(72) Inventors; and
(75) Inventors/Applicants (for US only): AHARONOV,
Ranit [IL/IL]; 10 Plaut St., 76706 Rehovot (IL). ROSEN-
FELD, Nitzan [IL/IL]; 10 Plaut St., 76706 Rehovot (IL).
BENJAMIN, Hila [IL/IL]; 10 Plaut St., 76706 Rehovot
(IL).

[Continued on next page]

(54) Title: CANCER-RELATED NUCLEIC ACIDS
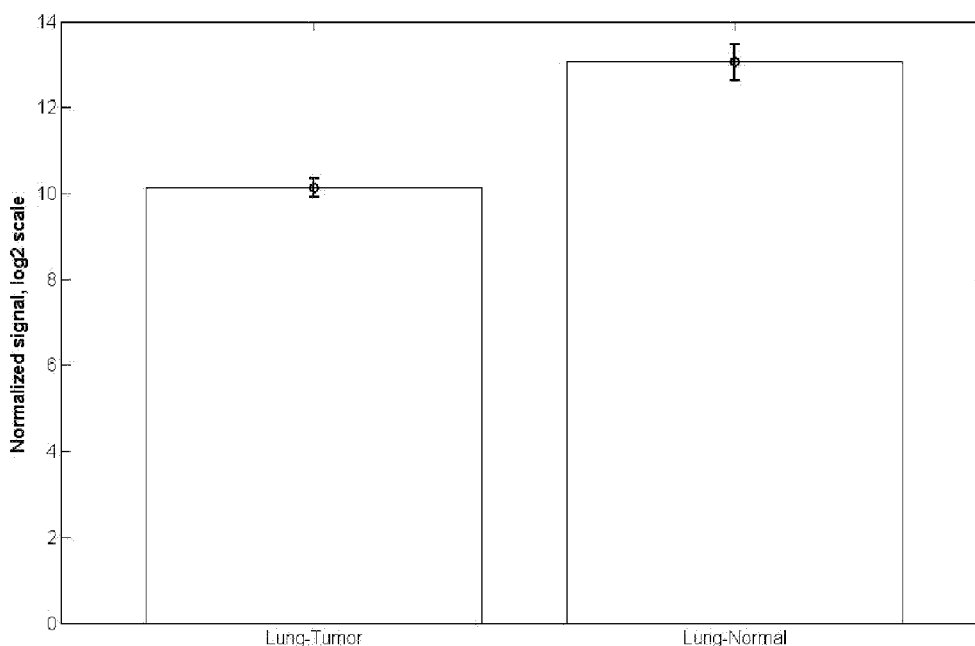
(57) Abstract: Described herein are polynucleotides associated with specific types of cancers. The polynucleotides are miRNAs,
miRNA precursors, and associated nucleic acids. Methods and compositions are described that can be used for diagnosis, prognosis,
and treatment of various cancers.

CANCER-RELATED NUCLEIC ACIDS

FIELD OF THE INVENTION

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

BACKGROUND OF THE INVENTION

[0002] In recent years, microarray analysis of gene expression patterns has provided a way to improve the diagnosis and risk stratification of many cancers. Extensive differences in gene expression patterns has identified molecularly distinct subtypes of cancer, that were previously considered homogeneous based on classical diagnostic methods. Such molecular subtypes are often associated with different clinical outcomes. Global gene expression pattern can also be examined for features that correlate with clinical behavior to create prognostic signatures.

[0003] The expression of many MicroRNAs are altered in numerous types of human cancer. MiRNAs are short RNA oligonucleotides of approximately 22 nucleotides that are involved in gene regulation. MiRNAs regulate gene expression by targeting mRNAs for cleavage or translational repression. Although miRNAs are present in a wide range of species including *C. elegans, Drosophila* 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. Alterations in miRNA expression have been implicated in playing a causative role in tumor progression.

[0004] The search for biomarkers for the early detection and diagnosis of various cancers has met with little success. Much emphasis has been placed on the discovery and characterization of a unique tumor marker. However, in the majority of cancers no marker has been identified that has adequate sensitivity or specificity to be clinically useful, although a combination of multiple markers has been shown to increase diagnostic accuracy. There is an unmet need for specific and accurate markers associated with specific types of cancers.

SUMMARY OF THE INVENTION

[0005] The present invention provides specific nucleic acid sequences that may be used for the identification and diagnosis of various cancers including but not limited to small intestine cancer, bladder cancer, prostate cancer, lung cancer, thyroid cancer, uterus cancer, liver cancer, kidney cancer, breast cancer, stomach cancer, testicular cancer, cervical cancer, esophageal cancer, gallbladder cancer, ovarian cancer, melanoma and colon cancer. The nucleic acid sequences can also be used for prognosis evaluation of a subject based on the expression pattern of a biological sample.

[0006] An isolated nucleic acid is also provided. The nucleic acid may comprise a sequence of any of SEQ ID NOS: 1-561, the complement thereof, or a sequence at least 81% identical to 21 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.

[0007] 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.

[0008] A method for detecting a cancer-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: 1-561. The nucleic acid may also comprise a sequence at least about 81% identical to about 21 contiguous nucleotides of any of SEQ ID NOS: 1-561. A level of the nucleic acid higher or lower than that of a control may be indicative of a specific cancer.

[0009] A method of diagnosing a subject with a specific cancer is also provided. The method comprising: providing a biological sample from the subject; and measuring the level of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 1-561, wherein a level of the nucleic acid different from a control is indicative of said cancer.

[0010] The invention further provides a method to distinguish between small intestine carcinoid tumor and small intestine stromal tumor, the method comprising determining a level of a nucleic acid sequence as set forth in Table 3 in a small intestine tumor sample obtained from a subject.

[0011] The invention further provides a method to distinguish between lung adenocarcinoma and squamous cell carcinoma, the method comprising determining a level of a nucleic acid sequence as set forth in Table 3 in a lung sample.

[0012] The invention further provides a method to distinguish between lung undifferentiated large cell carcinoma and squamous cell carcinoma, the method comprising determining a level of a nucleic acid sequence as set forth in Table 3 in a lung sample.

[0013] The invention further provides a method to distinguish between prostate adenocarcinoma and benign prostatic hyperplasia (BPH), the method comprising determining a level of a nucleic acid sequence as set forth in Table 3 in a prostate sample.

[0014] The invention further provides a method to distinguish between high grade carcinoma, medium grade carcinoma and low grade carcinoma, the method comprising determining a level of a nucleic acid sequence as set forth in Table 3 in a tumor sample obtained from a subject.

[0015] The invention further provides a method to distinguish between primary liver cancer and liver metastasis, the method comprising determining a level of a nucleic acid sequence as set forth in Table 3 in a liver sample.

[0016] A method for identifying a compound that modulates expression of a cancer-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: 1-561. 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: 1-561. 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.

[0017] A method of treating or preventing cancer in a subject in need thereof is also provided. The subject may suffer from a specific cancer as set forth in Table 1. The subject may be administered an effective amount of a composition comprising a nucleic acid. The nucleic acid may comprise a sequence of any of SEQ ID NOS: 1-561, 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: 1-561, 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.

BRIEF DESCRIPTION OF THE TABLES

[0018] **miR name**: is the miRBase registry name (release 9.1). Names which are not presented in the miRBase registry were cloned in Rosetta Genomics.

[0019] **HIDs** – Hairpins ID are the SEQ ID NOS of the microRNA hairpin sequences. Some microRNA can be derived from more than one hairpin, positioned at different genomic locations; in those cases the SEQ ID NOS of all related hairpins are presented.

[0020] **MID** - Mir ID, the SEQ ID NO of the mature microRNA sequence.

[0021] **RID**- Row ID. in decisions table, running number that is an identifier for further use as a reference.

[0022] **TID** - Tissue ID, the numeric code of the tissue tested, as described in Table 1.

[0023] **Signal** - the normalized mean signal read from the chip.

[0024] **DID** – Decision ID, describing the comparisons of one group of tissue(s) to another group of tissue(s) presented in Table 3.

[0025] **p-value** – 2-taied, un-paired, t-test that compare the means if Group 1 and Group 2. Only p-values which are lower than 0.05 are presented, When "0" is presented, the p-value is lower than 1.0e-38.

[0026] **n1**-the number of samples tested in Group 1.

[0027] **n2**- the number of samples tested in Group 2.

[0028] **Mean1-** the mean microRNA score of all the samples in Group 1.

[0029] **Mean2-** the mean microRNA score of all the samples in Group 2.

BRIEF DESCRIPTION OF THE DRAWINGS

[0030] Figure 1 shows the average normalized signal and standard deviation (STD) divided by the square root of the number of samples (n) of hsa-miR-126 (SEQ ID NO:204) in two lung sample sets: lung tumors and normal lungs. T-test p-value: 1.73e-008.

[0031] Figure 2 shows the average normalized signal and standard deviation (STD) divided by the square root of the number of samples (n) of hsa-miR-200c (SEQ ID NO:156) in two liver sample sets: primary tumors and metastasis. T-test p-value:1.48E-09.

[0032] Figure 3 shows the average normalized signal and standard deviation (STD) divided by the square root of the number of samples (n) of hsa-miR-375 (SEQ ID NO:273) in two small intestine sample sets: stromal tumors and carcinoid tumors. T-test p-value: 6.68E-08.

[0033] Figure 4 shows the average normalized signal and standard deviation (STD) divided by the square root of the number of samples (n) of 43_5 (SEQ ID NO:215) in two bladder cancer sample sets: high grade tumors and low grade tumors. T-test p-value: 2.49E-08.

DETAILED DESCRIPTION

[0034] Nucleic acids are provided related to miRNAs and precursors thereto. Such nucleic acids may be useful for diagnostic and prognostic purposes. Also provided are methods and compositions that may be useful, among other things, for diagnostic and prognostic purposes.

[0035] Other aspects of the invention will become apparent to the skilled artisan by the following description of the invention.

1. Definitions

[0036] 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.

[0037] "About" as used herein may mean +/-10%.

[0038] "Aberrant proliferation" as used herein may mean cell proliferation that deviates from the normal, proper, or expected course. For example, aberrant cell proliferation may include inappropriate proliferation of cells whose DNA or other cellular components have become damaged or defective. Aberrant cell proliferation may include cell proliferation whose characteristics are associated with an indication caused by, mediated by, or resulting in inappropriately high levels of cell division, inappropriately low levels of apoptosis, or both. Such indications may be characterized, for example, by single or multiple local abnormal proliferations of cells, groups of cells, or tissue(s), whether cancerous or non-cancerous, benign or malignant.

[0039] "Antisense" as used herein may mean nucleotide sequences that are complementary to a specific sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules may be produced by any method, including synthesis by ligating the gene(s) of interest in a reverse orientation to a viral promoter which permits the synthesis of a complementary strand. Once introduced into a cell, this

transcribed strand may combine with natural sequences produced by the cell to form duplexes. These duplexes then block either the further transcription or translation. In this manner, mutant phenotypes may be generated.

[0040] "Animal" as used herein may mean fish, amphibians, reptiles, birds, and mammals, such as mice, rats, rabbits, goats, cats, dogs, cows, apes and humans.

[0041] "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.

[0042] "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.

[0043] "Cancer" as used herein may mean all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. Examples of cancers include but are nor limited to solid tumors and leukemias, including: apudoma, choristoma, branchioma, malignant carcinoid syndrome, carcinoid heart disease, carcinoma (e.g., Walker, basal cell, basosquamous, Brown-Pearce, ductal, Ehrlich tumor, non-small cell lung, oat cell, papillary, bronchiolar, bronchogenic,

squamous cell, and transitional cell), histiocytic disorders, leukemia (e.g., B cell, mixed cell, null cell, T cell, T-cell chronic, HTLV-II-associated, lymphocytic acute, lymphocytic chronic, mast cell, and myeloid), histiocytosis malignant, Hodgkin disease, immunoproliferative small, non-Hodgkin lymphoma, plasmacytoma, reticuloendotheliosis, melanoma, chondroblastoma, chondroma, chondrosarcoma, fibroma, fibrosarcoma, giant cell tumors, histiocytoma, lipoma, liposarcoma, mesothelioma, myxoma, myxosarcoma, osteoma, osteosarcoma, Ewing sarcoma, synovioma, adenofibroma, adenolymphoma, carcinosarcoma, chordoma, craniopharyngioma, dysgerminoma, hamartoma, mesenchymoma, mesonephroma, myosarcoma, ameloblastoma, cementoma, odontoma, teratoma, thymoma, trophoblastic tumor, adeno-carcinoma, adenoma, cholangioma, cholesteatoma, cylindroma, cystadenocarcinoma, cystadenoma, granulosa cell tumor, gynandroblastoma, hepatoma, hidradenoma, islet cell tumor, Leydig cell tumor, papilloma, Sertoli cell tumor, theca cell tumor, leiomyoma, leiomyosarcoma, myoblastoma, myosarcoma, rhabdomyoma, rhabdomyosarcoma, ependymoma, ganglioneuroma, glioma, medulloblastoma, meningioma, neurilemmoma, neuroblastoma, neuroepithelioma, neurofibroma, neuroma, paraganglioma, paraganglioma nonchromaffin, angiokeratoma, angiolymphoid hyperplasia with eosinophilia, angioma sclerosing, angiomatosis, glomangioma, hemangioendothelioma, hemangioma, hemangiopericytoma, hemangiosarcoma, lymphangioma, lymphangiomyoma, lymphangiosarcoma, pinealoma, carcinosarcoma, chondrosarcoma, cystosarcoma, phyllodes, fibrosarcoma, hemangiosarcoma, leimyosarcoma, leukosarcoma, liposarcoma, lymphangiosarcoma, myosarcoma, myxosarcoma, ovarian carcinoma, rhabdomyosarcoma, sarcoma (e.g., Ewing, experimental, Kaposi, and mast cell), neurofibromatosis, and cervical dysplasia, and other conditions in which cells have become immortalized or transformed.

[0044] "Complement" or "complementary" as used herein may mean 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.

[0045] "Diagnostic" as used herein may mean classifying a pathology or a symptom, determining a severity of the pathology (grade or stage), monitoring pathology progression, forecasting an outcome of a pathology and/or prospects of recovery.

[0046] "Differential expression" as used herein 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.

[0047] "Gene" as 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.

[0048] "Host cell" as 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.

[0049] "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.

[0050] "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.

[0051] "Nucleic acid" or "oligonucleotide" or "polynucleotide" as 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.

[0052] 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.

[0053] 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, NH_2, NHR, NR_2 or CN, wherein R is C_1-C_6 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. Modified nucleotides and nucleic acids may also include locked nucleic acids (LNA), as described in U.S. Patent No. 20020115080, which is 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.

[0054] "Operably linked" as 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.

[0055] "Predisposition" as used herein may mean an increased likelihood that an individual will have a disorder. Although a subject with a predisposition does not yet have the disorder, there exists an increased propensity to the disease.

[0056] "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.

[0057] "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.

[0058] "Selectable marker" as 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.

[0059] "Stringent hybridization conditions" as 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.

[0060] "Substantially complementary" as 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.

[0061] "Substantially identical" as 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.

[0062] "Subject" as used herein may mean a mammal, including both human and other mammals.

[0063] "Target" as used herein may mean a polynucleotide that may be bound by one or more probes under stringent hybridization conditions.

[0064] "Terminator" as 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.

[0065] "Therapeutically effective amount" or "therapeutically efficient" as used herein in respect to a drug dosage, may refer to dosage that provides the specific pharmacological response for which the drug is administered in a significant number of subjects in need of such treatment. The "therapeutically effective amount" may vary according, for example, the physical condition of the patient, the age of the patient and the severity of the disease.

[0066] "Treatment regimen" as used herein may mean a treatment plan that specifies the type of treatment, dosage, schedule and/or duration of a treatment provided to a subject in need thereof (e.g., a subject diagnosed with a pathology). The selected treatment regimen can be an aggressive one which is expected to result in the best clinical outcome (e.g., complete cure of the pathology) or a more moderate one which may relieve symptoms of the pathology yet results in incomplete cure of the pathology. It will be appreciated that in certain cases the treatment regimen may be associated with some discomfort to the subject or adverse side effects (e.g., a damage to healthy cells or tissue). The type of treatment can include a surgical intervention

(e.g., removal of lesion, diseased cells, tissue, or organ), a cell replacement therapy, an administration of a therapeutic drug (e.g., receptor agonists, antagonists, hormones, chemotherapy agents) in a local or a systemic mode, an exposure to radiation therapy using an external source (e.g., external beam) and/or an internal source (e.g., brachytherapy) and/or any combination thereof. The dosage, schedule and duration of treatment can vary, depending on the severity of pathology and the selected type of treatment, and those of skills in the art are capable of adjusting the type of treatment with the dosage, schedule and duration of treatment.

[0067] "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.

[0068] "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.

[0069] "Wild type" as used herein the context of a sequence may refer to a coding, non-coding or interface sequence is an allelic form of sequence that performs the natural or normal function for that sequence. Wild type sequences include multiple allelic forms of a cognate sequence, for example, multiple alleles of a wild type sequence may encode silent or conservative changes to the protein sequence that a coding sequence encodes.

2. MicroRNA

[0070] A gene coding for a miRNA may be transcribed leading to production of a miRNA primary transcript known as the pri-miRNA. The pri-miRNA may form a hairpin with a stem and loop. The stem of the hairpin may comprise mismatched bases. The pri-miRNA may comprise several pri-miRNAs in a polycistronic structure.

[0071] 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 the 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.

[0072] 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 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.

[0073] 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 specifity 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.

[0074] 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.

[0075] The RISC may identify target nucleic acids based on high levels of complementarity between the miRNA and the mRNA, especially by nucleotides 2-7 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).

[0076] 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 at 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).

[0077] 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.

[0078] 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.

[0079] 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

[0080] Nucleic acids are provided herein. The nucleic acid may comprise the sequence of SEQ ID NOS: 1-561 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 that hybridizes under stringent conditions to the referenced nucleotide sequence, complements thereof, or nucleotide sequences substantially identical thereto.

[0081] 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

[0082] The nucleic acid may comprise a sequence of a pri-miRNA or a variant thereof. The pri-miRNA sequence may comprise from 45-30,000, 50-25,000, 100-20,000, 1,000-1,500 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: 1-561, the complement of a target gene binding site, or variants thereof.

[0083] The pri-miRNA may comprise a hairpin structure. The hairpin may comprise a first and second nucleic acid sequence that are substantially complementary. 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

[0084] 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: 1-561, the complement of a target gene binding site, or variants thereof.

c. MiRNA

[0085] 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 from the first 13-33 nucleotides of the pre-miRNA. The sequence of the miRNA may also be from the last 13-33 nucleotides of the pre-miRNA. The sequence of the miRNA may comprise the sequence of SEQ ID NOS: 1-274, the complement of a target gene binding site, or variants thereof.

d. Anti-miRNA

[0086] 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 complementary to the 5' of a miRNA and at least 5-12 nucleotides that are substantially identical to the flanking regions of the target site from the 5' end of the miRNA, for the purposes of binding to a miRNA and repressing its activity; or (b) at least 5-12 nucleotides that are substantially identical to the 3' of a miRNA and at least 5 nucleotide that are substantially complementary to the flanking region of the target site from the 3' end of the miRNA, for the

purposes of inhibiting the ability of a miRNA to bind to its target. The sequence of the anti-miRNA may comprise the complement of SEQ ID NOS: 1-561, the complement of a target gene binding site, or variants thereof.

e. Binding Site of Target

[0087] 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: 1-274.

4. Synthetic Gene

[0088] 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

[0089] A vector is also provided comprising a nucleic acid described herein, such as a pri-miRNA, pre-miRNA, miRNA, anti-miRNA, target gene binding site, or synthetic gene. 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 that 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

[0090] 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

[0091] 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.

[0092] 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

[0093] 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.

[0094] 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.

[0095] 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, TeflonJ, 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.

[0096] 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.

[0097] 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 linkers. The probes may be attached to the solid support by either the 5' terminus, 3' terminus, or via an internal nucleotide.

[0098] 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

[0099] A method of identifying a nucleic acid associated with a specific cancer or a related 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.

[0100] 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 cancerous 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.

[0101] The expression level of a cancer-associated nucleic acid is information in a number of ways. For example, a differential expression of a cancer-associated nucleic acid compared to a control may be used as a diagnostic that a patient suffers from the cancer. Expression levels of a cancer-associated nucleic acid may also be used to monitor the treatment and cancer state of a patient. Furthermore, expression levels of a cancer-associated miRNA may allow the screening of drug candidates for altering a particular expression profile or suppressing an expression profile associated with cancer.

[0102] 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.

[0103] 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 labeled 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.

[0104] 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.

[0105] 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.

[0106] 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.

[0107] 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

[0108] A method of diagnosis is also provided. The method comprises detecting a differential expression level of a specific cancer-associated nucleic acid in a biological sample. The sample may be derived from a patient. Diagnosis of a cancer 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 cancer-associated nucleic acids.

[0109] 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

[0110] A method of screening therapeutics is also provided. The method comprises contacting a pathological cell capable of expressing a cancer related nucleic acid with a candidate therapeutic and evaluating the effect of a drug candidate on the expression profile of the cancer 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 cancer associated nucleic acid. Modulation includes both an increase and a decrease in gene expression.

[0111] 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.

[0112] 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

[0113] 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.

[0114] 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

[0115] 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

[0116] A method for treating cancer is also provided. The cancer may be treated in vivo or ex vivo. The compositions of the present invention may be combined with a chemotherapeutic agent, a combination of chemotherapeutic agents and/or radiotherapy.

[0117] 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.

[0118] 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

[0119] 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.

[0120] The composition can be administered to cells by a variety of methods known to those familiar to the art, including, but not restricted to, encapsulation in liposomes, by iontophoresis, or by incorporation into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, and bioadhesive microspheres. The composition may also be locally delivered by direct injection or by use of an infusion pump. Other routes of delivery include, but are not limited to oral (tablet or pill form) and/or intrathecal delivery (Gold, 1997, Neuroscience, 76, 1153-1158). Other approaches for delivery include the use of various transport and carrier systems, for example, through the use of conjugates and biodegradable polymers. Other descriptions of nucleic acid delivery and administration are provided for example in WO93/23569, WO99/05094, WO99/04819, WO94/02595 and Akhtar et al., (Trends Cell Bio. 2,139,1992).

[0121] The composition can be introduced into tissues or host cells by any number of routes, including viral infection, microinjection, or fusion of vesicles. Jet injection may also be used for intra-muscular administration, as described by Furth et al. (Anal Biochem 115 205:365-368, 1992). The composition can be coated onto gold microparticles, and delivered intradermally by a particle bombardment device, or "gene gun" as described in the literature (see, for example, Tang

et al. Nature 356:152-154, 1992), where gold microprojectiles are coated with the DNA, then bombarded into skin cells.

[0122] The compositions of the present invention can be formulated into pharmaceutical compositions by combination with appropriate, pharmaceutically acceptable carriers or diluents, and can be formulated into preparations in solid, semi-solid, liquid or gaseous forms, such as tablets, capsules, powders, granules, ointments, solutions, suppositories, injections, inhalants and aerosols. As such, administration of the agents can be achieved in various ways, including oral, buccal, rectal, parenteral, intraperitoneal, intradermal, transdermal, intracheal, etc.

14. Kits

[0123] 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. The kit may further comprise a software package for data analysis of expression profiles.

[0124] For example, the kit may be a kit for the amplification, detection, identification or quantification of a target nucleic acid sequence. The kit may comprise a poly(T) primer, a forward primer, a reverse primer, and a probe.

15. Method of Synthesis

[0125] A method of synthesizing the reverse-complement of a target nucleic acid is also provided. The reverse complement may be synthesized according to methods outlined in U.S. Patent No. 11/384,049, the contents of which are incorporated herein by reference.

16. Method of Detection

[0126] A method of detecting a target nucleic acid in a biological sample is also provided. The target nucleic acid may be detected according to methods outlined in U.S. Patent No. 11/384,049, the contents of which are incorporated herein by reference.

a. Detectable Malignancies

[0127] **Liver cancer:** Primary liver cancer is the fifth most common cancer worldwide. Hepatocellular carcinoma (HCC) accounts for 80% of all liver cancer and the rates of HCC have increased by over 70% in the last two decades in the U.S. The fatality ratio (mortality/incidence) of liver cancer is approximately 1, indicating that the majority of patients live less than a year.

Late diagnosis due to lack of clinical symptoms is one of the main reasons for the high fatality ratio. Liver cancer can result from both viral infection and chemical exposure. Known risk factors include hepatitis B and C virus infection. It is not known whether distinct routes to liver cancer affect the same or different cellular pathways. No mutational model has yet been developed for liver cancer as it has been for other cancers. The molecular events that precede neoplastic transformation of the liver are not well understood. With no clearly identified cause, successful treatment options are lacking.

[0128] Nearly any primary tumor site can deposit metastases in the liver, since the liver filters blood from throughout the body. Most discussions related to the treatment of metastatic tumors in the liver focus on those originating from the colon. In fact, the most common cause of death from colorectal cancer is liver metastasis.

[0129] Up to 50% of liver metastases are of colorectal cancer origin, while the remainder metastasizes from a wide variety of primary cancer sites including sarcomas, breast and kidney, as well as neuroendocrine tumors.

[0130] HCC may be solitary or multicentric, and it may mimic liver metastases. Furthermore hemangiomas and liver metastases are often confused in imaging methods. In general, the imaging appearances of liver metastases are nonspecific, and biopsy specimens are required for histological diagnosis. Various biochemical markers have been proposed to indicate liver metastases. However, the diagnostic accuracy of tumor markers has not yet been defined.

[0131] **Lung cancer:** Lung cancer is the primary cause of cancer death among both men and women in the U.S., with an estimated 172,000 new cases being reported in 1994. The five-year survival rate among all lung cancer patients, regardless of the stage of disease at diagnosis, is only 13%. This contrasts with a five-year survival rate of 46% among cases detected while the disease is still localized. However, only 16% of lung cancers are discovered before the disease has spread. Lung cancers are broadly classified into small cell or non-small cell lung cancers. Non-small cell lung cancers are further divided into adenocarcinomas, bronchoalveolar-alveolar, squamous cell and large cell carcinomas. Approximately, 75-85 percents of lung cancers are non-small cell cancers and 15-25 percents are small cell cancers of the lung.

[0132] Lung large cell carcinoma often is discovered on a chest X-ray, computed tomography (CT) or magnetic resonance imaging (MRI) scans, This helps to find the best place to obtain a sample of the tumor to differentiate large cell carcinoma from other types of lung cancer such as

small cell carcinoma or adenocarcinoma of the lung. The type of cancer affects the type of treatment the patient should get. Circulating tumor markers could impact non –small cell lung cancer (NSCLC) patient outcomes through improved screening, diagnosis, staging and management. The search for biomarkers for the early detection and diagnosis of NSCLC has met with little success. Much emphasis has been placed on the discovery and characterization of a unique tumor marker. However, no marker has been identified that has adequate sensitivity or specificity to be clinically useful, although a combination of multiple markers has been shown to increase diagnostic accuracy.

[0133] **Prostate cancer:** Prostate cancer is a common disease in men above 50 years. Prostate cancer is the most frequent type of cancer in males, accounting for 25% of all cancers in the male population. Unfortunately, prostate carcinomas that are progressive in nature frequently have already metastasized by the time of clinical detection with available methods. Survival rates for individuals with metastatic prostate cancer are quite low. Between these two extremes are patients with prostate tumors that will metastasize during their lifetimes, but have not yet done so. For these patients, surgical removal of the prostate is curative and extends life expectancy. Therefore, accurate determination of which group a newly diagnosed patient falls into is critical in determining optimal treatment and patient survival.

[0134] Although clinical and pathologic stage and histological grading systems (e.g., Gleason's) have been used to indicate prognosis for groups of patients based on the degree of tumor differentiation or the type of glandular pattern (Diamond et al., J. Urol., 128: 729-734, 1982), these systems do not adequately predict the progression rate of the cancer. While the use of computer-system image analysis of histologic sections of primary lesions for "nuclear roundness" has been suggested as an aide in the management of individual patients (Diamond et al., 1982, J Urol., 128:729-734), this method is of limited use in studying the progression of the disease.

[0135] **Thyroid cancer:** It is estimated that between 4-7% of the population harbors thyroid nodules. Literature reports that 5-30% of these nodules are malignant. The main diagnostic consideration in these cases is the exclusion of malignancy. Currently, the method offering the best preoperative prediction of the nature of these nodules is the fine needle aspiration biopsy of the lesion (FNA). Use of FNA resulted in decrease in number of thyroidectomies performed and increase in the yield of malignancy in resected lesions. However, instances of inadequate

sampling of the lesion and overlapping cytological features of benign and malignant thyroid neoplasms are inherent limitations of this technique. The major FNA limitation is its inability to distinguish between well-differentiated follicular carcinomas and benign follicular adenomas. Patients with follicular thyroid neoplasm usually undergo thyroidectomy and about 15% have malignant lesion. Reliable preoperative diagnosis of benign lesion would greatly reduce number of unnecessary surgeries.

[0136] **Bladder cancer**: Bladder cancer is the fourth most prevalent human malignancy, with about 49,000 new cases and 9,700 deaths reported annually [Silverman, D. T. et al., Epidemiology of Bladder Cancer. In: "Hematology/Oncology Clinics of North America". P. W. Kantoff et al.. eds., W.B. Saunders Co., Philadelphia, p. 1 (1992)]. Ninety percent of bladder cancers are transitional cell carcinomas which are typically superficial at early stages but often become invasive at later stages, 5% of bladder cancers are squamous cell carcinomas (SCC), which are more prevalent in cases of chronic bladder irritation, and the remainders are rare tumors such as adenocarcinoma, carcinosarcoma. Bladder cancer is usually diagnosed via cystoscopy, an invasive procedure, wherein a fiber optic device is inserted into the bladder and lesions are detected visually by a urologist. Cystoscopy is performed on patients expressing the symptom complex characteristic of bladder cancer, i.e., hematuria, pain, or urinary obstruction. However, when symptoms appear, the tumor is usually progressed to a dangerous grade or stage. In addition, this type of macroscopic diagnostics fails to detect microscopic disease such as carcinoma in situ [Halachmi et al., (2001), Bladder cancer: genetic overview. Med. Sci. Monit. 7: 164-168].

[0137] **Breast cancer**: Breast cancer is by far the most common cancer in women worldwide. Current global incidence is in excess of 1,151,000 new cases diagnosed each year. Breast cancer incidence is highest in developed countries, particularly amongst populations of Northern European ethnic origin, and is increasing. In the United States the annual age-standardized incidence rate is approximately 131 cases per 100,000 population, more than three times the world average. Rates in Northern European countries are similarly high. In the year 2006 it is estimated that 214,640 new cases of invasive breast cancer will be diagnosed in the U.S.A. and 41,430 people will die from the disease. To this figure must be added a further 59,000 ductal and lobular carcinoma in situ diagnoses. From an individual perspective, the lifetime probability of developing breast cancer is 13.1% in U.S. women (i.e., 1 in 8 women will develop breast cancer

during their lives). As with most cancers, early detection and appropriate treatment are important factors. Overall, the 5-year survival rate for breast cancer is 88%. However, in individuals presenting with regionally invasive or metastatic disease, the rate declines to 80% and 26%, respectively.

[0138] No universally successful method for the treatment or prevention of breast cancer is currently available. Management of breast cancer currently relies on a combination of early diagnosis (e.g., through breast screening procedures, e.g., mammography) and treatments using surgery, chemotherapy, radiotherapy and hormonal therapies. Increasingly, the focus is falling on the identification individuals who are at high risk for primary or recurrent breast cancer. Such individuals can be managed by more intensive screening, preventative chemotherapies or hormonal therapies and, in cases of individuals at extremely high risk, prophylactic surgery. There is a significant need, therefore, for improved diagnostic methods and identification of risk for breast cancer.

[0139] **Colon cancer:** The deaths by colon cancer are increasing. The number of deaths by colon cancer is the fourth large among male, and the second large among female deaths in all cancer deaths (Statistics of Japanese cancer deaths in 1999). According to an estimation of cancer patients in 2015 in Japan, number of colon cancer patients is estimated to be the first in both male and female. Global measures to counter colon cancer including secondary prevention are thus required, and mass screening of cancer may be one of the most effective methods. For the mass screening of cancer, it is important that the detection method is easy and non-invasive. The only non-invasive method now available is the method to examine existence of occult blood in feces, that is, the fecal occult blood test, and is used extensively as a standard method of the mass screening of colon cancer. However, the fecal occult blood test has rather low sensitivity and specificity (the sensitivity: 30 to 90%, the specificity: 70 to 98%), because appearance of hemoglobin in feces is not specific to tumor. Therefore, there is a shortcoming that quite a few false negatives and false positives exist.

[0140] **Kidney cancer:** The American Cancer Society predicted that there would be about 31,200 new cases of kidney cancer in the year 2000 in the United States alone. About 11,900 people, adults and children, will die from this disease each year. The cure rate of advanced stage kidney cancer is only fair and has improved little in the last two decades. Kidney cancer remains difficult to diagnose and treat effectively.

[0141] **Small intestine cancer:** Symptoms of cancer of the small intestine typically include pain or cramps in the middle of the abdomen, weight loss without dieting, and a lump in the abdomen or blood in the stool. Procedures used for detecting, diagnosing, monitoring, staging, and prognosticating cancer of the small intestine are of critical importance to the outcome of the patient. Patients diagnosed with early stage cancer generally have a much greater survival rate as compared to the survival rate for patients diagnosed with distant metastasized cancers. New diagnostic methods which are more sensitive and specific for detecting cancer of the small intestine are clearly needed.

[0142] **Testicular cancer:** Male germ cell tumors have been histologically categorized into seminomas, which retain germ cell characteristics, and nonseminomas which can display characteristics of embryonal differentiation. Both seminomas and nonseminomas are thought to initiate from a preinvasive stage designated carcinoma in situ (CIS) (Murty, et al., 1998, Sem. Oncol. 25:133-144). Testicular tumors develop from Leydig cells with high frequency in transgenic mice expressing human papilloma virus 16 (HPV16) E6 and E7 oncogenes (Kondoh, et al., 1991, J. Virol. 65:3335-3339; Kondoh, et al., 1994, J. Urol. 152:2151-2154).

[0143] The following examples are presented in order to more fully illustrate some embodiments of the invention. They should, in no way be construed, however, as limiting the broad scope of the invention.

EXAMPLES

Experimental Procedures

1. miRdicator™ array platform

[0144] Custom microarrays were produced by printing DNA oligonucleotide probes representing 688 miRNAs (Sanger database, version 9 and additional Rosetta Genomics validated and predicted miRs). Each probe carried up to a 22-nt linker at the 3' end of the miRNA's complement sequence in addition to an amine group used to couple the probes to coated glass slides. 20μM of each probe were dissolved in 2X SSC + 0.0035% SDS and spotted in triplicate on Schott Nexterion® Slide E coated microarray slides using a Genomic Solutions® BioRobotics MicroGrid II according the MicroGrid manufacturer's directions. 64 negative control probes were designed using the sense sequences of different miRNAs. Two groups of positive control probes were designed to hybridize to miRdicator™ array (1) synthetic spikes

small RNA were added to the RNA before labeling to verify the labeling efficiency and (2) probes for abundant small RNA (e.g. small nuclear RNAs (U43, U49, U24, Z30, U6, U48, U44), 5.8s and 5s ribosomal RNA) were spotted on the array to verify RNA quality. The slides were blocked in a solution containing 50 mM ethanolamine, 1M Tris (pH 9.0) and 0.1%SDS for 20 min at 500C, then thoroughly rinsed with water and spun dry.

2. Cy-dye labeling of miRNA for miRdicator™ array

[0145] 15 μg of total RNA was labeled by ligation of a RNA-linker p-rCrU-Cy- dye (Thomson et al., 2004, Nat Methods 1, 47-53) (Dharmacon) to the 3' -end with Cy3 or Cy5. The labeling reaction contained total RNA, spikes (20-0.1 fmoles), 500ng RNA-linker-dye, 15% DMSO, 1x ligase buffer and 20 units of T4 RNA ligase (NEB) and proceeded at 40C for 1hr followed by 1hr at 37^{0}C. The labeled RNA was mixed with 3x hybridization buffer (Ambion), heated to 95^{0}C for 3 min and than added on top of the miRdicator™ array. Slides were hybridize 12-16hr, followed by two washes with 1xSSC and 0.2% SDS and a final wash with 0.1xSSC.

[0146] The array was scanned using an Agilent Microarray Scanner Bundle G2565BA (resolution of 10 μm at 100% power). The data was analyzed using SpotReader software.

3. RNA extraction

[0147] RNA was extracted from frozen or paraffin-embedded (FFPE) tissues originated from different tumor or normal tissues as indicated in Table 1, where TID - Tissue ID is the numeric code of each tissue or sub-type tissue.

[0148] Total RNA from frozen tissues was extracted with the *miR*vana miRNA isolation kit (Ambion) according to the manufacturer's instructions.

[0149] Total RNA from formalin fixed, paraffin-embedded (FFPE) tissues was extracted according to the following protocol:

[0150] 1 ml Xylene (Biolab) was added to 1-2 mg tissue, incubated at 57^{0}C for 5 min and centrifuged for 2 min at 10,000g. The supernatant was removed and 1 ml Ethanol (100%) (Biolab) was added. Following centrifugation for 10 min at 10,000g, the supernatant was discarded and the washing procedure was repeated. Following air drying for 10-15 min, 500 μl Buffer B (NaCl 10mM, Tris pH 7.6, 500 mM, EDTA 20mM, SDS 1%) and 5ul proteinase K (50mg/ml) (Sigma) were added. Following incubation at 45^{0}C for16 h, inactivation of the proteinase K at 100^{0}C for 7 min was preformed. Following extraction with acid phenol chloroform (1:1) (Sigma) and centrifugation for 10 min at maximum speed at 4^{0}C, the upper

phase was transferred to a new tube with the addition of 3 volumes of 100% Ethanol, 0.1 volume of NaOAc (BioLab) and 8 µl glycogen (Ambion) and left over night at -20°C.

[0151] Following centrifugation at maximum speed for 40min at 4°C, washing with 1ml Ethanol (85%), and drying, the RNA was re-suspended in 45µl DDW.

[0152] The RNA concentration was tested and DNase Turbo (Ambion) was added accordingly (1µl DNase/10 µg RNA). Following incubation for 30 min at room temperature and extraction with acid phenol chloroform, the RNA was re-suspend in 45 µl DDW. The RNA concentration was tested again and DNase Turbo (Ambion) was added accordingly (1µl DNase/10 µg RNA). Following incubation for 30 min at room temperature and extraction with acid phenol chloroform, the RNA was re-suspend in 20µl DDW.

4. Data normalization

[0153] The initial data set was based on signals measured for multiple probes for every sample. For the analysis, signals were used only for probes that were designed to measure the expression levels of known or validated human microRNAs. Data was normalized separately for each group of samples relative to a reference data set, containing one value for each probe. The vector of numbers for the reference data set is represented by R. The reference data set was calculated in one of two ways:

[0154] (1)   For the expression data (Table 2), each sample type was normalized separately, for example: Lung adenocarcinoma samples. For a given sample type, the value of the reference data set for a given probe was calculated as the median expression of this probe in all samples of the given type.

[0155] (2)   For the differential expression analysis (data in Table 3), data was normalized separately for each indication relative to a reference data set that was calculated as follows: Each indication consisted of two sample types between which to distinguish, for example: small intestine carcinoid tumor and small intestine stromal tumor. For a given indication, the value of the reference data set for a given probe was calculated as the mean expression of two values: the median expression of this probe in each of the two tumor/tissue types. For example, samples from small intestine carcinoid tumor and small intestine stromal tumor were normalized relative to a reference data set, where the reference value for each probe was the mean expression of the following two values: the median expression of this probe in the all carcinoid tumor samples and the median expression of this probe in all stromal tumor samples.

[0156] For each indication (or sample type), the signals of the reference data set (represented by R) were log2 transformed, so that $R2=\log2(R)$. For each sample, signals for all probes (represented by S) were log2 transformed, so that $S2=\log2(S)$. A second degree polynomial, represented by F, was found for each sample so as to provide the best fit between the sample data and the reference data, such that $R2 \approx F(S2)$. In this process, remote data points ("outliers") are not used for calculating the polynomial F. For each probe in the sample, the normalized value V is calculated from the initial value I by transforming it with the polynomial function F, so that $V=F(I)$.

5. Statistical analysis

a. Expression

[0157] The purpose of this statistical analysis was to identify microRNAs with expression level significantly higher than the background expression level. The background signal was calculated based on the signal levels of the negative control probes for each sample type. The background level was determined as the 90th percentile of the negative control probes signal values in all samples of the same tissue/tumor. For each probe, we compared the group of normalized signals obtained for a specific tissue/tumor sample set relative to the background signal level. We used a two-sided t-test of the hypothesis that the probe signal levels come from a distribution whose mean is the background signal level. A probe was considered "expressed" if it had a t-test p-value lower than 0.05.

b. Differential expression

[0158] The purpose of this statistical analysis was to find probes whose normalized signal levels differ significantly between the two sample sets being compared. Probes that had normalized signal levels below $\log2(500)$ in the two sample sets were not analyzed. For each probe, we compared two groups of normalized signals obtained for two sample sets. For each probe, we calculated the p-value, which is the probability that we would obtain by chance the measured signals or a more extreme difference between the groups if the two groups of signal come from distributions with equal mean values, using the statistical un-paired two-sided t-test method. We selected microRNAs whose probes had t-test p-values lower than 0.05. A p-value lower than 0.05 means that the probability that the two groups come from distributions with the same mean is lower than 0.05 or 5%. The two groups of signal are likely to result from distributions with

different means, and the relevant microRNA is likely to be differentially expressed between the two sets of samples.

c.   **The probes types**

[0159] The probes used are: mirVana™ miRNA Probe Set V2 (Ambion) and other probes designed by Rosetta Genomics. The 3' amino modifiers of the probes: /3AmMC6/. The chemical structure was as follows:

[0160] A primary amino group can be used to attach a variety of modifiers (such as fluorescent dyes) to an oligonucleotide or used to attach an oligonucleotide to a solid surface. Amino modifiers can be positioned at the 5'-end with either a standard (C6) or longer (C12) spacer arm. Amino modifications can be positioned at the 3'-end. Internal amino modifications can be introduced using an amino-dT base.

Example 1

Specific microRNAs are expressed in different types of cancers and normal tissues

[0161] The normalized miRdicator™ arrays mean signals are presented in Table 2. The results show a significant expression pattern of specific miRs in different tumors or normal tissues.

Example 2

Specific microRNAs are able to distinguish between lung tumor and normal lung tissue

[0162] The statistical analysis of the normalized miRdicator™ arrays results of lung tumor versus normal lung tissue are presented in Tables 3-4. The results show a significant difference in the expression pattern of several miRs, most prominent among them being hsa-miR-126 (SEQ ID NO: 204).

[0163] The normalized expression levels of hsa-miR-126 were found to be higher in lung normal tissue in comparison to lung tumor, as measured by miRdicator™ array (Figure 1). The normalized expression levels of hsa-miR-126 were found to decrease in lung tumors in comparison to normal lung tissue.

Example 3

Specific microRNAs are able to distinguish between liver metastases and primary liver tumors

[0164] The statistical analysis of the miRdicator™ arrays results of liver metastases vs. primary liver tumors are presented in Tables 3-4. The results show a significant difference in the expression pattern of several miRs, most prominent among them being hsa-miR-200c (SEQ ID NO: 156). The normalized expression levels of hsa-miR-200c were found to increase in liver metastasis in comparison to primary liver tumors (Figure 2).

Example 4

Specific microRNAs are able to distinguish between small intestine carcinoid tumor and small intestine stromal tumor

[0165] The statistical analysis of the normalized miRdicator™ arrays results of small intestine carcinoid tumor versus small intestine stromal tumor are presented in Tables 3-4. The results show a significant difference in the expression pattern of several miRs, most prominent among them being hsa-miR-375 (SEQ ID NO: 273).

[0166] The normalized expression levels of hsa-miR-375 were found to be higher in small intestine carcinoid tumors in comparison to small intestine stromal tumors, as measured by miRdicator™ array (Figure 3).

Example 5

Specific microRNAs are able to distinguish between different grades of bladder carcinoma

[0167] The statistical analysis of the miRdicator™ arrays results of high grade bladder carcinoma versus medium grade bladder carcinoma or low grade bladder carcinoma are presented in Tables 3-4. The results exhibited a significant difference in the expression pattern of 43_5 (SEQ ID NO: 215). For Table 4, "DID" is the Decision ID, describing the comparisons of one group of tissue(s) to another group of tissue(s).

[0168] The normalized expression levels of 43_5 were found to increase in medium grade bladder carcinoma in comparison to high grade bladder carcinoma, as measured by miRdicator™ array (Figure 4).

CLAIMS

1.   A probe comprising a nucleic acid sequence selected from the group consisting of:

      (a)   any one of SEQ ID NOS: 1-561.

      (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 probe of claim 1 wherein the nucleic acid sequence comprises a modified base.

3.   The probe of claim 1, wherein the nucleic acid is selected from the group consisting of:

      (a)   a hairpin

      (b)   a pri-miRNA

      (c)   a pre-miRNA; and

      (d)   a miRNA.

4.   A plurality of probes according to claim 1.

5.   The plurality of probes of claim 4, wherein the probes comprise nucleic acid sequences associated with a specific cancer, as designated in Table 4.

6.   A biochip comprising the probe of claim 1.

7.   A biochip comprising the plurality of probes of claim 4.

8.   A method for detecting a cancer-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 different from a control is indicative of a cancer-associated nucleic acid.

9.   The method of claim 8, wherein said cancer is selected from the group consisting of small intestine cancer, bladder cancer, lung cancer, thyroid cancer, uterus cancer, liver cancer, kidney cancer, breast cancer, stomach cancer, testicular cancer, cervical cancer, esophageal cancer, gallbladder cancer, ovarian cancer, colon cancer, melanoma and prostate cancer.

10. A method of diagnosing a subject with a specific cancer comprising:

    (a)   providing a biological sample from the subject; and

    (b)   measuring the level of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 1-561,

        wherein a level of the nucleic acid different from a control is indicative of said cancer.

11. The method of claim 10, wherein said cancer is selected from the group consisting of. small intestine cancer, bladder cancer, lung cancer, thyroid cancer, uterus cancer, liver cancer, kidney cancer, breast cancer, stomach cancer, testicular cancer, cervical cancer, esophageal cancer, gallbladder cancer, ovarian cancer, colon cancer, melanoma and prostate cancer.

12. The method of claim 11, wherein said prostate cancer is selected from the group consisting of prostate adenocarcinoma, prostate sarcoma and benign prostatic hyperplasia (BPH).

13. The method of claim 11, wherein said lung cancer is selected from the group consisting of lung squamous cell carcinoma, lung undifferentiated small cell carcinoma, lung undifferentiated large cell carcinoma, lung adenocarcinoma, nonsmall-cell lung cancer (NSCLC), alveolar carcinoma, bronchial adenoma, lung sarcoma, lung lymphoma, lung chondromatous hanlartoma and lung inesothelioma.

14. The method of claim 11, wherein said small intestine cancer is selected from the group consisting of small intestine adenocarcinoma, small intestine stromal tumor and small intestine carcinoid tumor.

15. The method of claim 10, wherein said biological sample is selected from the group consisting of bodily fluid, a cell line and a tissue sample.

16. The method of claim 15, wherein said tissue is a frozen, fixed, wax-embedded or formalin fixed paraffin-embedded (FFPE) tissue.

17. A method to distinguish between small intestine carcinoid tumor and small intestine stromal tumor, the method comprising determining a level of a nucleic acid sequence as set forth in Table 3 in a small intestine sample.

18. A method to distinguish between lung adenocarcinoma and squamous cell carcinoma, the method comprising determining a level of a nucleic acid sequence as set forth in Table 3 in a lung sample.

19. A method to distinguish between lung undifferentiated large cell carcinoma and squamous cell carcinoma, the method comprising determining a level of a nucleic acid sequence as set forth in Table 3 in a lung sample.

20. A method to distinguish between prostate adenocarcinoma and benign prostatic hyperplasia (BPH), the method comprising determining a level of a nucleic acid sequence as set forth in Table 3 in a prostate sample.
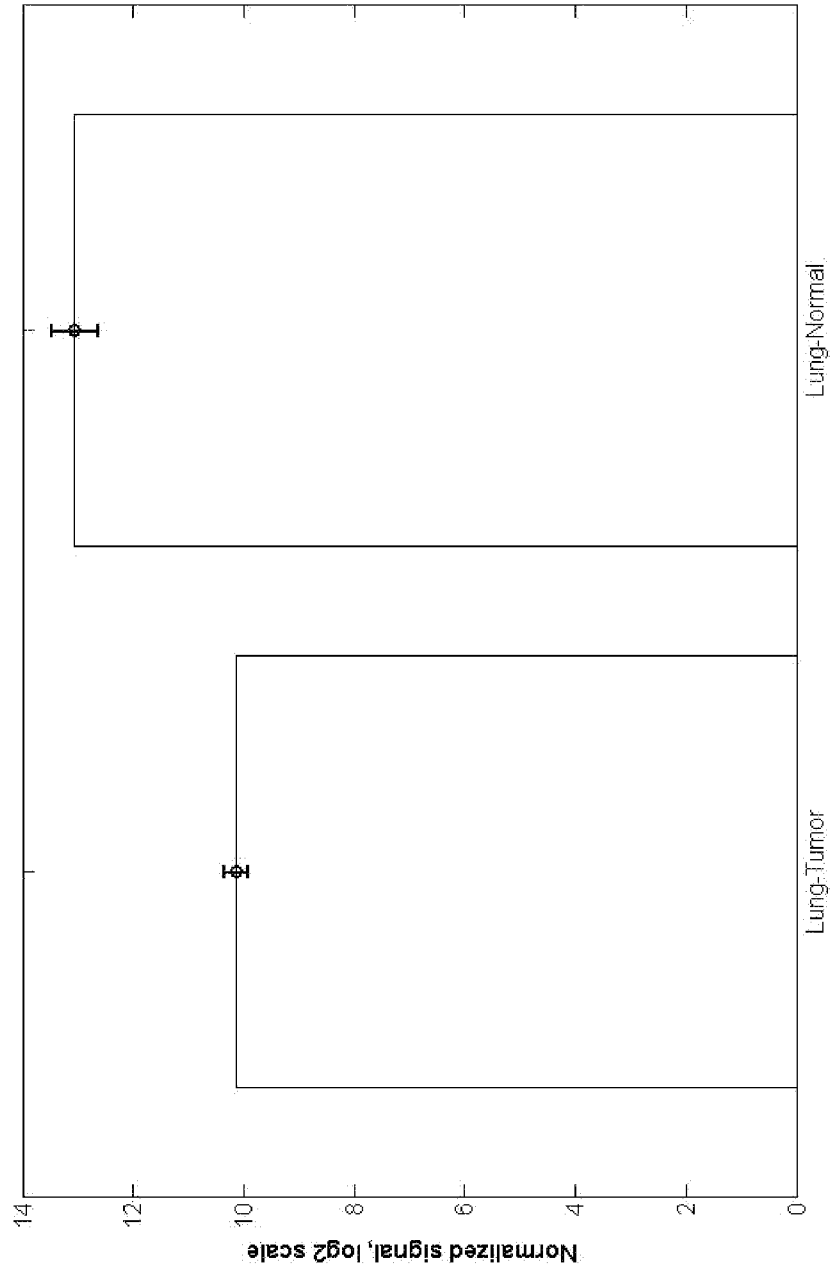
21. A method to distinguish between high grade carcinoma, medium grade carcinoma and low grade carcinoma, the method comprising determining a level of a nucleic acid sequence as set forth in Table 3 in a carcinoma sample.

22. The method of claim 21 wherein said carcinoma is selected from the group consisting of bladder carcinoma, breast carcinoma, colon carcinoma, lung carcinoma and prostate carcinoma.
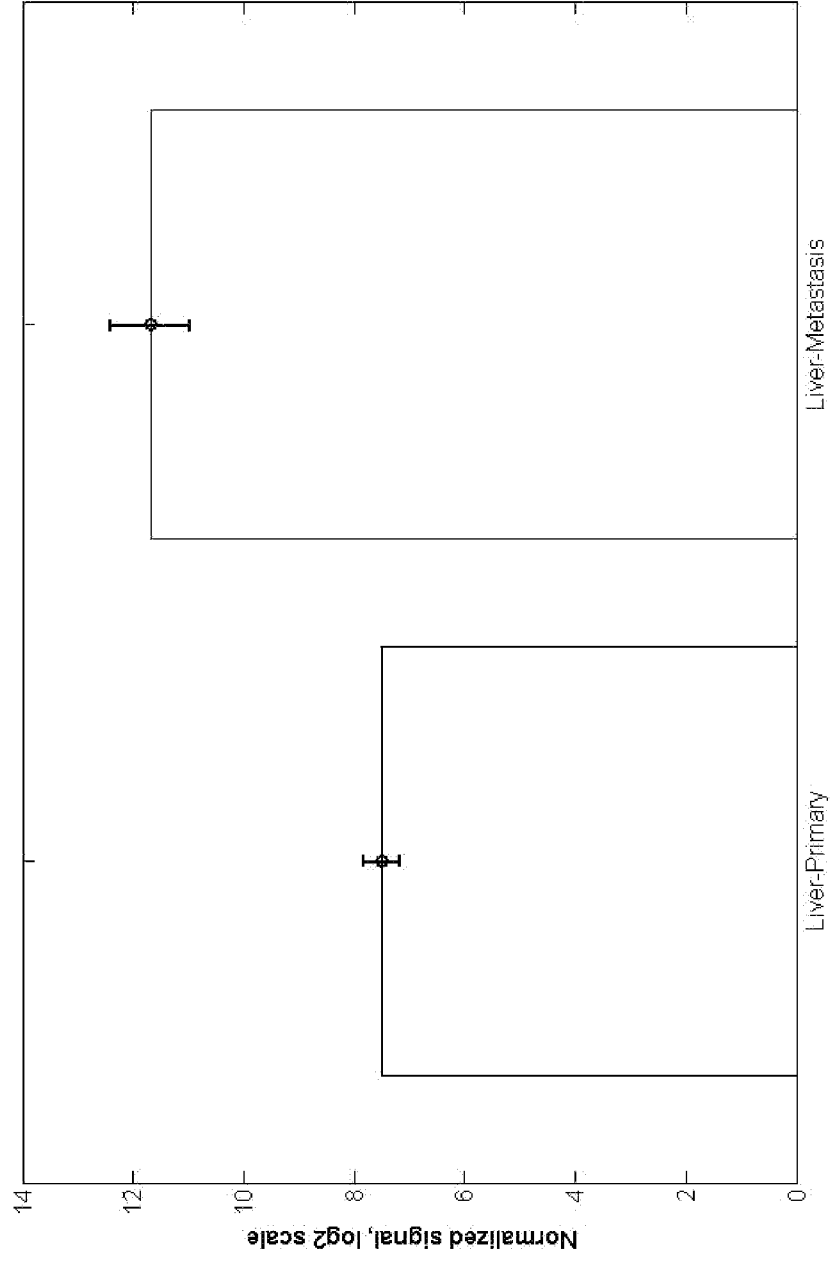
23. A method to distinguish between primary liver cancer and liver metastasis, the method comprising determining a level of a nucleic acid sequence in a liver sample.

24. A kit for diagnosing a subject with a specific cancer, said kit comprising the probe of claim 1.
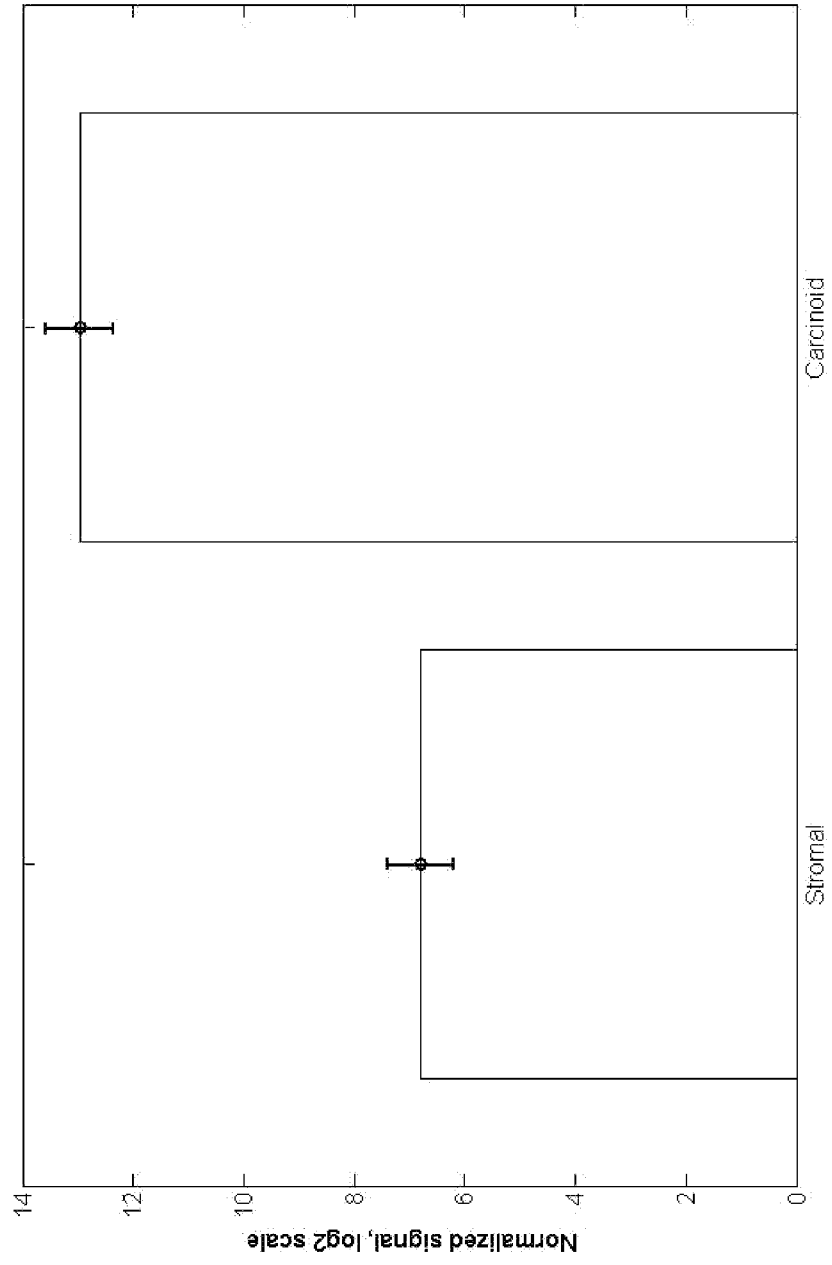
1/4

Figure 1

Figure 2

Figure 3

Figure 4