

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
8 September 2006 (08.09.2006)

PCT

(10) International Publication Number
WO 2006/092738 A2

- (51) International Patent Classification: Not classified
- (21) International Application Number:
PCT/IB2006/001400
- (22) International Filing Date: 7 February 2006 (07.02.2006)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:

60/593,696	7 February 2005 (07.02.2005)	US
60/728,161	19 October 2005 (19.10.2005)	US
60/739,522	23 November 2005 (23.11.2005)	US
60/743,098	5 January 2006 (05.01.2006)	US
11/275,628	19 January 2006 (19.01.2006)	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, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, 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.

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BENTWICH, Itzhak** [IL/IL]; 65 Kfar Daniel, 73125 Kfar Daniel (IL).
AVNIEL, Amir [IL/IL]; Ha Rav Herzog 23, Givatayim (IL).
KAROV, Yael [IL/IL]; Melchett 33, Tel Aviv (IL).
AHARONOV, Ranit [IL/IL]; Rabina 1, Tel Aviv (IL).

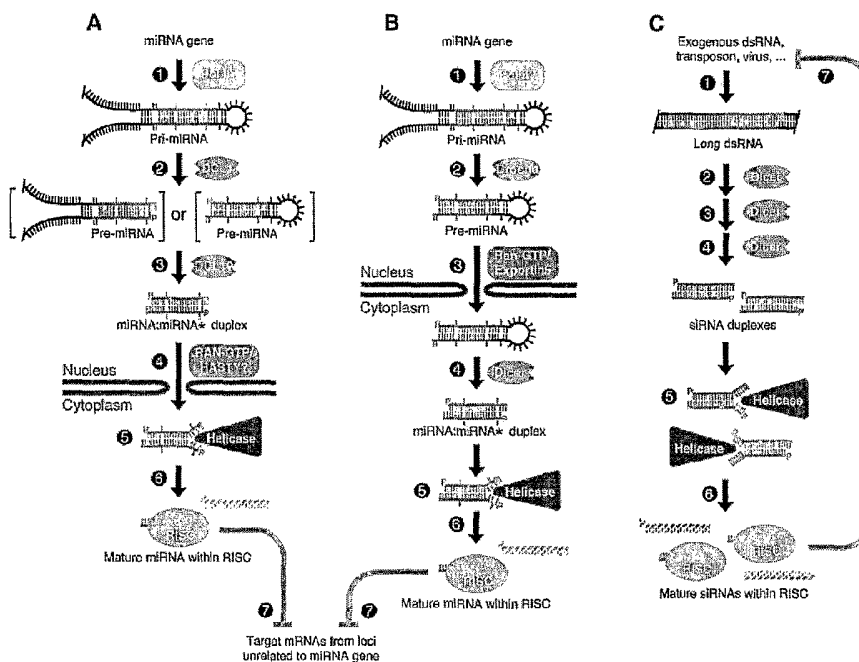
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, 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).

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: MICRORNAS AND RELATED NUCLEIC ACIDS



(57) Abstract: Described herein are novel polynucleotides associated with disease. The polynucleotides are miRNAs, miRNA precursors, and associated nucleic acids. Methods and compositions are described that can be used for diagnosis, prognosis, and treatment of disease. Also described herein are methods that can be used to identify modulators of the disease-associated polynucleotides. Also described herein are methods and compositions for linear amplification and labeling of a targeted nucleic acid. The amplified targeted molecules may be used in hybridization techniques like Luminex and Microarray analysis.

WO 2006/092738 A2

MICRORNAS AND RELATED NUCLEIC ACIDS

FIELD OF THE INVENTION

[0001] The invention relates in general to microRNA molecules as well as various nucleic acid molecules relating thereto or derived therefrom.

CROSS-REFERENCE TO RELATED APPLICATIONS

[0002] The present application claims the benefit of U.S. Provisional Patent Application No. 60/593,696, filed February 7, 2005, U.S. Provisional Patent Application No. 60/728,161, filed October 19, 2005, U.S. Provisional Patent Application No. 60/739,522, filed November 23, 2005, U.S. Provisional Patent Application No. 60/743,098, filed January 5, 2006, and U.S. Patent Application No. 11/275,628, filed January 19, 2006, the contents of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0003] MicroRNAs (miRNAs) are short RNA oligonucleotides of approximately 22 nucleotides that are involved in gene regulation. MicroRNAs 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. More importantly, the role of miRNAs in the development and progression of disease has only recently become appreciated. Deregulated miRNA expression is implicated in onset and progression of different diseases including, but not limited to embryonic malformations and cancers.

[0004] As a result of their small size, miRNAs have been difficult to identify using standard methodologies. A limited number of miRNAs have been identified by extracting large quantities of RNA. MiRNAs have also been identified that contribute to the presentation of visibly discernable phenotypes. Expression array data shows that miRNAs are expressed in different developmental stages or in different tissues. The restriction of miRNAs to certain tissues or at limited developmental stages indicates that the miRNAs identified to date are likely only a small fraction of the total miRNAs.

[0005] Computational approaches have recently been developed to identify the remainder of miRNAs in the genome. Tools such as MiRscan and MiRseeker have identified miRNAs that were later experimentally confirmed. Based on these computational tools, it has been estimated that the human genome contains 200-255 miRNA genes. These estimates are based on an assumption, however, that the miRNAs remaining to be identified will have the same properties as those miRNAs already identified. Based on the fundamental importance of miRNAs in mammalian biology and disease, the art needs to identify unknown miRNAs. The present invention satisfies this need and provides a significant number of miRNAs and uses therefore.

[0006] Moreover, because of their potential broad use in treating and diagnosing different diseases, there is a need in the art (yet unmet) to develop methods of identification, isolation and also quantitation of miRNAs. The present invention addresses the need by disclosing efficient and sensitive methods and compositions for isolating and quantitating miRNAs from different samples, including those wherein there is only minimum amount of a starting material available.

SUMMARY OF THE INVENTION

[0007] An isolated nucleic acid is provided. The nucleic acid may comprise a sequence of any of SEQ ID NOS: 1-42840282, the complement thereof, or a sequence at least 81% identical to 21 contiguous nucleotides thereof. The nucleic acid may also comprise the sequence of any of SEQ ID NOS: 3A-7A, the complement thereof, or a sequence at least 63% identical to 81 contiguous nucleotides thereof. The nucleic acid may also comprise the sequence of SEQ ID NOS: 1A or 2A, the complement thereof, or a sequence at least 63% identical to 81 contiguous nucleotides thereof. The nucleic acid may be from about 51 to about 250 nucleotides in length. The nucleic acid may comprise a modified base.

[0008] A probe comprising the nucleic acid is also provided. A composition comprising the probe is also provided. A biochip comprising the probe is also provided.

[0009] A method for detecting a disease-associated nucleic acid is also provided. A biological sample may be provided from which the level of a nucleic acid may be measured. The nucleic acid may comprise a sequence of any of SEQ ID NOS: 1-42840282 or 3A-7A. The nucleic acid may also comprise a sequence at least about 81% identical to about 21 contiguous nucleotides of any of SEQ ID NOS: 1-42840282 or 3A-7A. A level of the nucleic acid higher than that of a control may be indicative of a disease.

[0010] A method for identifying compound that modulates expression of a disease-associated miRNA is also provided. A cell is provided that is capable of expressing a nucleic acid comprising a sequence of any of SEQ ID NOS: 1-26518 or 3A-7A. A cell may also be provided that is capable of expressing a nucleic acid comprising a sequence at least about 81% identical to about 21 contiguous nucleotides of any of SEQ ID NOS: 1-26518 or 3A-7A. The cell may be contacted with a candidate modulator. The level of expression of the nucleic acid may then be measured. A difference in the level of the nucleic acid compared to a control identifies the compound as a modulator of expression of the miRNA.

[0011] A method of inhibiting expression of a target gene in a cell is also provided. A nucleic acid may be introduced into the cell in an amount sufficient to inhibit expression of the target gene. The target gene may comprise a binding site substantially identical to a binding site referred to in Table 5 or any of SEQ ID NOS: 26519-42840282. The nucleic acid may comprise a sequence of any of SEQ ID NOS: 3A-7A, SEQ ID NOS: 1-26518 or a variant thereof. The nucleic acid may also comprise a sequence at least about 81% identical to about 21 contiguous nucleotides of any of SEQ ID NOS: 3A-7A, SEQ ID NOS: 1-26518 or a variant thereof. Expression of the target gene may be inhibited in vitro or in vivo.

[0012] A method of increasing expression of a target gene in a cell is also provided. A nucleic acid may be introduced into the cell in an amount sufficient to increase expression of the target gene. The target gene may comprise a binding site substantially identical to a binding site referred to in Table 5 or any of SEQ ID NOS: 26519-42840282. The nucleic acid may comprise a sequence substantially complementary to any SEQ ID NOS: 3A-7A, SEQ ID NOS: 1-26518 or a variant thereof. The nucleic acid may also comprise a sequence substantially complementary to a sequence at least about 81% identical to about 21 contiguous nucleotides of any of SEQ ID NOS: 3A-7A, SEQ ID NOS: 1-26518 or a variant thereof. Expression of the target gene may be increased in vitro or in vivo.

[0013] A method of treating a patient is also provided. The patient may suffer from a disorder set forth in Table 8. The patient may be administered a composition comprising a nucleic acid. The nucleic acid may comprise a sequence of any of SEQ ID NOS: 3A-7A, SEQ ID NOS: 1-26518, the complement thereof, or a sequence at least 81% identical to 21 contiguous nucleotides thereof. The nucleic acid may also comprise the sequence of SEQ ID NOS: 1A or 2A, any of SEQ ID NOS: 1-26518, the complement thereof, or a sequence at least 63% identical

to 81 contiguous nucleotides thereof. The nucleic acid may be from about 51 to about 250 nucleotides in length. The nucleic acid may comprise a modified base.

[0014] Also provided is a method of detecting a target nucleic acid is provided. The targeted nucleic acid may be any nucleic acid, such as a miRNA. A nucleic acid comprising a short RNA sequence and a DNA sequence including a T7 RNA promoter sequence, may be ligated to the miRNA molecule. Another oligonucleotide may be annealed to the T7 promoter region to enable RNA polymerase binding to the double strand. Repeated cycles of initiation and product release may then be performed. As a result a linear amplification of the transcript that includes a complementary sequence to the original natural miRNA is achieved.

[0015] During the process of transcription, nucleotides that are labeled (e.g., biotinylated) may be incorporated to the transcript, which may be useful in the later detection of the transcript by a variety of assays. In the case of target nucleic acids expressed at low level, such as miRNAs, this method may provide detection of such target nucleic acids in methods such as Luminex. Using LNA (locked nucleic acid) in the probes that are bound to the Luminex microspheres may allow hybridization even more specific and increases signals.

BRIEF DESCRIPTION OF SEQUENCE LISTING AND TABLES

[0016] Reference is made to the appendix submitted on the compact discs submitted herewith. The compact discs contain the following: SEQ_01.txt (665,724 KB, 2/7/2006), SEQ_02.txt (673,696 KB, 2/6/2006), SEQ_03.txt (672,326 KB, 2/6/2006), SEQ_04.txt (673,986 KB, 2/6/2006), SEQ_05.txt (668,866 KB, 2/6/2006), SEQ_06.txt (666,389 KB, 2/6/2006), SEQ_07.txt (673,699 KB, 2/6/2006), SEQ_08.txt (668,601 KB, 2/6/2006), SEQ_09.txt (667,636 KB, 2/6/2006), SEQ_10.txt (669,467 KB, 2/6/2006), SEQ_11.txt (668,877 KB, 2/6/2006), SEQ_12.txt (670,181 KB, 2/6/2006), SEQ_13.txt (668,792 KB, 2/6/2006), and SEQ_14.txt (499,859 KB, 2/6/2006), together which form the Sequence Listing, and the following tables: Table01.txt (1,277 KB, 2/6/2006), Table02.txt (226 KB, 2/6/2006), Table03.txt (210 KB, 2/7/2006), Table04.txt (15 KB, 2/6/2006), Table05_1.txt (460,791 KB, 2/7/2006), Table05_2.txt (395,344 KB, 2/7/2006), Table06.txt (21,798 KB, 2/6/2006), Table07.txt (48 KB, 2/6/2006), Table08.txt (4 KB, 2/6/2006), Table09.txt (1 KB, 2/6/2006), Table10.txt (17 KB, 2/6/2006), Table11.txt (231 KB, 2/6/2006), and Table12.txt (2,005 KB, 2/6/2006), the contents of which are incorporated by reference herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] Figure 1 demonstrates a model of maturation for miRNAs.

[0018] Figure 2 shows a schematic illustration of the MC19 cluster on 19q13.42. Panel A shows the ~500,000bp region of chromosome 19, from 58,580,001 to 59,080,000 (according to the May 2004 USCS assembly), in which the cluster is located including the neighboring protein-coding genes. The MC19-1 cluster is indicated by a rectangle. Mir-371, mir-372, and mir-373 are indicated by lines. Protein coding genes flanking the cluster are represented by large arrow-heads. Panel B shows a detailed structure of the MC19-1 miRNA cluster. A region of ~102,000bp, from 58,860,001 to 58,962,000 (according to the May 2004 USCS assembly), is presented. MiRNA precursors are represented by a black bars. It should be noted that all miRNAs are at the same orientation from left to right. Shaded areas around miRNA precursors represent repeating units in which the precursor is embedded. The location of mir-371, mir-372, and mir-373, is also presented.

[0019] Figure 3 is a graphical representation of multiple sequence alignment of 35 human repeat units at distinct size of ~690nt (A) and 26 chimpanzees repeat units (B). The graph was generated by calculating a similarity score for each position in the alignment with an averaging sliding window of 10nt (Maximum score -1, minimum score-0). The repeat unit sequences were aligned by ClustalW program. Each position of the resulting alignment was assigned a score which represented the degree of similarity at this position. The region containing the miRNA precursors is bordered by vertical lines. The exact location of the mature miRNAs derived from the 5' stems (5p) and 3' stems (3p) of the precursors is indicated by vertical lines.

[0020] Figure 4 shows sequence alignments of the 43 A-type pre-miRNAs of the MC19-1 cluster. Panel A shows the multiple sequence alignment with the Position of the mature miRNAs marked by a frame. The consensus sequence is shown at the bottom. Conserved nucleotides are colored as follows: black-100%, dark grey- 80% to 99%, and clear grey- 60% to 79%. Panel B shows alignments of consensus mature A-type miRNAs with the upstream human cluster of mir-371, mir-372, miR-373. Panel C shows alignments of consensus mature A-type miRNAs with the hsa-mir-371-373 mouse orthologous cluster.

[0021] Figure 5 shows expression analysis of the MC19-1 miRNAs. Panel A shows a Northern blot analysis of two selected A-type miRNAs. Expression was analyzed using total RNA from human brain (B), liver (L), thymus (T), placenta (P) and HeLa cells (H). The expression of

mir-98 and ethidium bromide staining of the tRNA band served as control. Panel B shows RT-PCR analysis of the mRNA transcript containing the A-type miRNA precursors. Reverse transcription of 5 μ g total RNA from placenta was performed using oligo-dT. This was followed by PCR using the denoted primers (indicated by horizontal arrows). The region examined is illustrated at the top. Vertical black bars represent the pre-miRNA; shaded areas around the pre-miRNAs represent the repeating units; the location of four ESTs is indicated at the right side; the poly-A site, as found in the ESTs and located downstream to a AATAAA consensus, is indicated by a vertical arrow. The fragments expected from RT-PCR using three primer combinations are indicated below the illustration of the cluster region. The results of the RT-PCR analysis are presented below the expected fragments. Panel C shows the sequencing strategy of the FR2 fragment. The fragment was cloned into the pTZ57R/T vector and sequenced using external and internal primers.

[0022] Figure 6 shows the predicted hairpin structure of a miRNA precursor (SEQ ID NO: 2A) encoded by the 13_12 gene. The residues marked with underlines are predicted to be the mature miRNA (3') and miRNA* (5').

[0023] Figure 7 shows an alignment of predicted and/or validated miRNA precursors homologous to 13_12 which encode the following miRNAs: mmu-mir-193 (SEQ ID NO: 8A), rno-mir-193 (SEQ ID NO: 9A), dre-mir-193a (SEQ ID NO: 10A), dre-mir-193b (SEQ ID NO: 11A), has-mir-193a (SEQ ID NO: 12A) and has-mir-193b (13_12)(SEQ ID NO: 2A).

[0024] Figure 8 shows an alignment of the following 13_12-derived miRNAs that were cloned: 19061 (SEQ ID NO: 3A), 19062 (SEQ ID NO: 4A), 19063 (SEQ ID NO: 5A), (SEQ ID NO: 6A) and 19065 (SEQ ID NO: 7A).

[0025] Figure 9 shows a Northern blot analysis of 13_12 expression in brain (B), liver (L), thymus (T), placenta (P) and HeLa cells (H) using 40 μ g total RNA. Expression of the has-mir-98 was used as a control. Equal loading of the gel before transfer to membrane was monitored by ethidium bromide staining of the tRNA band (bottom). The expression of has-mir-98 was examined for reference.

[0026] Figure 10 shows the relative expression of 13_12 and other miRNA genes in prostate and lung tumors, with the Y-axis being shown in log base 2.

[0027] Figure 11 shows that inhibition of 13_12 caused a decrease in the proliferation of PC-3 cells.

[0028] Figure 12 shows the comparative inhibition of the 13_12 inhibitor and other miRNA inhibitors in PC-3 cells.

[0029] Figure 13 shows the comparative inhibition of the 13_12 inhibitor and other miRNA inhibitors in A549 cells.

[0030] Figure 13 shows linear amplification of microRNA. An adaptor containing a DNA sequence complementary to an RNA promoter is ligated to a microRNA. A primer containing the DNA sequence for the promoter is then annealed to the miRNA-adaptor hybrid. Labeled reverse complements to the miRNA are then generated by RNA polymerase.

[0031] Figure 14 shows sequences for an adaptor (SEQ ID NO: 1B), annealing primer (SEQ ID NO: 2B) and final transcript (SEQ ID NO: 3B).

DETAILED DESCRIPTION

[0032] Nucleic acids are provided related to miRNAs, precursors thereto, and targets thereof. Such nucleic acids may be useful for diagnostic and prognostic purposes, and also for modifying target gene expression. Also provided are methods and compositions that may be useful, among other things, for diagnostic and prognostic purposes. Other aspects of the invention will become apparent to the skilled artisan by the following description of the invention.

1. Definitions

[0033] Before the present compounds, 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.

a. animal

[0034] "Animal" as used herein may mean fish, amphibians, reptiles, birds, and mammals, such as mice, rats, rabbits, goats, cats, dogs, cows, apes and humans.

b. attached

[0035] "Attached" or "immobilized" as used herein to refer to a probe and a solid support may mean that the binding between the probe and the solid support is sufficient to be stable under conditions of binding, washing, analysis, and removal. The binding may be covalent or non-covalent. Covalent bonds may be formed directly between the probe and the solid support or

may be formed by a cross linker or by inclusion of a specific reactive group on either the solid support or the probe or both molecules. Non-covalent binding may be one or more of electrostatic, hydrophilic, and hydrophobic interactions. Included in non-covalent binding is the covalent attachment of a molecule, such as streptavidin, to the support and the non-covalent binding of a biotinylated probe to the streptavidin. Immobilization may also involve a combination of covalent and non-covalent interactions.

c. biological sample

[0036] "Biological sample" as used herein may mean a sample of biological tissue or fluid that comprises nucleic acids. Such samples include, but are not limited to, tissue or fluid isolated from animals. Biological samples may also include sections of tissues such as biopsy and autopsy samples, frozen sections taken for histologic purposes, blood, plasma, serum, sputum, stool, tears, mucus, hair, and skin. Biological samples also include explants and primary and/or transformed cell cultures derived from animal or patient tissues. A biological sample may be provided by removing a sample of cells from an animal, but can also be accomplished by using previously isolated cells (e.g., isolated by another person, at another time, and/or for another purpose), or by performing the methods described herein in vivo. Archival tissues, such as those having treatment or outcome history, may also be used.

d. complement

[0037] "Complement" or "complementary" as used herein to refer to a nucleic acid may mean Watson-Crick (e.g., A-T/U and C-G) or Hoogsteen base pairing between nucleotides or nucleotide analogs of nucleic acid molecules.

e. differential expression

[0038] "Differential expression" may mean qualitative or quantitative differences in the temporal and/or cellular gene expression patterns within and among cells and tissue. Thus, a differentially expressed gene may qualitatively have its expression altered, including an activation or inactivation, in, e.g., normal versus disease tissue. Genes may be turned on or turned off in a particular state, relative to another state thus permitting comparison of two or more states. A qualitatively regulated gene may exhibit an expression pattern within a state or cell type which may be detectable by standard techniques. Some genes may be expressed in one state or cell type, but not in both. Alternatively, the difference in expression may be quantitative, e.g., in that expression is modulated, either up-regulated, resulting in an increased amount of transcript, or

down-regulated, resulting in a decreased amount of transcript. The degree to which expression differs need only be large enough to quantify via standard characterization techniques such as expression arrays, quantitative reverse transcriptase PCR, northern analysis, and RNase protection.

f. gene

[0039] "Gene" used herein may be a natural (*e.g.*, genomic) or synthetic gene comprising transcriptional and/or translational regulatory sequences and/or a coding region and/or non-translated sequences (*e.g.*, introns, 5'- and 3'-untranslated sequences). The coding region of a gene may be a nucleotide sequence coding for an amino acid sequence or a functional RNA, such as tRNA, rRNA, catalytic RNA, siRNA, miRNA or antisense RNA. A gene may also be an mRNA or cDNA corresponding to the coding regions (*e.g.*, exons and miRNA) optionally comprising 5'- or 3'-untranslated sequences linked thereto. A gene may also be an amplified nucleic acid molecule produced *in vitro* comprising all or a part of the coding region and/or 5'- or 3'-untranslated sequences linked thereto.

g. host cell

[0040] "Host cell" used herein may be a naturally occurring cell or a transformed cell that may contain a vector and may support replication of the vector. Host cells may be cultured cells, explants, cells *in vivo*, and the like. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells, such as CHO and HeLa.

h. identity

[0041] "Identical" or "identity" as used herein in the context of two or more nucleic acids or polypeptide sequences, may mean that the sequences have a specified percentage of residues that are the same over a specified region. The percentage may be calculated by optimally aligning the two sequences, comparing the two sequences over the specified region, determining the number of positions at which the identical residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the specified region, and multiplying the result by 100 to yield the percentage of sequence identity. In cases where the two sequences are of different lengths or the alignment produces one or more staggered ends and the specified region of comparison includes only a single sequence, the residues of single sequence are included in the denominator but not the numerator of the calculation. When comparing DNA and RNA, thymine (T) and uracil (U) may be considered

equivalent. Identity may be performed manually or by using a computer sequence algorithm such as BLAST or BLAST 2.0.

i. label

[0042] "Label" as used herein may mean a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, chemical, or other physical means. For example, useful labels include ^{32}P , fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin, digoxigenin, or haptens and other entities which can be made detectable. A label may be incorporated into nucleic acids and proteins at any position.

j. nucleic acid

[0043] "Nucleic acid" or "oligonucleotide" or "polynucleotide" used herein may mean at least two nucleotides covalently linked together. The depiction of a single strand also defines the sequence of the complementary strand. Thus, a nucleic acid also encompasses the complementary strand of a depicted single strand. Many variants of a nucleic acid may be used for the same purpose as a given nucleic acid. Thus, a nucleic acid also encompasses substantially identical nucleic acids and complements thereof. A single strand provides a probe that may hybridize to a target sequence under stringent hybridization conditions. Thus, a nucleic acid also encompasses a probe that hybridizes under stringent hybridization conditions.

[0044] Nucleic acids may be single stranded or double stranded, or may contain portions of both double stranded and single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA, or a hybrid, where the nucleic acid may contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine and isoguanine. Nucleic acids may be obtained by chemical synthesis methods or by recombinant methods.

[0045] A nucleic acid will generally contain phosphodiester bonds, although nucleic acid analogs may be included that may have at least one different linkage, e.g., phosphoramidate, phosphorothioate, phosphorodithioate, or O-methylphosphoroamidite linkages and peptide nucleic acid backbones and linkages. Other analog nucleic acids include those with positive backbones; non-ionic backbones, and non-ribose backbones, including those described in U.S. Pat. Nos. 5,235,033 and 5,034,506, which are incorporated by reference. Nucleic acids containing one or more non-naturally occurring or modified nucleotides are also included within one definition of nucleic acids. 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. Modified nucleotides also include nucleotides conjugated with cholesterol through, e.g., a hydroxyprolinol linkage as described in Krutzfeldt et al., Nature (Oct. 30, 2005), Soutschek et al., Nature 432:173-178 (2004), and U.S. Patent Publication No. 20050107325, which are incorporated herein by reference. Additional modified nucleotides and nucleic acids are described in U.S. Patent Publication No. 20050182005, which is incorporated herein by reference. Modifications of the ribose-phosphate backbone may be done for a variety of reasons, e.g., to increase the stability and half-life of such molecules in physiological environments, to enhance diffusion across cell membranes, or as probes on a biochip. Mixtures of naturally occurring nucleic acids and analogs may be made; alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made.

k. operably linked

[0046] "Operably linked" used herein may mean that expression of a gene is under the control of a promoter with which it is spatially connected. A promoter may be positioned 5' (upstream) or 3' (downstream) of a gene under its control. The distance between the promoter and a gene may be approximately the same as the distance between that promoter and the gene it controls in the gene from which the promoter is derived. As is known in the art, variation in this distance may be accommodated without loss of promoter function.

l. probe

[0047] "Probe" as used herein may mean an oligonucleotide capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. Probes may bind target sequences lacking complete complementarity with the probe sequence depending

upon the stringency of the hybridization conditions. There may be any number of base pair mismatches which will interfere with hybridization between the target sequence and the single stranded nucleic acids described herein. However, if the number of mutations is so great that no hybridization can occur under even the least stringent of hybridization conditions, the sequence is not a complementary target sequence. A probe may be single stranded or partially single and partially double stranded. The strandedness of the probe is dictated by the structure, composition, and properties of the target sequence. Probes may be directly labeled or indirectly labeled such as with biotin to which a streptavidin complex may later bind.

m. promoter

[0048] "Promoter" as used herein may mean a synthetic or naturally-derived molecule which is capable of conferring, activating or enhancing expression of a nucleic acid in a cell. A promoter may comprise one or more specific transcriptional regulatory sequences to further enhance expression and/or to alter the spatial expression and/or temporal expression of same. A promoter may also comprise distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A promoter may be derived from sources including viral, bacterial, fungal, plants, insects, and animals. A promoter may regulate the expression of a gene component constitutively, or differentially with respect to cell, the tissue or organ in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, pathogens, metal ions, or inducing agents. Representative examples of promoters include the bacteriophage T7 promoter, bacteriophage T3 promoter, SP6 promoter, lac operator-promoter, tac promoter, SV40 late promoter, SV40 early promoter, RSV-LTR promoter, CMV IE promoter, SV40 early promoter or SV40 late promoter and the CMV IE promoter.

n. selectable marker

[0049] "Selectable marker" used herein may mean any gene which confers a phenotype on a host cell in which it is expressed to facilitate the identification and/or selection of cells which are transfected or transformed with a genetic construct. Representative examples of selectable markers include the ampicillin-resistance gene (Amp^r), tetracycline-resistance gene (Tc^r), bacterial kanamycin-resistance gene (Kan^r), zeocin resistance gene, the AURI-C gene which confers resistance to the antibiotic aureobasidin A, phosphinothricin-resistance gene, neomycin phosphotransferase gene ($nptII$), hygromycin-resistance gene, beta-glucuronidase (GUS) gene,

chloramphenicol acetyltransferase (CAT) gene, green fluorescent protein (GFP)-encoding gene and luciferase gene.

o. stringent hybridization conditions

[0050] "Stringent hybridization conditions" used herein may mean conditions under which a first nucleic acid sequence (e.g., probe) will hybridize to a second nucleic acid sequence (e.g., target), such as in a complex mixture of nucleic acids. Stringent conditions are sequence-dependent and will be different in different circumstances. Stringent conditions may be selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m may be the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium). Stringent conditions may be those in which the salt concentration is less than about 1.0 M sodium ion, such as about 0.01-1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., about 10-50 nucleotides) and at least about 60°C for long probes (e.g., greater than about 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal may be at least 2 to 10 times background hybridization. Exemplary stringent hybridization conditions include the following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42°C, or, 5x SSC, 1% SDS, incubating at 65°C, with wash in 0.2x SSC, and 0.1% SDS at 65°C.

p. substantially complementary

[0051] "Substantially complementary" used herein may mean that a first sequence is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% identical to the complement of a second sequence over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more nucleotides, or that the two sequences hybridize under stringent hybridization conditions.

q. substantially identical

[0052] "Substantially identical" used herein may mean that a first and second sequence are at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% identical over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65,

70, 75, 80, 85, 90, 95, 100 or more nucleotides or amino acids, or with respect to nucleic acids, if the first sequence is substantially complementary to the complement of the second sequence.

r. target

[0053] "Target" as used herein may mean a polynucleotide that may be bound by one or more probes under stringent hybridization conditions.

s. terminator

[0054] "Terminator" used herein may mean a sequence at the end of a transcriptional unit which signals termination of transcription. A terminator may be a 3'-non-translated DNA sequence containing a polyadenylation signal, which may facilitate the addition of polyadenylate sequences to the 3'-end of a primary transcript. A terminator may be derived from sources including viral, bacterial, fungal, plants, insects, and animals. Representative examples of terminators include the SV40 polyadenylation signal, HSV TK polyadenylation signal, CYC1 terminator, ADH terminator, SPA terminator, nopaline synthase (NOS) gene terminator of *Agrobacterium tumefaciens*, the terminator of the Cauliflower mosaic virus (CaMV) 35S gene, the zein gene terminator from *Zea mays*, the Rubisco small subunit gene (SSU) gene terminator sequences, subclover stunt virus (SCSV) gene sequence terminators, rho-independent *E. coli* terminators, and the lacZ alpha terminator.

t. variant

[0055] "Variant" used herein to refer to a nucleic acid may mean (i) a portion of a referenced nucleotide sequence; (ii) the complement of a referenced nucleotide sequence or portion thereof; (iii) a nucleic acid that is substantially identical to a referenced nucleic acid or the complement thereof; or (iv) a nucleic acid that hybridizes under stringent conditions to the referenced nucleic acid, complement thereof, or a sequences substantially identical thereto.

u. vector

[0056] "Vector" used herein may mean a nucleic acid sequence containing an origin of replication. A vector may be a plasmid, bacteriophage, bacterial artificial chromosome or yeast artificial chromosome. A vector may be a DNA or RNA vector. A vector may be either a self-replicating extrachromosomal vector or a vector which integrates into a host genome.

2. MicroRNA

[0057] While not being bound by theory, the current model for the maturation of mammalian miRNAs is shown in Figure 1. A gene coding for a miRNA may be transcribed leading to

production of an miRNA precursor known as the pri-miRNA. The pri-miRNA may be part of a polycistronic RNA comprising multiple pri-miRNAs. The pri-miRNA may form a hairpin with a stem and loop. As indicated on Figure 1, the stem may comprise mismatched bases.

[0058] The hairpin structure of the pri-miRNA may be recognized by Drosha, which is an RNase III endonuclease. Drosha may recognize terminal loops in the pri-miRNA and cleave approximately two helical turns into the stem to produce a 60–70 nt precursor known as the pre-miRNA. Drosha may cleave the pri-miRNA with a staggered cut typical of RNase III endonucleases yielding a pre-miRNA stem loop with a 5' phosphate and ~2 nucleotide 3' overhang. Approximately one helical turn of stem (~10 nucleotides) extending beyond the Drosha cleavage site may be essential for efficient processing. The pre-miRNA may then be actively transported from the nucleus to the cytoplasm by Ran-GTP and the export receptor Exportin-5.

[0059] The pre-miRNA may be recognized by Dicer, which is also an RNase III endonuclease. Dicer may recognize the double-stranded stem of the pre-miRNA. Dicer may also recognize the 5' phosphate and 3' overhang at the base of the stem loop. Dicer may cleave off the terminal loop two helical turns away from the base of the stem loop leaving an additional 5' phosphate and ~2 nucleotide 3' overhang. The resulting siRNA-like duplex, which may comprise mismatches, comprises the mature miRNA and a similar-sized fragment known as the miRNA*. The miRNA and miRNA* may be derived from opposing arms of the pri-miRNA and pre-miRNA. MiRNA* sequences may be found in libraries of cloned miRNAs but typically at lower frequency than the miRNAs.

[0060] Although initially present as a double-stranded species with miRNA*, the miRNA may eventually become incorporated as a single-stranded RNA into a ribonucleoprotein complex known as the RNA-induced silencing complex (RISC). Various proteins can form the RISC, which can lead to variability in specificity for miRNA/miRNA* duplexes, binding site of the target gene, activity of miRNA (repress or activate), and which strand of the miRNA/miRNA* duplex is loaded in to the RISC.

[0061] 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.

[0062] The RISC may identify target nucleic acids based on high levels of complementarity between the miRNA and the mRNA, especially by nucleotides 2-8 of the miRNA. Only one case has been reported in animals where the interaction between the miRNA and its target was along the entire length of the miRNA. This was shown for mir-196 and Hox B8 and it was further shown that mir-196 mediates the cleavage of the Hox B8 mRNA (Yekta et al 2004, Science 304-594). Otherwise, such interactions are known only in plants (Bartel & Bartel 2003, Plant Physiol 132-709).

[0063] A number of studies have looked at the base-pairing requirement between miRNA and its mRNA target for achieving efficient inhibition of translation (reviewed by Bartel 2004, Cell 116-281). In mammalian cells, the first 8 nucleotides of the miRNA may be important (Doench & Sharp 2004 GenesDev 2004-504). However, other parts of the microRNA may also participate in mRNA binding. Moreover, sufficient base pairing at the 3' can compensate for insufficient pairing at the 5' (Brennecke et al, 2005 PLoS 3-e85). Computation studies, analyzing miRNA binding on whole genomes have suggested a specific role for bases 2-7 at the 5' of the miRNA in target binding but the role of the first nucleotide, found usually to be "A" was also recognized (Lewis et al 2005 Cell 120-15). Similarly, nucleotides 1-7 or 2-8 were used to identify and validate targets by Krek et al (2005, Nat Genet 37-495).

[0064] The target sites in the mRNA may be in the 5' UTR, the 3' UTR or in the coding region. Interestingly, multiple miRNAs may regulate the same mRNA target by recognizing the same or multiple sites. The presence of multiple miRNA binding sites in most genetically identified targets may indicate that the cooperative action of multiple RISCs provides the most efficient translational inhibition.

[0065] 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 a certain degree of 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 the requisite degree of complementarity to the miRNA. Translational repression may be more

prevalent in animals since animals may have a lower degree of complementarity between the miRNA and binding site.

[0066] It should be noted that there may be variability in the 5' and 3' ends of any pair of miRNA and miRNA*. This variability may be due to variability in the enzymatic processing of Drosha and Dicer with respect to the site of cleavage. Variability at the 5' and 3' ends of miRNA and miRNA* may also be due to mismatches in the stem structures of the pri-miRNA and pre-miRNA. The mismatches of the stem strands may lead to a population of different hairpin structures. Variability in the stem structures may also lead to variability in the products of cleavage by Drosha and Dicer.

3. Nucleic Acid

[0067] Nucleic acids are provided herein. The nucleic acid may comprise the sequence of SEQ ID NOS: 1-42840282 or variants thereof. The variant may be a complement of the referenced nucleotide sequence. The variant may also be a nucleotide sequence that is substantially identical to the referenced nucleotide sequence or the complement thereof. The variant may also be a nucleotide sequence which hybridizes under stringent conditions to the referenced nucleotide sequence, complements thereof, or nucleotide sequences substantially identical thereto.

[0068] The nucleic acid may have a length of from 10 to 250 nucleotides. The nucleic acid may have a length of at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200 or 250 nucleotides. The nucleic acid may be synthesized or expressed in a cell (in vitro or in vivo) using a synthetic gene described herein. The nucleic acid may be synthesized as a single strand molecule and hybridized to a substantially complementary nucleic acid to form a duplex. The nucleic acid may be introduced to a cell, tissue or organ in a single- or double-stranded form or capable of being expressed by a synthetic gene using methods well known to those skilled in the art, including as described in U.S. Patent No. 6,506,559 which is incorporated by reference.

a. Pri-miRNA

[0069] The nucleic acid may comprise a sequence of a pri-miRNA or a variant thereof. The pri-miRNA sequence may comprise from 45-250, 55-200, 70-150 or 80-100 nucleotides. The sequence of the pri-miRNA may comprise a pre-miRNA, miRNA and miRNA*, as set forth

herein, and variants thereof. The sequence of the pri-miRNA may comprise the sequence of SEQ ID NOS: 1-8857-26518 or variants thereof.

[0070] The pri-miRNA may form a hairpin structure. The hairpin may comprise a first and second nucleic acid sequence that are substantially complimentary. The first and second nucleic acid sequence may be from 37-50 nucleotides. The first and second nucleic acid sequence may be separated by a third sequence of from 8-12 nucleotides. The hairpin structure may have a free energy less than -25 Kcal/mole as calculated by the Vienna algorithm with default parameters, as described in Hofacker et al., Monatshefte f. Chemie 125: 167-188 (1994), the contents of which are incorporated herein. The hairpin may comprise a terminal loop of 4-20, 8-12 or 10 nucleotides. The pri-miRNA may comprise at least 19% adenosine nucleotides, at least 16% cytosine nucleotides, at least 23% thymine nucleotides and at least 19% guanine nucleotides.

b. Pre-miRNA

[0071] The nucleic acid may also comprise a sequence of a pre-miRNA or a variant thereof. The pre-miRNA sequence may comprise from 45-90, 60-80 or 60-70 nucleotides. The sequence of the pre-miRNA may comprise a miRNA and a miRNA* as set forth herein, The sequence of the pre-miRNA may also be that of a pri-miRNA excluding from 0-160 nucleotides from the 5' and 3' ends of the pri-miRNA. The sequence of the pre-miRNA may comprise the sequence of SEQ ID NOS: 1-8857-26518 or variants thereof.

c. MiRNA

[0072] The nucleic acid may also comprise a sequence of a miRNA (including miRNA*) or a variant thereof. The miRNA sequence may comprise from 13-33, 18-24 or 21-23 nucleotides. The miRNA may also comprise a total of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 nucleotides. The sequence of the miRNA may be the first 13-33 nucleotides of the pre-miRNA. The sequence of the miRNA may also be the last 13-33 nucleotides of the pre-miRNA. The sequence of the miRNA may comprise the sequence of SEQ ID NOS: 8857-26518 or variants thereof.

d. Anti-miRNA

[0073] The nucleic acid may also comprise a sequence of an anti-miRNA that is capable of blocking the activity of a miRNA or miRNA*, such as by binding to the pri-miRNA, pre-miRNA, miRNA or miRNA* (e.g. antisense or RNA silencing), or by binding to the target binding site. The anti-miRNA may comprise a total of 5-100 or 10-60 nucleotides. The anti-

miRNA may also comprise a total of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 nucleotides. The sequence of the anti-miRNA may comprise (a) at least 5 nucleotides that are substantially identical or complimentary to the 5' of a miRNA and at least 5-12 nucleotides that are substantially complimentary to the flanking regions of the target site from the 5' end of the miRNA, or (b) at least 5-12 nucleotides that are substantially identical or complimentary to the 3' of a miRNA and at least 5 nucleotide that are substantially complimentary to the flanking region of the target site from the 3' end of the miRNA. The sequence of the anti-miRNA may comprise the compliment of SEQ ID NOS: 8857-26518 or variants thereof.

e. Binding Site of Target

[0074] The nucleic acid may also comprise a sequence of a target miRNA binding site, or a variant thereof. The target site sequence may comprise a total of 5-100 or 10-60 nucleotides. The target site sequence may also comprise a total of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62 or 63 nucleotides. The target site sequence may comprise at least 5 nucleotides of the sequence of SEQ ID NOS: 8857-26518 or a target gene binding site referred to in Table 5.

4. Synthetic Gene

[0075] A synthetic gene is also provided comprising a nucleic acid described herein operably linked to a transcriptional and/or translational regulatory sequence. The synthetic gene may be capable of modifying the expression of a target gene with a binding site for a nucleic acid described herein. Expression of the target gene may be modified in a cell, tissue or organ. The synthetic gene may be synthesized or derived from naturally-occurring genes by standard recombinant techniques. The synthetic gene may also comprise terminators at the 3'-end of the transcriptional unit of the synthetic gene sequence. The synthetic gene may also comprise a selectable marker.

5. Vector

[0076] A vector is also provided comprising a synthetic gene described herein. The vector may be an expression vector. An expression vector may comprise additional elements. For example, the expression vector may have two replication systems allowing it to be maintained in two organisms, e.g., in one host cell for expression and in a second host cell (e.g., bacteria) for

cloning and amplification. For integrating expression vectors, the expression vector may contain at least one sequence homologous to the host cell genome, and preferably two homologous sequences which flank the expression construct. The integrating vector may be directed to a specific locus in the host cell by selecting the appropriate homologous sequence for inclusion in the vector. The vector may also comprise a selectable marker gene to allow the selection of transformed host cells.

6. Host Cell

[0077] A host cell is also provided comprising a vector, synthetic gene or nucleic acid described herein. The cell may be a bacterial, fungal, plant, insect or animal cell.

7. Probes

[0078] A probe is also provided comprising a nucleic acid described herein. Probes may be used for screening and diagnostic methods, as outlined below. The probe may be attached or immobilized to a solid substrate, such as a biochip.

[0079] The probe may have a length of from 8 to 500, 10 to 100 or 20 to 60 nucleotides. The probe may also have a length of at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280 or 300 nucleotides. The probe may further comprise a linker sequence of from 10-60 nucleotides.

8. Biochip

[0080] A biochip is also provided. The biochip may comprise a solid substrate comprising an attached probe or plurality of probes described herein. The probes may be capable of hybridizing to a target sequence under stringent hybridization conditions. The probes may be attached at spatially defined address on the substrate. More than one probe per target sequence may be used, with either overlapping probes or probes to different sections of a particular target sequence. The probes may be capable of hybridizing to target sequences associated with a single disorder.

[0081] The probes may be attached to the biochip in a wide variety of ways, as will be appreciated by those in the art. The probes may either be synthesized first, with subsequent attachment to the biochip, or may be directly synthesized on the biochip.

[0082] The solid substrate may be a material that may be modified to contain discrete individual sites appropriate for the attachment or association of the probes and is amenable to at least one

detection method. Representative examples of substrates include glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon, etc.), polysaccharides, nylon or nitrocellulose, resins, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses and plastics. The substrates may allow optical detection without appreciably fluorescing.

[0083] The substrate may be planar, although other configurations of substrates may be used as well. For example, probes may be placed on the inside surface of a tube, for flow-through sample analysis to minimize sample volume. Similarly, the substrate may be flexible, such as a flexible foam, including closed cell foams made of particular plastics.

[0084] The biochip and the probe may be derivatized with chemical functional groups for subsequent attachment of the two. For example, the biochip may be derivatized with a chemical functional group including, but not limited to, amino groups, carboxyl groups, oxo groups or thiol groups. Using these functional groups, the probes may be attached using functional groups on the probes either directly or indirectly using a linker. The probes may be attached to the solid support by either the 5' terminus, 3' terminus, or via an internal nucleotide.

[0085] The probe may also be attached to the solid support non-covalently. For example, biotinylated oligonucleotides can be made, which may bind to surfaces covalently coated with streptavidin, resulting in attachment. Alternatively, probes may be synthesized on the surface using techniques such as photopolymerization and photolithography.

9. Expression Analysis

[0086] A method of identifying a nucleic acid associated with a disease or a pathological condition is also provided. The method comprises measuring a level of the nucleic acid in a sample that is different than the level of a control. Detection may be performed by contacting the sample with a probe or biochip described herein and detecting the amount of hybridization. PCR may be used to amplify nucleic acids in the sample, which may provide higher sensitivity.

[0087] The level of the nucleic acid in the sample may also be compared to a control cell (*e.g.*, a normal cell) to determine whether the nucleic acid is differentially expressed (*e.g.*, overexpressed or underexpressed). The ability to identify miRNAs that are differentially expressed in pathological cells compared to a control can provide high-resolution, high-sensitivity datasets which may be used in the areas of diagnostics, prognostics, therapeutics, drug development,

pharmacogenetics, biosensor development, and other related areas. An expression profile generated by the current methods may be a "fingerprint" of the state of the sample with respect to a number of miRNAs. While two states may have any particular miRNA similarly expressed, the evaluation of a number of miRNAs simultaneously allows the generation of a gene expression profile that is characteristic of the state of the cell. That is, normal tissue may be distinguished from diseased tissue. By comparing expression profiles of tissue in known different disease states, information regarding which miRNAs are associated in each of these states may be obtained. Then, diagnosis may be performed or confirmed to determine whether a tissue sample has the expression profile of normal or disease tissue. This may provide for molecular diagnosis of related conditions.

[0088] The expression level of a disease-associated nucleic acid is information in a number of ways. For example, a differential expression of a disease-associated nucleic acid compared to a control may be used as a diagnostic that a patient suffers from the disease. Expression levels of a disease-associated nucleic acid may also be used to monitor the treatment and disease state of a patient. Furthermore, expression levels of a disease-associated miRNA may allow the screening of drug candidates for altering a particular expression profile or suppressing an expression profile associated with disease.

[0089] A target nucleic acid may be detected and levels of the target nucleic acid measured by contacting a sample comprising the target nucleic acid with a biochip comprising an attached probe sufficiently complementary to the target nucleic acid and detecting hybridization to the probe above control levels.

[0090] The target nucleic acid may also be detected by immobilizing the nucleic acid to be examined on a solid support such as nylon membranes and hybridizing a labelled probe with the sample. Similarly, the target nucleic may also be detected by immobilizing the labeled probe to a solid support and hybridizing a sample comprising a labeled target nucleic acid. Following washing to remove the non-specific hybridization, the label may be detected.

[0091] The target nucleic acid may also be detected in situ by contacting permeabilized cells or tissue samples with a labeled probe to allow hybridization with the target nucleic acid. Following washing to remove the non-specifically bound probe, the label may be detected.

[0092] These assays can be direct hybridization assays or can comprise sandwich assays, which include the use of multiple probes, as is generally outlined in U.S. Pat. Nos. 5,681,702;

5,597,909; 5,545,730; 5,594,117; 5,591,584; 5,571,670; 5,580,731; 5,571,670; 5,591,584; 5,624,802; 5,635,352; 5,594,118; 5,359,100; 5,124,246; and 5,681,697, each of which is hereby incorporated by reference.

[0093] A variety of hybridization conditions may be used, including high, moderate and low stringency conditions as outlined above. The assays may be performed under stringency conditions which allow hybridization of the probe only to the target. Stringency can be controlled by altering a step parameter that is a thermodynamic variable, including, but not limited to, temperature, formamide concentration, salt concentration, chaotropic salt concentration pH, or organic solvent concentration.

[0094] Hybridization reactions may be accomplished in a variety of ways. Components of the reaction may be added simultaneously, or sequentially, in different orders. In addition, the reaction may include a variety of other reagents. These include salts, buffers, neutral proteins, e.g., albumin, detergents, etc. which may be used to facilitate optimal hybridization and detection, and/or reduce non-specific or background interactions. Reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors and anti-microbial agents may also be used as appropriate, depending on the sample preparation methods and purity of the target.

a. Diagnostic

[0095] A method of diagnosis is also provided. The method comprises detecting a differential expression level of a disease-associated nucleic acid in a biological sample. The sample may be derived from a patient. Diagnosis of a disease state in a patient may allow for prognosis and selection of therapeutic strategy. Further, the developmental stage of cells may be classified by determining temporarily expressed disease-associated nucleic acids.

[0096] In situ hybridization of labeled probes to tissue arrays may be performed. When comparing the fingerprints between an individual and a standard, the skilled artisan can make a diagnosis, a prognosis, or a prediction based on the findings. It is further understood that the genes which indicate the diagnosis may differ from those which indicate the prognosis and molecular profiling of the condition of the cells may lead to distinctions between responsive or refractory conditions or may be predictive of outcomes.

b. Drug Screening

[0097] A method of screening therapeutics is also provided. The method comprises contacting a pathological cell capable of expressing a disease related nucleic acid with a candidate therapeutic and evaluating the effect of a drug candidate on the expression profile of the disease associated nucleic acid. Having identified the differentially expressed nucleic acid, a variety of assays may be executed. Test compounds may be screened for the ability to modulate gene expression of the disease associated nucleic acid. Modulation includes both an increase and a decrease in gene expression.

[0098] The test compound or drug candidate may be any molecule, e.g., protein, oligopeptide, small organic molecule, polysaccharide, polynucleotide, etc., to be tested for the capacity to directly or indirectly alter the disease phenotype or the expression of the disease associated nucleic acid. Drug candidates encompass numerous chemical classes, such as small organic molecules having a molecular weight of more than 100 and less than about 500, 1,000, 1,500, 2,000 or 2,500 daltons. Candidate compounds may comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents may comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

[0099] Combinatorial libraries of potential modulators may be screened for the ability to bind to the disease associated nucleic acid or to modulate the activity thereof. The combinatorial library may be a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis by combining a number of chemical building blocks such as reagents. Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries encoded peptides, benzodiazepines, diversomers such as hydantoins, benzodiazepines and dipeptide, vinylogous polypeptides, analogous organic syntheses of small compound libraries, oligocarbamates, and/or peptidyl phosphonates, nucleic acid libraries, peptide nucleic acid libraries, antibody libraries, carbohydrate libraries, and small organic molecule libraries.

10. Gene Silencing

[0100] A method of reducing expression of a target gene in a cell, tissue or organ is also provided. Expression of the target gene may be reduced by expressing a nucleic acid described herein that comprises a sequence substantially complementary to one or more binding sites of the target mRNA. The nucleic acid may be a miRNA or a variant thereof. The nucleic acid may also be pri-miRNA, pre-miRNA, or a variant thereof, which may be processed to yield a miRNA. The expressed miRNA may hybridize to a substantially complementary binding site on the target mRNA, which may lead to activation of RISC-mediated gene silencing. An example for a study employing over-expression of miRNA is Yekta et al 2004, Science 304-594, which is incorporated herein by reference. One of ordinary skill in the art will recognize that the nucleic acids described herein may also be used to inhibit expression of target genes or inhibit activity of miRNAs using antisense methods well known in the art, as well as RNAi methods described in U.S. Patent Nos. 6,506,559 and 6,573,099, which are incorporated by reference.

[0101] The target of gene silencing may be a protein that causes the silencing of a second protein. By repressing expression of the target gene, expression of the second protein may be increased. Examples for efficient suppression of miRNA expression are the studies by Esau et al 2004 JBC 275-52361; and Cheng et al 2005 Nucleic Acids Res. 33-1290, which is incorporated herein by reference.

11. Gene Enhancement

[0102] A method of increasing expression of a target gene in a cell, tissue or organ is also provided. Expression of the target gene may be increased by expressing a nucleic acid described herein that comprises a sequence substantially complementary to a pri-miRNA, pre-miRNA, miRNA or a variant thereof. The nucleic acid may be an anti-miRNA. The anti-miRNA may hybridize with a pri-miRNA, pre-miRNA or miRNA, thereby reducing its gene repression activity. Expression of the target gene may also be increased by expressing a nucleic acid that is substantially complementary to a portion of the binding site in the target gene, such that binding of the nucleic acid to the binding site may prevent miRNA binding.

12. Therapeutic

[0103] A method of modulating a disease or disorder associated with developmental dysfunctions is also provided. The disease or disorder may be cancer, such as prostate or lung cancer. In general, the nucleic acid molecules described herein may be used as a modulator of

the expression of genes which are at least partially complementary to said nucleic acid. Further, miRNA molecules may act as target for therapeutic screening procedures, e.g. inhibition or activation of miRNA molecules might modulate a cellular differentiation process, e.g. proliferation or apoptosis.

[0104] Furthermore, existing miRNA molecules may be used as starting materials for the manufacture of sequence-modified miRNA molecules, in order to modify the target-specificity thereof, e.g. an oncogene, a multidrug-resistance gene or another therapeutic target gene. Further, miRNA molecules can be modified, in order that they are processed and then generated as double-stranded siRNAs which are again directed against therapeutically relevant targets. Furthermore, miRNA molecules may be used for tissue reprogramming procedures, e.g. a differentiated cell line might be transformed by expression of miRNA molecules into a different cell type or a stem cell.

13. Compositions

[0105] A pharmaceutical composition is also provided. The composition may comprise a nucleic acid described herein and optionally a pharmaceutically acceptable carrier. The compositions may be used for diagnostic or therapeutic applications. The pharmaceutical composition may be administered by known methods, including wherein a nucleic acid is introduced into a desired target cell in vitro or in vivo. Commonly used gene transfer techniques include calcium phosphate, DEAE-dextran, electroporation, microinjection, viral methods and cationic liposomes.

14. Kits

[0106] A kit is also provided comprising a nucleic acid described herein together with any or all of the following: assay reagents, buffers, probes and/or primers, and sterile saline or another pharmaceutically acceptable emulsion and suspension base. In addition, the kits may include instructional materials containing directions (e.g., protocols) for the practice of the methods described herein.

15. Method of Synthesis

[0107] A method of synthesizing the reverse-complement of a target nucleic acid is also provided. A first nucleic acid may be provided comprising the target nucleic acid and an adapter nucleic acid. The target nucleic acid may be 5' of the second nucleic acid. A second nucleic acid may then be provided. A portion of the second nucleic acid may be substantially complementary

to a portion of the adapter nucleic acid. The second nucleic acid may be annealed to the adapter nucleic acid of the first nucleic acid to form an annealed complex. The second nucleic acid may then be extended from its 3' end, which may lead to the synthesis of the reverse-complement of the target nucleic acid. Extension may occur in a solution comprising labeled nucleotides, such as biotinylated UTP or/and CTP.

[0108] The extended second nucleic acid may then be displaced from the first nucleic acid. The extended second nucleic acid may be displaced using heat denaturation. Another cycle of extension may be performed by providing a second nucleic acid, a portion of which may be substantially complementary to a portion of the adapter nucleic acid. The second nucleic acid may then be annealed to the adapter nucleic acid of the first nucleic acid to form an annealed complex. The second nucleic acid may then be extended from its 3' end, which may lead to the synthesis of another reverse-complement of the target nucleic acid. Additional cycles of displacement and extension may then be performed until a desirable amount of reverse-complement of the target nucleic acid has been produced. The number of cycles of displacement and extension performed may be at least 1, 10, 50, 100, 1000 or 10,000.

[0109] The resulting reverse-complement transcripts may be utilized for a number of different purposes. For example, the amplified sequence may be cloned. In addition, the reverse-complement transcripts transcript may be used in a variety of different methods. The amplification of the transcript may be linear, which may lead to improved quantitative analysis.

a. First Nucleic Acid

[0110] The first nucleic acid may be formed by ligating the target nucleic acid to the adapter nucleic acid. Any ligase may be used to ligate the adaptor to the target nucleic acid to the adaptor nucleic acid, such as T4 RNA ligase. The ligated target nucleic acid- adaptor nucleic acid hybrid may be purified by a variety of methods including, but not limited to, acrylamide gel electrophoresis. The hybrid molecules may then be excised from the gel and extracted by any method, for example, by dialysis.

(1) Target Nucleic Acid.

[0111] The target nucleic acid may be derived from a biological sample comprising nucleic acids. The target nucleic acid may be enriched by size-fractionation.

[0112] The target nucleic acid may any nucleic acid. The target nucleic acid may comprise RNA. The RNA may comprise 18 to 24 nucleotides. The RNA may be a miRNA or a shRNA.

(2) Adapter Nucleic Acid

[0113] The adapter nucleic acid may comprise DNA. The adapter nucleic acid may comprise 10 to 20 nucleotides of DNA. The adapter nucleic acid may also comprise RNA. The adapter nucleic acid may comprise 1 to 5 nucleotides of RNA. The adapter nucleic acid may also comprise DNA and RNA. The adapter nucleic acid may comprise 10 to 20 nucleotides of DNA. The RNA of the adapter nucleic acid may be 5' of the DNA.

[0114] The adapter nucleic acid may comprise a 5'-phosphate. The adapter nucleic acid may also comprise a 3'-IdT.

b. Second Nucleic Acid

[0115] The second nucleic acid may be any nucleic acid. The second nucleic acid may comprise DNA.

c. Polymerase

[0116] The second nucleic acid may be extended using a suitable polymerase. Representative examples of polymerases include, but are not limited to, T7 RNA polymerase, T3 RNA polymerase and SP6 RNA polymerase.

d. Promoter

[0117] A double-stranded portion of the annealed complex formed by the adapter nucleic acid and the second nucleic acid may comprise a promoter. Both strands of the promoter may be DNA.

[0118] The promoter may be a synthetic or naturally-derived molecule which is capable of conferring, activating or enhancing expression of a nucleic acid in a cell. The promoter may comprise one or more specific transcriptional regulatory sequences to further enhance expression and/or to alter the spatial expression and/or temporal expression of same. A promoter may also comprise distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A promoter may be derived from sources including viral, bacterial, fungal, plants, insects, and animals. A promoter may regulate the expression of a gene component constitutively, or differentially with respect to cell, the tissue or organ in which expression occurs or, with respect to the developmental stage at which expression occurs, or in response to external stimuli such as physiological stresses, pathogens, metal ions, or inducing agents. Representative examples of promoters include the T7 promoter, T3 promoter, SP6 promoter, lac operator-promoter, tac promoter, SV40 late promoter, SV40 early promoter,

RSV-LTR promoter, CMV JE promoter, SV40 early promoter or SV40 late promoter and the CMV JE promoter.

16. Method of Detection

[0119] A method of detecting a target nucleic acid in a biological sample is also provided. The reverse-complement of the target nucleic acid may be synthesized. The target nucleic acid may then be detected in the biological sample by detecting binding of a probe comprising the target sequence to the synthesized reverse-complement. Binding of the probe to the reverse-complement may be performed using a microarray or by Luminex analysis. The probe may be attached to a biochip. The probe may also comprise locked nucleic acids, which may increase the sensitivity of detection

[0120] The ability to detect the target nucleic acid provides a wealth of information about the targeted nucleic acid, including whether the targeted nucleic acid is expressed in a particular cell, the level of the targeted nucleic acid expression in a particular cell or in comparison to other cells, identify the exact sequence of the targeted nucleic acid. Examples of methods based on the detection of a target nucleic acid, such as diagnostics and prognostics, are disclosed in U.S. Patent Application Nos. 10/536,560, 10/543,164, 11/130,645, 60/728,161 and 60/739,522; and International Patent Application Nos. PCT/IL2003/00970, PCT/IL2003/00998, PCT/IB2005/002352, PCT/IB2005/002702, PCT/US2005/16986 and PCT/US2005/14213, the contents of which are incorporated herein by reference.

EXAMPLE 1

Prediction Of miRNAs

[0121] We surveyed the entire human genome for potential miRNA coding genes using computational approaches similar to those described in U.S. Patent Application Nos. 60/522,459, 10/709,577 and 10/709,572, the contents of which are incorporated herein by reference, for predicting miRNAs. Briefly, non-protein coding regions of the entire human genome were scanned for hairpin structures. The predicted hairpins and potential miRNAs were scored by thermodynamic stability, as well as structural and contextual features. The algorithm was calibrated by using miRNAs in the Sanger Database which had been validated.

[0122] Table 1 lists the SEQ ID NO for each predicted hairpin ("HID") from the computational screen. Table 1 also lists the genomic location for each hairpin ("Hairpin Location"). The

format for the genomic location is a concatenation of <chr_id><strand><start position>. For example, 19+135460000 refers chromosome 19, +strand, start position 135460000.

Chromosomes 23-25 refer to chromosome X, chromosome Y and mitochondrial DNA, respectively. The chromosomal location is based on the hg17 assembly of the human genome by UCSC (<http://genome.ucsc.edu>), which is based on NCBI Build 35 version 1 and was produced by the International Human Genome Sequencing Consortium.

[0123] Table 1 also lists whether the hairpin is conserved in evolution (“C”). The hairpins were identified as conserved (“Y”) or nonconserved (“N”) by using phastCons data. The phastCons data is a measure of evolutionary conservation for each nucleotide in the human genome against the genomes of chimp, mouse, rat, dog, chicken, frog, and zebrafish, based on a phylo-HMM using best-in-genome pair wise alignment for each species based on BlastZ, followed by multiZ alignment of the 8 genomes (Siepel et al, J. Comput. Biol 11, 413-428, 2004 and Schwartz et al., Genome Res. 13, 103-107, 2003). A hairpin is listed as conserved if the average phastCons conservation score over the 7 species in any 15 nucleotide sequence within the hairpin stem is at least 0.9 (Berezikov, E. et al. Phylogenetic Shadowing and Computational Identification of Human microRNA Genes. Cell 120, 21-24, 2005).

[0124] Table 1 also lists the genomic type for each hairpin (“T”) as either intergenic (“G”), intron (“I”) or exon (“E”). Table 1 also lists the SEQ ID NO (“MID”) for each predicted miRNA and miRNA*. Table 1 also lists the prediction score grade for each hairpin (“P”) on a scale of 0-1 (1 the hairpin is the most reliable), as described in Hofacker et al., Monatshefte f. Chemie 125: 167-188, 1994. If the grade is zero or null, they are transformed to the lower value of PalGrade that its p-value is <0.05. Table 1 also lists the p-value (“Pval”) calculated out of background hairpins for the values of each P scores. As shown in Table 1, there are few instances where the Pval is >0.05. In each of these cases, the hairpins are highly conserved or they have been validated (F=Y).

[0125] Table 1 also lists whether the miRNAs were validated by expression analysis (“E”) (Y=Yes, N=No), as detailed in Table 2. Table 1 also lists whether the miRNAs were validated by sequencing (“S”) (Y=Yes, N=No). If there was a difference in sequences between the predicted and sequenced miRNAs, the sequenced sequence is predicted. It should be noted that failure to sequence or detect expression of a miRNA does not necessarily mean that a miRNA does not exist. Such undetected miRNAs may be expressed in tissues other than those tested. In

addition, such undetected miRNAs may be expressed in the test tissues, but at a different stage or under different condition than those of the experimental cells.

[0126] Table 1 also listed whether the miRNAs were shown to be differentially expressed ("D") (Y=Yes, N=No) in at least one disease, as detailed in Table 2). Table 1 also lists whether the miRNAs were present ("F") (Y=Yes, N=No) in Sanger DB Release 7.1 (October 2005) (<http://nar.oupjournals.org/>) as being detected in humans or mice or predicted in humans. As discussed above, the miRNAs listed in the Sanger database are a component of the prediction algorithm and a control for the output.

[0127] Table 1 also lists a genetic location cluster ("LC") for those hairpins that are within 5,000 nucleotides of each other. Each miRNA that has the same LC share the same genetic cluster. Table 1 also lists a seed cluster ("SC") to group miRNAs by their seed of 2-7 by an exact match. Each miRNA that has the same SC have the same seed. For a discussion of seed lengths of 5-6 nucleotides being sufficient for miRNA activity, see Lewis et al., *Cell*, 120;15-20 (2005).

EXAMPLE 2

Prediction of Target Genes

[0128] The predicted miRNAs from the computational screen of Example 1 was then used to predict target genes and their binding sites using two computational approaches similar to those described in U.S. Patent Application Nos. 60/522,459, 10/709,577 and 10/709,572, the contents of which are incorporated herein by reference, for predicting miRNAs.

[0129] Table 5 lists the predicted target gene for each miRNA (MID) and its hairpin (HID) from the computational screen. The names of the target genes were taken from NCBI Reference Sequence release 9 (<http://www.ncbi.nlm.nih.gov>; Pruitt et al., *Nucleic Acids Res*, 33(1):D501-D504, 2005; Pruitt et al., *Trends Genet.*, 16(1):44-47, 2000; and Tatusova et al., *Bioinformatics*, 15(7-8):536-43, 1999). Target genes were identified by having a perfect complimentary match of a 6 nucleotide miRNA seed (positions 2-7) that have an "A" after the seed on the UTR and/or an exact match in the nucleotide before the seed (total=7 nucleotides). For a discussion on identifying target genes, see Lewis et al., *Cell*, 120: 15-20, (2005). For a discussion of the seed being sufficient for binding of a miRNA to a UTR, see Lim Lau et al., (*Nature* 2005) and Brenneck et al, (*PLoS Biol* 2005).

[0130] Binding sites were then predicted using a filtered target genes dataset by including only those target genes that contained a UTR of a least 30 nucleotides. The binding site screen only considered the first 8000 nucleotides per UTR and considered the longest transcript when there were several transcripts per gene. A total of 14,236 transcripts were included in the dataset. Table 5 lists the SEQ ID NO for the predicted binding sites for each target gene as predicted from each miRNA ("MID"). The sequence of the binding site includes the 20 nucleotides 5' and 3' of the binding site as they are located on the spliced mRNA.

[0131] Table 6 shows the relationship between the miRNAs ("MID")/hairpins ("HID") and diseases by their target genes. The name of diseases are taken from OMIM. For a discussion of the rationale for connecting the host gene the hairpin is located upon to disease, see Baskerville and Bartel, *RNA*, 11: 241–247 (2005) and Rodriguez et al., *Genome Res.*, 14: 1902–1910 (2004). Table 6 shows the number of miRNA target genes ("N") that are related to the disease. Table 6 also shows the total number of genes that are related to the disease ("T"), which is taken from the genes that were predicted to have binding sites for miRNAs. Table 6 also shows the percentage of N out of T ("P") and the p-value of hypergeometric analysis ("Pval"). Table 10 shows the disease codes for the diseases described in Tables 6-7 and 11-12. For a reference of hypergeometric analysis, see Schaum's Outline of Elements of Statistics II: Inferential Statistics.

[0132] Table 7 shows the relationship between the miRNAs ("MID")/hairpins ("HID") and diseases by their host genes. We defined hairpins genes on the complementary strand of a host gene as located on the gene: Intron_c as Intron and Exon_c as Exon. We choose the complementary strands as they can cause disease. For example, a mutation in the miRNA that is located on the complementary strand. In those case that a miRNA in on both strands, two statuses like when Intron and Exon_c Intron is the one chosen. The logic of choosing is Intron>Exon>Intron_c>Exon_c>Intergenic. Table 11 shows the relationship between the target sequences ("Gene Name") and disease ("Disease Code").

[0133] Table 12 shows the relationship between the miRNAs ("MID")/hairpins ("HID"), known SNPs and diseases. SNP were identified in the sequence of hairpins. For the miRNAs of these hairpins, all their target genes listed in Table 5 were collected. For these genes, we checked whether they are associates to disease(s) according to OMIM. The numeric code of the relevant diseases for each miRNA according to Table 10 are presented in Table 12. The disease codes are taken from Table 10. Each SNP ("SNP_Id ") is identified based on NCBI database dbSNP

BUILD 123 based on NCBI Human Genome Build 35. The genomic location for each SNP ("SNP_location") is also provided in a formation concatenating "<chr_id>:<start position>". For example, "19:135460569" means chr19 +strand, start position 135460569. Although the mutations are referred to as SNPs, a number of the mutations cover a few nucleotides (e.g., small insertions, deletions, micro-satellites, etc.) For a discussion on the connection between a SNP and disease, see Swibertus (Blood 1996) and Frittitta (Diabetes 2001).

EXAMPLE 3

Validation of miRNAs

1. Chip Expression

[0134] To confirm the hairpins and miRNAs predicted in Example 1, we detected expression in various tissues using the high-throughput microarrays similar to those described in U.S. Patent Application Nos. 60/522,459, 10/709,577 and 10/709,572, the contents of which are incorporated herein by reference. For each predicted precursor miRNA, mature miRNAs derived from both stems of the hairpin were tested.

[0135] Table 2 shows the hairpins ("HID") of the second prediction set that were validated by detecting expression of related miRNAs ("MID"), as well as a code for the tissue ("Tissue") that expression was detected. The tissue and diseases codes for Table 2 are listed in Table 8. Some of the tested tissues were cell lines. Lung carcinoma cell line (H1299) with/without P53: H1299 has a mutated P53. The cell line was transfected with a construct with P53 that is temperature sensitive (active at 32°C). The experiment was conducted at 32°C.

[0136] Table 2 also shows the chip expression score grade ("S")(range of 500-65000). A threshold of 500 was used to eliminate non-significant signals and the score was normalized by MirChip probe signals from different experiments. Variations in the intensities of fluorescence material between experiments may be due to variability in RNA preparation or labeling efficiency. We normalized based on the assumption that the total amount of miRNAs in each sample is relatively constant. First we subtracted the background signal from the raw signal of each probe, where the background signal is defined as 400. Next, we divided each miRNA probe signal by the average signal of all miRNAs, multiplied the result by 10000 and added back the background signal of 400. Thus, by definition, the sum of all miRNA probe signals in each experiment is 10400.

[0137] Table 2 also shows a statistical analysis of the normalized signal (“Spval”) calculated on the normalized score. For each miRNA, we used a relevant control group out of the full predicted miRNA list. Each miRNA has an internal control of probes with mismatches. The relevant control group contained probes with similar C and G percentage (abs diff < 5%) in order to have similar T_m. The probe signal P value is the ratio over the relevant control group probes with the same or higher signals. The results are p-value ≤ 0.05 and score is above 500. In those cases that the SPVal is listed as 0.0, the value is less than 0.0001. The data was obtained using a chip an internal control of probes with mismatches, which were checked for each significant signal that was affected in the mutated probes.

2. Northern Blot

[0138] To further confirm the hairpins and miRNAs predicted in Example 1, we detected expression in additional tissues by Northern Blot. Ten Northern blots were performed to assess the expression of certain miRNAs in three human tissues (brain, placenta and testis). 5 μ g of total RNA from human brain, placenta and testis (Ambion) was fractionated by PAGE using a 15% denaturing polyacrylamide gel. The RNA was transferred to positively charged nylon membranes by electroblotting at 200 mA in 0.5x TBE for 2 hours. The blots were dried and incubated overnight in separate hybridization bottles with 10 ml of ULTRAhyb-Oligo (Ambion) and 10⁷ cpm of radio-labeled oligonucleotides complementary to the predicted miRNAs. The blots were washed 3x10 min at room temperature in 2x SSC, 0.5% SDS and then 1x15 min at 42°C in 2x SSC, 0.5% SDS. Overnight phosphorimaging using the Storm system (Amersham) revealed probes hybridizing to the predicted miRNAs.

3. Sequencing

[0139] To further validate the hairpins (“HID”) of the second prediction, a number of miRNAs were validated by sequencing methods similar to those described in U.S. Patent Application Nos. 60/522,459, 10/709,577 and 10/709,572, the contents of which are incorporated herein by reference. Table 4 shows the hairpins (“HID”) that were validated by sequencing a miRNA (MID) in the indicated tissue (“Tissue”). Numeric codes for the tissues are shown in Table 8.

EXAMPLE 4

MiRNAs of Chromosome 19

[0140] A group of the validated miRNAs from Example 3 were highly expressed in placenta, have distinct sequence similarity, and are located in the same locus on chromosome 19 (Figure 2). These predicted miRNAs are spread along a region of ~100,000 nucleotides in the 19q13.42 locus. This genomic region is devoid of protein-coding genes and seems to be intergenic. Further analysis of the genomic sequence, including a thorough examination of the output of our prediction algorithm, revealed many more putative related miRNAs, and located mir-371, mir-372, and mir-373 approximately 25,000bp downstream to this region. Overall, 54 putative miRNA precursors were identified in this region. The miRNA precursors can be divided into four distinct types of related sequences (Figure 2). About 75% of the miRNAs in the cluster are highly related and were labeled as type A. Three other miRNA types, types B, C and D, are composed of 4, 2, and 2 precursors, respectively. An additional 3 putative miRNA precursors (S1 to S3) have unrelated sequences. Interestingly, all miRNA precursors are in the same orientation as the neighboring mir-371, mir-372, and mir-373 miRNA precursors.

[0141] Further sequence analysis revealed that the majority of the A-type miRNAs are embedded in a ~600bp region that is repeated 35 times in the cluster. The repeated sequence does not appear in other regions of the genome and is conserved only in primates. The repeating unit is almost always bounded by upstream and downstream Alu repeats. This is in sharp contrast to the MC14-1 cluster which is extremely poor in Alu repeats.

[0142] Figure 3-A shows a comparison of sequences of the 35 repeat units containing the A-type miRNA precursors in human. The comparison identified two regions exhibiting the highest sequence similarity. One region includes the A-type miRNA, located in the 3' region of the repeat. The second region is located ~100 nucleotides upstream to the A-type miRNA precursors. However, the second region does not show high similarity among the chimp repeat units while the region containing the A-type miRNA precursors does (Figure 3-B).

[0143] Examination of the region containing the A-type repeats showed that the 5' region of the miRNAs encoded by the 5' stem of the precursors (5p miRNAs) seem to be more variable than other regions of the mature miRNAs. This is matched by variability in the 3' region of the mature miRNAs derived from the 3' stems (3p miRNAs). As expected, the loop region is highly

variable. The same phenomenon can also be observed in the multiple sequence alignment of all 43 A-type miRNAs (Figure 4).

[0144] The multiple sequence alignment presented in Figure 4 revealed the following findings with regards to the predicted mature miRNAs. The 5p miRNAs can be divided into 3 blocks. Nucleotides 1 to 6 are C/T rich, relatively variable, and are marked in most miRNAs by a CTC motif in nucleotides 3 to 5. Nucleotides 7 to 15 are A/G rich and apart from nucleotides 7 and 8 are shared among most of the miRNAs. Nucleotides 16 to 23 are C/T rich and are, again, conserved among the members. The predicted 3p miRNAs, in general, show a higher conservation among the family members. Most start with an AAA motif, but a few have a different 5' sequence that may be critical in their target recognition. Nucleotides 8 to 15 are C/T rich and show high conservation. The last 7 nucleotides are somewhat less conserved but include a GAG motif in nucleotides 17 to 19 that is common to most members.

[0145] Analysis of the 5' region of the repeated units identified potential hairpins. However, in most repeating units these hairpins were not preserved and efforts to clone miRNAs from the highest scoring hairpins failed. There are 8 A-type precursors that are not found within a long repeating unit. Sequences surrounding these precursors show no similarity to the A-type repeating units or to any other genomic sequence. For 5 of these A-type precursors there are Alu repeats located significantly closer downstream to the A-type sequence.

[0146] The other miRNA types in the cluster showed the following characteristics. The four B group miRNAs are found in a repeated region of ~500bp, one of which is located at the end of the cluster. The two D-type miRNAs, which are ~2000 nucleotides from each other, are located at the beginning of the cluster and are included in a duplicated region of 1220 nucleotides. Interestingly, the two D-type precursors are identical. Two of the three miRNAs of unrelated sequence, S1 and S2, are located just after the two D-type miRNAs, and the third is located between A34 and A35. In general, the entire ~100,000 nucleotide region containing the cluster is covered with repeating elements. This includes the miRNA-containing repeating units that are specific to this region and the genome wide repeat elements that are spread in the cluster in large numbers.

EXAMPLE 5

Cloning Of Predicted MiRNAs

[0147] To further validate the predicted miRNAs, a number of the miRNAs described in Example 4 were cloned using methods similar to those described in U.S. Patent Application Nos. 60/522,459, 10/709,577 and 10/709,572, the contents of which are incorporated herein by reference. Briefly, a specific capture oligonucleotide was designed for each of the predicted miRNAs. The oligonucleotide was used to capture, clone, and sequence the specific miRNA from a placenta-derived library enriched for small RNAs.

[0148] We cloned 41 of the 43 A-type miRNAs, of which 13 miRNAs were not present on the original microarray but only computationally predicted, as well as the D-type miRNAs. For 11 of the predicted miRNA precursors, both 5p and 3p predicted mature miRNAs were present on the microarray and in all cases both gave significant signals. Thus, we attempted to clone both 5' and 3' mature miRNAs in all cloning attempts. For 27 of the 43 cloned miRNA, we were able to clone miRNA derived from both 5' and 3' stems. Since our cloning efforts were not exhaustive, it is possible that more of the miRNA precursors encode both 5' and 3' mature miRNAs.

[0149] Many of the cloned miRNAs have shown heterogeneity at the 3' end as observed in many miRNA cloning studies (Lagos-Quintana 2001, 2002, 2003) (Poy 2004). Interestingly, we also observed heterogeneity at the 5' end for a significant number of the cloned miRNAs. This heterogeneity seemed to be somewhat more prevalent in 5'-stem derived miRNAs (9) compared to 3'-stem derived miRNAs (6). In comparison, heterogeneity at the 3' end was similar for both 3' and 5'-stem derived miRNAs (19 and 13, respectively). The 5' heterogeneity involved mainly addition of one nucleotide, mostly C or A, but in one case there was an addition of 3 nucleotides. This phenomenon is not specific to the miRNAs in the chromosome 19 cluster. We have observed it for many additional cloned miRNAs, including both known miRNAs as well as novel miRNAs from other chromosomes (data not shown).

EXAMPLE 6

Analysis Of MiRNA Expression

[0150] To further examine the expression of the miRNAs of Example 4, we used Northern blot analysis to profile miRNA expression in several tissues. Northern blot analysis was performed using 40µg of total RNA separated on 13% denaturing polyacrylamide gels and using ³²P end

labeled oligonucleotide probes. The oligonucleotide probe sequences were 5'-ACTCTAAAGAGAAGCGCTTTGT-3' (A19-3p, NCBI: HSA-MIR-RG-21) and 5'-ACCCACCAAAGAGAAGCACTTT-3' (A24-3p, NCBI: HSA-MIR-RG-27). The miRNAs were expressed as ~22 nucleotide long RNA molecules with tissue specificity profile identical to that observed in the microarray analysis (Figure 5-A).

[0151] In order to determine how the MC19-1 cluster is transcribed. A survey of the ESTs in the region identified only one place that included ESTs with poly-adenylation signal and poly-A tail. This region is located just downstream to the A43 precursor. The only other region that had ESTs with poly-adenylation signal is located just after mir-373, suggesting that mir-371,2,3 are on a separate transcript. We performed initial studies focusing on the region around mir-A43 to ensure that the region is indeed transcribed into poly-adenylated mRNA. RT-PCR experiments using primers covering a region of 3.5kb resulted in obtaining the expected fragment (Figure 5-B). RT-PCR analysis was performed using 5µg of placenta total RNA using oligo-dT as primer. The following primers were used to amplify the transcripts: f1:

5'-GTCCCTGTACTGGAAGTTGAG-3'; f2: 5'-GTGTCCCTGTACTGGAACGCA-3'; r1: 5'-GCCTGGCCATGTCAGCTACG-3'; r2: 5'-TTGATGGGAGGCTAGTGTTTC-3'; r3: 5'-GACGTGGAGGCGTTCTTAGTC-3'; and r4: 5'-TGACAACCGTTGGGGATTAC-3'. The authenticity of the fragment was validated by sequencing. This region includes mir-A42 and mir-A43, which shows that both miRNAs are present on the same primary transcript.

[0152] Further information on the transcription of the cluster came from analysis of the 77 ESTs located within it. We found that 42 of the ESTs were derived from placenta. As these ESTs are spread along the entire cluster, it suggested that the entire cluster is expressed in placenta. This observation is in-line with the expression profile observed in the microarray analysis. Thus, all miRNAs in the cluster may be co-expressed, with the only exception being the D-type miRNAs which are the only miRNAs to be expressed in HeLa cells. Interestingly, none of the 77 ESTs located in the region overlap the miRNA precursors in the cluster. This is in-line with the depletion of EST representation from transcripts processed by Drosha.

[0153] Examination of the microarray expression profile revealed that miRNAs D1/2, A12, A21, A22, and A34, have a somewhat different expression profile reflected as low to medium expression levels in several of the other tissues examined. This may be explained by alternative

splicing of the transcript(s) encoding the miRNAs or by the presence of additional promoter(s) of different tissues specificity along the cluster.

[0154] Comparison of the expression of 3p and 5p mature miRNAs revealed that both are expressed for many miRNA precursors but in most cases at different levels. For most pre-miRNAs the 3p miRNAs are expressed at higher levels than the 5p miRNAs. However, in 6 cases (mir-D1,2, mir-A1, mir-A8, mir-A12, mir-A17 and mir-A33) both 3p and 5p miRNAs were expressed at a similar level, and in one case (mir-A32) the 5p miRNA was expressed at higher levels than the 3p miRNA.

EXAMPLE 7

Conservation

[0155] Comparison of the sequences from all four types of predicted miRNAs of Example 4 to that of other species (chimp, macaque, dog, chicken, mouse, rat, drosophila, zebra-fish, fungi, c. elegans) revealed that all miRNAs in the cluster, and in fact the entire region, are not conserved beyond primates. Interestingly, homologues of this region do not exist in any other genomes examined, including mouse and rat. Thus, this is the first miRNA cluster that is specific to primates and not generally shared in mammals. Homology analysis between chimp and human show that all 35 repeats carrying the A-type miRNAs are contiguous between the two species. Furthermore, the entire cluster seems to be identical between human and chimp. Thus, the multiple duplications leading to the emergence of the MC19-1 cluster must have occurred prior to the split of chimp and human and remained stable during the evolution of each species. It should be noted that human chromosome 19 is known to include many tandemly clustered gene families and large segmental duplications (Grimwood et al, 2004). Thus, in this respect the MC19-1 cluster is a natural part of chromosome 19.

[0156] In comparison, the MC14-1 cluster is generally conserved in mouse and includes only the A7 and A8 miRNAs within the cluster are not conserved beyond primates (Seitz 2004). In contrast all miRNAs in the MC19-1 cluster are unique to primates. A survey of all miRNAs found in Sanger revealed that only three miRNA, mir-198, mir-373, and mir-422a, are not conserved in the mouse or rat genomes, however, they are conserved in the dog genome and are thus not specific to primates. Interestingly, mir-371 and mir-372, which are clustered with mir-

373, and are located 25kb downstream to the MC19-1 cluster, are homologous to some extent to the A-type miRNAs (Figure 4), but are conserved in rodents.

[0157] Comparison of the A-type miRNA sequences to the miRNAs in the Sanger database revealed the greatest homology to the human mir-302 family (Figure 4-C). This homology is higher than the homology observed with mir-371,2,3. The mir-302 family (mir-302a, b, c, and d) are found in a tightly packed cluster of five miRNAs (including mir-367) covering 690 nucleotides located in the antisense orientation in the first intron within the protein coding exons of the HDCMA18P gene (accession NM_016648). No additional homology, apart from the miRNA homology, exists between the mir-302 cluster and the MC19-1 cluster. The fact that both the mir-371,2,3 and mir-302a,b,c,d are specific to embryonic stem cells is noteworthy.

EXAMPLE 8

Differential Expression of miRNAs

[0158] Using chip expression methods similar to those described in Example 3, microarray images were analyzed using Feature Extraction Software (Version 7.1.1, Agilent). Table 3 shows the ratio of disease related expression ("R") compared to normal tissues for the indicated diseases. Disease codes for the disease are shown in Table 9. Table 3 also shows the statistical analysis of the normalized signal ("RPval"). The signal of each probe was set as its median intensity. Signal intensities range from background level of 400 to saturating level of 66000. 2 channels hybridization was performed and Cy3 signals were compared to Cy5 signals, where fluor reversed chip was preformed (normal vs. disease), probe signal was set to be its average signal. Signals were normalized by dividing them with the known miRNAs average signals such that the sum of known miRNAs signal is the same in each experiment or channel. Signal ratios between disease and normal tissues were calculated. Signal ratio greater than 1.5 indicates a significant upregulation with a P value of 0.007 and signal ratio grater than 2 has P value of 0.003. P values were estimated based on the occurrences of such or greater signal ratios over duplicated experiments. The differential expression analysis in Table 3 indicates that the expression of a number of the miRNAs are significantly altered in disease tissue.

EXAMPLE 9

Expression Analysis Of Cancer-Associated MiRNAs

[0159] One of the predicted and validated human miRNA genes is the 13_12 gene (SEQ ID NO: 1A), which is located at residues 14305325-14305407 of the + strand of chromosome 16, with reference to version HG17 of the human genome (References to SEQ ID NOS or MID numbers ending with "A" refer to the correspond label without the "A" in U.S. Patent Application No. 11/275,628, which are incorporated herein by reference). Figure 6 shows the predicted hairpin formed by the pre-miRNA (SEQ ID NO: 2A) encoded by 13_12, with the residues of the predicted miRNA and miRNA* indicated with underlining.

[0160] The 13_12 gene (annotated as has-mir-193b at www.sanger.ac.uk) is homologous to the following predicted miRNAs: has-mir-193a (human), mmu-mir-193 (mouse), rno-mir-193 (rat), dre-mir-193a (zebra fish) and dre-mir-193b (zebra fish). An alignment of the genes encoding these miRNAs (Figure 7) indicates that both the 5' and 3' mature miRNAs have high similarity and the 3' mature miRNAs have identical seeds, which may be important for target recognition. Analysis of the hairpin encoded by the 13_12 gene using the UCSC Genome Browser (genome.ucsc.edu/cgi-bin/hgGateway) is highly conserved in humans, chimpanzees, dogs, mice, rats, chickens, fungi and zebra fish.

[0161] The physical location of the 13_12 gene is near has-mir-365-1 (has-mir-193a is also near has-mir-365-2). Interestingly, the 13_12 gene and has-mir-365-1 are apparently expressed from the same transcript based on reviewing the human mRNA and human EST databases using the UCSC Genome Browser. The coexpression of miRNAs on a single transcript has been previously reported (Baskerville et al., RNA, 11(3):241-7, 2005). CpG islands are also present near the beginning of the transcript.

[0162] We cloned a number of mature miRNAs derived from the 13_12 gene using a placenta library (Table 13). An alignment of the cloned sequences (Figure 4) shows that there is variability at the 5' and 3' ends, which is not unexpected.

Table 13

MID	Sequence	SEQ ID NO
19061A	AACUGGCCCUCAAAGUCCCGC	3A
19062A	AACUGGCCCUCAAAGUCCCGCU	4A
19063A	AACUGGCCCUCAAAGUCCCGCUUU	5A
19064A	ACUGGCCCUCAAAGUCCCGCUUU	6A
19065A	CCAACUGGCCCUCAAAGUCCCGCUU	7A

[0163] To further characterize gene 13_12, we performed a Northern blot analysis of its expression in several additional tissues. Figure 9 shows that 13_12 miRNA was expressed in brain, liver, thymus, HeLa cells (a uterus carcinoma cell line) and placenta. The 13_12 was also shown to be expressed in adipose.

[0164] The expression of 13_12 in cancer tissues was next investigated. Using methods similar to those described in U.S. Patent Application Nos. 60/522,459, 10/709,577 and 10/709,572, the contents of which are incorporated herein by reference, the level of expression of 13_12 and other miRNA genes was determined in human prostate adenocarcinoma and lung adenocarcinoma and compared to the level of expression in normal prostate or lung tissue.

[0165] Figure 10 indicates that 13_12 is overexpressed in both prostate and lung cancer tissues. 13_12 was expressed approximately 39.93-fold higher in prostate tumor cells (chip signal_{tumor}=196796.3 and chip signal_{normal}=4957.4). 13_12 was also overexpressed by approximately 9.04-fold in A549 lung tumor cells (chip signal_{tumor}=12389.3 and chip signal_{normal}=1369.0).

EXAMPLE 10

Antisense Inhibition of a Cancer-Associated miRNA in Prostate Cancer Cells

[0166] To gain a better understanding of the role that 13_12-encoded miRNA plays in cellular pathways, we utilized anti-sense molecules to inhibit precursors of 13_12-encoded miRNAs and compared the effects to inhibition of other miRNAs. The sequence of the antisense inhibitors tested is shown in Table 14, where all nucleotides in the inhibitors contain 2'-O-methyl (2'-OMe) modifications at every base.

Table 14

Name	Target	Sequence	SEQ ID NO
193	has-miR-193a	GAGAACUGGGACUUUGUAGGCCAGUUGAUCC	13A
14-3-5	14_3 14_5 14_5rel	CGGUAACMCUCUAAAAGGAWGCACUUKCUUUU	14A
13-12	has-miR-193b	CCCAAAGCGGGACUUUGAGGGCCAGUUGGAU	15A
3-7-8	33_8 33_8 rel 33_7	GACAASAGAAAGUGCUUCCCUCHAGAGGGUCA	16A
33-3-14-2	33_3 14_12	ACAGUAACACUCUAAAARGGAUGCACGAUCUUU	17A
24-2	24_2	CAAAGUCUCAGUUUCCUCUGCAAACAGUUGUA	18A
eGFP	GFP	CACCAACGGCAAGCUGACCCUGAAGUUCAUC	19A

M=A or C, W=A or U, K=G or U, S=C or G, H=A, C or U, R=A or G

[0167] The miRNA inhibitors were used to test the effect on cell growth by inhibiting miRNA activity in PC-3 prostate cancer cells. Although expression of 13_12 was not able to be detected in PC-3 cells using the methods described above, PC-3 prostate cancer cells were nonetheless used to assay the effects of the miRNA inhibitors because the detection methods used were not very sensitive. The inhibitors were tested by transfecting PC-3 cells in six replicate wells. For all transfections, wells contained 5,000-6,000 cells, 250 nM inhibitor and 0.3 µl LIPOFECTAMIN 2000 (#11668-019 Invitrogen). After 72 h the samples were assayed with MTT (CellTiter 96 non radioactive cell pro. Assay #G400, Promega), in order to determine the number of viable cells by measuring specific absorbance at 570 nm in an ELIZA reader. The results were graphed and normalized against a negative control miRNA inhibitor eGFP targeting green fluorescent protein (GFP). The eGFP inhibitor is the same sequence as a section of the GFP mRNA and functionality inhibits eGFP siRNA activity (data not shown).

[0168] We found that inhibition of 13_12 caused a significant decrease in the proliferation of PC-3 cells (Figure 11). The results were confirmed in a total of six separate experiments, showing that the 13_12 inhibitor reduced PC-3 cell growth by 30-70%. Figure 12 shows the comparative inhibition of the 13_12 inhibitor and other miRNA inhibitors in PC-3 cells. The 13_12 inhibitor demonstrated the highest level of inhibition of PC-3 cell growth.

EXAMPLE 11

Antisense Inhibition of a Cancer-Associated miRNA in Lung Cancer Cells

[0169] As discussed above, the 13_12 gene is homologous to the gene encoding has-mir-193a. Inhibitors of has-mir-193a have been shown to cause a decrease of 50%-80% in the relative cell growth of A549 lung cancer cells. *See* Cheng et al., *Nucleic Acids Res.* 33(4):1290-7 (2005). We next compared the inhibition of the 13_12-encoded miRNA to inhibition of has-mir-193a in A549 lung cancer cells.

[0170] A549 cells were transfected with inhibitors 193 and 13_12 and affects on cell growth measured using methods similar to those previously discussed in Example 2. The results in Figure 13 indicate that inhibition of 13_12 is at least as effective as inhibition of has-mir-193a in reducing proliferation of A549 lung cancer cells. These results are in contrast to the effect of the inhibitors in PC-3 cells discussed above, where the 193 inhibitor did not inhibit cell growth.

EXAMPLE 12

Direct Transcription of miRNA

[0171] Total RNA is fractioned onto YM-100 columns to separate the small RNA molecules. The small RNA molecules are then precipitated in EtOH overnight at 40°C and resuspended in a buffer suitable for ligation of the resuspended RNA. An adaptor with SEQ ID NO: 1B is then added to the mixture and ligation is carried on using T4 ligase (NEB) in ligase buffer including DMSO (15%)(References to SEQ ID NOS ending with "B" refer to the correspond label without the "B" in U.S. Provisional Patent Application No. 60/743,098, which are incorporated herein by reference). The reaction may be spiked with known RNA. The ligation products are then separated by electrophoresis in acrylamide gel. After the separation is completed, the ligation products are cut out from the gel and extracted out of the gel using GebaFlex electrophoration dialysis tubes. The extraction is followed by overnight precipitation in EtOH. A DNA oligo with SEQ ID NO: 2B is then annealed to T7 promoter to create a double stranded complex at the T7 promoter region. Annealing is controlled by step decrease of temperature from 70°C to 25°C, after 2 minutes at 85°C.

[0172] Transcription is then performed using the Megashortscript kit (Ambion) for transcription of short sequences. The transcription can be carried on using biotinylated UTP and or biotinylated CTP to produce a labeled RNA transcript. The adaptor is then digested with DNase

for 15 minutes at 37°C and final RNA transcript is then purified by chloroform/phenol phase extraction and EtOH precipitation overnight.

CLAIMS

1. An isolated nucleic acid comprising a sequence selected from the group consisting of:
 - (a) any one of SEQ ID NOS: 1-42840282
 - (b) complement of (a); and
 - (c) sequence at least about 81% identical to 21 contiguous nucleotides of (a) or (b);wherein the nucleic acid is from 17-250 nucleotides in length.
2. The nucleic acid of claim 1 comprising a sequence selected from the group consisting of:
 - (a) any one of SEQ ID NOS: 3A-7A;
 - (b) complement of (a); and
 - (c) sequence at least about 81% identical to 21 contiguous nucleotides of (a) or (b);wherein the nucleic acid is from 17-250 nucleotides in length.
3. The nucleic acid of claim 2 wherein the nucleic acid comprises a sequence selected from the group consisting of:
 - (a) SEQ ID NO: 1A;
 - (b) SEQ ID NO: 2A;
 - (c) complement of (a) or (b); and
 - (d) sequence at least about 63% identical to 81 contiguous nucleotides of (a), (b) or (c);wherein the nucleic acid is from 51-250 nucleotides in length.
4. The nucleic acid of any one of claims 1-3 wherein the nucleic acid comprises a modified base.
5. A probe comprising the nucleic acid of claim 4.
6. A composition comprising the probe of claim 5.
7. A biochip comprising the probe of claim 5.
8. A method for detecting a disease-associated nucleic acid comprising:
 - (a) providing a biological sample; and

(b) measuring the level of a nucleic acid according to claim 1, wherein a level of the nucleic acid higher than a control is indicative of (a) disease-associated nucleic acid is detected of the level of (b) is above that of a control being detected.

9. A method for identifying a compound that modulates expression of a disease-associated miRNA:

- (a) providing a cell that is capable of expressing a nucleic acid according to claim 1;
- (b) contacting the cell with a candidate modulator; and
- (c) measuring the level of expression of the nucleic acid,

wherein a difference in the level of the nucleic acid compared to a control identifies the compound as a modulator of expression of the miRNA.

10. A method of inhibiting expression of a target gene in a cell comprising introducing a nucleic acid into the cell in an amount sufficient to inhibit expression of the target gene, wherein the target gene comprises a binding site substantially identical to a binding site referred to in Table 5 or any one of SEQ ID NOS 26519-42840282, and wherein the nucleic acid is a nucleic acid according to claim 1.

11. The method of claim 9 wherein expression is inhibited in vitro or in vivo.

12. A method of increasing expression of a target gene in a cell comprising introducing a nucleic acid into the cell in an amount sufficient to increase expression of the target gene, wherein the target gene comprises a binding site substantially identical to a binding site referred to in 5 or any one of SEQ ID NOS 26519-42840282, wherein the nucleic acid comprises a sequence substantially complementary to a nucleic acid according to claim 1.

13. The method of claim 12 wherein expression is inhibited in vitro or in vivo.

14. A method of treating a patient with a disorder set forth on Table 10 comprising administering to a patient in need thereof a composition comprising the nucleic acid of claim 1.

15. A method of synthesizing the reverse-complement of a target nucleic acid comprising:

- (a) providing a first nucleic acid comprising the target nucleic acid and an adapter nucleic acid, wherein the target nucleic acid is 5' of the second nucleic acid;

- (b) providing a second nucleic acid, wherein a portion of the second nucleic acid is substantially complementary to a portion of the adapter nucleic acid;
- (c) annealing the second nucleic acid and the adapter nucleic acid of the first nucleic acid, whereby an annealed complex is formed;
- (d) extending the second nucleic acid from its 3' end, whereby the reverse-complement of the target nucleic acid is synthesized.

16. The method of claim 15, wherein first nucleic acid is formed by ligating the target nucleic acid and the adapter nucleic acid.

17. The method of claim 16, wherein the target nucleic acid is derived from a biological sample.

18. The method of claim 17, wherein the target nucleic acid is size-fractionated.

19. The method of claim 15 further comprising

- (a) displacing the extended second nucleic acid from the first nucleic acid;
- (b) providing a second nucleic acid, wherein a portion of the second nucleic acid is substantially complementary to a portion of the adapter nucleic acid;
- (c) annealing the second nucleic acid and the adapter nucleic acid of the first nucleic acid, whereby an annealed complex is formed;
- (d) extending the second nucleic acid from its 3' end, whereby the reverse-complement of the target nucleic acid is synthesized.

20. The method of claim 15, wherein the adapter nucleic acid comprises DNA.

21. The method of claim 20, wherein the DNA comprises 10 to 20 nucleotides.

22. The method of claim 15, wherein the adapter nucleic acid comprises RNA.

23. The method of claim 22, wherein the RNA comprises 1 to 5 nucleotides.

24. The method of claim 15, wherein the adapter nucleic acid comprises DNA and RNA, and wherein the RNA is 5' of the DNA.

25. The method of claim 15, wherein the adapter nucleic acid comprises a 5'-phosphate.

26. The method of claim 15, wherein the adapter nucleic acid comprises a 3' -IdT.

27. The method of claim 15, wherein the second nucleic acid comprises DNA.

28. The method of claim 19, wherein 5(a)-5(d) are repeated 10-1000 times.

29. The method of claim 15, wherein the displacing is carried out by heat denaturation.

30. The method of claim 15, wherein the second nucleic acid is extended using a suitable polymerase.

31. The method of claim 30, wherein the polymerase is selected from the group consisting of T7 RNA polymerase, T3 RNA polymerase or SP6 RNA polymerase.

32. The method of claim 15, wherein a double-stranded portion of the annealed complex comprises a promoter for T7 RNA polymerase.

33. The method of claim 31, wherein both strands of the promoter are DNA.

34. The method of claim 15, wherein the target nucleic acid comprises RNA.

35. The method of claim 34, wherein the RNA comprises 18 to 24 nucleotides.

36. The method of claim 35, wherein the RNA is a miRNA.

37. A method of detecting a target nucleic acid in a biological sample comprising:

- (a) providing a biological sample comprising nucleic acid;
- (b) synthesizing the reverse-complement of the nucleic acid according to the method of claim 15;
- (c) providing a biochip comprising a probe, wherein the probe comprises the target nucleic acid;
- (d) measuring the binding of the probe to reverse-complement synthesized nucleic acid, wherein a level of binding higher than a control identifies the target nucleic acid in the biological sample.

38. The method of claim 37 wherein the probe comprises locked nucleic acids.

39. The method of claim 37 wherein detection is carried out using Luminex.

FIGURE 1

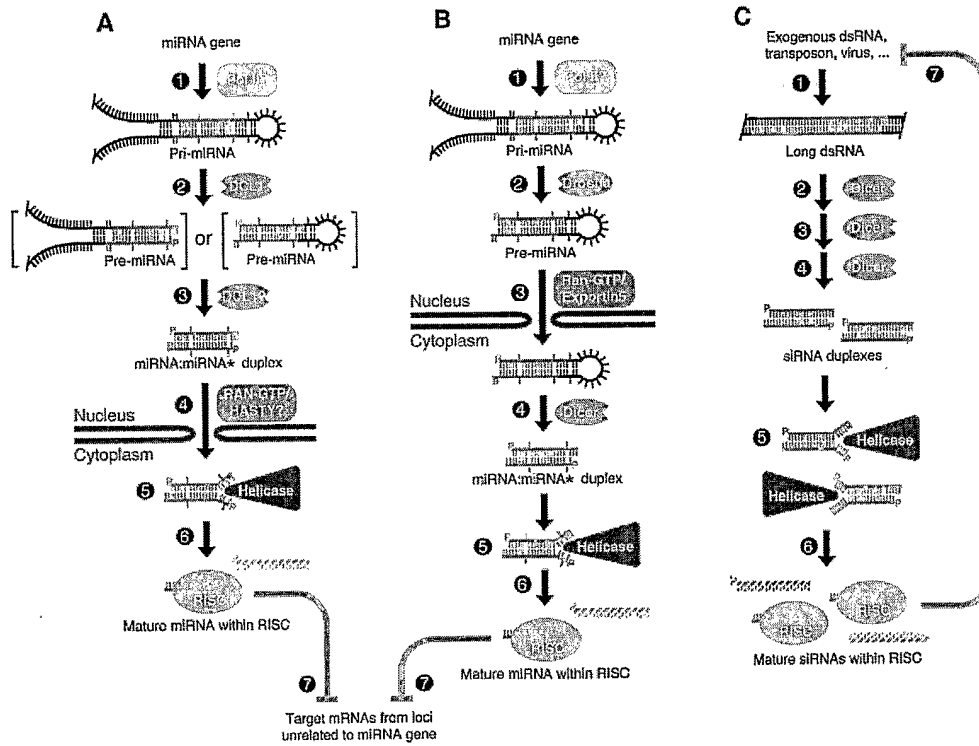


FIGURE 2

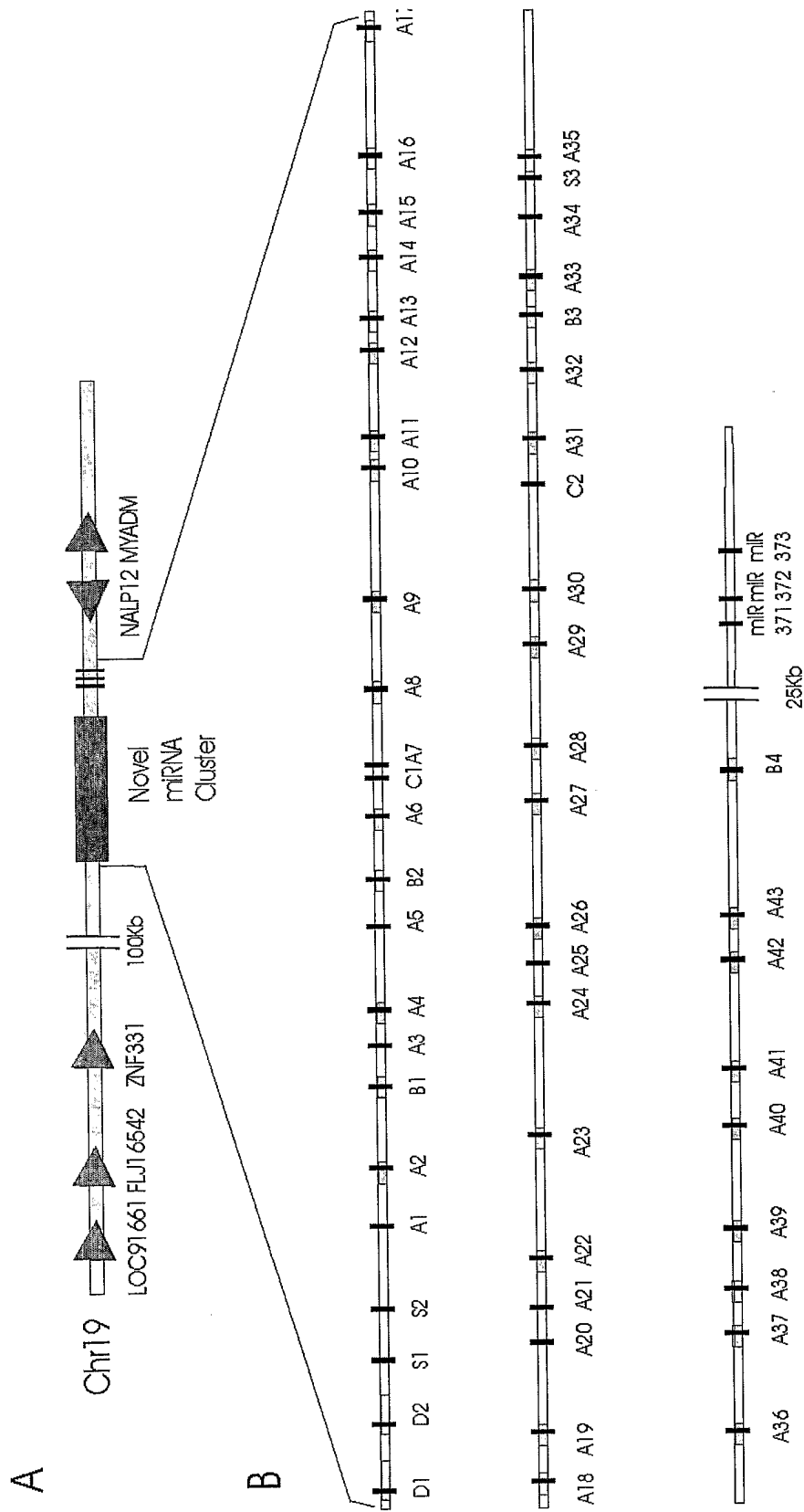


FIGURE 3

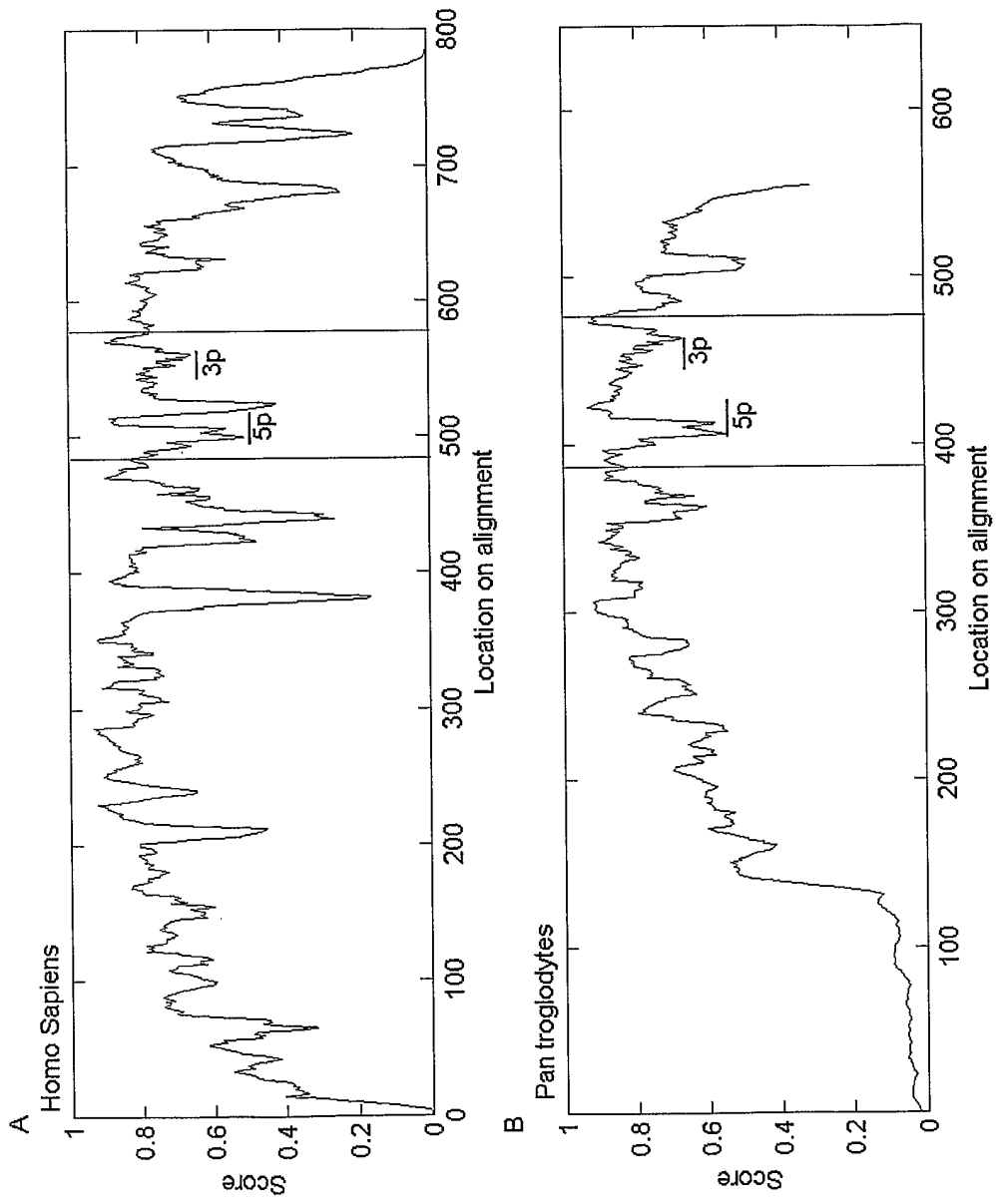


FIGURE 4

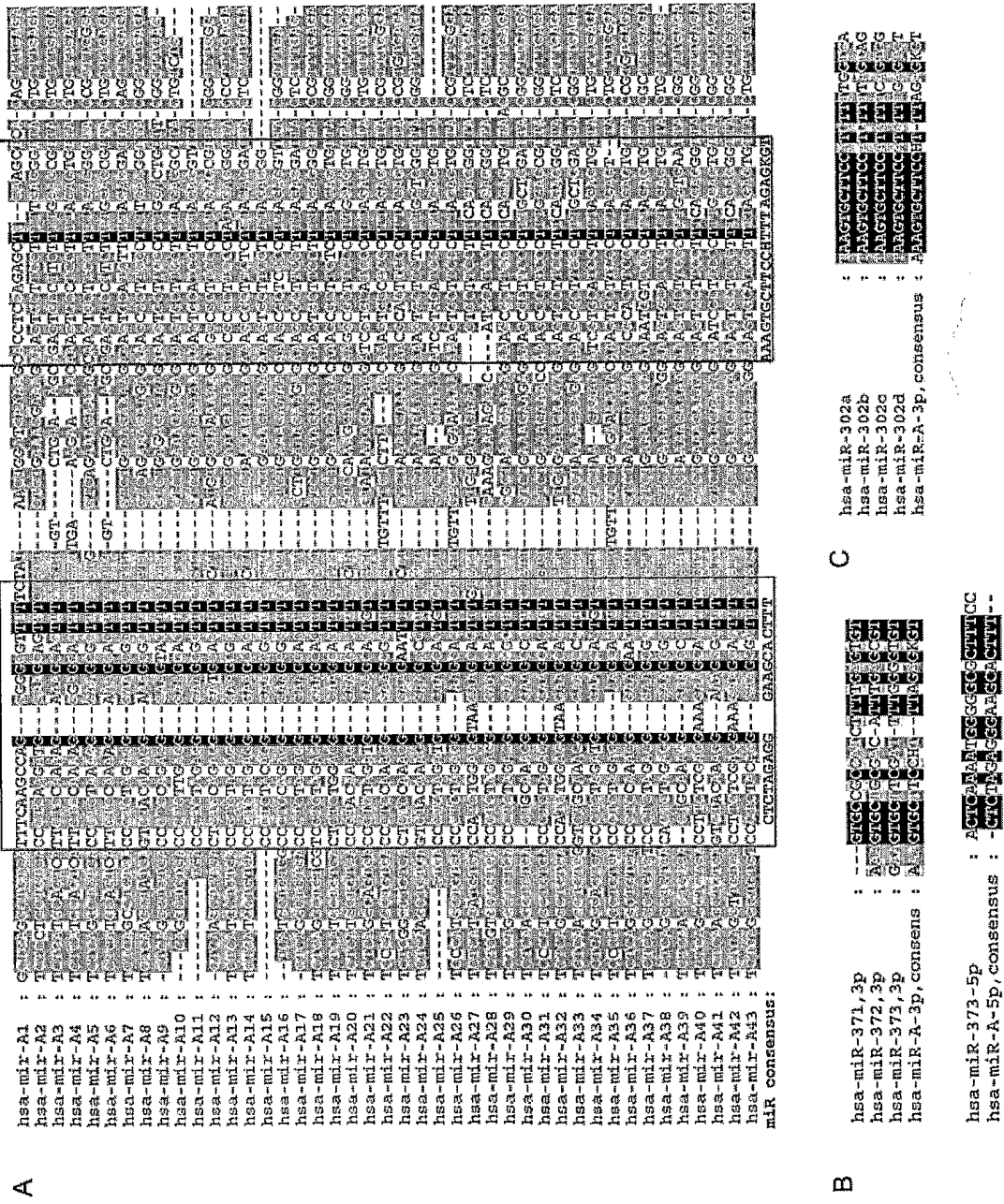


FIGURE 5

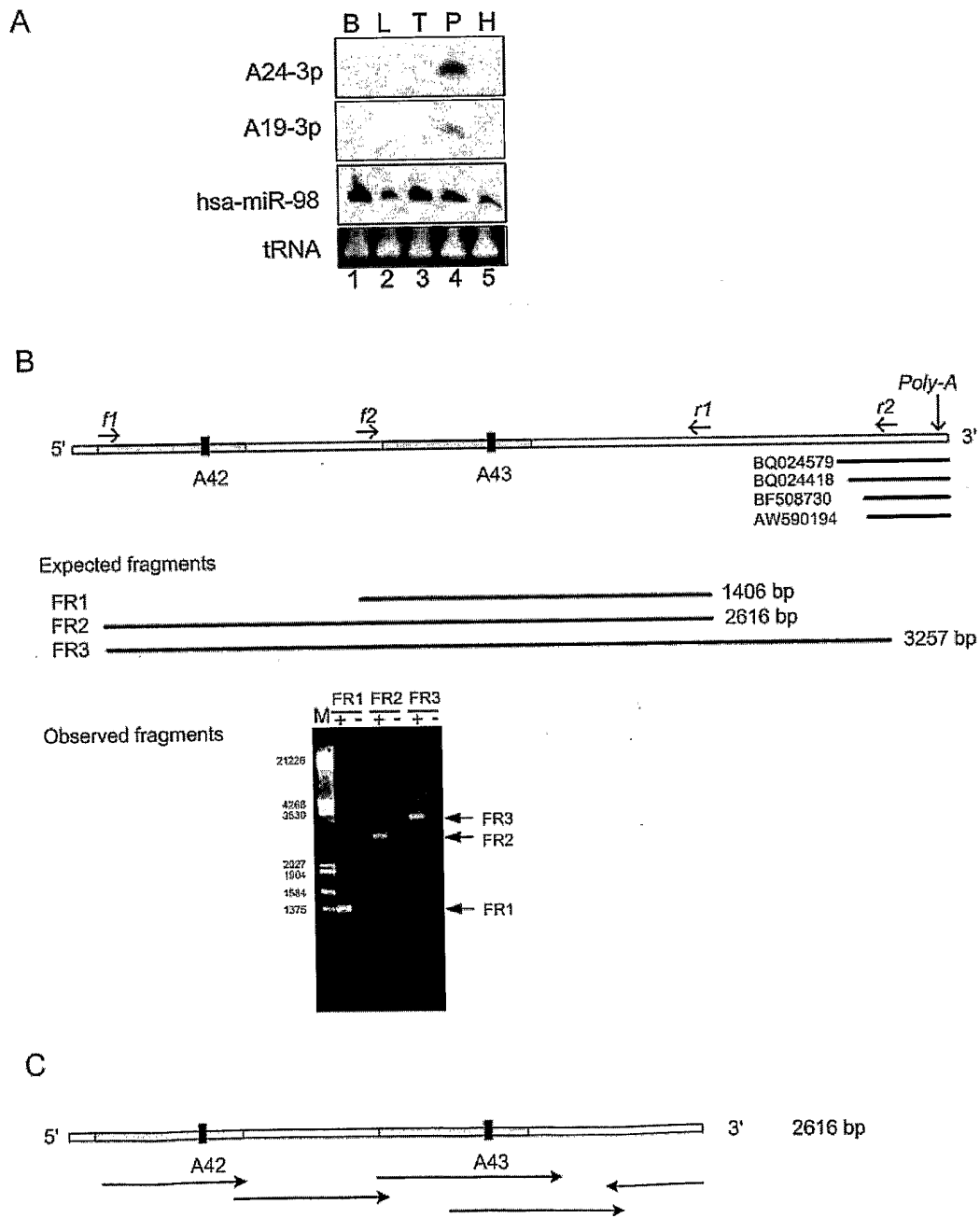


FIGURE 6

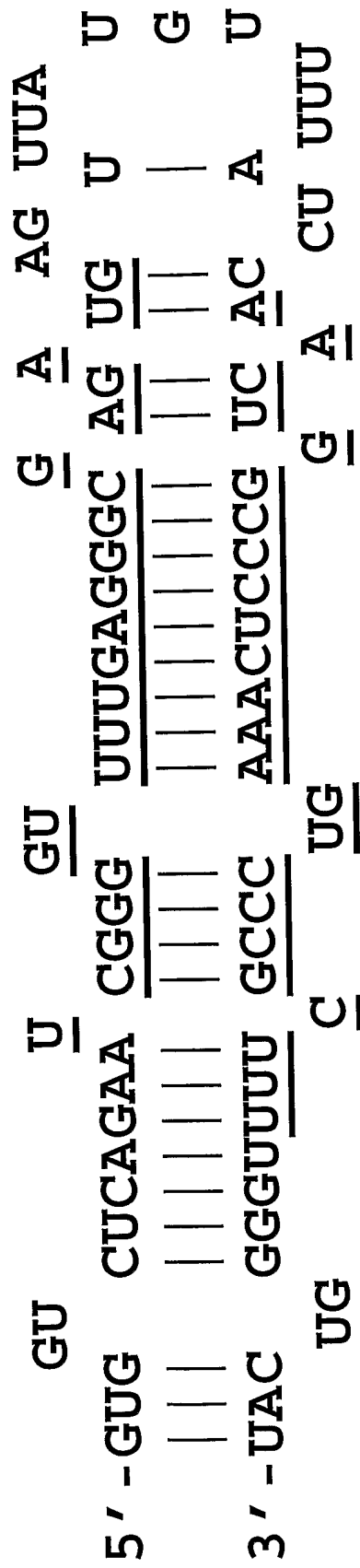


FIGURE 8

19061	AACUGGGCCCUCAAAGUCCCGC
19062	AACUGGGCCCUCAAAGUCCCGCU
19063	AACUGGGCCCUCAAAGUCCCGCUUU
19064	ACUGGGCCCUCAAAGUCCCGCUUU
19065	CCAAACUGGGCCCUCAAAGUCCCGCUU

FIGURE 9

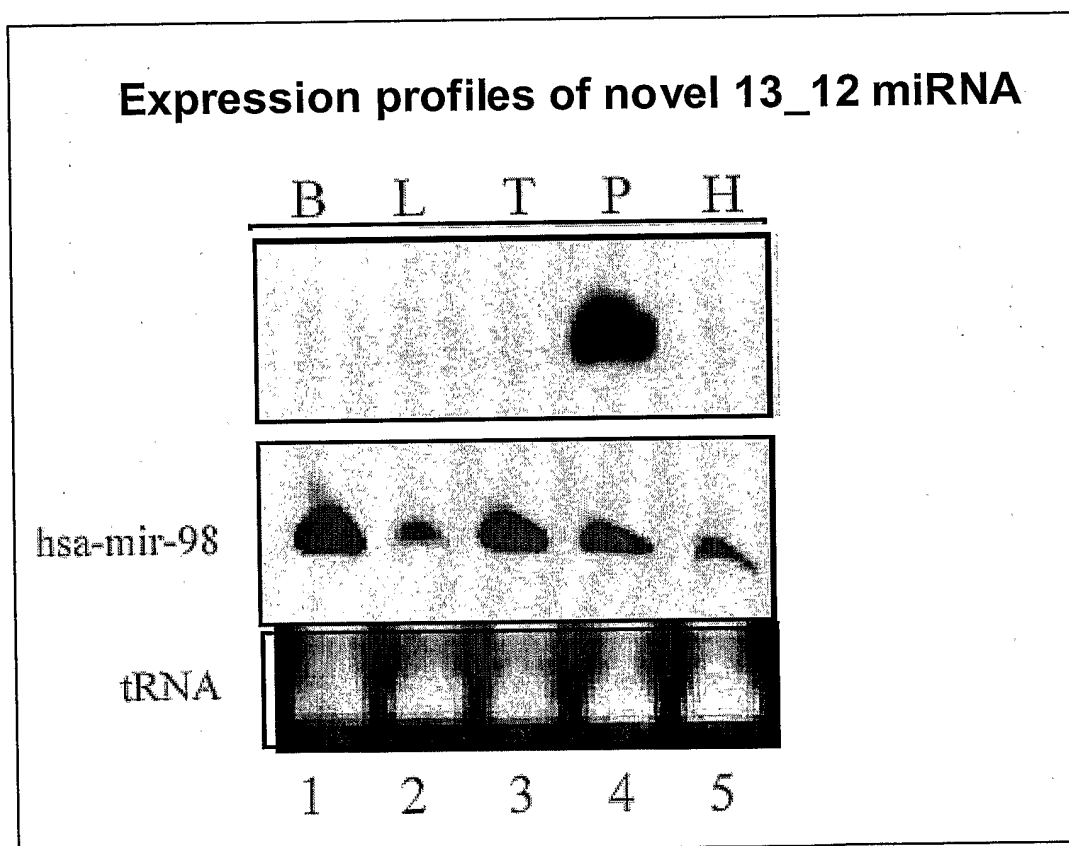


FIGURE 10

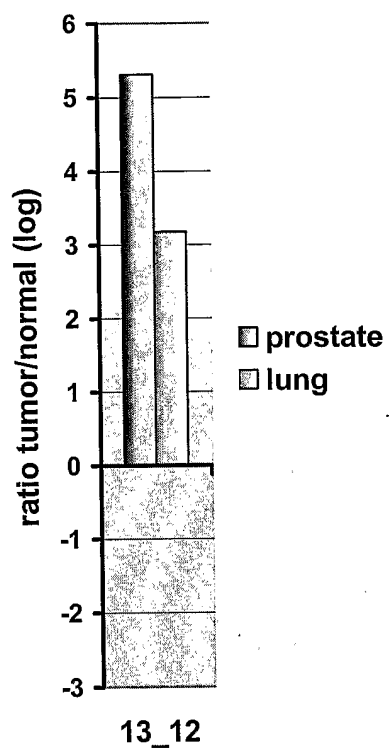


FIGURE 11

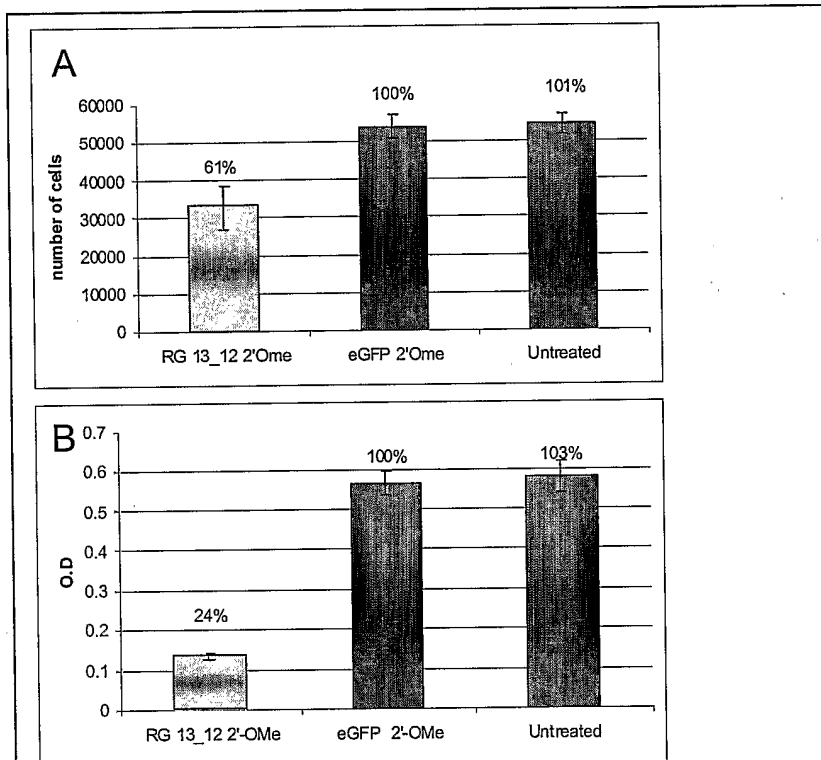


FIGURE 12

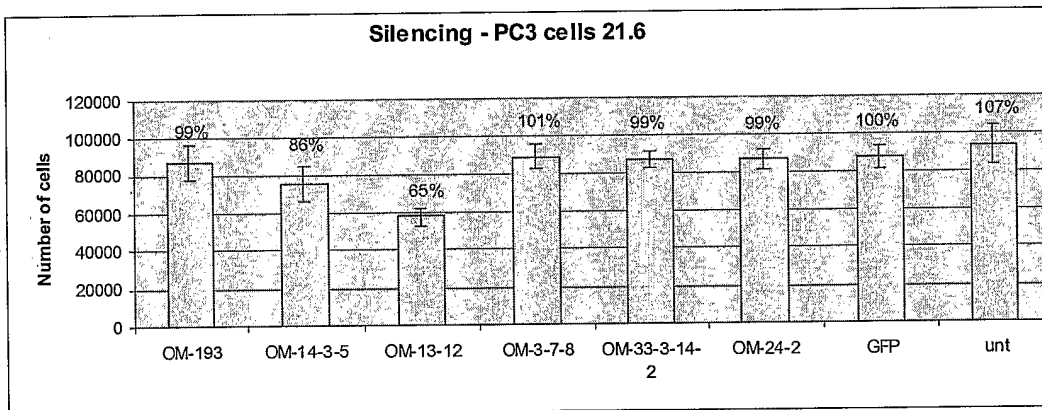


FIGURE 13

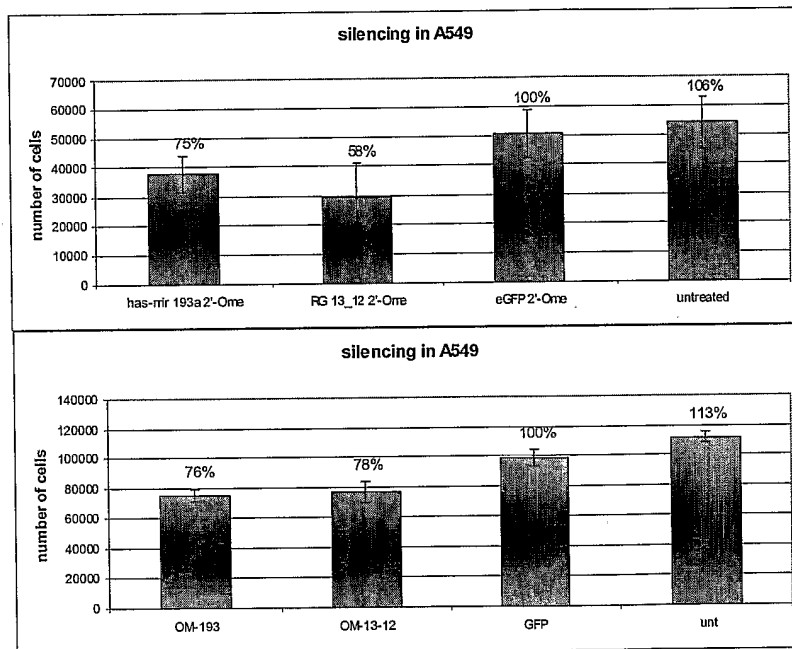


FIGURE 15

Detailed Methodology – outline protocol

1. Prepare total RNA.
2. Size fractionate the RNA onto YM-100 columns, to separate the small RNA molecules.
3. Precipitate RNA in EtOH overnight at 4°C.
4. Resuspend RNA is ready for ligation
5. Order adaptor for ligation. Adaptor should have the following features:
5' phosphate (for ligation) – **3 RNA nucleotides** (preferably starting with A (U), since ligation is RNA to RNA using RNA ligase, and is more efficient using U) – **15 DNA Nucleotide spacer** of non-relevant sequence (preferably including a restriction site) – **T7 promoter sequence** – **3' IdT** (to prevent self ligation).

For example:

RNA	DNA
5'-AGGAACAGAT	GAATTCACCTATAGTGAGTCGTATTATGCCT-3' IdT

- | Spacer | T7 promoter |
|--------|---|
| 6. | Ligate RNA and adaptor using T4 RNA ligase (NEB), in ligase buffer including DMSO (15%). Spike with known RNA. |
| 7. | Run ligation product onto acrylamide gel. |
| 8. | Cut out of the gel ligation product of right size |
| 9. | Ligation product is extracted out of the gel using GebaFlex electrophoration dialysis tubes. |
| 10. | Ligated product is then precipitated overnight in EtOH. |
| 11. | A DNA oligo is annealed to T7 promoter to create double strand at the T7 promoter region. Annealing is controlled by step decrease of temperature from 70°C to 25°C, after 2 minutes at 85°C. |
| 12. | Transcription is performed using the Megashortscript kit (Ambion) for transcription of short sequences. The kit uses T7 polymerize, nucleotides including biotinylated UTP and/or biotinylated CTP including the cold non-labeled nucleotide. Transcription occurs overnight at 37°C. |
| 13. | Adaptors are digested with DNase for 15 minutes at 37°C. |
| 14. | Final RNA Transcript is purified by Chloroform/Phenol phase extraction and EtOH precipitation overnight. |