We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Novel microRNA Cloning Using Bioinformatics

Yoshiaki Mizuguchi, Takuya Mishima, Eiji Uchida and Toshihiro Takizawa

Additional information is available at the end of the chapter

1. Introduction

MicroRNAs (miRNAs) participate in several biological processes, including development, differentiation, apoptosis, and proliferation (**1, 2**) through imperfect pairing with target messenger RNAs (mRNAs) of protein-coding genes and transcriptional or post-transcriptional regulation of their expression (**3, 4**). Approaches to miRNA detection, such as parallel sequencing technologies may replace conventional sequencing (**5**). The GS 454 technology can produce a similar number of longer (100–150-nucleotides (nt)) sequence reads in a single analysis run, with the advantage that this method can derive the complete sequence of the mature miRNA. Moreover, recent studies on miRNA profiling performed with cloning techniques suggest that sequencing methods are suitable for the detection of novel miRNAs, modifications, and precise compositions, and that cloning frequencies calculated by clone count analysis strongly correlate with the concentrations measured by Northern blotting, and are reproducible. The achievement of comprehensive profiling of miRNA in human diseases requires exhaustive qualitative and quantitative analyses. Here we show the techniques and the some of the results of the miRNA transcriptomes in the liver using sequencing. This serves as a critical step in clarifying the functional significance of specific miRNAs as they relate to liver diseases.

2. Techiniques of MicroRNA cloning and bioinformatics for MiRNAome

2.1. MicroRNA cloning

The method for microRNA cloning and sequencing that we moderated from the original ones are shown in Fig1. We cloned small RNA by a modification of the published miRNA cloning protocol of Lagos-Quintana et al. (**6**). In brief, total RNA samples were extracted using ISOGEN (Nippon Gene, Tokyo, Japan), separated in a denaturing polyacrylamide gel, and the 18–24 nt fraction was recovered. Next, 5'- and 3'-adapters were ligated to the RNAs Ligation of small RNAs with DNA_RNA chimera linkers at both termini [3' linker

oligonucleotide (5′/5 Phos/rCrUrGrUAGGCACCATCAATdi-deoxyC-3 ′) and the 5′ linker oligonucleotide (5′-ATCGTrArGrGrCrArCrCrUrGrArArA-3′)] and RT-PCR was carried out.
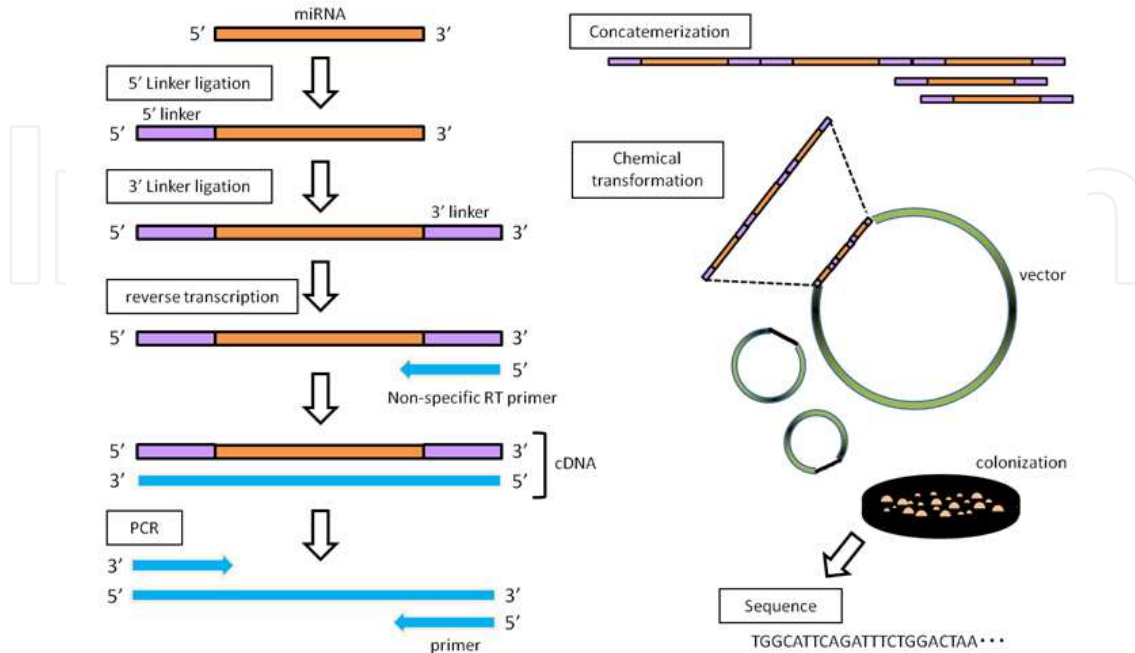

Figure 1. Overview of the miRNA cloning. See paper body for details.

Amplification of the cDNA fragments was obtained by two consecutive rounds of PCR. Specific restriction enzyme digestion of the adaptors allowed for concatemerization of the cDNA into larger fragments. These fragments were then cloned into a vector to create a cDNA library. Concatemerization increases the length of informative sequences obtainable from each clone. we concatenated more than 20 cDNAs into a single fragment using a BanI restriction enzyme (New England Biolabs, Ipswich, MA, USA), a DNA ligation kit ver. 2.1 (Takara Bio, Shiga, Japan), and a Geneclean III kit (Qbiogene, Irvine, CA, USA) prior to TA cloning. The concatenated products were then inserted into plasmids and sequenced (Fig1).

The sequences were compared to human DNA to determine the genomic origin of the small RNA. It was important to avoid contamination from other samples and molecular-weight makers during electrophoresis. Such contaminants considerably diminished the accuracy and efficiency of miRNA cloning. We avoided contamination by performing the cloning procedure separately for each sample, by using a special gel with a small plastic rod that divided the sample and marker lanes, and by using separate vats for each gel for ethidium bromide staining. We made small RNA libraries by excising a portion of a polyacrylamide gel containing species 18–24 nt in length to avoid contaminating our purified RNAs with piRNAs **(7)**.

2.2. Bioinformatics analysis of the sequence data

We performed a homology search for all cloned small RNAs and a secondary structural analysis for all novel miRNA candidates.

2.2.1. Comparing the cloned sequences with those of known RNAs

The vector sequence, the 5'and 3' linkers, and their coupled sequences (CTGTAGGCACCTGAAA) were removed. Those extracted sequences composed of 16–30 nt were defined as valid small RNAs and were subjected to followings. The small RNA sequences were analyzed for homology with known RNAs, including miRNA, piwi-interacting (pi) RNA, rRNA, tRNA, small nuclear (sn) RNA, small nucleolar (sno) RNA, and mRNA, and human genomic DNA sequences. The databases used were: miRNA (mature- and pre-), Sanger Database.; piRNA, the NCBI Entrez Nucleotide database; rRNA, the European ribosomal RNA database; tRNA, the Genomic tRNA database; sn/snoRNA, RNAdb and NONCODE; mRNA, NCBI Reference Sequence; and human genomic sequences, the UCSC Genome Bioinformatics Site. In our searches, we defined the cloned sequencing results that had higher than 90% homology as valid if they met our criteria for sequence error, erroneous PCR amplification, and 3′- and 5′-end variations. Clones with 100% homology with human genomic DNA but did not match known RNAs when compared to the above databases were termed novel miRNA candidates and were subjected to further analysis. The result is shown in table1.

Reads (%)	HCC	ANL	Sum
Total	314,359	268,708	583,067
miRNA	256,64 (81.6)	208,038 (77.4)	464,687 (79.7)
piRNA	2,983 (0.9)	1,440 (0.5)	4,423 (0.8)
rRNA	5,474 (1.7)	10,161 (3.8)	15,635 (2.7)
tRNA	1,703 (0.5)	621 (0.2)	2,324 (0.4)
snRNA	700 (0.2)	343 (0.1)	1,043 (0.2)
snoRNA	654 (0.2)	747 (0.3)	1,401 (0.2)
mRNA	6,053 (1.9)	7,279 (2.7)	13,332 (2.3)
Genome	2,799 (0.9)	3,149 (1.2)	5,948 (1.0)
Others	15,588 (5.0)	34,686 (12.9)	71,174 (12.2)

Table 1. Annotation of the sequenced small RNAs

2.2.2. Secondary structure analysis of Novel microRNAs

The two-dimensional precursor miRNA (pre-miRNA) configurations of our novel miRNA candidates were predicted according to the method described previously **(8)** with some modifications. Briefly, 196-nt of genomic sequence was added to the candidate sequences (88-nt at each end). Each candidate sequence was divided into 110-nt windows and subjected to two-dimensional analysis along its entire length, using the RNAfold software (Vienna RNA Secondary Structure Package **(9)**). The configurations that had the lowest free energy and that had a high conservation (described below) and met the following criteria were termed novel miRNAs: (*a*) contained a stem-loop configuration; (*b*) cloned mature

miRNA sequence portion consisted of more than 16-nt in its double-stranded region; (*c*) the loop contained fewer than 20-nt; (*d*) the internal loop contained fewer than 10-nt; and (*e*) the bulge contained fewer than 5-nt. Furthermore, novel sequences with overlapping positions in the genome were grouped together. Novel antisense miRNAs are defined with above criteria (a)-(e) but without conservation score if they are coded in same chromosomal region.

2.2.3. Determination of hairpin conservation of Novel MicroRNAs

We classified all the candidate miRNAs using the PhastCon database at the University of California at Santa Cruz.**(10, 11)** This database has scores for each nucleotide in the human genome relative to its degree of conservation when compared to nucleotides in the armadillo, bush baby, cat, chicken, chimpanzee, cow, dog, elephant, frog, fugu, guinea pig, hedgehog, horse, lizard, medaka, mouse, opossum, platypus, rabbit, rat, rhesus monkey, shrew, stickleback, tenrec, tetraodon, tree shrew, and zebrafish. The algorithm is based on a phylogenetic hidden Markov model that uses best-in-genome pairwise alignment for each species (based on BLASTZ), followed by multiple alignment of the twenty eight genomes. A hairpin was defined as conserved if the average PhastCon conservation score over the 28 species for any 15-nt sequence in the hairpin stem was at least 0.8 .**(12)**

3. Other techniques for MicroRNAome

3.1. Real-time PCR analysis of known miRNAs

Real-time PCR was performed on an ABI7300 (Applied Biosystems, Foster City, CA, USA) using various mirVana qRTPCR primer sets (Ambion, Austin, TX, USA) and a SYBR ExScript RT-PCR kit (Takara Bio), or with TaqMan miRNAs assays (Applied Biosystems), a High capacity cDNA archive kit (Applied Biosystems), and Absolute QPCR ROX mix (Abgene, Rochester, NY, USA), according to the manufacturers' instructions. As an endogenous control, 5SrRNA or U6 snRNA was used.

3.2. Ago2-immunoprecipitation and PCR analysis of novel miRNAs

After bioinformatic analysis of the sequence data, we further validated novel miRNAs by using a combination of Ago2-immunoprecipitation **(13)** followed by PCR-based miRNA detection **(14)**. Briefly, 50 ml Dynabeads protein G slurry (Invitrogen) was immobilized with

20 mg human P4 anti-mouse Ago2 monoclonal antibody (clone 2D4, Wako Pure Chemical Industries, Osaka, Japan). One hundred fifty micrograms of human tissue P4 were homogenized in 1.5 ml of a cell lysis solution (provided in miRNAs isolation kit, Wako) using a Polytron PT1200C homogenizer (Kinematica AG, Lucerne, Switzerland) for 10 s at 4 8C, and then 1.5 ml of the cell lysis solution was added into the homogenized solution. Following incubation for 15 min on ice, lysate was centrifuged at 20 000 g for 20 min at 4 8C and filtered through a 0.8 mmSupor Acrodisc syringe filter (Pall Corporation, Ann Arbor, MI, USA). One milliliter of the filtered lysate was incubated with 25 ml of the anti-Ago2-Dynabead protein G for incubation for 60 min at 4 8C. After immunoprecipitation, Ago2-

associated RNAs were isolated from the immunoprecipitate according to the manufacture's protocol (Wako). We confirmed that the immunoprecipitate contained human P5 Ago2 protein of w100 kDa in size by western blot (data not shown). Non-immune human IgG (Sigma) was used as a control for Ago2-immunoprecipitation. Preparation of the cDNA library using the Ago2-associated RNAs and semi-quantitative PCR analysis of the above-mentioned novel miRNA candidates were performed, as reported previously **(14)**. A small RNA-specific primer and a universal reverse primer RTQ-UNIr **(14)**, were used for amplification of each of the small RNAs. The PCR products were analyzed on a 12% polyacrylamide gel. The primers for the human GAPDH were used for negative control.

3.3. PCR analysis of novel miRNAs (Alternative method of 3.2)

After bioinformatic analysis of the sequence data, we further validated novel miRNAs by PCRbased miRNA detection **(14).** Briefly, small RNAs were isolated using the mirVana™ miRNA isolation kit (Ambion). Small RNA samples were polyadenylated with Poly(A) Tailing Kit (Ambion) and were purified with Acid-Phenol:Chloroform and with filter cartridge provided in the mirVana Probe & Marker Kit (Ambion). To generate a small RNA cDNA library, tailed RNA were reverse transcribed using RTQ primer**(14)** and the samples were purified using the QIAquick

spin PCR purification kit (QIAGEN). A small RNA-specific primer and a universal reverse primer RTQ-UNIr **(14)**, were used for amplification of each of the small RNAs. The PCR products were analyzed on a 12% polyacrylamide gel. The primers for the human GAPDH were used for negative control.

3.4. Real-time PCR-based miRNA expression profiling

Total miRNA (350 ng) was reverse-transcribed using Megaplex RT Primers (Applied Biosystems). The resulting cDNAs were pre-amplified using Megaplex PreAmp Primers (Applied Biosystems) and the pre-amplified products applied to a TaqMan Human MicroRNA Array Panel (A and B, v2.0).

## 3.5. siRNA, Pre-miR and anti-miR transfections

Cultured cells were transfected with precursor hsa-miR-200c and hsa-miR-141 (ID: PM11714; PM10860); Anti-miR™ 200c and 141 inhibitors (ID: MH11714; MH10860) (Ambion, Austin, TX) for 8 hours in serum free medium. Serum supplemented medium was added and gene and protein expression measured at the indicated time points.

4. Study designs

Sequencing using 454 sequencing and conventional cloning from 22 pair of HCC and adjacent normal liver (ANL) and 3 HCC cell lines identified reliable reads of more than 300000 miRNAs from HCC and more than 270000 from ANL for registered human miRNAs.

5. Detected novel microRNAs using cloning

Eleven novel opposite miRNAs (defined as miRNAs cloned from the other arm of precursors from which known miRNAs have been cloned) were identified from the annotated miRNAs (Figure 2A).Based on above criteria, a scan of all the novel miRNA candidates identified 245 novel precursors, representing putative 210 novel mature miRNAs (Figure 2B).
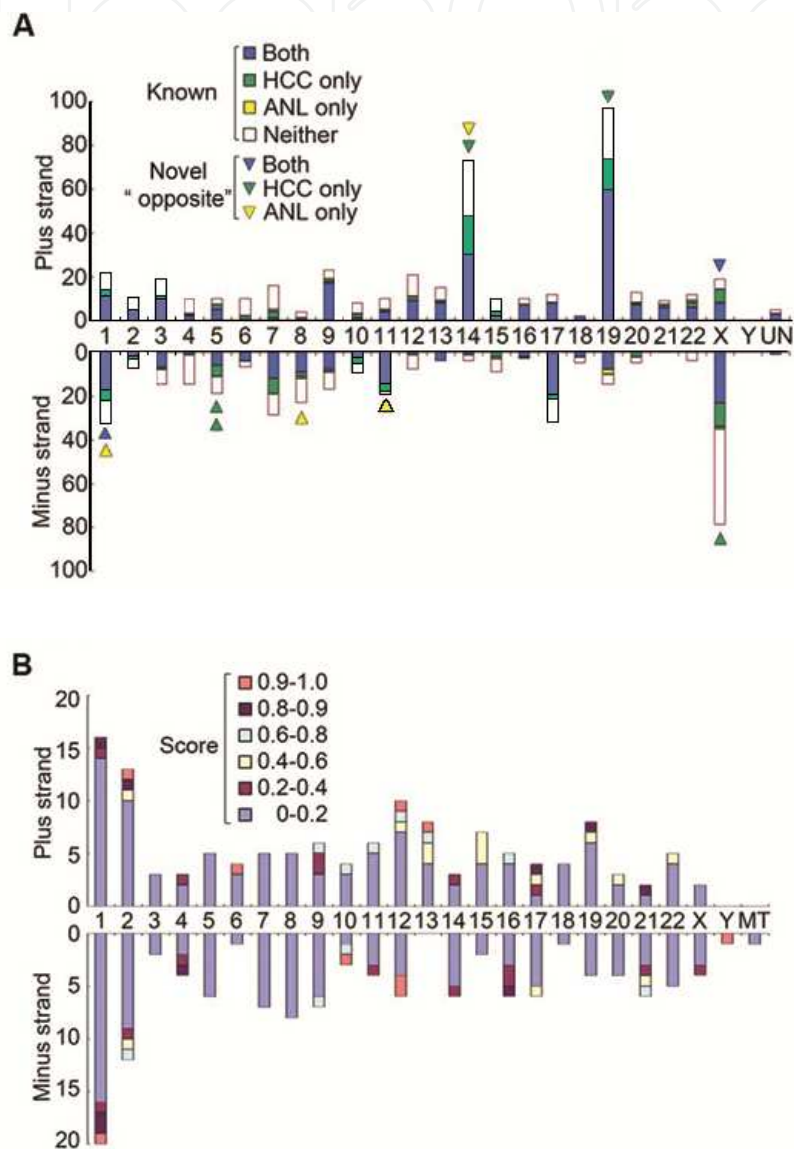
A: Chromosomal locations of our cloned mature miRNAs which correspond to previously identified pre-miRNAs. The bar graphs indicate the location of mature miRNAs (both our cloned miRNAs and previously identified miRNAs) based on known pre-miRNA sequences. Many of our cloned sequences were found in these regions. Arrowheads indicate novel "opposite" miRNAs cloned in our experiments. These sequences represent mature miRNAs which originate on the opposite DNA strand of the known precursor miRNA.

B: The 222 novel mature miRNAs identified in our study were checked for species conservation using the PhastCon database and their corresponding hairpin conservation scores charted in this figure.

Figure 2. Qualitative analysis of miRNAs in the human liver.

Accession Number	Most frequent sequence	Strand	Chr.	Precursor			
				Start	End	PhastCon score	dG (kcal/mol)
AB372573	TGCAATGGTTGTCCTGAGA	+	1	30294032	30294114	0.00	-28.6
AB372574	GTGGGGGAGAGGCTGTCA	+	1	150066781	150066870	0.00	-50.7
AB372575	TTGCACAACTCTCACATTC	+	1	152498870	152498979	0.39	-38.2
AB372576	CCCCTCACTTCTCCGTCATATC	+	1	153367241	153367332	0.05	-28.8
AB372577	CCCCAGATTCCACACC	+	1	154326362	154326427	0.85	-28.9

Table 2. Summary of the novel miRNAs

Summary of predicted novel miRNAs. The predicted novel miRNAs with their PhastCon scores are listed. And the most frequent cloned sequences, genome locations, clone counts, and conservation scores calculated with PhastCon are listed. The stem-loop is a 110-nt sequence derived by computational prediction. The precursor structures are listed in Table 3.

```
AB372573
UUGGA   U-       CAG     CCCC   AUUGA       CUUGUU
    CUC  GACAGCC   UGUA      UGG     GGGGGG
    GAG  CUGUUGG   ACGU      ACC     CCCCCU         A
A----   UC       UA-     ----   A----       ACUGUU

AB372574
---       ACA-          UG  -   UG     ---       G
   CCAGGC    GGCCUUUCCU  CC UUCC GGCCU   UCCUGG  G
   GGUCCG    UCGGAGAGGG  GG GAGG CCGGG   AGGACU  C
CUA       ACUG          -- U  UG     UCG       U

AB372575
A        CAGUGUAU   AG U      -     U    UAAC--     CCA   UA
 GGGGGGA          GCA  AG AAUGUG GAGUU GUGC       UCUAG   GCU
 CCUCCUU          UGU  UC UUACAC CUCAA CACG       AGGUC   UGA  A
G        UUU-----  CU  -      U     -    UUAAAU     AG-   UU

AB372576
CCAAACCCUGA    GAG       GUAG     C   U  -   C-  CUGUA
   GCU    CCUAGA    AUGAC GAGG AG GAGGG  GG
   CGA    GGGUCU    UACUG CUCU UC CUCCC  CC       G
A----------    ---      A---     C   -  A   CU  UUUUA

AB372577
UGG     AGA     A    -     -- C     G
   UCCCC   UUCC CACC CAGU  GG GCCA C
   AGGGG   AGGG GUGG GUCG  CU UGGU U
GG-     G--     A    C     AG U     C
```

Table 3. Samples of the Secondary structures of predicted novel miRNAs. The predicted novel miRNAs with their particular ID numbers are listed. The mature miRNAs are depicted in blue in the drawing of the precursor.

We cloned 210 novel microRNA candidates. Samples of the novel microRNAs that were detected from our study and its bioinformatics data are tabulated in Table 2 and 3. And those novel miRNAs have been deposited with DDBJ under consecutive accession codes from AB372573 to AB372814.

6. Future direction

We have been demonstrated the usefulness and accuracy of sequencing in genetic research of the liver. One of the main problems with applying sequencing to the miRNA transcription research is that sequencing is a time-consuming procedure. And an important consideration for the discovery of miRNA by sequencing is the difficultly in identifying miRNAs that are expressed at low levels, at highly specific stages or in rare cell types. Moreover, a serious problem is that some miRNAs are difficult to profile precisely due to their physical properties or post-transcriptional modifications, such as RNA editing. In principle, these limitations can be overcome by extensive sequencing of small RNA libraries from a broad range of samples. For differential display, the sequencing-based method has the theoretical advantage in that it has the capability to discover and detect novel miRNAs. Based on our sequence variability results, especially with regard to RNA modifications, the accuracy of the sequence-based method is expected to be superior to that of the hybridization-based method. For the prediction of novel miRNAs, methods that rely on phylogenetic 1 genes. To overcome this problem, we made use of a computational approach for structural conservation criteria using the thermodynamic stability and intrinsic structural features of miRNAs. In clinics, pathologists often meet difficult situations in which they cannot clearly tell whether the tissue specimens they are observing are malignant or benign. Thus, in our opinion, using some miRNAs as a tumor marker would help clinicians to clearly determine whether that tissue is cancerous. miRNA sequences followed by bioinformatics have greater power than individual miRNAs or other clinic-pathological variables for the detection of high risk patients' groups with poor prognoses. There is currently little data available as to how we can use each miRNA to predict high risk groups; however, additional future miRNA work and data accumulation will elucidate such criteria. And further investigation is warranted to clarify the mechanism of aberrant expression of miRNAs in cancer and its participation in carcinogenesis. Nevertheless, these findings show that sequence-based miRNA profiling has potential for the confirmation of precise miRNA dynamics in a specific disease. In addition, it will increase our understanding of the mechanisms and factors involved in human liver cancer.

Author details

Yoshiaki Mizuguchi and Eiji Uchida
Nippon Medical School, Department of Surgery for Organ Function and Biological Regulation, Japan

Takuya Mishima and Toshihiro Takizawa
Nippon Medical School, Department of Molecular Anatomy and Medicine, Japan

7. References

[1]   Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 116: 281–297.

[2]   Harfe BD (2005) MicroRNAs in vertebrate development. Curr Opin Genet Dev 15: 410–415.

[3]   Bartel DP, Chen CZ (2004) Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat Rev Genet 5: 396–400.

[4]   Rajewsky N (2006) microRNA target predictions in animals. Nat Genet 38 Suppl: S8–13.

[5]   Meyers BC, Souret FF, Lu C, Green PJ (2006) Sweating the small stuff: microRNA discovery in plants. Curr Opin Biotechnol 17: 139–146.

[6]   Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, et al. (2002) Identification of tissue-specific microRNAs from mouse. Curr Biol 12: 735–739.

[7]   Kim, VN. (2006) Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes. *Genes Dev.* 20: 1993-1997.

[8]   Mineno J, Okamoto S, Ando T, Sato M, Chono H, et al. (2006) The expression profile of microRNAs in mouse embryos. Nucleic Acids Res 34: 1765–1771.

[9]   Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31: 3429–3431.

[10] Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, et al. (2003) Human-mouse alignments with BLASTZ. Genome Res 13: 103–107.

[11] Siepel A, Haussler D (2004) Combining phylogenetic and hidden Markov models in biosequence analysis. J Comput Biol 11: 413–428.

[12] Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, et al. (2005) Phylogenetic shadowing and computational identification of human microRNA genes. Cell 120: 21–24.

[13] Azuma-Mukai A, Oguri H, Mituyama T, Qian ZR, Asai K, Siomi H, Siomi MC Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing. (2008) Proc Natl Acad Sci U S A 105: 7964-7969.

[14] Ro S, Park C, Jin J, Sanders KM, Yan W (2006) A PCR-based method for detection and quantification of small RNAs. Biochem Biophys Res Comm 351:756-763.

Websites for database used in this manuscript

NCBI Entrez Nucleotide database,
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide; European ribosomal RNA database, http://psb.ugent.be/rRNA/; Genomic tRNA database ,
http://lowelab.ucsc.edu/GtRNAdb/; RNAdb, http://research.imb.uq.edu.au/randb;

NONCODE, http://www.bioinfo.org.cn/NONCODE/; NCBI Reference Sequence, ftp://ftp.ncbi.nih.gov/refseq/; UCSC Genome Bioinformatics Site, http://genome.ucsc.edu; OncoDb HCC, http://oncodb.hcc.ibms.sinica.edu.tw/index.ht