(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 March 2011 (03.03.2011)

PCT

(10) International Publication Number
WO 2011/024157 A1

(51) International Patent Classification:
C07H 21/00 (2006.01) *G01N 33/48* (2006.01)
A61K 31/7088 (2006.01)

(21) International Application Number:
PCT/IL2010/000462

(22) International Filing Date:
10 June 2010 (10.06.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/236,090 23 August 2009 (23.08.2009) US
61/330,920 4 May 2010 (04.05.2010) US

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

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SITBON, Einat** [IL/IL]; Chaim Teper 6, 74042 Nes Ziona (IL). **LITHWICK YANAI, Gila** [IL/IL]; Hamelitz 8, apartment 4, 93116 Jerusalem (IL). **MEIRI, Eti** [IL/IL]; Lachish 131, 60850 Shoam (IL). **DROMI, Nir** [IL/IL]; Gorediski 14, 76227 Rehovot (IL). **LEVY, Asaf** [IL/IL]; Ofek 2, apartment 6, 74027 Nes Ziona (IL).

(74) Agent: **SHOSHANI-KUPITZ, Tzipora**; Rosetta Genomics Ltd., 10 Plaut Street, 76706 Rehovot (IL).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2011/024157 A1

(54) **Title**: NUCLEIC ACID SEQUENCES RELATED TO CANCER

(57) **Abstract**: Disclosed are microRNA molecules, as well as various nucleic acid molecules relating thereto or derived therefrom Further disclosed are methods and compositions that can be used for diagnosis of cancer.

NUCLEIC ACID SEQUENCES RELATED TO CANCER

CROSS REFERENCE TO RELATED APPLICATIONS

The present application claims priority under 35 U.S.C. § 119(e) to U.S. Provisional Applications No. 61/236,090 filed Aug. 23, 2009 and No. 61/330,920 filed May 4, 2010, which are herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

The invention relates to microRNA molecules, as well as various nucleic acid molecules relating thereto or derived therefrom. The invention also relates to methods and compositions that can be used for diagnosis of cancer.

BACKGROUND OF THE INVENTION

microRNAs (miRNAs, miRs) are endogenous non-coding small RNAs that negatively regulate gene expression by interfering with the translation of coding messenger RNAs (mRNAs) in a sequence-specific manner, thereby playing a critical role in the control of gene expression during development and tissue homeostasis (Yi *et al.*, Nat Genet 2006;38:356-362). Certain miRNAs have been shown to be deregulated in human cancer, and their specific over- or under-expression has been shown to correlate with particular tumor types (Calin and Croce, Nat Rev Cancer 2006;6:857-866), as well as predict patient outcome (Yu *et al.*, Cancer Cell 2008;13:48-57). In some cases miRNA over-expression results in reduced expression of tumor suppressor genes, while loss of miRNA expression often leads to oncogene activation. Recent work has shown an essential role for miRNA deregulation in breast cancer metastasis, and in other tumor types. Differential expression of miRNAs may therefore indicate specific disease states.

Deep sequencing is method of high-throughput DNA sequencing using a novel highly parallel sequencing-by-synthesis approach, which allows rapid sequencing of millions of bases, and even whole genomes. The technique can be used to sequence any double-stranded DNA and can be used for *de novo* whole genome sequencing, re-sequencing of whole genomes and target DNA regions, metagenomics and RNA analysis. It is based on an emulsion-based method, in which short adaptors are ligated onto the ends of sequence fragments, which are then immobilized onto beads. The beads are then emulsified with the amplification reagents in a water-in-oil mixture, and are clonally

amplified within the emulsion droplets. Sequencing-by-synthesis is then performed by pyrosequencing in wells on a fibreoptic slide.

Deep sequencing methods have been widely used in recent years. These high throughput and highly sensitive sequencing methods include Roche Applied Sciences (454) GS, Illumina's Solexa 1G sequencer, and Applied Biosystem's SOLiD system. Deep sequencing can be used for the discovery of novel miRNA species and other small RNAs that are missed by traditional sequencing of small RNA libraries. Human microRNAs were previously identified using deep sequencing (Bar, M. *et al.* (2008) *Stem Cells*, 26, 2496-2505). However, the miRNA content of solid human tumors has only been partially explored using these methods and yet-unknown miRNAs and other small RNAs may be part of the tumor transcriptome.

Deep sequencing may be used to identify miRNAs and their differential expression in tissue samples, and may thus aid in distinguishing between primary tumors and cancer metastasis. Being able to distinguish between primary tumors and cancer metastasis, as well as distinguishing between metastases of different origins, has practical importance for choice of therapy. Diagnosis of specific tumors is also of great importance when choosing appropriate treatment. miR analogs, as well as anti-sense sequences of miRs, were recently shown to be useful as a therapeutic agent in several cancers. The miRNA content of solid human tumors has only been partially explored using these methods and yet-unknown miRNAs and other small RNAs may be part of the tumor transcriptome. Thus, there exists a need to identify nucleic acid sequences which will aid in cancer diagnosis.

SUMMARY OF THE INVENTION

The present invention is based in part on deep sequencing analysis of miRNAs from tumor specimens of different types. A computational approach was used to identify known miRNA sequences, miRNA sequence variants (isomiRs), and novel small RNA species in these tumors. Subsequently, normal and tumor samples from various tissue types were hybridized to a miRNA-microarray containing the novel miRNAs and known miRNAs. Some of the novel miRNAs are abundantly expressed in different types of tumors and others are expressed differently between tumor and non-tumor samples, between different tumor stages or between different types of tumors. In addition, using RT-PCR as a third platform the expression of several novel small RNAs was confirmed

in normal human serum. These new cancer miRNA candidates can potentially be used as diagnostic biomarkers or therapeutic targets in different types of cancer.

The present invention provides nucleic acid sequences related to cancer, and methods and compositions that can be used for diagnosis of cancer.

In one embodiment, the present invention provides an isolated nucleic acid comprising a sequence selected from the group consisting of:

(a) SEQ ID NOS: 31, 92-93, 450-454, 1, 2, 4-7, 9-26, 28-30, 32-35, 37-43, 46-72, 74-81, 83-91, 95-102, 104-108, 110-114, 116-157, 159-175 and 472-495;

(b) a DNA encoding (a);

(c) the complementary sequence of any one of (a) and (b); and

(d) a sequence at least 80% identical to (a) - (c),

wherein said nucleic acid is 16-26 nucleotides in length.

In one embodiment, the present invention provides an isolated nucleic acid comprising a sequence selected from the group consisting of:

(a) SEQ ID NOS: 31, 92-93, 450-454, 1, 2, 4-7, 9-26, 28-30, 32-35, 37-43, 46-72, 74-81, 83-91, 95-102, 104-108, 110-114, 116-157, 159-175 and 472-495;

(b) a DNA encoding (a);

(c) the complementary sequence of any one of (a) and (b); and

(d) a sequence at least 90% identical to (a) - (c),

wherein said nucleic acid is 16-26 nucleotides in length.

In one embodiment, the present invention provides an isolated nucleic acid comprising a sequence selected from the group consisting of:

(a) SEQ ID NOS: 192, 232-242, 176-187, 189-191, 193-196, 198-231, 243-267, 269-326, 328-330, 332-340, 342, 343, 345-350, 352-361, 363-369, 371-387, 389-413, 415-432, 434-438, 440-449 and 496-514;

(b) a DNA encoding (a);

(c) the complementary sequence of any one of (a) and (b); and

(d) a sequence at least 80% identical to (a) - (c),

wherein said nucleic acid is 50-150 nucleotides in length.

In one embodiment, the present invention provides an isolated nucleic acid comprising a sequence selected from the group consisting of:

(a) SEQ ID NOS: 192, 232-242, 176-187, 189-191, 193-196, 198-231, 243-267, 269-326, 328-330, 332-340, 342, 343, 345-350, 352-361, 363-369, 371-387, 389-413, 415-432, 434-438, 440-449 and 496-514;

(b) a DNA encoding (a);

(c) the complementary sequence of any one of (a) and (b); and

(d) a sequence at least 90% identical to (a) - (c),

wherein said nucleic acid is 50-150 nucleotides in length.

In one embodiment, the present invention provides an isolated nucleic acid comprising an endogenous human siRNA. In one embodiment, the endogenous human siRNA comprises a sequence selected from the group consisting of:

(a) SEQ ID NOS: 450-454;

(b) a DNA encoding (a);

(c) the complementary sequence of any one of (a) and (b); and

(d) a sequence at least 80% identical to (a) - (c),

wherein said nucleic acid is 16-26 nucleotides in length.

In one embodiment, the present invention provides an isolated nucleic acid comprising an endogenous human siRNA. In one embodiment, the endogenous human siRNA comprises a sequence selected from the group consisting of:

(a) SEQ ID NOS: 450-454;

(b) a DNA encoding (a);

(c) the complementary sequence of any one of (a) and (b); and

(d) a sequence at least 90% identical to (a) - (c),

wherein said nucleic acid is 16-26 nucleotides in length.

In one embodiment, the present invention provides an isolated nucleic acid comprising a sequence selected from the group consisting of:

(a) SEQ ID NOS: 53 and 162;

(b) SEQ ID NOS: 70 and 110;

(c) SEQ ID NOS: 14 and 120;

(d) SEQ ID NOS: 63, 106 and 58; and

(e) SEQ ID NOS: 135 and 159.

In one embodiment, the isolated nucleic acid of the invention is a modified oligonucleotide.

In one embodiment, the invention provides a composition comprising an isolated nucleic acid of the invention. According to one embodiment the composition is suitable for diagnostic applications. According to another embodiment the composition is suitable for therapeutic applications. According to some embodiments the composition further comprises a pharmaceutically acceptable carrier. According to one embodiment the composition is a marker or modulator of cancer.

In one embodiment, the invention provides a recombinant expression vector comprising an isolated nucleic acid of the invention. According to another embodiment, the invention provides a probe comprising an isolated nucleic acid of the invention. According to another embodiment, the invention provides a biochip comprising the probe of the invention. According to another embodiment, the invention provides a host cell comprising an isolated nucleic acid of the invention.

According to another embodiment, the invention provides a method for diagnosing a cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 31, 4, 7, 11, 16, 17, 21, 22, 23, 26, 30, 33-35, 37, 39, 46, 47- 49, 51, 52, 53, 56, 58, 60, 63, 64, 66, 68, 71, 72, 74, 76-78, 83, 86-88, 90, 96, 98, 100, 101, 106, 110-114, 116, 117, 119-121, 127, 129, 130, 132-136, 138, 141, 144, 145, 147, 148, 152, 153, 157, 159-162, 165, 167, 169-174, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression levels of any of said nucleic acids in healthy controls,

wherein the comparison of said expression profile to said reference expression allows for diagnosis of said cancer.

According to some embodiments, the cancer is selected from the group consisting of colon, bladder, breast, lung, liver, kidney, ovarian, prostate, esophagus, cervix, and pancreatic cancer. According to some embodiments relatively high expression levels of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 31, 4, 7, 11, 17, 21, 22, 23, 26, 30, 33, 35, 37, 39, 46, 48, 49, 51, 52, 53, 56, 58, 60, 63, 64, 66, 68, 71, 72, 74, 76-78, 86-88, 90, 96, 98, 100, 101, 106, 110-114, 116,

117, 119-121, 127, 129, 130, 132-136, 138, 141, 144, 145, 147, 148, 152, 153, 157, 159-162, 165, 167 and 169-174, as compared to said reference expression profile, is indicative of cancer. According another embodiment, relatively low expression levels of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 16, 34, 47 and 83, as compared to said reference expression profile, is indicative of cancer.

According to another embodiment, the invention provides a method of diagnosing an increased risk of colon cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 21, 68, 111 and 174, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression levels of any of said nucleic acids in healthy controls,

wherein a high expression level of said nucleic acid sequence is indicative of an increased risk of colon cancer in a subject.

According to another embodiment, the invention provides a method of diagnosing an increased risk of colon cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of SEQ ID NO: 92, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein a low expression level of said nucleic acid sequence is indicative of an increased risk of colon cancer in a subject.

According to another embodiment, the invention provides a method of diagnosing colon cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 2, 15, 28, 29, 48, 58, 61, 70, 86, 97, 107, 110, 150, 156, 170 and 172, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of colon cancer.

According to another embodiment, the invention provides a method of diagnosing lung cancer in a subject a comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 1, 7, 19, 21, 22, 26, 30, 35, 37, 39, 43, 46, 51, 53, 58, 59, 63, 64, 68, 71, 72, 74, 78, 86, 87, 106, 110, 113, 116, 119, 125, 127, 132, 135, 141, 153, 159, 161, 164, 165 and 170-173, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of lung cancer.

According to another embodiment, the invention provides a method of diagnosing bladder cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 7, 26, 35, 37, 39, 53, 71, 72, 125, 127, 130, 132, 135, 159, 161, 165, 170 and 172, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of bladder cancer.

According to another embodiment, the invention provides a method of diagnosing bladder cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 34 and 83, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively low expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of bladder cancer.

According to another embodiment, the invention provides a method of diagnosing liver cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 11, 26, 71, 72, 77, 98, 134, 136, 153, 160, 170 and 171, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of liver cancer.

According to another embodiment, the invention provides a method of diagnosing liver cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 12, 14, 28-32, 34, 42, 43, 52-55, 67, 75, 84, 85, 89, 97,105, 123-126, 128, 129, 137 and 140 a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively low expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of liver cancer.

According to another embodiment, the invention provides a method of diagnosing an endometrial metastasis in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 22, 25, 29, 31, 37, 39, 53, 64, 68, 72, 76, 77, 78, 84, 113, 119, 121, 127, 130, 132, 133, 136, 153, 161, 170 and 171 a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of endometrial metastasis.

According to another embodiment, the invention provides a method of diagnosing an endometrial metastasis in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 16 and 57 a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively low expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of endometrial metastasis.

According to another embodiment, the invention provides a method of diagnosing kidney cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 26, 37, 39, 53, 71, 72, 98, 125, 127, 130, 135, 159 and 165, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of kidney cancer.

According to another embodiment, the invention provides a method of diagnosing kidney cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 16, 34, 83 and 140, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively low expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of kidney cancer.

According to another embodiment, the invention provides a method of diagnosing breast cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 79, 99, 122, 130, 153 and 154 a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively low expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of breast cancer.

According to some embodiments, the invention provides a method to distinguish between a primary lung tumor and a metastasis to the lung, said method comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 31, 92, 1, 2, 6, 7, 10, 11, 13, 17, 19-24, 28-30, 33, 40, 42, 43, 46, 49, 51, 56-59, 61, 63, 64, 68, 70, 71, 74, 75, 77, 78, 80, 81, 86-88, 90-91, 95, 96, 97, 100, 104, 106, 107, 110-113, 116, 117, 119, 122,129, 133, 138, 139, 141, 142, 144, 145-152, 154, 162, 164, 165, 169, 171-173, 175, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample;

(c) comparing said expression profile to a reference expression profile,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of a primary lung tumor.

According to some embodiments, the origin of the metastasis to the lung is selected from the group consisting of endometrium, kidney, larynx, melanocyte and salivary gland.

According to some embodiments, the subject of the invention is a human. According to some embodiments, the method of the invention is used to determine a course of treatment for said subject.

According to some embodiments, the biological sample obtained in the method of the invention is selected from the group consisting of bodily fluid, a cell line and a tissue sample. According to some embodiments, the bodily fluid is selected from the group consisting of whole blood and serum. According to some embodiments, the tissue is a fresh, frozen, fixed, wax-embedded or formalin fixed paraffin-embedded (FFPE) tissue.

According to some embodiments, the expression levels are determined by a method selected from the group consisting of nucleic acid hybridization, nucleic acid amplification, and a combination thereof. According to some embodiments, the nucleic acid hybridization is performed using a solid-phase nucleic acid biochip array or *in situ* hybridization. According to some embodiments, the nucleic acid amplification method is real-time PCR. According to some embodiments, the real-time PCR method comprises forward and reverse primers. According to some embodiments, the forward primer comprises a sequence selected from the group consisting of SEQ ID NOS: 466, 467, 469, a fragment thereof and a sequence at least about 80% identical thereto. According to some embodiments, the real-time PCR method further comprises a probe. According to some embodiments, the probe comprises a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 31, 4, 7, 11, 16, 17, 21, 22, 23, 26, 30, 33-35, 37, 39, 46, 47- 49, 51, 52, 53, 56, 58, 60, 63, 64, 66, 68, 71, 72, 74, 76-78, 83, 86-88, 90, 96, 98, 100, 101, 106, 110-114, 116, 117, 119-121, 127, 129, 130, 132-136, 138, 141, 144, 145, 147, 148, 152, 153, 157, 159-162, 165, 167, 169-174, a fragment thereof and a sequence at least about 80% identical thereto.

According to another embodiment, the invention provides a kit for diagnosing a cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NO: 31, 4, 7, 11, 16, 17, 21, 22, 23, 26, 30, 33-35, 37, 39, 46, 47- 49, 51, 52, 53, 56, 58, 60, 63, 64, 66, 68, 71, 72, 74, 76-78, 83, 86-88, 90, 96, 98, 100, 101, 106, 110-114, 116, 117, 119-121, 127, 129, 130, 132-136, 138, 141, 144, 145, 147, 148, 152, 153, 157, 159-162, 165, 167, 169-174, a fragment thereof and a sequence at least about 80% identical thereto.

According to another embodiment, the invention provides a kit for diagnosing an increased risk of colon cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 21, 68, 92, 111 and 174, a fragment thereof and a sequence at least about 80% identical thereto.

According to another embodiment, the invention provides a kit for diagnosing colon cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 2, 15, 28, 29, 48, 58, 61, 70, 86, 97, 107, 110, 150, 156, 170 and 172, a fragment thereof and a sequence at least about 80% identical thereto.

According to another embodiment, the invention provides a kit for diagnosing lung cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 1, 7, 19, 21, 22, 26, 30, 35, 37, 39, 43, 46, 51, 53, 58, 59, 63, 64, 68, 71, 72, 74, 78, 86, 87, 106, 110, 113, 116, 119, 125, 127, 132, 135, 141, 153, 159, 161, 164, 165 and 170-173, a fragment thereof and a sequence at least about 80% identical thereto.

According to another embodiment, the invention provides a kit for diagnosing bladder cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 7, 26, 34, 35, 37, 39, 53, 71, 72, 83, 125, 127, 130, 132, 135, 159, 161, 165, 170 and 172, a fragment thereof and a sequence at least about 80% identical thereto.

According to another embodiment, the invention provides a kit for diagnosing liver cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 11, 12, 14, 26, 28-32, 34, 42, 43, 52-55, 67, 71, 72, 75, 77, 84, 85, 89, 97, 89, 105, 123-

126, 128, 129, 134, 136, 137, 140, 153, 160, 170 and 171, a fragment thereof and a sequence at least about 80% identical thereto.

According to another embodiment, the invention provides a kit for diagnosing a subject with endometrial metastasis, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 16, 22, 25, 29, 31, 37, 39, 53, 57, 64, 68, 72, 76, 77, 78, 84, 113, 119, 121, 127, 130, 132, 133, 136, 153, 161, 170 and 171, a fragment thereof and a sequence at least about 80% identical thereto.

According to another embodiment, the invention provides a kit for diagnosing kidney cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 16, 26, 34, 37, 39, 53, 71, 72, 83, 98, 125, 127, 130, 135, 140, 159 and 165, a fragment thereof and a sequence at least about 80% identical thereto.

According to another embodiment, the invention provides a kit for diagnosing breast cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 79, 99, 122, 130, 153 and 154 a fragment thereof and a sequence at least about 80% identical thereto.

According to another embodiment, the invention provides a kit for distinguishing between a primary lung tumor and a metastasis to the lung, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 31, 92, 1, 2, 6, 7, 10, 11, 13, 17, 19-24, 28-30, 33, 40, 42, 43, 46, 49, 51, 56-59, 61, 63, 64, 68, 70, 71, 74, 75, 77, 78, 80, 81, 86-88, 90-91, 95, 96, 97, 100, 104, 106, 107, 110-113, 116, 117, 119, 122,129, 133, 138, 139, 141, 142, 144, 145-152, 154, 162, 164, 165, 169, 171-173, 175, a fragment thereof and a sequence at least about 80% identical thereto.

These and other embodiments of the present invention will become apparent in conjunction with the figures, description and claims that follow.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows differential expression of miRs (in $\log_2$ (fluorescence units)), comparing the median values of each miR in breast primary tumor (y-axis) with breast metastases into lymph nodes (x-axis). Median normalized fluorescence for each miRNA (black crosses) indicates expression levels as measured by microarray. Squares represent differentially expressed miRs. The parallel lines describe a fold change between groups of 1.5 in either direction.

Figure 2 shows differential expression of miRs (in $\log_2$(fluorescence units)), comparing the median values of each miR in colon tumors (y-axis) with the corresponding median for their adjacent tissues (x-axis). Median normalized fluorescence for each miRNA (black crosses) indicates expression levels as measured by microarray. Squares represent differentially expressed miRs. The parallel lines describe a fold change between groups of 1.5 in either direction.

Figure 3 shows differential expression of miRs (in $\log_2$(fluorescence units)), comparing the median values of each miR in lung tumors (y-axis) with the corresponding median for other tumors from the following tissues: bile duct, bladder, breast, colon, kidney, liver, lung, ovary, pancreas, and prostate. (x-axis). Median normalized fluorescence for each miRNA (black crosses) indicates expression levels as measured by microarray. Squares represent differentially expressed miRs. The parallel lines describe a fold change between groups of 1.5 in either direction.

Figure 4. Expression in RT-PCR of novel miRNAs and small RNAs, in human serum. RNA was measured in sera of 19 normal humans and in negative control not containing RNA. Shown is the median of expression signals (y-axis, in units of 42-*Ct*) for each miR in all tested samples. Black bars show expression in experimental samples and white bars show expression in negative controls.

DETAILED DESCRIPTION

The present invention extends the current knowledge of the tumor small RNA transcriptome and provides novel candidates for molecular biomarkers and drug targets.

This work demonstrates, using a sensitive method of next generation sequencing on RNA extracted from tumors, in addition to careful computational analysis and followed by verification experiments can identify yet unknown sequences such as the

new miRNAs, miRNA-offset RNAs (MORs), Y-RNA derived sequences and endogenous siRNAs presented in this analysis. The identification of such tumor-specific small RNAs, could lead to the development of new therapeutic targets, which may be utilized as a treatment more specific than the set of tools currently available.

The present invention provides methods and compositions for diagnosis of cancer and cancer metastasis. Other aspects of the invention will become apparent to the skilled artisan by the following description of the invention.

Before the present 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.

For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 6-9, the numbers 7 and 8 are contemplated in addition to 6 and 9, and for the range 6.0-7.0, the numbers 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

1. Definitions

Attached

"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.

Biological sample

"Biological sample" as used herein means a sample of biological tissue or fluid that comprises nucleic acids. Such samples include, but are not limited to, tissue or fluid isolated from subjects. Biological samples may also include sections of tissues such as biopsy and autopsy samples, FFPE samples, frozen sections taken for histological 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.

Biological samples may also be blood, a blood fraction, urine, effusions, ascitic fluid, saliva, cerebrospinal fluid, cervical secretions, vaginal secretions, endometrial secretions, gastrointestinal secretions, bronchial secretions, sputum, cell line, tissue sample, cellular content of fine needle aspiration (FNA) or secretions from the breast. 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.

Cancer

The term "cancer" is meant to include 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, small cell lung, non-small cell lung (e.g., lung squamous cell carcinoma, lung adenocarcinoma and lung undifferentiated large cell carcinoma), 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.

Cancer prognosis

A forecast or prediction of the probable course or outcome of the cancer. As used herein, cancer prognosis includes the forecast or prediction of any one or more of the following: duration of survival of a patient susceptible to or diagnosed with a cancer, duration of recurrence-free survival, duration of progression-free survival of a patient susceptible to or diagnosed with a cancer, response rate in a group of patients susceptible to or diagnosed with a cancer, duration of response in a patient or a group of patients susceptible to or diagnosed with a cancer. As used herein, "prognostic for cancer" means providing a forecast or prediction of the probable course or outcome of the cancer. In some embodiments, "prognostic for cancer" comprises providing the forecast or prediction of (prognostic for) any one or more of the following: duration of survival of a patient susceptible to or diagnosed with a cancer, duration of recurrence-free survival, duration of progression-free survival of a patient susceptible to or diagnosed with a cancer, response rate in a group of patients susceptible to or diagnosed with a cancer, and duration of response in a patient or a group of patients susceptible to or diagnosed with a cancer.

Chemotherapeutic

A drug used to treat a disease, especially cancer. In relation to cancer the drugs typically target rapidly dividing cells, such as cancer cells.

Complement

"Complement" or "complementary" as used herein means Watson-Crick (*e.g.*, A-T/U and C-G) or Hoogsteen base pairing between nucleotides or nucleotide analogs of nucleic acid molecules. A full complement or fully complementary may mean 100% complementary base pairing between nucleotides or nucleotide analogs of nucleic acid molecules. In some embodiments, the complementary sequence has a reverse orientation (5'-3').

C_T

C_T signals represent the first cycle of PCR where amplification crosses a threshold (cycle threshold) of fluorescence. Accordingly, low values of C_T represent high abundance or expression levels of the microRNA.

In some embodiments the PCR C_T signal is normalized such that the normalized C_T remains inversed from the expression level. In other embodiments the PCR C_T signal may be normalized and then inverted such that low normalized-inverted C_T represents low abundance or expression levels of the microRNA.

Detection

"Detection" means detecting the presence of a component in a sample. Detection also means detecting the absence of a component. Detection also means measuring the level of a component, either quantitatively or qualitatively.

Differential expression

"Differential expression" may mean qualitative or quantitative differences in the temporal and/or cellular gene expression patterns within and among cells and tissue. Thus, a differentially expressed gene can 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 will exhibit an expression pattern within a state or cell type that may be detectable by standard techniques. Some genes will 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, 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.

Expression profile

"Expression profile", as used herein, may mean a genomic expression profile, e.g., an expression profile of microRNAs. Profiles may be generated by any convenient means for determining a level of a nucleic acid sequence, e.g., quantitative hybridization of microRNA, labeled microRNA, amplified microRNA, cRNA, etc., quantitative PCR, ELISA for quantification, and the like, and allow the analysis of differential gene expression between two samples. A subject or patient tumor sample, e.g., cells or collections thereof, e.g., tissues, is assayed. Samples are collected by any convenient method, as known in the art. Nucleic acid sequences of interest are nucleic acid sequences that are found to be predictive, including the nucleic acid sequences provided above, where the expression profile may include expression data for 5, 10, 20, 25, 50, 100 or more, including all of the listed nucleic acid sequences. The term "expression profile" may also mean measuring the abundance of the nucleic acid sequences in the measured samples.

Expression ratio

"Expression ratio", as used herein, refers to relative expression levels of two or more nucleic acids as determined by detecting the relative expression levels of the corresponding nucleic acids in a biological sample.

FDR

When performing multiple statistical tests, for example, in comparing the signal between two groups in multiple data features, there is an increasingly high probability of obtaining false positive results, by random differences between the groups that can reach levels that would otherwise be considered statistically significant. In order to limit the proportion of such false discoveries, statistical significance is defined only for data features in which the differences reach a p-value (such as by a two-sided t-test) below a threshold, which is dependent on the number of tests performed and the distribution of p-values obtained in these tests. FDR or false discovery rate is the probability that one of the "significant" results was actually false.

Fold change

The larger signal value divided by the smaller signal value.

Hairpin

"Hairpin", as used herein, refers to an area where single-stranded DNA or RNA has folded back on itself and nucleotides from the two strands have base paired, so that the resulting structure appears as a hairpin structure. The hairpin may comprise a first and a 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.

Gene

"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.

Identity

"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.

Increased risk of cancer

As used herein, "increased risk of cancer" may mean that the probability that a subject will develop cancer in the future is higher than that of a control subject.

Inhibit

"Inhibit", as used herein, may mean prevent, suppress, repress, reduce or eliminate.

Label

"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.

Logistic regression

Logistic regression is part of a category of statistical models called generalized linear models. Logistic regression allows one to predict a discrete outcome, such as group membership, from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. The dependent or response variable is dichotomous, for example, one of two possible types of cancer. Logistic regression models the natural log of the odds ratio, i.e., the ratio of the probability of belonging to the first group (P) over the probability of belonging to the second group ($1–P$), as a linear combination of the different expression levels (in log-space) and of other explaining variables. The logistic regression output can be used as a classifier by prescribing that a case or sample will be classified into the first type if P is greater than 0.5 or 50%. Alternatively, the calculated probability P can be used as a variable in other contexts such as a 1D or 2D threshold classifier.

metastasis

"Metastasis", as used herein, means the process by which cancer spreads from the place at which it first arose as a primary tumor (origin) to other locations in the body. The metastatic progression of a primary tumor reflects multiple stages, including

dissociation from neighboring primary tumor cells, survival in the circulation, and growth in a secondary location. The name of a specific metastasis refers to its origin.

miRNA or miR

"miRNA" or "miR", as used herein, may mean a non-coding RNA between 18 and 25 nucleobases in length, which is the product of cleavage of a pre-miRNA by the enzyme Dicer. Examples of mature miRNAs are found in the miRNA database known as Sanger miRBase (release 10).

miRNA precursor

"miRNA precursor", as used herein, may mean a transcript that originates from a genomic DNA and that comprises a non-coding, structured RNA comprising one or more miRNA sequences. For example, in certain embodiments a miRNA precursor is a pre-miRNA. In certain embodiments, a miRNA precursor is a pri-miRNA.

Mismatch

"Mismatch" means a nucleobase of a first nucleic acid that is not capable of pairing with a nucleobase at a corresponding position of a second nucleic acid.

Modified oligonucleotide

"Modified oligonucleotide" as used herein means an oligonucleotide having one or more modifications relative to a naturally occurring terminus, sugar, nucleobase, and/or internucleoside linkage. According to one embodiment, the modified oligonucleotide is a miRNA or siRNA comprising a modification (*e.g.* labeled). According to another embodiment, the modified oligonucleotide is complementary to a miRNA or siRNA.

Nucleic acid

"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.

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.

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 2005;438:685-689, Soutschek *et al.*, Nature 2004;432:173-178, and U.S. Patent Publication No. 20050107325, which are incorporated herein by reference. Additional modified nucleotides and nucleic acids are described in U.S. Patent Publication No. 20050182005, which is incorporated herein by reference. Modifications of the ribose-phosphate backbone may be done for a variety of reasons, e.g., to increase the stability and half-life of such molecules in physiological environments, to enhance diffusion across cell membranes, or as probes on a biochip. The backbone modification

may also enhance resistance to degradation, such as in the harsh endocytic environment of cells. The backbone modification may also reduce nucleic acid clearance by hepatocytes, such as in the liver and kidney. 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.

Probe

"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.

Pseudogene

As used herein, pseudogenes are defunct relatives of known genes that are no longer expressed in the cell. Although most pseudogenes have some gene-like features (such as promoters, CpG islands, and splice sites), they are nonetheless considered nonfunctional, due to their lack of protein-coding ability resulting from various genetic disablements (stop codons, frameshifts, or a lack of transcription) or their inability to encode RNA (such as with rRNA pseudogenes).

Reference expression profile

As used herein, the term "reference expression profile" means a profile of values that statistically correlates to a particular outcome when compared to an assay result. In preferred embodiments the reference profile values are determined from statistical analysis of studies that compare microRNA expression with known clinical outcomes. The reference values may be a threshold score value or a cutoff score value. Typically, a reference value will be a threshold above which one outcome is more probable and below which an alternative outcome is more probable.

Sensitivity

"Sensitivity", as used herein, may mean a statistical measure of how well a binary classification test correctly identifies a condition, for example, how frequently it correctly classifies a cancer into the correct type out of two possible types. The sensitivity for class A is the proportion of cases that are determined to belong to class "A" by the test out of the cases that are in class "A", as determined by some absolute or gold standard.

siRNA

"siRNA", as used herein, refers to small inhibitory RNA duplexes (generally 16-30 base pairs) that induce the RNA interference (RNAi) pathway. These molecules contain varying degrees of complementarity to their target mRNA in the antisense strand. Some, but not all, siRNA have unpaired overhanging bases on the 5' or 3' end of the sense strand and/or the antisense strand. The term "siRNA" includes duplexes of two separate strands, as well as single strands that can form hairpin structures comprising a duplex region.

Specificity

"Specificity", as used herein, may mean a statistical measure of how well a binary classification test correctly identifies a condition, for example, how frequently it correctly classifies a cancer into the correct type out of two possible types. The specificity for class A is the proportion of cases that are determined to belong to class "not A" by the test out of the cases that are in class "not A", as determined by some absolute or gold standard.

Stem-loop sequence

"Stem-loop sequence", as used herein, may mean an RNA having a hairpin structure and containing a mature miRNA sequence. Pre-miRNA sequences and stem-loop sequences may overlap. Examples of stem-loop sequences are found in the miRNA database known as Sanger miRBase (release 10).

Stringent hybridization conditions

"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.

Substantially complementary

"Substantially complementary" as used herein means 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.

Substantially identical

"Substantially identical" as used herein means that a first and a 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.

Subject

As used herein, the term "subject" refers to a mammal, including both human and other mammals. The methods of the present invention are preferably applied to human subjects.

Threshold expression level

As used herein, the phrase "threshold expression level" refers to a criterion expression profile to which measured values are compared in order to determine the prognosis of a subject with cancer. The reference expression profile may be based on the expression of the nucleic acids, or may be based on a combined metric score thereof.

Tissue sample

As used herein, a tissue sample is tissue obtained from a tissue biopsy using methods well known to those of ordinary skill in the related medical arts. The phrase "suspected of being cancerous", as used herein, means a cancer tissue sample believed by one of ordinary skill in the medical arts to contain cancerous cells. Methods for obtaining the sample from the biopsy include gross apportioning of a mass, microdissection, laser-based microdissection, or other art-known cell-separation methods.

Treat

"Treat" or "treating", as used herein when referring to protection of a subject from a condition, may mean preventing, suppressing, repressing, or eliminating the condition. Preventing the condition involves administering a composition described herein to a subject prior to onset of the condition. Suppressing the condition involves administering the composition to a subject after induction of the condition but before its clinical appearance. Repressing the condition involves administering the composition to a subject after clinical appearance of the condition such that the condition is reduced or prevented from worsening. Elimination of the condition involves administering the composition to a subject after clinical appearance of the condition such that the subject no longer suffers from the condition.

Tumor

"Tumor", as used herein, refers to all neoplastic cell growth and proliferation, whether malignant or benign, and all pre-cancerous and cancerous cells and tissues.

Variant

"Variant", as 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.

Vector

"Vector", as 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.

Y-RNA

"Y RNAs", as used herein, are small non-coding RNA components of the Ro ribonucleoprotein particle (Ro RNP). These small RNAs are predicted to fold into a conserved stem formed by the 3' and 5' ends of the RNA and characterized by a single bulged cytosine. In some embodiments, Y RNAs are over-expressed in human tumours. In some embodiments, Y RNAs are required for cell proliferation.

1D/2D threshold classifier

"1D/2D threshold classifier", as used herein, may mean an algorithm for classifying a case or sample such as a cancer sample into one of two possible types such as two types of cancer or two types of prognosis (e.g., good and bad). For a 1D threshold classifier, the decision is based on one variable and one predetermined threshold value; the sample is assigned to one class if the variable exceeds the threshold and to the other class if the variable is less than the threshold. A 2D threshold classifier is an algorithm for classifying into one of two types based on the values of two variables. A score may be calculated as a function (usually a continuous function) of the two variables; the decision is then reached by comparing the score to the predetermined threshold, similar to the 1D threshold classifier.

2. MicroRNAs and their processing

A gene coding for a miRNA may be transcribed, leading to production of an miRNA precursor known as the pri-miRNA. The pri-miRNA may be part of a polycistronic RNA comprising multiple pri-miRNAs. The pri-miRNA may form a hairpin with a stem and loop. The stem may comprise mismatched bases.

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 30–200 nucleotide precursor known as the pre-miRNA. Drosha may cleave the pri-miRNA with a staggered cut typical of Rnase III endonucleases yielding a pre-miRNA stem loop with a 5'

phosphate and ~2 nucleotide 3' overhang. Approximately one helical turn of stem (~10 nucleotides) extending beyond the Drosha cleavage site may be essential for efficient processing. The pre-miRNA may then be actively transported from the nucleus to the cytoplasm by Ran-GTP and the export receptor Ex-portin-5.

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

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.

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.

The RISC may identify target nucleic acids based on high levels of complementarity between the miRNA and the mRNA, especially by nucleotides 2-8 of the miRNA. Only one case has been reported in animals where the interaction between the miRNA and its target was along the entire length of the miRNA. This was shown for miR-196 and Hox B8 and it was further shown that miR-196 mediates the cleavage of the Hox B8 mRNA (Yekta *et al.*, Science 2004; 304:594-596). Otherwise, such interactions are known only in plants (Bartel & Bartel, Plant Physiol 2003; 132:709-717).

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, Cell 2004;116:281-297). In mammalian cells, the first 8 nucleotides of the miRNA may be important (Doench & Sharp, GenesDev 2004; 18:504-511). 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.*, PloS Biol 2005; 3:e85). Computation studies, in which miRNA binding on whole genomes is analyzed, have suggested a specific role for bases 2-7 at the 5' of the miRNA in target binding, but the role of the first nucleotide, found usually to be "A", was also recognized (Lewis *et al.*, Cell 2005;120:15-20). Similarly, nucleotides 1-7 or 2-8 were used by Krek *et al.*, Nat Genet 2005; 37:495-500) to identify and validate targets.

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.

miRNAs may direct the RISC to down-regulate 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.

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.

2.1 Nucleic Acids

Nucleic acids are provided herein. The nucleic acid may comprise the sequence of SEQ ID NOS: 1-514, or variants thereof. The variant may be a complement of the referenced nucleotide sequence. The variant may also be a nucleotide sequence that is substantially identical to the referenced nucleotide sequence or the complement thereof. The variant may also be a nucleotide sequence which hybridizes under stringent conditions to the referenced nucleotide sequence, complements thereof, or nucleotide sequences substantially identical thereto.

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 may be 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.

Table 1: miR and miR hairpin sequence identity numbers

microRNA ID	miR SEQ ID NO	Hairpin SEQ ID NO
15268	1	408
15294	2	176
15425	4	328
15438	5	389
15503	6	177,178
15571	7	390
15641	9	179
15684	10	374
15701	11	364
15760	12	180,181
15867	13	365
15907	14	409
15946	15	182
15947	16	183
15986	17	184
15998	18	185
16065	19	410
16103	20	186
16179	21	352
16206	22	430
16207	23	375
16252	24	391
16279	25	187
16318	26	411
16469	28	189
16475	29	392
16481	30	190, 191
16489	31	192
16512	32	340
16533	33	366
16582	34	193, 194, 195, 196
16626	35	329
16748	37	198, 199, 200
16758	39	201, 202, 442, 443, 444

microRNA ID	miR SEQ ID NO	Hairpin SEQ ID NO
16822	40	445
16825	41	412
16844	42	413
16869	43	353
17329	46	354
17375	47	446
17398	48	447
17508	49	367
17544	50	431
17576	51	448, 449, 203, 204, 205, 206, 207,. 208, 209
17591	52	415
17592	53	393
17593	54	416
17594	55	342
17628	56	417
17644	57	418
17663	58	210
17770	59	441
17866	60	211, 212
17901	61	213, 214, 215, 216, 217
17915	62	355
17945	63	218
18002	64	432
18047	65	376
18102	66	377
18112	67	356
18160	68	368
18279	69	330
18307	70	219
18336	71	220
18395	72	221
18468	74	378
18603	75	357
18670	76	222, 223, 224, 225, 226, 227
18788	77	343

microRNA ID	miR SEQ ID NO	Hairpin SEQ ID NO
18876	78	369
19003	79	419
19149	80	420
19186	81	394
19208	83	421
19264	84	422
19265	85	395
19340	86	423
19342	87	228, 229
19355	88	396
19371	89	397
19397	90	332
19420	91	230, 231
19433	92	232, 233, 234
19434	93	235, 236, 237, 238, 239, 240, 241, 242
19473	95	243
19485	96	333
19533	97	379
19544	98	334
19545	99	380
19549	100	381
19562	101	434
19572	102	335
19638	104	358
19640	105	435
19667	106	244
19770	107	398
19790	108	436
19898	110	245, 246, 247, 248, 249, 250, 251
19899	111	382
19920	112	359
19962	113	252, 253
20158	114	437
20179	116	254
20508	117	255, 256, 257

microRNA ID	miR SEQ ID NO	Hairpin SEQ ID NO
20516	118	383
20524	119	258
20703	120	400
20931	121	384
21033	122	425
21269	123	336
21270	124	401
21271	125	345
21272	126	360
21296	127	259, 260, 261, 262, 263, 264
21330	128	402
21342	129	265
21368	130	266, 267
21466	131	323
21493	132	324, 325, 326, 318, 320, 321, 322
21518	133	319
21531	134	403
21622	135	258, 269
21716	136	426
21803	137	404
22037	138	337
22124	139	385
22132	140	270
22331	141	271, 272, 273, 274, 275, 276, 277, 278, 279
22335	142	346
22406	143	371
22429	144	338
22655	145	405
22664	146	372
22673	147	438
22701	148	339
22736	149	427
22860	150	347
22875	151	280
22876	152	373

microRNA ID	miR SEQ ID NO	Hairpin SEQ ID NO
22912	153	348
22920	154	361
22923	155	386
22950	156	281
22998	157	349, 350
23017	159	282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302
23168	160	303
23178	161	304, 305, 306, 307, 308
23233	162	363
23252	163	387
23256	164	309
23291	165	310
23350	166	428
23717	167	440
23751	168	311
23752	169	429
23794	170	312
23795	171	313
23815	172	314, 315, 316, 317
23835	173	406
23983	174	407
24496	175	363
24545	38	368
32015 (siRNA)	450, 451, 452, 453	
32016 (siRNA)	454	
15275	480	496
15613	481	497
16049	482	510
16251	483	514
16453	484	498
16521	485	499
17111	486	500
17903	487	512
17963	488	501, 502

microRNA ID	miR SEQ ID NO	Hairpin SEQ ID NO
18078	489	503
18483	490	504, 505, 506
19148	491	507
19404	492	511
23314	493	513
24263	494	509
24078	495	508
16087	472	
32017	473	
17711	474	
32018	475	
19667	476	
32020	477	
23406	478	
24036	479	

Predicted precursor positions, including genomic location, are presented in Table 2 below. In one embodiment, a miRNA may be identified at a single genome locus. In another embodiment, a miRNA may be identified within multiple genomic loci.

5

Table 2: Predicted precursor positions

SEQ ID NO	Chromosome name	Chromosome strand	Chromosome start	Chromosome end
1	chr9	+	125388643	125388755
2	chr3	-	186618946	186619020
4	chr2	+	183079734	183079789
5	chr2	+	51324258	51324322
6	chr1	-	169906295	169906352
6	chr17	-	4789957	4790014
7	chr11	-	122960021	122960105
9	chr1	-	203357669	203357727
10	chr12	-	124694681	124694784
11	chrX	+	46963545	46963647
12	chr7	-	140875084	140875162
12	chr8	+	122112924	122112983
13	chr13	+	26434224	26434283
14	chr19	-	54820985	54821050
15	chr12	-	131892299	131892359
16	chr15	-	43278112	43278184
17	chr14	-	19881248	19881377
18	chr22	-	29082775	29082903
19	chr22	-	36452045	36452108
20	chr2	-	181446437	181446536
21	chr9	-	132108275	132108330
22	chr9	-	118374328	118374384
23	chr7	-	20336775	20336829
24	chr7	+	47445639	47445716
25	chr1	-	1499795	1499924
26	chr1	-	149813800	149813902
28	chr11	+	8298713	8298804
29	chr16	+	4801896	4801988
30	chr17	+	41658728	41658833
30	chr17	+	42532854	42532976

SEQ ID NO	Chromosome name	Chromosome strand	Chromosome start	Chromosome end
31	chr1	-	202879413	202879470
32	chr1	+	181789401	181789480
33	chr5	-	89861173	89861239
34	chr1	-	556268	556343
34	chr14	+	32023781	32023851
34	chrMT	-	5856	5931
34	chr21_random	+	928122	928192
35	chrX	+	128949752	128949807
37	chr1	+	40188880	40188977
37	chr10	-	96926496	96926567
37	chr18	+	1681816	1681874
38	chr20	+	38744218	38744277
39	chr2	-	117191800	117191863
39	chr4	+	21950082	21950153
39	chr18	-	27327202	27327328
39	chr19	-	20125898	20125954
39	chrX	+	4703041	4703097
40	chr1	-	35795436	35795513
41	chr16	-	30884414	30884469
42	chr3	-	47011000	47011062
43	chr10	+	21764175	21764304
46	chr22	-	22881964	22882019
47	chrMT	+	10271	10389
48	chr6	+	134316075	134316175
49	chr12	-	6801731	6801802
50	chr4	+	80303810	80303902
51	chr21	+	44475221	44475301
51	chr21	-	31270258	31270324
51	chr21	-	38621570	38621648
51	chr21	-	43803994	43804051
51	chrY	+	1316200	1316280

SEQ ID NO	Chromosome name	Chromosome strand	Chromosome start	Chromosome end
51	chrY	+	4890454	4890534
51	chrY	-	12803715	12803770
51	chr9_random	-	977861	977985
51	chrX_random	-	1655866	1655977
52	chr1	-	36395618	36395691
53	chr1	+	32992461	32992515
54	chr2	+	146802108	146802163
55	chr9	-	122386320	122386382
56	chr17	+	38979195	38979257
57	chr8	+	51632206	51632266
58	chr20	-	26137777	26137869
59	chr11	-	102785578	102785639
60	chr11	-	92327459	92327548
60	chr15	-	73218705	73218834
61	chr8	-	81340880	81340973
61	chr10	+	98500024	98500097
61	chr11	+	14418885	14418940
61	chr13	-	53913404	53913520
61	chr21	+	25655995	25656068
62	chr1	+	27803998	27804061
63	chr20	-	26137996	26138062
64	chr4	-	182530292	182530355
65	chr1	+	149955955	149956009
66	chr13	+	35312134	35312220
67	chr22	+	25472767	25472840
68	chr3	-	14809957	14810011
69	chr2	-	176133862	176133938
70	chr12	-	19501592	19501685
71	chr1	+	79418034	79418107
72	chr2	-	76552889	76552955
74	chr12	+	83902942	83902996

SEQ ID NO	Chromosome name	Chromosome strand	Chromosome start	Chromosome end
75	chr12	-	54910008	54910078
76	chr7	-	151353230	151353319
76	chr8	+	35550396	35550477
76	chr8	+	141503824	141503885
76	chr14	+	49123238	49123321
76	chr14	-	49399044	49399127
76	chr17	-	73878566	73878620
77	chr15	+	18901193	18901250
78	chr15	-	64785918	64785978
79	chr3	-	161653575	161653698
80	chr18	+	14989764	14989847
81	chr15	+	59637880	59637935
83	chr14	-	105210457	105210525
84	chr10	-	114468974	114469070
85	chr17	+	45480066	45480131
86	chr11	-	76527050	76527112
87	chr7	-	22096507	22096567
87	chr18	-	9286869	9286953
88	chr12	+	52676921	52676977
89	chr16	-	66129558	66129617
90	chr14	-	68331032	68331118
91	chr1	-	166414834	166414962
91	chr2	-	99163785	99163877
92	chr7	-	6218863	6218932
92	chr7	-	148315246	148315329
92	chr11	-	13304318	13304447
93	chr7	+	148311780	148311877
93	chr10	+	128599366	128599443
93	chr11	-	47705094	47705169
93	chr13	-	98986577	98986652
93	chr15	-	73062670	73062772

SEQ ID NO	Chromosome name	Chromosome strand	Chromosome start	Chromosome end
93	chrX	+	19304811	19304894
93	chrX	-	16840336	16840397
93	chrX	-	116932237	116932338
95	chr17	-	15109319	15109447
96	chr1	-	216165065	216165135
97	chr19	-	50005472	50005529
98	chr2	-	111247733	111247797
99	chr9	-	133332875	133332929
100	chr17	-	59363451	59363510
101	chr8	-	98555568	98555627
102	chr2	+	60630994	60631048
104	chr3	+	156510506	156510564
105	chr17	-	7582029	7582106
106	chr20	-	26136807	26136929
107	chr10	-	112686897	112686970
108	chr20	+	10599686	10599748
110	chr1	+	192425468	192425594
110	chr2	-	106680721	106680840
110	chr5	-	14705626	14705728
110	chr5	-	43531461	43531534
110	chr6	-	74285020	74285122
110	chr7	-	22517340	22517398
110	chr12	-	19501028	19501157
111	chr12	-	108231400	108231489
112	chr18	-	48760420	48760489
113	chr7	+	142054426	142054491
113	chr9	+	5083206	5083294
114	chr14	+	60051330	60051466
116	chr11	+	124857238	124857327
117	chr4	+	58079062	58079116
117	chr6	+	30568718	30568839

SEQ ID NO	Chromosome name	Chromosome strand	Chromosome start	Chromosome end
117	chr16	-	19511659	19511786
118	chr22	+	16744004	16744071
119	chr1	-	142084816	142084943
120	chr19	-	54820985	54821050
121	chr1	+	112964451	112964513
122	chr10	+	80861887	80861945
123	chr1	+	38324290	38324361
124	chr17	-	34082627	34082744
125	chr12	-	53070806	53070906
126	chrX	-	11370539	11370600
127	chr4	+	84650503	84650580
127	chr7	+	152741746	152741822
127	chr7	-	102401297	102401389
127	chr14	-	101770256	101770322
127	chr14	-	101776378	101776436
127	chrX	-	83880564	83880640
128	chr1	+	27774334	27774396
129	chr9	+	126309476	126309532
130	chr6	-	23269623	23269723
130	chr12	+	94365484	94365562
131	chr10	-	105144001	105144149
132	chr21	-	27149306	27149409
132	chr2	-	75953092	75953150
132	chr2	-	106192559	106192636
132	chr1	+	46629053	46629108
132	chr19	-	55148085	55148142
132	chr12	+	57539008	57539107
132	chr17	+	50450249	50450351
133	chr3	+	177715568	177715650
134	chr22	+	31542120	31542205
135	chrY	+	23951513	23951644

SEQ ID NO	Chromosome name	Chromosome strand	Chromosome start	Chromosome end
135	chrY	-	26830046	26830177
136	chr10	-	21825066	21825128
137	chr19	-	51166540	51166594
138	chr17	-	77478870	77478950
139	chr18	+	10609992	10610048
140	chr12	+	105940477	105940597
141	chr2	-	132755168	132755232
141	chr3	+	112144699	112144809
141	chr3	+	122716195	122716274
141	chr4	+	61218026	61218127
141	chr7	+	24256821	24256918
141	chr14	-	19268163	19268248
141	chr15	-	19816998	19817103
141	chr19	-	23975877	23975989
141	chrX	+	82995155	82995224
142	chr19	-	34709332	34709386
143	chr11	-	16785015	16785097
144	chrX	-	139088531	139088585
145	chr2	+	72225641	72225695
146	chr3	-	73780369	73780468
147	chr1	-	36462318	36462399
148	chr17	-	69860934	69861021
149	chr7	-	149116341	149116410
150	chr22	+	23838153	23838226
151	chr13	+	48916447	48916576
152	chr11	-	43559833	43559952
153	chr19	+	51854636	51854690
154	chr10	+	42523386	42523443
155	chr1	+	219468887	219468944
156	chr11	-	123117497	123117600
157	chr11	+	8663630	8663684

SEQ ID NO	Chromosome name	Chromosome strand	Chromosome start	Chromosome end
157	chr16	-	2786418	2786472
159	chr1	+	167363556	167363627
159	chr1	+	200735860	200735939
159	chr1	+	247135060	247135188
159	chr1	-	17312376	17312440
159	chr1	-	144110615	144110734
159	chr2	+	3630392	3630468
159	chr2	+	205927151	205927228
159	chr4	+	86462607	86462682
159	chr4	+	94295222	94295322
159	chr4	-	76873460	76873521
159	chr4	-	128179046	128179106
159	chr5	+	118852323	118852391
159	chr5	+	139120702	139120766
159	chr9	+	89814961	89815032
159	chr9	-	115180239	115180333
159	chr10	-	43766444	43766519
159	chr11	-	62415645	62415761
159	chr19	+	60700470	60700553
159	chrX	-	20540225	20540324
159	chrY	+	23951513	23951644
159	chrY	-	26830046	26830177
160	chr19	+	10081477	10081532
161	chr2	-	147785974	147786050
161	chr3	-	192859742	192859825
161	chr7	+	113813041	113813123
161	chr7	+	156358447	156358550
161	chr13	-	68694495	68694577
162	chr1	+	32992461	32992515
163	chr4	+	25266489	25266543
164	chr17	-	34564241	34564326

SEQ ID NO	Chromosome name	Chromosome strand	Chromosome start	Chromosome end
165	chr4	+	22439329	22439486
166	chr9	-	125343650	125343745
167	chr9	-	128936093	128936166
168	chrMT	+	16024	16083
169	chr5	+	87282036	87282100
170	chr12	+	105816650	105816712
171	chr5	+	180625273	180625362
172	chr2	+	208970363	208970431
172	chr3	-	173867074	173867158
172	chr10	-	91885741	91885850
172	chr11	-	17348435	17348527
173	chr12	-	95494567	95494621
174	chr20	+	44093753	44093846
175	chr1	-	153454102	153454270
480	chr 20	+	42470174	42470257
481	chr 8	+	32990291	32990381
482	chr 16	+	69120903	69121003
483	chr21_random	-	1679120	1679299
484	chr12	-	1639742	1639807
485	chr5	-	38593361	38593420
486	chr6_random	+	813777	813876
487	chr12	+	102848335	102848463
488	chr22	-	28059147	28059256
489	chr16	+	69120888	69120943
490	chrX	+	114844386	114844441
491	chr5	+	140007613	140007695
492	chr11	+	118394864	118394932
493	chr12	-	110959786	110959902
494	chr1	+	164143782	164143837
495	chr1	-	43882114	43882188

Sequence variants are presented in Table 3 below, together with counts for both the most abundant sequence and sequence variants.

Table 3: Sequence variants of miR sequences

SEQ ID NO	Most abundant sequence	Count	Variant 1 sequence	Count	Variant 2 sequence	Count	Variant 3 sequence	Count
37	ATCCCACTCCTGACACCA	2	AATCCCACTCCTGACACCA	1	CGAATCCCACTCCTGACACCA	8	GAATCCCACTCCTGACACCA	3
39	ATCCTGTTCGTGACGCCA	7	GAATCCTGTTCGTGACGCCA	3	TCGTATCCTGTTCGTGACGCCA	0	NULL	0
59	CCTGGTGATAGCTGGTTGTCCA	0	GTGATAGCTGGTTGTCCAA	1	TGGTGATAGCTGGTTGTCCA	1	NULL	0
71	GAACCCTACTCCTGGTACCA	1	AACCCTACTCCTGGTACCA	1	NULL	0	NULL	0
72	GAATCCCACTTCTGACACCA	18	AATCCCACTTCTGACACCA	1	CGAATCCCACTTCTGACACC	1	CGAATCCCACTTCTGACACCA	5
76	GATCGCGCCTGTGAATAGCCA	1	GGATCGCGCCTGTGAATAGCC	1	NULL	0	NULL	0
87	GGCAGGGCGCCCTGGAATGGGTT	2	GGCAGGGCGCCCTGGAATGG	3	NULL	0	NULL	0
92	GGCTGGTCCGAAGGTAGTGAGTT	20	CTGGTCCGAAGGTAGTGAGTT	1	GGCTGGTCCGAAGGTAGTG	1	GGCTGGTCCGAAGGTAGTGA	1

SEQ ID NO	Most abundant sequence	Count	Variant 1 sequence	Count	Variant 2 sequence	Count	Variant 3 sequence	Count
93	GGCTGGTCCGAGT GCAGTGGTGTTT	200	CTGGTCCGAGTG CAGTGGTGTT	1	GCTGGTCCGAG TGCAGTGGTGTT	3	GGCTGGTCCGA GTGCAGTG	5
119	TAGGTCAAGGTGT AGCCCATA	1	TAGGTCAAGGTG TAGCCCATGAG	2	NULL	0	NULL	0
127	T©CCGGGTTTCGG CACCA	5	AATCCCGGGTTT CGGCACCA	9	ATCCCGGGTTTC GGCACCA	16	CAATCCCGGGTT TCGGCACCA	14
130	TCCTCACACGGGG CACCA	3	CGATCCTCACAC GGGGCACCA	3	TTCGATCCTCAC ACGGGGCACC	1	NULL	0
131	TCGACCGGACCTC GACTGGCTCG	7	TCGACCGGACCT CGACTGGCT	1	TCGACCGGACC TCGACTGGCTCA	2	NULL	0
132	TCGGGTCCCTTCG TGGTCGCCA	4	CGGGTCCCTTCG TGGTCGCCA	7	TTCGGGTCCCTT CGTGGTCGCCA	1	NULL	0
133	TCGTTTCCCGGCC AACGCACCA	6	CGTTTCCCGGCC AACGCACCA	2	TTCGTTTCCCGG CCAACGCACCA	0	NULL	0
135	TCTCTGGTGGTCT AGTGGT	5	TCTCTGGTGGTC TAGTGGC	2	TCTCTGGTGGTC TAGTGGTT	1	TCTCTGGTGGTC TAGTGGTTAG	5
141	TGCAGATCTTGGT GGTA	1	GCAGGTGCAGAT CTTGGTGGTA	1	NULL	0	NULL	0
159	TGGTGGTCTAGTG	1	CCCTGGTGGTCT	1	CCCTGGTGGTCT	1	CTCTGGTGGTCT	1

SEQ ID NO	Most abundant sequence	Count	Variant 1 sequence	Count	Variant 2 sequence	Count	Variant 3 sequence	Count
	GTTA		AGTGGTTA		AGTGGTTAG		AGTGGTTAGGA	
161	TGTCCCTTCGTGG TCGCCA	1	CGTGTCCCTTCG TGGTCGCCA	2	TCGTGTCCCTTC GTGGTCGCCA	1	TTGTCCCTTCGT GGTCGCCA	0
164	TGTGCTCCGGAGT TACCTCGTTT	5	TGTGCTCCGGAG TTACCTCGTT	0	NULL	0	NULL	0
171	TTCCCGGCCCATG CACCA	2	CGATTCCCGGCC CATGCACCA	8	CGTTTCCCGGCC CATGCACCA	1	GATTCCCGGCCC ATGCACCA	26
172	TTCCTGGCCAATG CACCA	1	CGATTCCTGGCC AATGCACCA	1	TCGATTCCTGGC CAATGCACCA	1	NULL	0

MiRNAs may, in one embodiment, form clusters in the genome. In one embodiment, a cluster is defined based on the criterion of a distance of not more than 5000 nucleotides between miRNAs within the human genome. Most miRNA genes within 50 kb of each other have highly correlated expression patterns, with the correlation dropping sharply beyond the 50-kb range. Relatively few miRNAs are found between 50 kb and 500 kb of each other, as described in Baskerville and Bartel (RNA 2005; 11:241-247). Thus, in one embodiment, clustered miRNAs are defined as falling within a range of 0.1 kb and 50 kb.

In one embodiment, miRNAs located at this distance are co-transcribed. In another embodiment, miRNAs located at this distance are co-regulated. In yet another embodiment, miRNAs located at this distance are co-transcribed and co-regulated. In one embodiment, SEQ ID NOS: 53 and 162 appear in a cluster. In one embodiment, SEQ ID NOS: 70 and 110 appear in a cluster, separated by 500 nucleotides on chromosome 12. In one embodiment, SEQ ID NOS: 14 and 120 appear in a cluster. In one embodiment, SEQ ID NOS: 63, 106 and 58 appear in a cluster separated by 1000 nucleotides on chromosome 20. In one embodiment, SEQ ID NOS: 135 and 159 appear in a cluster.

a. Nucleic acid complex

The nucleic acid may further comprise one or more of the following: a peptide, a protein, a RNA-DNA hybrid, an antibody, an antibody fragment, a Fab fragment, and an aptamer. The nucleic acid may also comprise a protamine-antibody fusion protein as described in Song *et al.* (Nature Biotechnology 2005;23:709-717) and Rossi (Nature Biotechnology 2005;23:682-684), the contents of which are incorporated herein by reference. The protamine-fusion protein may comprise the abundant and highly basic cellular protein protamine. The protamine may readily interact with the nucleic acid. The protamine may comprise the entire 51-amino acid protamine peptide or a fragment thereof. The protamine may be covalently attached to another protein, which may be a Fab. The Fab may bind to a receptor expressed on a cell surface.

b. Pri-miRNA

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. A sequence of the pri-miRNA may comprise the sequence of SEQ ID NOS: 1, 2, 4-7, 9-26, 28-35, 37-

43, 46-72, 74-81, 83-93, 95-102, 104-108, 110-114, 116-157, 159-187, 189-196, 198-267, 269-326, 328-330, 332-340, 342, 343, 345-350, 352-361, 363-369, 371-387, 389-413, 415-432, 434-438, 440-449, 472-514 or variants thereof.

The pri-miRNA may form a hairpin structure. The hairpin may comprise 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 1994; 125:167-188), 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.

c. Pre-miRNA

The nucleic acid may also comprise a sequence of a pre-miRNA or a variant thereof. The pre-miRNA sequence may comprise from 45-200, 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. A sequence of the pre-miRNA may comprise the sequence of SEQ ID NOS: 1, 2, 4-7, 9-26, 28-35, 37-43, 46-72, 74-81, 83-93, 95-102, 104-108, 110-114, 116-157, 159-187, 189-196, 198-267, 269-326, 328-330, 332-340, 342, 343, 345-350, 352-361, 363-369, 371-387, 389-413, 415-432, 434-438, 440-449, 472-514 or variants thereof.

d. miRNA

The nucleic acid may also comprise a sequence of a miRNA (including miRNA*) or a variant thereof. The miRNA sequence may comprise from 13-33, 18-24 or 21-23 nucleotides. The miRNA may also comprise a total of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 nucleotides. The sequence of the miRNA may be the first 13-33 nucleotides of the pre-miRNA. The sequence of the miRNA may also be the last 13-33 nucleotides of the pre-miRNA. The sequence of the miRNA may comprise the sequence of SEQ ID NOS: 1, 2, 4-7, 9-26, 28-35, 37-43, 46-72, 74-81, 83-93, 95-102, 104-108, 110-114, 116-157, 159-175 and 472-495 or variants thereof.

e. Anti-miRNA

The nucleic acid may also comprise a sequence of an anti-miRNA that is capable of blocking the activity of a miRNA or miRNA*, such as by binding to the pri-miRNA, pre-miRNA, miRNA or miRNA* (e.g., antisense or RNA silencing), or by binding to the target binding site. The anti-miRNA may comprise a total of 5-100 or 10-60 nucleotides. The anti-miRNA may also comprise a total of at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 nucleotides. The sequence of the anti-miRNA may comprise (a) at least 5 nucleotides that are substantially identical or complementary to the 5' of a miRNA and at least 5-12 nucleotides that are substantially complementary to the flanking regions of the target site from the 5' end of the miRNA, or (b) at least 5-12 nucleotides that are substantially identical or complementary to the 3' of a miRNA and at least 5 nucleotides that are substantially complementary to the flanking region of the target site from the 3' end of the miRNA. A sequence of the anti-miRNA may comprise the complement of the sequence of SEQ ID NOS: 1, 2, 4-7, 9-26, 28-35, 37-43, 46-72, 74-81, 83-93, 95-102, 104-108, 110-114, 116-157, 159-175 and 472-495 or variants thereof.

f. siRNA

The nucleic acid may also comprise a sequence of a double-stranded RNA (dsRNA) that is capable of suppressing specific transcripts in a sequence-dependent manner. The dsRNA may be processed to provide small interfering RNAs (siRNAs). The siRNA may comprise a total of 20-25 nucleotides. Endogenous siRNAs have been identified in nematodes, plants and mammalian cells, including mouse oocytes. The sequence of the siRNA may comprise SEQ ID NOS: 450-454 or variants thereof. In one embodiment, these sequences may be found in the introns of two genes: ERBB4 and AKAP6. ERBB4 is a member of the Tyr protein kinase family and the epidermal growth factor receptor subfamily. In one embodiment, mutations in ERBB4 may be associated with cancer. In one embodiment, AKAPs (A-kinase anchor proteins) bind to the regulatory subunit of protein kinase A (PKA). In one embodiment, the encoded protein is expressed in brain and cardiac and skeletal muscle. In one embodiment, it is specifically localized to the sarcoplasmic reticulum and nuclear membrane. In another embodiment, it is involved in anchoring PKA to the nuclear membrane or sarcoplasmic reticulum. In a further embodiment, the siRNA sequence may be found in pseudogenes derived from mitochondrial tRNA. In some embodiments, the siRNA may have a regulatory role in splicing of the genes in which it resides. In other embodiments, the siRNA may have a

role in removing unspliced mRNAs. Detection of such siRNAs in a tissue sample may be indicative of cancerous processes within the cells.

3. Probes

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

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.

4. Biochip

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 a 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, as 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.

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.

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.

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

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.

5. Diagnosis

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

In situ hybridization of labeled probes to tissue sections 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 nucleic acids 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.

6. miRNA expression analysis

Certain changes in miRNA expression patterns in cancer cells relative to non-cancerous cells, have been reported. Both increases and decreases in miRNA expression have been described in relation to cancer. The total number of miRNAs in the human genome is estimated to range from approximately 800 to several thousand. In view of

this high number of total miRNAs, identification of particular miRNAs linked to particular cancer types is necessary in order to identify miRNAs that could be targeted for cancer therapy, either through inhibition or augmentation of the miRNA.

Accordingly, there exists a need for the identification of miRNAs that can be inhibited for the treatment of cancer. Also needed are inhibitory agents useful for the treatment of cancer. Further, there exists a need for methods of treating cancer by administering to a subject in need thereof a pharmaceutical agent capable of inhibiting a miRNA identified as dysregulated in connection with cancer. As cancer is a disease caused by the uncontrolled proliferation of cells, as well as increased cell survival, desirable traits of pharmaceutical agents for the treatment of cancer include the ability to reduce cell proliferation, and/or induce apoptosis, which will in turn reduce tumor size, reduce tumor number, and/or prevent or slow the metastasis of cancer cells.

In certain embodiments, the methods provided herein are useful for the treatment of cancer. These methods may result in one or more clinically desirable outcomes in a subject having cancer, such as reduction in tumor number and/or size, reduced metastatic progression, prolonged survival time, and/or increased progression-free survival time. Also provided herein are pharmaceutical agents, such as modified oligonucleotides, that may be used for the treatment of cancer.

The present invention also relates to a method of identifying miRNAs that are associated with disease or a pathological condition comprising contacting a biological sample with a probe or biochip of the invention and detecting the amount of hybridization. PCR may be used to amplify nucleic acids in the sample, which may provide higher sensitivity.

The ability to identify miRNAs that are overexpressed or underexpressed in pathological cells compared to a control can provide high-resolution, high-sensitivity datasets which may be used in the areas of diagnostics, therapeutics, drug development, pharmacogenetics, biosensor development, and other related areas. An expression profile generated by the current methods may be a "fingerprint" of the state of the sample with respect to a number of miRNAs. While two states may have any particular miRNA similarly expressed, the evaluation of a number of miRNAs simultaneously allows the generation of a gene expression profile that is characteristic of the state of the cell. That is, normal tissue may be distinguished from diseased tissue. By comparing expression profiles of tissue in known different disease states, information regarding which miRNAs are associated in each of these states may be obtained. Then, diagnosis may be performed

or confirmed to determine whether a tissue sample has the expression profile of normal or disease tissue. This may provide for molecular diagnosis of related conditions.

7. Determinination of Expression Levels

The present invention also relates to a method of determining the expression level of a cancer-associated miRNA comprising contacting a biological sample with a probe or biochip of the invention and measuring the amount of hybridization. The expression level of a cancer-associated miRNA is information in a number of ways. For example, a differential expression of a cancer-associated miRNA compared to a control may be used as a diagnostic that a patient suffers from cancer. Expression levels of a cancer-associated miRNA 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.

A target nucleic acid may be detected 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.

The target nucleic acid may also be detected by immobilizing the nucleic acid to be examined on a solid support such as nylon membranes and hybridizing a labelled probe with the sample. Similarly, the target nucleic may also be detected by immobilizing the labeled probe to the 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.

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.

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.

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.

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.

The present invention also relates to a method of diagnosis comprising detecting a differential expression level of a cancer-associated miRNA in a biological sample. The sample may be derived from a patient. Diagnosis of cancer in a patient allows for prognosis and selection of therapeutic strategy. Further, the developmental stage of cells may be classified by determining temporarily expressed miRNA-molecules.

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.

8. Drug Screening

The present invention also relates to a method of screening therapeutics comprising contacting a pathological cell capable of expressing a disease related miRNA with a candidate therapeutic and evaluating the effect of a drug candidate on the expression profile of the disease associated miRNA. Having identified the differentially expressed miRNAs, a variety of assays may be executed. Test compounds may be screened for the ability to modulate gene expression of the disease associated miRNA. Modulation includes both an increase and a decrease in gene expression.

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 miRNA. 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.

Combinatorial libraries of potential modulators may be screened for the ability to bind to the disease associated miRNA 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.

9. Gene Silencing

The present invention also relates to a method of using the nucleic acids of the invention to reduce expression of a target gene in a cell, tissue or organ. Expression of the target gene may be reduced by expressing a nucleic acid of the invention 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.* (Science 2004; 304:594-596), which is incorporated herein by reference. One of ordinary skill in the art will recognize that the nucleic acids of the present invention may be used to inhibit expression of target genes 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.

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.* (JBC 2004; 275:52361) and Cheng *et al.* (Nucleic Acids Res 2005; 33:1290), which is incorporated herein by reference.

10. Gene Enhancement

The present invention also relates to a method of using the nucleic acids of the invention to increase expression of a target gene in a cell, tissue or organ. Expression of the target gene may be increased by expressing a nucleic acid of the invention 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 of the invention 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.

11. Therapeutic

The present invention also relates to a method of using the nucleic acids of the invention as modulators or targets of disease or disorders associated with developmental dysfunctions, such as cancer. In general, the claimed nucleic acid molecules 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. apoptosis.

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.

12. Compositions

A composition is also provided. The composition may comprise a nucleic acid described herein and optionally a pharmaceutically acceptable carrier. The composition may encompass modified oligonucleotides that are identical, substantially identical, substantially complementary or complementary to any nucleobase sequence version of the miRNAs described herein or a precursor thereof.

In certain embodiments, a nucleobase sequence of a modified oligonucleotide is fully identical or complementary to a miRNA nucleobase sequence listed herein, or a precursor thereof. In certain embodiments, a modified oligonucleotide has a nucleobase sequence having one mismatch with respect to the nucleobase sequence of the mature miRNA, or a precursor thereof. In certain embodiments, a modified oligonucleotide has a nucleobase sequence having two mismatches with respect to the nucleobase sequence of the miRNA, or a precursor thereof. In certain such embodiments, a modified oligonucleotide has a nucleobase sequence having no more than two mismatches with respect to the nucleobase sequence of the mature miRNA, or a precursor thereof. In certain such embodiments, the mismatched nucleobases are contiguous. In certain such embodiments, the mismatched nucleobases are not contiguous.

In certain embodiments, a modified oligonucleotide consists of a number of linked nucleosides that is equal to the length of the mature miRNA.

In certain embodiments, the number of linked nucleosides of a modified oligonucleotide is less than the length of the mature miRNA. In certain such embodiments, the number of linked nucleosides of a modified oligonucleotide is one less than the length of the mature miRNA. In certain such embodiments, a modified oligonucleotide has one less nucleoside at the 5' terminus. In certain such embodiments, a modified oligonucleotide has one less nucleoside at the 3' terminus. In certain such embodiments, a modified oligonucleotide has two fewer nucleosides at the 5' terminus.

In certain such embodiments, a modified oligonucleotide has two fewer nucleosides at the 3' terminus. A modified oligonucleotide having a number of linked nucleosides that is less than the length of the miRNA, wherein each nucleobase of a modified oligonucleotide is complementary to each nucleobase at a corresponding position in a miRNA, is considered to be a modified oligonucleotide having a nucleobase sequence that is fully complementary to a portion of a miRNA sequence.

In certain embodiments, a modified oligonucleotide consists of 15 to 30 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 19 to 24 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 21 to 24 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 15 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 16 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 17 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 18 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 19 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 20 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 21 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 22 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 23 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 24 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 25 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 26 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 27 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 28 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 29 linked nucleosides. In certain embodiments, a modified oligonucleotide consists of 30 linked nucleosides.

Modified oligonucleotides of the present invention may comprise one or more modifications to a nucleobase, sugar, and/or internucleoside linkage. A modified nucleobase, sugar, and/or internucleoside linkage may be selected over an unmodified form because of desirable properties such as, for example, enhanced cellular uptake, enhanced affinity for other oligonucleotides or nucleic acid targets and increased stability in the presence of nucleases.

In certain embodiments, a modified oligonucleotide of the present invention comprises one or more modified nucleosides. In certain such embodiments, a modified nucleoside is a stabilizing nucleoside. An example of a stabilizing nucleoside is a sugar-modified nucleoside.

In certain embodiments, a modified nucleoside is a sugar-modified nucleoside. In certain such embodiments, the sugar-modified nucleosides can further comprise a natural or modified heterocyclic base moiety and/or a natural or modified internucleoside linkage and may include further modifications independent from the sugar modification. In certain embodiments, a sugar modified nucleoside is a 2'-modified nucleoside, wherein the sugar ring is modified at the 2' carbon from natural ribose or 2'-deoxy-ribose. In certain embodiments, 2'-O-methyl group is present in the sugar residue.

The modified oligonucleotides of the present invention can be generated according to any oligonucleotide synthesis method known in the art, including both enzymatic syntheses and solid-phase syntheses. Equipment and reagents for executing solid-phase synthesis are commercially available from, for example, Applied Biosystems. Any other means for such synthesis may also be employed; the actual synthesis of the oligonucleotides is well within the capabilities of one skilled in the art and can be accomplished via established methodologies as detailed in, for example: Sambrook, J. and Russell, D. W. (2001), "Molecular Cloning: A Laboratory Manual"; Ausubel, R. M. *et al.*, eds. (1994, 1989), "Current Protocols in Molecular Biology," Volumes I-III, John Wiley & Sons, Baltimore, Md.; Perbal, B. (1988), "A Practical Guide to Molecular Cloning," John Wiley & Sons, New York; and Gait, M. J., ed. (1984), "Oligonucleotide Synthesis"; utilizing solid-phase chemistry, e.g. cyanoethyl phosphoramidite followed by deprotection, desalting, and purification by, for example, an automated trityl-on method or HPLC. It will be appreciated that an oligonucleotide comprising an RNA molecule can be also generated using an expression vector as is further described hereinbelow.

The compositions may be used for 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*.

Methods for the delivery of nucleic acid molecules are described in Akhtar *et al.* (Trends Cell Bio 1992; 2:139). WO 94/02595 describes general methods for delivery of RNA molecules. These protocols can be utilized for the delivery of virtually any nucleic

acid molecule. Nucleic acid molecules 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. Alternatively, the nucleic acid/vehicle combination is 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, Neuroscience, 76, 1153-1158, 1997). Other approaches include the use of various transport and carrier systems, for example, through the use of conjugates and biodegradable polymers. More detailed descriptions of nucleic acid delivery and administration are provided for example in WO93/23569, WO99/05094, and WO99/04819.

The nucleic acids 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 1992; 205:365-368). The nucleic acids 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 1992;356:152-154), where gold microprojectiles are coated with the DNA, then bombarded into skin cells.

Administration of a pharmaceutical composition of the present invention to a subject having cancer results in one or more clinically desirable outcomes. Such clinically desirable outcomes include reduction of tumor number or reduction of tumor size. Additional clinically desirable outcomes include the extension of overall survival time of the subject, and/or extension of progression-free survival time of the subject. In certain embodiments, administration of a pharmaceutical composition of the invention prevents an increase in tumor size and/or tumor number. In certain embodiments, administration of a pharmaceutical composition of the invention prevents metastatic progression. In certain embodiments, administration of a pharmaceutical composition of the invention slows or stops metastatic progression. In certain embodiments, administration of a pharmaceutical composition of the invention prevents the recurrence of tumors. In certain embodiments, administration of a pharmaceutical composition of the invention prevents recurrence of tumor metastasis.

Administration of a pharmaceutical composition of the present invention to cancer cells may result in desirable phenotypic effects. In certain embodiments, a modified oligonucleotide may stop, slow or reduce the uncontrolled proliferation of

cancer cells. In certain embodiments, a modified oligonucleotide may induce apoptosis in cancer cells. In certain embodiments, a modified oligonucleotide may reduce cancer cell survival.

A miRNA hybridizes to an mRNA to regulate expression of the mRNA and its protein product. Generally, the hybridization of a miRNA to its mRNA target inhibits expression of the mRNA. Thus, the inhibition of a miRNA may result in the increased expression of a miRNA nucleic acid target. In certain embodiments, the inhibition of a miRNA results in the increase of a protein encoded by a miRNA nucleic acid target.

The present invention also relates to a pharmaceutical composition comprising the nucleic acids of the invention and optionally a pharmaceutically acceptable carrier. The compositions may be used for diagnostic or therapeutic applications. The administration of the pharmaceutical composition may be carried out by known methods, wherein a nucleic acid is introduced into a desired target cell *in vitro* or *in vivo*. Commonly used gene transfer techniques include calcium phosphate, DEAE-dextran, electroporation, microinjection, viral methods and cationic liposomes.

Cancer treatments often comprise more than one therapy. As such, in certain embodiments the present invention provides methods for treating cancer comprising administering to a subject in need thereof a compound comprising a modified oligonucleotide complementary to a miRNA, or a precursor thereof, and further comprising administering at least one additional therapy.

In certain embodiments, an additional therapy may also be designed to treat cancer. An additional therapy may be a chemotherapeutic agent. Suitable chemotherapeutic agents include 5-fluorouracil, gemcitabine, doxorubicine, mitomycin c, sorafenib, etoposide, carboplatin, epirubicin, irinotecan and oxaliplatin. An additional suitable chemotherapeutic agent includes a modified oligonucleotide, other than a modified oligonucleotide of the present invention, that is used to treat cancer.

In certain embodiments, an additional therapy may be designed to treat a disease other than cancer.

In certain embodiments, an additional therapy is a treatment that includes interferons, for example, interferon alfa-2b, interferon alfa-2a, and interferon alfacon-1. Less frequent interferon dosing can be achieved using pegylated interferon (interferon attached to a polyethylene glycol moiety which significantly improves its pharmacokinetic profile). Combination therapy with interferon alfa-2b (pegylated and

unpegylated) and ribavarin has also been shown to be efficacious for some patient populations. Other agents currently being developed include RNA replication inhibitors (e.g., ViroPharma's VP50406 series), antisense agents, therapeutic vaccines, protease inhibitors, helicase inhibitors and antibody therapy (monoclonal and polyclonal).

In certain embodiments, an additional therapy may be a pharmaceutical agent that enhances the body's immune system, including low-dose cyclophosphamide, thymostimulin, vitamins and nutritional supplements (e.g., antioxidants, including vitamins A, C, E, beta-carotene, zinc, selenium, glutathione, coenzyme Q-10 and echinacea), and vaccines, e.g., the immunostimulating complex (ISCOM), which comprises a vaccine formulation that combines a multimeric presentation of antigen and an adjuvant.

In certain such embodiments, the additional therapy is selected to treat or ameliorate a side effect of one or more pharmaceutical compositions of the present invention. Such side effects include, without limitation, injection site reactions, liver function test abnormalities, renal function abnormalities, liver toxicity, renal toxicity, central nervous system abnormalities, and myopathies. For example, increased aminotransferase levels in serum may indicate liver toxicity or liver function abnormality. For example, increased bilirubin may indicate liver toxicity or liver function abnormality.

In certain embodiments, one or more pharmaceutical compositions of the present invention and one or more other pharmaceutical agents are administered at the same time. In certain embodiments, one or more pharmaceutical compositions of the present invention and one or more other pharmaceutical agents are administered at different times. In certain embodiments, one or more pharmaceutical compositions of the present invention and one or more other pharmaceutical agents are prepared together in a single formulation. In certain embodiments, one or more pharmaceutical compositions of the present invention and one or more other pharmaceutical agents are prepared separately.

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, but not limited to, oral, buccal, rectal, parenteral,

transmucosal, intestinal, enteral, topical, suppository, through inhalation, intraperitoneal, intradermal, transdermal, intracheal, intrathecal, intraventricular, intranasal, intraocular and iratumoral (e.g., intravenous, intramuscular, intramedullary, and subcutaneous). An additional suitable administration route includes chemoembolization. In certain embodiments, pharmaceutical intrathecals are administered to achieve local rather than systemic exposures. For example, pharmaceutical compositions may be injected directly in the area of desired effect (e.g., into a tumor).

In certain embodiments, a pharmaceutical composition of the present invention is administered in the form of a dosage unit (e.g., tablet, capsule, bolus, etc.). In certain embodiments, such pharmaceutical compositions comprise a modified oligonucleotide in a dose selected from 25 mg, 30 mg, 35 mg, 40 mg, 45 mg, 50 mg, 55 mg, 60 mg, 65 mg, 70 mg, 75 mg, 80 mg, 85 mg, 90 mg, 95 mg, 100 mg, 105 mg, 110 mg, 115 mg, 120 mg, 125 mg, 130 mg, 135 mg, 140 mg, 145 mg, 150 mg, 155 mg, 160 mg, 165 mg, 170 mg, 175 mg, 180 mg, 185 mg, 190 mg, 195 mg, 200 mg, 205 mg, 210 mg, 215 mg, 220 mg, 225 mg, 230 mg, 235 mg, 240 mg, 245 mg, 250 mg, 255 mg, 260 mg, 265 mg, 270 mg, 270 mg, 280 mg, 285 mg, 290 mg, 295 mg, 300 mg, 305 mg, 310 mg, 315 mg, 320 mg, 325 mg, 330 mg, 335 mg, 340 mg, 345 mg, 350 mg, 355 mg, 360 mg, 365 mg, 370 mg, 375 mg, 380 mg, 385 mg, 390 mg, 395 mg, 400 mg, 405 mg, 410 mg, 415 mg, 420 mg, 425 mg, 430 mg, 435 mg, 440 mg, 445 mg, 450 mg, 455 mg, 460 mg, 465 mg, 470 mg, 475 mg, 480 mg, 485 mg, 490 mg, 495 mg, 500 mg, 505 mg, 510 mg, 515 mg, 520 mg, 525 mg, 530 mg, 535 mg, 540 mg, 545 mg, 550 mg, 555 mg, 560 mg, 565 mg, 570 mg, 575 mg, 580 mg, 585 mg, 590 mg, 595 mg, 600 mg, 605 mg, 610 mg, 615 mg, 620 mg, 625 mg, 630 mg, 635 mg, 640 mg, 645 mg, 650 mg, 655 mg, 660 mg, 665 mg, 670 mg, 675 mg, 680 mg, 685 mg, 690 mg, 695 mg, 700 mg, 705 mg, 710 mg, 715 mg, 720 mg, 725 mg, 730 mg, 735 mg, 740 mg, 745 mg, 750 mg, 755 mg, 760 mg, 765 mg, 770 mg, 775 mg, 780 mg, 785 mg, 790 mg, 795 mg, and 800 mg. In certain such embodiments, a pharmaceutical composition of the present invention comprises a dose of modified oligonucleotide selected from 25 mg, 50 mg, 75 mg, 100 mg, 150 mg, 200 mg, 250 mg, 300 mg, 350 mg, 400 mg, 500 mg, 600 mg, 700 mg, and 800mg.

In certain embodiments, a pharmaceutical agent is sterile lyophilized modified oligonucleotide that is reconstituted with a suitable diluent, e.g., sterile water for injection or sterile saline for injection. The reconstituted product is administered as a subcutaneous injection or as an intravenous infusion after dilution into saline. The

lyophilized drug product consists of a modified oligonucleotide which has been prepared in water for injection, or in saline for injection, adjusted to pH 7.0-9.0 with acid or base during preparation, and then lyophilized. The lyophilized modified oligonucleotide may be 25-800 mg of a modified oligonucleotide. It is understood that this encompasses 25, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, and 800 mg of modified lyophilized oligonucleotide. The lyophilized drug product may be packaged in a 2 mL Type I, clear glass vial (ammonium sulfate-treated), stoppered with a bromobutyl rubber closure and sealed with an aluminum FLIP-OFF® overseal.

In certain embodiments, the compositions of the present invention may additionally contain other adjunct components conventionally found in pharmaceutical compositions, at their art-established usage levels. Thus, for example, the compositions may contain additional, compatible, pharmaceutically-active materials such as, for example, antipruritics, astringents, local anesthetics or anti-inflammatory agents, or may contain additional materials useful in physically formulating various dosage forms of the compositions of the present invention, such as dyes, flavoring agents, preservatives, antioxidants, opacifiers, thickening agents and stabilizers. However, such materials, when added, should not unduly interfere with the biological activities of the components of the compositions of the present invention. The formulations can be sterilized and, if desired, mixed with auxiliary agents, e.g., lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, colorings, flavorings and/or aromatic substances and the like which do not deleteriously interact with the oligonucleotide(s) of the formulation.

In certain embodiments, pharmaceutical compositions of the present invention comprise one or more modified oligonucleotides and one or more excipients. In certain such embodiments, excipients are selected from water, salt solutions, alcohol, polyethylene glycols, gelatin, lactose, amylase, magnesium stearate, talc, silicic acid, viscous paraffin, hydroxymethylcellulose and polyvinylpyrrolidone.

In certain embodiments, a pharmaceutical composition of the present invention is prepared using known techniques, including, but not limited to mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or tabletting processes.

In certain embodiments, a pharmaceutical composition of the present invention is a liquid (e.g., a suspension, elixir and/or solution). In certain of such embodiments, a liquid pharmaceutical composition is prepared using ingredients known in the art, including, but not limited to, water, glycols, oils, alcohols, flavoring agents, preservatives, and coloring agents.

In certain embodiments, a pharmaceutical composition of the present invention is a solid (e.g., a powder, tablet, and/or capsule). In certain of such embodiments, a solid pharmaceutical composition comprising one or more oligonucleotides is prepared using ingredients known in the art, including, but not limited to, starches, sugars, diluents, granulating agents, lubricants, binders, and disintegrating agents.

In certain embodiments, a pharmaceutical composition of the present invention is formulated as a depot preparation. Certain such depot preparations are typically longer acting than non-depot preparations. In certain embodiments, such preparations are administered by implantation (for example, subcutaneously or intramuscularly) or by intramuscular injection. In certain embodiments, depot preparations are prepared using suitable polymeric or hydrophobic materials (for example, an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

In certain embodiments, a pharmaceutical composition of the present invention comprises a delivery system. Examples of delivery systems include, but are not limited to, liposomes and emulsions. Certain delivery systems are useful for preparing certain pharmaceutical compositions including those comprising hydrophobic compounds. In certain embodiments, certain organic solvents such as dimethylsulfoxide are used.

In certain embodiments, a pharmaceutical composition of the present invention comprises one or more tissue-specific delivery molecules designed to deliver the one or more pharmaceutical agents of the present invention to specific tissues or cell types. For example, in certain embodiments, pharmaceutical compositions include liposomes coated with a tissue-specific antibody.

In certain embodiments, a pharmaceutical composition of the present invention comprises a co-solvent system. Certain of such co-solvent systems comprise, for example, benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. In certain embodiments, such co-solvent systems are used for hydrophobic compounds. A non-limiting example of such a co-solvent system is the VPD co-solvent system, which is a solution of absolute ethanol comprising 3% w/v

benzyl alcohol, 8% w/v of the nonpolar surfactant Polysorbate 80™ and 65% w/v polyethylene glycol 300. The proportions of such co-solvent systems may be varied considerably without significantly altering their solubility and toxicity characteristics. Furthermore, the identity of co-solvent components may be varied: for example, other surfactants may be used instead of Polysorbate 80™; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g., polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose.

In certain embodiments, a pharmaceutical composition of the present invention comprises a sustained-release system. A non-limiting example of such a sustained-release system is a semi-permeable matrix of solid hydrophobic polymers. In certain embodiments, sustained-release systems may, depending on their chemical nature, release pharmaceutical agents over a period of hours, days, weeks or months.

In certain embodiments, a pharmaceutical composition of the present invention is prepared for oral administration. In certain of such embodiments, a pharmaceutical composition is formulated by combining one or more compounds comprising a modified oligonucleotide with one or more pharmaceutically acceptable carriers. Certain of such carriers enable pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a subject. In certain embodiments, pharmaceutical compositions for oral use are obtained by mixing oligonucleotide and one or more solid excipient. Suitable excipients include, but are not limited to, fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). In certain embodiments, such a mixture is optionally ground and auxiliaries are optionally added. In certain embodiments, pharmaceutical compositions are formed to obtain tablets or dragee cores. In certain embodiments, disintegrating agents (e.g., cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof, such as sodium alginate) are added.

In certain embodiments, dragee cores are provided with coatings. In certain such embodiments, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to tablets or dragee coatings.

In certain embodiments, pharmaceutical compositions for oral administration are push-fit capsules made of gelatin. Certain of such push-fit capsules comprise one or more pharmaceutical agents of the present invention in admixture with one or more filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In certain embodiments, pharmaceutical compositions for oral administration are soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. In certain soft capsules, one or more pharmaceutical agents of the present invention are be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added.

In certain embodiments, pharmaceutical compositions are prepared for buccal administration. Certain of such pharmaceutical compositions are tablets or lozenges formulated in conventional manner.

In certain embodiments, a pharmaceutical composition is prepared for administration by injection (e.g., intravenous, subcutaneous, intramuscular, etc.). In certain of such embodiments, a pharmaceutical composition comprises a carrier and is formulated in aqueous solution, such as water or physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. In certain embodiments, other ingredients are included (e.g., ingredients that aid in solubility or serve as preservatives). In certain embodiments, injectable suspensions are prepared using appropriate liquid carriers, suspending agents and the like. Certain pharmaceutical compositions for injection are presented in unit dosage form, e.g., in ampoules or in multi-dose containers. Certain pharmaceutical compositions for injection are suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Certain solvents suitable for use in pharmaceutical compositions for injection include, but are not limited to, lipophilic solvents and fatty oils, such as sesame oil, synthetic fatty acid esters, such as ethyl oleate or triglycerides, and liposomes. Aqueous injection suspensions may contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, such suspensions may also contain suitable stabilizers or agents that increase the solubility of the pharmaceutical agents to allow for the preparation of highly concentrated solutions.

In certain embodiments, a pharmaceutical composition is prepared for transmucosal administration. In certain of such embodiments penetrants appropriate to

the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

In certain embodiments, a pharmaceutical composition is prepared for administration by inhalation. Certain of such pharmaceutical compositions for inhalation are prepared in the form of an aerosol spray in a pressurized pack or a nebulizer. Certain of such pharmaceutical compositions comprise a propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In certain embodiments using a pressurized aerosol, the dosage unit may be determined with a valve that delivers a metered amount. In certain embodiments, capsules and cartridges for use in an inhaler or insufflator may be formulated. Certain of such formulations comprise a powder mixture of a pharmaceutical agent of the invention and a suitable powder base such as lactose or starch.

In certain embodiments, a pharmaceutical composition is prepared for rectal administration, such as a suppositories or retention enema. Certain of such pharmaceutical compositions comprise known ingredients, such as cocoa butter and/or other glycerides.

In certain embodiments, a pharmaceutical composition is prepared for topical administration. Certain of such pharmaceutical compositions comprise bland moisturizing bases, such as ointments or creams. Exemplary suitable ointment bases include, but are not limited to, petrolatum, petrolatum plus volatile silicones, and lanolin and water in oil emulsions. Exemplary suitable cream bases include, but are not limited to, cold cream and hydrophilic ointment.

In certain embodiments, a pharmaceutical composition of the present invention comprises a modified oligonucleotide in a therapeutically effective amount. In certain embodiments, the therapeutically effective amount is sufficient to prevent, alleviate or ameliorate symptoms of a disease or to prolong the survival of the subject being treated. Determination of a therapeutically effective amount is well within the capability of those skilled in the art.

In certain embodiments, one or more modified oligonucleotides of the present invention is formulated as a prodrug. In certain embodiments, upon *in vivo* administration, a prodrug is chemically converted to the biologically, pharmaceutically or therapeutically more active form of a modified oligonucleotide. In certain embodiments, prodrugs are useful because they are easier to administer than the corresponding active form. For example, in certain instances, a prodrug may be more

bioavailable (e.g., through oral administration) than is the corresponding active form. In certain instances, a prodrug may have improved solubility compared to the corresponding active form. In certain embodiments, prodrugs are less water soluble than the corresponding active form. In certain instances, such prodrugs possess superior transmittal across cell membranes, where water solubility is detrimental to mobility. In certain embodiments, a prodrug is an ester. In certain such embodiments, the ester is metabolically hydrolyzed to carboxylic acid upon administration. In certain instances the carboxylic acid containing compound is the corresponding active form. In certain embodiments, a prodrug comprises a short peptide (polyaminoacid) bound to an acid group. In certain of such embodiments, the peptide is cleaved upon administration to form the corresponding active form.

In certain embodiments, a prodrug is produced by modifying a pharmaceutically active compound such that the active compound will be regenerated upon *in vivo* administration. The prodrug can be designed to alter the metabolic stability or the transport characteristics of a drug, to mask side effects or toxicity, to improve the flavor of a drug or to alter other characteristics or properties of a drug. By virtue of knowledge of pharmacodynamic processes and drug metabolism *in vivo*, those of skill in this art, once a pharmaceutically active compound is known, can design prodrugs of the compound (see, e.g., Nogrady (1985) Medicinal Chemistry A Biochemical Approach, Oxford University Press, New York, pages 388-392).

13. Kits

The present invention also relates to a kit, which may comprise 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.

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.

Having now generally described the invention, the same will be more readily understood through reference to the following examples, which are provided by way of illustration and are not intended to be limiting of the present invention.

EXAMPLES

Example 1: Materials and methods

Deep Sequencing

Samples in which miRNA expression was identified include tumors (colon, bladder, breast, lung, liver, kidney, ovarian, prostate, esophagus, cervix, and pancreas), normal tissues (colon, bladder, breast, lung, liver, kidney, brain, endometrium, lymph nodes, and heart), metastases (breast, lung, kidney, endometrium, salivary gland, larynx, tongue, and melanocytes) and blood samples (blood cells, whole blood). RNAs used for deep sequencing libraries are as follows: bladder pool, colon pool, breast pool and lung pool.

RNA Isolation and Enrichment Phase

Total RNA was extracted twice from each sample of 23 human formalin-fixed paraffin-embedded (FFPE) samples derived from cancerous tissue [breast (n=5), bladder (n=5), colon (n=7), and lung (n=6)] (producing a total of 46 samples). RNA was isolated using ten 10-μm-thick tissue sections using the extraction protocol developed at Rosetta Genomics. Briefly, the sample was incubated repeatedly in xylene at 57°C to remove excess paraffin, followed by washing in ethanol. Proteins were degraded by incubation in proteinase K solution at 45°C for a few hours. The RNA was extracted with acid phenol: chloroform followed by ethanol precipitation and DNAse digestion. Total RNA quantity and quality were checked by spectrophotometry (Nanodrop ND-1000). Pools of samples of the small RNA fraction within the total RNA were labeled and hybridized on arrays. After ensuring the presence and expression of more than 100 miRNAs per cancerous tissue pool, tissues were pooled together, resulting in a bladder+breast tumor pool and a colon+lung pool. Array expression revealed the presence of 157 miRNAs from bladder cancer FFPEs, 260 miRNAs from breast cancer FFPEs, 135 miRNAs from lung cancer FFPEs, and 239 miRNAs from colon cancer FFPEs. Total RNA (75 μg) of seven duplicate different colon cancer FFPEs were pooled together with 75 μg of six duplicate different lung cancer FFPEs, while 75 μg total RNA of five duplicate different bladder cancer FFPEs were pooled together with 75 μg of five duplicate different breast cancer FFPEs.

Cloning Linker Attachment Phase

The 3' and 5' cloning linkers (3' Linker: 5'-rAppCTGTAGGCACCATCAAT/3ddC/-3' (SEQ ID NO: 455); 5' Linker: 5'-TGGAATrUrCrUrCrGrGrGrCrArCrCrArArGrGrU-3' (SEQ ID NO: 456)) were ligated to purified small RNA species in preparation for cDNA synthesis and amplification.

Amplification

Reverse transcription of the linkered RNA species was carried out followed by PCR amplification. Primer sequences were as follows:

RT: 5'-GATTGATGGTGCCTACAG -3' (SEQ ID NO: 457) (Tm: 50.2°C)

Fwd tag 1 primer: (454 fwd1 –BL1 mm)

5' GCCTCCCTCGCGCCATCAGcagtTGTAATTCTCGGTCACCAA 3' (SEQ ID NO: 458)

Rev tag1 primer: (454-Rev1-BL1)

5' GCCTTGCCAGCCCGCTCAGcatgATTGACGGTGCCTACAG 3' (SEQ ID NO: 459)

Fwd tag2 primer: (454 fwd2 –BL1 mm)

5' GCCTCCCTCaCGCCATCAGtagtTGTAATTCTCGGTCACCAA 3' (SEQ ID NO: 460)

Rev tag2 primer: (454-Rev2-BL1)

5' GCCTTGCCAGCCCGCTCAGtagtATTGACGGTGCCTACAG 3' (SEQ ID NO: 461)

Small RNA Enrichment

The libraries were built using the MirCat kit, with several modifications, as described below. The mass of RNAs in the miRNA size range of 18 nt to 26 nt is very small relative to total RNA, so removal of as much competing mass as possible is essential. Therefore, enrichment of small RNA was carried out by recovering the small RNA fraction, identified by internal size markers, from a slice of a 12% denaturing (7 M Urea) polyacrylamide gel.

The synthetic RNA size markers run in the lane adjacent to the cancer samples were:

15 nt (5' GCAAAGCACACGGCC 3') (SEQ ID NO: 462),

22 nt (5' UAUGUAUCGAAUUUAAGCUCAA 3') (SEQ ID NO: 463) and

38 nt (5' GCAAGGAUGACACGCAAAUUCGUGAAGCGUUCCAUAUU 3') (SEQ ID NO: 464).

Cleaning the desired RNA from the gel was carried out by GeBAflex-tube-midi column using an electric current of 300 volt for 40 min until the nucleic acid exited from the gel slice, followed by applying reverse polarity of the current for 120 seconds. This step releases the nucleic acid from the membrane. Isolated RNA was precipitated by adding 8 µl of linear acrylamide, a 1/10 volume of NaOAc 3 M, pH 5.2, and three volumes of cold 100% ETOH, with vortexing after each addition. The isolated RNA was precipitated overnight at -20°C, centrifuged for 1 h at 4°C at 14000 rpm, followed by washing with 1 ml cold 85% ETOH and subsequent centrifugation for 5 min at 14000 rpm.

RNA linkering

Following recovery of the enriched small RNA fraction from the acrylamide gel slice, the small RNAs were ligated with a 3' and a 5' linker in two separate reactions. First, 3' ligation was performed in which the 3' linker was ligated to the small RNAs using T4 RNA ligase in the absence of ATP in order to avoid circularization of the RNA fragments, as described in Lau *et al.*, 2001. The ligated product was purified by recovering the desired band, identified using size markers, from a slice of a 12% denaturing (7 M Urea) polyacrylamide gel. Two synthetic RNAs (24 nt and 38 nt, described previously) and two synthetic RNA transcripts (53 nt and 83 nt) were run adjacent to the cancer samples. Purification and precipitation were carried out as described previously.

The 5' linker is ligated to the 3' linkered small RNAs in the presence of 1.0 mM ATP, followed by recovering the desired band from a slice of a 12% denaturing (7M urea) polyacrylamide gel with the same size markers. Purification and precipitation were done as described before.

Reverse Transcription

The 5' and 3' ligated RNAs contained both RNA and DNA regions which were converted to DNA using reverse transcriptase with RT primer, according to MirCat protocol.

PCR amplification

The PCR amplification step was carried out using primers different from those provided by the MirCat kit since the primers provided cause strong self- and heterodimers. PCR was carried out using *PfuUltra* high fidelity DNA polymerase (Stratagene #600380) and pairs of longer PCR primers (40-42-mers) containing sequences complementary to the linkers, tag sequences and sequences which were suitable for the 454 platform. Tag1-flagged colon and lung library and Tag2-flagged breast and bladder library followed by 454 sequences that will convert the small RNA libraries made to ones that can be directly sequenced on the 454 platform.

Samples from five PCR reactions were pooled, extracted with phenol: chloroform, followed by recovery of the desired band from slices of an 8% native polyacrylamide gel. The resulting library was sent for sequencing on the 454 platform.

Deep sequencing data analysis

The deep sequencing process yielded over 200,000 sequences from both libraries. Adaptors were removed using a Perl script allowing internal polyN sequences within the adaptors and 1 mismatch. About 1000 sequences were removed since they were too short after adaptor removal (<10 bp). The sequences were mapped to the human genome (UCSC hg18 build) using BLAST, allowing maximum three bps mismatched to the genome and maximum insertion/deletion (indels) of three bps. For each aligned sequence the highest scoring hit was retrieved. All sequences with position overlap were clustered together using a Perl script. For example, if sequence X was mapped to positions 1-20 within the plus strand of chromosome 1 and a sequence Y was mapped to positions 15-35 on the same chromosome and strand, then the two sequences were unified in the same genomic cluster of chromosome 1, plus strand, positions 1-35. The clusters of sequences represent segments of expressed genes. Each genomic cluster of sequences was assigned the most abundant sequence in this cluster and demanded that for candidate miRNAs, the most abundant sequence will be mapped precisely to the genome (not allowing any mismatches/indels).

The next step was to annotate known sequences. The following datasets were used for this task: RNA genes, sno/miRNA, RefSeq genes, and RepeatMasker tables were downloaded from the UCSC table browser, and known miRNA precursors were downloaded from miRBase in order to mark whether the sequence is part of a noncoding gene, a snoRNA, a protein-coding gene exon, a genomic repeat, or a known miRNA

precursor, respectively. The sequences of the novel miRNA candidates were extended by several hundred bp within their chromosomes in order to predict possible miRNA precursors. An extended sequence was intended to predict the folding of a pri-miRNA that contains a hairpin-folded pre-miRNA. The candidate pri-miRNAs were folded using the Vienna package {Hofacker, I.L. (2003) *Nucleic Acids Res*, **31**, 3429-3431} or mfold {Zuker, M. (2003) *Nucleic Acids Res*, **31**, 3406-3415} programs. All hairpin structures that had at least six base pairs, were at least 55 nucleotides long and had a loop not longer than 20 nucleotides were extracted from the minimum free energy fold of the predicted pri-miRNA (excluding overlapping hairpins). Each hairpin was assigned a Palgrade and conservation score. Predicted miRNA precursors have either Palgrade>0 (meaning it has structural characteristics of known miRNA) or have absolute value of conservation score>0.9 (conserved in mammals) {Bentwich, *et al.* (2005) *Nat Genet*, **37**, 766-770}. These criteria have a sensitivity of 86% for known miRNA precursors from miRBase 13.0. In addition, only sequences with ten or less genomic copies, with a length of 17-25 bp and a GC content in known miRNA range (15-90%) were chosen as miRNA candidates.

Microarray design

Custom microarrays (Biochips) were manufactured by Agilent Technologies by *in situ* synthesizing DNA oligonucleotide probes to 949 known microRNAs and 876 sequences printed in triplicate, and 8639 computationally predicted microRNAs printed in one copy. 44/49 of the novel miRNA and small RNAs were used in the microarray (five sequences were identified as novel miRNAs/small RNAs after the design of the microarray). Sequences from deep sequencing were characterized by:

1. Mapping to the human genome.
2. Being part of a predicted hairpin (folded by Vienna/Mfold).
3. Not being part of an annotated sequence (known miRNA, small RNA, coding exon).
4. Having less than 10 genomic occurrences.
5. MiRNA-sized (17-25 bp).
6. 10%<%GC content <90%.

Each probe comprised an antisense sequence of the relevant sequence, followed by a tail sequence (GCAATGCTAGCTATTGCTTGCTATTAAAAA) (SEQ ID NO: 465), trimmed so the final length of the probe would be 45 nucleotides. Seventeen

negative control probes were designed using the sense sequences of different microRNAs. Two groups of positive control probes were designed to hybridize to the array: (i) synthetic small RNA that were spiked to the RNA before labeling to verify the labeling efficiency, and (ii) probes for abundant small nuclear RNAs that were spotted on the array to verify RNA quality.

Microarray hybridization

Thirty-eight samples were hybridized to these microarrays. The samples divided to normal (n = 8), tumor (n = 15), tumor adjacent (n = 5) and metastasis indications (n = 8).

A total of 2-2.5 μg of total RNA was labeled by ligation of an RNA-linker, p-rCrU-Cy/dye (Eurogentec S.A.; Cy3 or Cy5), to the 3′ end. Synthetic small RNA was spiked into the RNA before labeling to verify the labeling efficiency. Slides were incubated with the labeled RNA for 12-16 h at 55°C and then washed according to Agilent GE washes for Agilent miRNA protocol. Arrays were scanned using Agilent DNA Microarray Scanner Bundle (Agilent Technologies, Santa Clara, CA) at a resolution of 5 micrometer, dual pass at 100% and 10% PMT power green and red Dye channel. Array images were analyzed using Agilent Feature Extraction software (version 9.5).

Array signal calculation and normalization:

Array images were analyzed using the Feature Extraction software (FE) 9.5.1 (Agilent, Santa Clara, CA). Triplicate spots were combined to produce one signal for each probe by taking the logarithmic mean of reliable spots. All data were log-transformed (natural base) and the analysis was performed in log-space. A reference data vector for normalization R was calculated by taking the median expression level of a subset of all probes (all miRs in mirbase 10) across samples. For each sample data vector S, a 2^{nd} degree polynomial F was found so as to provide the best fit between the sample data and the reference data, such that $R \approx F(S)$. For each probe in the sample (element Si in the vector S), the normalized value (in log-space) Mi was calculated from the initial value Si by transforming it with the polynomial function F, so that $Mi=F(Si)$. P-values were calculated using a two-sided t-test on the log-transformed normalized fluorescence signal. The fold-difference (ratio of the median normalized fluorescence) was calculated for each microRNA.

The signal of a sequence is defined as differential between sample "A" and sample "B" if the fold change between the signal in sample A and sample B is either larger than the 95th percentile of fold changes of all sequences expressed in both samples, or larger than 8.

For every tumor/tissue type, the tumor sample was compared to the median signals of all other tumors, to the normal sample from the same tissue type where available, the relevant tumor adjacent sample where available, and metastatic samples originating in the same tissue where available. Each metastatic sample was also compared to normal samples originating from the same site.

Sequences meeting the criteria of having differential expression (as defined above) in at least one comparison, GC content<75%, and differing from other previously identified miRs, were chosen.

Expression detection by qRT-PCR

RNA was subjected to a polyadenylation reaction. RNA was incubated in the presence of poly (A) polymerase (PAP; NEB M0276), $MnCl_2$, and ATP for 1h at 37^{0}C. Then, using an oligodT primer harboring a consensus sequence, reverse transcription was performed on total RNA using SuperScript II RT (Invitrogen). Next, the cDNA was amplified by real time PCR;

Sequences used in the reaction are microRNA-specific forward primers detailed in table 4 below, a universal TaqMan probe (complementary to the 3' end of the oligodT plus part of the tail, SEQ ID NO: 470), and the universal reverse primer (complementary to the consensus 3' sequence of the oligodT tail, SEQ ID NO: 471). For each miR, expression signals were calculated by the formula 42 – *Ct* (miR-X).

Table 4: microRNA-specific forward primers sequences used in the qRT-PCR reaction

miR name	SEQ ID NO of miR	Forward primer	SEQ ID NO of forward primer
MID-19433	92	GGCTGGTCCGAAGGTAGTGAGTT	466
MID-19434	93	GGCTGGTCCGAGTGCAGT	467
MID-16489	31	GCAGGGTGACAGGGAACAGTAGAT	469

Example 2: Identified miR expression in tumor tissues, normal tissues and blood and metastatic tissues, for candidate miRs

Table 5 presents expression data for miRNAs identified in tumor tissues.

Table 5: log_2 expression per tumor tissue per miR for candidate miRs

SEQ ID NO	Colon Tumor	Bladder Tumor	Breast Tumor	lung Tumor	Liver Tumor	Kidney Tumor	Ovary Tumor	Prostate Tumor	Esophagus Tumor	Cervix Tumor	Pancreas Tumor
1	9.8012	6.06	9.4418	8.9957	5.6439	5.6439	5.7206	9.1141	5.7662	6.6784	7.4804
2	9.973	6.2411	6.2896	9.4598	5.6439	5.6439	5.6439	7.3124	5.881	7.1761	8.3106
4	8.3527	8.5728	9.3289	9.0806	8.1863	8.5888	7.3135	10.7365	8.0832	7.9904	7.9967
5	5.6439	9.3122	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
6	6.5673	5.6439	5.6439	8.9349	5.6439	5.6439	5.6439	6.1018	5.6439	5.6439	7.1879
7	12.0591	10.6437	11.3155	11.7173	10.9781	10.8283	9.2509	11.7531	10.2874	9.3214	10.8017
9	8.2551	6.3662	6.3854	7.7073	5.6439	5.6439	6.463	8.0354	5.8435	6.5753	6.0086
10	13.5372	12.014	11.3714	12.4486	10.9083	8.2677	9.8728	11.2934	11.9017	10.8942	13.7688
11	13.5292	11.7668	10.785	12.4112	10.1851	8.7616	9.0083	11.519	11.5392	10.4039	12.4106
12	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
13	14.5109	11.0286	10.6714	12.857	8.8664	8.5217	9.4297	12.3598	11.6731	10.4197	12.7931
14	11.9284	9.4233	10.2116	9.8346	8.0719	8.0069	8.6565	9.4273	9.8978	8.5517	11.1332
15	11.03	5.6439	5.8213	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	8.7708
16	5.6439	6.2677	6.1297	6.7133	5.6439	5.6439	8.5414	6.1506	5.8053	6.5154	5.6439
17	13.7056	10.7088	11.2178	11.9362	10.8132	9.3766	9.6134	11.235	10.7931	9.7668	13.0122
18	12.1399	9.9496	9.0253	10.1961	8.134	7.5875	8.644	11.4423	9.6427	8.2754	10.3167

SEQ ID NO	Colon Tumor	Bladder Tumor	Breast Tumor	lung Tumor	Liver Tumor	Kidney Tumor	Ovary Tumor	Prostate Tumor	Esophagus Tumor	Cervix Tumor	Pancreas Tumor
19	9.2706	6.6835	8.1412	9.4144	5.6439	5.6439	5.6686	7.7511	5.6439	6.3867	7.8791
20	9.0988	7.4742	8.3048	10.0963	7.6245	6.8606	7.2389	8.6135	8.2137	7.8594	10.4849
21	12.3657	9.5942	8.9931	11.3538	5.6439	7.2224	6.5634	8.8733	10.4945	9.5773	12.103
22	8.4684	7.6565	10.4655	10.6184	8.3398	6.6285	7.2901	10.7301	8.1845	7.8888	9.3513
23	10.274	8.4969	7.9667	9.2515	6.8065	6.4869	6.6604	8.3666	8.4301	7.8481	10.0587
24	10.6601	8.0871	7.1111	8.9807	6.1498	5.6439	7.245	8.4429	6.338	6.7315	8.5532
25	15.4676	13.9095	13.3726	14.992	15.162	13.7366	12.3128	13.9697	14.688	14.3756	15.9049
26	12.3116	10.2533	11.4782	10.9514	8.8804	10.1087	7.8417	11.9683	9.2761	8.8586	10.0667
28	13.1399	7.4313	6.9702	9.853	7.1581	5.6439	6.1316	7.7596	9.6878	7.5135	12.0653
29	12.8677	7.2323	8.068	9.7605	6.2031	5.6439	7.2171	8.8249	9.0889	8.2167	12.1695
30	14.0462	11.1042	11.1619	13.463	10.7862	9.6757	8.4008	11.434	12.3686	9.9949	13.3871
31	14.6095	10.2367	12.2543	13.3379	9.4181	8.5682	9.7681	13.403	12.6714	10.1992	11.3593
32	11.402	8.8905	9.8591	9.3855	7.3081	7.2224	8.1754	9.1766	9.4425	7.6933	9.8353
33	14.499	11.8433	11.6019	13.6666	11.2771	9.9069	9.5559	12.3973	12.5264	11.7815	14.5518
34	12.9267	12.9395	12.9826	13.3536	12.7154	13.199	11.4823	13.939	13.4488	12.3371	13.5398
35	10.7476	11.7003	12.092	12.337	11.1964	11.1009	11.3993	11.9255	10.0087	11.4475	10.7946
37	14.803	15.0756	15.904	15.9746	15.1121	14.4139	13.7085	16.5649	14.5566	15.3914	14.0933
39	13.8797	12.106	12.9358	13.1525	12.1266	12.0256	10.314	14.0547	12.8395	12.3393	11.5817
40	12.2837	9.9963	9.4818	11.3806	9.4373	7.6535	7.9745	10.4079	10.1496	9.2654	11.355

SEQ ID NO	Colon Tumor	Bladder Tumor	Breast Tumor	lung Tumor	Liver Tumor	Kidney Tumor	Ovary Tumor	Prostate Tumor	Esophagus Tumor	Cervix Tumor	Pancreas Tumor
41	11.613	8.1162	8.7609	9.1231	6.6365	6.5237	7.7694	10.1575	7.7985	7.2484	8.9006
42	11.3845	8.4895	8.0279	9.6821	5.9478	5.6439	6.8539	8.344	8.0832	7.6243	9.9514
43	13.7016	10.543	12.8683	12.5291	7.6456	7.7767	9.5845	12.7978	11.8632	8.9231	10.0945
46	11.9763	9.0116	10.3697	11.1929	9.0632	8.1378	8.0483	10.3554	9.5234	8.3211	9.9673
47	5.6439	8.1374	7.057	5.9837	7.656	7.5271	6.5114	7.207	7.2953	7.0529	6.5175
48	10.4932	8.9184	7.9182	9.0443	7.4918	7.969	6.5967	8.3383	8.2799	7.9493	10.0958
49	11.5669	9.1319	10.3839	11.5772	9.2222	8.215	7.639	10.5559	10.0837	8.6938	11.0021
50	10.0823	6.5818	6.757	8.0464	5.6439	5.6439	7.3192	5.6439	5.6439	5.6439	7.7512
51	12.1581	10.0582	10.5507	12.426	10.0439	9.059	8.0653	10.5753	10.6906	10.6432	12.4812
52	10.3667	8.6128	8.3354	8.4042	6.9272	8.0255	8.0941	8.0424	8.8628	8.0498	10.1646
53	11.6614	9.793	9.2288	9.2915	8.056	9.3812	9.1004	8.8851	9.6804	9.0186	11.3213
54	10.6039	8.5245	9.4704	8.7676	6.9105	6.9962	7.7908	8.8084	8.3787	7.0323	9.195
55	8.6611	7.2549	7.2074	7.1249	5.6439	5.6439	6.7352	7.2781	6.5294	6.0838	7.7445
56	12.3162	10.3774	9.2704	12.2291	10.0564	7.5533	8.5796	10.6156	10.5874	9.9533	11.207
57	11.4107	9.2732	8.0388	9.6934	5.9167	6.7102	6.5463	8.1105	9.4512	8.3933	10.7087
58	13.0171	8.1759	9.3988	9.9918	7.2552	6.3084	6.6396	9.3749	7.5149	7.6933	11.3703
59	13.3621	10.328	13.5345	13.8702	11.3826	10.1835	9.4013	13.6832	11.9899	11.4049	11.4101
60	11.0078	10.2848	11.152	12.058	10.4568	9.6112	8.0467	11.9304	10.3445	9.8773	11.0636
61	10.5309	6.2989	7.0936	8.5301	6.3532	5.6439	6.8187	6.854	6.1238	6.8395	8.3554

SEQ ID NO	Colon Tumor	Bladder Tumor	Breast Tumor	lung Tumor	Liver Tumor	Kidney Tumor	Ovary Tumor	Prostate Tumor	Esophagus Tumor	Cervix Tumor	Pancreas Tumor
62	10.274	7.8191	6.4712	7.8187	5.6439	5.6439	5.9766	7.2665	6.9437	6.6784	8.9971
63	12.2211	9.0483	9.3249	11.8824	7.8324	5.8343	7.1853	10.1768	8.9195	9.0042	11.1207
64	11.3668	9.4586	11.9569	12.5922	10.6861	9.1462	8.7299	11.5886	11.5429	10.3313	11.9857
65	10.274	7.4557	7.3112	8.3323	5.6439	5.6439	6.9453	8.0354	7.377	6.4091	8.6995
66	8.6611	8.347	8.0416	9.2334	8.3791	6.3084	7.3889	7.9928	8.1648	7.7225	9.4699
67	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
68	15.0288	9.0582	10.4462	12.1295	8.9305	7.108	7.7255	10.1591	12.464	9.3189	14.3376
69	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
70	11.2703	7.1003	8.2804	10.5035	7.3466	6.5417	7.3392	9.4831	6.5033	8.4522	9.8068
71	9.8012	10.1692	11.2299	10.6562	8.7243	9.9576	7.0446	11.8659	7.6149	8.6694	7.4561
72	14.1792	12.7934	13.8541	14.4743	13.5468	12.4171	10.953	15.1311	13.5281	12.8931	11.8206
74	10.9853	8.0762	7.8743	9.7903	5.6439	5.9207	5.8286	8.49	8.4046	7.7868	10.6044
75	10.8307	8.9365	8.1684	9.9977	8.1416	6.9701	6.9333	9.1862	9.2712	8.4736	10.2398
76	14.2748	12.1316	12.9343	13.8781	13.0934	11.7667	9.9593	13.0832	13.9682	12.9077	14.2219
77	10.8554	8.8485	11.5587	11.3666	8.7397	7.7241	8.5203	11.9097	7.2443	8.6978	9.9369
78	11.499	9.1497	11.6697	11.6051	9.1194	8.8725	8.2778	11.3161	7.1005	7.5654	9.3666
79	8.0446	7.361	7.0815	7.9326	5.6439	5.6439	6.463	8.0904	6.3089	6.9687	7.873
80	12.1735	9.1221	8.3116	9.6613	6.9105	6.1036	8.7603	9.4986	9.6729	8.5717	8.9743
81	8.0993	8.1187	7.6966	9.5957	8.3464	5.9207	7.1392	8.723	7.377	7.814	8.9772

SEQ ID NO	Colon Tumor	Bladder Tumor	Breast Tumor	lung Tumor	Liver Tumor	Kidney Tumor	Ovary Tumor	Prostate Tumor	Esophagus Tumor	Cervix Tumor	Pancreas Tumor
83	8.3763	6.4006	6.4712	8.3614	6.6759	5.6439	6.0705	6.8692	6.6055	7.0731	7.5888
84	9.4145	5.6439	6.4636	7.9656	5.6439	5.6439	6.359	7.7255	7.1559	5.6439	8.1671
85	9.2405	6.0456	6.2502	7.4675	6.1767	5.6439	6.5802	7.2781	7.5295	6.0696	7.7973
86	13.1224	7.7713	8.5343	10.1621	7.3842	5.6439	6.6396	9.4674	8.1747	7.5421	12.1654
87	14.5382	11.7535	13.1129	14.1716	11.7276	9.9007	11.1248	13.3753	12.8445	12.1812	11.4776
88	11.594	8.3802	8.6408	9.3897	8.5091	6.9164	6.913	8.5156	8.8173	7.6679	10.7321
89	9.0492	7.6648	8.3099	7.8689	5.9783	5.8343	7.0699	7.9191	7.9527	6.8475	8.6135
90	11.3623	8.8097	8.3036	9.6074	7.0546	5.8925	6.5577	9.0239	8.4635	7.4343	10.8307
91	8.2551	7.7359	7.0168	8.9805	5.7517	5.6439	5.9466	7.7427	5.9537	6.7824	8.1469
92	13.378	15.2795	14.4614	15.632	14.0261	13.3865	14.0715	13.2315	13.5864	14.353	17.0467
95	10.3502	8.5166	8.679	9.6534	8.1187	7.0839	8.0775	10.1164	8.6056	7.9904	8.682
96	10.2912	8.9741	9.473	10.0353	8.2995	8.3134	7.3221	9.5715	8.5362	7.767	10.0179
97	13.7096	6.3469	6.534	8.5217	6.1767	5.6439	5.8056	7.1948	8.8041	6.7741	12.086
98	12.4594	10.0717	10.3044	10.3562	9.6983	9.6495	8.3396	11.6868	9.5109	9.0042	10.5288
99	9.4825	8.9719	6.8729	8.9305	6.2291	5.6439	6.4442	7.8885	7.6009	7.1823	9.6166
100	9.4556	8.5227	7.3711	9.3891	6.9272	5.6439	6.7962	8.2622	7.7985	7.5039	8.6961
101	10.8493	8.8417	8.2123	8.9087	7.4088	5.9207	7.1016	9.144	8.2706	8.2309	9.8983
102	10.8979	6.9927	7.1732	8.3803	5.7517	5.6439	6.0615	8.8331	7.3284	6.9171	12.0863
104	11.3977	8.7187	7.3693	9.2308	6.4002	5.6439	7.6258	8.5605	7.455	7.1192	9.4012

SEQ ID NO	Colon Tumor	Bladder Tumor	Breast Tumor	lung Tumor	Liver Tumor	Kidney Tumor	Ovary Tumor	Prostate Tumor	Esophagus Tumor	Cervix Tumor	Pancreas Tumor
105	7.7785	5.7822	5.9706	7.2597	7.5261	5.6439	5.6439	5.9725	6.4767	6.2812	8.9364
106	14.0641	10.0281	12.1422	13.519	10.4363	8.8523	8.6811	12.7108	10.5834	10.4103	11.8685
107	11.3078	6.2937	7.2656	8.3253	5.8851	5.6439	6.3173	7.0541	6.2492	6.8789	8.3106
108	11.0134	8.2246	6.7478	7.8869	7.0998	5.6439	6.8143	7.2896	7.2785	7.5421	10.9037
110	13.0666	8.0834	9.0937	11.3631	8.3265	7.2334	7.4725	10.048	8.9379	8.7625	10.6153
111	14.5785	11.842	11.2835	13.3761	10.3993	8.9183	8.6294	12.0409	11.9296	10.7826	14.3811
112	14.421	12.0082	11.8516	12.7973	10.9552	9.7032	9.3557	11.9949	12.2206	10.8306	13.0341
113	13.5974	8.8448	11.1565	12.3849	11.6798	9.2812	8.1091	10.8358	11.6714	10.5961	11.7683
114	11.5233	8.3586	7.5872	9.3025	7.6766	5.9754	7.7694	8.8084	8.5362	8.3105	11.1412
116	14.638	11.4588	12.754	14.4902	11.6766	9.4463	10.3847	13.7783	12.495	11.7851	13.5932
117	12.5976	10.327	10.3498	12.3163	9.7809	8.1879	7.8781	10.9256	10.309	9.5656	11.7317
118	9.6364	8.5769	7.4702	9.4251	7.0998	5.6439	7.2872	7.7082	7.8595	7.9211	10.3596
119	12.4574	9.3428	12.7833	13.0742	10.2998	9.7213	8.7035	13.0928	10.5132	10.7358	10.5713
120	13.5049	10.8281	10.3352	11.6092	9.8966	8.7724	8.4901	10.1752	11.7333	10.1772	13.3711
121	9.5991	11.0338	11.1076	11.2364	11.34	10.4877	11.375	11.0213	9.3242	10.8279	9.8787
122	9.2706	7.8146	7.5665	9.2924	5.6439	5.6439	6.303	8.9506	7.9299	6.9687	8.6428
123	9.1312	7.6959	8.0922	7.9814	6.1498	5.9483	6.9768	7.8573	8.0621	6.9687	8.7305
124	10.9624	9.1027	8.8848	9.0797	7.3339	8.5847	8.3603	7.6093	9.206	8.527	10.7437
125	13.0054	10.5828	10.8547	10.9166	8.8988	10.1719	9.6399	10.4819	11.0937	10.1293	12.1949

SEQ ID NO	Colon Tumor	Bladder Tumor	Breast Tumor	lung Tumor	Liver Tumor	Kidney Tumor	Ovary Tumor	Prostate Tumor	Esophagus Tumor	Cervix Tumor	Pancreas Tumor
126	8.4457	7.3439	7.3756	7.3847	5.6439	5.6439	6.5406	7.432	6.9022	6.6874	8.1671
127	15.4676	17.4618	16.1213	16.2695	16.0009	15.4167	14.5459	17.6877	15.4593	17.338	15.3197
128	9.6609	7.9839	8.1206	8.1747	6.3768	6.002	7.4341	7.9341	8.0408	7.0323	8.8823
129	14.4349	10.1814	11.5669	13.1433	10.694	9.0355	8.813	12.1889	12.1035	11.123	13.4447
130	15.3673	18.0647	15.8358	16.6333	15.8367	15.6885	16.1655	18.4404	15.5735	17.6292	16.0259
131	8.3287	8.9805	7.1223	8.2725	9.4718	7.8555	8.8273	7.3459	9.4814	9.2345	9.3819
132	15.4676	15.1152	14.8484	16.1625	15.3753	14.5624	12.9322	16.5773	15.476	15.5958	15.1852
133	15.4676	17.5968	17.6191	17.0193	16.02	15.5081	15.0995	18.1902	15.4217	17.5952	17.0154
134	11.9314	9.951	8.8774	10.0648	8.6774	7.0594	7.8618	9.0523	10.0949	8.8586	11.0265
135	13.7306	14.3109	13.1136	13.6632	12.6849	13.5273	11.0087	12.6275	13.0433	12.9134	14.2216
136	8.7607	7.8601	7.1783	8.7763	9.0174	7.1199	6.4995	9.3	9.2613	8.5405	8.8915
137	7.8711	7.1515	6.4044	7.2131	5.6439	6.0789	6.6809	6.9996	7.3929	6.8789	8.5142
138	14.905	12.7162	12.0062	14.2904	12.0251	10.0295	10.141	12.9564	13.0495	12.5772	14.918
139	9.1312	7.2797	7.3119	8.997	5.6806	5.6439	5.6439	8.2379	7.377	7.132	9.3732
140	7.7144	7.1399	5.7719	6.3456	6.6365	5.6439	5.6439	5.6439	5.6439	6.4201	5.6439
141	7.9884	8.8416	11.3211	10.8937	8.7699	8.0439	7.7396	11.5153	8.4301	8.4206	9.2219
142	11.2512	8.0158	8.1314	9.716	7.0849	6.4299	6.4995	8.5751	8.8107	8.0204	11.2047
143	10.9393	9.3804	7.7225	9.4481	7.8231	5.8043	7.1981	9.4461	9.0666	8.245	10.9654
144	11.613	9.4095	7.8883	10.0792	6.2799	6.8171	7.1392	8.344	9.6389	8.1343	11.0419

SEQ ID NO	Colon Tumor	Bladder Tumor	Breast Tumor	lung Tumor	Liver Tumor	Kidney Tumor	Ovary Tumor	Prostate Tumor	Esophagus Tumor	Cervix Tumor	Pancreas Tumor
145	12.8318	10.0949	8.9427	10.9465	10.6304	7.7767	7.9958	9.3833	10.0205	9.2153	12.1474
146	11.1725	7.4554	7.1491	9.0598	6.2799	5.6439	5.9365	7.8092	8.0832	7.3928	10.2666
147	12.2162	9.8383	9.0519	11.2225	8.8711	7.1549	7.5552	9.0203	9.8946	8.5473	11.7309
148	10.5383	8.7341	7.9341	9.2518	7.3211	6.1752	6.6656	7.8011	8.85	8.0754	9.7512
149	11.019	8.3964	8.1021	9.6528	6.512	5.6439	6.9768	8.1883	8.3165	7.763	9.6546
150	8.3287	7.9013	6.259	8.9851	5.8851	5.6439	5.6949	8.0494	7.8833	6.3525	7.9034
151	11.1419	7.3325	7.4332	8.2155	6.6563	5.6439	6.0246	7.6551	7.9752	6.5356	10.6144
152	10.9393	9.1536	8.3579	9.4027	7.0698	6.8024	6.7778	8.07	9.5643	7.9734	10.3757
153	12.6481	8.8107	7.3851	9.7057	11.2587	8.062	7.517	9.6822	11.0283	10.1605	9.5179
154	9.9211	9.65	8.0724	10.9188	7.0849	5.6439	7.0482	9.7421	9.1109	7.4844	9.633
155	10.167	8.0411	8.0271	9.2705	5.6439	5.6439	6.6859	7.357	7.5295	6.9974	9.0056
156	8.3527	6.0673	6.0344	7.3641	5.6439	5.6439	5.6439	7.7255	5.6439	5.6439	10.4869
157	9.3725	8.5433	8.4506	8.1711	7.6766	8.3184	8.3723	8.9278	5.6855	7.4137	10.022
159	13.5608	13.9052	13.0655	13.1729	12.8963	13.1091	11.1078	12.5236	12.9793	12.6494	13.1611
160	13.748	11.3159	12.5756	12.962	12.5216	11.8907	10.6159	14.0784	11.7459	11.1326	11.8282
161	15.4676	15.1324	15.005	16.217	15.6551	14.6884	13.219	16.7494	15.5755	15.7807	15.2571
162	11.8176	9.8601	9.613	10.9518	8.7243	7.9234	7.7255	9.4515	9.5763	8.4736	11.3771
163	9.5611	7.2777	6.0601	6.5283	5.6439	5.6439	5.8399	5.8894	6.338	6.5154	9.6776
164	11.2025	8.5733	9.0464	10.6458	6.9437	7.5182	7.6192	11.0836	8.104	8.4132	8.7575

SEQ ID NO	Colon Tumor	Bladder Tumor	Breast Tumor	lung Tumor	Liver Tumor	Kidney Tumor	Ovary Tumor	Prostate Tumor	Esophagus Tumor	Cervix Tumor	Pancreas Tumor
165	11.0245	8.6939	10.1272	10.4516	8.5673	8.7688	6.4995	10.4883	6.4225	7.2188	8.6572
166	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
167	12.7629	10.4779	9.3585	10.924	8.7192	7.91	8.2195	10.1672	10.5756	8.9468	11.792
168	5.6439	9.116	6.1729	6.2064	5.6806	8.3479	5.9365	5.6439	5.6439	7.3928	5.6439
169	10.2827	8.5615	8.0495	9.0878	7.2819	6.8888	7.608	9.2179	7.8951	7.914	9.5257
170	11.9314	10.4001	8.6292	10.019	11.1225	8.7543	7.8193	10.7777	9.1218	9.9933	10.1547
171	14.6241	12.2942	13.0724	14.7834	12.0826	12.1531	10.6362	13.4203	14.0064	13.3808	13.2508
172	10.2302	9.2709	7.7505	8.9361	8.5268	7.883	6.303	8.7405	7.1739	8.5017	7.9214
173	9.4959	7.5339	10.1399	11.0444	8.3791	6.4104	6.7207	9.3861	9.7845	9.05	9.3064
174	11.6206	9.117	8.7372	9.3874	7.2281	6.5237	7.105	8.0143	9.6804	8.1494	11.5946
175	12.2407	9.268	8.9287	10.2097	6.6759	6.1036	7.5784	8.9163	9.6465	8.1404	10.6207
38	12.3457	10.3143	9.0927	10.8194	8.1863	7.3282	8.1311	9.199	9.8322	8.9717	11.4383

Table 6 presents expression data for miRNAs identified in normal tissues.

Table 6: $\log_2$ expression per normal tissue per miR for all candidate miRs

SEQ ID NO	Normal Colon	Normal Bladder	Normal Breast	Normal lung	Normal Liver	Normal Kidney	Normal Brain	Normal Endo-metrium	Normal Lymph Node	Normal Heart
1	9.2819	5.6439	10.2943	8.0659	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439

SEQ ID NO	Normal Colon	Normal Bladder	Normal Breast	Normal lung	Normal Liver	Normal Kidney	Normal Brain	Normal Endo-metrium	Normal Lymph Node	Normal Heart
2	5.6439	5.738	6.7535	6.5466	6.2106	5.6439	5.6439	5.6439	5.9767	5.6439
4	9.0253	8.4887	10.871	8.6643	7.1131	5.6439	5.6439	6.1998	6.8229	5.6802
5	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
6	5.6439	5.6439	6.0822	5.942	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
7	11.9098	10.221	11.8633	10.2643	12.4042	9.4536	9.3553	8.8474	9.019	7.7343
9	7.1213	5.7498	7.9163	7.5465	5.6439	5.6439	5.6439	5.6439	9.1697	7.782
10	12.7601	8.9689	12.5083	12.2771	11.2413	10.0164	8.4355	9.1538	13.1809	11.2778
11	12.9496	9.023	11.9066	10.671	11.0265	7.6078	6.4043	10.4124	10.7344	9.291
12	5.6439	6.0059	5.6439	6.2809	9.3114	6.1151	7.8079	8.1266	6.9143	6.5969
13	13.5593	8.9638	11.5822	11.7657	10.9314	7.8889	7.4608	9.2777	11.7536	9.8308
14	12.2389	7.4451	10.5153	8.9969	12.2504	6.2433	7.4608	8.2943	7.9735	8.5166
15	7.9209	5.8013	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
16	6.0851	5.6439	7.2845	7.3715	9.6839	8.8687	7.623	9.77	6.5227	5.759
17	12.0422	8.6287	11.541	10.3426	12.6994	8.8915	9.6269	8.9628	10.6083	8.5233
18	12.5086	7.9082	9.925	10.0641	9.1287	5.6439	5.8417	7.8178	9.793	8.4146
19	8.0659	5.6439	8.6536	7.5257	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
20	9.171	7.1615	9.1135	8.7592	9.544	9.3036	10.3051	9.2185	9.566	9.0201
21	11.3179	6.7697	10.2301	9.271	5.6439	5.6439	5.6439	7.6155	10.9959	7.5033
22	8.642	5.7288	11.2578	9.1746	8.6068	5.6439	5.6439	5.6439	7.0921	5.6439

SEQ ID NO	Normal Colon	Normal Bladder	Normal Breast	Normal lung	Normal Liver	Normal Kidney	Normal Brain	Normal Endo-metrium	Normal Lymph Node	Normal Heart
23	10.7032	7.4174	9.0157	8.428	7.5136	5.6439	5.6439	9.4282	7.0697	7.124
24	8.1119	5.8722	8.7389	8.019	5.6439	5.6439	5.6439	5.7708	7.4945	6.1371
25	14.4949	14.3293	13.3073	13.1219	14.9701	12.8778	12.7225	11.9442	14.4993	14.1737
26	11.1221	9.9989	12.6484	10.2597	8.1837	5.6439	5.6439	6.0825	7.3999	6.5829
28	9.8293	6.0713	8.307	7.5086	9.5234	5.6439	5.6439	6.3453	7.0697	5.6439
29	10.2068	6.3527	9.3256	8.9997	11.4094	5.6439	5.6439	5.6439	9.6893	7.3751
30	12.5359	10.0213	11.6049	10.4795	13.2526	9.3925	9.6328	9.2033	11.221	10.5505
31	14.1422	8.8174	12.8978	11.9631	11.5287	7.2563	5.7736	8.5019	11.1996	8.9562
32	11.9387	6.7738	9.7328	8.3159	12.0724	5.6439	7.5972	7.8217	7.5354	8.3888
33	13.3028	10.5607	12.6381	11.9703	12.8761	9.8943	9.1086	11.5836	10.9003	11.9815
34	11.8653	15.1678	12.9734	12.8092	15.5463	17.5859	15.6043	15.9338	13.2551	18.6627
35	10.7858	10.3801	12.1031	10.3516	11.5257	8.9752	8.0948	8.7407	9.3649	9.2663
37	14.0863	14.1874	16.2949	14.5567	14.6169	11.3461	10.0663	10.5184	10.6396	11.2564
39	12.6118	11.9072	13.6403	11.5764	13.2962	8.5998	8.157	7.6471	7.5027	7.251
40	11.5264	7.7844	10.3501	9.605	10.9966	8.0073	8.5994	9.358	10.116	10.0053
41	11.7358	6.5999	10.0945	9.238	6.4035	5.6439	5.6439	6.8281	9.4161	6.4022
42	10.3386	6.8011	8.9925	8.0914	9.0757	5.6439	6.295	7.8333	7.6289	7.7878
43	13.1345	7.7263	13.2701	10.9755	11.4373	5.6439	5.6777	7.0033	10.1996	7.8109
46	11.6468	7.4363	11.0092	9.3701	10.4504	7.2441	6.0671	7.5594	8.6333	5.6529

SEQ ID NO	Normal Colon	Normal Bladder	Normal Breast	Normal lung	Normal Liver	Normal Kidney	Normal Brain	Normal Endo-metrium	Normal Lymph Node	Normal Heart
47	5.6439	7.2044	5.927	6.8433	9.5767	10.479	9.2327	7.1896	8.0139	9.7826
48	7.7084	7.7851	8.6	7.8881	8.1214	5.6439	5.6439	6.6834	6.8093	5.6802
49	11.2681	7.8709	10.9829	10.2448	10.7295	7.2441	8.2082	8.8285	9.4673	8.7825
50	10.9038	6.235	5.6439	6.8989	5.6439	5.6439	7.7215	8.3101	8.5631	7.2682
51	11.8717	7.3082	11.3567	9.3938	11.1572	8.2599	8.3513	7.3908	9.3753	7.9419
52	10.3945	6.9077	8.4065	7.3559	11.5276	5.6439	5.6439	5.9532	5.9006	6.057
53	11.5095	7.6906	9.1971	8.141	12.7271	5.6439	5.6439	5.6898	5.6439	5.6439
54	11.3081	6.4013	9.2412	7.9193	11.5359	5.6439	6.8961	7.2883	7.0001	7.8109
55	9.1837	5.7139	7.4327	6.4982	9.3423	5.6439	5.6439	6.1128	5.9264	6.3022
56	11.3342	7.4417	10.8521	9.8548	9.8885	8.0215	6.6899	8.8457	9.6562	8.4292
57	10.4006	8.3247	8.8812	8.9987	6.9736	5.6439	5.6439	9.7357	8.363	7.7098
58	9.8465	6.827	9.5433	8.2565	5.9444	5.6439	5.6439	5.8283	6.1832	5.6439
59	12.4947	10.0546	13.4148	12.502	12.5872	10.9912	11.4537	10.0533	11.7046	11.7652
60	10.6159	9.9869	11.4431	10.2429	11.702	8.7533	10.5922	8.6973	10.2614	9.6102
61	5.6439	6.0264	6.8489	6.4775	5.6439	5.6439	5.6439	6.8109	7.8638	5.9937
62	8.8514	6.3052	7.878	7.2279	5.6439	5.6439	5.6439	7.2703	6.5227	5.6439
63	9.7404	6.3371	10.1419	8.5995	8.4517	5.6805	5.8193	6.3943	8.7205	6.6517
64	11.0994	7.7787	12.5806	10.7654	12.2038	8.23	8.3959	8.2455	9.5891	6.2672
65	8.1568	5.7439	8.4672	7.4652	5.6439	5.6439	5.6439	5.6439	6.0952	5.6439

SEQ ID NO	Normal Colon	Normal Bladder	Normal Breast	Normal lung	Normal Liver	Normal Kidney	Normal Brain	Normal Endo-metrium	Normal Lymph Node	Normal Heart
66	7.9209	6.6163	8.5632	7.6654	8.0701	5.6439	6.8194	7.8711	8.2193	7.8449
67	5.6439	5.6439	5.6439	5.6439	10.433	5.6439	5.6439	5.6439	8.6758	5.6439
68	13.1228	8.1077	10.9619	9.3617	12.4302	5.6439	5.6439	6.5418	9.4673	5.6439
69	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
70	8.2006	6.4545	9.2293	8.3689	9.0495	7.0623	6.3271	7.196	8.2579	6.3364
71	7.7909	9.8177	11.9551	9.7851	8.9411	5.6439	5.6439	5.6439	5.6439	5.6439
72	13.0229	12.7663	14.7515	12.8461	13.9995	8.0562	7.6037	7.6692	6.9269	7.9778
74	8.9974	6.1622	9.5304	8.3102	5.6439	5.6439	5.6439	5.9361	5.6439	5.6439
75	10.5675	7.1326	9.4018	8.4113	10.8155	7.2802	8.6819	8.6522	8.4647	9.42
76	13.4494	11.8882	12.7874	11.8087	14.157	10.422	12.4849	10.5996	11.4863	11.4982
77	10.2136	6.8257	12.4679	10.4075	9.4294	6.5112	5.6439	5.8833	8.9496	5.8813
78	10.9803	7.4382	12.1202	10.716	8.1955	6.4273	6.9382	6.5524	8.2531	6.4651
79	9.329	5.8013	8.9925	7.7758	5.6439	5.6439	5.6439	5.9187	6.0723	5.6439
80	11.2579	7.0743	9.0111	8.3817	7.8514	5.6439	5.6439	7.8061	8.1645	6.73
81	6.6481	6.7784	8.6121	8.0405	9.6064	7.0201	8.1829	8.8576	8.7864	9.7644
83	8.5185	8.8543	7.256	8.908	6.2697	9.734	5.8637	6.9266	8.4233	9.6249
84	8.8207	5.6575	7.4198	7.7886	9.5636	5.6439	5.6439	5.6439	5.6439	5.6439
85	8.8666	5.7498	6.9907	7.3513	8.7963	5.6439	5.6439	5.6439	5.6439	5.6439
86	10.4187	6.3884	9.7072	8.8193	8.0571	5.6439	5.6439	7.1896	7.2396	5.6802

SEQ ID NO	Normal Colon	Normal Bladder	Normal Breast	Normal lung	Normal Liver	Normal Kidney	Normal Brain	Normal Endo-metrium	Normal Lymph Node	Normal Heart
87	13.585	10.0809	13.8577	11.9989	13.6126	9.3257	8.3922	9.3804	11.483	9.9233
88	10.2608	6.8176	9.6093	8.0885	9.261	5.6439	5.7502	6.6544	7.7165	5.6439
89	10.0929	5.9238	8.4803	7.1722	10.1982	5.6439	5.7967	6.0977	5.9264	6.8856
90	11.501	6.8123	9.7299	8.5504	8.4337	5.6439	5.6439	8.2147	7.6513	7.5803
91	7.7909	5.7957	7.693	6.532	5.8288	5.6439	5.6439	5.6439	6.5881	5.6439
92	13.241	15.3544	14.1076	14.5971	14.3357	13.1502	15.8011	14.0108	17.3367	15.7123
95	10.1579	7.267	10.0181	8.4562	8.2041	5.6439	5.6439	5.7708	7.0001	6.5544
96	10.274	7.9219	10.068	8.5523	9.3122	6.088	5.6439	6.32	7.7165	6.5969
97	9.1962	5.7468	7.8086	6.7626	9.5234	5.6439	5.6439	5.6439	5.6439	5.6439
98	10.8045	9.4086	11.9496	9.8975	9.8823	5.6439	6.1046	7.2942	8.6042	7.5803
99	8.2645	6.4007	8.7608	8.4433	7.8396	5.6439	6.7143	7.2213	9.4301	8.0565
100	10.1002	6.0531	8.4605	7.6692	7.079	5.6439	6.3585	6.9345	7.9967	7.3431
101	10.9594	6.8262	9.3256	8.1256	7.1466	5.6439	5.6439	7.4232	7.4777	6.1756
102	10.9678	5.9469	7.2845	7.3809	5.6439	5.6439	6.0286	6.2546	9.3333	5.6439
104	8.9404	6.1717	8.6477	8.0363	5.7071	5.6439	5.6439	6.0825	7.4608	6.3022
105	5.6439	5.6439	7.1974	6.125	9.7961	5.6439	5.6439	5.6439	6.0723	5.6439
106	12.7277	8.6617	13.0016	11.2094	12.2843	7.5598	7.4537	8.5064	9.5821	8.8394
107	5.8028	5.8762	6.7535	6.124	5.9444	5.6439	5.6439	5.8833	7.3079	5.6439
108	8.9259	7.1156	8.0258	8.7483	6.4493	5.6439	5.6439	7.8256	8.493	7.7761

SEQ ID NO	Normal Colon	Normal Bladder	Normal Breast	Normal lung	Normal Liver	Normal Kidney	Normal Brain	Normal Endo-metrium	Normal Lymph Node	Normal Heart
110	9.7765	6.4562	9.8876	8.653	10.0765	6.5514	5.8417	7.1237	8.1178	6.2493
111	14.0882	9.6314	12.583	11.3959	12.5717	8.6721	8.3589	11.0185	11.0571	11.0076
112	13.8189	10.4644	12.4589	11.6363	12.1972	9.1997	7.6669	12.2667	11.3062	10.6001
113	13.0716	8.3185	11.5918	10.3774	12.5553	8.7197	9.8549	8.1388	9.6512	9.079
114	10.54	7.051	8.8761	8.3538	8.0131	5.6439	5.6439	8.2537	7.1359	6.9733
116	13.7931	9.3598	13.7409	12.4056	13.2185	8.7616	9.0956	9.8268	10.8005	10.3431
117	11.5459	8.0889	10.913	9.6505	11.6203	8.254	9.0824	8.3815	10.0677	9.462
118	7.8176	6.246	8.8507	8.6642	6.1597	5.6439	6.5199	7.3176	9.8185	7.7643
119	12.3481	8.3817	12.7089	11.2561	11.0184	9.4183	8.4671	6.7396	9.8288	9.6889
120	12.3654	9.1228	10.6364	10.338	11.45	8.5038	8.4461	9.7332	10.3564	10.3186
121	9.1962	9.8175	10.9713	9.6558	10.9483	8.543	7.8191	7.6909	9.1079	8.985
122	9.2211	6.0744	9.454	8.236	6.1077	5.6439	5.6439	6.067	6.604	6.7919
123	9.8805	5.8484	8.3996	7.1814	10.0184	5.6439	5.6439	6.2943	6.1399	6.4651
124	10.9125	7.1418	9.0918	7.8614	11.8973	5.6439	5.6777	6.7304	6.3443	6.6246
125	12.803	8.7203	11.0527	9.4105	13.6078	6.3151	8.5216	8.0221	7.9197	7.9829
126	9.1062	5.6439	8.0947	6.8507	9.6736	5.6439	5.6439	6.1128	5.7637	6.1371
127	15.0709	14.8495	17.2067	16.3782	14.6943	11.504	10.8752	11.6976	11.4143	12.1558
128	10.0174	5.9838	8.5879	7.3146	10.517	5.6439	5.8853	6.7024	6.5881	6.7426
129	13.3231	9.938	12.586	11.5835	12.8155	9.5169	6.4633	12.5594	11.1336	10.4131

SEQ ID NO	Normal Colon	Normal Bladder	Normal Breast	Normal lung	Normal Liver	Normal Kidney	Normal Brain	Normal Endo-metrium	Normal Lymph Node	Normal Heart
130	15.0284	15.3832	17.7493	16.9838	14.8502	11.7679	12.5253	12.5032	12.8907	13.9005
131	9.0114	8.838	7.3263	7.9724	10.3102	9.5097	9.3535	8.0286	8.5089	9.4115
132	15.0582	13.5545	15.6512	14.4541	14.7414	12.0419	11.4857	11.6524	10.5399	13.0346
133	14.9746	15.3177	17.3572	16.5501	15.0364	12.8728	12.3145	12.7368	13.2446	15.2601
134	11.3502	7.787	9.4018	8.8857	8.8113	6.2186	5.6439	9.3513	8.9944	7.3909
135	11.8888	12.7133	12.2625	11.593	12.6996	9.8979	11.1621	9.7037	10.5662	10.6766
136	8.8666	6.7658	8.3787	7.5165	8.9961	5.6439	5.6439	5.7708	5.6439	5.6439
137	7.9209	5.8223	6.8115	6.2227	8.9826	5.6439	5.6439	5.6439	5.6439	5.6439
138	13.768	11.1028	13.0351	12.6969	13.3827	10.6055	9.6985	12.2921	12.1212	12.3959
139	6.8339	5.7585	8.6768	6.9983	6.3091	5.6439	5.6439	5.6439	6.8759	5.6439
140	8.4816	6.7149	5.6439	7.617	11.5743	8.9431	9.8474	9.2932	8.9468	8.6937
141	8.3262	7.1461	11.7154	10.0769	7.8739	7.0058	6.8418	7.9151	7.7444	7.1145
142	9.8206	6.9697	9.0111	8.382	7.079	5.6439	6.2116	7.9645	7.6807	6.434
143	10.234	6.8962	9.3947	8.9392	7.0775	6.1151	6.2787	7.9294	9.3396	8.1926
144	10.2674	6.79	9.1092	8.4292	8.4155	6.0037	5.6439	5.9012	8.5009	7.7404
145	10.864	7.8001	9.5907	8.6128	9.6323	7.169	7.0382	8.6934	8.9468	7.3672
146	9.4195	6.7389	8.2698	7.8045	6.6239	5.6439	5.6439	7.6471	7.6061	5.707
147	10.723	7.6918	9.8774	8.885	9.9529	6.2433	6.3585	8.4373	9.1818	8.0373
148	9.6561	7.4694	8.7166	8.1669	8.0173	5.6439	6.2116	8.5993	7.812	7.8671

SEQ ID NO	Normal Colon	Normal Bladder	Normal Breast	Normal lung	Normal Liver	Normal Kidney	Normal Brain	Normal Endo-metrium	Normal Lymph Node	Normal Heart
149	9.4523	6.5652	8.9925	8.0348	8.0404	5.6439	5.6439	7.2942	7.8316	7.2768
150	5.6439	5.6439	8.0606	8.178	7.4853	5.6439	5.6439	5.6439	7.3999	5.6439
151	9.3861	5.9433	8.4672	7.6029	6.3091	5.6439	5.6439	6.7396	6.3443	5.6439
152	10.6053	6.9203	8.9879	8.2439	6.9736	5.6439	5.6439	7.897	8.0859	6.057
153	10.234	8.0441	9.1971	7.636	11.8285	5.6439	5.6439	5.6439	5.6439	5.6439
154	9.1582	6.8984	9.9972	9.775	10.238	5.6439	5.6439	6.37	10.882	8.0848
155	10.6159	6.2743	7.6488	7.4611	6.3915	5.6439	5.6439	6.9807	6.8759	6.434
156	5.6439	5.6439	5.682	6.2496	5.6439	5.6439	5.6439	5.6473	8.3183	5.6439
157	7.1213	7.5461	8.5382	7.5914	5.6439	5.6439	5.6439	5.6439	6.4364	5.6439
159	11.7942	12.5428	12.5704	11.5285	12.4844	9.8835	10.8401	9.7546	10.6083	10.2331
160	13.0472	10.4382	13.4156	12.1586	13.6192	9.284	9.6017	9.6444	11.9325	10.1306
161	15.0706	13.6316	15.7441	14.5842	14.6599	12.0507	11.373	11.6857	10.4587	13.1395
162	11.4924	7.6863	10.2828	9.3199	9.3539	5.6439	6.2455	8.4396	8.5894	8.4146
163	7.1591	6.1295	6.5663	7.7005	5.6439	5.6439	5.6439	6.8618	7.2595	6.5399
164	12.0366	6.205	10.202	9.5018	8.2743	8.0766	6.2787	8.7701	5.6439	5.9273
165	10.1365	8.9009	10.623	8.9991	9.4186	5.6439	5.6439	5.6439	5.6439	5.6439
166	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
167	11.5705	7.9437	10.3281	10.1262	9.7017	6.1151	7.1673	9.6128	9.6274	8.788
168	5.6439	6.4937	7.2845	6.1307	7.7882	8.4838	7.2277	5.6439	7.0809	5.6439

SEQ ID NO	Normal Colon	Normal Bladder	Normal Breast	Normal lung	Normal Liver	Normal Kidney	Normal Brain	Normal Endo-metrium	Normal Lymph Node	Normal Heart
169	9.5264	6.5503	9.4675	8.2031	5.6439	5.6439	5.6439	7.5304	6.9143	5.7333
170	5.6439	8.4202	9.8387	8.4147	9.4718	6.7194	5.6439	5.6439	5.6439	7.1048
171	12.2484	11.5711	13.624	11.3476	11.1767	9.4003	9.4426	8.4746	7.6661	10.2628
172	5.6439	7.6796	8.8862	7.452	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
173	8.9974	6.8131	9.8123	8.8543	9.0445	6.4273	6.5744	7.2024	7.2595	6.9411
174	11.0291	7.1942	8.594	8.2356	7.4853	5.6439	5.6439	7.9506	7.8638	5.6802
175	11.1737	7.2686	9.8042	8.991	8.852	6.1934	5.6439	7.9851	9.3354	7.9979
38	10.7181	8.2904	10.5521	10.6211	9.8555	7.0343	7.6669	9.1187	11.8912	9.7878

Table 7 presents expression data for miRNAs identified in metastatic tissue originating from primary tumors at the indicated sites and normal blood tissues.

Table 7: $\log_2$ expression per metastatic tissue and blood per miR for all candidate miRs

SEQ ID NO	Breast Metas.	Lung Metas.	Kidney Metas.	Endometrium Metas.	Salivary Gland Metas.	Larynx Metas.	Tongue Metas.	Melano-cyte Metas.	Blood Cells	Whole Blood
1	5.6439	10.1283	5.6439	5.6439	5.6439	5.6439	5.8988	5.6439	5.6439	5.6439
2	5.6439	7.66	5.6439	6.7748	5.9437	5.9007	5.6439	5.6439	5.6439	5.6439
4	9.2888	9.6066	6.9344	7.088	7.7328	6.5996	7.9694	7.6585	5.6439	5.6439
5	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
6	5.6439	7.3028	5.6439	5.7575	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439

SEQ ID NO	Breast Metas.	Lung Metas.	Kidney Metas.	Endometrium Metas.	Salivary Gland Metas.	Larynx Metas.	Tongue Metas.	Melano-cyte Metas.	Blood Cells	Whole Blood
7	10.5048	11.2897	10.9159	10.6565	9.2833	9.404	9.1443	8.9273	7.0998	6.6593
9	5.6439	8.5483	5.6439	5.6439	6.5768	5.8404	9.9662	5.9413	5.6439	5.6439
10	8.7716	11.0055	10.3475	10.3152	11.2993	10.4848	10.8351	9.7614	5.6439	5.6439
11	10.652	11.679	9.0355	10.296	9.8523	10.3615	11.2195	10.4282	6.0853	5.6439
12	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
13	10.3295	11.6489	9.6773	11.028	9.5246	11.0095	13.0913	11.4803	6.006	6.7833
14	9.6482	8.8229	9.2317	9.0389	8.515	8.0843	8.9977	9.5649	5.6439	5.6439
15	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
16	6.4587	6.0371	5.6439	5.6439	9.208	6.6182	6.6869	7.8574	10.024	5.6439
17	9.7419	11.9163	9.8492	11.4978	8.8419	9.9086	10.8905	8.8444	7.5121	7.1606
18	8.9227	9.4284	7.321	8.8885	8.428	8.3714	11.4564	9.3511	5.6439	5.6439
19	5.6439	10.7286	5.6439	5.6439	6.0449	5.6439	6.4819	5.7338	5.6439	5.6439
20	7.242	9.7048	7.8156	9.2553	7.6746	7.9395	8.842	8.6876	9.6705	9.0221
21	6.2082	10.5863	5.6439	6.6959	7.9921	6.8407	9.4847	8.3997	5.6439	5.6439
22	5.9107	11.1633	7.7777	9.4984	7.3244	8.5523	8.2481	7.43	5.6439	5.6439
23	8.2583	7.9258	5.6439	8.4382	7.6213	7.3601	7.9287	8.4153	5.6439	5.6439
24	5.6439	8.72	5.6439	6.7748	6.5026	6.2435	9.8305	6.989	5.6439	5.6439
25	13.6792	13.7387	15.1259	15.0908	12.6929	15.4024	12.957	12.5781	13.0024	11.2352
26	10.7455	12.2942	8.1404	7.9961	8.7547	7.589	8.4147	8.4127	5.6439	5.6439
28	5.6439	9.1788	5.6439	6.9735	6.3927	7.6915	7.818	6.2478	5.6439	5.6439
29	5.6439	8.9981	6.0014	10.4813	7.062	7.7958	7.9287	6.8527	5.6439	5.7554
30	9.9804	12.2984	11.2632	10.9983	9.8676	11.0754	10.716	9.5762	9.0505	9.6229
31	11.0461	12.8375	9.2217	11.8973	10.3645	11.3516	13.6745	13.3555	7.0998	6.6593
32	8.8553	8.3364	9.1019	8.21	8.0895	7.6082	8.334	9.2904	5.6439	5.6439

SEQ ID NO	Breast Metas.	Lung Metas.	Kidney Metas.	Endometrium Metas.	Salivary Gland Metas.	Larynx Metas.	Tongue Metas.	Melano-cyte Metas.	Blood Cells	Whole Blood
33	11.2073	12.8597	10.3064	10.7318	12.3251	11.4888	12.455	12.0554	9.8612	10.739
34	13.7701	13.3197	13.0532	13.2581	13.3333	12.5936	12.8408	14.7239	13.7661	10.739
35	11.4925	12.4034	10.9766	11.4406	10.5316	10.6839	10.7555	10.6864	5.6439	6.5259
37	16.1928	16.1843	14.6638	15.0441	15.2421	15.2216	15.5842	16.1327	13.6959	14.1187
39	13.4405	13.1073	11.7066	11.609	12.0108	12.5245	12.9526	12.3284	9.0233	7.877
40	9.7314	10.359	8.4341	9.8184	8.9386	9.3984	9.9109	9.2612	7.7012	5.6439
41	6.4874	9.6663	5.6439	7.8286	7.2164	5.6784	11.8019	6.0414	5.6439	5.6439
42	7.7757	8.8273	5.6439	7.0722	7.368	7.5199	9.0855	8.7662	5.6439	5.6439
43	9.3891	13.4306	7.4835	9.8396	8.31	9.243	15.5575	10.7935	5.6439	7.6676
46	8.1134	11.4034	7.8031	9.7048	8.4074	7.8289	9.5159	8.7024	5.6439	5.6439
47	8.4971	6.7025	9.3966	9.5913	8.7891	8.2473	8.2055	9.0307	5.6439	5.6439
48	7.6447	9.5496	5.961	6.2011	6.8788	6.4833	6.9525	7.0484	5.6439	5.6439
49	8.7109	11.9088	8.8365	9.1301	8.9468	9.3647	10.3497	9.8165	6.8243	5.9587
50	5.6439	8.8099	5.6439	8.4047	7.3869	8.1842	7.8308	7.5442	5.6439	5.6439
51	9.8035	10.9968	8.9302	9.6521	9.0182	9.2461	9.7556	8.3865	6.3694	5.8601
52	8.9699	7.7858	8.3658	7.6408	7.0921	6.7077	8.0635	8.3597	5.6439	5.6439
53	10.509	9.0096	9.632	8.7705	7.9658	7.7873	9.3584	9.1605	5.6439	5.6439
54	8.5361	7.9949	8.6808	7.5612	7.6558	6.8565	7.9049	8.796	5.6439	5.6439
55	6.2747	6.3944	6.9561	5.6439	6.4638	5.6439	6.2811	7.2624	5.6439	5.6439
56	8.6831	10.1372	8.7775	9.196	9.0525	9.6545	11.3701	8.7439	5.6439	5.6439
57	6.174	9.1617	5.6439	6.6755	7.7004	7.1554	9.6599	7.2928	5.6439	5.6439
58	5.6439	9.9126	5.6439	6.8315	6.4638	6.1955	7.6403	7.0983	5.6439	5.6439
59	12.0255	14.3695	10.7866	12.7677	12.3199	11.8371	13.9489	11.6643	11.89	9.3323
60	11.1956	11.7803	10.6566	11.3005	10.2563	11.3706	10.7261	10.0746	8.5345	9.6733

SEQ ID NO	Breast Metas.	Lung Metas.	Kidney Metas.	Endometrium Metas.	Salivary Gland Metas.	Larynx Metas.	Tongue Metas.	Melano-cyte Metas.	Blood Cells	Whole Blood
61	5.6439	7.3379	5.6439	6.3589	6.1886	5.6439	5.6439	5.6706	5.6439	5.6439
62	5.6916	8.3542	5.6439	5.8289	5.9735	5.6784	7.6905	6.6518	5.6439	5.6439
63	5.6439	10.1389	6.5647	7.224	7.7851	7.1683	8.7869	6.5381	5.6439	5.6439
64	8.9582	12.9212	10.6213	12.2691	9.6294	11.1658	10.5743	9.7878	9.8855	7.9734
65	5.7829	7.0735	5.6439	6.2011	5.6439	6.29	7.4611	7.5689	5.6439	5.6439
66	7.3609	9.172	7.1396	8.4967	7.2271	7.3713	8.1108	7.0843	5.6439	5.6439
67	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
68	7.0746	11.5863	8.4996	10.0249	7.7986	10.3629	11.1328	8.0712	5.8349	5.6439
69	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
70	5.9511	9.4681	7.2147	8.7598	7.8612	7.8931	6.9033	6.7832	5.6439	5.6439
71	11.4946	11.3779	6.2584	6.0259	8.836	7.4475	7.9346	8.0779	5.6439	5.6439
72	14.8263	14.3433	12.4833	12.2552	13.1776	13.6676	13.3762	13.6849	8.8354	9.4626
74	5.6439	9.7072	5.6439	5.6439	6.8928	5.6439	8.4553	6.4485	5.6439	5.6439
75	8.849	8.6094	6.9561	8.6542	7.9146	8.3428	8.4986	8.7682	5.6439	5.6439
76	13.0982	13.5203	13.4444	14.0998	11.6728	14.3921	12.9118	12.1993	11.6394	13.5344
77	8.0619	12.2502	6.9561	9.7249	8.4643	8.9175	10.5613	8.496	5.6439	5.6439
78	8.1534	12.522	8.0359	9.7696	8.6696	7.1162	9.7434	8.0779	5.6439	5.6439
79	7.1134	8.2437	5.6439	8.1473	6.3927	6.7922	9.8447	8.0645	5.6439	5.6439
80	8.7716	8.1853	5.6439	9.4378	7.6482	8.1907	11.245	8.9057	5.6439	5.6439
81	8.093	8.6872	6.5923	8.1231	7.368	7.8452	7.9463	7.3343	6.7215	5.6439
83	5.6439	7.5432	5.6439	7.1499	7.5859	6.9482	7.9168	7.5689	5.6439	5.6439
84	5.6439	8.7656	6.1892	8.7758	5.6439	8.3136	6.4648	6.104	5.6439	5.6439
85	5.6439	7.9722	5.6439	8.471	5.6439	7.8124	6.878	5.8325	5.6439	5.6439
86	6.6235	10.2718	5.6439	6.716	6.6123	6.8085	7.2929	7.019	5.6439	5.6439

SEQ ID NO	Breast Metas.	Lung Metas.	Kidney Metas.	Endometrium Metas.	Salivary Gland Metas.	Larynx Metas.	Tongue Metas.	Melano-cyte Metas.	Blood Cells	Whole Blood
87	11.7568	14.4741	10.2717	10.7589	11.2245	12.0451	14.3117	11.4553	10.3307	11.1111
88	6.9093	9.6833	6.7232	8.1632	6.8136	6.9629	8.5254	6.6907	5.6439	5.6439
89	6.9523	6.8836	7.0197	6.7358	6.8858	6.0685	7.0458	7.8455	5.6439	5.6439
90	7.6447	9.0804	5.9194	7.7479	8.2596	7.6177	8.0635	8.8763	5.6439	5.6439
91	7.1134	7.1796	5.6439	7.9694	5.9587	6.2669	6.532	6.2064	5.6439	5.6439
92	14.8617	14.3529	15.0529	14.8674	13.3626	14.4205	15.8466	13.9991	16.0761	16.438
95	8.4971	9.3089	6.4191	8.1946	7.5657	7.8694	10.8822	8.3597	5.6439	5.6439
96	8.093	9.9146	7.2864	7.6185	7.3045	7.0485	7.3483	7.564	5.6439	5.6439
97	5.6439	8.6144	5.6439	6.5251	5.6439	6.7757	7.5575	5.6439	5.6439	5.6439
98	10.5602	10.7824	8.6228	8.9179	8.9386	8.5372	9.3502	9.1191	5.6439	5.6439
99	5.7829	7.5126	5.6439	6.9564	7.4009	7.2552	7.9521	7.5689	5.6439	5.6439
100	7.546	8.0173	5.6439	7.7787	6.9997	7.4368	8.0581	8.1173	5.6439	5.6439
101	9.5282	8.6488	5.6439	9.5536	7.211	8.2034	8.6625	9.1919	5.6439	5.6439
102	5.6439	7.7685	5.6439	5.6826	8.0842	7.6825	6.8652	5.9924	5.6439	5.6439
104	6.3384	8.218	5.6439	6.5475	6.4135	6.725	10.0596	6.8943	5.6439	5.6439
105	5.6439	6.868	5.6439	5.6439	5.8335	5.6439	5.6439	5.6439	5.6439	5.6439
106	9.4067	13.3138	9.271	9.9065	9.6546	10.2837	11.2313	9.4479	6.161	5.6439
107	5.6439	7.7944	5.6439	5.7936	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
108	6.1031	7.9646	5.6439	7.8088	6.7211	6.6725	7.1749	7.1862	5.6439	5.6439
110	6.7728	10.2637	7.321	8.6939	7.8054	7.9991	6.9882	6.9266	5.6439	5.6439
111	10.3827	12.2537	9.7569	10.3133	11.7863	10.9722	11.0343	11.6455	8.3612	5.6439
112	11.8155	12.2631	9.8722	12.4937	10.5799	10.8945	11.5991	11.4761	9.6143	5.6439
113	8.7716	12.2782	10.5626	11.8953	9.2591	11.678	10.6752	9.6367	9.0233	5.6439
114	8.6619	8.4123	6.3568	8.6656	7.7505	8.2898	9.6518	8.7024	5.6439	5.6439

SEQ ID NO	Breast Metas.	Lung Metas.	Kidney Metas.	Endometrium Metas.	Salivary Gland Metas.	Larynx Metas.	Tongue Metas.	Melano-cyte Metas.	Blood Cells	Whole Blood
116	11.1865	14.3469	10.3828	11.4721	11.1806	12.3001	14.8843	10.9485	10.4968	11.1692
117	9.1711	11.1539	9.5775	10.0684	9.2387	9.8268	9.7899	8.6919	7.8982	5.6439
118	5.9511	8.3183	6.3247	7.4773	7.9747	7.4996	8.8746	6.8861	5.6439	5.6439
119	10.8746	13.522	9.8949	11.258	9.6779	10.2955	11.5086	10.0213	5.6439	5.6439
120	9.4998	11.3345	9.3045	10.1019	9.718	9.8309	10.6001	9.3824	8.1453	5.6439
121	10.5501	11.3742	10.7118	11.274	9.9818	10.1855	9.8837	9.9522	5.6439	5.6439
122	6.0664	9.1957	5.6439	6.9217	6.7987	6.4215	10.6417	8.0944	5.6439	5.6439
123	7.3111	6.754	6.8206	6.6128	6.7759	5.7777	6.9882	7.6676	5.6439	5.6439
124	9.5242	8.172	8.723	8.1393	7.6444	7.2186	8.4553	8.75	5.6439	5.6439
125	11.2779	9.8365	10.5626	10.0156	9.0968	9.0686	10.0218	10.4673	5.6439	5.6439
126	6.5432	6.1406	6.7482	5.9947	6.493	5.8404	6.5644	7.3285	5.6439	5.6439
127	16.6337	16.9773	15.788	16.3557	15.6967	16.0567	17.3276	16.614	14.6088	14.6372
128	7.8612	7.0457	7.544	6.9041	6.9607	6.1955	7.2834	8.1043	5.6439	5.6439
129	10.4208	13.0187	9.3786	10.294	10.1888	10.7124	11.9381	10.2952	5.8349	6.7224
130	17.7273	17.4311	16.3047	16.6584	18.308	16.7695	18.0674	18.0651	14.7186	13.563
131	8.83	7.2051	7.4835	8.8536	8.0706	8.2595	8.6762	8.3185	8.82	5.6439
132	16.1928	15.5401	14.9035	15.444	14.787	15.4116	15.6516	15.6333	12.3032	13.0789
133	17.5848	16.7444	15.7908	16.4014	16.3208	16.4331	18.0038	17.639	14.187	13.1774
134	8.9406	10.0235	7.251	8.7864	8.0975	8.4378	9.728	8.7722	5.6439	5.6439
135	13.254	13.3872	12.7759	12.1665	11.9948	11.7144	12.4026	11.7258	9.2167	8.8932
136	8.6761	8.2627	7.0607	8.9889	7.4009	8.5321	8.4672	8.0678	5.6439	5.6439
137	7.0746	6.0636	6.2584	5.8289	5.9286	5.6439	6.5155	7.1796	5.6439	5.6439
138	11.9931	13.3137	10.8829	11.3699	12.415	12.4835	13.7178	12.9277	10.6415	11.7167
139	5.6439	8.743	5.6439	5.6439	6.0586	5.6439	6.4988	5.6439	5.6439	5.6439

SEQ ID NO	Breast Metas.	Lung Metas.	Kidney Metas.	Endometrium Metas.	Salivary Gland Metas.	Larynx Metas.	Tongue Metas.	Melano-cyte Metas.	Blood Cells	Whole Blood
140	5.6916	5.6439	5.6439	7.4017	7.2482	7.1029	6.3006	6.7003	8.3612	5.6439
141	8.8801	11.5424	8.0032	8.6426	9.2387	7.9918	9.7987	8.746	5.6439	5.6439
142	6.6749	9.6953	6.4493	7.465	7.2586	7.0205	7.6256	6.9737	5.6439	5.6439
143	6.174	9.2091	6.1892	7.3491	7.8897	7.3713	9.0319	7.8807	5.6439	5.6439
144	6.3384	9.2711	7.4523	7.9873	7.6406	7.0205	9.622	6.288	5.6439	5.6439
145	9.5242	10.0574	7.9239	10.0796	7.9115	8.7702	9.4623	8.3758	5.6439	5.6439
146	5.6439	8.2245	5.6439	6.456	6.705	6.2669	7.6403	6.4133	5.6439	5.6439
147	7.546	10.424	7.388	8.6368	8.3301	7.5794	8.7931	8.1173	5.6439	5.6439
148	7.0146	9.0948	5.6439	7.4145	7.062	7.0895	8.0256	7.2988	5.6439	5.6439
149	7.8965	8.6872	5.6439	7.465	7.2586	7.6825	8.4105	8.9626	5.6439	5.6439
150	5.6439	7.0032	5.6439	5.6439	6.6123	5.6439	7.2547	5.6439	5.6439	5.6439
151	5.7829	8.6243	5.6439	6.4082	5.6439	5.6439	7.0571	6.2342	5.6439	5.6439
152	7.1882	9.5121	5.9194	6.7939	6.934	6.9333	8.8979	7.0041	5.6439	5.6439
153	10.3033	9.3729	8.9241	9.5214	8.3843	11.2932	10.2834	9.3511	5.6439	5.6439
154	6.3384	9.3579	8.1404	6.634	8.562	7.6177	9.7972	7.5689	5.6439	5.6439
155	6.5705	7.1407	5.6439	7.0722	6.737	6.759	8.1958	8.006	5.6439	5.6439
156	5.6439	5.6439	5.6439	5.6439	7.6406	6.1709	5.6439	5.6439	5.6439	5.6439
157	7.0746	9.5336	7.2147	6.2011	6.5026	5.6439	5.8988	6.6318	5.6439	5.6439
159	13.0335	13.1307	12.5629	12.3452	12.3214	11.8911	12.2676	12.3372	9.7944	9.1259
160	11.8629	13.2426	12.3496	11.0102	11.6345	11.7102	10.6924	11.9329	9.9792	5.6439
161	16.3987	15.777	15.1063	15.7814	15.0787	15.6631	15.9364	16.064	12.1979	13.0341
162	7.7375	11.0684	8.3125	8.749	7.9053	7.458	9.9015	7.8416	5.6439	5.6439
163	5.6439	6.5561	5.6439	6.2555	5.6439	5.6439	6.1142	5.9924	5.6439	5.6439
164	9.2064	9.5809	5.6439	7.5138	9.8887	8.6491	8.6487	8.9433	8.9535	10.2955

SEQ ID NO	Breast Metas.	Lung Metas.	Kidney Metas.	Endometrium Metas.	Salivary Gland Metas.	Larynx Metas.	Tongue Metas.	Melano-cyte Metas.	Blood Cells	Whole Blood
165	9.326	11.2983	7.1007	6.0259	6.9272	5.8094	5.8461	6.4829	5.6439	5.6439
166	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439	5.6439
167	8.4971	10.4153	8.3745	8.7917	8.9345	8.635	10.8975	9.4748	5.6439	8.9496
168	5.6439	7.4386	6.3568	5.8635	6.0168	5.6439	6.8652	8.8058	5.6439	5.6439
169	7.3772	8.874	5.6439	7.9331	7.0312	6.759	8.5518	7.7296	5.6439	5.6439
170	12.4699	9.6784	7.1588	8.9703	8.0037	9.3446	9.4847	8.1461	5.6439	5.6439
171	14.4685	11.8911	12.3521	13.8545	11.2449	12.7631	13.7948	12.2547	5.7422	12.6118
172	10.9644	8.874	5.6439	5.6439	6.5858	5.8404	6.7443	6.5806	5.6439	5.6439
173	6.8425	11.2826	7.8031	8.7436	7.3195	8.731	8.1159	7.4461	5.6439	5.6439
174	7.4562	9.4847	6.7727	7.9784	7.0802	7.3713	8.8863	7.7122	5.6439	5.6439
175	9.2793	9.0096	6.646	9.2476	8.1368	8.2286	9.3685	9.2612	5.6439	5.6439
38	7.8006	9.8981	8.3832	8.797	9.403	8.7616	10.4606	8.1525	5.6439	5.6439

Example 3: Differential expression of miRNAs

Example 3.1: Differential expression of miRNAs in tumor vs. normal tissue

Table 8 presents expression for of a list of miRNAs identified in tumor and normal tissues, measured by $\log_2$(signal). "Expression in the tumor" refers to median of expression in all tumor tissues tested. "Normal" refers to median of expression in all normal tissues tested. "P-value" refers to the significance of the change between all normal and all tumor tissues tested according to two-sided Student's t-test. "Fold change" refers to the fold change between all normal and all tumor tissues tested.

Table 8: Differential expression of miRNAs in median of all tumors vs. median of all normal tissues

SEQ ID NO	Expression in Tumor	Expression in Normal Tissue	P-value	Fold change
4	8.7022	6.3478	0.000166	4.2505
7	11.4268	8.838	0.020517	8.1591
11	11.4961	8.6861	0.001198	8.1199
16	5.8969	8.5107	0.018056	7.7914
17	11.4755	9.1745	0.005703	4.5265
21	9.3175	6.7714	0.015275	7.3171
22	8.9666	5.8923	0.004663	6.5683
23	8.3652	6.7028	0.004859	3.0459
26	10.6786	6.5978	4.26E-05	11.5525
30	12.0296	9.9232	0.01282	3.9484
31	11.5973	9.612	0.043977	5.1373
33	12.4108	10.8263	0.013722	3.1756
34	12.8498	15.5002	0.033077	8.6193
35	11.7237	9.1606	0.008423	5.9611
37	15.0822	11.2096	0.000672	14.2391
39	12.5627	8.0425	0.000948	29.5597
46	10.187	7.3096	0.005953	7.9478
47	6.9198	8.687	0.041366	2.6349
48	8.6676	6.5392	0.004988	3.7932
49	10.2674	8.6856	0.02351	2.3743
51	11.0349	8.3781	0.00323	6.2472
52	8.5787	6.2161	0.033347	5.7689
53	9.6784	5.9351	0.00307	14.497
56	10.5727	8.6774	0.047774	4.1742
58	9.0081	5.6439	0.017772	8.4795
60	10.9154	9.2432	0.040817	4.015
63	9.5498	6.9751	0.019608	5.0647

SEQ ID NO	Expression in Tumor	Expression in Normal Tissue	P-value	Fold change
64	11.0904	8.3297	0.013667	7.947
66	8.3536	7.1524	0.015706	2.1565
68	10.6491	6.7095	0.007627	10.7174
71	10.0558	5.6439	0.000117	18.8378
72	13.5426	7.8222	2.20E-05	58.3977
74	8.2698	6.0318	0.027674	5.6204
76	13.103	10.8669	0.012129	6.5822
77	10.0754	7.3403	0.000563	9.5317
78	10.4686	6.9634	0.000493	14.6391
83	7.0229	9.0922	0.038293	7.4343
86	8.9989	6.3686	0.023022	4.5969
87	12.8099	10.0322	0.008374	7.5894
88	9.0949	6.5108	0.002512	4.7874
90	8.8082	6.6211	0.008106	5.6219
96	9.4723	6.9593	0.000913	5.8761
98	10.5581	7.0803	7.30E-05	7.975
100	8.1262	6.5926	0.034875	2.8866
101	8.5028	6.5674	0.011831	3.7603
106	11.7466	8.4098	0.00467	8.6575
110	9.693	7.0724	0.013016	3.4138
111	12.0577	10.0163	0.026345	4.3573
112	12.242	10.246	0.005072	5.3172
113	11.258	8.4543	0.011849	7.0497
114	8.4259	6.7998	0.047278	2.6997
116	12.6904	10.724	0.032184	3.9115
117	10.9194	8.9669	0.033883	3.8266
119	11.5162	8.3563	0.01293	10.4879
120	11.028	9.774	0.037388	2.2551
121	10.9221	8.8869	0.010086	5.7758
127	15.756	12.0285	0.000419	14.6845
129	11.7289	10.2146	0.030579	2.8449
130	16.0554	13.5661	0.013987	4.5166
132	15.1844	12.6411	0.000347	6.5607
133	16.035	14.4383	0.019141	2.1796
134	9.5646	7.311	0.0061	3.8267
135	13.5494	10.8019	0.00242	7.7777
136	7.9898	6.0667	0.018432	5.6063
138	12.9455	11.8627	0.041937	2.7185
141	9.5973	7.222	0.015545	3.4415
144	8.7651	6.8397	0.046269	3.3787
145	10.419	7.7834	0.006281	7.9693

SEQ ID NO	Expression in Tumor	Expression in Normal Tissue	P-value	Fold change
147	10.0094	7.9474	0.030208	4.0375
148	8.4409	7.0259	0.025794	2.2417
152	8.7991	6.4613	0.003923	5.5344
153	9.6272	5.795	0.004838	14.3832
157	8.2504	5.8636	0.000766	6.0434
159	13.2526	10.1683	0.000503	10.0062
160	12.4788	9.619	0.000322	8.4277
161	15.2408	12.5652	0.000219	7.1339
162	10.0643	7.3606	0.000144	6.1185
165	9.7086	5.6439	0.000414	16.2081
167	10.2711	8.3019	0.024997	3.9295
169	8.4429	6.3771	0.005517	3.5679
170	10.3442	6.1659	0.000688	25.3474
171	13.3656	8.7261	0.00051	21.6952
172	8.8214	5.6439	0.000785	9.7292
173	9.0348	6.8835	0.022078	3.5269
174	8.8238	6.7106	0.006262	4.0259

Example 3.2: Differential expression of miRNAs in tumor vs. adjacent and normal tissue

Table 9 presents data comparing expression of a list of miRNAs identified in tumor vs. adjacent and normal tissues, measured by $\log_2$(signal). The term "adjacent" refers to an area of the same tissue and the same patient without tumor growth. "Expression in normal tissue" refers to expression in non-tumor tissue of the same origin.

SEQ ID NOS: 7, 26, 34, 35, 37, 39, 53, 71, 72, 125, 127, 130, 132, 135, 159, 161, 165, 170 and 172 exhibit relatively high expression, and SEQ ID NOS: 34 and 38 exhibit relatively low expression, when comparing expression in a bladder tumor vs. normal bladder tissue.

SEQ ID NOS: 79, 99, 122, 130, 153, 154 exhibit relatively low expression, when comparing expression in a breast tumor vs. adjacent breast tissue.

SEQ ID NOS: 2, 15, 28, 29, 48, 58, 61, 70, 86, 97, 107, 110, 150, 156, 170 and 172 exhibit relatively high expression when comparing expression in a colon tumor vs. adjacent colon tissue.

SEQ ID NOS: 22, 25, 29, 31, 37, 39, 53, 64, 68, 72, 76, 77, 78, 84, 113, 119, 121, 127, 130, 132, 133, 136, 153, 161, 170 and 171 exhibit relatively high expression, and SEQ ID NOS: 16 and 57 exhibit relatively low expression, when comparing expression in an endometrium met.(metastasis) vs. normal endometrium tissue.

SEQ ID NOS: 26, 37, 39, 53, 71, 72, 98, 125, 127, 130, 135, 159 and 165, exhibit relatively high expression, and SEQ ID NOS: 16, 34, 83 and 140 exhibit relatively low expression, when comparing expression in kidney tumor vs. normal kidney tissue.

SEQ ID NOS: 11, 26, 71, 72, 77, 98, 134, 136, 153, 160, 170 and 171, exhibit relatively high expression, and SEQ ID NOS: 12, 14, 32, 34, 54, 89, 123, 126, 128 and 140, exhibit relatively low expression, when comparing expression in liver tumor vs. normal liver tissue. SEQ ID NOS: 14, 28-32, 42, 43, 52-55, 67, 75, 84, 85, 89, 97,105, 123-126, 128, 129, 137, 144 and 154 exhibit relatively low expression when comparing expression in liver tumor vs. adjacent liver tissue.

SEQ ID NOS: 1, 7, 19, 21, 22, 26, 30, 35, 37, 39, 43, 46, 53, 58, 59, 63, 64, 68, 71, 72, 74, 78, 86, 87, 106, 110, 113, 116, 119, 125, 127, 132, 135, 141, 153, 159, 161, 164, 165 and 170-173 exhibit relatively high expression when comparing expression in lung tumor vs. normal lung tissue. SEQ ID NOS: 30, 51, 63, 171 exhibit relatively high expression when comparing expression in lung tumor vs. adjacent lung tissue.

Table 9: Differential expression of miRNAs in tumor vs. adjacent and normal tissue from same origin

Tumor name	SEQ ID NO.	expression in tumor (normalized with adjacent tissue)	expression in edjacent tissue	fold change tumor vs. adjacent tissue	expression in tumor (normalized with normal tissue from same origin)	expression in normal tissue	fold change tumor vs. normal tissue
bladder tumor	7				10.5402	7.3842	8.9136
bladder tumor	26				10.2533	6.6105	12.4907
bladder tumor	34				12.5694	16.0552	11.2033
bladder tumor	35				11.6526	8.5998	8.298
bladder tumor	37				14.9269	10.2265	25.999
bladder tumor	39				11.9746	7.208	27.2218
bladder tumor	53				9.5284	5.8094	13.1682
bladder tumor	71				10.0714	5.6439	21.5186
bladder tumor	72				12.7717	7.0575	52.499
bladder tumor	83				6.2691	9.7742	11.3537
bladder tumor	125				10.1682	6.9132	9.5467
bladder tumor	127				16.9051	10.5801	80.171
bladder tumor	130				17.3101	13.2392	16.8058
bladder tumor	132				15.09	12.0516	8.2157
bladder tumor	135				14.3024	10.7953	11.3696
bladder tumor	159				13.9031	9.9468	15.5227
bladder tumor	161				15.0636	11.9567	8.6151
bladder tumor	165				8.6862	5.6439	8.2383
bladder tumor	170				10.159	6.7522	10.6065
bladder tumor	172				9.2336	5.6439	12.04
breast tumor	79	6.7155	8.9925	4.8468			
breast tumor	99	6.7248	8.7608	4.1012			

Tumor name	SEQ ID NO.	expression in tumor (normalized with adjacent tissue)	expression in edjacent tissue	fold change tumor vs. adjacent tissue	expression in tumor (normalized with normal tissue from same origin)	expression in normal tissue	fold change tumor vs. normal tissue
breast tumor	122	7.5453	9.454	3.7545			
breast tumor	130	15.7314	17.7493	4.05			
breast tumor	153	7.3036	9.1971	3.7156			
breast tumor	154	8.0717	9.9972	3.7985			
colon tumor	2	9.973	5.6439	20.1006			
colon tumor	15	11.03	7.9209	8.6289			
colon tumor	28	13.1399	9.8293	9.9219			
colon tumor	29	12.8677	10.2068	6.3246			
colon tumor	48	10.4932	7.7084	6.8913			
colon tumor	58	13.0171	9.8465	9.0041			
colon tumor	61	10.5309	5.6439	29.5898			
colon tumor	70	11.2703	8.2006	8.3959			
colon tumor	86	13.1224	10.4187	6.5145			
colon tumor	97	13.7096	9.1962	22.8376			
colon tumor	107	11.3078	5.8028	45.4114			
colon tumor	110	13.0666	9.7765	9.7818			
colon tumor	150	8.3287	5.6439	6.4301			
colon tumor	156	8.3527	5.6439	6.5378			
colon tumor	170	11.9314	5.6439	78.1183			
colon tumor	172	10.2302	5.6439	24.0236			
endometrium met.	16				5.6439	9.77	17.4624
endometrium met.	22				9.4984	5.6439	14.466
endometrium met.	25				15.0908	11.9442	8.8556

Tumor name	SEQ ID NO.	expression in tumor (normalized with adjacent tissue)	expression in edjacent tissue	fold change tumor vs. adjacent tissue	expression in tumor (normalized with normal tissue from same origin)	expression in normal tissue	fold change tumor vs. normal tissue
endometrium met.	29				10.4813	5.6439	28.5897
endometrium met.	31				11.8973	8.5019	10.522
endometrium met.	37				15.0441	10.5184	23.0337
endometrium met.	39				11.609	7.6471	15.5823
endometrium met.	53				8.7705	5.6898	8.4604
endometrium met.	57				6.6755	9.7357	8.3407
endometrium met.	64				12.2691	8.2455	16.2646
endometrium met.	68				10.0249	6.5418	11.1824
endometrium met.	72				12.2552	7.6692	24.017
endometrium met.	76				14.0998	10.5996	11.3151
endometrium met.	77				9.7249	5.8833	14.3356
endometrium met.	78				9.7696	6.5524	9.2998
endometrium met.	84				8.7758	5.6439	8.7664
endometrium met.	113				11.8953	8.1388	13.5158
endometrium met.	119				11.258	6.7396	22.9174
endometrium met.	121				11.274	7.6909	11.9837
endometrium met.	127				16.3557	11.6976	25.2479
endometrium met.	130				16.6584	12.5032	17.8175
endometrium met.	132				15.444	11.6524	13.8487
endometrium met.	133				16.4014	12.7368	12.6808
endometrium met.	136				8.9889	5.7708	9.3057
endometrium met.	153				9.5214	5.6439	14.6978
endometrium met.	161				15.7814	11.6857	17.0977

Tumor name	SEQ ID NO.	expression in tumor (normalized with adjacent tissue)	expression in edjacent tissue	fold change tumor vs. adjacent tissue	expression in tumor (normalized with normal tissue from same origin)	expression in normal tissue	fold change tumor vs. normal tissue
endometrium met.	170				8.9703	5.6439	10.0311
endometrium met.	171				13.8545	8.4746	41.6408
kidney tumor	16				5.6439	8.8687	9.3495
kidney tumor	26				10.1087	5.6439	22.0822
kidney tumor	34				13.199	17.5859	20.9212
kidney tumor	37				14.4139	11.3461	8.3853
kidney tumor	39				12.0256	8.5998	10.7467
kidney tumor	53				9.3812	5.6439	13.3372
kidney tumor	71				9.9576	5.6439	19.8869
kidney tumor	72				12.4171	8.0562	20.5475
kidney tumor	83				5.6439	9.734	17.0315
kidney tumor	98				9.6495	5.6439	16.063
kidney tumor	125				10.1719	6.3151	14.4885
kidney tumor	127				15.4167	11.504	15.061
kidney tumor	130				15.6885	11.7679	15.1439
kidney tumor	135				13.5273	9.8979	12.3756
kidney tumor	140				5.6439	8.9431	9.8437
kidney tumor	159				13.1091	9.8835	9.3539
kidney tumor	165				8.7688	5.6439	8.7236
liver tumor	11				10.1851	6.9693	9.2912
liver tumor	12				5.6439	10.2427	24.2328
liver tumor	14	8.0719	12.823	26.928	8.0719	11.2871	9.287
liver tumor	16				5.6439	10.6394	31.901

Tumor name	SEQ ID NO.	expression in tumor (normalized with adjacent tissue)	expression in edjacent tissue	fold change tumor vs. adjacent tissue	expression in tumor (normalized with normal tissue from same origin)	expression in normal tissue	fold change tumor vs. normal tissue
liver tumor	26				8.8804	5.6439	9.4253
liver tumor	28	7.1581	10.4735	9.9548			
liver tumor	29	6.2031	12.3961	73.1598			
liver tumor	30	10.7862	14.1898	10.5823			
liver tumor	31	9.4181	12.444	8.1452			
liver tumor	32	7.3081	12.6351	40.1398	7.3081	11.1373	14.2133
liver tumor	34				12.7154	16.4206	13.0437
liver tumor	42	5.9478	9.4573	11.3887			
liver tumor	43	7.6456	12.2411	24.1769			
liver tumor	52	6.9272	12.4499	45.9735			
liver tumor	53	8.056	13.711	50.3891			
liver tumor	54	6.9105	12.0933	36.3222	6.9105	10.6155	13.0413
liver tumor	55	5.6439	9.8362	18.2816			
liver tumor	67	5.6439	11.4067	54.2979			
liver tumor	68	8.9305	13.4236	22.5199	8.9305	5.6439	9.7582
liver tumor	71				8.7243	5.6439	8.459
liver tumor	72				13.5468	8.4289	34.7247
liver tumor	75	8.1416	11.1872	8.2573			
liver tumor	77				8.7397	5.6439	8.5495
liver tumor	84	5.6439	10.5151	29.2685			
liver tumor	85	6.1767	9.7128	11.6002			
liver tumor	89	5.9783	10.6724	25.8867	5.9783	9.487	11.3827
liver tumor	97	6.1767	10.4735	19.6546			

Tumor name	SEQ ID NO.	expression in tumor (normalized with adjacent tissue)	expression in edjacent tissue	fold change tumor vs. adjacent tissue	expression in tumor (normalized with normal tissue from same origin)	expression in normal tissue	fold change tumor vs. normal tissue
liver tumor	98				9.6983	5.9651	13.2986
liver tumor	105	7.5261	10.755	9.3756			
liver tumor	123	6.1498	10.4629	19.8783	6.1498	9.3726	9.3359
liver tumor	124	7.3339	12.7512	42.7337			
liver tumor	125	8.8988	14.4045	45.4338			
liver tumor	126	5.6439	10.1131	22.1508	5.6439	9.0383	10.5154
liver tumor	128	6.3768	11.1309	26.9851	6.3768	9.4265	8.28
liver tumor	129	10.694	13.7695	8.4298			
liver tumor	134				8.6774	5.6439	8.1879
liver tumor	136				9.0174	5.6439	10.3644
liver tumor	137	5.6439	9.6968	16.5978			
liver tumor	140				6.6365	12.4385	55.794
liver tumor	144	6.2799	9.3058	8.1451			
liver tumor	153				11.2587	5.6439	49.0037
liver tumor	154	7.0849	11.2078	17.4237			
liver tumor	160				12.5216	8.1619	20.5301
liver tumor	170				11.1225	5.6439	44.5913
liver tumor	171				12.0826	8.4124	12.7299
lung tumor	1				8.9182	5.6439	9.6755
lung tumor	7				11.5932	8.0833	11.3912
lung tumor	19				9.3524	5.6439	13.0731
lung tumor	21				11.1545	7.9081	9.4902
lung tumor	22				10.5157	5.6439	29.2805

Tumor name	SEQ ID NO.	expression in tumor (normalized with adjacent tissue)	expression in edjacent tissue	fold change tumor vs. adjacent tissue	expression in tumor (normalized with normal tissue from same origin)	expression in normal tissue	fold change tumor vs. normal tissue
lung tumor	26				10.9013	7.108	13.8646
lung tumor	30	12.9579	10.8355	4.3542			
lung tumor	35				12.213	9.0957	8.6773
lung tumor	37				15.8972	11.0328	29.1303
lung tumor	39				13.1493	7.7314	42.7495
lung tumor	43				12.4715	8.2997	18.0242
lung tumor	46				11.1614	6.9494	18.5325
lung tumor	51	11.8655	9.7318	4.3886			
lung tumor	53				9.2535	5.6439	12.2067
lung tumor	58				9.9916	5.6439	20.3612
lung tumor	59				13.6929	10.3181	10.3735
lung tumor	63	11.2934	9.087	4.6152	11.2934	7.8578	10.8199
lung tumor	64				12.5269	7.7205	27.9817
lung tumor	68				12.0464	7.5027	23.3222
lung tumor	71				10.6558	5.6439	32.2662
lung tumor	72				14.4623	7.8499	97.8444
lung tumor	74				9.5731	6.4024	9.0048
lung tumor	78				11.605	8.0965	11.3804
lung tumor	86				9.9302	6.9072	8.1286
lung tumor	87				14.0192	10.9769	8.238
lung tumor	106				13.3418	8.843	22.6093
lung tumor	110				10.7156	7.5664	8.8714
lung tumor	113				12.3539	7.753	24.2654

Tumor name	SEQ ID NO.	expression in tumor (normalized with adjacent tissue)	expression in edjacent tissue	fold change tumor vs. adjacent tissue	expression in tumor (normalized with normal tissue from same origin)	expression in normal tissue	fold change tumor vs. normal tissue
lung tumor	116				14.359	10.8875	11.0921
lung tumor	119				12.8314	7.9024	30.4649
lung tumor	125				10.8698	7.8238	8.2593
lung tumor	127				16.2655	12.0203	18.9638
lung tumor	132				16.1443	12.5898	11.7496
lung tumor	135				13.5779	10.544	8.1897
lung tumor	141				10.6952	7.3439	10.206
lung tumor	153				9.7056	5.6439	16.7001
lung tumor	159				13.1578	9.6393	11.4599
lung tumor	161				16.2074	12.384	14.1568
lung tumor	164				10.6052	7.4602	8.8461
lung tumor	165				10.3966	5.6439	26.9598
lung tumor	170				9.8619	5.6439	18.61
lung tumor	171	14.6973	12.2251	5.549	14.6973	8.7291	62.6024
lung tumor	172				8.8633	5.6439	9.314
lung tumor	173				10.9475	7.2544	12.934

Example 3.4: Differential expression of miRNAs in a specific tumor vs. other tumors

Table 10 presents data comparing expression of a list of miRNAs identified in a specific tumor vs. their expression in other tumors, measured by log_2(signal).

SEQ ID NOS: 2, 13, 15, 21, 24, 28, 29, 31, 41, 42, 50, 58, 61, 62, 65, 68, 70, 80, 86, 90, 97, 102, 104, 107, 108, 110, 111, 114, 129, 144-146, 151, 153 and 175 exhibit high fold change when comparing expression in a colon tumor vs. other tumors. SEQ ID NOS: 5, 10, 21, 29, 43, 47, 51, 63, 68-70, 86, 110, 129, 143, 154, 165 and 166 exhibit high fold change when comparing expression in bile vs. other tumors. SEQ ID NOS: 13, 19, 21, 63, 78, 106, 116, 154, 171 and 173 exhibit high fold change when comparing expression in lung tumor vs. other tumors. This is also shown graphically in Figure 3. SEQ ID NOS: 21, 28, 29, 68, 86, 92, 97, 102, 108, 111, 151, 156, 163 and 174 exhibit high fold change when comparing expression in a pancreatic tumor vs. other tumors. SEQ ID NOS: 21, 68, 92, 111 and 174 exhibit high fold change when comparing expression in whole blood vs. other tumors. SEQ ID NOS: 21, 68, 92, 111 and 174 exhibit high fold change when comparing expression in blood cells vs. other tumors.

mets. = metastasis.

Table 10: Differential expression of miRNAs in a specific tumor vs. other tumors

SpecificTumor	SEQ ID NO	Expression in specific tumor	Median of Expression in Other Tumors	Fold Change of Specific Tumor vs. Other Tumors
bile	5	13.4139	5.932	218.286
bile	10	8.672	11.4606	6.6954
bile	21	6.272	9.5663	9.3768
bile	29	5.6439	8.6868	6.8476
bile	43	8.4387	11.1079	8.0178
bile	47	10.0201	6.9687	8.2031
bile	51	8.0217	10.6445	6.0105
bile	63	6.5714	9.6781	7.8443
bile	68	6.6757	10.7649	14.1454
bile	69	10.3447	5.6439	26.007
bile	70	5.8652	8.5374	5.3025
bile	86	5.6439	8.9754	6.2047
bile	110	6.5893	9.482	5.1669

SpecificTumor	SEQ ID NO	Expression in specific tumor	Median of Expression in Other Tumors	Fold Change of Specific Tumor vs. Other Tumors
bile	129	9.0376	11.6442	5.6286
bile	143	6.2496	8.8157	6.0445
bile	154	5.6439	8.7962	9.0922
bile	165	6.2496	9.1646	5.8657
bile	166	11.6982	5.6439	66.4558
bladder tumor	113	8.8277	11.2976	5.4921
bladder tumor	119	9.2888	11.5872	2.9939
bladder tumor	168	8.8754	6.6181	7.6684
blood cells	21	5.6439	9.0919	10.829
blood cells	68	5.8423	10.2021	15.5596
blood cells	92	15.2629	14.033	2.7218
blood cells	111	8.4548	11.2199	7.3816
blood cells	174	5.6439	8.6536	7.2615
blood whole	21	5.6439	9.1124	10.9809
blood whole	68	5.6439	10.2249	18.4282
blood whole	92	15.2816	14.042	2.7482
blood whole	111	5.6439	11.2439	53.2975
blood whole	174	5.6439	8.6752	7.3282
breast tumor	1	9.0819	7.4141	5.2908
breast tumor	71	11.1529	9.5178	3.071
breast tumor	78	11.4927	9.7564	4.4623
breast tumor	131	6.7404	8.5801	3.7164
breast tumor	141	11.0327	9.3245	4.7993
breast tumor	153	7.3036	9.6295	5.2003
cervix tumor pool	43	8.9231	11.0777	5.7309
cervix tumor pool	78	7.5654	10.1104	3.6179
colon tumor	2	9.973	6.8384	10.0797
colon tumor	13	14.5109	11.0047	10.1335
colon tumor	15	11.03	5.9315	41.8219
colon tumor	21	12.3657	9.1855	9.2963
colon tumor	24	10.6601	7.5592	10.3756
colon tumor	28	13.1399	8.0246	43.2745
colon tumor	29	12.8677	8.2353	24.7804
colon tumor	31	14.6095	11.5027	6.9829
colon tumor	41	11.613	8.4324	8.2056
colon tumor	42	11.3845	8.0765	10.16

SpecificTumor	SEQ ID NO	Expression in specific tumor	Median of Expression in Other Tumors	Fold Change of Specific Tumor vs. Other Tumors
colon tumor	50	10.0823	6.4983	21.6827
colon tumor	58	13.0171	8.7694	21.3052
colon tumor	61	10.5309	6.9473	13.9441
colon tumor	62	10.274	7.0724	9.8857
colon tumor	65	10.274	7.2635	7.7966
colon tumor	68	15.0288	10.2429	31.8857
colon tumor	70	11.2703	8.1996	9.5479
colon tumor	80	12.1735	8.7072	9.828
colon tumor	86	13.1224	8.508	30.2317
colon tumor	90	11.3623	8.2779	9.4699
colon tumor	97	13.7096	7.4669	110.075
colon tumor	102	10.8979	7.5072	10.5193
colon tumor	104	11.3977	7.9096	14.4928
colon tumor	107	11.3078	6.9196	22.7611
colon tumor	108	11.0134	7.5852	12.8681
colon tumor	110	13.0666	9.0772	18.59
colon tumor	111	14.5785	11.3948	9.8341
colon tumor	114	11.5233	8.3525	9.5834
colon tumor	129	14.4349	11.3069	10.1277
colon tumor	144	11.613	8.5073	9.7668
colon tumor	145	12.8318	9.7407	10.3359
colon tumor	146	11.1725	7.6549	11.8882
colon tumor	151	11.1419	7.5768	12.0451
colon tumor	153	12.6481	9.1501	9.5641
colon tumor	175	12.2407	8.7369	10.0856
endometrium met.	21	6.6959	9.3726	6.6296
endometrium met.	47	9.5913	7.1482	5.8094
endometrium met.	71	6.0259	9.7102	13.6929
endometrium met.	165	6.0259	8.9931	6.6942
endometrium met.	172	5.6439	8.3155	6.7606
esophagus tumor pool	68	12.464	10.4032	5.389
esophagus tumor pool	70	6.5033	8.4976	3.4073
esophagus tumor pool	71	7.6149	9.8411	4.8051
esophagus tumor pool	77	7.2443	10.0736	5.1808
esophagus tumor pool	78	7.1005	10.1394	4.9935
esophagus tumor pool	157	5.6855	8.5004	6.0769

SpecificTumor	SEQ ID NO	Expression in specific tumor	Median of Expression in Other Tumors	Fold Change of Specific Tumor vs. Other Tumors
esophagus tumor pool	165	6.4225	9.1538	5.2033
kidney tumor	10	8.2677	11.4859	8.8612
kidney tumor	29	5.6439	8.6868	6.8476
kidney tumor	31	8.5682	11.8803	12.8715
kidney tumor	43	7.7767	11.1493	12.6864
kidney tumor	58	6.3084	9.1887	7.0266
kidney tumor	63	5.8343	9.7242	13.0747
kidney tumor	68	7.108	10.7379	10.483
kidney tumor	80	6.1036	9.0866	8.4054
kidney tumor	86	5.6439	8.9754	6.2047
kidney tumor	116	9.4463	12.6382	7.3442
kidney tumor	143	5.8043	8.8436	8.2303
kidney tumor	154	5.6439	8.7962	9.0922
kidney tumor	175	6.1036	9.1205	7.1333
larynx met.	21	6.8407	9.3726	5.9964
larynx met.	26	7.589	10.3592	6.3109
larynx met.	41	5.6784	8.6195	8.304
larynx met.	58	6.1955	9.0193	6.9732
larynx met.	63	7.1683	9.4954	5.0157
larynx met.	71	7.4475	9.7102	5.1115
larynx met.	78	7.1162	9.9607	4.7583
larynx met.	157	5.6439	8.3348	6.1284
larynx met.	162	7.458	9.6529	3.9818
larynx met.	165	5.8094	8.9931	7.778
larynx met.	172	5.8404	8.3155	5.8994
liver tumor	13	8.8664	11.3575	6.7896
liver tumor	21	5.6439	9.6056	14.4923
liver tumor	29	6.2031	8.6519	4.6471
liver tumor	31	9.4181	11.8272	7.1413
liver tumor	43	7.6456	11.1575	13.8931
liver tumor	57	5.9167	8.8467	6.3789
liver tumor	144	6.2799	8.8406	4.4191
liver tumor	175	6.6759	9.0847	4.7975
lung tumor	13	12.8465	10.9929	4.2671
lung tumor	19	9.3524	7.1474	4.6438
lung tumor	21	11.1545	9.135	4.8608

SpecificTumor	SEQ ID NO	Expression in specific tumor	Median of Expression in Other Tumors	Fold Change of Specific Tumor vs. Other Tumors
lung tumor	63	11.2934	9.2556	4.4726
lung tumor	78	11.605	9.7414	4.8233
lung tumor	106	13.3418	11.1365	6.7664
lung tumor	116	14.359	12.1959	5.1327
lung tumor	154	10.6807	8.3348	5.9107
lung tumor	171	14.6973	12.7752	3.148
lung tumor	173	10.9475	8.8186	3.7255
melanocyte met.	41	6.0414	8.6195	6.4567
melanocyte met.	63	6.5381	9.4954	7.763
melanocyte met.	68	8.0712	10.5244	4.2514
melanocyte met.	110	6.9266	9.3119	4.0313
melanocyte met.	144	6.288	8.69	4.1583
melanocyte met.	165	6.4829	8.9931	4.8767
ovary tumor	21	6.5634	9.5481	7.6619
ovary tumor	30	8.4008	11.568	8.7972
ovary tumor	59	9.4013	12.3332	7.45
ovary tumor	60	8.0467	10.9593	6.9747
ovary tumor	68	7.7255	10.6993	6.8329
ovary tumor	76	9.9593	13.2121	8.7485
ovary tumor	111	8.6294	11.7666	8.6303
ovary tumor	113	8.1091	11.1881	8.4256
pancreas tumor	21	12.103	9.2019	7.7486
pancreas tumor	28	12.0653	8.0917	20.5467
pancreas tumor	29	12.1695	8.279	15.2722
pancreas tumor	68	14.3376	10.2861	19.7476
pancreas tumor	86	12.1654	8.5678	15.5732
pancreas tumor	92	17.0467	14.2347	7.6442
pancreas tumor	97	12.086	7.5684	35.7222
pancreas tumor	102	12.0863	7.433	23.9737
pancreas tumor	108	10.9037	7.5921	11.926
pancreas tumor	111	14.3811	11.4071	8.5767
pancreas tumor	151	10.6144	7.6098	8.3564
pancreas tumor	156	10.4869	6.4422	24.8635
pancreas tumor	163	9.6776	6.678	9.9674
pancreas tumor	174	11.5946	8.7629	8.3758
prostate tumor	71	11.8659	9.5754	4.5892

SpecificTumor	SEQ ID NO	Expression in specific tumor	Median of Expression in Other Tumors	Fold Change of Specific Tumor vs. Other Tumors
prostate tumor	77	11.9097	9.782	6.1592
prostate tumor	130	18.4404	16.2807	5.6929
salivary gland met.	16	9.208	6.4101	8.8769
salivary gland met.	17	8.8419	11.143	4.73
salivary gland met.	43	8.31	10.9509	8.5724
salivary gland met.	58	6.4638	9.0193	5.7899
salivary gland met.	68	7.7986	10.5244	5.1355
salivary gland met.	86	6.6123	8.7794	3.0775
tongue met.	9	9.9662	6.8501	10.4902
tongue met.	41	11.8019	8.6195	8.3956
tongue met.	43	15.5575	10.9509	17.7258
tongue met.	79	9.8447	7.2277	7.3411
tongue met.	116	14.8843	12.4504	6.6506
tongue met.	122	10.6417	7.8044	7.2028
tongue met.	157	5.8988	8.3348	5.1356
tongue met.	165	5.8461	8.9931	7.5827

Example 3.5: Differential expression of miRNAs in a specific tumor vs. sites of metastases

Tables 12A-C presents data comparing expression of a list of miRNAs identified in a specific tumor vs. their expression in other tumors, measured by $\log_2$(signal). Table 11A details the differential expression of miRNAs in a specific primary tumor vs. metastisis originating from the same tissue or the metastasis vs. the median of metasteses originating from other tissues; table 11B details miR expression in a tumor originating from a specific tissue, in metastases of same tissue to the lung, in normal lung tissue, in lung primary tumors (normalized to the metasteses), and in median of metastasis of other origins; table 11C details miR expression in a tumor originating from a specific tissue, in metastasis of same tissue to the lymph node, in normal lymph node tissue, in lymph node primary tumors, and in median of metastasis of other origins;

"Specific tumor signal" refers to the expression in a specific tumor tissue. "Mets signal" refers to expression in the metastasis originating from the specific tissue. "Other Mets Signal" refers to expression in median of metastasis of various origins.

Figure 1 shows differential expression of miRs, comparing the median values of each miR in breast primary tumor with breast metastases into lymph nodes. Figure 2 shows differential expression of miRs, comparing the median values of each miR in colon tumors with the corresponding median for their adjacent tissues. Figure 3 shows differential expression of miRs, comparing the median values of each miR in lung tumors with the corresponding median for other tumors from the following tissues: bile duct, bladder, breast, colon, kidney, liver, lung, ovary, pancreas, and prostate.

Table 11A: miR expression in tumor originating from a specific tissue, in metastasis originating from same tissue, and in median of metastasis originating from other tissues

Specific Tumor	SEQ ID NO	Specific Tumor Signal	Mets Signal	Fold-Change Specific Tumor Vs Mets	Other Mets Signal	Fold-Change Mets Vs Other Mets Signal
breast tumor	1	9.0819	5.6439	10.8382		
	22	10.0955	5.9107	18.1869		
	58	9.3914	5.6439	13.4314		
	63	9.3135	5.6439	12.7254		
	68	10.3729	7.0746	9.8372		
	77	11.3245	8.0619	9.5968		
	78	11.4927	8.1534	10.1217		
	173	9.9531	6.8425	8.6373		
kidney tumor	10	8.2677	10.3475	4.2275		
	101		5.6439		8.7142	8.1041
lung tumor	21		10.5863		7.3236	13.4133
	46		11.4034		8.5823	7.9782
	50	7.2735	8.8099	2.9007		
	56	11.6372	10.1372	2.8285		
	63		10.1389		7.1016	7.8391
	139		8.743		5.8253	8.569
	150	8.6558	7.0032	3.1441		
	155	9.1188	7.1407	3.9397		
endometrium metastasis	29		10.4813		7.1832	10.6977
larynx metastasis	25		15.4024		13.6947	3.3016

Table 11B: miR expression in tumor originating from a specific tissue, in metastasis of same tissue to the lung, in normal lung tissue, in lung primary tumors, and in median of metastasis of other origins

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
endometrium met.	1		5.6439		9.0722	10.7655	8.7946	8.8811		
endometrium met.	2		6.5282				9.6594	8.7617		
endometrium met.	6		5.7289				9.2209	11.2511		
endometrium met.	19		5.6439				9.1461	11.3316		
endometrium met.	21				10.1207	9.9087	11.1907	26.4761		
endometrium met.	26		8.1894		11.3487	8.934				
endometrium met.	28		6.6907				9.7436	8.2991		
endometrium met.	30		10.2284				13.516	9.7647		
endometrium met.	33		9.9839				13.6414	12.6188		
endometrium met.	47				6.6177	9.1858				
endometrium met.	51		9.0059				12.4736	11.0627		
endometrium met.	56		8.5993				12.286	12.8766		
endometrium met.	63		6.898				11.9377	32.8928		
endometrium met.	71				10.913	28.5829	9.9969	16.7176		
endometrium met.	74		5.6439		9.2542	12.2131	9.7105	16.7565		
endometrium met.	86		6.8337		9.771	7.6598	10.0742	12.0736		
endometrium met.	87		10.0087				13.9957	15.856		
endometrium met.	106		9.2344				13.3445	17.2685		
endometrium met.	110		8.1566				11.4419	9.7491		

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
endometrium met.	111		9.6023				13.3857	13.7692		
endometrium met.	116		10.6661				14.3029	12.439		
endometrium met.	129		9.5849				13.0295	10.8874		
endometrium met.	138		10.5714				14.2966	13.225		
endometrium met.	139		5.6439				9.028	10.4409		
endometrium met.	147		8.1066				11.2164	8.6329		
endometrium met.	150		5.6439				9.0446	10.5612		
endometrium met.	154		6.7449		9.7827	8.2124	10.804	20.9657		
endometrium met.	164		7.1409				10.2439	8.5923		
endometrium met.	165				10.088	16.1352	10.1001	17.957		
endometrium met.	172						8.731	8.4982		
kidney tumor	1		5.6439		8.948	9.8778				
kidney tumor	2		5.6439				9.4507	13.9949		
kidney tumor	6		5.6439				9.0188	10.3741		
kidney tumor	11		8.3067				11.9813	12.7696		
kidney tumor	13		8.864				12.0435	9.0601		
kidney tumor	18		7.3239		10.8625	11.6208				
kidney tumor	19		5.6439				8.9452	9.8585		
kidney tumor	21		5.6439		9.9833	20.2442	10.967	40.0332		
kidney tumor	23		5.6439		8.8096	8.9741	8.9209	9.6936		
kidney tumor	24		5.6439		8.7022	8.33				
kidney tumor	26				11.2091	8.3113				

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
kidney tumor	28		5.6439				9.5338	14.8249		
kidney tumor	29				9.5764	11.9394	9.4095	11.4969		
kidney tumor	30		10.2765				13.2892	8.0704		
kidney tumor	31		9.2217		12.6468	10.512	12.7891	19.9977	12.0753	6.3889
kidney tumor	33		9.4189				13.4151	15.9579		
kidney tumor	40		7.7939				11.1058	9.931		
kidney tumor	41		5.6439		9.9302	19.5134				
kidney tumor	42		5.6439				9.2121	11.8615		
kidney tumor	43		7.4835		11.8996	21.2798	11.9044	29.7676	10.9376	5.1201
kidney tumor	46		7.2685				10.5174	9.5065		
kidney tumor	51		8.2161				12.2454	16.3274		
kidney tumor	56		8.0854				12.058	15.6989		
kidney tumor	57		5.6439		9.1783	11.5869	9.0159	10.3535		
kidney tumor	58				9.1566	11.4137	9.1583	11.4274		
kidney tumor	59		9.8473				13.4845	12.4424		
kidney tumor	63						11.7105	42.7219		
kidney tumor	68						11.5654	13.1429		
kidney tumor	70		6.7942				10.3653	11.8854		
kidney tumor	71	9.9576	6.2584	12.9886	10.7725	22.9006	9.7839	13.1346		
kidney tumor	74		5.6439		9.127	11.1826	9.5011	14.4928		
kidney tumor	75		6.5917				9.6908	8.5687		
kidney tumor	77		6.9561		11.2185	19.1976	10.7114	17.3847	9.4966	3.8941

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
kidney tumor	78		8.048		11.6316	11.9888	10.6711	9.2565		
kidney tumor	80		5.6439		8.7348	8.5206	9.0556	10.6424	8.912	8.7406
kidney tumor	81		6.3142				9.4866	9.0153		
kidney tumor	86				9.6369	15.9233	9.8603	18.5894		
kidney tumor	87		9.3881				13.771	20.8639		
kidney tumor	90		5.9172		9.1547	9.4319	9.3784	11.7033		
kidney tumor	91		5.6439				8.9927	10.1884		
kidney tumor	95		6.1856				9.2221	8.2047		
kidney tumor	99		5.6439		8.7594	8.667				
kidney tumor	100		5.6439				9.2857	12.4823		
kidney tumor	104		5.6439				8.6852	8.2324		
kidney tumor	106		8.51				13.1172	24.3717		
kidney tumor	108		5.6439		9.344	12.9969				
kidney tumor	110		6.8786				11.2167	20.2264		
kidney tumor	111		8.9338				13.1585	18.6969		
kidney tumor	112		9.0351				12.3062	9.6541		
kidney tumor	116						14.0801	24.1387		
kidney tumor	117		8.7767				12.1335	10.2445		
kidney tumor	119		9.0551				12.7252	12.7297		
kidney tumor	122		5.6439		8.9733	10.0522	8.981	10.1059		
kidney tumor	129		8.6035				12.8014	18.3534		
kidney tumor	133		14.5413				17.6014	8.3406		

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
kidney tumor	138		9.9338				14.0737	17.6297		
kidney tumor	139		5.6439				8.8291	9.0963		
kidney tumor	141		7.4335				10.5226	8.5098		
kidney tumor	142		6.2079				9.3351	8.7371		
kidney tumor	145		7.3679				10.6711	9.871		
kidney tumor	146		5.6439				8.8678	9.3433		
kidney tumor	147		6.9321				10.9926	16.6847		
kidney tumor	148		5.6439				8.6743	8.1708		
kidney tumor	149		5.6439				9.2246	11.965		
kidney tumor	150		5.6439				8.8454	9.1992		
kidney tumor	152		5.8295				8.9055	8.4323		
kidney tumor	154						10.583	8.1991		
kidney tumor	164		5.6439		10.3408	25.9376	10.0281	20.8832	8.9187	9.8453
kidney tumor	165						9.8859	9.0727		
kidney tumor	169		5.6439		8.9288	9.7469	8.6671	8.1297		
kidney tumor	171		11.2711				14.314	8.2415		
kidney tumor	173		7.2685				10.5349	9.6221		
kidney tumor	175						9.4507	8.5512		
larynx met.	1		5.6439		9.1668	11.495	8.8488	9.221		
larynx met.	2		5.8529				9.7458	14.8548		
larynx met.	4		6.7035		9.6895	7.9227				
larynx met.	6		5.6439				9.2898	12.5182		

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
larynx met.	19		5.6439		8.5333	7.4098	9.2123	11.8633		
larynx met.	21				10.2398	9.7006	11.3584	26.3094		
larynx met.	26				11.4977	13.4395				
larynx met.	30		10.4953				13.8711	10.3801		
larynx met.	33		10.8948				14.0089	8.6584		
larynx met.	41				10.1851	22.6515				
larynx met.	43		8.7743				12.3661	12.0574		
larynx met.	46		7.5004				10.8782	10.3952		
larynx met.	51		8.7771				12.7348	15.5379		
larynx met.	56		9.1539				12.532	10.3974		
larynx met.	58				9.3837	8.6943	9.4369	10.1297		
larynx met.	61		5.6439				8.779	8.7857		
larynx met.	63						12.157	37.6432		
larynx met.	70		7.5571				10.7161	8.9321		
larynx met.	71				11.0512	10.9252				
larynx met.	74		5.6439		9.353	13.0787	9.7991	17.8179		
larynx met.	78		7.1162		11.9281	25.533	11.0421	17.9323	9.2817	2.9352
larynx met.	86		6.9274		9.8819	7.7514	10.1794	11.8434		
larynx met.	91		6.1559				9.2623	8.6124		
larynx met.	96		6.8192				9.8862	8.3802		
larynx met.	106		9.742				13.6831	15.3597		
larynx met.	108		6.7819		9.5783	6.9472				

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
larynx met.	110		7.6511				11.6261	15.7248		
larynx met.	111		10.3963				13.7282	10.0692		
larynx met.	116		11.6913				14.7397	8.2728		
larynx met.	117		9.314				12.6138	9.8474		
larynx met.	119		9.7531				13.2557	11.3344		
larynx met.	129		10.1481				13.3387	9.1303		
larynx met.	139		5.6439				9.09	10.8992		
larynx met.	141		8.1668		11.2789	8.6464	10.8837	9.4421		
larynx met.	147		7.2807				11.3858	17.2088		
larynx met.	150		5.6439				9.1071	11.0291		
larynx met.	151		5.6439		8.4331	6.9124				
larynx met.	153		11.2932		8.6652	7.1846			9.4486	3.785
larynx met.	154		7.3144				10.9482	12.4132		
larynx met.	157				8.4331	6.9124				
larynx met.	162						10.8245	12.5527		
larynx met.	165				10.2064	20.7304	10.2066	21.5322		
melanocyte met.	1		5.6439		9.5073	14.5549	9.4712	14.1956		
melanocyte met.	2		5.6439				10.3993	27.0104		
melanocyte met.	6		5.6439				9.9326	19.5453		
melanocyte met.	7		8.9273				12.237	8.9662	10.1712	2.9845
melanocyte met.	10		9.7614				13.0178	8.62	10.437	1.651
melanocyte met.	17		8.8444				12.2351	9.4898	10.378	2.091

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
melanocyte met.	19		5.8409		8.858	8.0953	9.8522	16.9029		
melanocyte met.	21		8.3997				11.9733	10.8183	7.636	2.9464
melanocyte met.	22		7.43		10.7569	7.3738	11.1258	11.9637	8.3536	1.7632
melanocyte met.	26		8.9744		11.8623	7.4019				
melanocyte met.	28		6.3154				10.4881	18.0353		
melanocyte met.	29		7.2064		10.1719	7.8112	10.3552	10.6121		
melanocyte met.	30		9.5762				14.2085	22.3718	10.8856	2.6798
melanocyte met.	38		8.687		11.6406	7.7468				
melanocyte met.	41				10.5427	19.8969				
melanocyte met.	47		9.647		6.9138	6.6494				
melanocyte met.	51		8.3865				13.2292	26.0771	9.6289	2.4044
melanocyte met.	56		8.7439				13.049	17.9005	9.553	1.368
melanocyte met.	61		5.706				9.3973	12.9178		
melanocyte met.	63				9.6795	7.1905	12.7114	68.209		
melanocyte met.	64		9.7878				12.9804	8.2473	10.877	1.782
melanocyte met.	68		8.0712				12.569	20.6184	9.4971	3.8737
melanocyte met.	70		6.8752				11.3601	22.3918		
melanocyte met.	71		8.6042		11.4154	7.019				
melanocyte met.	74		6.7259		9.6974	7.8437	10.4532	15.2149		
melanocyte met.	77		8.496		11.8719	6.9937	11.7146	8.4492	9.2766	1.3392
melanocyte met.	78		8.0779		12.2912	12.8793	11.6736	11.0315	9.1443	1.5071
melanocyte met.	84		6.3057		9.1784	7.3243				

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
melanocyte met.	86		7.4009		10.2355	7.1336	10.834	13.1199		
melanocyte met.	87		11.4553				14.6466	8.3433	12.1204	1.2325
melanocyte met.	88		6.7789				9.8256	8.2634		
melanocyte met.	91		6.2718				9.9042	12.4009		
melanocyte met.	97		5.6439				8.8005	8.9175		
melanocyte met.	106		9.4479		12.7695	6.3874	14.0499	21.9112	10.4382	1.3742
melanocyte met.	107		5.6439				9.1142	11.0838		
melanocyte met.	110						12.2236	36.7392		
melanocyte met.	116		10.9485				14.9231	14.272	12.2505	1.4375
melanocyte met.	117		8.6919				13.1218	19.5274	9.8323	2.1406
melanocyte met.	119		10.0213				13.6845	11.4327	11.0045	1.8066
melanocyte met.	129		10.2952				13.756	9.9473	10.8502	1.091
melanocyte met.	139		5.6439				9.7248	16.9237		
melanocyte met.	142		7.0732				10.2752	9.2024		
melanocyte met.	144				9.643	8.643	10.3924	16.3918		
melanocyte met.	145		8.3758				11.6736	8.9379	9.1042	2.1235
melanocyte met.	146		6.4889				9.7673	9.7033		
melanocyte met.	147		8.1173				11.999	13.443	8.3853	1.1589
melanocyte met.	150		5.6439		8.3427	6.4928	9.7427	17.1345		
melanocyte met.	154		7.6883				11.5837	14.881		
melanocyte met.	162		7.8416				11.4648	11.2868	8.7332	1.386
melanocyte met.	165				10.5643	13.9012	10.861	19.6877		

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
melanocyte met.	173		7.5619				11.5344	15.6978		
salivary gland met.	1		5.6439		9.6632	16.2155	9.7919	17.7288		
salivary gland met.	2		6.1021				10.7906	25.7852		
salivary gland met.	6		5.6439				10.288	25.0042		
salivary gland met.	7		9.6584				12.78	8.7039		
salivary gland met.	11		10.2652				13.5845	9.982		
salivary gland met.	13		10.2947		13.6122	9.9689	13.6502	13.3108		
salivary gland met.	16		9.208		6.1402	13.999			6.4208	6.724
salivary gland met.	17						12.778	12.044		
salivary gland met.	19		6.2096				10.2015	15.9099		
salivary gland met.	20		7.9441				11.2355	9.791		
salivary gland met.	21		8.2823				12.4935	18.5223		
salivary gland met.	22		7.8607		10.9753	8.6614	11.575	16.0422		
salivary gland met.	24		6.6966				9.7124	8.0885		
salivary gland met.	28		6.5797				10.8863	19.7896		
salivary gland met.	29		7.5665		10.3599	6.9327	10.743	10.9367		
salivary gland met.	30		10.2815				14.9374	25.209		
salivary gland met.	31		10.8117				14.4272	12.2569		
salivary gland met.	40		9.2908				12.645	10.2261		
salivary gland met.	41		7.7397		10.7498	8.0563				
salivary gland met.	43		8.31		12.8882	15.2495	13.503	29.4886	10.8195	2.887
salivary gland met.	46		8.7246				11.9979	9.6685		

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
salivary gland met.	49		9.2996				12.6585	10.2597		
salivary gland met.	51		9.3757				13.8625	22.4204		
salivary gland met.	56		9.4123				13.6654	19.0693		
salivary gland met.	58				9.895	8.0221	10.4512	13.8901		
salivary gland met.	61		6.3625				9.7124	10.1958		
salivary gland met.	63		8.0617				13.2968	37.6637		
salivary gland met.	64		10.0274				13.5905	11.8196		
salivary gland met.	65		5.7769		8.7444	7.8214				
salivary gland met.	68						13.1416	33.4853		
salivary gland met.	70		8.1428				11.8285	12.8684		
salivary gland met.	74		7.1118				10.8487	13.333		
salivary gland met.	77		9.1282		12.1546	8.148	12.2127	10.7579		
salivary gland met.	78		9.3549		12.6003	9.4835	12.1682	8.9632		
salivary gland met.	81		7.6176				10.832	9.2815		
salivary gland met.	84		5.6439		9.3193	12.7769				
salivary gland met.	85		5.6439		8.6868	8.2418				
salivary gland met.	86				10.4267	10.3198	11.2596	21.801		
salivary gland met.	87		11.7297				15.4207	12.9158		
salivary gland met.	88		7.0275				10.1728	8.8477		
salivary gland met.	91		6.118				10.2574	17.6224		
salivary gland met.	92		14.0143				17.1625	8.8657		
salivary gland met.	96		7.55				10.9434	10.5076		

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
salivary gland met.	97		5.6439				9.0717	10.762		
salivary gland met.	100		7.2255				10.5995	10.3679		
salivary gland met.	104		6.6017				9.8936	9.7939		
salivary gland met.	106		10.0543				14.7629	26.1468		
salivary gland met.	107		5.6439				9.4085	13.5914		
salivary gland met.	108		7.1824		10.1028	7.5708				
salivary gland met.	110		8.0833				12.7655	25.6736		
salivary gland met.	113		9.6325				13.2619	12.3752		
salivary gland met.	116		11.6828				15.7265	16.4914		
salivary gland met.	117		9.6107				13.745	17.561		
salivary gland met.	119		10.4625		13.2424	6.8679	14.3614	19.4568		
salivary gland met.	122		7.0116				10.2436	9.3955		
salivary gland met.	129		10.6242				14.4399	14.0812		
salivary gland met.	139		6.2243				10.0645	14.322		
salivary gland met.	142		7.5012				10.6569	8.9117		
salivary gland met.	145		8.1964				12.1682	15.6901		
salivary gland met.	146		6.9118				10.1102	9.1789		
salivary gland met.	147		8.6423				12.5214	14.7137		
salivary gland met.	149		7.5012				10.5285	8.1533		
salivary gland met.	150		6.8133				10.0837	9.6491		
salivary gland met.	151		5.8981		8.8773	7.8856	8.8052	8.4021		
salivary gland met.	154		8.8894				12.0707	9.0713		

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lung) signal	fold-change specific tumor vs mets to lung	signal of normal lung tissue	fold-change mets vs normal lung tissue	lung primary tumor signal	fold-change mets vs primary lung tumor	other mets signal	fold-change mets vs other mets signal
salivary gland met.	162		8.1898				11.9419	13.4741		
salivary gland met.	163		5.9211		8.8411	7.5683				
salivary gland met.	165		7.4148		10.7725	10.2506	11.2887	17.6347		
salivary gland met.	171		11.7514				15.9556	18.4325		
salivary gland met.	173		7.5659				12.0173	21.8773		

Table 11C: miR expression in tumor originating from a specific tissue, in metastasis of same tissue to the lymph node, in normal lymph node tissue and in median of metastasis of other origins

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lymph node) signal	fold-change specific tumor vs mets to the lymph node	signal of normal lymph node tissue	fold-change mets vs Normal lymph node	other mets signal	fold-change mets vs other mets signal
breast tumor	9		5.6439		8.9933	10.1927		
breast tumor	10		8.7077		13.0896	20.8486		
breast tumor	21		6.1646		10.8422	25.5919		
breast tumor	26		10.745		7.2376	11.3724		
breast tumor	29		5.6439		9.5162	14.6456		
breast tumor	37		16.0999		10.4792	49.2044		
breast tumor	38		7.7221		11.7586	16.4102		
breast tumor	39		13.5516		7.3382	74.2043		
breast tumor	52		8.9108		5.8037	8.6168		
breast tumor	53		10.4996		5.6439	28.9546		
breast tumor	71		11.4946		5.6439	58.8841	7.994	11.7941

WO 2011/024157 PCT/IL2010/000462

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lymph node) signal	fold-change specific tumor vs mets to the lymph node	signal of normal lymph node tissue	fold-change mets vs Normal lymph node	other mets signal	fold-change mets vs other mets signal
breast tumor	72		14.998		6.7779	298.184		
breast tumor	99				9.255	11.197		
breast tumor	102		5.6439		9.1576	11.422		
breast tumor	118		5.9245		9.6467	13.1971		
breast tumor	124		9.4809		6.2203	9.5838		
breast tumor	125		11.2983		7.7482	11.7132		
breast tumor	127		16.6339		11.2697	41.1889		
breast tumor	130				12.789	30.6617		
breast tumor	132		16.0615		10.3778	51.4007		
breast tumor	133		17.5848		13.1556	21.5439		
breast tumor	140		5.687		8.7698	8.4728		
breast tumor	143		6.1323		9.164	8.1776		
breast tumor	153				5.6439	24.9769		
breast tumor	154				10.726	21.6812		
breast tumor	161		16.3989		10.2953	68.7665		
breast tumor	164		9.1537		5.6439	11.3909		
breast tumor	165		9.2767		5.6439	12.4053		
breast tumor	170	8.6277	12.4699	14.3423	5.6439	119.066	8.6838	11.3111
breast tumor	171		14.6244		7.4984	139.682		
breast tumor	172	7.716	10.9644	9.5033	5.6439	40.1841	6.559	20.8766
lung tumor	1		10.1283		5.6439	23.6963	5.6803	22.3854
lung tumor	19		10.7286		5.6439	36.579	5.8337	33.9348
lung tumor	22		11.1633		7.0595	18.7769	7.8202	10.4512
lung tumor	26		12.2942		7.3614	34.4819	8.579	14.7383
lung tumor	35				9.3045	9.7026		
lung tumor	37				10.5833	60.6402		
lung tumor	39				7.4624	57.8007		

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lymph node) signal	fold-change specific tumor vs mets to the lymph node	signal of normal lymph node tissue	fold-change mets vs Normal lymph node	other mets signal	fold-change mets vs other mets signal
lung tumor	43		13.4306		10.1401	11.4047	10.088	16.4665
lung tumor	53				5.6439	10.5516		
lung tumor	58		9.9126		6.1705	14.0719	6.5025	10.919
lung tumor	64				9.5283	12.0721		
lung tumor	71		11.3779		5.6439	58.4965	8.0107	10.8775
lung tumor	72				6.8977	208.191		
lung tumor	74		9.7072		5.6439	17.4732	6.3389	16.718
lung tumor	77		12.2502		8.8913	11.5679	8.7403	13.4933
lung tumor	78		12.522		8.2015	22.7042	8.5094	20.6583
lung tumor	84		8.7889		5.6439	8.8461		
lung tumor	86		10.2718		7.2042	8.9132	6.6737	11.7596
lung tumor	87				11.4388	9.8248		
lung tumor	106		13.3138		9.5212	16.1002	9.886	12.6336
lung tumor	116				10.746	14.4933		
lung tumor	119		13.522		9.7681	15.7637	10.5044	9.3599
lung tumor	127				11.3688	48.7888		
lung tumor	130		17.4311		12.8851	23.3618		
lung tumor	132				10.4827	40.9783		
lung tumor	133		16.7444		13.2524	11.2514		
lung tumor	135				10.5092	8.5579		
lung tumor	140		5.6439		8.8884	9.478		
lung tumor	141				7.7001	15.8416		
lung tumor	153				5.6439	13.7203		
lung tumor	157		9.5336		6.4178	9.013	6.4525	8.1737
lung tumor	161				10.4008	51.4032		
lung tumor	164				5.6439	15.9481		
lung tumor	165		11.2983		5.6439	55.2264	6.7883	28.157

Specific Tumor Name	SEQ ID NO	specific tumor signal	mets (to the lymph node) signal	fold-change specific tumor vs mets to the lymph node	signal of normal lymph node tissue	fold-change mets vs Normal lymph node	other mets signal	fold-change mets vs other mets signal
lung tumor	170				5.6439	17.1137		
lung tumor	171	14.6973	11.8911	6.9943	7.623	21.499		
lung tumor	172				5.6439	9.5662		
lung tumor	173		11.2826		7.2236	18.2668	7.8574	11.1543
tongue metastasis	37		15.5842		10.889	35.0695	15.5259	1.2676
tongue metastasis	39		12.9526		7.529	59.7729	12.3896	1.5414
tongue metastasis	43		15.5575		10.4176	47.7721	9.7842	71.923
tongue metastasis	53		9.3584		5.6439	16.2684	8.9764	1.2734
tongue metastasis	67		5.6439		8.783	8.8097		
tongue metastasis	72		13.3762		6.9188	122.682	13.4912	1.2238
tongue metastasis	74		8.4553		5.6439	8.2645	6.5177	7.0198
tongue metastasis	79		9.8447		6.0265	17.9181	7.1997	6.6408
tongue metastasis	80		11.245		8.2352	10.8272	8.1119	8.3071
tongue metastasis	95		10.8822		6.996	19.6306	8.0306	6.4425
tongue metastasis	116		14.8843		11.0612	19.5062	11.6882	12.9761
tongue metastasis	122		10.6417		6.5793	21.9783	7.0203	14.3499
tongue metastasis	127		17.3276		11.7174	48.8468	16.3032	1.9613
tongue metastasis	130		18.0674		13.2891	27.4422	17.3235	1.5543
tongue metastasis	132		15.6516		10.7822	39.4848	15.416	1.1547
tongue metastasis	133		18.0038		13.6641	20.2488	16.702	2.9706
tongue metastasis	136		8.4672		5.6439	8.3392	8.1413	1.1523
tongue metastasis	153		10.2834		5.6439	32.3079	9.5929	1.8798
tongue metastasis	161		15.9364		10.6952	50.5892	15.6956	1.1168
tongue metastasis	164		8.6487		5.6439	9.5596	8.4894	1.2265
tongue metastasis	170		9.4847		5.6439	17.8757	9.1103	1.4284
tongue metastasis	171		13.7948		7.7029	95.2131	12.6898	2.7184

Example 3.6: Differential expression of miRNAs in blood vs. colon tumor or adjacent colon tissue

Table 12 presents data comparing expression of a list of miRNAs measured by $\log_2$(signal). "Colon tumor expression" refers to expression in a colon tumor. "Fold change blood vs. colon tumor signal" refers to the fold change of expression in a blood tissue vs. colon tumor. "Expression in adjacent colon" refers to expression in colon tissue adjacent to the specific tumor. This is shown graphically in Figure 2. "Fold change blood vs. adjacent site" refers to the fold change obtained upon comparison of the blood tissue vs. adjacent colon tissue.

Blood cells and whole blood samples were collected from normal individuals. miRNA expression was measured in both sample types and compared with miRNA expression in colon tumors and in normal colon tissue adjacent to a tumor site. SEQ ID NOS: 21, 68, 111 and 174 exhibit high fold change in expression in blood cells or whole blood or both when compared to expression in both colon tumor tissue and colon tissue adjacent to the tumor site. SEQ ID NO: 92 exhibit low fold change in expression in both blood cells and whole blood when compared to expression in both colon tumor tissue and colon tissue adjacent to the tumor site.

Table 12: Differential expression of miRNAs in blood vs. colon tumor or adjacent colon tissue

SEQ ID NO	blood tissue	blood expression	colon tumor expression	fold change blood vs. colon tumor expression	expression in adjacent colon	fold change blood vs. adjacent colon
21	blood cells	5.6439	11.7925	70.9461	10.8645	37.2871
21	blood whole	5.6439	11.8239	72.5092	10.8433	36.7443
68	blood cells	5.8423	14.4778	397.6962	12.6822	114.5514
68	blood whole	5.6439	14.4994	463.2309	12.6645	129.8426
92	blood cells	15.2629	12.7751	5.6093	12.8053	5.493
92	blood whole	15.2816	12.8086	5.5522	12.7875	5.6338

SEQ ID NO	blood tissue	blood expression	colon tumor expression	fold change blood vs. colon tumor expression	expression in adjacent colon	fold change blood vs. adjacent colon
111	blood cells	8.4548	13.9891	46.3451	13.7154	38.335
111	blood whole	5.6439	14.0177	331.7221	13.6951	265.2569
174	blood cells	5.6439	11.0818	43.3491	10.5812	30.6392
174	blood whole	5.6439	11.1099	44.2027	10.5591	30.1732

Example 4: Deep sequencing of small RNAs from solid tumor samples

Example 4.1: Expression and sequence variability of known microRNAs

The sequencing process yielded 141,023 sequences from the bladder+breast tumor pool and 90,986 sequences from the colon+lung pool. After combining identical sequences, 27,968 unique sequences remained, 81% of which are 17-26 nt long, accounting for 93% of all redundant sequences. As described in Example 1, these sequences were further analyzed by aligning each sequence, using BLAST, to the human genome allowing a maximum of three nucleotides mismatched relative to the genome and a maximum insertion/deletion of three base pairs. This yielded ~723,000 genomic loci of mapped sequences; 83% of the unique sequences were mapped to the human genome using these criteria and 59% of the unique sequences were mapped with a maximum of one nucleotide mismatch. Subsequently, ~565,000 clusters of sequences with position overlap were created.

When the mapped sequenced reads were mapped to known miRNAs from Sanger miRBase registry (release 10), according to genomic position overlap, inclusion of sequenced reads in mature miRNA sequences or inclusion of mature miRNA sequences in the sequenced reads, the small RNA libraries were found to be enriched with human miRNAs. Known miRNAs occupied 61% (140,255 of 230,740) of the total small RNA reads. Three hundred and eighty-seven out of 885 (44%) human miRNAs were sequenced in at least one read in the different tumor libraries.

Most miRNAs were sequenced in several sequence variants that were previously referred to as isomiRs. The different isomiRs were predominantly variable in the 3' end of the mature miRNA sequence, a region which is less precisely defined than the miRNA 5' end. For numerous known miRNAs the most abundant isomiR in the cancer tissue survey was much more abundant (at least 20%) than the reference miRNA sequence from miRBase database. This suggests that the relative abundance of isomiRs may be inherently different between normal tissue and tumors. Additionally, several known miRNAs had an abundant isomiR with at least one mismatch to the human genome sequence, suggestive of the discovery of novel miRNA-related SNPs/cancer mutations or post-transcriptional modification of the miRNAs. All the isomiRs were expressed in at least the same number of reads as the miRBase isomiR. Most of the sequence modifications (69%) occurred in the 3' end of the miRNA and involved either DNA base modification, 3' uridylation or 3' adenylation. 3' additions of G or C were completely absent. The high abundance and the specificity of the 3' terminal single nucleotide insertions suggest that these are regulated post-transcriptional modifications and not DNA-level changes (SNPs/mutations), which are expected to occur in a more random manner. Several sequence modifications that occur internally within the miRNA sequence were also noted. These isomiRs demonstrated primarily (77%) C->T or A->G nucleotide modifications, again suggesting involvement of post-transcriptional RNA editing by cytidine deaminase or ADAR enzymes, respectively, contrary to DNA level changes.

Example 4.2: Deep sequence identification of novel miRNAs and miRNA-like small RNAs

Example 4.2.1: miRNAs derived from known miRNA precursors

One group of novel miRNAs and miRNA-like small RNAs expressed in the tested cancer tissues, includes miRNAs derived from known miRNA precursors, SEQ ID NOS: 472-479, mostly miRNA star sequences of known human miRNAs. miRNA star sequences are ~22-nt RNA species nearly complementary to a known miRNA, which are located within the miRNA precursor and which may have an inhibitory activity.

In several cases the novel complementary mature miRNA was more abundant than the known miRNA, suggesting that the identified miRNA is the major active product of the miRNA precursor, at least in the tested tumor samples. In addition several

cases of miRNA-offset RNAs (MORs), a miRNA-like group that was recently characterized {Shi, W., et al. (2009) *Nat Struct Mol Biol*, **16**, 183-189} were identified. MORs are part of the miRNA precursor and are processed from a ~22 bp dsRNA region directly upstream to the miRNA- miRNA star dsRNA region. All MORs sequenced in the human tumors are highly conserved, derived exclusively from the 5' stem of the miRNA precursor directly upstream to the 5' miRNA, and lowly expressed relative to the main miRNA product of the precursor. The MORs identified here tend to be located in a region of lower dsRNA stability than the main miRNA- miRNA star pair of the miRNA precursor. Therefore, the miRNA precursor of a MOR may switch between different folded RNA structures, only part of which accommodates the MOR in a dsRNA region that would be processed by the canonical miRNA pathway. This may explain the relatively low expression of MORs in comparison to the main mature miRNAs of the precursors.

Example 4.2.2: miRNAs derived from novel miRNA precursors

An additional group of novel miRNAs and miRNA-like small RNAs expressed in the tested cancer tissues includes completely novel miRNAs from novel miRNA precursors. Only reads that were exactly mapped to the genome were used. Reads that were mapped to more than 10 loci were filtered out, since human miRNAs rarely map to more than a few genomic loci. Other reasons for which sequences were discarded include rare occurrence (*i.e.* very few reads), length exceeding normal miRNA length and %GC higher than the %GC of known miRNAs. After filtering out by these criteria, as well as filtering sequences located within already annotated sequences (known miRNAs, other small RNAs, transposons, coding exons), miRNA precursors were predicted by folding several hundred bp flanking the final miRNA candidates using RNAfold {Denman, R.B. (1993) *Biotechniques*, 15, 1090-1095}. In order to reduce the number of false positive predictions, only predicted miRNA precursors that were either evolutionarily conserved or had structural features of known miRNAs were kept. Such structural features include limited length of bulges and loops and low folding energy. A miRNA precursor score was computed by integrating these parameters {Bentwich, I., *et al.* (2005) *Nat Genet*, **37**, 766-770}. This process resulted in the identification of small RNAs SEQ ID NOS.480-494 and precursors SEQ ID NOS.496-507, 509-514.

Example 4.2.3: miRNA-like sequences derived from annotated small RNAs and genomic repeats

Another group of novel miRNAs and miRNA-like small RNAs, expressed in the tested cancer tissues, contained miRNA-like sequences derived from annotated small RNAs and genomic repeats. Several miRNAs were previously described as having been derived from such genetic elements. Sequences whose length exceeded the conventional size of miRNAs (17-25 bp) were discarded. MiRNA precursors were predicted using RNAfold and mFold and the precursor score described above. Finally, only sequences with at least 10 reads were taken, in order to ensure that the identified novel miRNAs were likely to be consistent products of enzymatic excision and not rare degradation products. This strict criterion was used for this group only as these derive from known RNA species that are often highly expressed and their degradation products are expected to be found in the cell, therefore their re-annotation as miRNAs needs stronger evidence. This process revealed 3 miRNA-like sequences, SEQ ID NOS: 92, 93 and 495.

One of the candidates in this group, MID-24078 (SEQ ID NO: 495), is derived from a local hairpin-fold of an *Alu* repeat. The other two [MID-19433 (SEQ ID NO: 92), MID-19434 (SEQ ID NO: 93)] are, interestingly, derived from Y RNAs. Y RNAs are relatively unexplored noncoding RNA species that are implicated in chromosomal DNA replication {Krude, T., *et al*, *J Cell Sci*, **122**, 2836-2845} and RNA quality control {Sim, *et al*, (2009) *Mol Biol Cell*, **20**, 1555-1564}. Y RNA have been shown to be over-expressed in solid tumors (Christov *et al.*, Br J Cancer 2008;98(5):981-988), and thus may have potential for the diagnosis of cancer.

MID-19434 (SEQ ID NO: 93) is a 25-nt long RNA derived from a ~100 nucleotide-long hY3 RNA-like sequence. This sequence was highly expressed, with 200 sequenced reads, which is more abundant than over 300 known miRNAs sequenced in the analyzed tumor samples. The predicted well-folded precursor of this miRNA (SEQ ID NOS: 235-242) is precisely aligned to the hY3 RNA (genebank number NR_004392.1), suggesting that the Y RNA is processed, possibly by Dicer, to yield a 25 bp mature miRNA. MID-19433 (SEQ ID NO: 92) is derived from hairpin-folded hY1 Y RNAs (genebank number NR_004391.1).

Example 4.2.4: human siRNA

Endogenous siRNA are ~21-bp-long RNA species that are processed from a dsRNA by Dicer and assembled in the RNA induced silencing complex (RISC). These were recently described in mouse oocytes {Watanabe, T., *et al.* (2008) *Nature*, **453**, 539-543}, but have not yet been identified in the human transcriptome. The identified candidate human endogenous siRNA is a ~20-nt dsRNA that could be derived from bi-directional transcription of the same locus. Six sequenced reads (SEQ ID NOs: 450-454) are transcribed in the same orientation as a mitochondrial tRNA as well as a tRNA-derived pseudogene in one of the chromosomes, which is the more likely source of the siRNA sequences. Their transcription starts in the transcription start site of the tRNA, suggesting that these sequences are processed from the tRNA transcripts. Two antisense reads create a dsRNA with a short 5' overhang, as opposed to common siRNA which are characterized by a 3' overhang.The sense and antisense reads are mapped to nine different genomic loci. Therefore, it is also possible that the complementary sequences were derived from independent single-stranded RNAs and not from a hybridized dsRNA. This is thought to be the first time that endogenous siRNAs have been identified in the human genome.

Example 5: RT-PCR validation

As indicated in Figure 4, several of the newly identified miRNAs: MID-19433 (SEQ ID NO: 92), MID-19434 (SEQ ID NO: 93) and MID-16489 (SEQ ID NO: 31) were found to be expressed in serum of healthy people, using a third platform, RT-PCR.

The two novel small RNAs that are most abundantly expressed in different tumors in all platforms (high throughput sequencing, microarray, and RT-PCR), MID-19433 (SEQ ID NO: 92) and MID-19434 (SEQ ID NO: 93) are derived from small cytoplasmic Y RNA. The potential importance of these two novel miRNA-like sequences in tumorgenesis is supported by a recent work reporting that the Y RNAs hY1 and hY3, that are the unprocessed Y RNAs of MID-19433 and MID-19434, respectively, are overexpressed in carcinomas of the bladder, cervix, colon, kidney, lung and prostate {Christov, C.P., et al, (2008) Br J Cancer, 98, 981-988}. Therefore, measuring the highly expressed small RNAs from these Y RNAs can be used for molecular diagnostics of these cancers. The fact that these were also detected in serum confers them potential usage in non-invasive assays.

The present invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are intended to fall within the scope of the appended claims.

All patents, applications, publications, test methods, literature, and other materials cited herein are hereby incorporated by reference in their entirety as if physically present in this specification.

CLAIMS:

1. An isolated nucleic acid comprising a sequence selected from the group consisting of:
   (a) SEQ ID NOS: 31, 92-93, 450-454, 1, 2, 4-7, 9-26, 28-30, 32-35, 37-43, 46-72, 74-81, 83-91, 95-102, 104-108, 110-114, 116-157, 159-175 and 472-495;
   (b) a DNA encoding (a);
   (c) the complementary sequence of any one of (a) and (b); and
   (d) a sequence at least 80% identical to (a) - (c),
   wherein said nucleic acid is 16-26 nucleotides in length.
2. The nucleic acid of claim 1, wherein said sequence in part (d) has an identity of at least 90% to the sequence in parts (a) - (c).
3. An isolated nucleic acid comprising a sequence selected from the group consisting of:
   (a) SEQ ID NOS: 192, 232-242, 176-187, 189-191, 193-196, 198-231, 243-267, 269-326, 328-330, 332-340, 342, 343, 345-350, 352-361, 363-369, 371-387, 389-413, 415-432, 434-438, 440-449 and 496-514;
   (b) a DNA encoding (a);
   (c) the complementary sequence of any one of (a) and (b); and
   (d) a sequence at least 80% identical to (a) - (c),
   wherein said nucleic acid is 50-150 nucleotides in length.
4. The nucleic acid of claim 3, wherein said sequence in part (d) has an identity of at least 90% to the sequence in parts (a) - (c).
5. An isolated nucleic acid comprising an endogenous human siRNA.
6. The isolated nucleic acid of claim 5, wherein said nucleic acid comprises a sequence selected from the group consisting of:
   (a) SEQ ID NOS: 450-454;
   (b) a DNA encoding (a);
   (c) the complementary sequence of any one of (a) and (b); and
   (d) a sequence at least 80% identical to (a) - (c),
   wherein said nucleic acid is 16-26 nucleotides in length.

7. The nucleic acid of claim 6, wherein said sequence in part (d) has an identity of at least 90% to the sequence in parts (a) - (c).

8. An isolated nucleic acid of claim 1, wherein said nucleic acid comprises a sequence selected from the group consisting of:
   (a) SEQ ID NOS: 53 and 162;
   (b) SEQ ID NOS: 70 and 110;
   (c) SEQ ID NOS: 14 and 120;
   (d) SEQ ID NOS: 63, 106 and 58; and
   (e) SEQ ID NOS: 135 and 159.

9. The isolated nucleic acid of any of claims 1, 3 and 6, wherein said nucleic acid is a modified oligonucleotide.

10. A composition comprising an isolated nucleic acid of any of claims 1, 3 and 6.

11. The composition of claim 10, wherein said composition is suitable for diagnostic applications.

12. The composition of claim 10, wherein said composition is suitable for therapeutic applications.

13. The composition of claim 12, further comprising a pharmaceutically acceptable carrier.

14. The composition of claim 10, as a marker or modulator of cancer.

15. A recombinant expression vector comprising an isolated nucleic acid of any of claims 1, 3 and 6.

16. A probe comprising an isolated nucleic acid of any of claims 1, 3 and 6.

17. A biochip comprising the probe of claim 16.

18. A host cell comprising an isolated nucleic acid of any of claims 1, 3 and 6.

19. A method for diagnosing a cancer in a subject comprising:
   (a) obtaining a biological sample from said subject;
   (b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 31, 4, 7, 11, 16, 17, 21, 22, 23, 26, 30, 33-35, 37, 39, 46, 47- 49, 51, 52, 53, 56, 58, 60, 63, 64, 66-68, 71, 72, 74, 76-78, 83, 86-88, 90, 96, 98, 100, 101, 106, 110-114,

116, 117, 119-121, 127, 129, 130, 132-136, 138, 141, 144, 145, 147, 148, 152, 153, 157, 159-162, 165, 167, 169-174, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression levels of any of said nucleic acids in healthy controls,

wherein the comparison of said expression profile to said reference expression profile allows for diagnosis of said cancer.

20. The method of claim 19, wherein said cancer is selected from the group consisting of colon, bladder, breast, lung, liver, kidney, ovarian, prostate, esophagus, cervix, and pancreatic cancer.

21. The method of claim 19, wherein relatively high expression levels of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 31, 4, 7, 11, 17, 21, 22, 23, 26, 30, 33, 35, 37, 39, 46, 48, 49, 51, 52, 53, 56, 58, 60, 63, 64, 66, 68, 71, 72, 74, 76-78, 86-88, 90, 96, 98, 100, 101, 106, 110-114, 116, 117, 119-121, 127, 129, 130, 132-136, 138, 141, 144, 145, 147, 148, 152, 153, 157, 159-162, 165, 167 and 169-174, as compared to said reference expression profile, is indicative of cancer.

22. The method of claim 19, wherein relatively low expression levels of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 16, 34, 47 and 83, as compared to said reference expression profile, is indicative of cancer.

23. A method of diagnosing an increased risk of colon cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 21, 68, 111 and 174, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression levels of any of said nucleic acids healthy controls,

wherein a high expression level of said nucleic acid sequence is indicative of an increased risk of colon cancer in a subject.

24. A method of diagnosing an increased risk of colon cancer in a subject comprising:
    (a) obtaining a biological sample from said subject;
    (b) determining an expression profile of SEQ ID NO: 92, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and
    (c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

    wherein a low expression level of said nucleic acid sequence is indicative of an increased risk of colon cancer in a subject.

25. A method of diagnosing colon cancer in a subject comprising:
    (a) obtaining a biological sample from said subject;
    (b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 2, 15, 28, 29, 48, 58, 61, 70, 86, 97, 107, 110, 150, 156, 170 and 172, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and
    (c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

    wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of colon cancer.

26. A method of diagnosing lung cancer in a subject a comprising:
    (a) obtaining a biological sample from said subject;
    (b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 1, 7, 19, 21, 22, 26, 30, 35, 37, 39, 43, 46, 51, 53, 58, 59, 63, 64, 68, 71, 72, 74, 78, 86, 87, 106, 110, 113, 116, 119, 125, 127, 132, 135, 141, 153, 159, 161, 164, 165

and 170-173, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of lung cancer.

27. A method of diagnosing bladder cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 7, 26, 35, 37, 39, 53, 71, 72, 125, 127, 130, 132, 135, 159, 161, 165, 170 and 172, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of bladder cancer.

28. A method of diagnosing bladder cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 34 and 83, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively low expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of bladder cancer.

29. A method of diagnosing liver cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 11, 26, 71, 72, 77, 98, 134, 136, 153, 160, 170 and 171, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of liver cancer.

30. A method of diagnosing liver cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 12, 14, 28-32, 34, 42, 43, 52-55, 67, 75, 84, 85, 89, 97,105, 123-126, 128, 129, 137 and 140 a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively low expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of liver cancer.

31. A method of diagnosing an endometrial metastasis in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 22, 25, 29, 31, 37, 39, 53, 64, 68, 72, 76, 77, 78, 84, 113, 119, 121, 127, 130, 132, 133, 136, 153, 161, 170 and 171 a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of endometrial metastasis.

32. A method of diagnosing an endometrial metastasis in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 16 and 57 a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively low expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of endometrial metastasis.

33. A method of diagnosing kidney cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 26, 37, 39, 53, 71, 72, 98, 125, 127, 130, 135, 159 and 165, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of kidney cancer.

34. A method of diagnosing kidney cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 16, 34, 83 and 140, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively low expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of kidney cancer.

35. A method of diagnosing breast cancer in a subject comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 79, 99, 122, 130, 153 and 154 a fragment thereof, and a sequence having at least about 80% identity thereto from said sample; and

(c) comparing said expression profile to a reference expression profile representing the expression level of said nucleic acid in healthy controls,

wherein relatively low expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of breast cancer.

36. A method to distinguish between a primary lung tumor and a metastasis to the lung, said method comprising:

(a) obtaining a biological sample from said subject;

(b) determining an expression profile of a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 31, 92, 1, 2, 6, 7, 10, 11, 13, 17, 19-24, 28-30, 33, 40, 42, 43, 46, 49, 51, 56-59, 61, 63, 64, 68, 70, 71, 74, 75, 77, 78, 80, 81, 86-88, 90-91, 95, 96, 97, 100, 104, 106, 107, 110-113, 116, 117, 119, 122,129, 133, 138, 139, 141, 142, 144, 145-152, 154, 162, 164, 165, 169, 171-173, 175, a fragment thereof, and a sequence having at least about 80% identity thereto from said sample;

(c) comparing said expression profile to a reference expression profile, wherein relatively high expression levels of any of said nucleic acid sequences, as compared to said reference expression profile, is indicative of a primary lung tumor.

37. The method of claim 36, wherein the origin of said metastasis to the lung is selected from the group consisting of endometrium, kidney, larynx, melanocyte and salivary gland.

38. The method of claim 19, wherein said subject is a human.

39. The method of claim 19, wherein said method is used to determine a course of treatment for said subject.

40. The method of claim 19, wherein said biological sample is selected from the group consisting of bodily fluid, a cell line and a tissue sample.

41. The method of claim 40, wherein said bodily fluid is selected from the group consisting of whole blood and serum.

42. The method of claim 40, wherein said tissue is a fresh, frozen, fixed, wax-embedded or formalin fixed paraffin-embedded (FFPE) tissue.

43. The method of claim 19, wherein said expression levels are determined by a method selected from the group consisting of nucleic acid hybridization, nucleic acid amplification, and a combination thereof.

44. The method of claim 43, wherein said nucleic acid hybridization is performed using a solid-phase nucleic acid biochip array or in situ hybridization.

45. The method of claim 43, wherein said nucleic acid amplification method is real-time PCR.

46. The method of claim 45, wherein said real-time PCR method comprises forward and reverse primers.

47. The method of claim 45, wherein said forward primer comprises a sequence selected from the group consisting of SEQ ID NOS: 466, 467, 469, a fragment thereof and a sequence at least about 80% identical thereto.

48. The method of claim 45, wherein said real-time PCR method further comprises a probe.

49. The method of claim 48, wherein the probe comprises a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 31, 4, 7, 11, 16, 17, 21, 22, 23, 26, 30, 33-35, 37, 39, 46, 47- 49, 51, 52, 53, 56, 58, 60, 63, 64, 66, 68, 71, 72, 74, 76-78, 83, 86-88, 90, 96, 98, 100, 101, 106, 110-114, 116, 117, 119-121, 127, 129, 130, 132-136, 138, 141, 144, 145, 147, 148, 152, 153, 157, 159-162, 165, 167, 169-174, a fragment thereof and a sequence at least about 80% identical thereto.

50. A kit for diagnosing a cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NO: 31, 4, 7, 11, 16, 17, 21, 22, 23, 26, 30, 33-35, 37, 39, 46, 47- 49, 51, 52, 53, 56, 58, 60, 63, 64, 66, 68, 71, 72, 74, 76-78, 83, 86-88, 90, 96, 98, 100, 101; 106, 110-114, 116, 117, 119-121, 127, 129, 130, 132-136, 138, 141, 144, 145, 147, 148, 152, 153, 157, 159-162, 165, 167, 169-174, a fragment thereof and a sequence at least about 80% identical thereto.

51. A kit for diagnosing an increased risk of colon cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 21, 68, 92, 111 and 174, a fragment thereof and a sequence at least about 80% identical thereto.

52. A kit for diagnosing colon cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 2, 15, 28, 29, 48, 58, 61, 70, 86, 97, 107, 110, 150, 156, 170 and 172, a fragment thereof and a sequence at least about 80% identical thereto.

53. A kit for diagnosing lung cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 1, 7, 19, 21, 22, 26, 30, 35, 37, 39, 43, 46, 51, 53, 58, 59, 63, 64, 68, 71, 72, 74, 78, 86, 87, 106, 110,

113, 116, 119, 125, 127, 132, 135, 141, 153, 159, 161, 164, 165 and 170-173, a fragment thereof and a sequence at least about 80% identical thereto.

54. A kit for diagnosing bladder cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 7, 26, 34, 35, 37, 39, 53, 71, 72, 83, 125, 127, 130, 132, 135, 159, 161, 165, 170 and 172, a fragment thereof and a sequence at least about 80% identical thereto.

55. A kit for diagnosing liver cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 11, 12, 14, 26, 28-32, 34, 42, 43, 52-55, 67, 71, 72, 75, 77, 84, 85, 89, 97, 89, 105, 123-126, 128, 129, 134, 136, 137, 140, 153, 160, 170 and 171, a fragment thereof and a sequence at least about 80% identical thereto.

56. A kit for diagnosing a subject with endometrial metastasis, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 16, 22, 25, 29, 31, 37, 39, 53, 57, 64, 68, 72, 76, 77, 78, 84, 113, 119, 121, 127, 130, 132, 133, 136, 153, 161, 170 and 171, a fragment thereof and a sequence at least about 80% identical thereto.

57. A kit for diagnosing kidney cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 16, 26, 34, 37, 39, 53, 71, 72, 83, 98, 125, 127, 130, 135, 140, 159 and 165, a fragment thereof and a sequence at least about 80% identical thereto.

58. A kit for diagnosing breast cancer in a subject, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID NOS: 79, 99, 122, 130, 153 and 154 a fragment thereof and a sequence at least about 80% identical thereto.

59. A kit for distinguishing between a primary lung tumor and a metastasis to the lung, said kit comprising a probe comprising a nucleic acid sequence that is complementary to a sequence selected from the group consisting of SEQ ID

NOS: 31, 92, 1, 2, 6, 7, 10, 11, 13, 17, 19-24, 28-30, 33, 40, 42, 43, 46, 49, 51, 56-59, 61, 63, 64, 68, 70, 71, 74, 75, 77, 78, 80, 81, 86-88, 90-91, 95, 96, 97, 100, 104, 106, 107, 110-113, 116, 117, 119, 122,129, 133, 138, 139, 141, 142, 144, 145-152, 154, 162, 164, 165, 169, 171-173, 175, a fragment thereof and a sequence at least about 80% identical thereto.

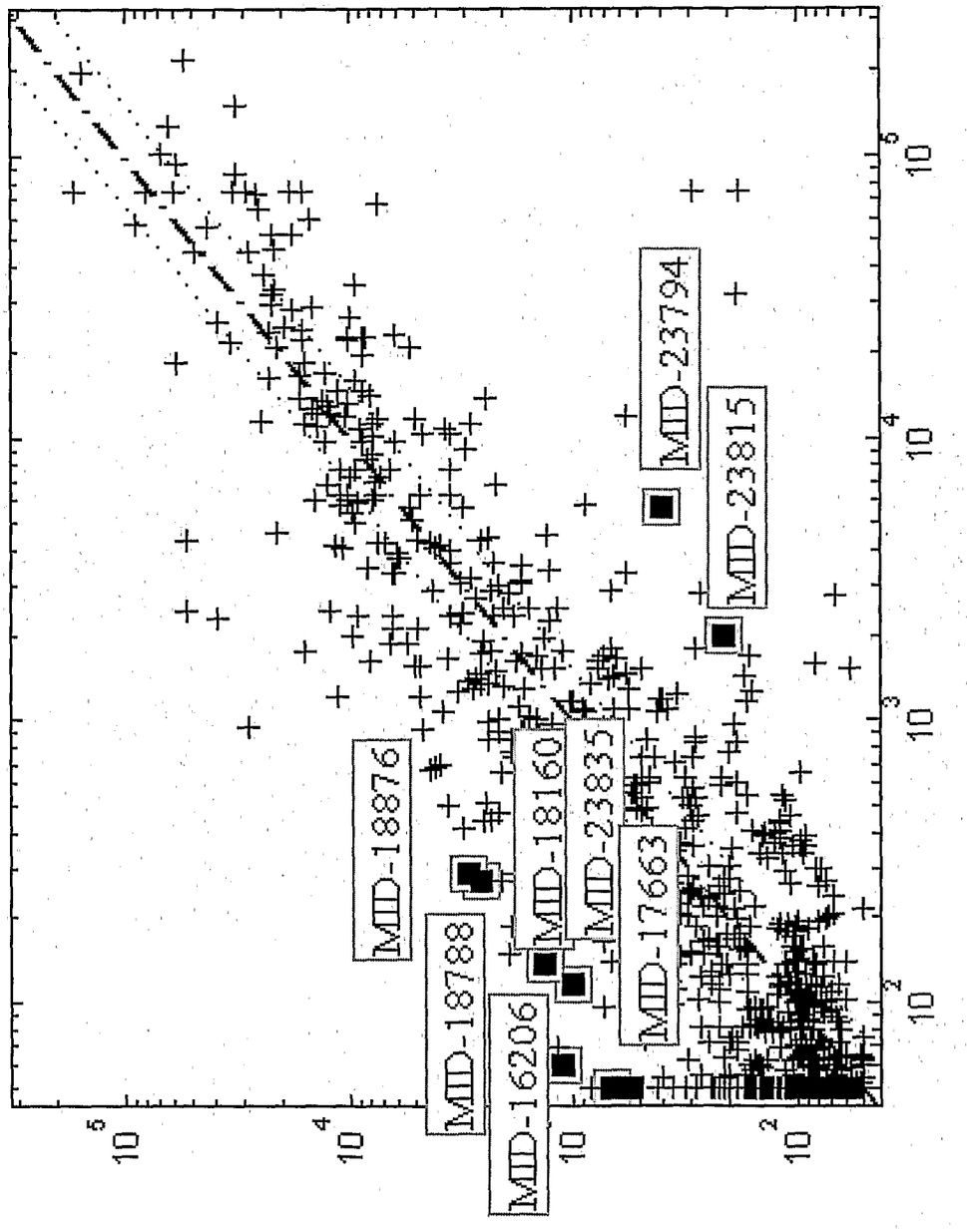

MID-18876
MID-18788
MID-16206
MID-18160
MID-23835
MID-17663
MID-23794
MID-23815
10
10
10
10
10
10
10
10
2
3
4
5

Figure 1

Figure 2

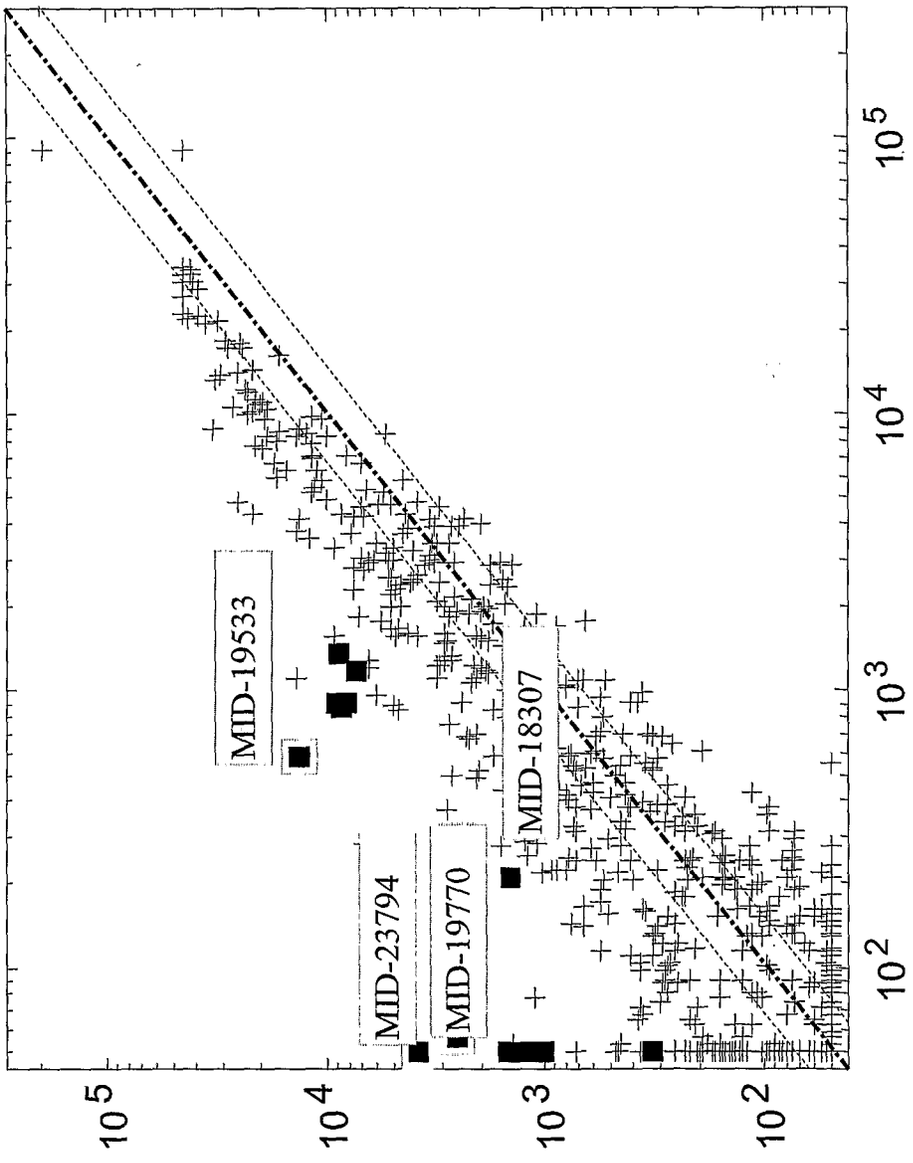

MID-19533
MID-23794
MID-19770
MID-18307
10^5
10^4
10^3
10^2
10^5
10^4
10^3
10^2

Figure 3

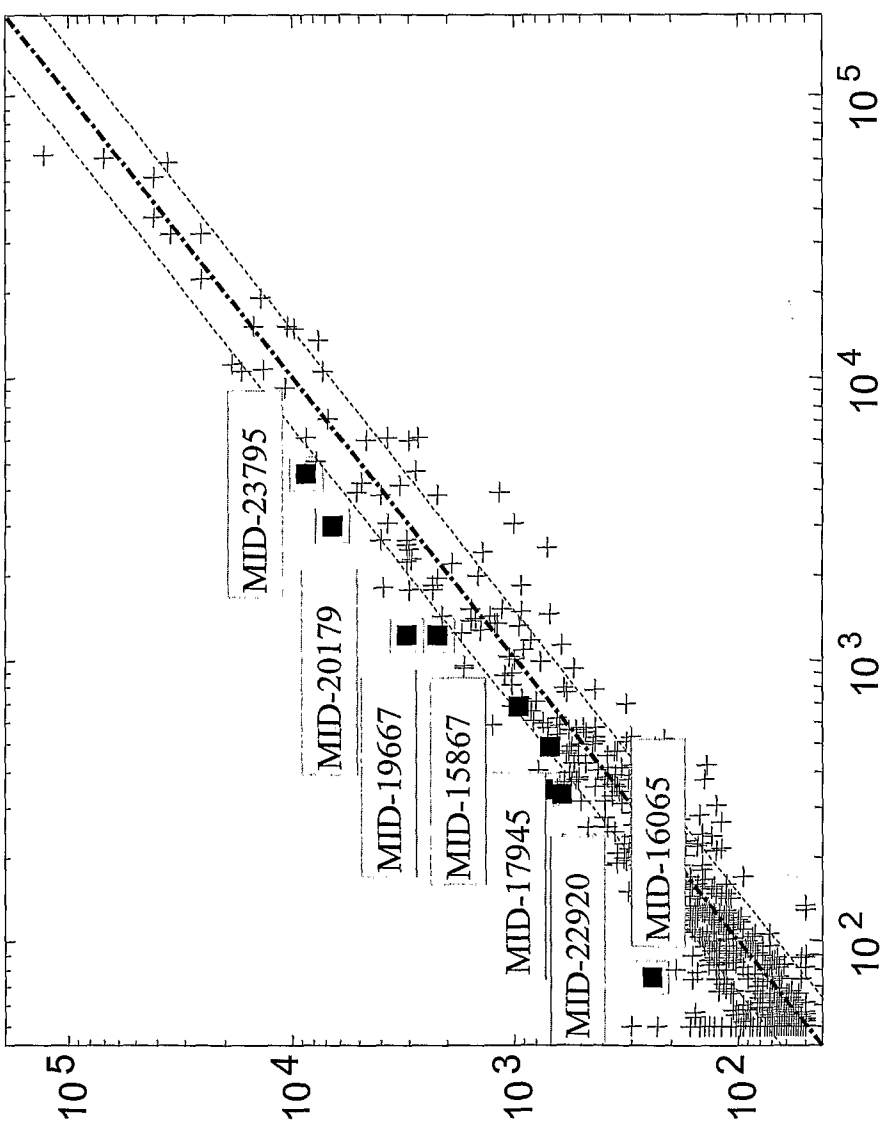
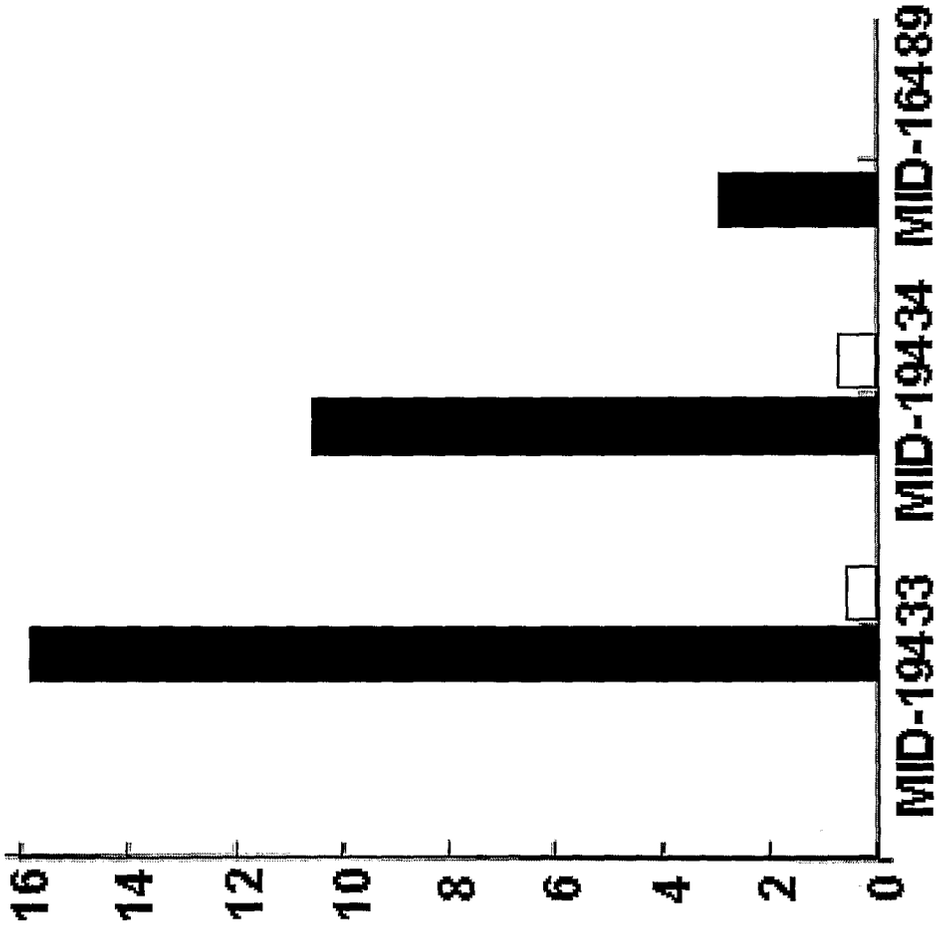

MID-23795
MID-20179
MID-19667
MID-15867
MID-17945
MID-22920
MID-16065
10^5
10^4
10^3
10^2
10^5
10^4
10^3
10^2

Figure 4



INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL 10/00462

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C07H 21/00; A61K 31/7088; G01N 33/48 (2010.01)
USPC - 536/23.1; 514/44R; 436/64
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8): C07H 21/00; A61K 31/7088; G01N 33/48 (2010.01)
USPC: 536/23.1; 514/44R; 436/64

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 536/23.1; 514/44R; 436/64; 436/94

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WEST (PGPB,USPT,EPAB,JPAB): cancer, Rosetta Genomics, therapeutic, diagnostic, expression profile, real time PCR, biochip, probe, modulator, RP11-430C7, MDM4

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	GenBank AL512306.16. Human DNA sequence from clone RP11-430C7 on chromosome 1 Contains the MDM4 gene for Mdm4 (transformed 3T3 cell double minute 4) p53 binding protein (mouse), a PEF protein with a long N-terminal hydrophobic domain (peflin) (PEF) pseudogene, two novel genes and the 3' end of the LRRN5 gene for leucine rich repeat neuronal 5. 13-JAN-2009. [Retrieved from the Internet 14 December 2010: <http://www.ncbi.nlm.nih.gov/nuccore/18491332>].	1-4, 9-21, and 38-46, 48-50
Y	DANOVI et al. Amplification of Mdmx (or Mdm4) Directly Contributes to Tumor Formation by Inhibiting p53 Tumor Suppressor Activity. Molecular and Cellular Biology, 2004, Vol 24, Pages 5835-5843, especially pg 5835, abstract; pg 5836, left column, seventh para, pg 5842, Fig 6.	1-4, 9-21, 38-46, and 48-50
Y	RIEMENSCHNEIDER et al. Amplification and Overexpression of the MDM4 (MDMX) Gene from 1q32 in a Subset of Malignant Gliomas without TP53 Mutation or MDM2 Amplification. Cancer Research. 1999. Vol 59, Pages 6091-6096, especially page 6091, abstract; page 6093, left col, second paragraph; page 6095, right col., fourth paragraph.	3, 4
Y	WO 2007/148235 A2 (AHARANOV et al.) 27 December 2007 (27.12.2007), para [0005], [0008], [0009], [0042], [0048], [0053], [0056], [0062], [0066], [0089], [0091], [0093], [0099], [0100], [0101], [0108], [0109], [0112], [0116], [0119], [0121], [0122], [0123], [0147], [0157].	9-21, 38-46, 48-50

☒ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 December 2010 (15.12.2010)	06 JAN 2011
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL 10/00462

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2002/0009738 A1 (HOUGHTON et al.) 24 January 2002 (24.01.2002), abstract; para [0012], [0203].	45, 46, 48, 49
A	US 2003/0130485 A1 (MEYERS et al.) 10 July 2003 (10.07.2003), para [0028], [0040], [0313], [0352], [0410], [0855]-[0864], [0875], [1009], [1047],[1048].	1-4, 9-21, and 38-46, 48-50

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL 10/00462

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: claims 1-4, 9-22 and 38-50 directed to an isolated nucleic acid sequence, wherein the sequence may be used in a method of diagnosing cancer comprising evaluating the expression of said sequence, and comparing the expression to the level in a healthy control, selected from the group consisting of SEQ ID NOs: 1, 2, 4-7, 9-26, 28-35, 37-43, 46-72, 74-81, 83-93, 95-102, 104-108, 110-114, 116-157, 159-175, 450-454 and 472-495, or the complement thereof, or a sequence having at least 80% sequence identity with any of the above, wherein the nucleic acid is 16-26 nucleotides in length or 50-150 nucleotides in length, limited to SEQ ID NOs: 31 and 192 (the first sequences indicated in claims 1 and 3, wherein 31 is a subsequence of 192).
Group II: claims 1 and 8-18, directed to an isolated nucleic acid SEQ ID NOs, selected from the groups consisting of: a) 53 and 162; b) 70 and 110; c) 14 and 120; d) 63, 106 and 58 or 135 and 159; limited to 53 and 162.

- Please see extra sheet for continuation -

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-4, 9-22 and 38-50; limited to SEQ ID NO: 31 and 192.

Note: claims 22 and 47 are withdrawn for being drawn to non-elected subject matter.

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL 10/00462

Continuation of Box III: Lack of Unity of Invention

Group III: claims 5-7 and 9-18, directed to an endogenous siRNA, further comprising a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 450-454.
Group IV: claims 23 and 51, directed to a method of diagnosing an increased risk of colon cancer in a subject, comprising obtaining a biological sample from a subject, determining the expression of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 21, 68, 11 and 174; and comparing the expression in the sample to the level of expression in a control sample, wherein a high expression level indicates an increased risk of colon cancer; limited to SEQ ID NO: 21.
Group V: claims 24 and 51, directed to a method of diagnosing an increased risk of colon cancer in a subject, comprising: obtaining a biological sample from a subject; determining the expression profile of a nucleic acid having SEQ ID NO: 92 in the sample and comparing said expression profile to a control profile, wherein a decreased amount of expression relative to control is indicative of an increased risk of colon cancer.
Group VI: claims 25 and 52, directed to a method of diagnosing colon cancer in a subject, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 2, 15, 28, 29, 48, 58, 61, 70, 86, 97, 107, 110, 150, 156, 170 and 172; wherein an increased expression level relative to control is indicative of colon cancer; limited to SEQ ID NO: 2.
Group VII: claims 26 and 53, directed to a method of diagnosing lung cancer in a subject, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 1, 7, 19, 1, 7, 19, 21, 22, 26, 30, 35, 37, 39, 43, 46, 51, 53, 58, 59, 63, 64, 68, 71, 72, 74, 78, 86, 87, 106, 110, 113, 116, 119, 125, 127, 132, 135, 141, 153, 159, 161, 164, 165 and 170-172; wherein an increased expression level relative to control is indicative of lung cancer; limited to SEQ ID NO: 1.
Group VIII: claims 27 and 54, directed to a method of diagnosing bladder cancer in a subject, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 7, 26, 35, 37, 39, 53, 71, 72, 125, 127, 130, 132, 135, 159, 161, 165, 170 and 172, wherein an increased expression level relative to control is indicative of bladder cancer; limited to SEQ ID NO: 7.
Group IX: claims 28 and 54, directed to a method of diagnosing bladder cancer in a subject, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 34 and 83, wherein a decreased level of expression relative to a control is indicative of bladder cancer; limited to SEQ ID NO: 34.
Group X: claims 29 and 55, directed to a method of diagnosing liver cancer in a subject, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 11, 26, 71, 72, 77, 98, 134, 136, 153, 160, 170 and 171, wherein an increased expression level relative to control is indicative of liver cancer; limited to SEQ ID NO: 11.
Group XI: claims 30 and 55, directed to a method of diagnosing liver cancer in a subject, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 12, 14, 28-32, 34, 42, 43, 52-55, 67, 75, 84, 85, 89, 97,105, 123-126, 128, 129, 137 and 140, wherein a decreased level of expression relative to a control is indicative of bladder cancer; limited to SEQ ID NO: 12.
Group XII: claims 31 and 56, directed to a method of diagnosing an endometrial metastasis in a subject, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 22, 25, 29, 31, 37, 39, 53, 64, 68, 72, 76, 77, 78, 84, 113, 119, 121, 127, 130, 132, 133, 136, 153, 161, 170 and 171, wherein an increased expression level relative to control is indicative of endometrial metastasis; limited to SEQ ID NO: 22.
Group XIII: claims 32 and 56, directed to a method of diagnosing an endometrial metastasis in a subject, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 16 and 57, wherein a decreased level of expression relative to a control is indicative of endometrial metastasis; limited to SEQ ID NO: 16.
Group XIV: claims 33 and 57, directed to a method of diagnosing kidney cancer in a subject, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 26, 37, 39, 53, 71, 72, 98, 125, 127, 130, 135, 159 and 165, wherein an increased expression level relative to control is indicative of kidney cancer; limited to SEQ ID NO: 26.
Group XV: claims 34 and 57, directed to a method of diagnosing kidney cancer in a subject, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 16, 34, 83 and 140, wherein a decreased level of expression relative to a control is indicative of kidney cancer; limited to SEQ ID NO: 16.
Group XVI: claims 35 and 58, directed to a method of diagnosing breast cancer in a subject, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 79, 99, 122, 130, 153 and 154, wherein a decreased level of expression relative to a control is indicative of breast cancer; limited to SEQ ID NO: 79.
Group XVII: claims 36, 37, and 59, directed to a method to distinguish between a primary lung tumor and a metastasis, comprising obtaining a biological sample from a subject, determining the expression profile of a nucleic acid comprising a sequence selected from the group consisting of SEQ ID NOs: 31, 92, 1, 2, 6, 7, 10, 11, 13, 17, 19-24, 28-30, 33, 40, 42, 43, 46, 49, 51, 56-59, 61, 63, 64, 68, 70, 71, 74, 75, 77, 78, 80, 81, 86-88, 90-91, 95, 96, 97, 100, 104, 106, 107, 110-113, 116, 117, 119, 122, 129, 133, 138, 139, 141, 142, 144, 145-152, 154, 162, 164, 165, 169, 171-173, and 175, wherein an increased expression level relative to control is indicative of a primary lung cancer; limited to SEQ ID NO: 31.

The inventions listed as Groups I - XVII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of the claims of Groups I-XVII is an isolated nucleic acid sequence, wherein the sequence may be used in a method of diagnosing cancer comprising evaluating the expression of said sequence, and comparing the expression to the level in a healthy control, and wherein the nucleic acid may be from 15 nucleotides to 150 nucleotides in length, wherein each Group is directed either to specific sequences, or diagnosis of specific cancers based upon particular changes in expression of specific nucleic acid sequences.

- Please see next extra sheet for continuation -

Continuation of Box III: Lack of Unity of Invention (page 2)

The only common technical element shared by the above groups is that they are related to alterations of nucleic acid expression in cancers. This common technical element does not represent an improvement over the prior art of US 2003/0130485 A1 to Meyers et al. (see para [0028], [0040], [0313], [0352], [0410], [0855]-[0864], [0875], [1009], [1047]-[1048] and SEQ ID NO: 68 in comparison to Applicants' SEQ ID NO: 91.) Regarding Group II, the nucleic acids claimed in the groupings do not improve upon the prior art of WO 2000/026401 A1 (see abstract, Fig. 6a; claim 61, pg 12, ln 24- pg 13, ln 5; pg 16, ln 24-26; pg 22, ln 17-23). Still further in view of WO 2008/046911 A2 to Litman et al. (see abstract, pg. 2, ln 12-15 and SEQ ID NO: 380 in comparison to Applicants' SEQ ID NO: 450) (regarding endogenous human siRNA molecules). Still further in view (regarding diagnosis of bladder cancer via determination of expression of nucleic acid sequences) of US 2009/0186353 A1 to Aharonov et al. (see para [0005], [0016]). Further in view of (with respect to endometrial metastases) US 2008/0226554 A1 to Colgan et al. (see abstract, para [0001], [0017], [0018], [0020], [0031] and [0048]). Still further in view of US 2008/0139402 A1 to Pressman et al. (see abstract, para [0056], and [0180]). Therefore, the inventions of Group I and Group II lack unity of invention under PCT Rule 13 because they do not share a same or corresponding special technical feature.

Note: Applicants may elect for additional nucleic acid sequences to be searched in accordance with the above Groups by indicating the SEQ ID NO (or, in the case wherein one sequence is a subsequence of another, the related SEQ ID NO's) and by paying an additional invention search fee for each sequence (All groups for which a search has been paid that include each elected sequence will be searched in relation to that sequence).