

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 March 2006 (30.03.2006)

PCT

(10) International Publication Number
WO 2006/033020 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/IB2005/003493

(22) International Filing Date: 26 April 2005 (26.04.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/521,433 26 April 2004 (26.04.2004) US
60/521,563 25 May 2004 (25.05.2004) US
60/522,300 14 September 2004 (14.09.2004) US
60/522,453 3 October 2004 (03.10.2004) US
60/522,454 3 October 2004 (03.10.2004) US
60/522,762 4 November 2004 (04.11.2004) US

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (for all designated States except US):
ROSETTA GENOMICS LTD. [IL/IL]; 10 Plaut Street,
76706 Rehovot (IL).

(72) Inventor; and

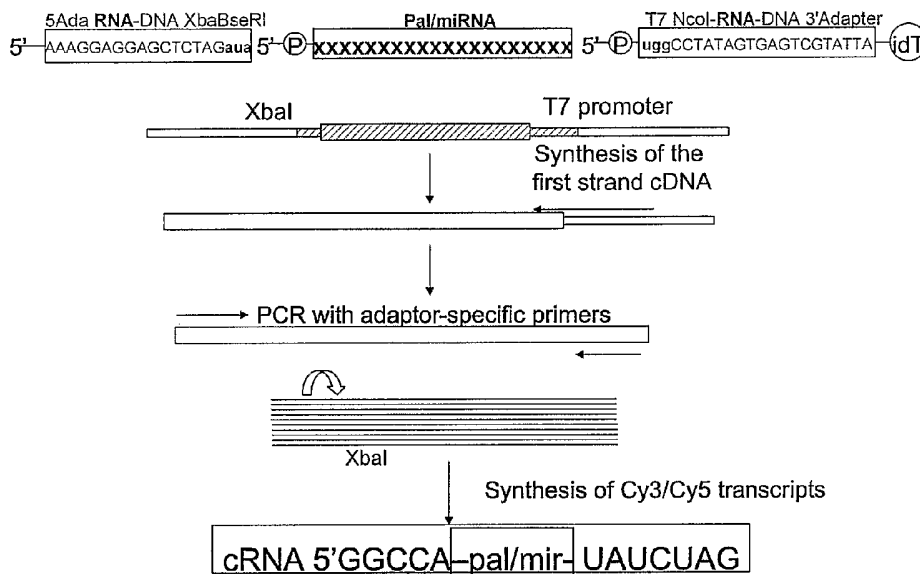
(75) Inventor/Applicant (for US only): BENTWICH, Itzhak
[IL/IL]; 76802 Kvutzat Shiler (IL).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND APPARATUS FOR THE DETECTION AND VALIDATION OF MICRORNAS



(57) Abstract: The present invention provides methods and apparatus for detecting, sequencing, cloning and otherwise validating nucleic acids, including microRNAs. The methods make use of a plurality of sequence-specific recognition reagents which can also be used for classification of biological samples, and to characterize the expression of certain microRNAs.

WO 2006/033020 A2

**METHODS AND APPARATUS FOR THE DETECTION
AND VALIDATION OF MICRORNAS**

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims the benefit of U.S. Provisional Patent Application No. 60/521,433, filed April 26, 2004, U.S. Provisional Patent Application No. 60/521,563, filed May 25, 2004, U.S. Provisional Patent Application No. 60/522,300, filed September 14, 2004, U.S. Provisional Patent Application No. 60/522,454, filed October 3, 2004, U.S. Provisional Patent Application No. 60/522,453, filed October 3, 2004 and U.S. Provisional Patent Application No. 60/522,762, filed November 4, 2004.

FIELD OF THE INVENTION

[0002] The present invention is related to methods for identifying nucleic acids, such as microRNAs.

BACKGROUND OF THE INVENTION

[0003] MicroRNAs (miRNAs) are short RNA oligonucleotides of approximately 22 nucleotides that play an important role in gene regulation. MiRNAs regulate gene expression by targeting mRNAs for cleavage or translational repression. Although miRNAs are present in a wide range of species including *C. elegans*, *Drosophilla* and humans, they have only recently been identified. Although a limited number of miRNAs have been identified by extracting large quantities of RNA, miRNAs are difficult to identify using standard methodologies as a result of their small size.

[0004] Computational approaches have recently been developed to identify the remainder of miRNAs in the genome. Tools such as MiRscan, MiRseeker and those described in U.S. Patent Application Nos. 60/522,459, 10/709,577 and 10/709,572 have predicted a large number of miRNAs in the human genome. It would be beneficial to validate those predicted miRNAs that exist *in vivo*, and to determine the expression profiles of these miRNAs.

[0005] Microarrays allow the high throughput analysis of gene expression. Microarray technology is based on measuring the hybridization of a target sequence to a probe sequence attached to a substrate. A limitation of microarrays is that only hybridization is measured,

without any indication of the degree of complementarity between the probe and the target gene. This indirect evidence of a target sequence is of little concern when the target sequence is a relatively long sequence that can be positively identified by using multiple probe sequences per target. Such a practice is of little benefit for the identification and confirmation of short nucleic acid sequences, such as miRNAs.

SUMMARY OF THE INVENTION

[0006] The present invention is related to a method of detecting a miRNA. An array may be provided comprising a solid substrate and a plurality of positionally distinguishable polynucleotides attached to the solid substrate. Each polynucleotide may comprise a miRNA. The array may be contacted with a plurality of target polynucleotides comprising a complement of a miRNA under conditions permitting hybridization. Hybridization of a target sequence to the miRNA may be detected. A miRNA may be detected when hybridization is above background.

[0007] The plurality of target polynucleotides may be produced by providing RNA comprising a plurality of miRNA. The RNA may be less than 160 nucleotides in length. Adapters may then be ligated to the 5' and 3' ends of the RNA. The adapters may comprise a restriction site, which may be used later to remove the adapters. The adapters may be DNA-RNA hybrids. First strand cDNA of the 5'-adapter-miRNA-adapter-3' may then be prepared. The adapter-miRNA-adapter may then be amplified. cRNA may then be prepared using a promoter complementary to the 3' adapter.

[0008] A miRNA may also be detected by providing a plurality of target polynucleotides comprising a miRNA, a labeled oligo that is complementary to a portion of the target nucleotides, and substrate comprising a capture oligonucleotide comprising at least 16 nucleotides of a miRNA complementary sequence. The target nucleotides may then be contacted with the labeled oligo and substrate. Hybridization of the target nucleotides, labeled oligo and substrate may then be detected.

[0009] The present invention is related to a method of isolating a miRNA. A solid substrate may be provided comprising a capture oligonucleotide comprising at least 16 nucleotides of a miRNA sequence. The capture oligonucleotide may be contacted with a plurality of target polynucleotides comprising a complement of a miRNA under conditions permitting

hybridization. The target polynucleotides may then be eluted from the capture oligonucleotide. The eluted target polynucleotide may be sequenced. The eluted target polynucleotide may also be sequenced.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] Fig. 1 demonstrates model for maturation of mammalian miRNAs.

[0011] Fig. 2 demonstrates the preparation of a target cDNA library.

[0012] Fig. 3 shows results using a MIRChip. Panel A shows the signal ratio on a MIRChip using probes that include a miRNA precursor, a miRNA in the 5' portion of a probe, and a miRNA precursor with no more than 16 nucleotides of the miRNA sequence. Panel B shows a mismatch analysis of pre-miRNA-125b showing the location on the WT sequence of the mismatches that were included.

[0013] Fig. 4 shows results using different types of probes. Panel A shows results of probes containing miRNAs in the 5', middle, or 3' end of the probe. Panel B shows results of probes containing one, two, or three miRNA copies. Panel C shows results of probes containing two or three miRNA copies containing no mismatches, 2 mismatches in the 5' miRNA, 2 mismatches in the middle or 3' miRNA, or miRNA 2 mismatches in each of 2 miRNAs.

[0014] Fig. 5 shows the effect of the number of mismatches in the probe on the signal intensity of the MIRChip at either 50°C or 60°C.

[0015] Fig. 6 shows expression of the 150 human miRNAs in five tissues and HeLa cells.

[0016] Fig. 7 shows expression of tissue-specific or highly enriched miRNAs in five human tissues using MIRChip.

[0017] Fig. 8 shows an illustration of the MIRAClone method.

[0018] Fig. 9 shows the cloning of human mir-21. Panel A shows amplification of recovered library molecules using primers matching to the adaptors. The PCR product was detected from cRNA as target (lane 1), amplified cDNA (lane 2) or non-amplified cDNA (lane 3). No PCR products were observed when the entire procedure was performed without the addition of library molecules (lane 4), or in mock PCR (lane 5). Panel B shows PCR on ligation. The presence of the candidate miRNAs in the ligated vector was verified using a primer specific to mir-21 and a primer located downstream (lanes 1-4) or upstream (lanes 5-8) of the MCS on the vector. Amplified PCR products were detected only when specific ligation was tested (lanes 1, 5

amplified cDNA, and lanes 2, 6 non-amplified cDNA). Controls of ligation from another recovered miRNA gave no amplification (lanes 3 and 7). Panel C shows colony hybridization with DIG labeled oligonucleotide probes complementary to mir-21. Panel D shows colony PCR on ten colonies with a mir-21 specific primer and a primer located upstream or downstream to the MCS of the vector. Only clones that were positive in the colony hybridization assay (lanes 1, 4, 6, 7, and 9, also marked in Panel C) gave the expected PCR products. Mock PCR gave no amplification (lane 11).

[0019] Fig. 10 shows clones containing the authentic sequence of miRNAs expressed above background levels in a microarray analysis of placenta-derived miRNAs.

[0020] Fig. 11 describes the elongation of capture oligonucleotide. The top capture oligonucleotide is composed of the longer mature mir-21 sequence. Capture oligonucleotides containing extra nucleotides derived from the precursor sequence are found below. The sequence of mir-21 is underlined.

[0021] Fig. 12 shows the structure and sequence of the six predicted hairpins and miRNAs used as templates for MIRclone cloning. The sequence composing the capture oligonucleotide is boxed. The sequence of the predicted miRNA is in bold. The arrows mark the boundary of the actual sequence of the cloned miRNAs. For mir-RG-27 and mir-RG-21, the multiple arrows mark the various 3' ends observed in different clones.

[0022] Fig. 13 shows a chromosomal cluster analysis of novel miRNAs. The location of the novel miRNAs relative to each other or to published miRNAs is depicted for chromosomes 14, 17, and X. The boxes represent miRNA precursors and the thick line within the box represents the location of the mature miRNA sequence within the precursor.

DETAILED DESCRIPTION

[0023] While not being bound by theory, the current model for maturation of mammalian miRNAs is shown in Figure 1. The first step may be the nuclear cleavage of the pri-miRNA. The pri-miRNA may be part of a polycistronic RNA comprising multiple miRNAs. Cleavage of the pri-miRNA may liberate a 60–70 nt stem loop intermediate, known as the miRNA precursor, or the pre-miRNA. The processing of the pri-miRNA may be performed by the Drosha RNase III endonuclease, which may cleave both strands of the stem at sites near the base of the primary stem loop. Drosha may cleave the RNA duplex with a staggered cut typical of RNase III

endonucleases, and thus the base of the pre-miRNA stem loop may have a 5' phosphate and ~2 nt 3' overhang. The pre-miRNA may then be actively transported from the nucleus to the cytoplasm by Ran-GTP and the export receptor Exportin-5.

[0024] The cleavage by Drosha may define one end of the mature miRNA. The other end of the miRNA may be processed in the cytoplasm by the enzyme Dicer. Dicer, also an RNase III endonuclease, may also be involved in generating the small interfering RNAs (siRNAs) that mediate RNA interference (RNAi). Dicer perform may perform an activity in metazoan miRNA maturation similar to that which it performs in cleaving double-stranded RNA during RNAi. Dicer may first recognize the double-stranded portion of the pre-miRNA, perhaps with particular affinity for a 5' phosphate and 3' overhang at the base of the stem loop. Then, at about two helical turns away from the base of the stem loop, Dicer may cut both strands of the duplex. The cleavage by Dicer may cleave off the terminal base pairs and loop of the pre-miRNA, leaving the 5' phosphate and ~2 nt 3' overhang characteristic of an RNase III and producing an siRNA-like imperfect duplex that comprises the mature miRNA and a similar-sized fragment derived from the opposing arm of the pre-miRNA. The fragments from the opposing arm, called the miRNA* sequences, are found in libraries of cloned miRNAs but typically at much lower frequency than the miRNAs.

[0025] The specificity of the initial cleavage mediated by Drosha may determine the correct register of cleavage within the miRNA precursor and thus may define both mature ends of the miRNA. The determinants of Drosha recognition may include a larger terminal loop (~10 nt). From the junction of the loop and the adjacent stem, Drosha may cleave approximately two helical turns into the stem to produce the pre-miRNA. Beyond the pre-miRNA cleavage site, approximately one helical turn of stem extension (~10 nt) may be essential for efficient processing.

[0026] Following cleavage, the miRNA pathway may be identical to the RNA silencing pathways known as posttranscriptional gene silencing. Although initially present as a double-stranded species with miRNA*, the miRNA may eventually become incorporated as single-stranded RNAs into a ribonucleoprotein complex, known as the RNA-induced silencing complex (RISC). When the miRNA strand of the miRNA:miRNA* duplex is loaded into the RISC, the miRNA* may be removed and degraded. The strand of the miRNA:miRNA* duplex that is loaded into the RISC may be the strand whose 5' end is less tightly paired. In cases where both

ends of the miRNA:miRNA* have roughly equivalent 5' pairing, both miRNA and miRNA* may have gene silencing activity.

[0027] The RISC may identify target messages based on a high levels of complementarity between the miRNA and the mRNA. The target sites in the mRNA may be in the 5' UTR, the 3' UTR or in the coding region. Interesting multiple miRNAs may regulate the same mRNA target by recognizing the same or multiple sites. The presence of multiple miRNA complementarity sites in most genetically identified targets may indicate that the cooperative action of multiple RISCs provides the most efficient translational inhibition.

[0028] MiRNAs may direct the RISC to downregulate gene expression by either of two mechanisms: mRNA cleavage or translational repression. The miRNA may specify cleavage of the mRNA if the mRNA has sufficient complementarity to the miRNA. When a miRNA guides cleavage, the cut may be between the nucleotides pairing to residues 10 and 11 of the miRNA. Alternatively, the miRNA may repress translation if the miRNA does not have sufficient complementarity to the mRNA. Translational repression may be more prevalent in animals since animals may have a lower degree of complementarity.

[0029] We have developed apparatus and methods for identifying miRNAs. We have also developed apparatus and methods for sequencing miRNAs. Moreover, we have developed apparatus and methods for cloning miRNAs.

[0030] Before the present apparatus, products and compositions and methods are disclosed and described, it is to be understood that the terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting. It must be noted that, as used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise.

1. Detection

[0031] The present invention is related to a microarray comprising a solid substrate comprising a plurality of capture sequences, which may be used for detecting the presence of a target nucleic acid in a sample. A nucleic acid-containing sample may be contacted with the array. Binding of the target nucleic acid to the capture sequence may be detected, and the extent thereof may be measured.

a. Substrate

[0032] The solid substrate may be any of the many materials available in the art. Representative examples of solid substrates include glass, plastic or a polymeric substrate.

b. Capture Sequences**(1) First Nucleic Acid**

[0033] Each capture sequence comprises a first nucleic acid. The first nucleic acid may be a miRNA, a miRNA*, a pre-miRNA, a pri-miRNA, the complement thereof, a nucleic acid substantially identical thereto, or a portion thereof at least 12, 15, 17, 18, 19, 20, 21, 22 or 23 nucleotides, or a DNA encoding said sequence. A substantially identical nucleic acid may have greater than 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to the reference nucleic acid.

[0034] Mature miRNAs usually have a length of 19-24 nucleotides, particularly 21, 22 or 23 nucleotides. The miRNAs may also be provided as a precursor which may have a length of 50-90 nucleotides, particularly 60-80 nucleotides. It should be noted that the precursor may be produced by processing of a primary transcript which may have a length of >100 nucleotides.

[0035] The nucleic acids may be selected from RNA, DNA or nucleic acid analog molecules, such as sugar- or backbone-modified ribonucleotides or deoxyribonucleotides. It should be noted, however, that other nucleic analogs, such as peptide nucleic acids (PNA) or locked nucleic acids (LNA), are also suitable.

[0036] The nucleic acids may be an RNA- or DNA molecule, which contains at least one modified nucleotide analog, i.e. a naturally occurring ribonucleotide or deoxyribonucleotide is substituted by a non-naturally occurring nucleotide. The modified nucleotide analog may be located for example at the 5'-end and/or the 3'-end of the nucleic acid molecule. Representative examples of nucleotide analogs may be selected from sugar- or backbone-modified ribonucleotides. It should be noted, however, that also nucleobase-modified ribonucleotides, i.e. ribonucleotides, containing a non-naturally occurring nucleobase instead of a naturally occurring nucleobase such as uridines or cytidines modified at the 5-position, e.g. 5-(2-amino)propyl uridine, 5-bromo uridine; adenosines and guanosines modified at the 8-position, e.g. 8-bromo guanosine; deaza nucleotides, e.g. 7-deaza-adenosine; O- and N-alkylated nucleotides, e.g. N6-methyl adenosine are suitable. The 2'-OH-group may be replaced by a group selected from H, OR, R, halo, SH, SR, NH₂, NHR, NR₂ or CN, wherein R is C₁-C₆ alkyl, alkenyl or alkynyl and

halo is F, Cl, Br or I. The phosphoester group of backbone-modified ribonucleotides connecting to adjacent ribonucleotides may be replaced by a modified group, e.g. of phosphothioate group. It should be noted that the above modifications may be combined.

(2) Second Nucleic Acid

[0037] Each capture sequence also comprises a second nucleic acid of at least 20, 25, 30, 35, 40, 45, 50, 55 or 60 nucleotides. The second nucleic acid may be used to anchor the first nucleic acid to the substrate. The second nucleic acid may have features that minimize background hybridization of sample nucleic acids to the capture sequence. For example, the second nucleic acid may not appear in the genome of the organism from which the sample is derived. The second nucleic acid may have also less than 25%, 30%, 35%, 40%, 45%, 50%, or 55% identity to any sequence in the genome of the organism from which a sample is derived. Each 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotide window of the second nucleic acid may have less than 80% identity to any sequence in the genome of the organism from which a sample is derived. Such properties of the second nucleic acid sequence may yield better specificity compared to the triplet method, which cannot differentiate between binding of a target sequence to the first or second nucleic acid.

(3) Control Sequences

[0038] The microarray may comprise one or more negative control sequences. Representative examples of such negative controls include the second nucleic acid sequence by itself, palindrome sequences, mRNA for coding genes, adaptors added in the preparation of the library, tRNA and snoRNA.

[0039] The microarray may also comprise mismatch probes. For any given capture sequence, multiple mismatch sequences may be generated by changing nucleotides in different positions of the capture sequence. For example, one or more nucleotide may be replaced with its respective complementary nucleotides (A \longleftrightarrow T/U, G \longleftrightarrow C, and vice versa). Mismatch control sequences may be used to determine the degree of complementarity between the binding between the target sequence and the first nucleic acid. Mismatches in the second nucleic acid may not generate a significant change in the intensity of the probe signal, while mismatches in the first nucleic acid may induce a significant decrease in the probe intensity signal. Mismatches in the first nucleic acid may be used to determine that a particular position does not represent a perfect complementary match between the first nucleic acid and the target sequence.

c. Nucleic Acid Sample

[0040] The nucleic acid sample comprises a plurality or library of target sequences. The target sequences may comprise sequences that are substantially complementary to the first nucleic acid. The target sequences may be DNA, RNA or a hybrid thereof.

[0041] The target sequences may be prepared by one of many methodologies available in the art. For example, total RNA may be size fractionated to isolate RNA sequences less than or equal to 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 or 200 nucleotides. In one embodiment, the isolated RNA sequences are approximately 20 nucleotides.

[0042] Adapters may then be ligated to the 5'- and 3'-ends of the size-fractionated RNA. The 3' adapter may have a T7 promoter. The 5'- and 3'-adapter may each have restriction sites that allow later cleavage of the adapter. The adapters may be DNA-RNA hybrids. The RNA sequence of the DNA-RNA hybrids may be adjacent to the size-fractionated RNA after ligation.

[0043] First strand cDNA may then be produced by reverse transcription. The resulting double-stranded product may then be amplified using the polymerase chain reaction (PCR). PCR may be carried out using labeled nucleotides. The adapters may then be removed from the amplified sequences by using restriction enzymes that are specific for sites present in the design of the adapters. The resulting cDNA products may then be converted to cRNA.

[0044] In order to reduce the presence of tRNA sequences in the library, the 3'-adapter may be designed to have a 5'-sequence that yields a restriction site after ligation to the 3'-end of a tRNA. For example, a 5'-adapter sequence of GGT ligated to the 3'-end of a tRNA (ACC at their 3' end) yields a restriction site for NcoI. Such a restriction enzyme may be used to cleave the tRNA containing sequences prior to or after PCR.

d. Hybridization Analysis

[0045] The microarray may be contacted with the nucleic acid sample under stringent or moderately stringent hybridization conditions, thereby allowing a target sequence to hybridize to a sufficiently complementary probe sequence. The intensity at each probe sequence is then measured. The probe signals may be evaluated using parameters including, but not limited to, background signals, controls signals, comparison to signals from mismatch probe sequences.

2. Alternative Detection

[0046] The present invention is also related to a method of detecting a target sequence by contacting a plurality of target sequences with a labeled nucleic acid, whereby a labeled nucleic

acid may hybridize to a first portion of the target sequence to yield a partial duplex. The partial duplex may then be contacted with a solid substrate comprising a plurality of capture sequences, which may be coupled to color-coded microspheres, whereby a capture sequence may hybridize to a second portion of the target sequence to yield a captured duplex. Binding of the partial duplex to the capture sequence may be detected by measuring the signal of binding at the capture sequence.

3. Sequencing and Cloning

[0047] The present invention is also related to a method of sequencing or cloning a target nucleic acid in a sample that hybridizes to a capture sequence coupled to a solid substrate, such as a magnetic bead. In the preparation of the nucleic acid sample, the adapters are not removed from the library of target sequences. The plurality of target sequences comprising 5'- and 3'- adapters is contacted with one or more solid substrates each individually comprising a capture sequence, thereby allowing hybridization of a probe sequence to a target sequence of sufficient complementarity. The bound target sequences may then be dislodged from the solid substrate using any chemical or physical method in the art. The dislodged target sequences may be amplified using primers that hybridize to the 5'- and 3'-adapters. The amplified target sequence may then be cloned or sequenced using any method in the art.

[0048] The present invention has multiple aspects, illustrated by the following non-limiting examples.

Example 1

Design of a Microarray

[0049] Microarray chips were produced by attaching various probe sequences of 60 nucleotides to a substrate. The probes contained known or predicted miRNAs, as well as various controls. Known miRNAs were attached to MIRChip1 and predicted miRNAs were attached to MIRChip2.

1. Single miRNA probes

[0050] From each miRNA precursor we took a 26-mer containing the miRNA, then assigned 3 probes for each extended miRNA sequence: the 26-mer at the 5' of the 60-mer probe, the 26-mer at the 3' of the 60-mer probe, and the 26-mer in the middle of the 60-mer probe. Two different 34-mer sequences which do not appear in the human genome (NHG-sequences) were attached to

the 26-mer to complete a 60-mer probe. The NHG-sequences were a combination of 10-mer sequences which are very rare in the human genome. Each potential 34-mer sequence was compared to the human genome by the Blast program and we ended up with 2 different rare NHG-sequences that have an identity of no more than 40% and have no 15-mer sub-sequences with are more than 80% identical.

[0051] For a subset of 32 of single miRNA probes we designed an additional 6 mismatch mutation probes: a) a block of 4 mismatches at the 5' end of the miRNA; b) a block of 6 mismatches at 3' end of the miRNA; c) one mismatch at position 10 of the miRNA; d) two mismatches at positions 8 and 17 of the miRNA; e) three mismatches at positions 6, 12 and 18 of the miRNA; and f) six mismatches at different positions outside of the miRNA.

2. Duplex miRNA probes

[0052] From each precursor we took a 30-mer containing the miRNA, and then duplicated it to obtain a 60-mer probe. For a subset of 32 of miRNAs we designed additional 3 mismatch mutation probes: a) two mismatches on the first miRNA; b) two mismatches on the second miRNA; and c) two mismatches on each of the miRNAs.

3. Triplex miRNA probes

[0053] Similar to methods described in Krichevsky et al (2003), we attached ~22 nucleotide long miRNA sequences head-to-tail to obtain 60-mer probes containing up to 3 times the same miRNA sequence. For a subset of 32 probes, we designed an additional 3 mismatch mutation probes: a) two mismatches on the first miRNA; b) two mismatches on the second miRNA; and c) two mismatches on each of the miRNA copies.

4. Precursor with miRNA probes

[0054] For each precursor we took a 60-mer sequence containing the entire miRNA.

5. Precursor without miRNA probes

[0055] For each precursor we took a 60-mer sequence containing no more then 16 nucleotides of the miRNA. For a subset of 32 probes, we designed additional mismatch probes containing 4 mismatches.

6. Controls

[0056] General control included the following: 100 probes for mRNAs, representing mostly genes expressed in a wide variety of cell types, 85 representative tRNAs, and 19 representative snoRNA probes. Negative controls included one group composed of 294 randomly chosen 26-

mer sequences from the human genome not contained in published precursors sequences, placed at the 5' and complemented with a 34-mer NHG-sequence. A second group was composed of 182 different 60-mer probes containing different combinations of 10-mer rare sequences.

Example 2

Design of a Target Library

[0057] A cDNA target library was made using a procedure similar to that described in Elbashir et al., Genes Dev. 2001 15:188-200. Briefly, total RNA was size-fractionated using an YM-100 column to isolate RNA of about 200 nucleotides. Adaptor sequences were then ligated to the 5'- and 3'- ends of the size-fractionated RNA (Fig. 2). Both adaptors were RNA-DNA hybrids with the RNA portion ligated directly to the size-fractionated total RNA. The 3'-adaptor included a T7 promoter. Either the first or second pair of the following adaptors was used:

Table 1

5' Adapter	
5'-AAAGGAGGAGCTCTAG _{aua} -3'	SEQ ID NO: 1
3' Adapter	
5'-P- _{ugg} CCTATAGTGAGTCGTATTA-idT-3'	SEQ ID NO: 2
5' Adapter	
5'-CCTAGGAGGAGGACGTCTG _{cag} -3'	SEQ ID NO: 3
3' Adapter	
5'-P- _{ccu} ATAGTGAGTCGTATTATCT-idT-3'	SEQ ID NO: 4

*nucleotides in lowercase represent ribonucleotides and the nucleotides in uppercase represent deoxyribonucleotides

[0058] After ligating the adaptors to the RNA, the product was converted to first strand cDNA by reverse transcription. The resulting cDNA was then amplified by polymerase chain reaction (PCR) using one of the following pairs of primers:

Table 2

5' Primer	
5'-TAATACGACTCACTATAGGCCA-3'	SEQ ID NO: 5
3' Primer	
5'-AAAGGAGGAGCTCTAGATA-3'	SEQ ID NO: 6
5' Primer	
5'-GCTAGCACTAGTTAATACGACTCACTATAGGCCA-3 '	SEQ ID NO: 7
3' Primer	
5'-GCTCTAGGATAATACGACTCACTATAGG-3 '	SEQ ID NO: 8
5' Primer	
5'-TGACCTGCAGAAAGGAGGAGCTCTAGATA-3'	SEQ ID NO: 9
3' Primer	
5'-ATCCTAGGAGGAGGACGTCTGCAG-3'	SEQ ID NO: 10

[0059] After amplification, the amplified DNA was digested with Xba1 or Pst to remove the adaptor sequences that were added to the initial RNA. Using the first set of RNA-DNA hybrid adaptors listed above, the first set of primers listed above, and Xba1 yielded cRNA-1. Using the second set of RNA-DNA hybrid adaptors listed above, the second set of primers listed above, and PstI yielded cRNA-2. The resulting cDNA products were then converted to labeled cRNA (lcRNA) incorporating either 3-CTP (Cy3-CTP) or cyanine 5-CTP (Cy5-CTP). The lcRNA was purified using a G-50 column.

Table 3

cRNA Products	
5'-GGCCA - pallindrome/miRNA- UAUCUAG-3'	cRNA-1
5'-GG- pallindrome/miRNA-C-3'	cRNA-2

Example 3

Expression Analysis of Known MiRNAs

[0060] To examine the ability of miRNAs or pre-miRNAs in the lcRNA to hybridize to theMIRChip1, we examined hybridizations with 5, 17 or 50 µg of lcRNA derived from HeLa cells. Hybridization solutions that contained the indicated amount of each lcRNA from either the control or the test samples were prepared using the *In situ* Hybridization Reagent Kit (Agilent). Hybridized microarrays were scanned using the Agilent LP2 DNA Microarray Scanner at 10 µm resolution. Microarray images were visually inspected for defects.

[0061] Microarray images were analyzed using Feature Extraction Software (Version 7.1.1, Agilent). We set the signal of each probe as its median intensity. We observed a nearly constant background intensity signal of 430. Using NHG-sequence negative control probes, the threshold for reliable probe signals was set at 1500. No NHG-sequence probes with signals higher than 1500 were observed in HeLa, Brain, liver and thymus and less than 0.5% of these probes gave signals higher than 1500 in testes and placenta. In all hybridization experiments a high correlation of 0.96 to 0.98 was observed between the Cy5-labeled common control lcRNA. In addition, lcRNAs derived from the same RNA source and hybridized to MIRChip1 and MIRChip2 (below) gave a correlation coefficient of 0.98 when identical probes on the two chips were compared.

[0062] The hybridization results showed that 17 µg of lcRNA gave the optimal outcome. In general, signal intensity of miRNA containing probes followed their known abundance in HeLa cells. In contrast, the antisense and 4 palindrome probes outside the miRNA gave no signal above background. This shows that the signals were derived from miRNAs and not from their hairpin precursors.

[0063] Of the other controls, signals of tRNA probes were at most similar to those of the most abundant miRNAs, while probes for abundant mRNAs gave only background signals. Hybridizations of MIRChip1 with total RNA oligo-dT derived lcRNA resulted in the expected pattern of signals from the mRNA probes but no signals above background were observed from the miRNA-containing probes.

Example 4

Expression Analysis of Predicted MiRNAs

[0064] MIRChip2 was hybridized with 17 μ g of lcRNA derived from HeLa cells. A comparison of 60-mer probes containing miRNAs within their precursor sequence to those in which the miRNAs were embedded in NHG-sequences show that both give similar signal levels (Fig. 3A). In contrast, probes containing precursor sequences without miRNAs or with truncated miRNAs gave low or background signals (Fig. 3A). Moreover, a similar hybridization on MIRChip1, which included mismatches either in the miRNA or in the non-miRNA precursor regions, showed that mismatches within the miRNA sequence result in significant reduction in signal intensity while no change is observed in mismatches outside the miRNA (Fig. 3B). Control 60-mer probes composed only of the NHG-sequences gave only background signals. This shows that miRNAs, and not their hairpin precursors, are responsible for the observed signals.

[0065] MIRChip2 included miRNAs in three locations along the 60-mer probes to examine the importance of miRNA location. Fig. 4A shows that miRNAs located at the 5' end of the 60-mer probes result in significantly higher signals than miRNAs located in the middle, with miRNAs located at the 3' end giving the lowest signals. Comparison of the 60-mer probes containing a single miRNA to the duplex and triplex 60-mer probes show that the inclusion of additional miRNA copies in the 60-mer probes results, at most, in a minor increase in signal intensity (Fig. 4B). Moreover, analysis of duplex and triplex 60-mer probes containing mismatches revealed that, in general, mismatches within the 5' miRNA cause a significant reduction in signal levels while mismatches in miRNAs located in the middle or in the 3' ends had a significantly lower effect on signal intensity (Fig. 4C). This shows that a single miRNA located at the 5' end of the probe, furthest from the surface of the chip, is sufficient to obtain high signals.

[0066] An important control of hybridization specificity is the effect of mismatches on observed signals. Fig. 5 shows the results of hybridizations at temperatures of 50°C and 60°C for a subset of 32 miRNAs, each in two different settings of NHG-sequences, for which mismatch-probes were included. While mismatches outside the miRNA sequences did not change signal levels, one, two or three mismatches within the miRNA significantly reduced the signal. In the 60°C hybridization, even a single mismatch reduced the signal to close to background levels compared to a significantly lower reduction of signal intensity in the 50°C hybridization, in accordance with the lower stringency of these conditions. Similarly, mismatches in either the 5' or 3' regions

of the miRNA significantly reduced the signal intensity with higher effects at the 60°C hybridization temperature. Thus, under the standard conditions using hybridizations at 60°C, specificity was high.

[0067] Our findings that single mismatches in the middle of the miRNA sequence, or 2 mismatches in either side, reduce the signal to background levels suggest that the signals are specific. Moreover, miRNAs that are different by few nucleotides from each other often show different expression patterns. As shown below, miRNAs let-7A and let-7B, which differ from each other in only 2 nucleotides, have a very similar pattern, while let-7c, which is one nucleotide different from both let-7A and let-7B, has a different expression pattern with significantly lower expression in placenta and brain but not in the other tissues. Taken together, our data strongly support the specificity of the signals observed using the MIRChip.

Example 5

Expression Profile of Predicted MiRNAs

[0068] We next hybridized the MIRChip2 with lcrnAs derived from human brain, liver, thymus, testes, and placenta and examined the tissue specificity of the various miRNAs. The results obtained from the HeLa-cell hybridization mentioned above were included in the analysis. The full set of results can be found in Fig. 6. A comparison was made to results obtained by Sempere et al (2004) that examined the expression of 119 miRNAs by Northern blots in brain and liver as well as other tissues not examined in the present study. Comparison was also made, when relevant, to the oligonucleotide array results of Krichevsky et al (2003) and to the cloning data of Lagos-Quintana et al (2002). MicroRNAs showing distinct brain (e.g. miRNA-9 and miRNA-124A) or liver (miRNA-122A and miRNA-194) tissue specificity gave identical results on our MIRChip hybridizations (Fig. 7). Also, the findings that certain miRNAs, such as let-7A, let-7B, and miRNA-30C (Sempere et al. 2004), are expressed at high levels in many tissues were confirmed using our microarrays, extending the results to the thymus, testes and placenta (Fig. 7). An overall correlation of approximately 0.6 was found between our results and those of Sempere et al. (2004).

[0069] We also found distinct differences between our study and those of others. For example, we found very high expression of miRNA-149 in the brain and high expression in the liver whereas Sempere et al. (2004) found low levels in the brain and no signal in the liver. Similarly,

we detected significant expression levels of miRNA-20 in both brain and liver compared to no signals on the Northern blots reported by Sempere et al. (2004). On the other hand, miRNA-203 and miRNA-137 showed only background signals in our study compared to high levels of expression in both brain and liver or in the brain, respectively, observed by Sempere et al. (2004).

Example 6

Further Validation of Expression of Predicted MiRNAs

[0070] As an additional way of validating expression of the miRNAs, we used a fluorescence-based hybridization method developed by Luminex (Yang et al. 2001) termed "miRNAMASA." The miRNAMASA technology uses a specific capture-oligo for each targeted miRNA. The capture oligo was covalently coupled onto color-coded microspheres (beads), and was used together with a detection-oligo that was labeled with biotin (Fig. 8). Both capture and detection oligos are spiked with Locked Nucleic Acid (LNA) nucleotides to increase specificity and sensitivity (Petersen and Wengel, 2003). Following hybridization of the capture and detection oligos with the RNA, streptavidine-phycoerythrin is added. The fluorescence associated with the color-coded beads provides a measure for miRNA expression level.

[0071] We have focused the miRNAMASA validation study on those miRNAs showing distinct differences between MIRChip1 and the previously published Northern blot data. The analysis was done by multiplexing in two groups. One group included let-7b, miRNA-127, miRNA-129, miRNA-137, miRNA-203 and 5sRNA control. The second group included miRNA-20, miRNA-199a, miRNA-141 and 5sRNA control. The analysis of each group was done on 1 µg of total RNA. A negative bead-control was performed for each group, shown as "blank" in Fig. 9. As shown in Fig. 9, the expression of miRNA-20 was detected in brain and liver, as well in the other three tissues, compared to no signals observed in the Northern blot analysis of Sempere et al. (2004). On the other hand, no expression is observed in any of the tissues for miRNA-137 and miRNA-203 compared to expression of miRNA-137 in the brain and of miRNA-203 in both brain and liver, observed by Sempere et al. (2004). In addition, a good correlation was observed between the miRNAMASA and the MIRChip results in the expression of miRNA-141 and let-7B in all five tissues. These results validate the expression patterns observed in the MIRChip experiments. Examination of the expression of miRNA-127 and miRNA-129 show no signals

(Fig. 9) compared to clear expression predicted from the MIRChip experiments. These results are in agreement with the Northern blot data of Sempere et al. (2004).

Example 7 Clustering Analysis

[0072] Clustering analysis was performed on 150 of the miRNAs. For each miRNA, the background signal of 500 was first subtracted from the values observed in all 6 different tissues. A threshold of 30 was set as a minimal value. A log₂ transformation was applied, and the Euclidian distance matrix was calculated. A hierarchical clustering using Average Linkage algorithm was performed with an output of a dendrogram. A distance threshold of 6 was used to distinguish between the most significant clusters.

[0073] Clustering analysis revealed that miRNAs are expressed in almost every conceivable pattern (Fig. 10). This includes miRNAs expressed in all tissues, miRNAs expressed in some tissues, tissues specific miRNAs and miRNAs undetectable in any of the tissues examined. The analysis revealed distinct clusters of miRNAs specifically expressed in brain, liver, and thymus, while clusters of miRNAs that are specifically expressed in testes and placenta are more obscure. A thorough analysis of the testes, and placenta hybridization data revealed miRNAs that are specific, or highly enriched, in these tissues. Fig. 7 shows, in addition to the brain and liver data, the miRNAs that are tissue-specific or highly enriched in the three tissues not examined before by others: miRNA-96, miRNA-182, and miRNA 183 in the thymus, miRNA-10b, miRNA-212, and miRNA-299 in the testes, and miRNA-141, miRNA-200c, and miRNA-320 in the placenta. Some miRNAs were expressed in two of these tissues (e.g. miRNA-197 and miRNA-205) and others were expressed in all three tissues (e.g. miRNA-26a, miRNA-100, and miRNA-222). Interestingly, we observed an overall low expression of miRNAs in HeLa cells. Only 44 of the miRNAs show signal levels above background compared to 86 to 119 in the five tissues. In addition, none of the miRNAs was found to be specifically enriched in HeLa cells and the vast majority of the miRNAs showing significant signals were expressed at lower levels than in the five tissues. These results are compatible with other reports observing lower expression levels of various miRNAs in cancer cells (Calin et al. 2004, Michael et al. 2003).

Example 8

Sequence-Directed miRNA Cloning

7. Capture Sequences

[0074] Predicted miRNAs were cloned by the MIRAclone method using biotin-labeled capture oligonucleotides which are in reverse-complementary orientation to the library of target molecules. A schematic illustration of the MIRAclone method is presented in Fig. 8. The biotin moiety was added to the 5' end of the capture sequence.

8. Single-Stranded Library

[0075] To construct a library of enriched miRNAs, endogenous 18 to 24 nucleotide RNAs were size-fractionated from total RNA of human placenta tissue. The cDNA preparation procedure was similar to that of Example 2. The 5'- and 3'- adaptors shown below were ligated to the size-fractionated RNA.

Table 4

5' Adapter	
5'-AACTGCAGAAAGGAGGAGCTCTAGata-3'	SEQ ID NO: 11
3' Adapter	
5'-P-uggAACAGATGAATTCTACC-idT-3'	SEQ ID NO: 12

[0076] Reverse-transcription was then performed using 3µg of the adapter-ligated RNA. PCR amplification was then performed using the following primers using an excess of the reverse primer (1:50 ratio) 5'-TAATACGACTCACTATAGGTAGAATTTCATCTGTTCCA-3' (SEQ ID NO: 13). Alternatively, the cRNA was produced by PCR using the same forward primer and a modified reverse primer (5'-ACTGGTGCCTAATACGACTCACTATAGGTAGAAT-3') (SEQ ID NO: 14) that contained a T7 promoter. This served as a template for *in-vitro* transcription with T7 RNA polymerase.

9. Hybridization

[0077] Hybridization was conducted using 5µl of the single-stranded PCR products and ~0.5µg capture oligonucleotide added to 200µl TEN buffer (10mM tris ph=8.0; 1mM EDTA; 100mM NaCl). Following hybridization, µMACS Streptavidin Microbeads were added and incubated for 2 minutes at the hybridization temperature. The mixture was then loaded onto a magnetized

μ MACS Streptavidin Kit columns (130-074-101; Miltenyi Biotec, Gladbach, Germany) and processed according to the manufacturer instructions. The hybridized single-stranded library molecules were eluted by adding 150 μ l H₂O pre-heated to 80°C.

10. Sequencing

[0078] The recovered single-stranded cDNA library molecules were amplified by PCR using primers for the adaptor sequences. When cRNA was used the PCR was preceded by an RT reaction.

11. Cloning

[0079] The recovered single-stranded cDNA library molecules were amplified by PCR using primers for the adaptor sequences. When cRNA was used the PCR was preceded by an RT reaction. PCR products were ligated into a pTZ57/T vector (#k1214, MBI Fermentas, Hanover, MD, USA). The presence of the candidate miRNAs in the ligation products was confirmed by PCR using a primer specific for the candidate miRNAs and a primer located on the 5' region (FV-primer-5'-CTTCGCTATTACGCCAGCTG-3') or to the 3' region (RV-primer-5'-GTTAGCTCACTCATTAGGCACC 3') of the multiple cloning site of the vector. Positive ligations were transformed into competent JM109 E. coli (L2001, Promega) and plated onto LB-Ampicillin plates with IPTG and Xgal. White and light blue colonies were transferred to duplicate grid-plates, one of which was blotted onto a membrane (Biodyne Plus, Pall) for hybridization with DIG tailed oligonucleotide probes complementary to the expected miRNAs according to manufacturer's instructions (Roche). Positive clones were examined by colony PCR with a miRNAs-specific primer and vector primers as described above. The positive clones were amplified with two external primers on the vector (FV and RV primers). Plasmid DNA from positive colonies was sequenced with a nested primer (5' GATGTGCTGCAAGGCGATTAAG 3').

Example 9

MIRAClone Detection of *mir-21*

[0080] Initially, we tested the MIRAClone method described in Example 8 on human *mir-21*, which is highly expressed in several tissues (Barad O, 2004). Amplified and non-amplified cDNA resulted in efficient recovery of *mir-21*. Importantly, no background was observed in the controls, including amplification using the capture oligo itself as template, indicating that the

PCR products were derived from the captured and recovered library molecules. Following ligation of the amplified recovered material into the cloning vector we conducted a quality control PCR with *mir-21* primer and vector primers to ensure that *mir-21* sequences were ligated. As shown in Fig. 9B, *mir-21* was present in the ligated vector. Next, the ligation products were transformed into bacteria and positive clones were identified by colony filter hybridization with a *mir-21* specific oligonucleotide (Fig. 9C). PCR on colonies, using a *mir-21* specific primer and a primer flanking the multiple cloning site in the plasmid, revealed that these clones carried *mir-21* sequences (Fig. 9D). Finally, sequencing analysis revealed authentic *mir-21* sequences in these clones.

Example 10

Analysis of MIRAclone Sensitivity

[0081] To examine the sensitivity of the MIRAclone method, we tested additional published miRNAs that are expressed at varying levels. As shown in Fig. 10 we were able to obtain clones containing the authentic sequence of all miRNAs that were expressed above background levels in a microarray analysis of placenta-derived miRNAs. This included miRNAs expressed at levels just above the threshold of significant microarray values. Interestingly, *MIR-23a* and *MIR-23b*, which differ in only one nucleotide, were each specifically cloned by the respective capture oligo. Thus, the MIRAclone method is highly specific and sensitive, allowing the cloning of miRNAs expressed at relatively low levels.

[0082] Several of the published miRNAs were predicted solely by homology to miRNAs in other species, but were never cloned in humans. These include *mir-23b*, *mir-34b*, *mir-135b*, *mir-154*, and *mir-203* (Fig. 10). Thus, the MIRAclone method was able to provide the first evidence for their expression in human cells. In addition, for *mir-21*, *mir-34b*, *mir-96*, *mir-135b*, and *mir-203* we found variations from the published sequences (Fig. 10). While for the first four miRNAs the variations were in the 3' end, which is commonly observed among many miRNAs, the cloned *mir-203* sequence lacked the first 5' "G" found in the predicted sequence.

[0083] An additional feature we have examined is the flexibility and sensitivity of the method when longer capturing oligonucleotides are used. This is important since prediction algorithms cannot always predict the precise location of the mature miRNA within the hairpin precursor. We created variations of the *mir-21* capture oligonucleotides with 8 additional nucleotides from

the precursor sequence on the 5' side of *mir-21* (5' 8nt), 8 additional nucleotides from the precursor sequence on the 3' side of *mir-21* (3' 8nt), and 4 nucleotides on the 5' and 4 nucleotides on the 3' side of *mir-21* (5' 4nt 3' 4nt) (Fig. 11). In addition, the sequence of the capture oligonucleotide was shifted by 2 nucleotides, leaving only 20 nucleotides that match the mature miRNA (Fig. 11). We found that the cloning efficiency of all three longer capture oligo was comparable to that of the parental capture oligo. In addition, the reduction of match-length of *mir-21* capture oligo by two nucleotides did not change the cloning efficiency.

Example 11

Cloning of Computationally Predicted MicroRNAs.

[0084] We selected for cloning 55 miRNAs predicted using methods similar to those described in U.S. Patent Application Nos. 60/522,459, 10/709,577 and 10/709,572. The design of the 26-30 nucleotide long capture oligonucleotides was based on the prediction of miRNA location within the predicted hairpin precursors. 2-4 nucleotides were added on each side of the 22-mer predicted miRNA. As described above, these additional nucleotides do not impede miRNA detection. The capture oligo designed for *mir-RG-2* is identical to both *mir-RG-2-1* and *mir-RG-2-2* precursors as these two precursors share an identical 3' stem (Fig. 12). Interestingly, due to differences in the 5' stem sequence the miRNAs predicted from these two precursors vary slightly in position along the 3' stem. In addition, it should be noted that these two precursors are found within 700bp of each other on chromosome 19.

[0085] Using the designed capture oligonucleotides we successfully cloned 45 novel miRNAs, including all five predicted miRNAs shown on Fig. 12. Fig. 13 presents the full set of cloned miRNAs. The 45 novel miRNAs were derived from 35 capture oligonucleotides as clones derived from some of the capture oligonucleotides resulted in the cloning of two or more miRNAs of related sequence. The expression of the five miRNAs shown in Fig. 12 was further validated by Northern blot analysis, which showed three of the miRNAs (*mir-RG-27*, *mir-RG-21*, and *mir-RG-2*) are expressed predominantly in placenta, *mir-RG-24* is expressed both in placenta and brain, and *mir-RG-3* is expressed in the brain, liver and thymus.

[0086] Many of the cloned miRNAs have shown 3' sequence length heterogeneity as found in previous cloning studies. However, for four miRNAs we have also found heterogeneity in the 5' end. Two miRNAs, *mir-RG-31* and *mir-RG-36*, can also be regarded as the same miRNA

showing 5' sequence length heterogeneity; however, *mir-RG-31* is uniquely encoded by another hairpin precursor and thus regarded as a unique miRNA. Two of the cases showing apparent 3' sequence heterogeneity may actually be interpreted as two different miRNAs processed from two different precursors. Thus, the 22 nucleotide long *mir-RG-18* can be processed from two different precursors while *mir-RG-35*, which have the same 22 nucleotides but the later is 2 nucleotides longer, is encoded by only one of these palindromes. Similarly, *mir-RG-31* and *mir-RG-36*, which differ in both 3' and 5' sequences, share the same precursor though *mir-RG-31* is found also on another precursor.

[0087] In three cases we have found two miRNAs encoded by the different stems of the precursor. This includes *mir-RG-9* and *mir-RG-37*, *mir-RG-10* and *mir-RG-40*, and *mir-RG-15* and *mir-RG-28*. Thus, one the miRNAs in these pairs can be regarded as miRNAs*. In four cases we have found a miRNAs that matches two precursors. This includes the *mir-RG-9* and *mir-RG-37* pair that are encoded by two identical precursors, as well as *mir-RG-2*, *mir-RG-4*, and *mir-RG-14*.

[0088] For some of the cloned miRNAs a match between the predicted and cloned sequences was observed, exemplified by *mir-RG-27* and *mir-RG-2-2* in Fig. 12. For other miRNAs a minimal difference of 1 nucleotide in the 5' (e.g. *mir-RG-21* and *mir-RG-3* in Fig. 12) or both 5' and 3' (e.g. *mir-RG-24* in Fig. 12) was observed. Thus, in general, our miRNA-prediction algorithm seems accurate.

CLAIMS

1. A method of detecting a miRNA comprising
 - (a) providing an array comprising a solid substrate and a plurality of positionally distinguishable polynucleotides attached to the solid substrate, wherein each polynucleotide comprises a miRNA;
 - (b) contacting the array of (a) with a plurality of target polynucleotides comprising a complement of a miRNA under conditions permitting hybridization; and
 - (c) detecting hybridization of a target sequence to the miRNA of (a),wherein a miRNA is detected when the hybridization of (c) is above background.
2. The method of claim 1 wherein the plurality of target polynucleotides is produced by a method comprising:
 - (a) providing RNA comprising a plurality of miRNA;
 - (b) ligating adapters to the 5' and 3' ends of the miRNA;
 - (c) preparing first strand cDNA of the 5'-adapter-miRNA-adapter-3';
 - (d) amplifying the adapter-miRNA-adapter; and
 - (e) preparing cRNA using a promoter complementary to the 3' adapter.
3. The method of claim 2 wherein the RNA is less than 160 nucleotides in length.
4. The method of claim 2 wherein the adapters are DNA-RNA hybrids.
5. The method of claim 2 wherein the adapters comprise a restriction site.
6. The method of claim 5 wherein either prior to or after amplification, the adapters are removed by restriction digestion.
7. A method of detecting a miRNA, comprising:
 - (a) providing a plurality of target polynucleotides comprising a miRNA;
 - (b) providing a labeled oligo that is complementary to a portion of the target nucleotides;
 - (c) providing a substrate comprising a capture oligonucleotide comprising at least 16 nucleotides of a miRNA complementary sequence;
 - (d) contacting the target nucleotides, labeled oligo and substrate; and

- (e) detecting hybridization of the target nucleotides, labeled oligo and substrate.

8. A method of isolating a miRNA comprising:

- (a) providing a solid substrate comprising a capture oligonucleotide comprising at least 16 nucleotides of a miRNA sequence;
- (b) contacting the capture oligonucleotide with a plurality of target polynucleotides comprising a complement of a miRNA under conditions permitting hybridization; and
- (c) eluting the target polynucleotides from the capture oligonucleotide.

9. The method of sequencing an miRNA, comprising sequencing a miRNA isolated by the method of claim 8;

10. The method of cloning an miRNA, comprising cloning a miRNA isolated by the method of claim 8;

FIGURE 1

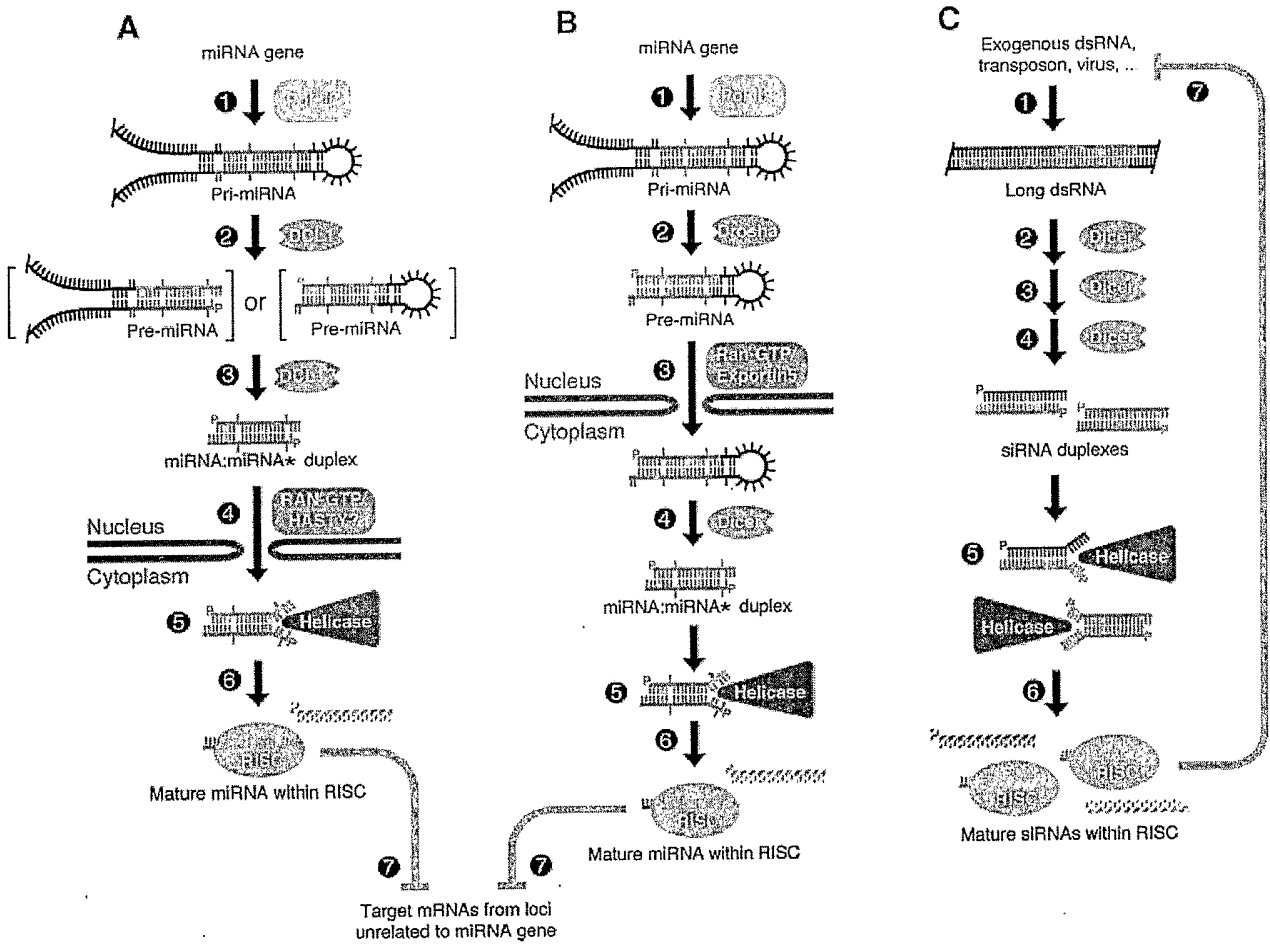


FIGURE 2

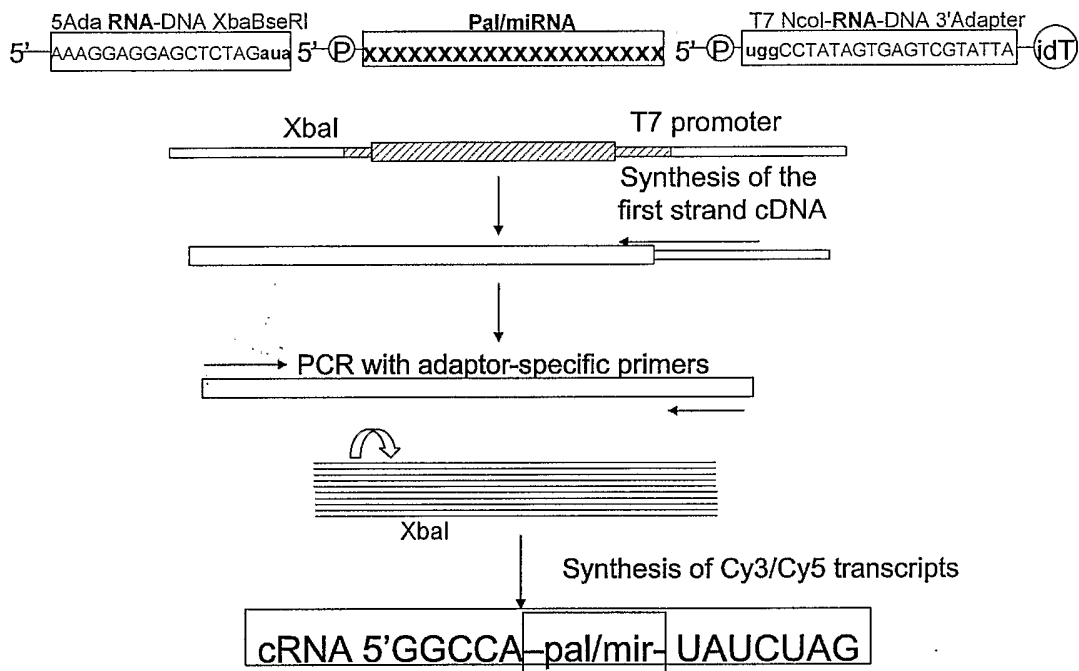
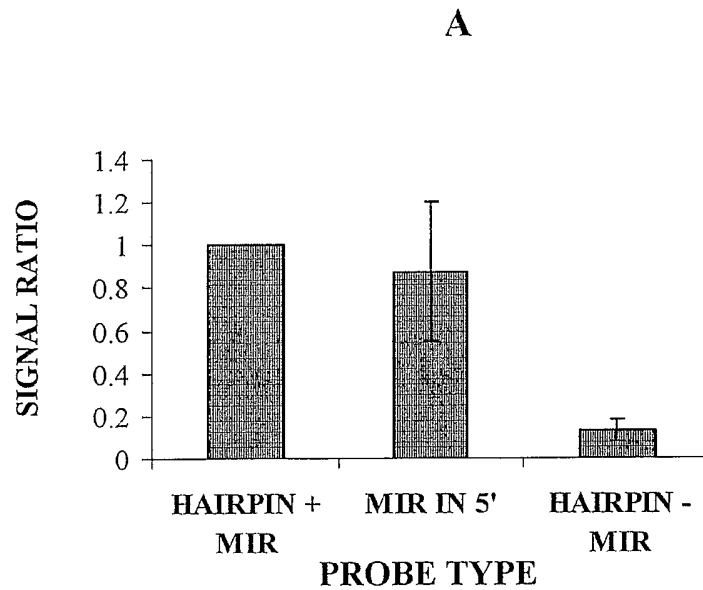


FIGURE 3



B

PROBE	SIGNAL	SEQUENCE
WT	65517	1 GTCCCTGAGACCCTAACTTGTGATGTTTACC GTTTAAATCCACGGGTTAGGCTCTTGGGA 60
MM1	65517ACCCCT
MM2	65517	CAG.....CT
MM3	65517	.A.....A.....T.....GA.....
MM4	65517CAA.....
MM5	65517ATGGC.....
MM6	65517CAATCCGAGAACCCT
MM7	65517CGAGAACCCT
MM8	64744ACT.....CCT
MM9	43431A.....G.....A.....
MM10	21450	CAG.....A.....G.....A.....G.....C.....
MM11	2579	CAGGG.....
MM12	2497	CAGGG.....ACCCCT
MM13	1557A.....A.....A.....A.....T.....C.....A.....ACC..
MM14	905	CAGGGACT.....GAACCCT
MM15	756CT.....C.....CCT
MM16	637A.....GG.....AC.C.CAAA.GG.....A.T.CC.A.T.....AA.....
MM17	565	CAGGGACTCT.....
MM18	564	CAGGGACTCTGGGAT.....
MM19	564	C...GA...G.G.T...CT...A.....T...G...C.T...A.A....
MM20	544	CAGGGACTCT.....CGAGAACCCT

4/17
FIGURE 4

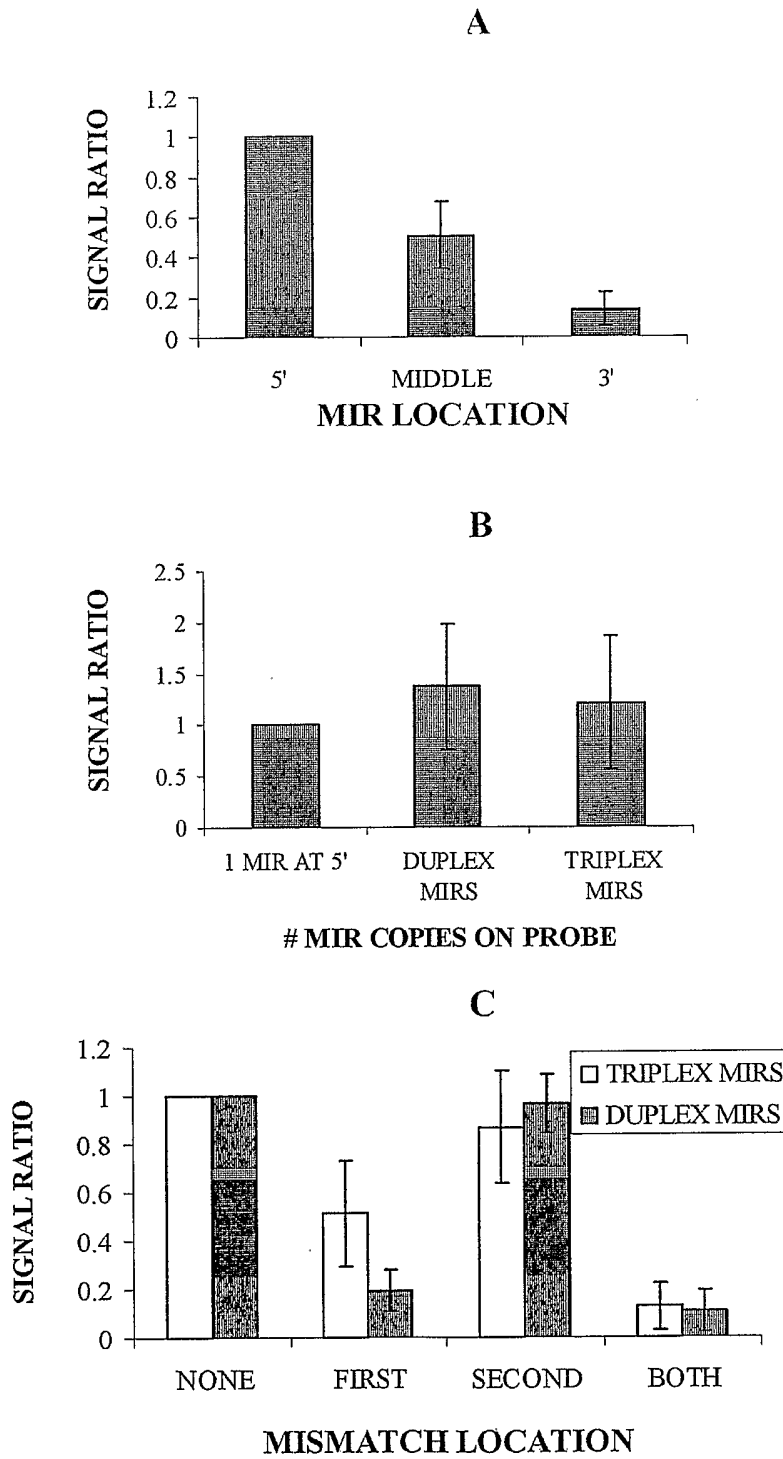


FIGURE 5

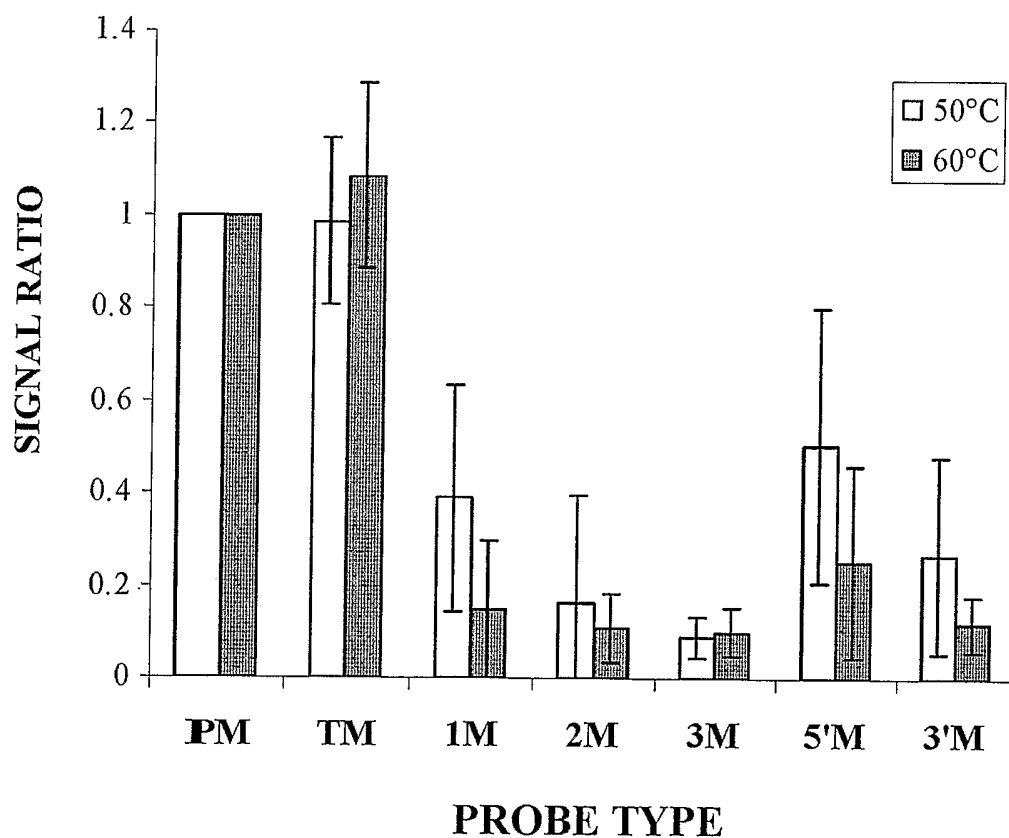


FIGURE 6A

DATA ON THE EXPRESSION OF THE 150 HUMAN MIRS IN FIVE TISSUES
AND HELA CELLS

HUMAN MIR NUMBER	HELA	BRAIN	LIVER	THYMUS	TESTES	PLACENTA
HSA-LET-7A	3666	24544	18500	63972	65518	43407
HSA-LET-7B	7578	42483	38260	65518	65518	34929
HSA-LET-7C	1886	9698	16448	43867	65518	6278
HSA-LET-7D	737	3091	2419	10946	11011	3360
HSA-LET-7E	557	2723	1367	4805	10147	4293
HSA-LET-7F	439	1294	966	2546	3625	1119
HSA-LET-7G	459	2631	2147	10296	10890	3801
HSA-LET-7I	3088	6079	3407	19230	24832	14161
HSA-MIR-1	397	410	387	381	390	386
HSA-MIR-7	459	620	588	1038	712	576
HSA-MIR-9	642	42659	3504	4455	4485	2313
HSA-MIR-9*	399	464	438	418	420	415
HSA-MIR-10A	408	509	1403	4983	3600	2024
HSA-MIR-10B	410	433	477	3871	23083	738
HSA-MIR-15A	1659	4508	11846	31759	29151	46424
HSA-MIR-15B	3650	2829	6592	9089	12885	11833
HSA-MIR-16	1154	4959	10718	45909	38217	61713
HSA-MIR-18	819	1033	1488	1756	2033	3987
HSA-MIR-19A	412	516	459	492	536	554
HSA-MIR-19B	434	1034	890	1430	1973	2556
HSA-MIR-20	1309	3496	6541	11775	9270	19044
HSA-MIR-21	27260	65517	65518	65518	65518	65518
HSA-MIR-22	1133	7491	58986	64041	62713	65518
HSA-MIR-23A	1312	3492	2990	6021	11173	40076
HSA-MIR-23B	584	4790	9103	3629	9914	9797
HSA-MIR-24	7778	20139	65518	65518	65518	65518
HSA-MIR-25	2389	18893	30895	59486	52717	64858
HSA-MIR-26A	603	1760	4508	41382	52010	28504
HSA-MIR-26B	417	471	514	1209	1430	1373
HSA-MIR-27A	12384	16443	33247	65518	65518	65518
HSA-MIR-27B	12259	26025	44791	62789	65518	65518
HSA-MIR-28	1023	7811	20014	43302	44562	30652
HSA-MIR-29A	704	6964	6195	40476	62304	20813
HSA-MIR-29B	510	1423	1429	4957	8244	8405
HSA-MIR-29C	462	958	1139	3459	4770	2448
HSA-MIR-30A	799	8548	7085	12426	26213	42248
HSA-MIR-30A*	6848	15361	34360	55512	65518	65518
HSA-MIR-30B	2070	8641	41938	65518	65518	65518
HSA-MIR-30C	2668	34655	63926	65518	65518	65518
HSA-MIR-30D	5159	14236	44003	64387	65518	65518

FIGURE 6B

DATA ON THE EXPRESSION OF THE 150 HUMAN MIRS IN FIVE TISSUES
AND HELA CELLS

HUMAN MIR NUMBER	HELA	BRAIN	LIVER	THYMUS	TESTES	PLACENTA
HSA-MIR-30E	598	1574	5154	6504	10392	15363
HSA-MIR-31	30128	22000	8327	14142	26517	26850
HSA-MIR-32	430	472	651	675	763	1566
HSA-MIR-33	396	591	548	538	1344	834
HSA-MIR-34A	3614	7312	12914	26645	65499	34344
HSA-MIR-34B	403	452	452	591	1681	1225
HSA-MIR-34C	418	497	465	731	2524	2397
HSA-MIR-92	2091	7924	30214	53596	65518	48819
HSA-MIR-93	4137	3745	9461	39473	31972	62807
HSA-MIR-95	392	1042	512	634	620	446
HSA-MIR-96	887	3100	1477	44800	2266	5466
HSA-MIR-98	549	605	958	1140	1111	838
HSA-MIR-99A	511	4540	21812	65518	65518	14807
HSA-MIR-99B	1101	7606	5216	19528	46932	65518
HSA-MIR-100	742	2633	4836	40001	50900	61755
HSA-MIR-101	403	428	549	501	662	544
HSA-MIR-103	793	8589	8801	14488	20325	30685
HSA-MIR-105	627	1301	760	1190	1557	1858
HSA-MIR-106A	2524	5809	15861	35942	26829	55540
HSA-MIR-106B	530	558	896	1544	1492	2537
HSA-MIR-107	573	4189	4029	6808	9224	11716
HSA-MIR-122A	1051	447	65518	2644	617	570
HSA-MIR-124A	1879	65517	7025	3099	2672	2498
HSA-MIR-125A	2435	54390	27187	65518	65518	65518
HSA-MIR-125B	8694	65517	65518	65518	65518	65518
HSA-MIR-126	547	1569	6498	36867	21885	29777
HSA-MIR-126*	414	682	1485	2531	3654	3487
HSA-MIR-127	758	9792	3199	10965	54584	65037
HSA-MIR-128A	2015	27701	4940	4876	5166	2495
HSA-MIR-128B	1168	21969	3954	4819	5383	2027
HSA-MIR-129	503	22573	1175	2213	5364	2017
HSA-MIR-130A	432	489	622	1151	1479	2930
HSA-MIR-130B	392	458	450	430	453	498
HSA-MIR-132	453	1684	721	1982	3453	833
HSA-MIR-133A	1089	1992	2121	20533	19667	15402
HSA-MIR-134	448	617	698	763	2250	997
HSA-MIR-135	967	1645	1828	2765	10880	7966
HSA-MIR-136	465	725	709	776	3100	8840
HSA-MIR-137	395	453	507	414	384	421
HSA-MIR-138	496	1102	634	1011	1024	1069

FIGURE 6C

DATA ON THE EXPRESSION OF THE 150 HUMAN MIRs IN FIVE TISSUES AND HELA CELLS

HUMAN MIR NUMBER	HELA	BRAIN	LIVER	THYMUS	TESTES	PLACENTA
HSA-MIR-139	472	12415	7186	25388	16439	5477
HSA-MIR-140	631	7552	7366	40601	55476	35013
HSA-MIR-141	696	805	1220	4063	2000	46845
HSA-MIR-142-3P	660	1797	16460	65518	33441	62213
HSA-MIR-142-5P	419	437	493	663	626	550
HSA-MIR-143	661	1289	2057	13887	19822	20344
HSA-MIR-144	410	445	449	480	668	1345
HSA-MIR-145	1419	6734	25815	65518	65518	65518
HSA-MIR-146	386	445	489	1421	658	477
HSA-MIR-147	399	436	464	526	530	499
HSA-MIR-148	413	620	38436	5250	6204	2711
HSA-MIR-149	4128	65517	20784	65518	65518	65518
HSA-MIR-150	648	1463	5295	65518	29728	5280
HSA-MIR-152	481	805	3001	6681	7286	5338
HSA-MIR-153	389	453	395	394	413	390
HSA-MIR-154	438	733	1914	3309	14750	9637
HSA-MIR-155	432	512	719	2640	1688	560
HSA-MIR-181A	966	16159	5462	42154	14183	64827
HSA-MIR-181B	660	8091	1414	14894	6793	18934
HSA-MIR-181C	405	789	501	1188	777	1673
HSA-MIR-182	662	1944	1091	25771	2034	3683
HSA-MIR-182*	442	529	665	939	1290	864
HSA-MIR-183	1026	1123	1286	8754	1681	2138
HSA-MIR-184	401	408	410	386	489	405
HSA-MIR-185	524	4121	2956	4920	5118	10982
HSA-MIR-186	485	663	1227	2857	3684	2148
HSA-MIR-187	510	752	851	962	3641	827
HSA-MIR-188	440	467	730	1343	2891	4532
HSA-MIR-189	8152	42303	29353	65518	65518	44055
HSA-MIR-190	385	434	394	395	395	389
HSA-MIR-191	3161	11542	14980	36754	43251	45574
HSA-MIR-192	452	606	20650	1628	1263	2607
HSA-MIR-193	534	623	1772	1686	2941	2614
HSA-MIR-194	501	910	65518	4737	2342	7952
HSA-MIR-195	418	492	700	1248	2198	625
HSA-MIR-196	1193	3096	1124	1554	2258	1845
HSA-MIR-197	533	3125	1777	10637	23508	3896
HSA-MIR-198	434	447	514	1320	813	644
HSA-MIR-199A	3556	2438	65518	65518	65518	65518
HSA-MIR-199A*	477	589	3835	22042	61830	25657

FIGURE 6D

DATA ON THE EXPRESSION OF THE 150 HUMAN MIRs IN FIVE TISSUES
AND HELA CELLS

HUMAN MIR NUMBER	HELA	BRAIN	LIVER	THYMUS	TESTES	PLACENTA
HSA-MIR-199B	1109	1197	9791	64693	65518	65518
HSA-MIR-200A	415	429	547	2708	540	675
HSA-MIR-200B	452	490	792	2176	1045	855
HSA-MIR-200C	556	595	1027	10636	1478	33532
HSA-MIR-203	385	419	403	417	411	422
HSA-MIR-204	525	3898	1757	6535	64859	6233
HSA-MIR-205	551	615	1646	65518	2645	39072
HSA-MIR-206	400	437	805	743	624	612
HSA-MIR-208	406	419	386	383	388	384
HSA-MIR-210	844	2280	10703	6864	15288	62452
HSA-MIR-211	712	1249	2738	5031	9841	5053
HSA-MIR-212	608	1410	1414	3849	6425	1478
HSA-MIR-213	405	956	563	1150	771	2997
HSA-MIR-214	638	861	5322	40479	65518	17601
HSA-MIR-215	393	426	580	401	409	407
HSA-MIR-216	435	485	522	603	776	619
HSA-MIR-217	381	412	400	390	397	394
HSA-MIR-218	572	6938	1502	12307	7182	12293
HSA-MIR-219	392	483	452	470	564	472
HSA-MIR-220	409	445	643	727	758	657
HSA-MIR-221	625	9325	3520	20212	10608	54287
HSA-MIR-222	968	4740	5858	54036	25630	59213
HSA-MIR-223	450	786	4762	15344	12702	14884
HSA-MIR-224	3233	11061	7684	32305	5377	65518
HSA-MIR-296	1613	833	2305	1879	2088	2075
HSA-MIR-299	411	776	742	941	5921	3684
HSA-MIR-301	396	461	442	421	468	943
HSA-MIR-302	408	451	499	510	582	835
HSA-MIR-320	928	2048	5287	10814	16586	35478
HSA-MIR-321	8374	65431	65518	65518	65518	65518

FIGURE 7

EXPRESSION OF TISSUE-SPECIFIC OR HIGHLY ENRICHED MIRs IN THE FIVE HUMAN TISSUES

	HELA	BRAIN	LIVER	THYMUS	TESTES	PLACENTA
HSA-MIR-9	642	^A 42659	3504	4455	4485	2313
HSA-MIR-124A	1879	65517	7025	3099	2672	2498
HSA-MIR-128A	2015	27701	4940	4876	5166	2495
HSA-MIR-128B	1168	21969	3954	4819	5383	2027
HSA-MIR-129	503	22573	1175	2213	5364	2017
HSA-MIR-194	501	910	65518	4737	2342	7952
HSA-MIR-122A	1051	447	65518	2644	617	570
HSA-MIR-148	413	620	38436	5250	6204	2711
HSA-MIR-192	452	606	20650	1628	1263	2607
HSA-MIR-96	887	3100	1477	44800	2266	5466
HSA-MIR-182	662	1944	1091	25771	2034	3683
HSA-MIR-183	1026	1123	1286	8754	1681	2138
HSA-MIR-200A	415	429	547	2708	540	675
HSA-MIR-10B	410	433	477	3871	23083	738
HSA-LET-7E	557	2723	1367	4805	10147	4293
HSA-MIR-212	608	1410	1414	3849	6425	1478
HSA-MIR-187	510	752	851	962	3641	827
HSA-MIR-134	448	617	698	763	2250	997
HSA-MIR-141	696	805	1220	4063	2000	46845
HSA-MIR-23A	1312	3492	2990	6021	11173	40076
HSA-MIR-136	465	725	709	776	3100	8840

^A SIGNAL VALUES ARE IN BOLD IN THE TISSUE IN WHICH THE MIR WAS FOUND TO BE EXPRESSED AT THE HIGHEST LEVEL.

FIGURE 8

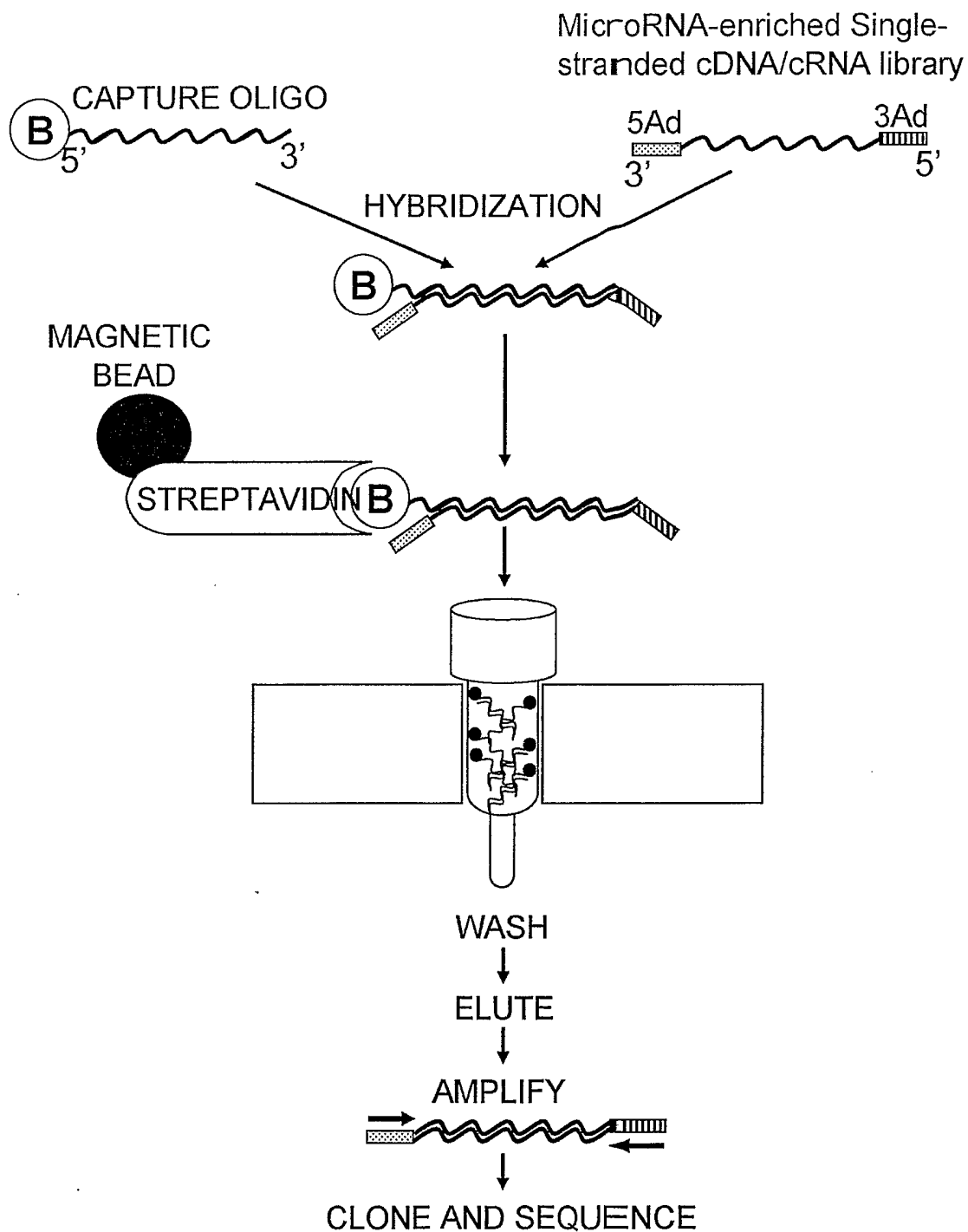


FIGURE 9

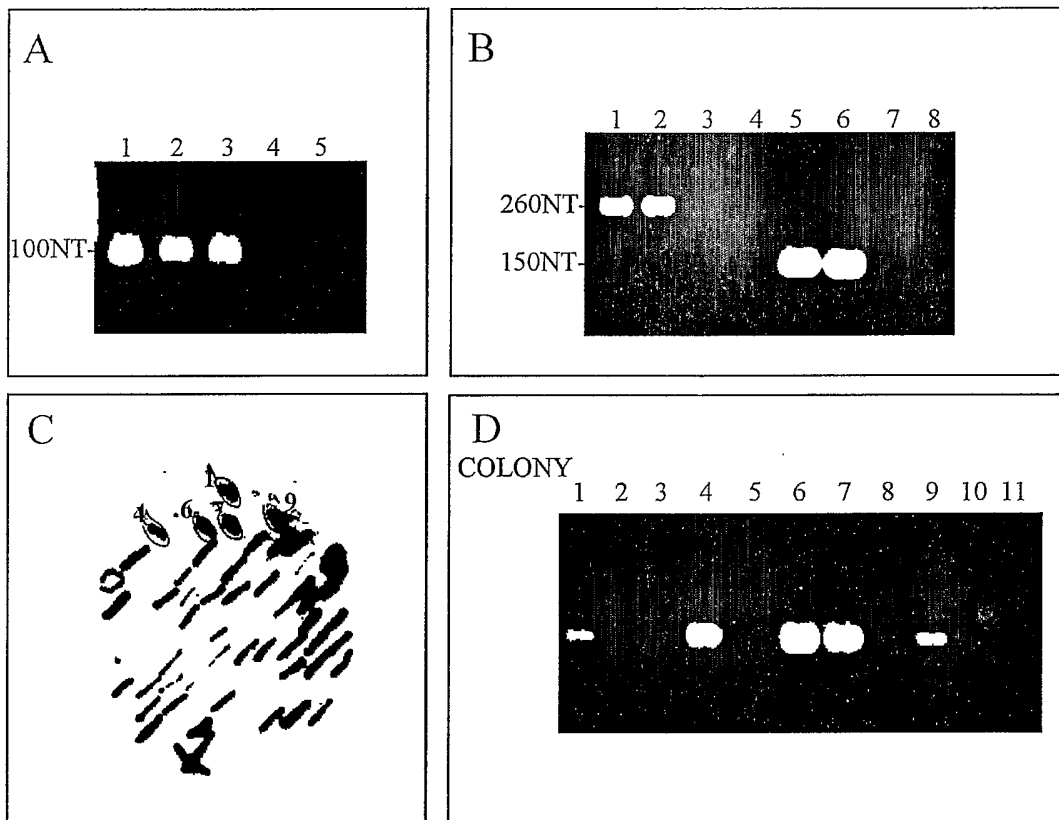


FIGURE 10

MIR NAME	SIGNAL ^A	STATUS ^B	SEQUENCE	SIZE	REF
HSA-MIR-21	65517	CLONED	TAGCTTATCAGACTGATGTTGAC	21-22	1
HSA-MIR-23A	65517	CLONED	ATCACATTCGCCAGGGATTTTC	20	1
<u>HSA-MIR-23B</u>	<u>65517</u>	PREDICTED	ATCACATTCGCCAGGGATTACCA	21-22	2
HSA-MIR-135B	14962	PREDICTED	TATGGCTTTTCATTCCTATGTGA	22-23	3
HSA-MIR-154	13926	PREDICTED	TAGGTTATCCGTTGTTGCCCTTCG	22	2
HSA-MIR-96	9073	CLONED	TTTGGCAGCTAGCACATTTTTCCT	23	4
HSA-MIR-29C	3537	CLONED	TAGCACCATTTGAAATCCGTTA	21-22	5
HSA-MIR-203	2257	PREDICTED	-TGAATGTTTAGGACCACCTAG	21	6
HSA-MIR-135A	1581	CLONED	TATGGCTTTTATTCCTATGTGA	23	7
HSA-MIR-34B	1282	PREDICTED	AGGCAGTGTCAATTAGCTGATTTGT	23	8
HSA-MIR-137	621	PREDICTED	NONE		2
HSA-MIR-124A	643	CLONED	NONE		5

THE UNDERLINE NUCL EOTIDES DENOTE NUCL EOTIDES NOT FOUND IN THE SANGER DATABASE SEQUENCES. THE “ ” IN THE BEGINNING OF THE HSA-MIR-203 SEQUENCE STANDS FOR A MISSING “G” FOUND IN THE PREDICTED SEQUENCE IN THE SANGER DATABASE.

A FLUORESCENT SIGNAL FROM MICROARRAY ANALYSIS.

B THE STATUS IS TAKEN FROM THE SANGER MICRORNA REPOSITORY DATABASE. REFERENCES:

1 LAGOS-QUINTANA M, ET AL., SCIENCE 2001;294:853-858.

FIGURE 11

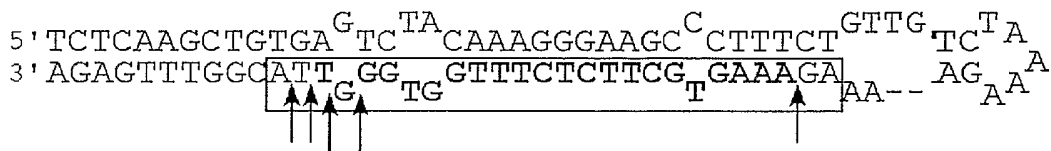
```

miR-21
5' 8NT 1 -----TAGCTTATCAGACTGATGTTGA----- 22
3' 8NT 1 TTGTCGGGTAGCCTTATCAGACTGATGTTGA----- 30
5' 4NT 1 -----TAGCCTTATCAGACTGATGTTGACTGTGAA 30
3' 4NT 1 -----CGGGTAGCCTTATCAGACTGATGTTGACTGT----- 30
5' 2NT 1 -----GGTAGCCTTATCAGACTGATGTT----- 22

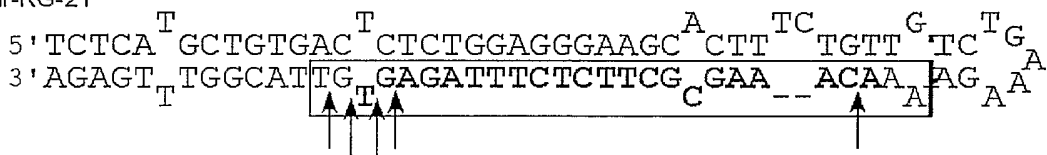
```

FIGURE 12

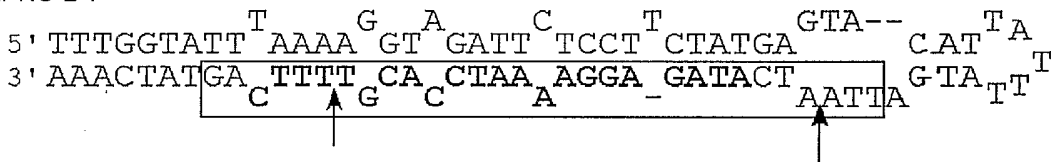
mir-RG-27



mir-RG-21



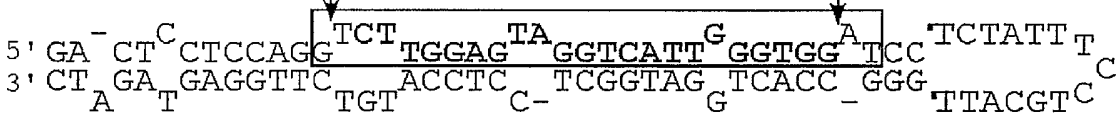
mir-RG-2-1



mir-RG-2-2



mir-RG-24



mir-RG-3

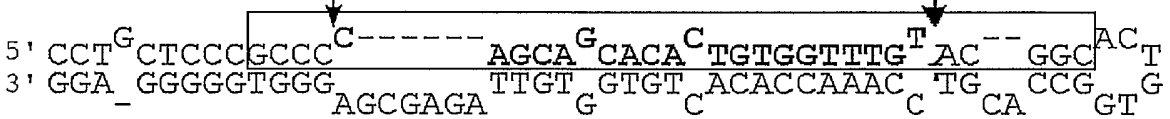


FIGURE 13A

NCBI_NAME	MIR_SEQ	LENGTH	CF-IR_ID
HSA-MIR-RG-2	ATCATAGAGGAAAATCCACGT	21	14
HSA-MIR-RG-3	CAGCAGCACACTGTGGTTTGT	21	17
HSA-MIR-RG-4	AAAGCGCTTCCCTTTGCTGGA	21	19
HSA-MIR-RG-5	GAAAGCGCTTCTCTTTAGA	19	19
HSA-MIR-RG-5	*AAAGCGCTTCTCTTTAGAGGA	21	19
HSA-MIR-RG-6	CAAAGCGCTTCCCTTTGGAGC	21	19
HSA-MIR-RG-7	GAAGGCGCTTCCCTTTAGAGC	21	19
HSA-MIR-RG-8	GAAGGCGCTTCCCTTTGGAGT	21	19
HSA-MIR-RG-9	GAGTGCCTTCTTTGGAGCGT	21	19
HSA-MIR-RG-10	TCAGTTCATCTGCAAAGAAG	21	X
HSA-MIR-RG-11	AAACAAACATGGTGCCTTCTT	22	14
HSA-MIR-RG-12	TATAACAAGGGCAAGCTCTCTGT	22	14
HSA-MIR-RG-13	AAAGCGCTTCCCTTCAGAGTGT	21-22	19
HSA-MIR-RG-14	AAAGTGCATCCTTTTAGAGTGT	22	19
HSA-MIR-RG-15	AAAGTGCCTCCTTTTAGAGTGT	22	19
HSA-MIR-RG-16	AAAGTGCTTCCCTTTGGACTGT	22	19
HSA-MIR-RG-17	AAAGTGCTTCCCTTTAGAGGGT	22	19
HSA-MIR-RG-18	ACAAAGTGCTTCCCTTTAGAGT	22	19
HSA-MIR-RG-19	ATCGTGCATCCTTTTAGAGTGT	21-22	19
HSA-MIR-RG-19	***GTGCATCCTTTTAGAGTGT	19	19
HSA-MIR-RG-20	CAAAGCGCTCCCCTTTAGAGGT	22	19
HSA-MIR-RG-21	CAAAGCGCTTCTCTTTAGAGTG	19-22	19
HSA-MIR-RG-22	AATTGCACGGTATCCATCTGTA	20-22	X
HSA-MIR-RG-23	GTCAACACTTGCTGGTTTCCTC	21-22	X
HSA-MIR-RG-24	TCTTGGAGTAGGTCATTGGGTGG	23	14
HSA-MIR-RG-25	TGAAACATACACGGGAAACCTCT	21-23	14
HSA-MIR-RG-26	AAAATGGTTCCCTTTAGAGTGTT	23	19
HSA-MIR-RG-27	AAAGTGCTTCTCTTTGGTGGGT	20-23	19
HSA-MIR-RG-28	ATTCTCCAAAAGGGAGCACTTTC	21-23	19
HSA-MIR-RG-29	CAAAGTGCTCCCTTTAGAGTGT	23	19
HSA-MIR-RG-30	CAAGTGCTTCCCTTTAGAGGGTT	23	19
HSA-MIR-RG-31	CATCTGGAGGTAAGAAGCACTTT	23	19
HSA-MIR-RG-32	GAACGCGCTTCCCTATAGAGGGT	23	19
HSA-MIR-RG-32	*AACGCGCTTCCCTATAGAGG	20	19
HSA-MIR-RG-33	CAAAGTGCTCATAGTGCAGGTAG	23	X
HSA-MIR-RG-34	AACATTCATTGTTGTTCGGTGGGT	24	19
HSA-MIR-RG-35	ACAAAGTGCTTCCCTTTAGAGTGT	22-24	19

FIGURE 13B

NCBI_NAME	MIR_SEQ	LENGTH	CHR_ID
HSA-MIR-RG-36	*ATCTGGAGGTAAGAAGCACTTTCT	23-24	19
HSA-MIR-RG-37	TTCTCCAAAAGAAAGCACTTTCTG	23-24	19
HSA-MIR-RG-38	AATATAACACAGATGGCCTGT	21	14
HSA-MIR-RG-39	AAAGTGCTTCCTTTTTGAGGG	21	19
HSA-MIR-RG-40	AACTGTTTGCAGAGGAAA	18	X
HSA-MIR-RG-40	***TGTTTGCAGAGGAACTGAGAC	22	X
HSA-MIR-RG-41	AATCCTTGGAACCTAGGTGTGAGTGC	22-26	X
HSA-MIR-RG-42	AATCCTTTGTCCCTGGGTGAGA	21-22	X
HSA-MIR-RG-43	AGACCCTGGTCTGCACTCTAT	20-21	X
HSA-MIR-RG-44	TCCCCAGGTGTGATTCTGATT	22	X
HSA-MIR-RG-45	AGCTCGGTCTGAGGCCCTCAGT	23	17
HSA-MIR-RG-46	AATTGCTTGAACCCAGGAAG	20	X