Chromatograms were analyzed with Phred (14), and resulting sequences were screened for vector and linker sequences. Sequences were oriented based on the 3′- and 5′-end specific linker sequences and were used only if the inserts were 17–34 bases with a minimum base quality of 20 for all bases. Distinct clone sequences were clustered and assembled to obtain the longest sequence, along with member counts within the cluster. Some of the clusters were further collapsed by manual intervention that allowed for single base mismatches, especially toward the ends of a cloned sequence. The reduced set of sequences from the above analysis was annotated by matching against miRBase for known miR. Sequences not having matches to miRBase were screened against rRNA, tRNA (http://lowelab.ucsc.edu/GtRNAdb/Hsapi/Hsapi-summary.html) and snoRNAs (http://www.snorna.biotoul.fr/browse.php) to remove contaminating sequences that could interfere with identification of novel bovine miR. Bases flanking the bovine miR were obtained by BLAST analysis (1) against the bovine genome and were used to check for hairpin conformation using mfold (27). Human matches and conservation of bases across different species was determined using UCSC genome browser (http://genome.ucsc.edu). Those sequences identified as mir were submitted to the miRBase registry web site for official annotation. Statistical analysis of miR expression was conducted using χ^2 analysis to compare global and individual miR abundance among the libraries. MicroRNA expression was defined as the number of sequences for each miR in a library, divided by the total number of sequences for that library (Supplemental Table S2). (The online version of this article contain supplemental material.) Quantitative RT-PCR was conducted using human TaqMan miR probes that had the exact same sequence as the bovine miR. Reactions were conducted following manufacturer's recommendations (Applied Biosystems). Hierarchical clustering of miR expression was performed on data from miR sequenced at least four times and with the program GeneSpring version 7.2 (Agilent, Foster City, CA) and Pearson correlation.

RESULTS AND DISCUSSION

Identification of bovine mir by database searches. The bovine genome draft sequence provided a resource to generate in silico mir predictions for comparison to experimental results. Analysis of the current bovine genome draft sequence (August 2006) was performed by comparison with experimentally observed and predicted data sets from other species. Pair-wise comparison of the bovine genome to either hairpin sequences of mirBase (*set 1*; Ref. 18) or two alternative mir prediction data sets (*sets 2* and *3*; Refs. 10 and 21, respectively) resulted in 334, 908 and 2,076 alignments, respectively (Supplemental Table S1). The number of matches to the bovine genome for these data sets can be improved by homology base method that takes into account sequence divergence between species. Combining all data sets resulted in 2,793 nonoverlapping unique putative bovine mir with 131 mir shared by all three data sets. Pairwise comparison yielded 182, 168, and 306 common predictions between *sets 1* and *2*, *1* and *3*, and *2* and *3*, respectively. Prediction *set 3* may have overestimated the number of mir, because it yielded more putative mir but fewer matches to known mir than did *set 2* and contained several putative mir that mapped to the same bovine locus. Considering that mirBase contains validated mir and that the overlaps for the bovine genome matches between *sets 2* and *3* to mirBase were only ~50%, comparison against mirBase appears to be the most conservative way to identify mir in an unannotated genome. A count of matches between predicted bovine mir and mir from other species was also determined (Supplemental Table S2). As in other species (8, 34), some mir were found clustered in specific genomic regions, suggesting potential coexpression or coregulation. A large cluster with 38 mir was observed from BLAST alignment in Chr21 (59.4- to 59.6-Mb region). A comprehensive list of clusters is presented in Supplemental Table S2. Full-scale genomic inferences about mir clusters in cattle is still incomplete, as the current genome sequence assembly has assigned 80% of the scaffolds to chromosomes, and some of these have not been oriented. Further comparison of mir conserved across different species will also be useful in understanding common regulation of gene expression among species.

Identification of expressed bovine miR. To validate predicted mir and begin characterizing the miR portion of the transcriptome, five cDNA libraries were constructed from size-fractioned bovine RNA (15–30 bases). A library from early embryo (d30) RNA was constructed to capture the presumed diversity of miR expression during somite differentiation, and four libraries were constructed from tissues of the immune-gut axis. These latter libraries represent tissues important to current animal health and food safety studies.

A total of 3,209 clones were processed to yield 2,617 sequences. These sequences collapsed into 412 clusters that represented potential unique small RNAs and yielded an overall novelty index of 15.7% (Table 1). The 412 unique sequences were evaluated by several criteria (described below) to better determine if these sequences actually represented expression of bovine miR.

Comparative analyses of bovine miR. BLAST analysis of the 412 small RNA-derived sequences against mirBase (18) identified 100 sequence clusters with high or perfect homology to

Table 1. *Description of the small RNA libraries*

Library	Quality Sequence	Unique Sequence	Novelty (%)	Sequence of Known miR	Known miR (%)	Unique Known miR
THY	579	100	17.3	525	90.7	54
MLN	542	91	16.8	495	91.3	48
ALN	540	94	17.4	492	91.1	47
SI	559	187	33.5	399	71.4	64
EMB	397	131	33.0	317	79.8	54
Overall	2,617	412	15.7	2,228	85.1	100

Tissues used for library construction: THY, thymus; MLN, mesenteric lymph node, ALN, abomasum lymph node; SI, small intestine; and EMB, embryo. Novelty was calculated as a percentage of unique sequences from the valid sequences obtained. Known microRNAs were determined by comparison against the miRBase data set (15). miR, mature microRNA.

miR from other species (Table 1), indicating that library construction had successfully cloned mature miR. A perfect alignment with human miR was observed for 96 of the 100 sequences, while four had one or two bases altered in the center of the sequence for the mature transcript (bta-miR-20b, bta-miR-34b, bta-miR-140 and bta-miR-380-3p). Positions in the bovine genome assembly were determined for 98 of the 100 sequences (Supplemental Table S2). The genomic DNA sequences flanking the miR were obtained from the bovine genome assembly and aligned against human mir for further validation. All bovine miR with flanking sequence were found to have a potential fold-back precursor structure typical of known mir transcripts. These 100 sequences representing conserved miR accounted for 24% of the sequence diversity and 85% (2,228 sequences) of the total count of sequences from the libraries (Table 1). This is the first report of successful cloning of bovine miR.

The remaining 312 sequence clusters were aligned against the bovine genome to detect potential cloning artifacts. A total of 177 sequence clusters (sequence count of 190) did not match, suggesting only a small percentage of the sequences from the libraries were potential artifacts of small RNA cloning (7%). Alternatively, some of these sequences may not yet be represented in the current draft genome assembly. For the 135 sequence clusters that did match (sequence count of 199), 40 matched tRNA and snoRNA. These sequences along with those not matching the genome sequence were discarded from subsequent analyses.

Secondary structures that incorporated flanking genomic sequence were generated from each of the remaining 95 potential miR sequences. These folded structures were examined for stem-loop motifs of an mir transcript capable of producing an miR (some examples in Fig. 1). A total of 28 sequences clusters could be identified as potentially novel miR (Table 2). The miR-derived sequences that generated *structure 7* (Table 2) probably represents a true miR, because this sequence was observed in more than one tissue (Supplemental Table S2). The other 27 miR sequences were only observed in one tissue. These sequences were further characterized by analysis of miR sequence conservation among other species. Such a comparison would be expected to yield several types of results. Those miR sequences not residing within a region of conserved genome sequence would be suggestive of mir loci unique to cattle. This was the case for 12 of the potential 28 unique miR sequences (Table 2). In contrast, the entire stem-loop sequence for the other 16 putative mir were highly similar to sequences found in other animal genomes supporting classification as probable bovine orthologs to mir not yet identified (Table 2). This class of sequences may also represent potential artifacts arising from cloning degradation products of longer cellular RNAs.

Example alignments for two of the 17 miR whose mir sequence aligned to the human genome are provided in Supplemental Figs. S1 and S2. One of these examples, the miR contained within *structure 7* was near the previously described hsa-miR-188 located in an intergenic region of human chromosome X with relatively high sequence conservation between the human, chimp, dog, mouse, and rat genomes. Since submission of this manuscript, this miR has been identified in humans (hsa-miR-532)(33), thus confirming our prediction. This miR and the miR representing *structures 3* and *8* were also predicted from the in silico comparative sequence analysis against data *sets 2* and *3* described above. A second novel miR, *structure 9* (Table 2), is located in a region on bovine chromosome 19 sharing conserved synteny with human chromosome 17 (Supplemental Fig. S2). This miR resides in the third intron of *MAP2K4*, a genomic region that is well conserved across the human, chimp, dog, rat, and bovine genomes. Furthermore, an miR with 100% identity to the one in *structure 9* was observed in a library produced from a primary culture of porcine myoblasts (R. Wiedmann, M. Doumit, L. Matukumalli, T. Sonstegard, L. Coutinho, D. Nonneman, and T. Smith, unpublished data). The *MAP2K4* region of the porcine genome has not been sequenced to permit evaluation of the entire mir sequence, but detection of the miR in porcine cells was sufficient to confirm its identity as a novel miR expressed in two species. The results of comparative analysis for *structures 7* and *9* underscore the utility of the direct sequencing approach of miR detection in model organism tissues.

Seven of the 16 novel bovine miR that aligned to regions of conserved genome actually matched stem-loop sequences of previously identified human mir but did not match the corresponding mature miR associated with that sequence (Fig. 2). Our sequences were located on either the complementary 5′- or the 3′-strand of the stem. An analysis of the abundance of these miR across libraries (Fig. 2) revealed that in some cases only the complementary strand to the known human miR was cloned (bta-miR-455-3p, bta-miR-545-5p, and bta-miR-22-5p), while in other cases both strands were observed (bta-miR-21-5p, bta-miR-21-3p, bta-miR-425-5p, bta-miR-425-3p, bta-miR-127-5p, bta-miR-127-3p bta-miR-193-5p, and bta-miR-193-3p). Similar observations were made by Suh et al. (36) for miR expressed in human embryonic stem cells. The expression of the complementary miR sequences for miR-455, miR-127, miR-193, and miR-22 was intriguing, because the mature and stem-loop sequences were perfect matches to the corresponding human mir and the mature miR observed in cattle have yet to be detected from human samples. However, since the submission of this manuscript, miR-455-3p was cloned in mouse (28) and miR-425-5p was cloned in human (2). To our knowledge there are no studies explaining why different stems of the loop might be selected in different species, but it may be reasonable to predict that the resulting miR will not have the same range of target genes.

The situation was more complicated for miR-21, miR-127, and miR-193, for which detection of both strands depended upon tissue sampled. In the case of bta-miR-21, both strands of the mir were observed as mature miR, whereas in human only the hsa-miR-21-5p has been reported. While bta-miR-21-5p was observed 82 times among the five libraries, bta-miR-21-3p was observed only once, in the abomasal lymph node library (Fig. 2). Analysis of 5′-end hairpin stability by the nearest-neighbor method (42) for bta-miR-21 indicates that 5′-end of bta-miR-21-3p is slightly less stable (−8.1 kcal/mol for the bta-miR-21-5p vs. −7.3 kcal/mol for the bta-miR-21-3p). Studies conducted by Khvorova et al. (22) concluded that the strand with the less stable 5′-end will became the mature miR, so one would expect the mir-21-3p strand to be the mature miR. In human and bovine tissues, mir-21 does not follow this rule. This however is not novel, as Suh et al. (36) reported that only 69% of the novel human miR identified in their study followed this rule. For mir-127, both strands were observed as mature forms in embryo. In the SI, only bta-miR-127-3p was

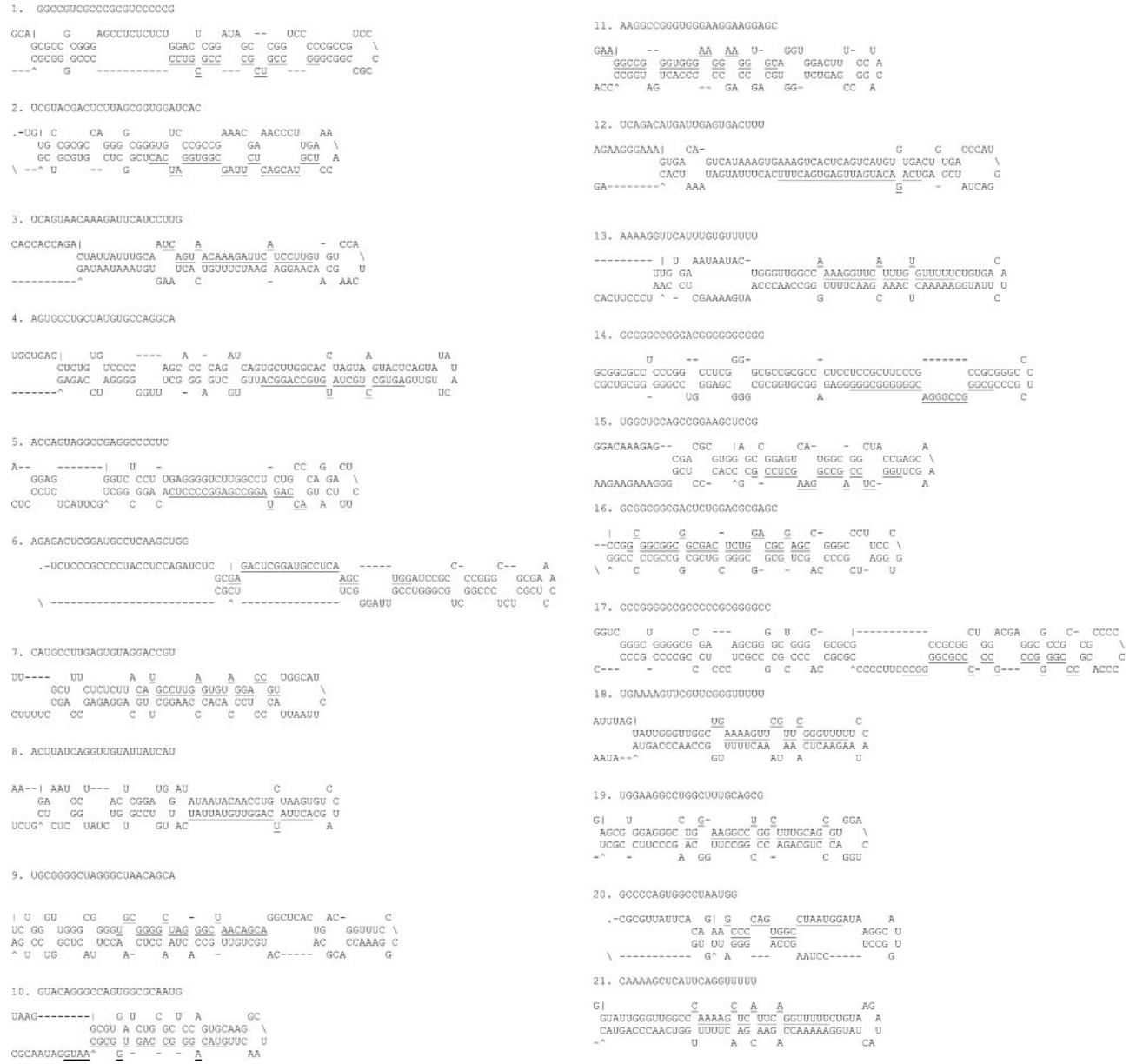

Fig. 1. Secondary structure of the bovine mir presented in Table 2 as predicted by the program mFold. The miR sequence is underlined. mir, microRNA; miR, mature microRNA.

observed. The presence of both forms in SI could not be ruled out, because the level of expression sampled was not sufficient for complete detection of all expressed miR. Similarly, bta-miR-193-3p was expressed in the THY and SI, and bta-miR-193-5p was expressed in the lymph node. These observations expand the results obtained by Thomson et al. (37), in which one strand of miR-30a was detected in adult mouse tissue while the other was detected in embryo. The differential expression of complementary strand miR in cattle is further evidence that the "rules" for mir processing (22) are not fully understood.

Overall expression profile of bovine miR. A total of 308 small RNAs were observed only once, while 107 were present multiple times. A biologically significant level of miR expression was not determined, but the sequence identity count data suggested that the depth of sequencing was sufficient to support estimation of expression levels for the more highly expressed miR within a specific tissue. Deeper sequencing from each sample would be necessary to fully characterize less prevalent miR based on a recent report that indicated identification of all expressed miR was saturated at ~40,000 sequences (13). Because miR modulate expression by binding to target mRNA and appear to act stoichiometrically rather than catalytically, it might be argued that the subtle effects caused by miR expressed with few molecules per cell would be difficult to elucidate with current expression profiling platforms. In any case, the data obtained from these bovine libraries provide an initial survey for comparison of the most abundant miR transcripts in the five tissues sampled and provide the first analysis of miR expression in bovids.

The diversity of observed miR sequences varied between different tissues. The EMB and SI libraries had the highest number of unique sequences (131 and 187 respectively, Table

Table 2. *New miR cloned from bovine tissues*

miR	BTA[1]	Start[2]	End[3]	HSA[4]	dG[5]	Conservation[6]	Structure[7]
GGCCGTCGCCCGCGTCCCCCG	ChrUn.003.3914	16036	16016	12	−47.4	H, C, D, M, R, Ch, Z	1
TCGTACGACTCTTAGCGGTGGATCAC	ChrUn.003.19	662725	662740	Y	−24.4	H, C, D, M, R, Ch, Z	2
TCAGTAACAAAGATTCATCCTTG	ChrUn.003.1645	10776	10798	21	−25.1	H, C, D, M, R	3
AGTGCCTGCTATGTGCCAGGCA	ChrUn.003.2276	12006	11985	5	−28.4	H, C, D, M, R	4
ACCAGTAGGCCGAGGCCCCTC	21	59455854	59455874	14	−31.5	H, C, D, M, R	5
AGAGACTCGGATGCCTCAAGCTGG	4	95921071	95921093		−27.3		6
CATGCCTTGAGTGTAGGACCGT	X	60739762	60739783	X	−19.0	H, C, D, M, R	7
ACTTATCAGGTTGTATTATCAT	X	54991862	54991841	X	−26	+, H, C, D, M, R	8
TGCGGGGCTAGGGCTAACAGCA	19	29850200	29850221	17	−22.9	H, C, D, M	9
GTACAGGGCCAGTGGCGCAATG	22	52050469	52050490	6	−18.8	H, C	10
AAGGCCGGGTGGGAAGGAAGGAGC	ChrUn.003.3153	2279	2256		−22.9		11
TCAGACATGATTGAGTGACTTT	ChrUn.003.5316	5301	5280		−36.3		12
AAAAGGTTCATTTGTGTTTTT	ChrUn.003.558	189490	189510		−26.1		13
GCGGGCCGGGACGGGGGGCGGG	ChrUn.003.3914	15479	15458		−65.8		14
TGGCTCCAGCCGGAAGCTCCG	13	66577414	66577394		−31.3		15
GCGGCGGCGACTCTGGACGCGAGC	26	1342399	1342377		−30.2		16
CCCGGGGCCGCCCCCGCGGGGCC	1	3493549	3493528		−41		17
TGAAAAGTTCGTTCGGGTTTT	25	32480995	32481016		−24.4		18
TGGAAGGCCTGGCTTTGCAGCG	ChrUn.003.1018	88333	88316		−32.1		19
GCCCCAGTGGCCTAATGG	19	50536929	50536909		−19		20
CAAAAGCTCATTCAGGTTTTT	X	55063231	55063211		−27.4		21
TCAGTAAATGTTTATTGGATG	19	16583101	16583080	14	*		
TGGGTCTTTGCGGGCGAGATGA	19	21406602	21406581	17	*		
AGTTCTTCAGTGGCAAGCTTTA	8	85357789	85357810	17	*		
GCAGTCCATGGGCATATACACT	19	10141770	10141749	9	*		
AACAGCAGTCGATGGGCTGTCT	22	49016911	49016932	17	*		
ATGACACGATCACTCCCGTTGA	21	59462128	59462150	3	*		
CTGAAGCTCAGAGGGCTCTGATT	ChrUn.003.3914	16036	16016	14	*		

[1]Chromosome assignments on *Bos taurus* genome build 3.0. [2]Position on BTA where microRNA sequence starts. [3]Position on BTA where microRNA sequence ends. [4]Chromosome assignment on human genome. [5]Free energy of the microRNA structure as calculated by mFold. [6]Conservation of the microRNA in H (human), C (chimp), D (dog), M (mouse), R (rat), Ch (chicken), and Z (zebrafish) genomes. [7]The folding structure of the microRNA is represented in Fig. 1. *These bovine miR are located in the hairpin sequence of human mir and are shown in Fig. 2. +This bovine miR aligned with two positions on human chromosome X. In one position it aligned with hsa-mir-374, and in the position reported in the table it aligned to a conserved region with other species, but this position does not have a human microRNA reported.

1), and the highest novelty rate (33.0% and 33.5% respectively, Table 1). In addition, the 10 most observed miR sequences in EMB and SI represented a smaller proportion of the total sequences collected (53% and 46% respectively, Supplemental Table S2) than the top 10 observed in ALN and MLN samples (75% and 85% respectively, Supplemental Table S2). THY had a novelty rate similar to ALN and MLN (17%, Table 1) and a top 10 proportion intermediate between the ALN/MLN and EMB/SI groups (67%, Supplemental Table S2). These results were consistent with the idea that embryos have a high relative diversity of miR expression, which might be explained in part by the varied emerging tissue types within a whole embryo. However, our results also suggested that the SI expresses a diverse repertoire of miR involved in posttranscriptional regulation. The latter observation was supported by a recent study of a colorectal sample (13) that indicated a high diversity of miR expression in alimentary tissue, perhaps indicative of multiple and divergent tissue types contained in the sample similar to those in a whole embryo sample.

To further characterize the abundance of some of the cloned bovine miR across different tissues, quantitative RT-PCR was employed (Table 3). The U6B small nuclear RNA (RNU6B) was used as an internal control. The constant expression level of RNU6B across all tissues examined indicated an equivalent loading of total RNA for all reactions. Thus, the calculated threshold values (C_t) for all miR are presented in Table 3 without correction for the internal control RNU6B. Efficiency of amplification was determined for all miR from serial dilutions of the specific miR cDNA, and values ranged from 1.79 to 1.94. The data obtained by quantitative RT-PCR had correlations between normalized miR counts (normalized across libraries to an equivalent number of clones sequenced per library) and C_t value that ranged from −0.59 to −0.88 (a negative correlation was expected, since the greater the abundance of the miR, the lower the C_t value). Despite the good agreement between quantitative RT-PCR results and normalized library counts, one must consider that the quantitative RT-PCR results reflect the abundance of each miR in relation to the total RNA for each tissue, while the normalized counts reflect the relative abundance of miR identified in each library. As a consequence, comparison of library counts and quantitative RT-PCR results are qualitative and not quantitative in nature.

The most abundant miR across the five tissues was miR-26a (Table 3 and Fig. 3), accounting for 13% (EMB), 12% (SI), 30% (THY), 34% (MLN), and 39% (ALN) of sequences (obtained by dividing values in Supplemental Table S2 by total sequences per library from Table 1). The human miR-26a ortholog has previously been detected in multiple tissues (29, 35) including THY, where it was expressed in naïve T cells and has been implicated in development of immune cells. However, previous studies using microarray or Northern blot analysis of miR expression in human tissues indicated that while miR-26a was present at intermediate to high levels in the THY (6, 8, 35), it was only present at a level defined as "average" in lymph node or "low" in the SI (8). In addition, a direct cloning approach similar to the one used for our study found miR-26a

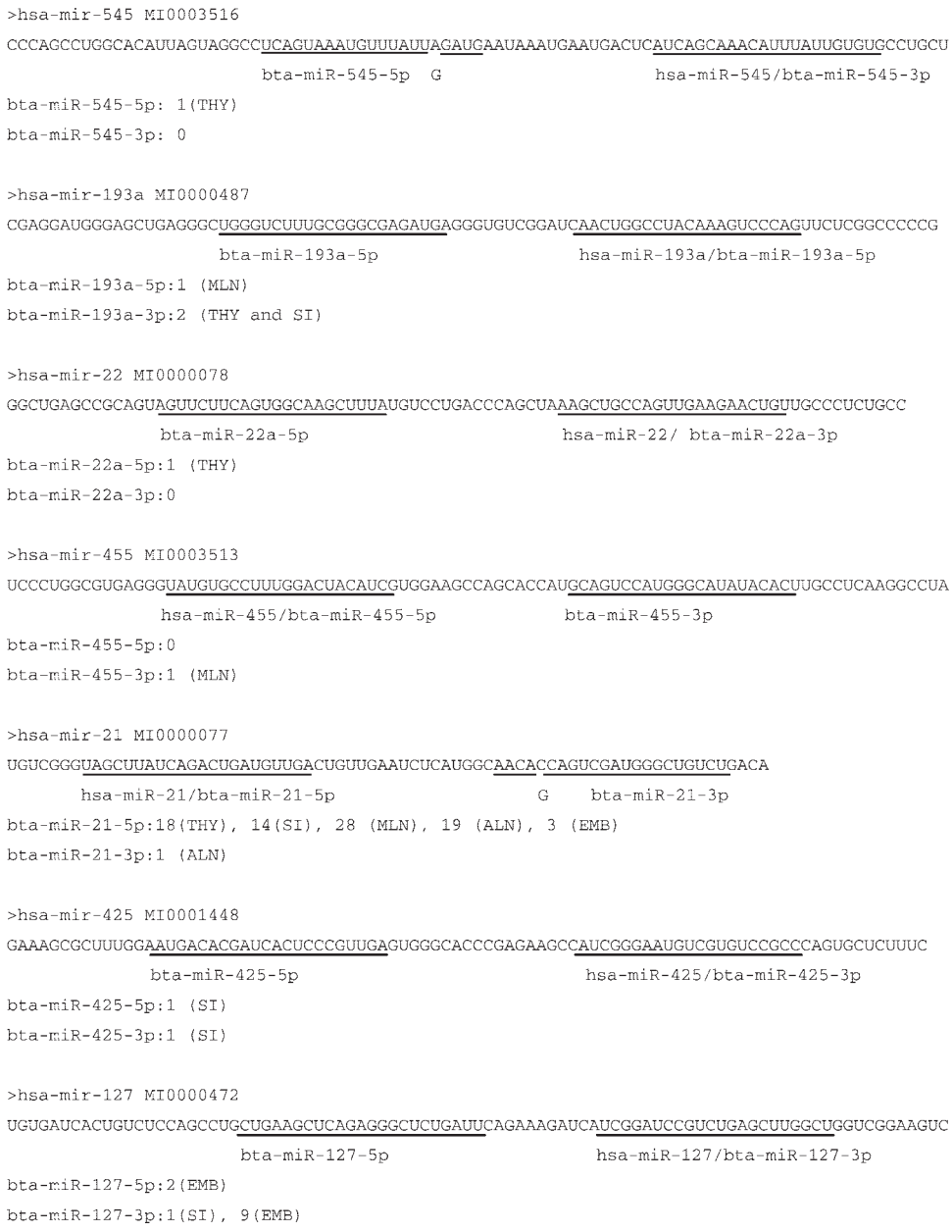

Fig. 2. Alignment of novel bovine miR to human mir. The stem-loop sequences of the human mir are shown, and the mature sequences of bovine and human miR are underlined. The tissue and the number of times each of the miR were sequenced are also indicated. On bta-miR-545-5p, the letter G indicates the difference in sequence observed in the bovine miR.

present at ~3% of unique tags in colorectal tissue (13). Together, these data suggest that significant differences exist in expression levels of miR between similar tissues of different species.

We attempted to determine whether this difference in miR-26a expression was caused by differences in genome copy number. As in humans, there were two loci that align with miR-26a in the cattle genome. The positions in cattle were on

Table 3. *Calculated threshold (Ct ± standard deviation) of selected bovine miRs determine by quantitative RT-PCR*

Tissues	*n*	RNU6B	bta-miR-26a	bta-miR-103	bta-miR-29a	bta-miR-125b	bta-miR-150	bta-miR-122a
THY	3	28.0±0.3	19.8±0.2	23.6±0.2	23.1±0.4	22.4±0.4	22.1±0.9	33.6±2.0
MLN	3	28.2±0.1	20.7±0.7	24.0±1.3	22.0±1.1	25.0±0.4	19.9±0.9	36.7±0.8
ALN	2	28.2±0.4	21.0±0.7	23.9±0.1	21.7±0.3	24.9±0.3	19.9±0.3	>40
SI	2	28.0±0.2	23.0±3.4	24.2	22.1±0.2	22.8±0.0	23.5±0.2	>40
EMB	1	28.6	21.3	23.7	29.0	23.0	28.4	26.8
Amplification		n/a	1.88	1.85	1.94	1.79	1.88	1.87
Correlation		n/a	−0.59	−0.66	−0.85	−0.63	−0.73	−0.88

Values are cycle threshold (Ct) ± SD. Tissues used for library construction were THY, MLN, ALN, SI, and EMB. Amplification refers to PCR amplification efficiency; correlation is correlation of Ct; *n* is number of animals used in the assay.

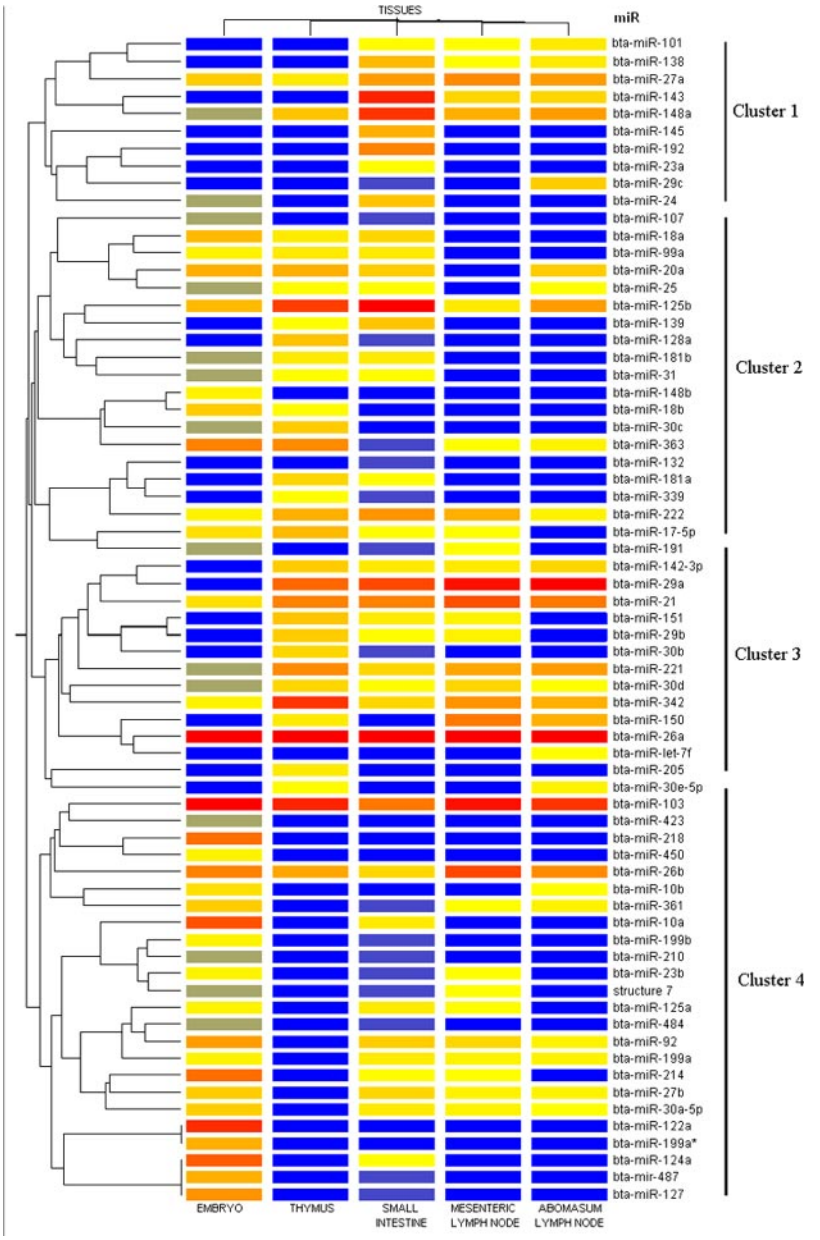

Fig. 3. Hierarchical clustering of tissues and miR using Pearson correlation. Blue indicates low expression and red high expression. EMB, embryo; THY, thymus; SI, small intestine; MLN, mesenteric lymph node; and ALN, abomasum lymph node.

chromosomes BTA 5 and 22 (orthologous to HSA 12 and 3, respectively). In both genomes miR-26a-1 was located in an intron of carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase-like (*CTDSPL*, HSA 3), and miR-26a-2 in an intron of carboxy-terminal domain RNA polymerase II polypeptide A small phosphatase 2 (*CTDSP2*, HSA 12). In the human, a strong correlation between the expression of hsa-miR-26a and the mRNA expression of *CTDSP2* was found, indicating that intronic miR were coexpressed with the respective gene (8). Coexpression appears to be a general feature of mammalian miR genes that lie in introns of protein-coding genes, as observed for numerous intronic miR of human genes (8). Human *CTDSP2* was expressed at a higher level than *CTDSPL*, suggesting that most of hsa-miR-26a comes from the mir located on human chromosome 12, and in fact the expression of hsa-miR-26a was not correlated with expression of *CTDSPL* (8). This did not appear to be the case in cattle, especially for the SI. Analysis of SI mRNA expression in cattle by microarray revealed that *CTDSP2* and *CTDSPL* have approximately the same expression level (data not shown), suggesting that in cattle both genes may contribute to miR-26a synthesis. This would provide a possible explanation for the higher level of this miR in bovine tissues.

The only other highly expressed miR (Table 3 and Fig. 3) present in all tissues was miR-103. Human microarray and Northern blot data support widespread expression of miR-103 in adult tissues, but data from human lymph node, THY, and SI did not indicate that miR-103 was highly expressed (6, 8, 35). The contrast between the bovine data for miR-103 and human data further supports the notion of species-specific variability in level of expression of miR.

There were 17 miR expressed in all five libraries, disregarding overall expression level (including miR-26a and miR-103 discussed above), and these represented 14% of the combined

set of 100 miR matching mirBase plus 28 putative new miR (listed in Supplemental Table S2). The remaining miR showed various patterns of expression (Supplemental Table S2). Of the 128 miR, 50 were represented by a single sequence from one library, making it difficult to assign a particular miR as tissue specific or enriched. All the libraries had approximately the same frequency of these singletons.

Comparison of expression profiles between tissues. Tissue clustering based on miR expression showed that different tissues have diverse expression profiles, while similar tissues such as abomasum and mesenteric lymph nodes have very similar patterns of miR expression (Fig. 3). Based on function, this result was somewhat expected, even though both types of lymph nodes reside in different positions along the gut axis. SI and THY were less similar to the lymph nodes, whereas embryonic tissues had a more distinct profile of miR expression. The clustering of THY and SI proximal to lymph nodes could reflect the presence of developing T-cells in the THY and the infiltration of immune cells into the SI. The tissue with the greatest diversity in miR expression was SI (74 distinct miR, Supplemental Table S2). Again, this observation could reflect cell type diversity of this tissue and the rapid turnover of intestinal epithelial cells. Our tissue clustering was consistent with results of a previous study based on the expression profile of several miR from 26 human tissues (35).

Clustering of the miR by expression profile resulted in four major groups (Fig. 3). miR that were present in *cluster 1* were preferentially expressed in SI and lymph nodes; *cluster 2* were preferentially expressed in embryo, THY, and SI with lower expression in lymph nodes; *cluster 3* were expressed in most tissues; and *cluster 4* were preferentially expressed in embryo.

Examination of *cluster 4* (Fig. 3 and Supplemental Table S2) uncovered two apparently embryo-specific miR, miR-122a and miR-199a*. The largest disparity was found for miR-122a, which occurred with 8% frequency in the EMB library sequences but was not observed in any other libraries. Quantitative RT-PCR results (Table 3) corroborate this observation, since expression of bta-miR122a was at least 36 times higher in embryo than the other tissues examined. Low expression of miR-122a in adult THY, lymph node, and SI was consistent with previous microarray and Northern analysis of human RNA (8, 35), which showed very low levels of hybridization to miR-122a probes (although high expression was observed in adult liver). Expression in embryo was also consistent with detection in early mouse embryos (11).

The other EMB-specific miR (excluding singletons) was miR-199a*, present at 2% frequency in EMB and not observed (i.e., present at $<0.2\%$ frequency) in any of the other libraries. Detection of miR-199* in d30 cattle embryo was consistent with expression observed in zebrafish embryos (39). However, microarray and Northern blot studies suggested expression of miR-199* at low to moderate levels in human lymph node, THY, and SI (8, 35). The most straightforward interpretation of our data was that miR-199a* does not have the same expression pattern in cattle as in human THY, lymph node, and SI. We also observed six miR sequences that were not EMB specific but had higher expression in the embryo library, with relative expression ratios to the next highest expressing tissue of 7.7 (miR-10a), 9.4 (miR-124a), 14 (miR-127), 8.4 (miR-214), 21 (miR-218), and 9.0 (miR-487).

There were 17 nonsingleton miR for which expression was detected in two or more nonembryonic samples. The most dramatic example of this was bta-miR-29a, which was in high relative abundance (4–12%) in the adult tissue libraries, but not observed in EMB. Low level bta-miR-29a expression was confirmed by RT-PCR results, where a 31-fold difference in expression was observed for this miR between embryonic and THY tissues. This was consistent with microarray results found for human samples that showed widespread, relatively high expression of miR-29a (6, 8). The data do not rule out a role for miR-29a in embryogenesis, because the EMB library only represents a narrow snapshot of development shortly after completion of somitogenesis. There were two other noteworthy instances of between-tissue variation in miR expression, involving miR-145 and miR-150. In our study, bta-miR-145 was found only in SI. This is consistent with previous expression studies in mice indicating miR-145 has a higher expression in SI than THY or lymph nodes (35). However, a previous study of zebrafish embryos (39) indicated expression of miR-145, which was not observed in our EMB library. This could mean that expression of miR-145 does not occur in d30 embryos or that the level of expression is $<0.3\%$, the level at which we would expect to see at least one clone in the number of sequences obtained. Similarly, bta-miR-150 had higher expression in tissues implicated in immune response (THY and lymph nodes), which agrees with results in mice where miR-150 is highly expressed in lymph node and THY (35). This miR is involved in maturation and differentiation of T and B cells by being up regulated in T and B cells and repressed in Th1 and Th2 cells (29). Our quantitative RT-PCR results are in general agreement with library counts (61-fold higher expression in lymph nodes than in the embryo), but RT-PCR also revealed expression of bta-miR-150 in SI.

The construction of size-fractioned RNA libraries from bovine tissues allowed discovery and expression profile characterization of over 100 bovine miR. This represents about one-third of the 334 mir predicted with the miRBase data set, which was somewhat surprising, considering that some miR are cell type specific and only a few bovine tissues were sampled. The identification of 28 potential new miR indicates the importance of animal model systems, since our study resulted in the identification of miR not previously identified in the human genome. However, it was evident that a thorough identification of bovine miR will require a broader sampling of tissues and more in depth sequencing. Certainly, the simple sampling approaches used in this study must be complemented in future studies aimed at determining miR function in bovids, thereby establishing which bovine mir loci are functional as well as structural orthologs of their counterparts in other mammalian species.

ACKNOWLEDGMENTS

We thank Kevin Tennill, Steve Simcox, Bob Lee, D. Larry Wood, and Renee Godtel for technical assistance. L. L. Coutinho is a recipient of a scholarship from the Brazilian Government (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Fundação de Amparo a Pesquisa do Estado de São Paulo, and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). The Beltsville Area Research Center Animal Care and Use Committee approved protocols using experimental animals.

DISCLOSURES

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

REFERENCES

1. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** Basic local alignment search tool. *J Mol Biol* 215: 403–410, 1990.
2. **Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, Brownstein MJ, Tuschl T, Margalit H.** Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res* 33: 2697–2706, 2005.
3. **Alvarez-Garcia I, Miska EA.** MicroRNA functions in animal development and human disease. *Development* 132: 4653–4662, 2005.
4. **Ambros V.** microRNAs: tiny regulators with great potential. *Cell* 107: 823–826, 2001.
5. **Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, Matzke M, Ruvkun G, Tuschl T.** A uniform system for microRNA annotation. *RNA* 9: 277–279, 2003.
6. **Barad O, Meiri E, Avniel A, Aharonov R, Barzilai A, Bentwich I, Einav U, Gilad S, Hurban P, Karov Y, Lobenhofer EK, Sharon E, Shiboleth YM, Shtutman M, Bentwich Z, Einat P.** MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res* 14: 2486–2494, 2004.
7. **Bartel DP.** MicroRNAs: genomics, biogenesis, mechanism, function. *Cell* 116: 281–297, 2004.
8. **Baskerville S, Bartel DP.** Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA* 11: 241–247, 2005.
9. **Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E, Sharon E, Spector Y, Bentwich Z.** Identification of hundreds of conserved and nonconserved human microRNAs. *Nat Genet* 37: 766–770, 2005.
10. **Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E.** Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* 120: 21–24, 2005.
11. **Chan JA, Krichevsky AM, Kosik KS.** MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res* 65: 6029–6033, 2005.
12. **Chen CZ, Li L, Lodish HF, Bartel DP.** MicroRNAs modulate hematopoietic lineage differentiation. *Science* 303: 83–86, 2004.
13. **Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA Jr, Sjoblom T, Barad O, Bentwich Z, Szafranska AE, Labourier E, Raymond CK, Roberts BS, Juhl H, Kinzler KW, Vogelstein B, Velculescu VE.** The colorectal microRNAome. *Proc Natl Acad Sci USA* 103: 3687–3692, 2006.
14. **Ewing B, Hillier L, Wendl MC, Green P.** Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175–185, 1998.
15. **Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP.** The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science* 310: 1817–1821, 2005.
16. **Gibs R, Weinstock G, Kappes S, Schook L, Skow L, Womack J.** Bovine Genomic Sequencing Initiative. www.genome.gov/Pages/Research/Sequencing/SeqProposals/BovineSEQpdf, 2006.
17. **Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF.** MicroRNAs regulate brain morphogenesis in zebrafish. *Science* 308: 833–838, 2005.
18. **Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ.** miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 34: D140–D144, 2006.
19. **Harfe BD.** MicroRNAs in vertebrate development. *Curr Opin Genet Dev* 15: 410–415, 2005.
20. **He L, Hannon GJ.** MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* 5: 522–531, 2004.
21. **Hsu PW, Huang HD, Hsu SD, Lin LZ, Tsou AP, Tseng CP, Stadler PF, Washietl S, Hofacker IL.** miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes. *Nucleic Acids Res* 34: D135-D139, 2006.
22. **Khvorova A, Reynolds A, Jayasena SD.** Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115: 209–216, 2003.
23. **Lecellier CH, Dunoyer P, Arar K, Lehmann-Che J, Eyquem S, Himber C, Saib A, Voinnet O.** A cellular microRNA mediates antiviral defense in human cells. *Science* 308: 557–560, 2005.
24. **Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM.** Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433: 769–773, 2005.
25. **Liu CG, Calin GA, Meloon B, Gamliel N, Sevignani C, Ferracin M, Dumitru CD, Shimizu M, Zupo S, Dono M, Alder H, Bullrich F, Negrini M, Croce CM.** An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc Natl Acad Sci USA* 101: 9740–9744, 2004.
26. **Lu C, Tej SS, Luo S, Haudenschild CD, Meyers BC, Green PJ.** Elucidation of the small RNA component of the transcriptome. *Science* 309: 1567–1569, 2005.
27. **Mathews DH, Sabina J, Zuker M, Turner DH.** Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288: 911–940, 1999.
28. **Mineno J, Okamoto S, Ando T, Sato M, Chono H, Izu H, Takayama M, Asada K, Mirochnitchenko O, Inouye M, Kato I.** The expression profile of microRNAs in mouse embryos. *Nucleic Acids Res* 34: 1765–1771, 2006.
29. **Monticelli S, Ansel KM, Xiao C, Socci ND, Krichevsky AM, Thai TH, Rajewsky N, Marks DS, Sander C, Rajewsky K, Rao A, Kosik KS.** MicroRNA profiling of the murine hematopoietic system. *Genome Biol* 6: R71, 2005.
30. **Pasquinelli AE, Hunter S, Bracht J.** MicroRNAs: a developing story. *Curr Opin Genet Dev* 15: 200–205, 2005.
31. **Pillai RS.** MicroRNA function: multiple mechanisms for a tiny RNA? *RNA* 11: 1753–1761, 2005.
32. **Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, Macdonald PE, Pfeffer S, Tuschl T, Rajewsky N, Rorsman P, Stoffel M.** A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* 432: 226–230, 2004.
33. **Rassenti L, Lu DP, Read RL, Humphreys DT, Battah FM, Martin DIK, Rasko JEJ.** PCR-based expression analysis and identification of microRNAs. *J RNAi Gene Silencing* 1: 44–49, 2005.
34. **Sewer A, Paul N, Landgraf P, Aravin A, Pfeffer S, Brownstein MJ, Tuschl T, van Nimwegen E, Zavolan M.** Identification of clustered microRNAs using an ab initio prediction method. *BMC Bioinformatics* 6: 267, 2005.
35. **Shingara J, Keiger K, Shelton J, Laosinchai-Wolf W, Powers P, Conrad R, Brown D, Labourier E.** An optimized isolation and labeling platform for accurate microRNA expression profiling. *RNA* 11: 1461–1470, 2005.
36. **Suh MR, Lee Y, Kim JY, Kim SK, Moon SH, Lee JY, Cha KY, Chung HM, Yoon HS, Moon SY, Kim VN, Kim KS.** Human embryonic stem cells express a unique set of microRNAs. *Dev Biol* 270: 488–498, 2004.
37. **Thomson JM, Parker J, Perou CM, Hammond SM.** A custom microarray platform for analysis of microRNA gene expression. *Nat Methods* 1: 47–53, 2004.
38. **Washietl S, Hofacker IL, Stadler PF.** Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci USA* 102: 2454–2459, 2005.
39. **Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, de Bruijn E, Horvitz HR, Kauppinen S, Plasterk RH.** MicroRNA expression in zebrafish embryonic development. *Science* 309: 310–311, 2005.
40. **Wienholds E, Plasterk RH.** MicroRNA function in animal development. *FEBS Lett* 579: 5911–5922, 2005.
41. **Zhao Y, Samal E, Srivastava D.** Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* 436: 214–220, 2005.
42. **Zuker M.** Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406–3415, 2003.