

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 November 2011 (03.11.2011)

PCT

(10) International Publication Number
WO 2011/137288 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2011/034451

(22) International Filing Date:
29 April 2011 (29.04.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/329,945 30 April 2010 (30.04.2010) US

(71) Applicant (for all designated States except US): **THE OHIO STATE UNIVERSITY** [US/US]; 1216 Kinnear Road, Columbus, OH 43212 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **CROCE, Carlo, M.** [US/US]; 2140 Cambridge Blvd., Columbus, OH 43221 (US). **VOLINIA, Stefano** [IT/IT]; Corso Piave, 17, 44100 Ferrara (IT).

(74) Agent: **MARTINEAU, Catherine, B.**; MacMillan, Sobanski & Todd, LLC, One Maritime Plaza, 5th Floor, 720 Water Street, Toledo, OH 43604 (US).

(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, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

— with information concerning refusal to authorize the rectification of an obvious mistake under Rule 91.1 (Rule 91.3(d))

(54) Title: MIRNA NETWORKS IN CANCERS AND LEUKEMIAS AND USES THEREOF

(57) Abstract: Methods for identification of cancer variations in miRNA networks, including comparing normal to cancer networks by generating a comprehensive miRNA alteration map in cancer, and superimposing DNA variations onto expression data, databases containing the same and uses thereof are disclosed.



WO 2011/137288 A2

TITLE

miRNA Networks in Cancers and Leukemias and Uses Thereof

Inventors: Carlo M. Croce, Stefano Volinia

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of United States Provisional Application Number 61/329,945 filed April 30, 2010, the entire disclosure of which is expressly incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under -----, awarded by National Cancer Institute Grant. The government has certain rights in this invention.

TECHNICAL FIELD AND
INDUSTRIAL APPLICABILITY OF THE INVENTION

[0003] This invention relates generally to the field of molecular biology. More particularly, it concerns miRNA networks in cancers and leukemias and uses thereof.

[0004] Certain aspects of the invention include application in diagnostics, therapeutics, and prognostics of cancers and leukemias associated disorders.

BACKGROUND

[0005] There is no admission that the background art disclosed in this section legally constitutes prior art.

[0006] Much of the current effort in miRNA studies is focused on the elucidation of their function. Typically miRNAs have been studied by using the gene profiling approach. Each miRNA has been studied for its single contribution to differential expression or to a compact predictive signature. However, the effect of miRNAs on cell pathology and physiology is likely to be complex for two reasons: (1) their activity is exerted in a one-to-many fashion, such that each miRNA can control translation of tens or even hundreds of different coding messengers and (2) a single messenger can be controlled by more than one miRNA.

[0007] Thus, there is a need for a paradigm shift in the study of miRNAs in cancer.

[0008] There is also a need for miRNA gene networks that include expression data.

SUMMARY

[0009] The present invention is based, at least in part, on the following information and discoveries, as described herein.

- [0010] In a broad aspect, there is provided herein a method for conducting an analysis of miRNA profiles, comprising analyzing tissue specificity and cancer-type specificity, accessing an expression database to determine the presence or absence of coordinated microRNA (miRNA) activities.
- [0011] In another broad aspect, there is provided herein an miRNA network for different solid tumors and leukemias and uses thereof. In certain embodiments, the nonmalignant tissues and cancer networks display a change in hubs, the most connected miRNAs.
- [0012] In another broad aspect, there is provided herein a method for identifying one or more miRNA cliques in cancer, comprising using the miRNA network as described and claimed herein.
- [0013] In another broad aspect, there is provided herein a method for determining miRNAs with comprehensive roles in cancer, comprising combining one or more of: differential expression, genetic networks, and DNA copy number alterations.
- [0014] In another broad aspect, there is provided herein a use of a miRNA networks comprised from miRNA expression data that includes a least one miRNA network from normal tissues, and/or at least one miRNA network for coupled cancerous and noncancerous tissues.
- [0015] In another broad aspect, there is provided herein a method for identification of cancer variations in miRNA networks, comprising comparing normal miR networks to cancer miR networks.
- [0016] In another broad aspect, there is provided herein a method for generating a comprehensive miRNA alteration map in cancer, comprising superimposing DNA variations onto expression data.
- [0017] In a particular aspect, there is provided herein a method for diagnosing, adenocarcinoma in a subject by identifying a set of miRs in a sample from the subject, and determining whether the set of miRs fall within the normal miRNA network substantially as shown in Figure 4A, or within the cancer miRNA network substantially as shown in Figure 4B.
- [0018] In certain embodiments, at least one miRNA network comprising the miRNA network substantially as shown in Figure 4B, for lung adenocarcinoma.
- [0019] In certain embodiments, at least one miRNA network comprising the miRNA network substantially as shown in Figure 6, for acute myeloid leukemia (AML).
- [0020] In certain embodiments, at least one miRNA network comprising the miRNA network substantially as shown in Figure 7, for chronic lymphocytic leukemia (CLL).
- [0021] In certain embodiments, at least one miRNA network comprising the miRNA network substantially as shown in Figure 8, for leukemias.
- [0022] In certain embodiments, at least one miRNA network comprising the miRNA network substantially as shown in Figure 16, for colon adenocarcinomas.
- [0023] In certain embodiments, at least one miRNA network comprising the miRNA network

substantially as shown in Figure 17, for breast cancer.

[0024] In certain embodiments, at least one miRNA network comprising the miRNA network substantially as shown in Figure 18, for prostate cancer.

[0025] In another broad aspect, there is provided herein a method for diagnosing a subject suspected of having a solid cancer, comprising screening a sample from the subject for the set of differentially up-regulated miRNAs listed in Supplemental Table IV.

[0026] In another broad aspect, there is provided herein a method for diagnosing a subject suspected of having a solid cancer, comprising screening a sample from the subject for the set of differentially down-regulated miRNAs listed in Supplemental Table V.

[0027] In another broad aspect, there is provided herein a method for diagnosing a subject suspected of having a solid cancer, comprising screening a sample from the subject for one or more of the differentially regulated miRNAs listed in Supplemental Table VI.

[0028] In another broad aspect, there is provided herein a method for diagnosing a subject suspected of having a solid cancer, comprising screening a sample from the subject for one or more of the differentially regulated miRNAs listed in Supplemental Table VIII.

[0029] In another broad aspect, there is provided herein a method for diagnosing a subject suspected of having a solid cancer, comprising screening a sample from the subject for one or more of the differentially regulated miRNAs listed in Supplemental Table IX.

[0030] In another broad aspect, there is provided herein a use of the method as described herein, where a change in miRNA expression in a subject is indicative of the subject being predisposed to, or having, a solid cancer and/or hematological cancer.

[0031] In certain embodiments, the cancer is one or more of lung, breast, colon, prostate, pancreas, and nasopharyngeal carcinomas, glioblastoma, melanoma, Ewing sarcoma, osteosarcoma, T-cell acute lymphoblastic leukemia (T-ALL), AML, CLL, myelodysplasia, various lymphomas, and mucosa-associated lymphoid tissue (MALT).

[0032] In another broad aspect, there is provided herein a method for detecting cancer differentiation in a subject, comprising screening for at least one of hsa-miR-221 and hsa-miR-218.

[0033] In another broad aspect, there is provided herein a method for compiling and/or analyzing a network of biological functions in a subject, comprising: i) obtaining information on one or more miR present in the subject, wherein at least one miR reflects a biological function; and ii) subjecting the obtained information to analysis of coherent groups of nodes by using clustering algorithms to calculate a relationship between the functional reporters to generate a network relationship of the miRs to the biological function.

[0034] In certain embodiments, the subject is human.

[0035] Other systems, methods, features, and advantages of the present invention will be or will become apparent to one with skill in the art upon examination of the following drawings and

detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present invention, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE FIGURES

[0036] The patent or application file contains one or more drawings executed in color and/or one or more photographs. Copies of this patent or patent application publication with color drawing(s) and/or photograph(s) will be provided by the Patent Office upon request and payment of the necessary fee.

[0037] **Figure 1.** The miRNA network in normal tissues (1107 samples, 50 tissues, 115 miRNAs). The network was inferred for all expressed and varying miRNAs, without preselecting for differential expression. Standard Banjo parameters were adopted with a q6 discretization policy. The consensus graph depicted here was obtained from the best 100 nets (after searching through 8.3 3 10⁹ networks). MCL graph-based clustering algorithm was applied to clusters extraction (miRNAs with highly related expression pattern have edges of the same color); thus, different clusters in the network are linked by different color edges. yEd graph editor (yFiles software) was employed for graphs visualization.

[0038] **Figure 2.** Cellular pathways regulated by differentially expressed miRNAs in cancer. KEGG analysis by ClueGO (Bindea et al. 2009) of pathways (score = 3, P-value < 1 3 10⁻³) simultaneously targeted by both up-regulated and down-regulated miRNAs (listed in **Table 1; Supplemental Tables IV, V**). The KEGG pie-chart shows the functional effect of differentially expressed miRNAs on cellular pathways in cancer. The large majority of the affected pathways is related to cancer or signal transduction (i.e., Wnt, VEGF, TGF-beta, insulin, and phosphatidylinositol signaling, focal adhesion, and colorectal cancer). Target genes selection was performed with DIANA-miRpath, microT-V4.0. The union of the target mRNAs with a score above 3 was used as an input to ClueGO. Right-sided hypergeometric test yielded the enrichment for GO terms. Benjamini-Hochberg correction for multiple testing controlled the P-values. GO term fusion was applied for redundancy reduction.

[0039] **Figure 3.** The miRNA network in solid cancers (2532 samples, 31 cancer types, 120 miRNAs). The network was inferred for all expressed and varying miRNAs, without preselecting for differential expression. Standard Banjo parameters were adopted with a q6 discretization policy. The consensus graph depicted here was obtained from the best 100 nets (after searching through 8.4 3 10⁹ networks). MCL expression clusters are linked by different color edges. yEd graph editor (yFiles software) was employed for graphs visualization. The miRNAs expressed differentially in the tumors are color-coded and in the graph miRNA neighbors are of the same color code: Closely clustered miRNAs are either red (overexpressed) or green (down-regulated).

The node labels, for which expression and physical alteration (CGH, see “miRNA copy number variations in cancer and leukemia” section, *infra*) were concordant (i.e., overexpression and amplification), were emboldened and visually reinforced with a hexagonally shaped border.

[0040] **Figures 4A-4B.** Comparison of miRNA networks in normal lung and adenocarcinoma. (**Fig. A - top**) Normal lung (71 samples). A single complete miRNA network is shown. (**Fig. 4B - bottom**) Lung adenocarcinoma (125 samples). In this graph, one major and eight minor sub-networks were detected. For example, hsa-miR-10a/b, has-miR-29a/b, hsa-miR-107, and hsa-miR-103 are in minor independent sub-networks disjoint from the main one.

[0041] **Figure 5.** The KEGG functional analysis of eight disjointed minor miRNA networks in lung adenocarcinoma. The miRNA present in the unconnected cliques target genes are involved in many cancer-related terms, such as focal adhesion, small cell lung cancer, and calcium signaling. The detailed list of significant GO terms is shown in **Figure 15**.

[0042] **Figure 6.** The miRNA network in acute myeloid leukemia (589 samples, two sub-networks). Standard Banjo parameters were adopted with a q6 discretization policy. The consensus graph depicted here was obtained from the best 100 nets (after searching through 8.5×10^9 networks). The miRNA network in AML has disjointed cliques. hsa-miR-155 and hsa-miR-181, two miRNAs with clinical relevance are in two separated sub-networks, as expected from their prognostic independence. hsa-miR-181 is associated with hsa-miR-146a in a detached yellow miniclique. mir-155 belongs to the main sub-network, in the same red MCL clique of hsa-miR-223, hsa-miR-92a, hsa-miR-25, and hsa-miR-32. Finally, hsa-miR-29b has a key role in AML and acts as a hub in the AML net.

[0043] **Figure 7.** The miRNA network in chronic lymphocytic leukemia (254 samples, three sub-networks). Standard Banjo parameters were adopted with a q6 discretization policy. The consensus graph depicted here was obtained from the best 100 nets (after searching through $>1 \times 10^{10}$ networks). The network graph shows a major net and two separated minicliques: hsa-miR-23a/b and the hsa-miR-15/16 cluster. hsa-miR-15 and hsa-miR-16, two miRNAs frequently deleted in CLL, have been showed to regulate apoptosis via BCL2. The key hsa-miR-29b, acting as a hub in AML, is only a branch in CLL. AML prognostic hsa-miR-181 is disjointed in AML but not in CLL, while the reverse happens in CLL for prognostic hsa-miR-15/16 genes.

[0044] **Figure 8.** Deregulated miRNAs in leukemia from Mir155 transgenic mice are preferentially located close to hsa-miR-155 in the cancer network. The inventors compared miRNA profiles of three leukemia samples from Mir155 transgenes to controls from wild-type mice. The deregulated miRNAs (**Supplemental Table X**) were mapped onto the cancer network and highlighted in yellow. Most of the other miRNAs are concentrated around hsa-miR-155 node (black). When a diagonal is drawn and the two sides compared the difference between

yellow nodes is significant (Fisher's exact test, two-tail P-value < 0.009).

[0045] **Figure 9.** miRNA specificity in 50 normal tissues grouped by system. The tissue specificity was calculated by using the information content (IC), value expressed on y-axis; each color represents a system. hsa-miR-302 cluster is the most representative for embryo tissues.

[0046] **Figure 10.** The miRNA specificity during ES cell differentiation. miRNA specificity in 7 different types of embryonic tissues (embryonic stem cells, 7 days and 14 days embryonic bodies, trophoblasts, endoderm, induced pluripotent stem cells (iPS), spontaneously differentiating monolayers). The specificity was calculated by using the information content (IC).

[0047] **Figures 11A-11B.** The node degree distribution of the normal tissues and solid cancers miRNA networks. The figure illustrates the apparent scale-freeness of the graphs for normal tissues (Fig. 11A) and solid cancers (Fig. 11B). The blue curve represents the absolute frequency of node degrees and the green curve the inverse cumulative frequency. The exponential decrease of both curves shows that there are a lot more poorly connected nodes than highly connected (hubs). Both miRNA graphs thus present a scale free behavior.

[0048] **Figure 12.** miRNA specificity in 31 solid tumors and 20 leukemia types. The specificity was calculated by using the information content (IC), value expressed on y-axis; each color represents a cancer type.

[0049] **Figure 13.** Functions repressed in cancer by up-regulated miRNAs. KEGG analysis by ClueGO (Bindea et al. 2009) of terms ($p\text{-value} < 1 \times 10^{-3}$) targeted by up-regulated miRNAs (listed in Table I and Supplemental Table IV). Target genes selection was performed with DIANA-miRpath, microT-V4.0 (Papadopoulos et al. 2009). The union of the target mRNAs with a score above 3 was used as an input to ClueGO. Right-sided hyper-geometric test yielded the enrichment for GO-terms. Benjamini-Hochberg correction for multiple testing controlled the p-values. GO term fusion was applied for redundancy reduction.

[0050] **Figure 14.** Functions activated in cancer by down-regulated miRNAs. KEGG analysis by ClueGO of terms ($p\text{-value} < 1 \times 10^{-3}$) targeted by down-regulated miRNAs (listed in Table I and Supplemental Table V). Target genes selection was performed with DIANA-miRpath, microT-V4.0. The union of the target mRNAs with a score above 3 was used as an input to ClueGO. Right-sided hyper-geometric test yielded the enrichment for GO-terms. Benjamini-Hochberg correction for multiple testing controlled the p-values. GO term fusion was applied for redundancy reduction.

[0051] **Figure 15.** Functions controlled in lung adenocarcinoma by miRNAs from disjointed minor clusters. The colored groups are assigned the name of the most prominent GO group. The chart presents the specific terms and information related to the miRNA targets differentially expressed in cancer. The bars represent the number of the genes from the analyzed cluster found to be associated with the term, and the label displayed on the bars is the percentage of found genes compared to all the genes associated with the term. Term significance information is included in the

chart.

[0052] **Figure 16.** MicroRNA genetic network in colon adenocarcinoma (245 samples, 10 subnets). The network was inferred for all expressed and varying miRNAs, without preselecting for differential expression. Standard Banjo parameters were adopted with a q6 discretization policy. The consensus graph depicted here was obtained from the best 100 nets (after searching through $>1 \times 10^{10}$ networks). MCL graph-based clustering algorithm was applied to clusters extraction (miRNAs with highly related expression pattern have edges of the same color); thus, different clusters in the network are linked by different color edges. yEd graph editor (yFiles software, Tubingen, Germany) was employed for graphs visualization.

[0053] **Figure 17** MicroRNA genetic network in breast cancer (159 samples, 14 subnets). Same procedure described in legend of **Supplemental Figure 14**.

[0054] **Figure 18.** MicroRNA genetic network in prostate cancer (170 samples and 5 subnets). Same procedure described in legend of **Figure 14**.

DETAILED DESCRIPTION

[0055] The inventors built our miRNA networks exclusively from miRNA expression data. Here, the inventors report the first miRNA network from normal tissues. In parallel, the inventors built miRNA networks for coupled cancerous and noncancerous tissues. By comparing normal to cancer networks the inventors attained a second goal: the identification of cancer variations in miRNA networks. Also, inventors superimposed DNA variations onto expression data to generate a comprehensive miRNA alteration map in cancer.

[0056] Described herein is the use of a miRNA networks comprised from miRNA expression data that includes a least one miRNA network from normal tissues, and/or at least one miRNA network for coupled cancerous and noncancerous tissues.

[0057] The method for identification of cancer variations in miRNA networks, includes comparing normal to cancer networks by generating a comprehensive miRNA alteration map in cancer, and superimposing DNA variations onto expression data. As further described herein, at least one miRNA network comprising the miRNA network substantially as shown in one or more of Figure 4B, Figure 6, Figure 7; Figure 8, Figure 16, Figure 17, and Figure 18.

[0058] A method for diagnosing a subject suspected of having a solid cancer as also described herein which includes screening a sample from the subject for one or more of the differentially regulated miRNAs listed in Supplemental Table IV, Supplemental Table V, Supplemental Table VIII and/or Supplemental Table IX.

[0059] The methods are useful to show where a change in miRNA expression in a subject is indicative of the subject being predisposed to, or having, a solid cancer and/or hematological cancer. In certain embodiments, the cancer is one or more of lung, breast, colon, prostate, pancreas, and

nasopharyngeal carcinomas, glioblastoma, melanoma, Ewing sarcoma, osteosarcoma, T-cell acute lymphoblastic leukemia (T-ALL), AML, CLL, myelodysplasia, various lymphomas, and mucosa-associated lymphoid tissue MALT. In a particular embodiment, the method for detecting cancer differentiation in a subject, can include screening for at least one of hsa-miR-221 and hsa-miR-218.

[0060] As further described herein, a method for compiling and/or analyzing a network of biological functions in a subject, includes:

i) obtaining information on one or more miR present in the subject, wherein at least one miR reflects a biological function; and

ii) subjecting the obtained information to analysis of coherent groups of nodes by using clustering algorithms to calculate a relationship between the functional reporters to generate a network relationship of the miRs to the biological function. In certain embodiments, the subject is human.

[0061] Also described is a method of profiling miR expression in a human subject having, or suspected of having, a cancer-related disease, the method comprising:

determining, for each miR of a set of miRs, a level of the miRs in a sample of the subject;

and

determining whether the miRs are identified as miRs in separate and un-linked miRNA subnets,

wherein, the presence of the separate and un-linked miRNA subnets, is indicative of the subject having the cancer-related disease.

[0062] In certain embodiments, the method includes detecting miRNA networks in cancer cells that are independently regulated miRNAs.

[0063] In certain embodiments, the method includes detecting target genes of uncoordinated miRs that are involved in specific cancer-related pathways.

[0064] For example, the method can include wherein the miRs are up-regulated microRNAs found in solid cancers, the set of miRs comprising one or more of: the miRs in Table IV, namely: hsa-miR-21, hsa-miR-25, hsa-miR-20a, hsa-miR-17, hsa-miR-106a, hsa-miR-106b, hsa-miR-146a, hsa-miR-92a, hsa-miR-103, hsa-miR-130b, hsa-miR-30c, hsa-miR-93, hsa-miR-107, hsa-miR-30e, hsa-miR-15a, hsa-miR-181b, hsa-miR-15b, hsa-miR-181a, hsa-miR-32, hsa-miR-345, hsa-miR-34a, hsa-miR-374a.

[0065] For example, the method can include wherein the miRs are down-regulated microRNAs found in solid cancers, the set of miRs comprising one or more of: the miRs in Table V, namely: miR-203, hsa-miR-145, hsa-miR-205, hsa-miR-206, hsa-miR-33b, hsa-miR-193a, hsa-miR-204, hsa-miR-143, hsa-miR-326, hsa-miR-338, hsa-miR-9, hsa-miR-95, hsa-miR-138, hsa-miR-183, hsa-miR-202, hsa-miR-128a, hsa-miR-214, hsa-miR-132, hsa-miR-299, hsa-miR-129, hsa-miR-133a, hsa-miR-139, hsa-miR-339, hsa-miR-1, hsa-miR-133b, hsa-miR-323, hsa-miR-218, hsa-miR-335.

[0066] Also described is a method for characterizing a cancer comprising:

- i) determining an expression level of one or more miRNAs in a biological sample of a subject; and,
- ii) characterizing the cancer based on at least a subset of the miRNAs selected from: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or, the network shown in Figure 18.

[0067] Also described is a method for characterizing a cancer in a subject comprising:

- i) determining an expression level of each of a plurality of miRNAs in a biological sample of the subject, and
- ii) characterizing the cancer based on the expression level of each of the plurality of miRNAs,

wherein the characterizing is with increased sensitivity or specificity as compared to characterizing by detecting an expression level of less than each of the plurality of miRNAs selected from: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or the network shown in Figure 18.

[0068] In certain embodiments, the plurality comprises at least 10 miRNAs. The method of claim 24 or 25, wherein at least a subset of the plurality of miRNAs is selected from Supplemental Table IV. In certain embodiments, the, wherein at least a subset of the plurality of miRNAs is selected from Supplemental Table V.

[0069] In certain embodiments, the method can include classifying a cell as non-cancerous, by:

- i) determining an expression level of one or more miRNAs in a biological sample of a subject; and
- ii) classifying the cell as non-cancerous when less than about 3000 copies per microliter of at least a subset of the miRNAs is detected in the sample.

[0070] In certain embodiments, the method can include classifying a cell as non-cancerous, by:

- i) determining an expression level of one or more miRNAs in a biological sample of a subject; and
- ii) classifying the cell as cancerous when greater than about 9000 copies per microliter of at least a subset of the miRNAs is detected in the sample.

[0071] In certain embodiments, the method can further include selecting a therapy or treatment regimen based on the classification. In certain embodiments, the biological sample is selected from the group consisting of: a heterogeneous cell sample, sputum, blood, blood cells, serum, biopsy, urine, peritoneal fluid, and pleural fluid.

[0072] Also described herein is a detection system configured to assess miRNAs selected from

the group consisting of: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; the network shown in Figure 18; miRs in Supplemental Table IV; or miRs in Supplemental Table V.

[0073] In certain embodiments, the system comprises a set of probes that selectively hybridizes to the two or more miRNAs.

[0074] Also described herein is a kit comprising a set of probes that selectively hybridizes to two or more miRNAs selected from the group consisting of: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; the network shown in Figure 18; miRs in Supplemental Table IV; or miRs in Supplemental Table V.

[0075] In certain embodiments, each of the probes can be coupled to a substrate.

[0076] Also described herein is a method for analyzing a network of biological functions in a biological entity, comprising:

- i) obtaining information on at least two miRs in the biological entity, wherein the miRs reflect a biological function; and
- ii) subjecting the obtained information to processing to calculate a relationship between the miRs to generate a network relationship of the biological functions.

[0077] In certain embodiments, the biological entity is a cell.

[0078] Further, in certain embodiments, the network can be used for a use selected from the group consisting of identification of a biomarker, analysis of a drug target, analysis of a side effect, diagnosis of a cellular function, analysis of a cellular pathway, evaluation of a biological effect of a compound, and diagnosis of an infectious disease.

[0079] In certain embodiments, the system can further include means for analyzing the generated network by conducting an actual biological experiment. For example, the means for analyzing can comprise a regulation agent specific to the function.

[0080] Also described herein is a computer program for implementing in a computer, a method for analyzing a network of biological functions in a biological entity, comprising the steps of:

- i) obtaining information on at least two miRs in the biological entity, wherein the miRs reflect a biological function; and
- ii) subjecting the obtained information to processing to calculate a relationship between the miRs to generate a network relationship of the biological functions.

[0081] Also described herein is storage medium comprising a computer program for implementing in a computer, a method for analyzing a network of biological functions in a biological entity, comprising the steps of:

- i) obtaining information on at least two miRs in the biological entity, wherein the miRs

reflect a biological function; and

ii) subjecting the obtained information to processing to calculate a relationship between the miRs to generate a network relationship of the biological functions.

[0082] Also described herein is a transmission medium comprising a computer program for implementing in a computer, a method for analyzing a network of biological functions in a biological entity, comprising the steps of:

i) obtaining information on at least two miRs in the biological entity, wherein the miRs reflect a biological function; and

ii) subjecting the obtained information to processing to calculate a relationship between the miRs to generate a network relationship of the biological functions.

[0083] Also described herein is an assay kit for detecting a risk that a cell will become malignant, comprising reagents for determining a cell signature, wherein the signature comprises the presence and/or level of two or more miRs selected from: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or, the network shown in Figure 18.

[0084] In certain embodiments, the miR is differentially expressed in the cells shows that the cell is at risk of becoming malignant.

[0085] Also described herein is method of determining a risk of developing cancer in a subject, comprising:

providing a biological sample from the subject;

determining a cell signature for the biological sample, wherein the signature comprises a collection of measurements of at least the presence and/or level of two or more miRs selected from: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or the network shown in Figure 18;

comparing the cell signature of the biological sample with a cell signature of a control sample; and

determining the risk of developing cancer.

[0086] In certain embodiments, the biological sample is selected from the group consisting of a living cell, a dead cell, any non-cellular liquid samples comprising nipple aspirate fluid, urine, blood, serum, plasma and a lavage sample, and any combinations thereof. Also, in certain embodiments, the cell signature of the control sample is obtained from a database.

[0087] In certain embodiments, at least part of the method is performed by automated means.

In certain embodiments, the automated means comprises a computer-based system configured to carry out at least one of the following: measuring the presence of absence of a substance in the sample; recording data obtained from the measurement; analyzing the data; determining the risk of

cancer; and generating a report. For example, the computer-based system can comprise suitable any hardware, software, firmware, processor, and any combinations thereof. In certain embodiments, the computer-based system is configured to access and/or use a database comprising a cell signature of a pre-cancerous cell and/or a control epithelial cells via a network system.

[0088] Also described herein is a method of making a medical report related to the risk of developing cancer in a subject, comprising:

- providing a biological sample from the subject;
- determining a signature for the biological sample, wherein the signature comprises a collection of measurements of at least two characteristics of the sample selected from one or more of the following: presence and/or level of at least one miRNA;
- comparing the signature of the biological sample with a mammary epithelial cell signature of a control sample;
- determining the risk of developing cancer; and
- generating a report related to the risk of developing cancer.

[0089] In certain embodiments the automated means comprises a computer-based system configured to carry out at least one of the following: measuring the cell signature; recording data obtained from the measurement; analyzing the data; determining the risk of developing cancer; and generating a report. For example, the computer-based system can comprise any suitable hardware, software, firmware, processor, and any combinations thereof. In certain embodiments, the computer-based system is configured to access and/or use a database comprising a cell signature of a pre-cancerous cell and/or a control cell via a network system.

[0090] Also described herein is a method of making a medical report related to the risk of developing cancer in a subject, comprising:

- providing a biological sample from the subject;
- determining a cell signature for the biological sample, wherein the signature comprises a collection of measurements of at least two characteristics of the cell, the at least two miRs selected from: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or the network shown in Figure 18;
- comparing the cell signature of the biological sample with a cell signature of a control sample;
- determining the risk of developing cancer; and
- generating a report related to the risk of developing cancer.

[0091] Also described herein is a method for compiling and/or analyzing an miRNA database, comprising one or more of:

- using F635-background values;

removing bad spots;
 averaging nonexpressed spots for each gpr files;
 for each mature miRNA, computing the geometric mean of its multiple reporters;
 assigning a NaN value to miRNAs with more than 50% of corrupted spots;
 log2-transforming the results;
 normalizing by using the quantiles normalization;
 performing t-test over two classes' experiments of F-tests over multiple classes (i.e.,
 different normal tissues);
 performing target genes selection, wherein the union of the target mRNAs with a score >3
 is used as an input to ClueGO;
 using ClueGO to relate differential expression in cancer to functional pathways (KEGG),
 wherein: ClueGO visualizes the selected terms in a functionally grouped annotation network that
 reflects the relationships between the terms based on the similarity of their associated genes, the
 size of the nodes reflects the statistical significance of the terms; and the degree of connectivity
 between terms (edges) is calculated using kappa statistics;
 using the calculated kappa score for defining functional groups, wherein: a term can be
 included in several groups, the reoccurrence of the term is shown by adding "n", and the group
 leading term is the most significant term of the group,
 whereby the network integrates only the positive kappa score term associations.

[0092] In certain embodiments, the expression values are preprocessed to only filter out nonvarying miRNAs, according to the following parameters:

#filter.flag = filter (variation filter and thresholding flag);
 #preprocessing.flag = no disk or norm (discretization and normalization flag);
 #minchange = 10 (minimum fold change for filter);
 #mindelta = 512 (minimum delta for filter);
 #threshold = 64 (value for threshold);
 #ceiling = 20,000 (value for ceiling);
 #max. sigma.binning = 1 (maximum sigma for binning);
 #prob.thres = 1 (value for uniform probability threshold filter);
 #num.excl = 2% of total chips (number of experiments to exclude (max and min) before
 applying variation filter);
 #log.base.two = no (whether to take the log base two after thresholding);
 #number.of.columns. above.threshold = 1% of total chips (remove row if n columns not ≥
 than given threshold above.threshold); and
 #column.threshold = 512 (threshold for removing rows).

[0093] Also described herein is a system for analyzing a an miRNA database, comprising one

or more means for:

- using F635-background values;
- removing bad spots;
- averaging nonexpressed spots for each gpr files;
- for each mature miRNA, computing the geometric mean of its multiple reporters;
- assigning a NaN value to miRNAs with more than 50% of corrupted spots;
- log2-transforming the results;
- normalizing by using the quantiles normalization;
- performing t-test over two classes' experiments of F-tests over multiple classes (i.e., different normal tissues);
- performing target genes selection, wherein the union of the target mRNAs with a score >3 is used as an input to ClueGO;
- using ClueGO to relate differential expression in cancer to functional pathways (KEGG), wherein: ClueGO visualizes the selected terms in a functionally grouped annotation network that reflects the relationships between the terms based on the similarity of their associated genes, the size of the nodes reflects the statistical significance of the terms; and the degree of connectivity between terms (edges) is calculated using kappa statistics;
- using the calculated kappa score for defining functional groups, wherein: a term can be included in several groups, the reoccurrence of the term is shown by adding "n", and the group leading term is the most significant term of the group,
- whereby the network integrates only the positive kappa score term associations.

[0094] In certain embodiments, the subject is a human.

[0095] Also described herein is a computer program for implementing in a computer, a method for compiling and/or analyzing an miRNA database, comprising one or more of:

- using F635-background values;
- removing bad spots;
- averaging nonexpressed spots for each gpr files;
- for each mature miRNA, computing the geometric mean of its multiple reporters in the chip;
- assigning a NaN value to miRNAs with more than 50% of corrupted spots;
- log2-transforming the results;
- normalizing by using the quantiles normalization;
- performing t-test over two classes' experiments of F-tests over multiple classes (i.e., different normal tissues);
- performing target genes selection, wherein the union of the target mRNAs with a score >3 is used as an input to ClueGO;
- using ClueGO to relate differential expression in cancer to functional pathways (KEGG),

wherein: ClueGO visualizes the selected terms in a functionally grouped annotation network that reflects the relationships between the terms based on the similarity of their associated genes, the size of the nodes reflects the statistical significance of the terms; and the degree of connectivity between terms (edges) is calculated using kappa statistics;

using the calculated kappa score for defining functional groups, wherein: a term can be included in several groups, the reoccurrence of the term is shown by adding “n”, and the group leading term is the most significant term of the group,

whereby the network integrates only the positive kappa score term associations.

[0096] Also described herein is a storage medium comprising a computer program for implementing in a computer, a method for compiling and/or analyzing an miRNA database, comprising one or more of:

using F635-background values; removing bad spots;

averaging nonexpressed spots for each gpr files;

for each mature miRNA, computing the geometric mean of its multiple reporters;

assigning a NaN value to miRNAs with more than 50% of corrupted spots;

log2-transforming the results;

normalizing by using the quantiles normalization;

performing t-test over two classes' experiments of F-tests over multiple classes (i.e., different normal tissues);

performing target genes selection, wherein the union of the target mRNAs with a score >3 is used as an input to ClueGO;

using ClueGO to relate differential expression in cancer to functional pathways (KEGG), wherein: ClueGO visualizes the selected terms in a functionally grouped annotation network that reflects the relationships between the terms based on the similarity of their associated genes, the size of the nodes reflects the statistical significance of the terms; and the degree of connectivity between terms (edges) is calculated using kappa statistics;

using the calculated kappa score for defining functional groups, wherein: a term can be included in several groups, the reoccurrence of the term is shown by adding “n”, and the group leading term is the most significant term of the group,

whereby the network integrates only the positive kappa score term associations.

[0097] Also described herein is a transmission medium comprising a computer program for implementing in a computer, a method for compiling and/or analyzing an miRNA database, comprising one or more of:

using F635-background values;

removing bad spots;

averaging nonexpressed spots for each gpr files;

for each mature miRNA, computing the geometric mean of its multiple reporters;
assigning a NaN value to miRNAs with more than 50% of corrupted spots;
log2-transforming the results;
normalizing by using the quantiles normalization;
performing t-test over two classes' experiments or F-tests over multiple classes (i.e.,
different normal tissues);
performing target genes selection, wherein the union of the target mRNAs with a score >3
is used as an input to ClueGO;
using ClueGO to relate differential expression in cancer to functional pathways (KEGG),
wherein: ClueGO visualizes the selected terms in a functionally grouped annotation network that
reflects the relationships between the terms based on the similarity of their associated genes, the
size of the nodes reflects the statistical significance of the terms; and the degree of connectivity
between terms (edges) is calculated using kappa statistics;
using the calculated kappa score for defining functional groups, wherein: a term can be
included in several groups, the reoccurrence of the term is shown by adding "n", and the group
leading term is the most significant term of the group,
whereby the network integrates only the positive kappa score term associations.

[0098] Also described herein is an assay kit for analyzing an miRNA database, comprising
reagents one or more of:

using F635-background values;
removing bad spots;
averaging nonexpressed spots for each gpr files;
for each mature miRNA, computing the geometric mean of its multiple reporters;
assigning a NaN value to miRNAs with more than 50% of corrupted spots;
log2-transforming the results;
normalizing by using the quantiles normalization;
performing t-test over two classes' experiments or F-tests over multiple classes (i.e.,
different normal tissues);
performing target genes selection, wherein the union of the target mRNAs with a score >3
is used as an input to ClueGO;
using ClueGO to relate differential expression in cancer to functional pathways (KEGG),
wherein: ClueGO visualizes the selected terms in a functionally grouped annotation network that
reflects the relationships between the terms based on the similarity of their associated genes, the
size of the nodes reflects the statistical significance of the terms; and the degree of connectivity
between terms (edges) is calculated using kappa statistics;
using the calculated kappa score for defining functional groups, wherein: a term can be

included in several groups, the reoccurrence of the term is shown by adding “n”, and the group leading term is the most significant term of the group,

whereby the network integrates only the positive kappa score term associations.

[00099] In certain embodiments, the substance detected is differentially expressed in a sample from a subject at risk of becoming malignant.

[00100] In certain embodiments, the assay kit detects wherein one or more miRNAs are differentially transcribed in the sample at risk of becoming malignant.

[00101] In certain embodiments, the assay kit detects wherein one or more of the miRNAs may have an expression level and/or copy number in a sample from a subject at risk of becoming malignant compared to a normal cell.

[00102] In certain embodiments, the assay kit detects wherein the miRNAs is selected from the group consisting miRs shown in: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or the network shown in Figure 18.

[00103] In certain embodiments, the assay kit detects wherein the miRNAs is selected from the group consisting miRs shown in: Supplemental Table IV or Supplemental Table V.

[00104] In certain embodiments, the sample is selected from one or more of: a living cell, a dead cell, any non-cellular liquid samples comprising aspirate fluid, urine, blood, serum, plasma and a lavage sample, and any combinations thereof.

[00105] The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

[00106] EXAMPLE I

[00107] The sequence data from this study have been submitted to the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession nos. GSE8126, GSE17155, GSE20099, GSE3467, GSE7828, GSE7055, GSE6857, GSE16654, GSE8126, and GSE14936, and the microarray data from this study have been submitted to ArrayExpress (ebi.ac.uk/microarray-as/ae) under accession nos. E-TABM-866, E-TABM-762, E-TABM-508, E-TABM-429, E-TABM-434, E-TABM-405, E-TABM-343, E-TABM-42, E-TABM-48, E-TABM-49, E-TABM-50, E-MEXP-1796, and E-GEOD-3292.

[00108] Results

[00109] The miRNA network in normal tissues

- [00110] The inventors assayed mature miRNAs in 17 groups of normal human tissues, from a total of 1107 chips. Tissue specificity was calculated by using the information content (IC) (Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129: 1401–1414, by measuring expression levels by sequencing cloned miRNAs. The most tissue-specific miRNAs are the members of the hsa-miR-302 cluster, as shown in **Figure 9** and **Supplemental Tables I and III**.
- [00111] hsa-miR-302a/b/c were expressed in embryonic samples. Complexity of genetic regulatory mechanisms in higher organisms is thought to also be achieved through controlled and coordinated networks of miRNAs. The inventors developed and used a microarray database to generate miRNA networks based exclusively on expression data.
- [00112] The inventors applied Banjo (Smith VA, Yu J, Smulders TV, Hartemink AJ, Jarvis ED. 2006. Computational inference of neural information flow networks. *PLoS Comput Biol* 2: e161. doi: 10.1371/journal.pcbi.0020161) to infer the Bayesian network for normal tissues. miRNA relations were modeled as graphs where nodes represent the miRNAs and colored edges the relationships between them. The node degree distribution of the normal miRNA network is illustrated in **Figure 11A**.
- [00113] The exponential decrease of both absolute frequency and inverse cumulative frequency curves shows that there were a lot more poorly connected nodes than highly connected (hubs). More than 40% of the nodes had degrees of 1 and almost 75% had degrees of #2. The normal tissues miRNA graph thus presented a scale free behavior. The highest degree hub was hsa-miR-16, followed by hsa-miR-215. To discover miRNA groups with highly related expression patterns the inventors extracted coherent groups of nodes by adopting clustering algorithms, using the MCL graph-based algorithm (Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30: 1575–1584). This algorithm, which was implemented in Neat, (Brohee S, Faust K, Lima-Mendez G, Vanderstocken G, van Helden J. 2008. Network Analysis Tools: From biological networks to clusters and pathways. *Nat Protoc* 3: 1616–1629) enables good performance in extracting coregulated genes from transcriptomes.
- [00114] **Figure 1** displays the miRNA network of normal tissues, obtained from over 1000 samples and 50 cell types/tissues. The inventors used all of the expressed miRNAs to build the Bayesian networks (rather than only the differentially expressed ones). The MCL clusters with high coexpression patterns throughout normal tissues are linked by specific colored edges. miRNAs are generally connected as expected from the published literature.
- [00115] For example, hsa-miR-133a/b was in a cluster with hsa-miR-1 (light orange) and all were involved in skeletal muscle proliferation and differentiation. A close cluster is hsa-miR-10a/b and hsa-miR-214 (green). hsa-miR-214 is expressed during early segmentation stages in somites

and can modulate the expression of genes regulated by Hedgehog. Inhibition of hsa-miR-214 results in a reduction or loss of slow-muscle cell types. These muscle/differentiation clusters are linked to hsa-miR-143/145, a miRNA capable of pushing ES cells to differentiate. The top, right proliferation cluster hsa-miR-106a/b/93 to hsa-miR-20a/17 and to hsa-miR-25/92a (MYC-associated) includes hsa-miR-223 and is involved in cell cycle progression. The hsa-miR-145 node links the proliferation clusters described above to the muscle differentiation clusters.

[00116] While not wishing to be bound by theory, the inventors herein now believe that this link explains why the loss of hsa-miR-145 in cancer leads to an undifferentiated cellular state. The hsa-miR-29 family, that targets the anti-apoptotic protein MCL1 and plays a role in the TP53 pathway is linked to hsa-miR-30 and to hsa-miR-15/16, miRNAs that target the anti-apoptotic protein BCL2. Overexpression of hsa-miR-29a leads to epithelial-to-mesenchymal transition (EMT) and metastasis, in cooperation with oncogenic Ras signaling. hsa-miR-221-222, regulators of the cell cycle, together with hsa-miR-206, hsa-miR-155 (pre-B cell proliferation), and hsa-miR-130a/b are in a yellow cluster. In the top center of the graph, hsa-miR-194 and hsa-miR-192 connect to hsa-miR-215, within a purple cluster associated with TP53 activation. In a close branch the hsa-miR-200 family, hsa-miR-203 and hsa-miR-205 are directly involved in TGF-beta-mediated EMT and differentiation. hsa-miR-181 family members are involved in B- and T-lineage and myoblast differentiation. hsa-miR-181, hsa-miR-200, hsa-miR-205, and hsa-miR-215 are all linked to hsa-miR-145.

[00117] Thus, again while not wishing to be bound by theory, the inventors herein now believe that the loss of the hsa-miR-145 hub also impacts the TP53/EMT/ differentiation branch. Another important hub appears to be hsa-miR-16, which is located in the other half of the net and coordinates hsa-miR-29 and hsa-miR-221/206/155/130 clusters. The hsa-miR-16 hub also feeds the hsa-miR-26/let-7/hsa-miR-302 branch. hsa-miR-26 is a key gene in hepatocellular carcinoma and its expression is associated with survival and response to adjuvant therapy with interferon alpha. Let-7 regulates Ras and hsa-miR-302 are expressed in ES cells and other early embryonic tissues.

[00118] *The miRNA networks in solid cancers are reprogrammed*

[00119] Fifty-one (51) types of cancer in 3312 samples (2532 solid cancers and 780 leukemia samples) were investigated. The results of miRNA differential expression in solid cancers are listed in Table 1 and in Supplemental Tables IV and V. These results establish a miRNA profiling in cancer. For example, hsa-miR-21 and hsa-miR-17/20/25/92/103/106/146a were overexpressed, whereas hsa-miR-203/205 and hsa-miR-143/145 were down-regulated.

[00120] The KEGG pie chart in Figure 2 and the corresponding network graphs (Figure 13, Figure 14) show the functional effect of differentially expressed miRNAs on cellular pathways. The large majority of the affected pathways is related to cancer or signal transduction (i.e., Wnt, VEGF, TGF-beta, insulin, phosphatidylinositol signaling, focal adhesion, and colorectal cancer).

The inventors also applied the IC measure to identify cancer-specific miRNAs. The ICs were almost as high as those measured for the normal tissues, indicating that there were miRNAs with high cancer-type specificity (**Figure 12; Supplemental Table VI**).

[00121] The inventors generated a global miRNA expression network for solid cancers (**Figures 3A-3B**). It is important to note that, to build the Bayesian net, the inventors used all the expressed and varying miRNAs as input, rather than only using the differentially expressed ones (miRNAs with low variation were excluded from the analysis).

[00122] The node degree distribution of the solid cancer miRNA network is illustrated in **Figure 3B**. Like the normal tissues miRNA graph, the solid cancer net also presented a scale-free behavior. In cancer, the most connected hub was hsa-miR-30c (degree 10), followed by has-miR-16 (degree 6). Whereas, in nonmalignant tissues, hsa-miR-16 was the most connected node (degree 8) and hsa-miR-30c had only a low degree of 3. Opposite behavior had TP53 regulated hsa-miR-215 (degree 6 in normal tissues and degree 3 in cancer) and has-miR-103/106a (degree 5 in normal tissues and only degree 1 in cancer). The exchanges of hubs between nonmalignant and cancer tissues were the first notable sign of divergences in their respective miRNA programs.

[00123] The MCL clustering algorithm was employed to map the sub-networks with high coexpression patterns (these MCL clusters, or cliques, are linked by specific colored edges). Additionally, the inventors color-coded the miRNA nodes according to their differential expression in tumors (red, overexpressed; green, down-regulated). Neighbors preferentially appeared with the same trend, such that clustered miRNAs were either overexpressed or down-regulated. For example, hsa-miR-17/20a (chr 13q31.3), hsa-miR-106a/b (chr Xq26.2 and chr 7q22.1), and hsa-miR-93 (chr 7q22.1) were all up-regulated in cancers. Conversely, hsa-miR-143/145 (chr 5q32), hsa-miR-133a/b (chr 18q11.2, chr 20q13.3 and chr 6p12.2), hsa-miR-214 (chr 1q24.3), and hsa-miR-138 (chr 3p21.33 and chr 16q13), all in the same coexpression clique, were down-regulated.

[00124] In investigating certain cancer types (lung, colon, breast, prostate), the inventors discovered unique miRNA co-expression clusters, or networks. Normal lung miR-Nome was represented by a single complete miRNA network (**Figure 4A**), while adenocarcinomas were by one major and eight un-connected sub-networks (**Figure 4B**). KEGG functional analysis of the miRNAs in the eight minor sub-networks unconnected in cancer showed that they target genes involved in cancer-related pathways (**Figure 5**).

[00125] Strikingly, a similar situation of disjointed miRNA cliques was also present in the other cancers, for example: colon, breast, and prostate cancers (**Figure 15, Figure 17, Figure 18**, respectively). In particular, the inventors identified a number of notable, and often common, miRNAs in the unconnected cancer clusters. has-miR-10a/b was identified in lung, colon, and breast cancers; miR 26a/b in colon and prostate cancers; hsa-miR-29a/b in breast, colon, and lung cancers; hsa-miR-181 family members in colon and breast cancers; and hsa-miR-107/ hsa-miR-103

in breast, prostate, lung, and colon cancers. The let-7c/a miRNAs were prominent in colon, lung, and prostate cancers, and hsa-miR-106a/b, linked to hsa-miR-17 and hsa-miR-20, in colon and lung cancers. Other miRNA cliques included hsa-miR200c, linked to TP53-associated hsa-miR-192/215, in a colon sub-network.

[00126] Thus, miRNA networks were reprogrammed in solid cancer and the expression of few notable miRNAs was independent from the major network. While not wishing to be bound by theory, the inventors herein now believe that the single graph in the overall solid cancer net can be explained only by the same miRNAs having variable roles in a range of cancers, such that a miRNA regulates different targets in different cell types.

[00127] Leukemias confirm that miRNA networks are aberrant in neoplasia

[00128] In addition to looking at miRNA networks in solid cancers, the inventors generated the networks for two hematological cancers, acute myeloid leukemia (AML) and chronic lymphocytic leukemia (CLL). Both of these leukemias have been well characterized in terms of miRNA profiles and their relations to prognosis. The miRNA network in AML also had disjointed cliques (**Figure 6**).

[00129] One prominent finding here was that hsa-miR-155 and hsa-miR-181, two miRNAs with clinical relevance, were positioned in two separated sub-networks, as expected from their prognostic independence. In fact, hsa-miR-181 was associated to hsa-miR-146a in a detached yellow mini-clique, while hsa-miR-155 belonged to the main sub-network, in the same red MCL clique as hsa-miR-223, hsa-miR-92a, hsa-miR-25, and hsa-miR-32. Finally, hsa-miR-29b has a key role in AML and, in accordance, it acts as a hub in the AML net.

[00130] In chronic lymphocytic leukemia (CLL) two small cliques were separated from the main net: hsa-miR-23a/b and a second one embracing the hsa-miR-15/16 pair. hsa-miR-15 and hsa-miR-16, two miRNAs frequently deleted in CLL, have been showed to regulate apoptosis via BCL2. Thus, the network topologies for these two leukemias could recapitulate their respective molecular pathology, with the key AML hsa-miR-29b acting as a hub in AML, but only a branch in CLL. AML prognostic hsa-miR-181 was disjointed in AML, but not in CLL, with the reverse being true for the CLL prognostic hsa-miR-15/16 pair.

[00131] miRNA copy number variations in cancer and leukemia

[00132] miRNAs are differentially expressed in human cancer, but little is known about their chromosomal alterations, such as amplifications (hsa-miR-17/92) and deletions (hsa-miR-15a/16-1). To systematically study miRNA copy number alterations in cancer, the inventors investigated 744 samples (solid cancers and leukemia), at medium resolution (150 kb). The inventors used data from array comparative genomic hybridization (aCGH) and calculated, for each of 20,000 different chromosomal locations, two P-values, one for deletion and one for amplification. To measure miRNA copy number alterations the inventors used their respective host genes or, when unavailable, their two flanking genes. In addition, to focus on the functional role of miRNAs, to

increase the statistical power of the approach described herein, and to possibly dilute the contribution of the host/ associated genes, the inventors considered miRNAs as families. miRNA families have similar targeting properties and thus their members are expected to have similar impacts on oncogenesis.

[00133] The inventors worked on aCGH samples from lung, pancreas, breast, colon, and nasopharyngeal carcinomas, glioblastoma, melanoma, Ewing sarcoma, osteosarcoma, T-cell acute lymphoblastic leukemia (T-ALL), AML, CLL, myelodysplasia, various lymphomas, and mucosa-associated lymphoid tissue MALT (**Supplemental Table VII**).

[00134] The inventors used aCGH from the NCBI Gene Expression Omnibus (GEO) and Stanford Microarray Database (SMD). CDKN2A and CDKN2B were identified as the most deleted genes in cancers, followed by other tumor suppressors PTEN, ATM, and TP53. Oncogenes, like EGFR, MYC, LYN, MET, and MOS, were amplified. **Supplemental Tables VIII and IX** list amplified and deleted miRNA families. The detection of amplified hsa-miR-17-5p/20/93/106 family was a successful validation of our approach. It is also noteworthy that the MIR17HG host gene for the hsa-miR-17/92 cluster was not present in the arrays, but its flanking genes successfully compensated for its absence. The top deleted miRNA family was hsa-miR-204/211, followed by other families including hsa-miR-200b/c/429, hsa-miR-141/200a, hsa-miR-125/351, and hsa-miR-218. Down-regulation of hsa-miR-200a/b/c/429 and 141 have been linked to breast cancer stem cells by targeting BMI1, a stem cell self-renewal regulator. Likewise hsa-miR-211 is involved in stem cells as it shows the highest Information content in an ES cell differentiation series (**Figure 9; Supplemental Table III**).

[00135] Thus, while not wishing to be bound by theory, the inventors herein now believe that loss of hsa-miR-211 is involved in regulation of cancer differentiation. The inventors herein now also believe this is the same for hsa-miR-218, which is deleted in cancer and highly expressed in spontaneously differentiated monolayers. The results from aCGH were overlaid on the expression network in solid cancers (**Figure 3**). The node labels, for which expression and physical alteration were concordant (i.e., overexpression and amplification), were emboldened and visually reinforced with a hexagonally shaped border.

[00136] *Deregulated miRNAs in a MMSS-induced leukemia are preferentially located around hsa-miR-155 in the miRNA network*

[00137] The inventors generated two cases of leukemias in *Em/VH* Mir155 transgenic mice. These leukemias were positive for CD43 and T-cell markers (CD3, CD8) and negative for B220. Both cases exhibited VDJ and TCR oligoclonal rearrangement. This T-cell immunophenotype might be caused by the proliferation of lymphoid progenitors that atypically differentiated into T cells. The disease started early, at 2 and 4 mo of age, respectively, and had a rapid course with the mice dying 2 wk later. Their autopsy revealed a widespread leukemic infiltration, with

organomegaly and lymphadenopathy, histologically diagnosed as an aggressive malignant lymphoproliferation similar to Burkitt lymphoma (data not shown). The injection of single sick splenocytes into 30 syngeneic mice was sufficient to reproduce the full blown malignancy.

[00138] The inventors compared the miRNA profiles of three leukemia samples from these *Mir155* transgenes to controls from wild-type mice. Then the inventors located the positions in the network for the miRNAs regulated in the transgene's leukemias (**Supplemental Table X**). The inventors did not have an acute lymphocytic leukemia miRNA network as reference, therefore the inventors mapped the deregulated miRNAs onto the generic cancer network and highlighted the nodes in yellow (**Figure 8**). The yellow nodes appeared concentrated around the has-miR-155 node (black). When a diagonal, separating the has-miR-155 half from the other one, was drawn and the two sides compared the difference in yellow nodes concentrations was significant (14 vs. 43, 4 vs. 57, Fisher's exact test, two-tail P-value < 0.009). The topological distribution was even more skewed if has-miR-29s and has-miR-181s were not considered as has-miR-155 regulated. In fact, has-miR-181 overexpression and has-miR-29 down-regulation are hallmarks miRNA. In leukemias, thus, they are likely to be independent events in cellular transformation and not directly related to *Mir155* transgene.

[00139] Discussion

[00140] The inventors have presented a thorough analysis of miRNA tissue specificity in 50 different normal tissues grouped by 17 systems, corresponding to 1107 human samples. A small set of miRNAs were tissue-specific, while many others were broadly expressed. The inventors also studied 51 oncologic or hemato-oncologic disorders and identified cancer-type-specific miRNAs. Then the inventors inferred genetic networks for miRNAs in normal tissues and in their pathological counterparts. Normal tissues were represented by single complete miRNA networks. Cancers instead were portrayed by separate and un-linked miRNA subnets.

[00141] Intriguingly, miRNAs independent from the general transcriptional program were often known as cancer-related. While not wishing to be bound by theory, the inventors herein now believe that this "egocentric" behavior of cancer miRNAs is the result of positive selection during cancer establishment and progression, as supported by aCGH. Leukemias were also rewired, but to a much lower extent. Nevertheless, miRNAs related to AML and CLL pathogenesis, such as has-miR-155, has-miR-181, and hsa-miR-15/16, were still removed from coordinated control.

[00142] Also, while not wishing to be bound by theory, the inventors herein now believe that the dissimilar behavior of solid cancers and leukemia may be due to the diverging pathogenetic mechanisms, which include differing oncogenic miRNA networks. In the former, complex chromosomal aberrations are frequent, whereas in the latter, translocations often represent the major driving force.

[00143] Overall, miRNA networks in cancer cells defined independently regulated miRNAs.

The target genes of these uncoordinated miRNA were involved in specific cancer-related pathways

[00144] EXAMPLE 2

[00145] Methods

[00146] miRNA expression Arrays

[00147] Microarray analysis was performed as previously described (Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M. et al. 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. Proceedings of the National Academy of Sciences of the United States of America 103: 2257-2261). Briefly, 5 mg of total RNA were used for hybridization of miRNA microarray chips. These chips contain gene-specific oligonucleotide probes, spotted by contacting technologies and covalently attached to a polymeric matrix. The microarrays were hybridized in 63 SSPE (0.9 M NaCl, 60 mM NaH₂PO₄ • H₂O, 8 mM EDTA at pH 7.4), 30% formamide at 25°C for 18 h, washed in 0.753 TNT (Tris-HCl, NaCl, Tween 20) at 37°C for 40 min, and processed by using a method of detection of the biotin-containing transcripts by streptavidin-Alexa647 conjugate. Processed slides were scanned using a microarray scanner (Axon), with the laser set to 635 nm, at fixed PMT setting, and a scan resolution of 10 mm. Microarray images were analyzed by using GenePix Pro and post-processing was performed essentially as described in Volinia et al. (2006). Briefly, average values of the replicate spots of each miRNA were background-subtracted and subject to further analysis. miRNAs were retained, when present, in at least 20% of samples and when at least 20% of the miRNA had a fold change of more than 1.5 from the gene median. Absent calls were thresholded prior to normalization and statistical analysis. Normalization was performed by using the quantiles method. MiRNA nomenclature was according to the miRNA database at Sanger Center.

[00148] Data analysis

[00149] An SQL miRNA internal database was built with the data retrieved from a large number of different experiments performed in our laboratory. Briefly, the F635-background values were used. Bad spots were removed. Nonexpressed spots were averaged for each gpr files (chip). For each mature miRNA, the inventors computed the geometric mean of its multiple reporters in the chip. A NaN value was assigned to miRNAs with more than 50% of corrupted spots, as reported by the GenePix image analysis software. All the results were log₂-transformed. The normalization was performed by using the quantiles normalization, as implemented in Bioconductor “affy” package (Bolstad et al. 2003). BRB Arraytools was used to perform t-test over two classes’ experiments of F-tests over multiple classes (i.e., different normal tissues). Target genes selection was performed by DIANA-miRpath, microT-V4. The union of the target mRNAs with a score >3 was used as an input to ClueGO.

[00150] ClueGO was used to relate differential expression in cancer to functional pathways (KEGG). ClueGO visualizes the selected terms in a functionally grouped annotation network that

reflects the relationships between the terms based on the similarity of their associated genes. The size of the nodes reflects the statistical significance of the terms. The degree of connectivity between terms (edges) is calculated using kappa statistics. The calculated kappa score is also used for defining functional groups. A term can be included in several groups. The reoccurrence of the term is shown by adding "n." The not grouped terms are shown in white color. The group leading term is the most significant term of the group. The network integrates only the positive kappa score term associations and is automatically laid out using Organic layout algorithm supported by Cytoscape. Right-sided hypergeometric test yielded the enrichment for GO-terms. Benjamini-Hochberg correction for multiple testing controlled the P-values.

[00151] Network generation and clustering

[00152] Banjo was used to infer the Bayesian network for the different tissues and diseases. For each tissue or disease all the mature expressed and varying miRNAs were used as input to Banjo. The expression values were preprocessed with Gene Pattern to only filter out nonvarying miRNAs, according to the following parameters: #filter.flag = filter (variation filter and thresholding flag); #preprocessing.flag = no disk or norm (discretization and normalization flag); #minchange = 10 (minimum fold change for filter); #mindelta = 512 (minimum delta for filter); #threshold = 64 (value for threshold); #ceiling = 20,000 (value for ceiling); #max. sigma.binning = 1 (maximum sigma for binning); #prob.thres = 1 (value for uniform probability threshold filter); #num.excl = 2% of total chips (number of experiments to exclude (max and min) before applying variation filter); #log.base.two = no (whether to take the log base two after thresholding); #number.of.columns.above.threshold = 1% of total chips (remove row if n columns not $\geq$ than given threshold above.threshold); and #column.threshold = 512 (threshold for removing rows).

[00153] The inventors then performed a quality control step to remove chips with abnormal expression distribution across miRNAs: Chips were retained only if less than 25% of miRNAs were absent (expression value < 64). Similarly, miRNAs were retained only when less than 25% of samples had absent expression (value < 64). The static Bayesian network inference algorithm was run on the miRNA expression matrix by using standard parameters, with a discretization policy of q6. Consensus graphs, based on top 100 networks, were obtained from at least 8 3 109 searched networks. The inventors applied the MCL graph-based clustering algorithm to extraction of clusters (i.e., groups of densely connected nodes) from miRNA networks. MCL (Neat) has been shown to enable good performances in extracting coregulated genes from transcriptome networks. yEd graph editor (yFiles software, Tubingen, Germany) was employed for graphs visualization.

[00154] miRNA family array CGH I

[00155] Seven hundred forty-four comparative genomic hybridization arrays were studied (537 samples from GEO and 207 from SMD). All platforms were two-channel based, data were downloaded as normalized values, and genes were annotated according to the gene symbol. All

normalized log ratios were converted to log2 ratios, with the cancer value at the numerator and the control value at the denominator. Bootstrap analysis was used (10,000 random swaps of cancer and control channels) to obtain P-values and confidence limits for deletion and amplifications.

- [00156] The inventors investigated 306 miRNA loci; 168 miRNA loci were associated to a host gene, and 138 miRNA loci to flanking genes. miRNA families were defined according to TargetScan. The threshold P-value for a miRNA family was set at 0.05 to the number of family members, n (0.05^n). To control for multiple testing, the inventors performed 100 bootstrapping cycles and used the results to calculate the false discovery rate (FDR).
- [00157] The resampling analysis was executed by randomly assigning the original P-values to the miRNA loci, while all family structures and chromosomal locations were kept unchanged. The FDR was defined as the percentage of families in the simulation evaluating better (lower P-values) than in the original test. Since the number of family member was variable (from a minimum of 2 to 7), FDRs were computed for each family according to its size (n , number of miRNA members).
- [00158] EXAMPLE III
- [00159] Reprogramming of the miRNA networks in cancer and leukemia
- [00160] Datasets
- [00161] The expression data have been submitted to the NCBI Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession Nos.
- [00162] GSE8126 and GSE7055 (Ambs et al. 2008 Ambs, S., Prueitt, R.L., Yi, M., Hudson, R.S., Howe, T.M., Petrocca, F., Wallace, T.A., Liu, C.G., Volinia, S., Calin, G.A. et al. 2008. Genomic profiling of microRNA and messenger RNA reveals deregulated microRNA expression in prostate cancer. *Cancer research* 68: 6162-6170),
- [00163] GSE17155 (Fassan, M., Baffa, R., Palazzo, J.P., Lloyd, J., Crosariol, M., Liu, C.G., Volinia, S., Alder, H., Rugge, M., Croce, C.M. et al. 2009. MicroRNA expression profiling of male breast cancer. *Breast Cancer Res* 11: R58),
- [00164] GSE3467 (He, H., Jazdzewski, K., Li, W., Liyanarachchi, S., Nagy, R., Volinia, S., Calin, G.A., Liu, C.G., Franssila, K., Suster, S. et al. 2005. The role of microRNA genes in papillary thyroid carcinoma. *Proceedings of the National Academy of Sciences of the United States of America* 102: 19075-19080),
- [00165] GSE7828 (Schetter, A.J., Leung, S.Y., Sohn, J.J., Zanetti, K.A., Bowman, E.D., Yanaihara, N., Yuen, S.T., Chan, T.L., Kwong, D.L., Au, G.K. et al. 2008. MicroRNA expression profiles associated with prognosis and therapeutic outcome in colon adenocarcinoma. *Jama* 299: 425-436),
- [00166] GSE6857 (Budhu, A., Jia, H.L., Forgues, M., Liu, C.G., Goldstein, D., Lam, A., Zanetti, K.A., Ye, Q.H., Qin, L.X., Croce, C.M. et al. 2008. Identification of metastasis-related microRNAs in hepatocellular carcinoma. *Hepatology* (Baltimore, Md 47: 897-907),

- [00167] GSE16654 (Chin, M.H., Mason, M.J., Xie, W., Volinia, S., Singer, M., Peterson, C., Ambartsumyan, G., Aimiwu, O., Richter, L., Zhang, J. et al. 2009. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell stem cell* 5: 111-123), and
- [00168] GSE14936 (Seike, M., Goto, A., Okano, T., Bowman, E.D., Schetter, A.J., Horikawa, I., Mathe, E.A., Jen, J., Yang, P., Sugimura, H. et al. 2009. MiR-21 is an EGFR-regulated anti-apoptotic factor in lung cancer in never-smokers. *Proceedings of the National Academy of Sciences of the United States of America* 12085-12090),
- [00169] and to ArrayExpress (ebi.ac.uk/microarray-as/ae) under accession Nos.:
- [00170] E-TABM-866 (Pineau, P., Volinia, S., McJunkin, K., Marchio, A., Battiston, C., Terris, B., Mazzaferro, V., Lowe, S.W., Croce, C.M., and Dejean, A. miR-221 overexpression contributes to liver tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America* 107: 264-269),
- [00171] E-TABM-664 (Bloomston, M., Frankel, W.L., Petrocca, F., Volinia, S., Alder, H., Hagan, J.P., Liu, C.G., Bhatt, D., Taccioli, C., and Croce, C.M. 2007. MicroRNA expression patterns to differentiate pancreatic adenocarcinoma from normal pancreas and chronic pancreatitis. *Jama* 297: 1901-1908),
- [00172] E-TABM-762 and E-TABM-763 (Visone, R., Rassenti, L.Z., Veronese, A., Taccioli, C., Costinean, S., Aguda, B.D., Volinia, S., Ferracin, M., Palatini, J., Balatti, V. et al. 2009. Karyotype-specific microRNA signature in chronic lymphocytic leukemia. *Blood* 114: 3872-3879),
- [00173] E-TABM-508 (Pichiorri, F., Suh, S.S., Ladetto, M., Kuehl, M., Palumbo, T., Drandi, D., Taccioli, C., Zanesi, N., Alder, H., Hagan, J.P. et al. 2008. MicroRNAs regulate critical genes associated with multiple myeloma pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 12885-12890),
- [00174] E-TABM-429 (Garzon, R., Garofalo, M., Martelli, M.P., Briesewitz, R., Wang, L., Fernandez-Cymering, C., Volinia, S., Liu, C.G., Schnittger, S., Haferlach, T. et al. 2008a. Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin. *Proceedings of the National Academy of Sciences of the United States of America* 105: 3945-3950),
- [00175] E-TABM-434 (Petrocca, F., Visone, R., Onelli, M.R., Shah, M.H., Nicoloso, M.S., de Martino, I., Iliopoulos, D., Piloizzi, E., Liu, C.G., Negrini, M. et al. 2008. E2F1-regulated microRNAs impair TGFbeta-dependent cellcycle arrest and apoptosis in gastric cancer. *Cancer cell* 13: 272-286),
- [00176] E-TABM-405 (Garzon, R., Volinia, S., Liu, C.G., Fernandez-Cymering, C., Palumbo, T., Pichiorri, F., Fabbri, M., Coombes, K., Alder, H., Nakamura, T. et al. 2008b. MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. *Blood* 111: 3183-

3189,

- [00177] E-TABM-343 (Iorio, M.V., Visone, R., Di Leva, G., Donati, V., Petrocca, F., Casalini, P., Taccioli, C., Volinia, S., Liu, C.G., Alder, H. et al. 2007. MicroRNA signatures in human ovarian cancer. *Cancer research* 67: 86998707),
- [00178] E-TABM-41 and E-TABM-42 (Calin, G.A., Ferracin, M., Cimmino, A., Di Leva, G., Shimizu, M., Wojcik, S.E., Iorio, M.V., Visone, R., Sever, N.I., Fabbri, M. et al. 2005. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *The New England journal of medicine* 353: 1793-1801),
- [00179] E-TABM-48 (Roldo, C., Missiaglia, E., Hagan, J.P., Falconi, M., Capelli, P., Bersani, S., Calin, G.A., Volinia, S., Liu, C.G., Scarpa, A. et al. 2006. MicroRNA expression abnormalities in pancreatic endocrine and acinar tumors are associated with distinctive pathologic features and clinical behavior. *J Clin Oncol* 24: 4677-4684),
- [00180] E-TABM-22 (Yanaihara, N., Caplen, N., Bowman, E., Seike, M., Kumamoto, K., Yi, M., Stephens, R.M., Okamoto, A., Yokota, J., Tanaka, T. et al. 2006. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer cell* 9: 189-198),
- [00181] E-TABM-23 (Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M. et al. 2005. MicroRNA gene expression deregulation in human breast cancer. *Cancer research* 65: 7065-7070),
- [00182] ETABM-46, E-TABM-47, E-TABM-49, and E-TABM-50 (Volinia, S., Calin, G.A., Liu, C.G., Ambis, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M. et al. 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proceedings of the National Academy of Sciences of the United States of America* 103: 2257-2261), and
- [00183] E-MEXP-1796 (Godlewski, J., Nowicki, M.O., Bronisz, A., Williams, S., Otsuki, A., Nuovo, G., Raychaudhury, A., Newton, H.B., Chiocca, E.A., and Lawler, S. 2008. Targeting of the Bmi-1 oncogene/stem cell renewal factor by microRNA-128 inhibits glioma proliferation and self-renewal. *Cancer research*, 68: 9125-913),
- [00184] E-TABM-37 (Ciafre, S.A., Galardi, S., Mangiola, A., Ferracin, M., Liu, C.G., Sabatino, G., Negrini, M., Maira, G., Croce, C.M., and Farace, M.G. 2005. Extensive modulation of a set of microRNAs in primary glioblastoma. *Biochemical and biophysical research communications* 334: 1351-1358),
- [00185] E-TABM-341 (Ueda, T., Volinia, S., Okumura, H., Shimizu, M., Taccioli, C., Rossi, S., Alder, H., Liu, C.G., Oue, N., Yasui, W. et al. Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis. *Lancet Oncol* 11: 136-146).
- [00186] The additional samples we used here are unpublished or have been previously reported

in other context-specific papers, such as:

- [00187] Baffa, R., Fassan, M., Volinia, S., O'Hara, B., Liu, C.G., Palazzo, J.P., Gardiman, M., Rugge, M., Gomella, L.G., Croce, C.M. et al. 2009. MicroRNA expression profiling of human metastatic cancers identifies cancer gene targets. *The Journal of pathology* 219: 214-221;
- [00188] Garzon, R., Pichiorri, F., Palumbo, T., Iuliano, R., Cimmino, A., Aqeilan, R., Volinia, S., Bhatt, D., Alder, H., Marcucci, G. et al. 2006. MicroRNA fingerprints during human megakaryocytopoiesis. *Proceedings of the National Academy of Sciences of the United States of America* 103: 50785083;
- [00189] Garzon, R., Pichiorri, F., Palumbo, T., Visentini, M., Aqeilan, R., Cimmino, A., Wang, H., Sun, H., Volinia, S., Alder, H. et al. 2007. MicroRNA gene expression during retinoic acid-induced differentiation of human acute promyelocytic leukemia. *Oncogene* 26: 4148-4157; and
- [00190] Zhang, L., Volinia, S., Bonome, T., Calin, G.A., Greshock, J., Yang, N., Liu, C.G., Giannakakis, A., Alexiou, P., Hasegawa, K. et al. 2008. Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer. *Proceedings of the National Academy of Sciences of the United States of America* 105: 7004-7009.
- [00191] Results and Discussion
- [00192] The miRNA specificity and network in normal tissues and ES differentiation.
- [00193] Embryonic cell types, such as embryonic bodies, trophoblasts, endoderm, or other stem cells including induced pluripotent stem cells (iPS) also harbored high levels of hsa-miR-302. Since hsa-miR-302 expression only partially decreased throughout these stages of ES differentiation, its information content (IC) in embryonic tissues was low (0.5). The miRNAs which had highest tissue specificity in embryos were: hsa-miR-211 in 14 day embryoid bodies (EBs), hsa-miR-10b, hsa-miR-218, hsa-miR-122, and hsa-miR-148a in spontaneous differentiated monolayers, hsa-miR-138 and hsa-miR-338-3p in 7 day EBs, and hsa-miR-99a in trophoblast (**Figure 10 and Supplemental Table III**).
- [00194] The information content (IC) of hsa-miR-302 genes is in the datasets even higher than that found by Landgraf, who considered a lower number of tissues/organs/systems: After hsa-miR-302, the most specific miRNAs are hsa-miR-338-5p, hsa-miR-323-3p, hsa-miR-335, with highest expression in epidermis/nervous system, nervous system and breast respectively (**Figure 9 and Supplemental Table I**).
- [00195] hsa-miR-122 is the miRNA with the highest IC difference between ours and the sequencing dataset. hsa-miR-371-5p is here well represented in embryonic cell types, since we have high hsa-miR-371-5p expression in embryoid bodies, not assayed in the clones' dataset. hsa-miR-129-3p is specific for nervous system in both datasets. The inventors detected hsa-miR-142-5p in hematopoietic system and connective tissues. hsa-miR-9, hsa-miR-128 and hsa-miR-138 were nervous system specific in both datasets. The inventors now report twice as many different tissue

groups that in the clones' dataset, thus the inventors expected less tissue specific miRNAs and lower information contents.

[00196] This is confirmed by the IC table (**Supplemental Table I**). Nevertheless, the IC distributions are very similar: 25 miRNAs with IC higher than 1.35 for the microarrays, in comparison to 29 for the clones. The only notable exception to the lower IC is hsa-miR-302a/b/c. This variation might be due to the fact that we assayed different cell types of embryonic origin, in comparison to only ES cells for the clones.

[00197] Methods

[00198] Tissue specificity. First, all samples were classified according to their organ-, tissue- and cell-type; then the normal samples were grouped in specific systems and the disease samples in specific pathological states. To assess the specificity of miRNA expression across groups, the inventors estimated what fraction of the total, for a given miRNA belonged to each single group. Therefore, the inventors used the procedure described in the first miRNA expression atlas (Landgraf et al. 2007), but modified by what the inventors called Em , the value of miRNA m in the group t referred to as "mean expression value (subtracted of the background value, 100). From here onwards, the inventors proceeded as the reference stated.

[00199] The specificity score, which varies between 0, when the expression level of the miRNA m is the same across all tissues, and $\log_2$ (number of tissue types), when only one tissue expresses the miRNA. To minimize artifacts from miRNAs or tissues with very small expression levels, the inventors considered only miRNAs with a total expression value above 10 times the number of normal tissues and above 100 times the number of cancer types; with a minimal expression value (after background subtraction) of 100. For the calculation of overall specificity, the thus included 130 and 133 different mature miRNAs for normal tissues and cancer, respectively.

[00200] Array CGH.

[00201] Each cancer sample was compared to a healthy control on a two channel normalized $\log_2$ ratio (cancer over control). Different probes related to the same gene were averaged (gene symbols were used as keys). Data were normalized according to the providers. As a pre-processing step the inventors retained only those genes with high variability (standard deviation > 0.2). For each gene the inventors computed the 5th and 95th percentiles (only for genes measured in at least 300 samples). A gene harboring recurrent deletions in tumors would result in a low 5th percentile $\log_2$ ratio (negative), while one with amplifications would display a high 95th percentile (positive).

[00202] To identify the miRNAs with structural alterations in cancer, the inventors followed the following procedure (illustrated here for amplifications): the inventors selected all miRNA families where at least 1 member was significantly amplified ($p < 0.05$).

[00203] The inventors defined the family p-value for amplification, as the product of the amplification p-values for each family member (including also the non significant miRNAs), with

the following exceptions : i) replicated identical mature miRNAs were considered only when mapping on different loci (i.e. represented by different host genes or flanking genes); ii) physically clustered family members were scored only once.

[00204] The distinct human genes we assayed in the aCGH dataset were 19,654 in total. The inventors studied 530 distinct miRNA precursors in 308 chromosomal loci, corresponding to 471 different mature miRNAs and 356 distinct miRNA families. The average distance between the miRNAs and their flanking genes was of 188 Kb for the 5' and of 240 Kb for the 3' gene. The number of miRNA loci deleted/amplified in cancer was not significantly different from expectation, when compared to the whole coding genome (in both cases $p \gg 0.05$).

[00205] Bootstrap analysis (random swap between cancer and control channels) was used to simulate gene specific 5th and 95th percentiles. Gene-specific p-values for deletions were calculated as the percentage of resampled 5th percentiles which exceeded the original 5th percentile.

[00206] The inventors took in consideration two phenomena, associated to aCGH, but not linked to cancer: sex chromosomes and polymorphic copy number variations (CNV). Since the control sample was more frequently from male, while roughly half of the tumors were of female origin, the Y-chromosome genes were incorrectly expected to appear as deleted. Conversely, the inventors expected the X chromosome genes, except for those belonging to the pseudo-autosomal region, to incorrectly appear as amplified. Genes located in the sex chromosomes were indeed behaving exactly as expected (data not shown). Polymorphic CNVs could also display large fold-changes, resulting in high 95th or low 5th percentiles. The inventors herein found that such CNVs, not associated to cancer, would not display significant p-values. Indeed, most polymorphic sites for copy number variations (CNVs) did not filtered through the aCGH assay, since the different alleles were balanced in the cancer and control groups. Only a small percentage of miRNA coincided with polymorphic CNVs and that fraction was not enriched in the cancer subset.

[00207]

Table 1. Differentially regulated miRNA in solid cancers

Parametric P-value	FDR	Intensities in solid cancers	Intensities in normal tissues	Fold-change	miRNA	Chromosomal location
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	967	617.9	1.57	hsa-miR-21	17q23.1
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	1378.2	917.7	1.5	hsa-miR-25	7q22.1
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	902.7	626.3	1.44	hsa-miR-20a	13q31.3
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	925.7	646.9	1.43	hsa-miR-17	13q31.3
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	652.3	469	1.39	hsa-miR-106a	Xq26.2
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	410	297.8	1.38	hsa-miR-106b	7q22.1
1.30×10^{-6}	1.1×10^{-5}	918.8	697.9	1.32	hsa-miR-146a	5q34
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	11893.3	9370.9	1.27	hsa-miR-92a	13q31.3,Xq26.2
1.60×10^{-6}	1.2×10^{-5}	2354.9	1919.4	1.23	hsa-miR-103	5q35.1,20p13
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	289.4	237.8	1.22	hsa-miR-130b	22q11.21
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	452.7	372	1.22	hsa-miR-93	7q22.1
3.90×10^{-6}	2.7×10^{-5}	2116.4	1743.6	1.21	hsa-miR-107	10q23.31
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	297.7	248	1.2	hsa-miR-30e	1p34.2
7.4×10^{-6}	4.8×10^{-5}	225.6	259.4	0.87	hsa-miR-138	16q13.3,p21.33
1×10^{-7}	1×10^{-6}	220.5	260.2	0.85	hsa-miR-326	11q13.4
3.7×10^{-6}	2.6×10^{-5}	419.9	507.8	0.83	hsa-miR-193a	17q11.2
2×10^{-7}	1.8×10^{-6}	382.5	471.2	0.81	hsa-miR-206	6p12.2
3×10^{-6}	2.2×10^{-5}	273.7	349.6	0.78	hsa-miR-205	1q32.2
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	714.9	1015.1	0.70	hsa-miR-145	5q32
$<1 \times 10^{-7}$	$<1 \times 10^{-7}$	275.7	403.0	0.68	hsa-miR-203	14q32.33

Solid cancer samples numbering 2532 vs. 806 corresponding normal samples, at least one class with intensity >250 , P-value $<1 \times 10^{-5}$.

[00208]

Supplemental Table I. Expression levels of tissue specific microRNAs in normal tissues.
Sorted by information content (IC)

Name	Tissue Specificity	Expression Value	Information Content	Chromosomal Location
hsa-miR-302a	EMBRYO	4380	4.00	4:113569339-113569407
hsa-miR-302c	EMBRYO	1709	3.99	4: 113569519-113569586
hsa-miR-302b	EMBRYO	4929	3.73	4: 113569641-113569713
hsa-miR-338-5p	EPIDERMIS / NERVOUS SYSTEM	573 / 451	2.51	17: 79099683-79099749
hsa-miR-323-3p	NERVOUS SYSTEM	167	2.47	14: 101492069-101492154
hsa-miR-335	BREAST	322	2.37	7: 130135952-130136045
hsa-miR-124	NERVOUS SYSTEM	533	2.33	8:9,798,308-9,798,392 8:65,454,260-65,454,368
hsa-miR-139-5p	NERVOUS SYSTEM	245	2.27	11: 72326107-72326174
hsa-miR-371-5p	EMBRYO	443	1.93	19: 54290929-54290995
hsa-miR-325	EPIDERMIS	129	1.91	X: 76225829-76225926
hsa-miR-133a	CARDIOVASCULAR SYSTEM / SKELETAL MUSCLE	464/ 661	1.88	18:17,659,657-17,659,744 20:60,572,564-60,572,665
hsa-miR-184	RESPIRATORY SYSTEM	259	1.86	15: 79502130-79502213
hsa-miR-133b	CARDIOVASCULAR SYSTEM / SKELETAL MUSCLE	394 / 371	1.83	6: 52013721-52013839
hsa-miR-211	NERVOUS SYSTEM	180	1.78	15: 31357235-31357344
hsa-miR-9	NERVOUS SYSTEM	1346	1.73	1:154656757-154656845 5:87998427-87998513
hsa-miR-370	NERVOUS SYSTEM	146	1.72	14: 101377476-101377550
hsa-miR-376c	EPIDERMIS	121	1.61	14: 101506027-101506092
hsa-miR-132	NERVOUS SYSTEM	552	1.59	17: 1953202-1953302
hsa-miR-1	SKELETAL MUSCLE / CARDIOVASCULAR SYSTEM	500 / 350	1.56	20:60,561,958-60,562,028 18:17,662,963-17,663,047
hsa-miR-215	KIDNEY	2630	1.55	1: 220291195-220291304
hsa-miR-339-5p	LIVER	107	1.54	7: 1062569-1062662
hsa-miR-330-3p	LIVER	125	1.53	19: 46142252-46142345
hsa-miR-342-5p	EPIDERMIS	326	1.49	14: 100575992-100576090
hsa-miR-199b-5p	RESPIRATORY SYSTEM	144	1.39	9: 131007000-131007109
hsa-miR-129-3p	NERVOUS SYSTEM	266	1.36	11: 43602944-43603033

[00209]

Supplemental Table II. Normal Tissues grouped by systems

Normal Organism Part	Normal System	N of Samples
ADIPOSE TISSUE	Adipose	5
BREAST	Breast	4
HEART	Cardiovascular System	6
DERMAL FIBROBLAST	Connective Tissue	11
EMBRYONIC STEM CELLS / DEFINITIVE ENDODERM / SPONTANEOUS DIFFERENTIATED MONOLAYERS / EMBRYONIC BODIES / INDUCED PLURIPOTENT STEM CELLS/	Embryo	68
THYROID / PANCREAS / PITUITARY /	Endocrine Gland	107
SKIN	Epidermis	20
ESOPHAGUS / COLON / STOMACH / RECTUM / SIGMA / SMALL INTESTINE	Gastrointestinal System	394
BLOOD / BONE MARROW / LYMPHONODE / THYMUS / TONSIL / SPLEEN	Hematopoietic System	225
KIDNEY	Kidney	19
LIVER	Liver	38
OCCIPITAL CORTEX / TEMPORAL CORTEX / PRIMARY VISUAL CORTEX / THALAMUS / BRAIN / MEDULLA / CEREBELLUM / PREFRONTAL CORTEX / AUDITORY CORTEX / AMYGDALA/ HYPOTHALAMUS / MOTOR CORTEX / CAUDATE NUCLEUS / PARIETAL CORTEX	Nervous System	58
PLACENTA	Placenta	6
FALLOPIAN TUBE / CERVIX / OVARY / PROSTATE / TESTIS	Reproductive System	59
LUNG / TRACHEA	Respiratory System	75
SKELETAL MUSCLE	Skeletal Muscle	6
BLADDER	Urinary System	6

[00210]

Supplemental Table III. Expression levels of tissue specific microRNAs during differentiation of embryonic stem cells (ES) . Sorted by information content (IC)

Name	Embryonic tissue Specificity	Expression Value	IC	Chromosomal Location
hsa-miR-211	EB 14 day	165	2.65	15: 31357235-31357344
hsa-miR-10b	Monolayer	909	2.20	2: 177015031-177015140
hsa-miR-138	EB 7 day	181	1.88	3:44,130,708-44,130,806
hsa-miR-218	Monolayer	665	1.65	16:55,449,931-55,450,014
hsa-miR-122	Monolayer	3660	1.57	4:20,138,996-20,139,105
hsa-miR-99a	Trophoblast	197	1.54	5:168,127,729-168,127,838
hsa-miR-215	Monolayer /Definitive Endoderm	1673 / 325	1.42	18: 56118306-56118390
hsa-miR-338-3p	EB 7 day	432	1.26	21: 17911409-17911489
let-7e	iPS / Embryonic Stem Cells	284 / 181	1.23	1: 220291195-220291304
hsa-miR-371-5p	EB 14 day / EB 7 day / ES	1424 / 6274 / 848	1.20	17: 79099683-79099749
hsa-miR-148a	Monolayer	158	1.18	19: 52196039-52196117
hsa-miR-181c	Trophoblast / Monolayer	1500 / 927	1.07	19: 54290929-54290995
hsa-miR-203	EB 14 day / Monolayer / Trophoblast	974 / 750 / 934	1.05	7: 25989539-25989606
hsa-miR-181a	Trophoblast / Monolayer	2943 / 1106	0.99	19: 13985513-13985622
hsa-miR-181b	Trophoblast / Monolayer	8290 / 3397	0.93	14: 104583742-104583851
hsa-miR-192	Monolayer / Definitive Endoderm	1804 / 474	0.84	1:197,094,796-197,094,905
hsa-miR-150	EB 7 day / iPS	501 / 347	0.83	9:126,494,542-126,494,651
hsa-miR-139-3p	iPS / ES	477 / 250	0.82	1:197,094,625-197,094,734
hsa-miR-27b	Trophoblast / Monolayer	554 / 317	0.77	9:126,495,810-126,495,898
hsa-miR-10a	Monolayer	2547	0.75	11: 64658609-64658718

Legend: Embryonic Bodies: EB; Spontaneous Differentiated Monolayer: Monolayer; Induced Pluripotent Stem Cells: iPS; Embryonic Stem Cells: ES.

[00211]

Supplemental Table IV. Up-regulated microRNAs in 31 types of solid cancers (2532 cancers samples vs. 806 corresponding normal samples)

Parametric p-value	FDR	intensities in Solid Cancers	intensities in Normal Tissues	Fold change-	microRNA	Chromosomal location
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	967	617.9	1.57	hsa-miR-21	17q23.1
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	1378.2	917.7	1.5	hsa-miR-25	7q22.1
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	902.7	626.3	1.44	hsa-miR-20a	13q31.3
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	925.7	646.9	1.43	hsa-miR-17	13q31.3
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	652.3	469	1.39	hsa-miR-106a	Xq26.2
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	410	297.8	1.38	hsa-miR-106b	7q22.1
1.30×10^{-08}	1.08×10^{-05}	918.8	697.9	1.32	hsa-miR-146a	5q34
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	11893.3	9370.9	1.27	hsa-miR-92a	13q31.3,Xq26.2
1.60×10^{-08}	1.29×10^{-05}	2354.9	1919.4	1.23	hsa-miR-103	5q35.1,20p13
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	289.4	237.8	1.22	hsa-miR-130b	22q11.21
6.81×10^{-05}	0.000387	750.4	615.7	1.22	hsa-miR-30c	1p34.2,6q13
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	452.7	372	1.22	hsa-miR-93	7q22.1
3.90×10^{-08}	2.74×10^{-05}	2116.4	1743.6	1.21	hsa-miR-107	10q23.31
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	297.7	248	1.2	hsa-miR-30e	1p34.2
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	227	191.9	1.18	hsa-miR-15a	13q14.2
0.000433	0.002	1209.2	1031.6	1.17	hsa-miR-181b	1q32.1,9q33.3
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	191.3	164.8	1.16	hsa-miR-15b	3q25.33
0.000605	0.002648	508.2	438.5	1.16	hsa-miR-181a	1q32.1,9q33.3
0.000442	0.002002	333.8	287	1.16	hsa-miR-32	9q31.3
5.10×10^{-08}	3.40×10^{-05}	160	147.3	1.09	hsa-miR-345	14q32.2
0.000494	0.0022	208.1	193.5	1.08	hsa-miR-34a	1p36.22
0.000864	0.00372	154.1	146.2	1.05	hsa-miR-374a	Xq13.2

*Intensities (expression levels) are reported as geometric means after quantiles normalization.

[00212]

- **Supplemental Table V.** Down-regulated microRNAs in 31 types of solid cancers (2532 cancer samples vs. 806 corresponding normal samples)

Parametric p-value	FDR	Intensities in Solid Cancers	Intensities in Normal Tissues	Fold Change	microRNA	Chromosomal location
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	275.7	403	0.68	hsa-miR-203	14q32.33
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	714.9	1015.1	0.7	hsa-miR-145	5q32
0.000003	2.23×10^{-05}	273.7	349.6	0.78	hsa-miR-205	1q32.2
2.00×10^{-07}	1.80×10^{-06}	382.5	471.2	0.81	hsa-miR-206	6p12.2
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	204	247.9	0.82	hsa-miR-33b	17p11.2
3.70×10^{-08}	2.67×10^{-05}	419.9	507.8	0.83	hsa-miR-193a	17q11.2
0.000042	0.000249	175.1	208.5	0.84	hsa-miR-204	9q21.12
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	173.9	204.2	0.85	hsa-miR-143	5q32
1.00×10^{-07}	0.000001	220.5	260.2	0.85	hsa-miR-326	11q13.4
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	150.2	176.9	0.85	hsa-miR-338	17q25.3
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	164.8	193.6	0.85	hsa-miR-9	1q22.5q14.3, 15q26.1
0.000355	0.001789	190.6	220.4	0.86	hsa-miR-95	4p16.1
7.40×10^{-06}	4.82×10^{-05}	225.6	259.4	0.87	hsa-miR-138	16q13.3p21.33
0.000435	0.002	160.4	183.8	0.87	hsa-miR-183	7q32.2
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	150.4	172.9	0.87	hsa-miR-202	10q26.3
8.24×10^{-05}	0.000458	529.1	603.1	0.88	hsa-miR-128a	2q21.3,3p22.3
0.000921	0.003905	617	694.2	0.89	hsa-miR-214	1q24.3
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	139	153.9	0.9	hsa-miR-132	17p13.3
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	140.7	156.7	0.9	hsa-miR-299	14q32.31
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	140.7	154.8	0.91	hsa-miR-129	7q32.1,11p11.2
2.00×10^{-07}	1.80×10^{-06}	139.6	152.7	0.91	hsa-miR-133a	18q11.2,20q13.33
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	129.8	141.1	0.92	hsa-miR-139	11q13.4
8.80×10^{-08}	5.46×10^{-05}	153.4	165	0.93	hsa-miR-339	7p22.3
4.50×10^{-08}	3.08×10^{-05}	140	149.6	0.94	hsa-miR-1	20q13.33,18q11.2
9.68×10^{-05}	0.000528	140.4	149.6	0.94	hsa-miR-133b	6p12.2
$< 1 \times 10^{-07}$	$< 1 \times 10^{-07}$	129.7	137.8	0.94	hsa-miR-323	14q32.31
0.000394	0.001946	144.7	152.7	0.95	hsa-miR-218	4p15.31,5q34
0.000193	0.00101	130.2	133	0.98	hsa-miR-335	7q32.2

*Intensities (expression levels) are reported as geometric means after quantiles normalization.

[00213]

Supplemental Table VI. Expression levels of tissue specific microRNAs in cancer and leukemia

Name	Cancer Specificity	Expression	IC	Chromosomal location
hsa-miR-369-3p	OVARIAN ADENOCARCINOMA	2737	3.95	14:101531935-101532004
hsa-miR-325	PANCREAS NON-FUNCTIONING ENDOCRINE CANCER / INSULINOMA	1625 / 365	3.67	X: 76225829-76225926
hsa-miR-208	CC / BCC / EPIDERMIS SCC	256 / 244 / 418	3.36	14:22,927,645-22,927,715 14:22,957,036-22,957,112
hsa-miR-190	TPC / HEPATOBLASTOMA	211 / 220	3.04	15: 63116156-63116240
hsa-miR-184	AMoL / HEPATOBLASTOMA / TPC / MM	448 / 324 / 214 / 171	3.01	15: 79502130-79502213
hsa-miR-338-5p	EPIDERMIS SCC / BCC / CC	1928 / 808 / 472	2.70	17: 79099683-79099749
hsa-miR-105	HGSIL	234	2.57	X:151,311,347-151,313,620
hsa-miR-376c	LYMPHOMA CUTANEOUS / EPIDERMIS SCC / MM	1134 / 626 / 1223 / 525	2.50	14: 101506027-101506092
hsa-miR-302b	EPIDERMIS SCC / LYMPHOMA CUTANEOUS	336 / 110	2.48	4: 113569641-113569713
hsa-miR-299-5p	HEPATOBLASTOMA / AMoL / TPC	287 / 122 / 148	2.03	14: 101490131-101490193
hsa-miR-19a	LYMPHOMA CUTANEOUS / CML	388 / 367	1.83	13: 92003145-92003226
hsa-miR-200c	AMoL / HEPATOBLASTOMA	4460 / 4138	1.99	12: 7072862-7072929
hsa-miR-143	AMoL / HEPATOBLASTOMA	1043 / 1098	1.94	5: 148808481-148808586
hsa-miR-211	EPIDERMIS SCC / NSCLC	129 / 89	1.86	15: 31357235-31357344
hsa-miR-374a	BL SPORADIC / CML / APL	619 / 169 / 212	1.69	X: 73507121-73507192
hsa-miR-339-5p	BL ENDEMIC / BL SPORADIC	616 / 738	1.62	7: 1062569-1062662
hsa-miR-148a	LYMPHOMA CUTANEOUS	1161	1.64	7: 25989539-25989606
hsa-miR-370	BREAST, DUCTAL, LOBULAR, MALE CARCINOMA	157, 227, 239	1.69	14: 101377476-101377550
hsa-miR-199a-5p	HEPATOBLASTOMA	1122	1.52	19:10,789,102-10,789,172 1:170,380,298-170,380,407
hsa-miR-215	KIDNEY CARCINOMA / HEPATOBLASTOMA	7247 / 2461	1.45	1: 220291195-220291304
hsa-miR-19b	CML	232	1.33	13:90,801,447-90,801,533 X: 133,131,367-133,131,462
hsa-miR-148b	LYMPHOMA CUTANEOUS	657	1.39	12: 54731000-54731098
hsa-miR-18a	AMoL / FOLLICULAR ADENOMA	166 / 76	1.38	13: 92003005-92003075
hsa-miR-202	AMoL / HEPATOBLASTOMA	362 / 287	1.32	10: 135061015-135061124
hsa-miR-205	AMoL / HEPATOBLASTOMA / LUNG SCC	2029 / 2275 / 2665	1.31	1: 209605478-209605587

[00214]

Supplemental Table VII. Array CGH datasets

GEO datasets	# of samples	Cancer type	Platform
GSE4659	32	AML	GPL2873
GSE6472	6	Nasopharyngeal carcinoma	GPL2879
GSE7077	4	Osteosarcoma	GPL2879
GSE7344	16	Glioma	GPL2873
GSE7482	25	ACC	GPL2879
GSE7615	298	Pancreas, glioblastoma, T-ALL, melanoma, colon	GPL2879 GPL4091
GSE7822	14	Melanoma	GPL2879
GSE8398	25	Ewing Sarcoma	GPL2879
GSE8804	13	Myelodysplasia	GPL2879
GSE8918	87	Follicular Lymphoma, CLL, Mantle Cell Lymphoma, Nodal Marginal Zone Lymphoma, Lymphoplasmacytic Lymphoma, Splenic Marginal Zone Lymphoma, MALT	GPL2879
GSE9015	7	Breast	GPL4091
GSE9654	10	Osteosarcoma	GPL2879
SMD	207	Lung, Pancreas, Breast, Melanoma, Fibrous Histiocytoma	Stanford

- GPL2873 Agilent- Human Genome CGH Microarray 44A G4410A
- GPL2879 Agilent- Human Genome CGH Microarray 44B G4410B
- GPL4091 Agilent- Human Genome CGH Microarray 244A G4411B

Legend: Burkitt's Lymphoma: BL; Thyroid Papillary Cancer: TPC; Multiple Myeloma: MM; Squamous Cell Carcinoma: SCC; Basal Cell Carcinoma: BCC; Chronic Myelogenous Leukemia: CML; Acute Promyelocytic Leukemia: APL; High Grade Squamous Intraepithelial Lesion: HGSIL; Non-Small Cell Lung Cancer: NSCLC; Acute Monocytic Leukemia: AmoL; Cervix Carcinoma: CC.

[00215]

Supplemental Table VIII. miRNA families amplified in cancer

miRNA family	p-value	FDR	Amplified (p<0.05)	Family members
<i>hsa-miR-550</i>	1.25×10^{-09}	0.006	<i>hsa-miR-550-2</i> ; <i>hsa-miR-550-1</i>	2
<i>hsa-miR-9</i>	7.00×10^{-11}	0.012	<i>hsa-miR-9-1</i> ; <i>hsa-miR-9-2</i> ; <i>hsa-miR-9-3</i>	3
<i>hsa-miR-133</i>	2.50×10^{-09}	0.060	<i>hsa-miR-133a-2</i> ; <i>hsa-miR-133b</i>	3
<i>hsa-miR-153</i>	5.00×10^{-07}	0.062	<i>hsa-miR-153-1</i> ; <i>hsa-miR-153-2</i>	2
<i>hsa-miR-548d</i>	1.40×10^{-08}	0.073	<i>hsa-miR-548d-1</i> ; <i>hsa-miR-548d-2</i>	2
<i>hsa-miR-1/206</i>	7.55×10^{-10}	0.114	<i>hsa-miR-1-1</i> ; <i>hsa-miR-206</i>	4
<i>hsa-miR-146</i>	1.97×10^{-05}	0.116	<i>hsa-miR-146a</i>	2
<i>hsa-miR-148/152</i>	2.00×10^{-07}	0.133	<i>hsa-miR-152</i> ; <i>hsa-miR-148a</i>	3
<i>hsa-miR-196</i>	5.54×10^{-07}	0.150	<i>hsa-miR-196b</i> ; <i>hsa-miR-196a-2</i>	3
<i>hsa-miR-17/20/93/106/519</i>	3.70×10^{-06}	0.194	<i>hsa-miR-106b</i>	3
<i>hsa-miR-30-5p</i>	1.06×10^{-05}	0.226	<i>hsa-miR-30d</i>	3
<i>hsa-miR-192/215</i>	4.95×10^{-05}	0.239	<i>hsa-miR-192</i>	2
<i>hsa-miR-194</i>	4.95×10^{-05}	0.239	<i>hsa-miR-194-2</i>	2
<i>hsa-miR-551</i>	4.97×10^{-05}	0.239	<i>hsa-miR-551b</i>	2

* FDR < 0.3 (i.e. less than 30% of the identified miRNA families were expected to be false positives)

[00216]

Supplemental Table IX. miRNA families deleted in cancer

miRNA family	p-value	FDR	Deleted (p<0.05)	Family members
<i>hsa-miR-204/211</i>	6.25×10^{-10}	0.003	<i>hsa-miR-211</i> ; <i>hsa-miR-204</i>	2
<i>hsa-miR-103/107</i>	2.45×10^{-06}	0.064	<i>hsa-miR-107</i> ; <i>hsa-miR-103-2</i>	3
<i>hsa-miR-380-5p</i>	2.85×10^{-06}	0.109	<i>hsa-miR-563</i>	2
<i>hsa-miR-128</i>	3.25×10^{-06}	0.113	<i>hsa-miR-128a</i>	2
<i>hsa-miR-141/200a</i>	1.52×10^{-05}	0.170	<i>hsa-miR-200a</i>	2
<i>hsa-miR-200bc/429</i>	1.52×10^{-05}	0.170	<i>hsa-miR-429</i>	2
<i>hsa-miR-125/351</i>	3.50×10^{-07}	0.189	<i>hsa-miR-125b-1</i> ; <i>hsa-miR-125a</i>	3
<i>hsa-miR-218</i>	2.41×10^{-05}	0.193	<i>hsa-miR-218-1</i>	2
<i>hsa-miR-101</i>	2.70×10^{-05}	0.230	<i>hsa-miR-101-1</i> ; <i>hsa-miR-101-2</i>	2
<i>hsa-miR-99/100</i>	4.38×10^{-05}	0.241	<i>hsa-miR-99b</i>	2
<i>hsa-miR-23</i>	8.80×10^{-05}	0.290	<i>hsa-miR-23b</i> ; <i>hsa-miR-23a</i>	2
<i>hsa-miR-24</i>	8.80×10^{-05}	0.290	<i>hsa-miR-24-1</i> ; <i>hsa-miR-24-2</i>	2
<i>hsa-miR-27</i>	8.80×10^{-05}	0.290	<i>hsa-miR-27b</i> ; <i>hsa-miR-27a</i>	2

* FDR < 0.3 (i.e. less than 30% of the identified miRNA families were expected to be false positives)

[00217]

Supplemental Table X. Deregulated miRNAs in leukemia from Mir155 transgenic mice.

Parametric p-value	FDR	Intensity in hsa-miR-155 Leukemia	Intensity in Wt	Fold Change	miRNA
1.00x10 ⁻⁰⁶	2.31x10 ⁻⁰⁵	832.1	10.08	82.57	hsa-miR-224
4.20x10 ⁻⁰⁶	5.67x10 ⁻⁰⁵	3426.09	132.36	25.89	hsa-miR-9
1.00x10 ⁻⁰⁷	7.20x10 ⁻⁰⁶	21760.18	1005.44	21.64	hsa-miR-217
5.70x10 ⁻⁰⁶	7.39x10 ⁻⁰⁵	3387.97	219.75	15.42	hsa-miR-131
1.00x10 ⁻⁰⁷	7.20x10 ⁻⁰⁶	5334.39	375.03	14.22	hsa-miR-128b
1.00x10 ⁻⁰⁷	7.20x10 ⁻⁰⁶	31460.89	2922.63	10.76	hsa-miR-181b
2.00x10 ⁻⁰⁷	7.20x10 ⁻⁰⁶	17040.98	1615.73	10.55	hsa-miR-181c
2.00x10 ⁻⁰⁷	7.20x10 ⁻⁰⁶	18356.22	1976.64	9.29	hsa-miR-128a
1.28x10 ⁻⁰⁵	0.000148	759.87	2695.34	0.28	hsa-miR-140
0.000659	0.003556	228.42	874.66	0.26	hsa-miR-135a
3.69x10 ⁻⁰⁵	0.000332	8491.07	34327.47	0.25	hsa-miR-26a
0.000128	0.000984	528.06	2433.76	0.22	hsa-miR-10a
3.50x10 ⁻⁰⁶	5.55x10 ⁻⁰⁵	590.74	2779.66	0.21	hsa-miR-425
0.000193	0.001351	514.05	2467.2	0.21	hsa-miR-340
0.000474	0.002742	81.06	403.52	0.2	hsa-miR-24
2.00x10 ⁻⁰⁶	3.78x10 ⁻⁰⁵	656.68	3304.56	0.2	hsa-miR-140
0.000579	0.003178	30.24	196.77	0.15	hsa-miR-218
2.55x10 ⁻⁰⁵	0.000243	1006.31	6657.6	0.15	hsa-miR-29b
6.00x10 ⁻⁰⁷	1.62x10 ⁻⁰⁵	1007.33	7465.65	0.13	hsa-miR-29a
2.54x10 ⁻⁰⁵	0.000243	357.49	3183.78	0.11	hsa-miR-29c
< 1x10 ⁻⁰⁷	< 1x10 ⁻⁰⁷	2176.08	37081.81	0.06	hsa-miR-146
2.70x10 ⁻⁰⁶	4.60x10 ⁻⁰⁵	3339.45	59257.53	0.06	hsa-miR-150

[00218] Testing Systems

[00219] The microRNAs (miRNAs, miRs) are short RNA strands approximately 21-23 nucleotides in length. mRNAs are encoded by genes that are transcribed from DNA but not translated into protein (non-coding RNA). MiRs are processed from primary transcripts known as pri-miRNA to short stem-loop structures called pre-miRNA and finally to functional miRNA, as the precursors typically form structures that fold back on each other in self-complementary regions. The miRs are then processed by the nuclease Dicer in animals or DCL1 in plants. Mature miRNA molecules are partially complementary to one or more messenger RNA (mRNA) molecules. The sequences of miRNA can be accessed at publicly available databases, such as "microRNA.org."

[00220] A number of miRNAs are involved in gene regulation, and miRNAs are part of a growing class of non-coding RNAs that is now recognized as a major tier of gene control. In some

cases, miRNAs can interrupt translation by binding to regulatory sites embedded in the 3'-UTRs of their target mRNAs, leading to the repression of translation. Target recognition involves complementary base pairing of the target site with the miRNA's seed region (positions 2-8 at the miRNA's 5' end), although the exact extent of seed complementarity is not precisely determined and can be modified by 3' pairing. In other cases, miRNAs function like small interfering RNAs (siRNA) and bind to perfectly complementary mRNA sequences to destroy the target transcript.

[00221] In some embodiments, a single miRNA is assessed to characterize a cancer. In yet other embodiments, at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 1000, or more miRNAs are assessed. A change in the expression level, such as absence, presence, underexpression or overexpression of the miRNA as compared to a reference level, such as a level determined for a subject without the cancer, can be used to characterize a cancer for the subject.

[00222] For example, a reference level for classifying a cell as benign or malignant can include obtaining a sample from a subject, determining an amount of a miRNA in the subject's sample, and comparing the amount of the miRNA to one or more controls. The step of comparing the amount of the miRNA to one or more controls may include the steps of obtaining a range of the miRNA found in the sample for a plurality of subjects having a benign condition, or normal cells to arrive at a first control range, obtaining a range of the miRNA found in the sample for a plurality of subjects having malignant cancer cells to arrive at a second control range, and comparing the amount of the miRNA in the subject's sample with the first and second control ranges to determine if the subject's sample is classified as benign, or normal, or is a cancer.

[00223] Detection System and Kits

[00224] Also provided is a detection system configured to determine one or more RNAs for characterizing a cancer. For example, the detection system can be configured to assess 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 1000, 2500, 5000, 7500, 10,000, 100,000, 150,000, 200,000, 250,000, 300,000, 350,000, 400,000, 450,000, 500,000, 750,000, 1,000,000 or more miRNAs, wherein one or more of the miRNAs are selected from the data described herein.

[00225] The detection system can be a low density detection system or a high density detection system. For example, a low density detection system can detect up to about 100, 200, 300, 400, 500, or 1000 RNA, whereas a high density detection system can detect at least about 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9,000, 10,000, 15,000, 20,000; 25,000, 50,000, or 100,000 miRNAs.

[00226] The detection system can comprise a set of probes that selectively hybridizes to the one or more of the miRNAs. For example, the detection system can comprise a set of probes that selectively hybridizes to at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 1000, 2500, 5000, 7500, 10,000, 100,000, 150,000, 200,000, 250,000, 300,000, 350,000, 400,000, 450,000, 500,000, 750,000, or 1,000,000 miRNAs. For example, the set of probes can selectively hybridize to or more miRNAs selected from the data described herein.

[00227] The probes may be attached to a solid substrate, such as an array or bead. Alternatively, the probes are not attached. The detection system may be an array based system, a sequencing system, a PCR-based system, or a bead-based system.

[00228] The detection system may be part of a kit. Alternatively, the kit may comprise the one or more probe sets described herein. For example, the kit may comprise probes for detecting one or more of the miRNAs selected from data described herein.

[00229] Computer System

[00230] Also provided herein, is a computer system for characterizing a cancer. The computer system can include a logic device through which a phenotype profile and report may be generated. For example, the computer system (or digital device) can be configured to receive the expression level data from a biological sample, analyze the expression levels, determine a characteristic for a cancer (such as, but not limited to, classifying a cancer, determining whether a second sample should be obtained, providing a diagnosis, providing a prognosis, selecting a treatment, determining a drug efficacy), and produce the results, such as an output on the screen, printed out as a report, or transmitted to another computer system.

[00231] The computer system may be understood as a logical apparatus that can read instructions from a media and/or a network port, which can optionally be connected to a server having fixed media.

[00232] Data communication can be achieved through a communication medium to a server at a local or a remote location. The communication medium can include any means of transmitting and/or receiving data. For example, the communication medium can be a network connection, a wireless connection or an internet connection. Such a connection can provide for communication over the World Wide Web. It is envisioned that data relating to the present invention, such as the expression levels of the one or more miRNAs, the results of the analysis of the expression levels (such as the characterizing or classifying of the cancer), can be transmitted over such networks or connections for reception and/or review by a party. The receiving party can be, but is not limited, to a subject, a health care provider or a health care manager. In some embodiments, the information is stored on a computer-readable medium.

[00233] While the invention has been described with reference to various and preferred embodiments, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted for elements thereof without departing from the essential scope of the invention. In addition, many modifications may be made to adapt a particular situation or material to the teachings of the invention without departing from the essential scope thereof.

CLAIMS

What is claimed is:

1. Use of a miRNA networks comprised from miRNA expression data that includes a least one miRNA network from normal tissues, and/or at least one miRNA network for coupled cancerous and noncancerous tissues.
2. A method for identification of cancer variations in miRNA networks, comprising comparing normal to cancer networks by generating a comprehensive miRNA alteration map in cancer, and superimposing DNA variations onto expression data.
3. The method of claim 2, wherein at least one miRNA network comprising the miRNA network substantially as shown in Figure 4B.
4. The method of claim 2, wherein at least one miRNA network comprising the miRNA network substantially as shown in Figure 6.
5. The method of claim 2, wherein at least one miRNA network comprising the miRNA network substantially as shown in Figure 7.
6. The method of claim 2, wherein at least one miRNA network comprising the miRNA network substantially as shown in Figure 8.
7. The method of claim 2, wherein at least one miRNA network comprising the miRNA network substantially as shown in Figure 16.
8. The method of claim 2, wherein at least one miRNA network comprising the miRNA network substantially as shown in Figure 17.
9. The method of claim 2, wherein at least one miRNA network comprising the miRNA network substantially as shown in Figure 18.
10. A method for diagnosing a subject suspected of having a solid cancer, comprising screening a sample from the subject for one or more of the differentially regulated miRNAs listed in Supplemental Table IV.

11. A method for diagnosing a subject suspected of having a solid cancer, comprising screening a sample from the subject for one or more of the differentially regulated miRNAs listed in Supplemental Table V.
12. A method for diagnosing a subject suspected of having a solid cancer, comprising screening a sample from the subject for one or more of the differentially regulated miRNAs listed in Supplemental Table VIII.
13. A method for diagnosing a subject suspected of having a solid cancer, comprising screening a sample from the subject for one or more of the differentially regulated miRNAs listed in Supplemental Table IX.
14. Use of the method of claims 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein a change in miRNA expression in a subject is indicative of the subject being predisposed to, or having, a solid cancer and/or hematological cancer.
15. The method of claims 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein the cancer is one or more of lung, breast, colon, prostate, pancreas, and nasopharyngeal carcinomas, glioblastoma, melanoma, Ewing sarcoma, osteosarcoma, T-cell acute lymphoblastic leukemia (T-ALL), AML, CLL, myelodysplasia, various lymphomas, and mucosa-associated lymphoid tissue MALT.
16. A method for detecting cancer differentiation in a subject, comprising screening for at least one of hsa-miR-221 and hsa-miR-218.
17. A method for compiling and/or analyzing a network of biological functions in a subject, comprising:
 - i) obtaining information on one or more miR present in the subject, wherein at least one miR reflects a biological function; and
 - ii) subjecting the obtained information to analysis of coherent groups of nodes by using clustering algorithms to calculate a relationship between the functional reporters to generate a network relationship of the miRs to the biological function.
18. The method of claim 17, wherein the subject is human.
19. A method of profiling miR expression in a human subject having, or suspected of having, a cancer-related disease, the method comprising:

determining, for each miR of a set of miRs, a level of the miRs in a sample of the subject;
and

determining whether the miRs are identified as miRs in separate and un-linked miRNA subnets,

wherein, the presence of the separate and un-linked miRNA subnets, is indicative of the subject having the cancer-related disease.

20. The method of claim 19, including detecting miRNA networks in cancer cells that are independently regulated miRNAs.

21. The method of claim 19, including detecting target genes of uncoordinated miRs that are involved in specific cancer-related pathways.

22. The method of claim 19, wherein the miRs are up-regulated microRNAs found in solid cancers, the set of miRs comprising one or more of: the miRs in Table IV, namely:

hsa-miR-21, hsa-miR-25, hsa-miR-20a, hsa-miR-17, hsa-miR-106a, hsa-miR-106b, hsa-miR-146a, hsa-miR-92a, hsa-miR-103, hsa-miR-130b, hsa-miR-30c, hsa-miR-93, hsa-miR-107, hsa-miR-30e, hsa-miR-15a, hsa-miR-181b, hsa-miR-15b, hsa-miR-181a, hsa-miR-32, hsa-miR-345, hsa-miR-34a, hsa-miR-374a.

23. The method of claim 19, wherein the miRs are down-regulated microRNAs found in solid cancers, the set of miRs comprising one or more of: the miRs in Table V, namely:

miR-203, hsa-miR-145, hsa-miR-205, hsa-miR-206, hsa-miR-33b, hsa-miR-193a, hsa-miR-204, hsa-miR-143, hsa-miR-326, hsa-miR-338, hsa-miR-9, hsa-miR-95, hsa-miR-138, hsa-miR-183, hsa-miR-202, hsa-miR-128a, hsa-miR-214, hsa-miR-132, hsa-miR-299, hsa-miR-129, hsa-miR-133a, hsa-miR-139, hsa-miR-339, hsa-miR-1, hsa-miR-133b, hsa-miR-323, hsa-miR-218, hsa-miR-335.

24. A method for characterizing a cancer comprising:
i) determining an expression level of one or more miRNAs in a biological sample of a subject; and,
ii) characterizing the cancer based on at least a subset of the miRNAs selected from: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or, the network shown in Figure 18.

25. A method for characterizing a cancer in a subject comprising:
- i) determining an expression level of each of a plurality of miRNAs in a biological sample of the subject, and
 - ii) characterizing the cancer based on the expression level of each of the plurality of miRNAs,
- wherein the characterizing is with increased sensitivity or specificity as compared to characterizing by detecting an expression level of less than each of the plurality of miRNAs selected from: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or the network shown in Figure 18.
26. The method of claim 24 or 25, wherein the plurality comprises at least 10 miRNAs.
27. The method of claim 24 or 25, wherein at least a subset of the plurality of miRNAs is selected from Supplemental Table IV.
28. The method of claim 24 or 25, wherein at least a subset of the plurality of miRNAs is selected from Supplemental Table V.
29. The method of claim 24 or 25, including classifying a cell as non-cancerous, by:
- i) determining an expression level of one or more miRNAs in a biological sample of a subject; and
 - ii) classifying the cell as non-cancerous when less than about 3000 copies per microliter of at least a subset of the miRNAs is detected in the sample.
30. The method of claim 24 or 25, including classifying a cell as non-cancerous, by:
- i) determining an expression level of one or more miRNAs in a biological sample of a subject; and
 - ii) classifying the cell as cancerous when greater than about 9000 copies per microliter of at least a subset of the miRNAs is detected in the sample.
31. The method of claim 29 or 30, further comprising selecting a therapy or treatment regimen based on the classification.

32. The method of claim 31, wherein the biological sample is selected from the group consisting of: a heterogeneous cell sample, sputum, blood, blood cells, serum, biopsy, urine, peritoneal fluid, and pleural fluid.

33. A detection system configured to assess miRNAs selected from the group consisting of: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; the network shown in Figure 18; miRs in Supplemental Table IV; or miRs in Supplemental Table V.

34. The detection system of claim 33, wherein the system comprises a set of probes that selectively hybridizes to the two or more miRNAs.

35. A kit comprising a set of probes that selectively hybridizes to two or more miRNAs selected from the group consisting of: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; the network shown in Figure 18; miRs in Supplemental Table IV; or miRs in Supplemental Table V.

36. The kit of claim 35, wherein each of the probes is coupled to a substrate.

37. A method for analyzing a network of biological functions in a biological entity, comprising the steps of:

- i) obtaining information on at least two miRs in the biological entity, wherein the miRs reflect a biological function; and
- ii) subjecting the obtained information to processing to calculate a relationship between the miRs to generate a network relationship of the biological functions.

38. The method according to claim 37, wherein the biological entity is a cell.

39. The method according to claim 37, wherein the network is used for a use selected from the group consisting of identification of a biomarker, analysis of a drug target, analysis of a side effect, diagnosis of a cellular function, analysis of a cellular pathway, evaluation of a biological effect of a compound, and diagnosis of an infectious disease.

40. The system according to claim 39, further comprising means for analyzing the generated network by conducting an actual biological experiment.

41. The system according to claim 40, wherein the means for analyzing comprises a regulation agent specific to the function.

42. A computer program for implementing in a computer, a method for analyzing a network of biological functions in a biological entity, comprising the steps of:

- i) obtaining information on at least two miRs in the biological entity, wherein the miRs reflect a biological function; and
- ii) subjecting the obtained information to processing to calculate a relationship between the miRs to generate a network relationship of the biological functions.

43. A storage medium comprising a computer program for implementing in a computer, a method for analyzing a network of biological functions in a biological entity, comprising the steps of:

- i) obtaining information on at least two miRs in the biological entity, wherein the miRs reflect a biological function; and
- ii) subjecting the obtained information to processing to calculate a relationship between the miRs to generate a network relationship of the biological functions.

44. A transmission medium comprising a computer program for implementing in a computer, a method for analyzing a network of biological functions in a biological entity, comprising the steps of:

- i) obtaining information on at least two miRs in the biological entity, wherein the miRs reflect a biological function; and
- ii) subjecting the obtained information to processing to calculate a relationship between the miRs to generate a network relationship of the biological functions.

45. An assay kit for detecting a risk that a cell will become malignant, comprising reagents for determining a cell signature, wherein the signature comprises the presence and/or level of two or more miRs selected from: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or, the network shown in Figure 18.

46. The assay kit of claim 45, wherein the miR is differentially expressed in the cells at risk of becoming malignant.

47. A method of determining a risk of developing cancer in a subject, comprising:
providing a biological sample from the subject;
determining a cell signature for the biological sample, wherein the signature comprises a collection of measurements of at least the presence and/or level of two or more miRs selected from: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or the network shown in Figure 18;
comparing the cell signature of the biological sample with a cell signature of a control sample; and
determining the risk of developing cancer.
48. The method of claim 47, wherein the biological sample is selected from the group consisting of a living cell, a dead cell, any non-cellular liquid samples comprising nipple aspirate fluid, urine, blood, serum, plasma and a lavage sample, and any combinations thereof.
49. The method of claim 47, wherein the cell signature of the control sample is obtained from a database.
50. The method of claim 47, wherein at least part of the method is performed by automated means.
51. The method of claim 47, wherein the automated means comprises a computer-based system configured to carry out at least one of the following: measuring the presence of absence of a substance in the sample; recording data obtained from the measurement; analyzing the data; determining the risk of cancer; and generating a report.
52. The method of claim 47, wherein the computer-based system comprises any hardware, software, firmware, processor, and any combinations thereof.
53. The method of claim 52, wherein the computer-based system is configured to access and/or use a database comprising a cell signature of a pre-cancerous cell and/or a control epithelial cells via a network system.
54. A method of making a medical report related to the risk of developing cancer in a subject, comprising:
providing a biological sample from the subject;

determining a signature for the biological sample, wherein the signature comprises a collection of measurements of at least two characteristics of the sample selected from one or more of the following: presence and/or level of at least one miRNA;

comparing the signature of the biological sample with a mammary epithelial cell signature of a control sample;

determining the risk of developing cancer; and

generating a report related to the risk of developing cancer.

55. The method of claim 54, wherein the automated means comprises a computer-based system configured to carry out at least one of the following:

measuring the cell signature;

recording data obtained from the measurement;

analyzing the data;

determining the risk of developing cancer; and

generating a report.

56. The method of claim 54, wherein the computer-based system comprises any hardware, software, firmware, processor, and any combinations thereof.

57. The method of claim 54, wherein the computer-based system is configured to access and/or use a database comprising a cell signature of a pre-cancerous cell and/or a control cell via a network system.

58. A method of making a medical report related to the risk of developing cancer in a subject, comprising:

providing a biological sample from the subject;

determining a cell signature for the biological sample, wherein the signature comprises a collection of measurements of at least two characteristics of the cell, the at least two miRs selected from: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or the network shown in Figure 18;

comparing the cell signature of the biological sample with a cell signature of a control sample;

determining the risk of developing cancer; and

generating a report related to the risk of developing cancer.

59. A method for compiling and/or analyzing an miRNA database, comprising one or more of:

- using F635-background values;
- removing bad spots;
- averaging nonexpressed spots for each gpr files;
- for each mature miRNA, computing the geometric mean of its multiple reporters;
- assigning a NaN value to miRNAs with more than 50% of corrupted spots;
- log2-transforming the results;
- normalizing by using the quantiles normalization;
- performing t-test over two classes' experiments or F-tests over multiple classes (i.e.,

different normal tissues);

- performing target genes selection, wherein the union of the target mRNAs with a score >3 is used as an input to ClueGO;

- using ClueGO to relate differential expression in cancer to functional pathways (KEGG), wherein:

- i) ClueGO visualizes the selected terms in a functionally grouped annotation network that reflects the relationships between the terms based on the similarity of their associated genes,
- ii) the size of the nodes reflects the statistical significance of the terms; and
- iii) the degree of connectivity between terms (edges) is calculated using kappa statistics;

using the calculated kappa score for defining functional groups, wherein:

- i) a term can be included in several groups,
- ii) the reoccurrence of the term is shown by adding "n", and
- iii) the group leading term is the most significant term of the group,

whereby the network integrates only the positive kappa score term associations.

60. The method of claim 59, wherein expression values are preprocessed to only filter out nonvarying miRNAs, according to the following parameters:

- #filter.flag = filter (variation filter and thresholding flag);
- #preprocessing.flag = no disk or norm (discretization and normalization flag);
- #minchange = 10 (minimum fold change for filter);
- #mindelta = 512 (minimum delta for filter);
- #threshold = 64 (value for threshold);
- #ceiling = 20,000 (value for ceiling);
- #max. sigma.binning = 1 (maximum sigma for binning);

#prob.thres = 1 (value for uniform probability threshold filter);
 #num.excl = 2% of total chips (number of experiments to exclude (max and min) before applying variation filter);
 #log.base.two = no (whether to take the log base two after thresholding);
 #number.of.columns.above.threshold = 1% of total chips (remove row if n columns not $\geq$ than given threshold above.threshold); and
 #column.threshold = 512 (threshold for removing rows).

61. A system for analyzing a an miRNA database, comprising one or more means for:
- using F635-background values;
 - removing bad spots;
 - averaging nonexpressed spots for each gpr files;
 - for each mature miRNA, computing the geometric mean of its multiple reporters;
 - assigning a NaN value to miRNAs with more than 50% of corrupted spots;
 - log2-transforming the results;
 - normalizing by using the quantiles normalization;
 - performing t-test over two classes' experiments of F-tests over multiple classes (i.e., different normal tissues);
 - performing target genes selection, wherein the union of the target mRNAs with a score >3 is used as an input to ClueGO;
 - using ClueGO to relate differential expression in cancer to functional pathways (KEGG), wherein:
 - i) ClueGO visualizes the selected terms in a functionally grouped annotation network that reflects the relationships between the terms based on the similarity of their associated genes,
 - ii) the size of the nodes reflects the statistical significance of the terms; and
 - iii) the degree of connectivity between terms (edges) is calculated using kappa statistics;
 - using the calculated kappa score for defining functional groups, wherein:
 - i) a term can be included in several groups,
 - ii) the reoccurrence of the term is shown by adding "n", and
 - iii) the group leading term is the most significant term of the group,
 - whereby the network integrates only the positive kappa score term associations.

62. The system of claim 61, wherein the subject is a human.

63. A computer program for implementing in a computer, a method for compiling and/or analyzing an miRNA database, comprising one or more of:

- using F635-background values;
- removing bad spots;
- averaging nonexpressed spots for each gpr files;
- for each mature miRNA, computing the geometric mean of its multiple reporters in the chip;
- assigning a NaN value to miRNAs with more than 50% of corrupted spots;
- log2-transforming the results;
- normalizing by using the quantiles normalization;
- performing t-test over two classes' experiments of F-tests over multiple classes (i.e., different normal tissues);
- performing target genes selection, wherein the union of the target mRNAs with a score >3 is used as an input to ClueGO;
- using ClueGO to relate differential expression in cancer to functional pathways (KEGG), wherein:

- i) ClueGO visualizes the selected terms in a functionally grouped annotation network that reflects the relationships between the terms based on the similarity of their associated genes,
- ii) the size of the nodes reflects the statistical significance of the terms; and
- iii) the degree of connectivity between terms (edges) is calculated using kappa statistics;

using the calculated kappa score for defining functional groups, wherein:

- i) a term can be included in several groups,
- ii) the reoccurrence of the term is shown by adding "n", and
- iii) the group leading term is the most significant term of the group,

whereby the network integrates only the positive kappa score term associations.

64. A storage medium comprising a computer program for implementing in a computer, a method for compiling and/or analyzing an miRNA database, comprising one or more of:

- using F635-background values;
- removing bad spots;
- averaging nonexpressed spots for each gpr files;
- for each mature miRNA, computing the geometric mean of its multiple reporters;
- assigning a NaN value to miRNAs with more than 50% of corrupted spots;
- log2-transforming the results;
- normalizing by using the quantiles normalization;

performing t-test over two classes' experiments of F-tests over multiple classes (i.e., different normal tissues);

performing target genes selection, wherein the union of the target mRNAs with a score >3 is used as an input to ClueGO;

using ClueGO to relate differential expression in cancer to functional pathways (KEGG), wherein:

- i) ClueGO visualizes the selected terms in a functionally grouped annotation network that reflects the relationships between the terms based on the similarity of their associated genes,
- ii) the size of the nodes reflects the statistical significance of the terms; and
- iii) the degree of connectivity between terms (edges) is calculated using kappa statistics;

using the calculated kappa score for defining functional groups, wherein:

- i) a term can be included in several groups,
- ii) the reoccurrence of the term is shown by adding "n", and
- iii) the group leading term is the most significant term of the group,

whereby the network integrates only the positive kappa score term associations.

65. A transmission medium comprising a computer program for implementing in a computer, a method for compiling and/or analyzing an miRNA database, comprising one or more of:

using F635-background values;

removing bad spots;

averaging nonexpressed spots for each gpr files;

for each mature miRNA, computing the geometric mean of its multiple reporters;

assigning a NaN value to miRNAs with more than 50% of corrupted spots;

log2-transforming the results;

normalizing by using the quantiles normalization;

performing t-test over two classes' experiments of F-tests over multiple classes (i.e., different normal tissues);

performing target genes selection, wherein the union of the target mRNAs with a score >3 is used as an input to ClueGO;

using ClueGO to relate differential expression in cancer to functional pathways (KEGG), wherein:

- i) ClueGO visualizes the selected terms in a functionally grouped annotation network that reflects the relationships between the terms based on the similarity of their associated genes,

- ii) the size of the nodes reflects the statistical significance of the terms; and
- iii) the degree of connectivity between terms (edges) is calculated using kappa statistics;

using the calculated kappa score for defining functional groups, wherein:

- i) a term can be included in several groups,
 - ii) the reoccurrence of the term is shown by adding “n”, and
 - iii) the group leading term is the most significant term of the group,
- whereby the network integrates only the positive kappa score term associations.

66. An assay kit for analyzing an miRNA database, comprising reagents one or more of:

- using F635-background values;
- removing bad spots;
- averaging nonexpressed spots for each gpr files;
- for each mature miRNA, computing the geometric mean of its multiple reporters;
- assigning a NaN value to miRNAs with more than 50% of corrupted spots;
- log2-transforming the results;
- normalizing by using the quantiles normalization;
- performing t-test over two classes' experiments of F-tests over multiple classes (i.e., different normal tissues);
- performing target genes selection, wherein the union of the target mRNAs with a score >3 is used as an input to ClueGO;
- using ClueGO to relate differential expression in cancer to functional pathways (KEGG),

wherein:

- i) ClueGO visualizes the selected terms in a functionally grouped annotation network that reflects the relationships between the terms based on the similarity of their associated genes,
- ii) the size of the nodes reflects the statistical significance of the terms; and
- iii) the degree of connectivity between terms (edges) is calculated using kappa statistics;

using the calculated kappa score for defining functional groups, wherein:

- i) a term can be included in several groups,
 - ii) the reoccurrence of the term is shown by adding “n”, and
 - iii) the group leading term is the most significant term of the group,
- whereby the network integrates only the positive kappa score term associations.

67. The assay kit of claim 66, wherein the substance detected is differentially expressed in a sample from a subject at risk of becoming malignant.

68. The assay kit of claim 66, wherein one or more miRNAs are differentially transcribed in the sample at risk of becoming malignant.

69. The assay kit of claim 66, wherein one or more of the miRNAs may have an expression level and/or copy number in a sample from a subject at risk of becoming malignant compared to a normal cell.

70. The assay kit of claim 66, wherein the miRNAs is selected from the group consisting of miRs shown in: the network shown in Figure 4B; the network shown in Figure 6; the network shown in Figure 7; the network shown in Figure 8; the network shown in Figure 16; the network shown in Figure 17; or the network shown in Figure 18.

71. The assay kit of claim 66, wherein the miRNAs is selected from the group consisting of miRs shown in: Supplemental Table IV or Supplemental Table V.

72. The assay kit of claim 66, wherein the sample is selected from one or more of: a living cell, a dead cell, any non-cellular liquid samples comprising aspirate fluid, urine, blood, serum, plasma and a lavage sample, and any combinations thereof.



Figure 1

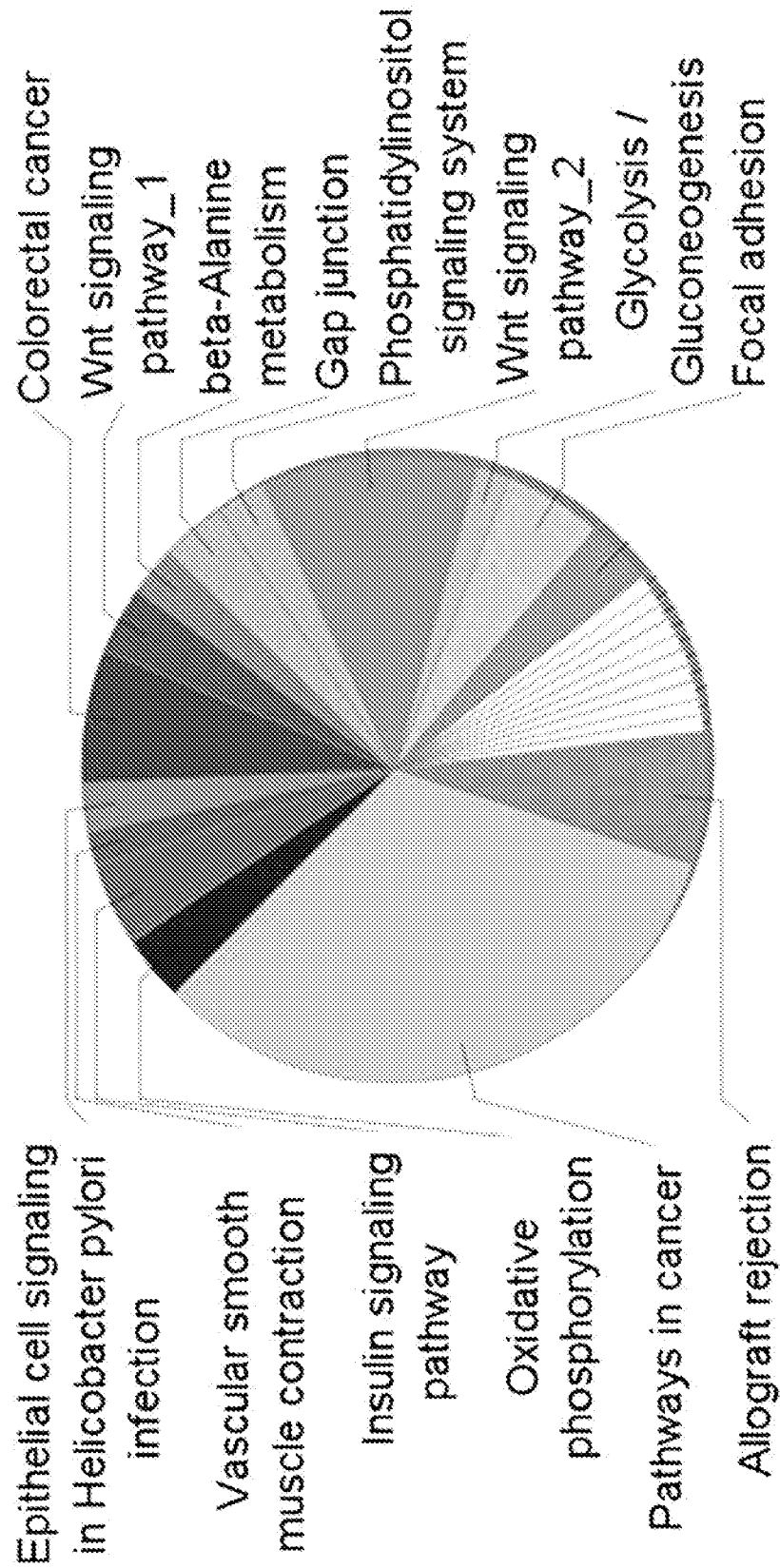


Figure 2

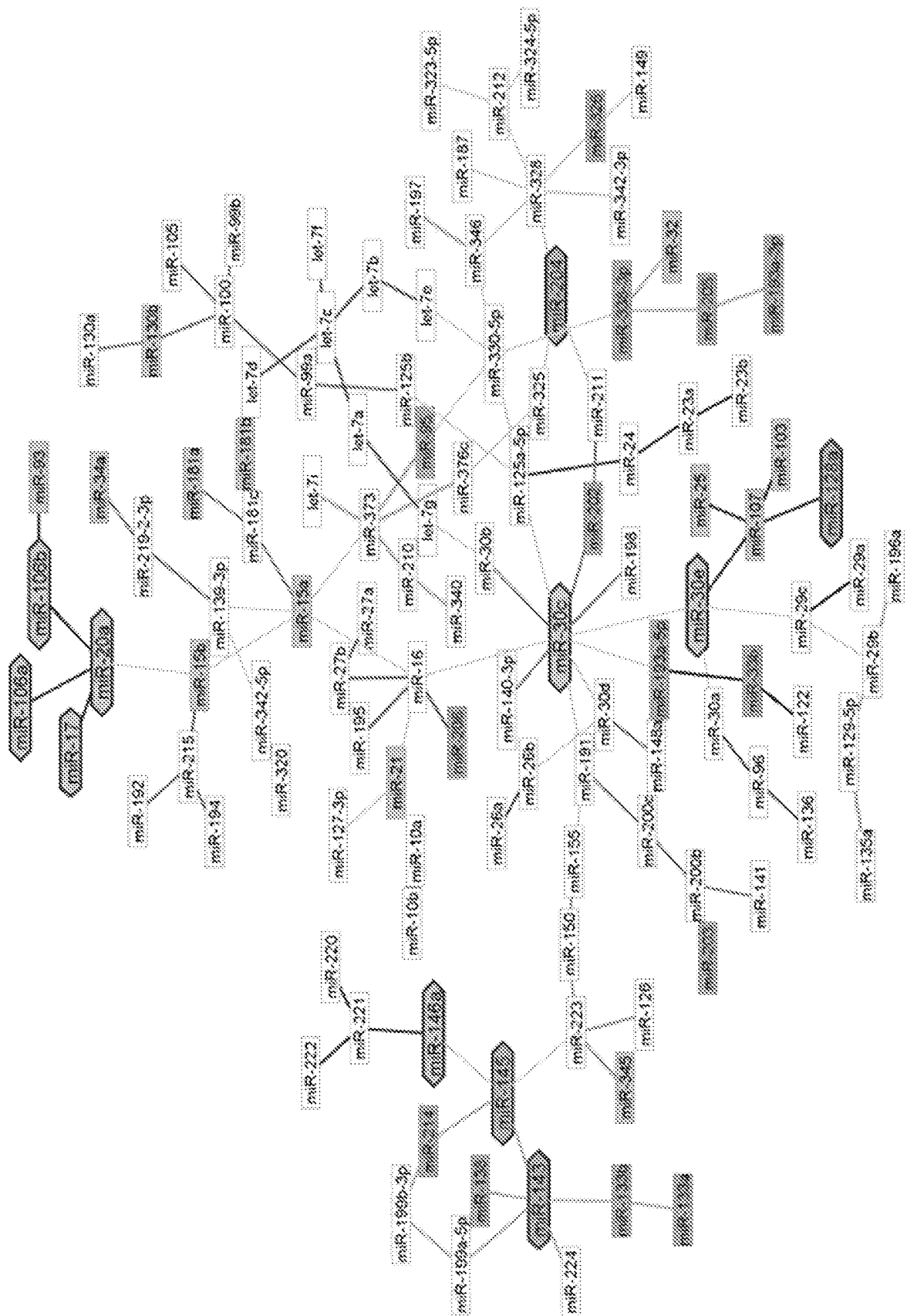


Figure 3

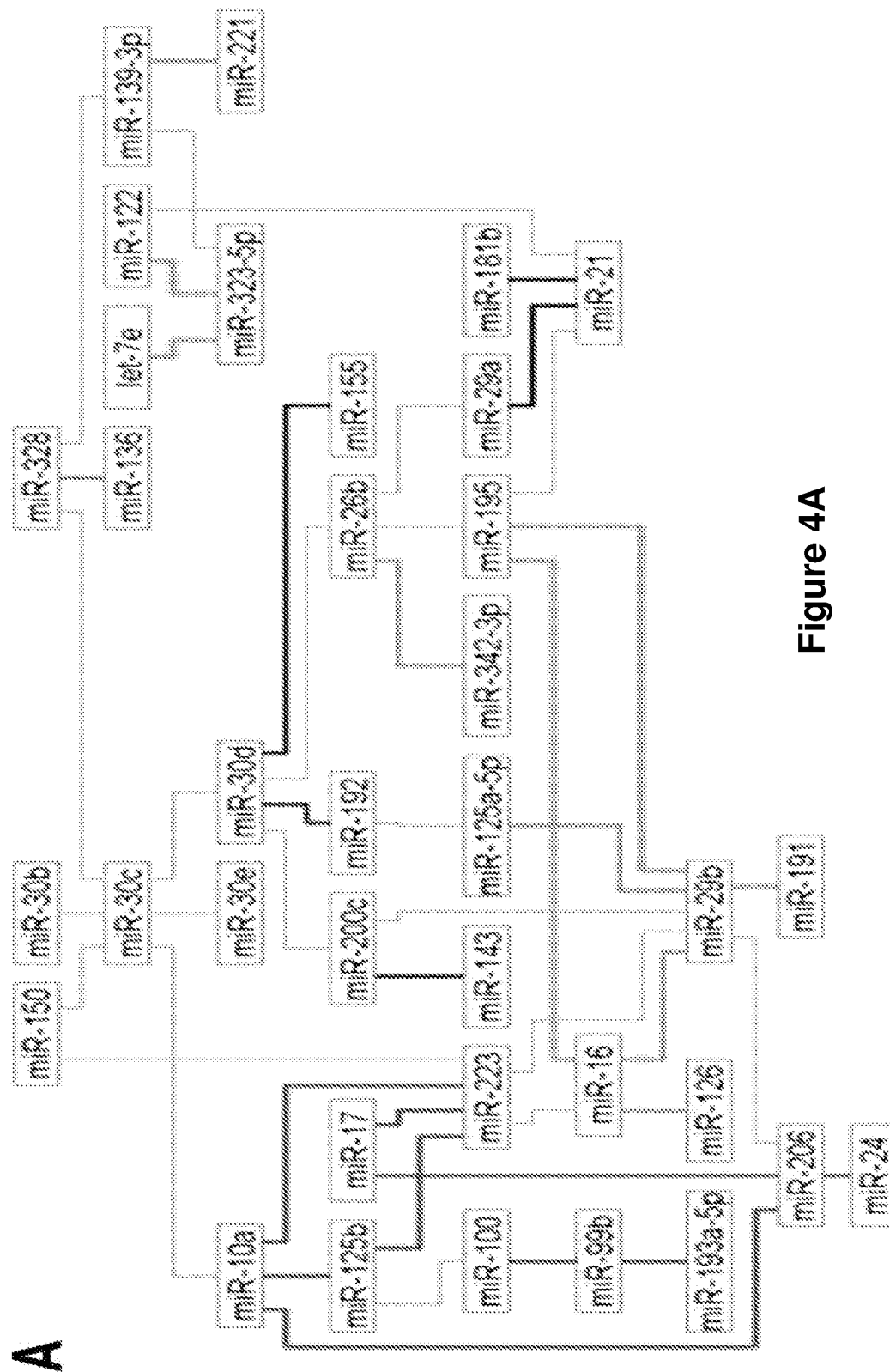


Figure 4A

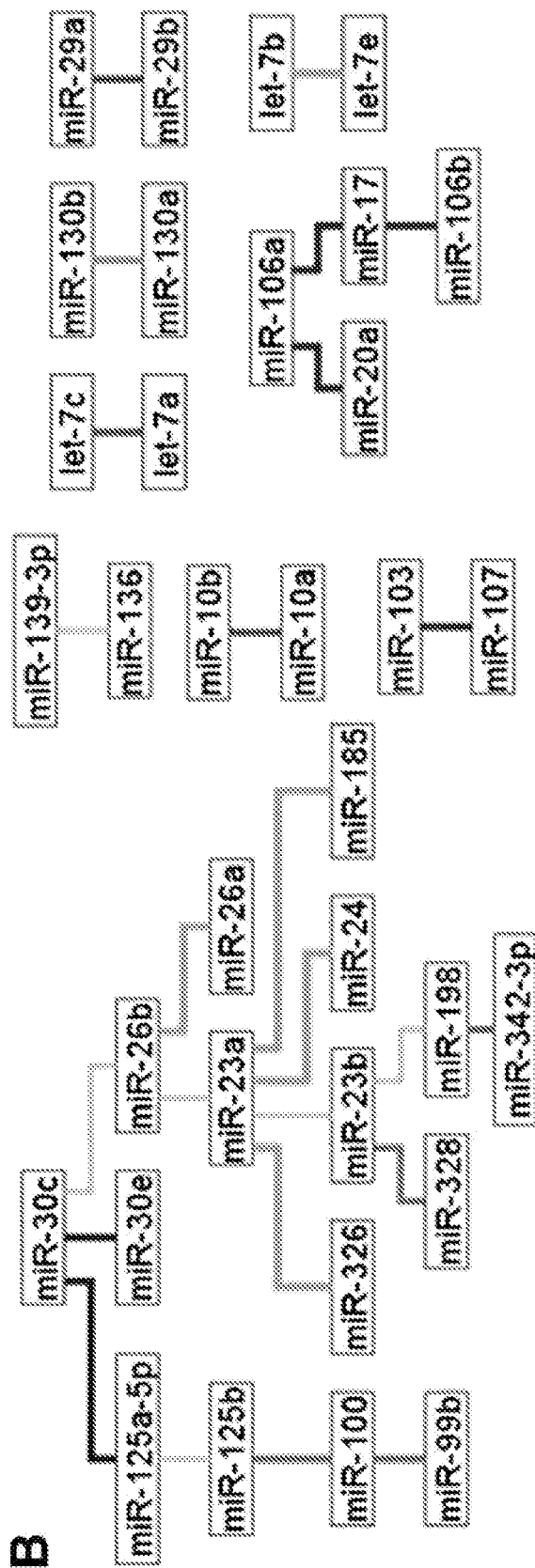


Figure 4b

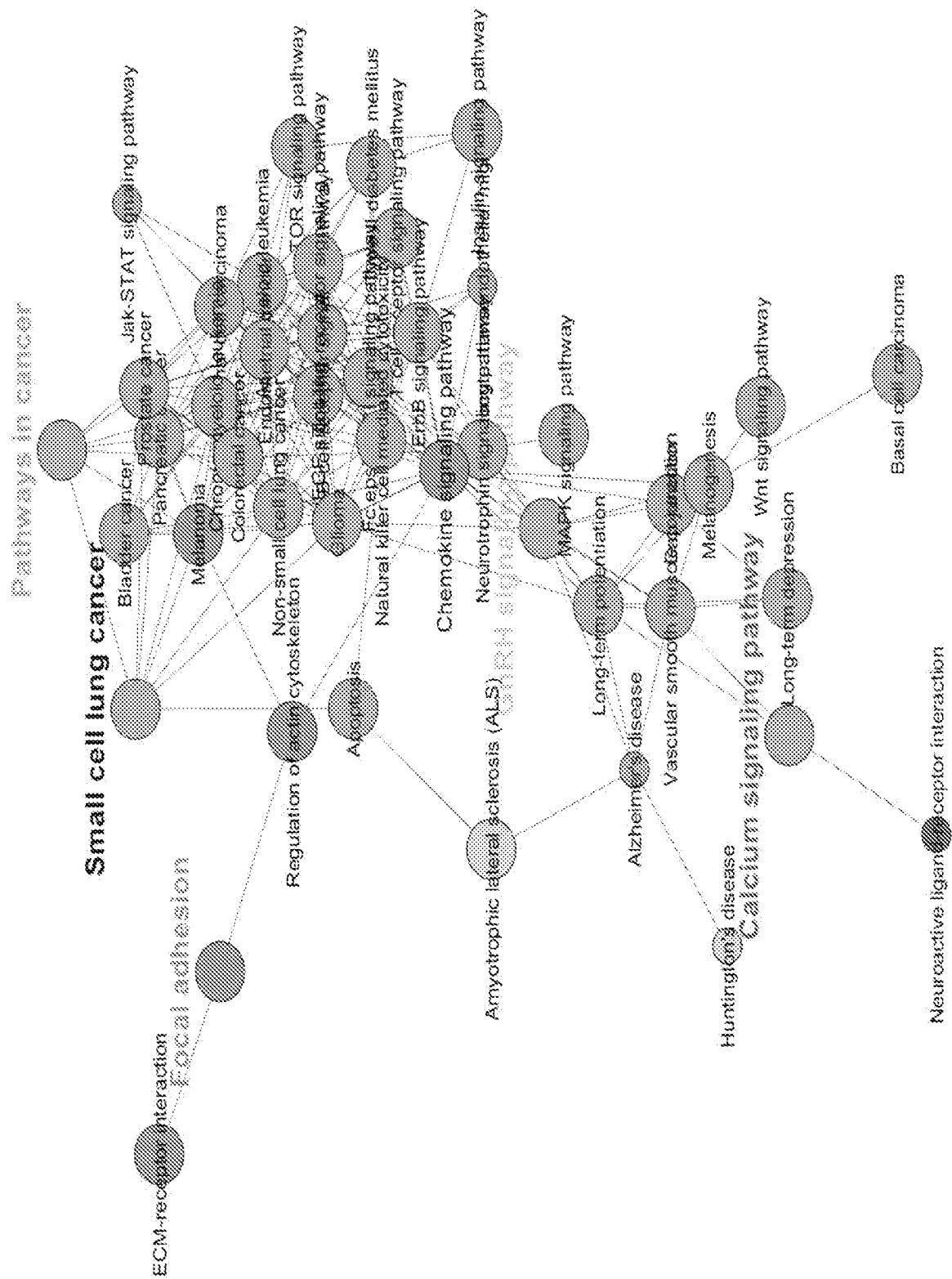


Figure 5

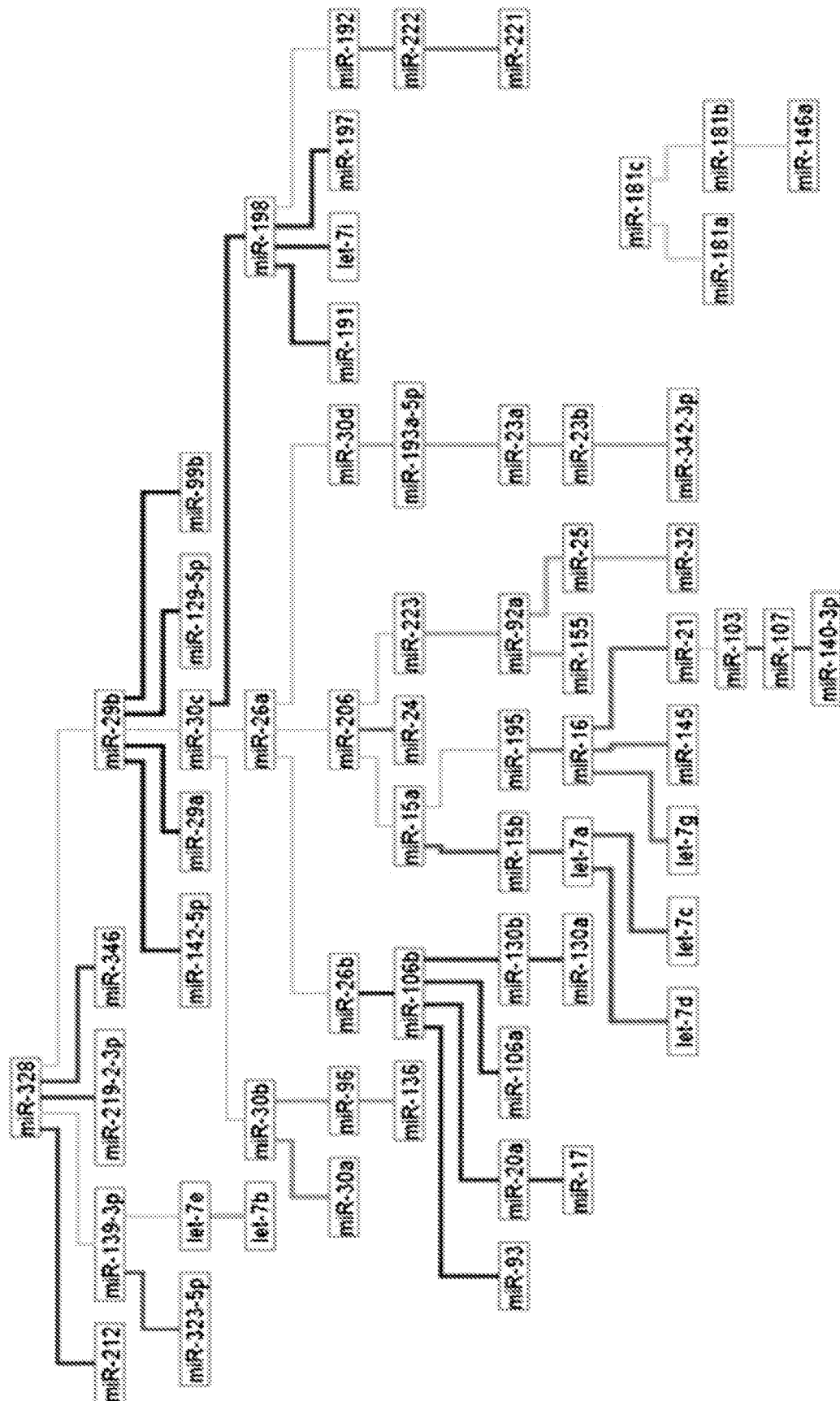


Figure 6

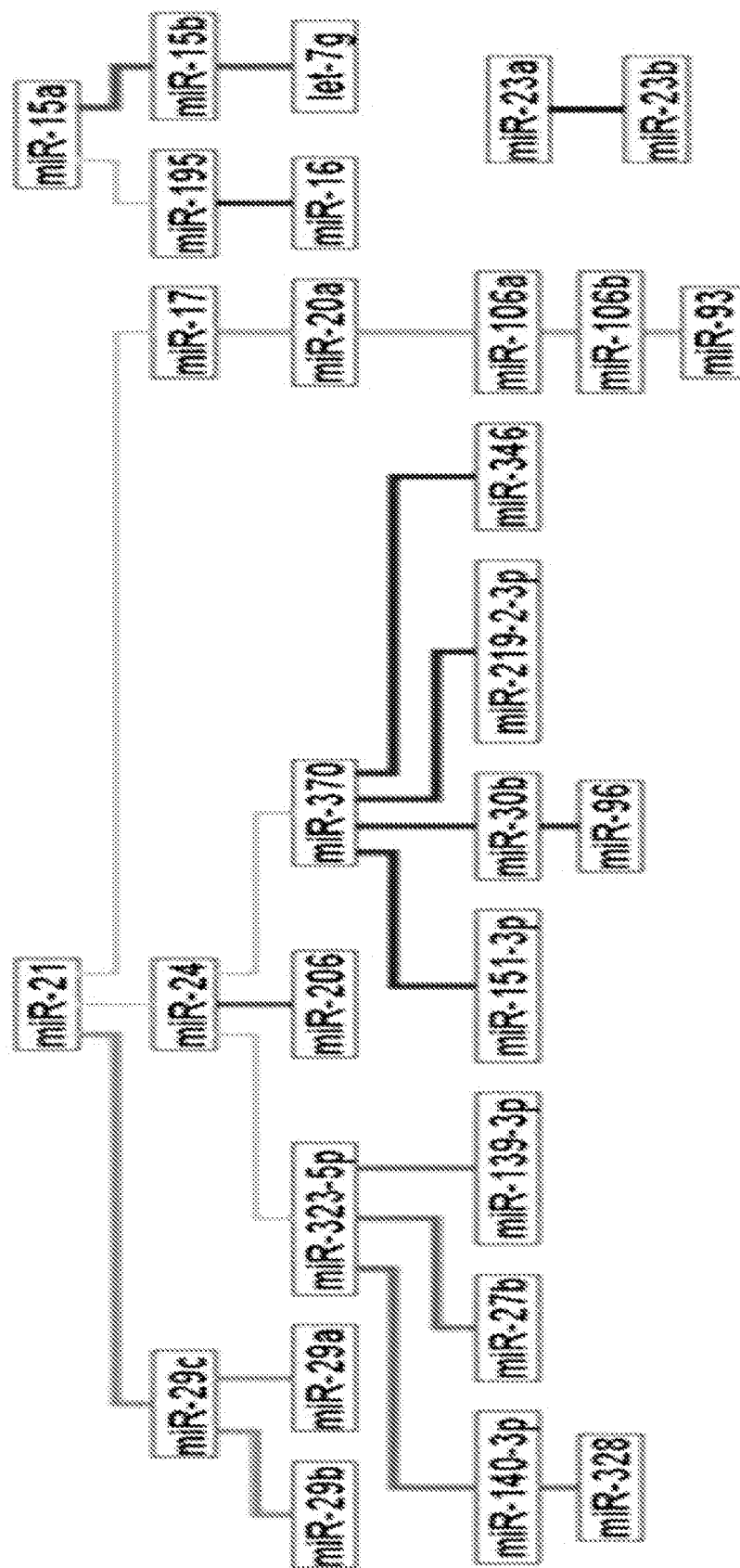


Figure 7

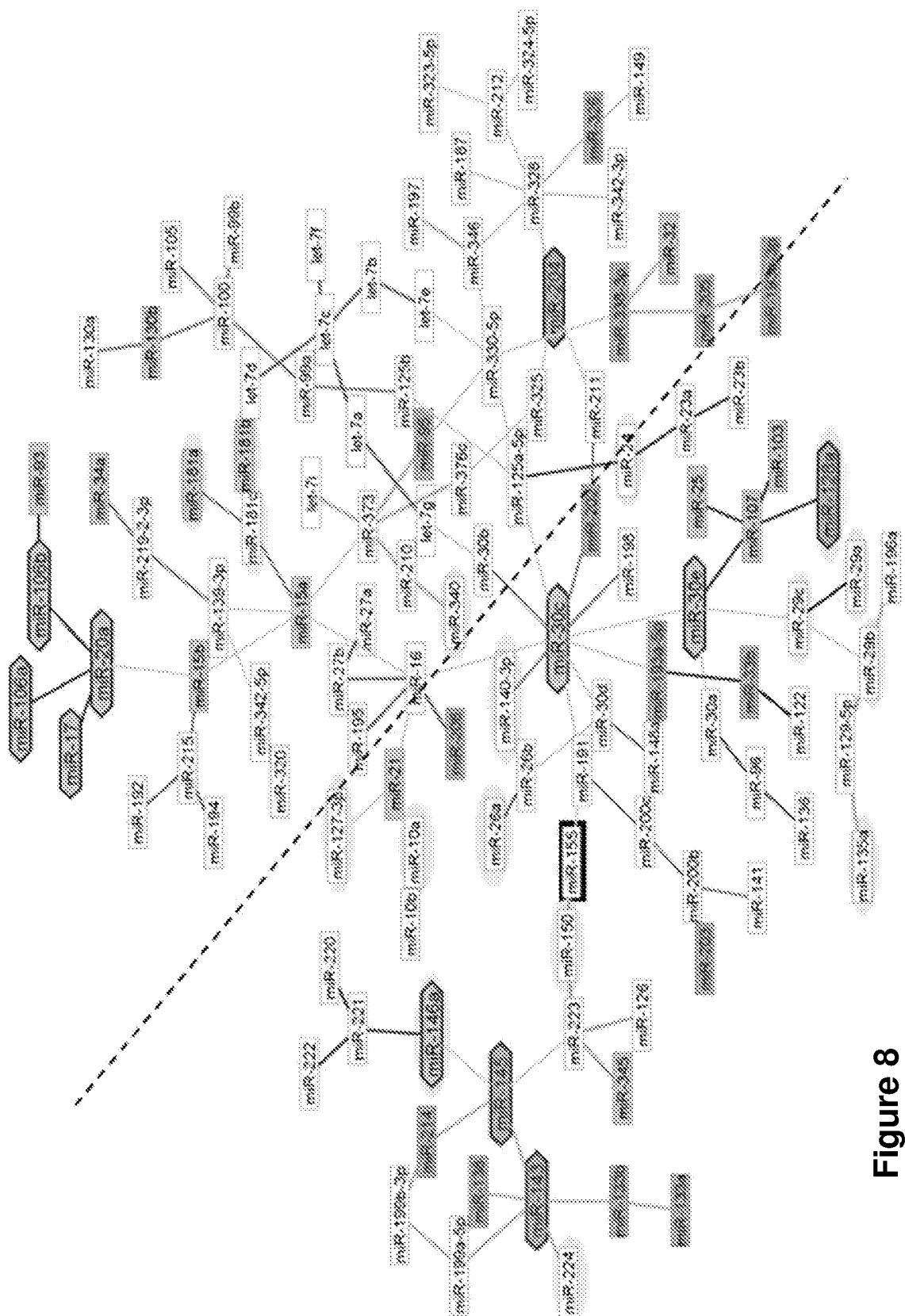


Figure 8

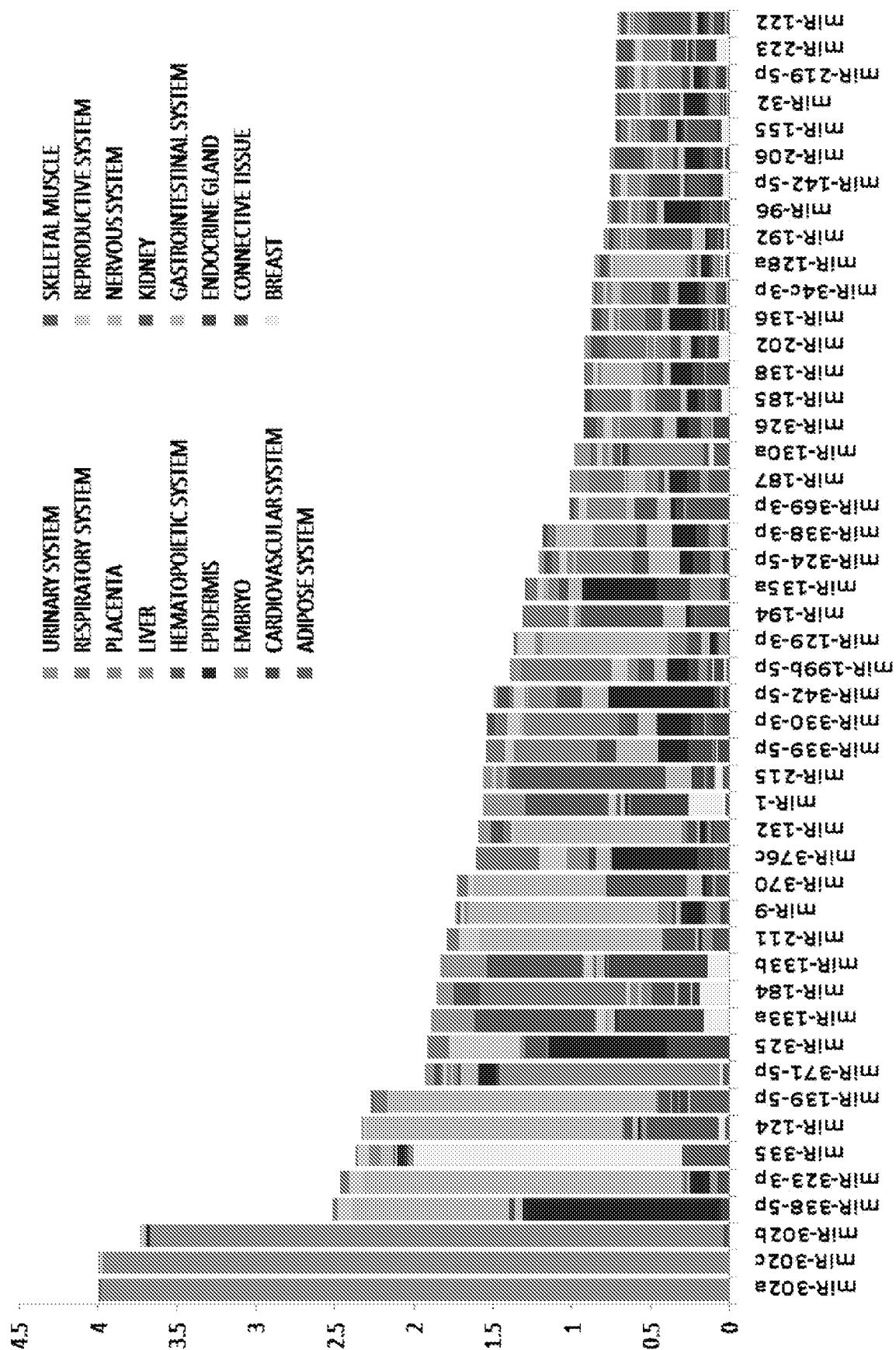


Figure 9

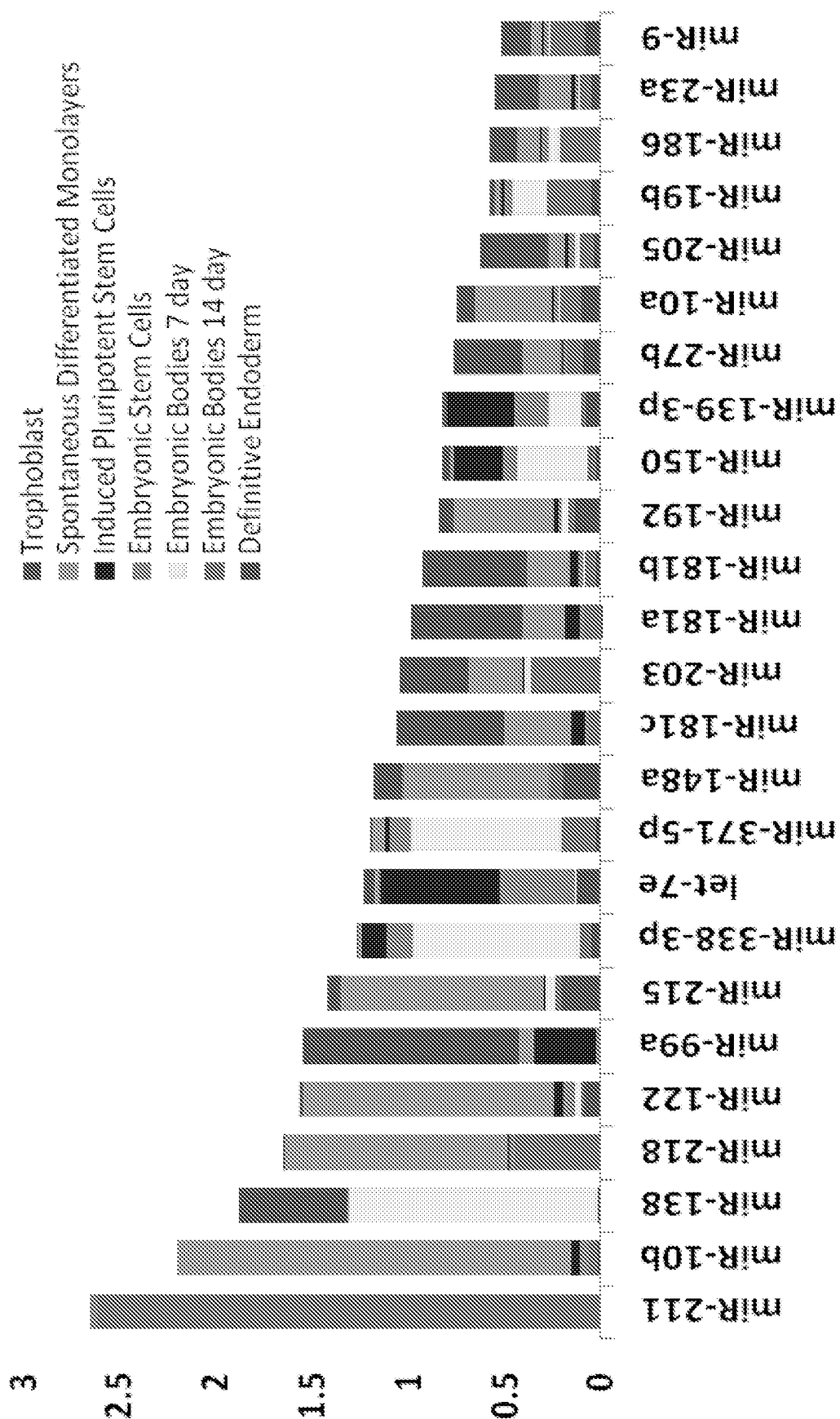
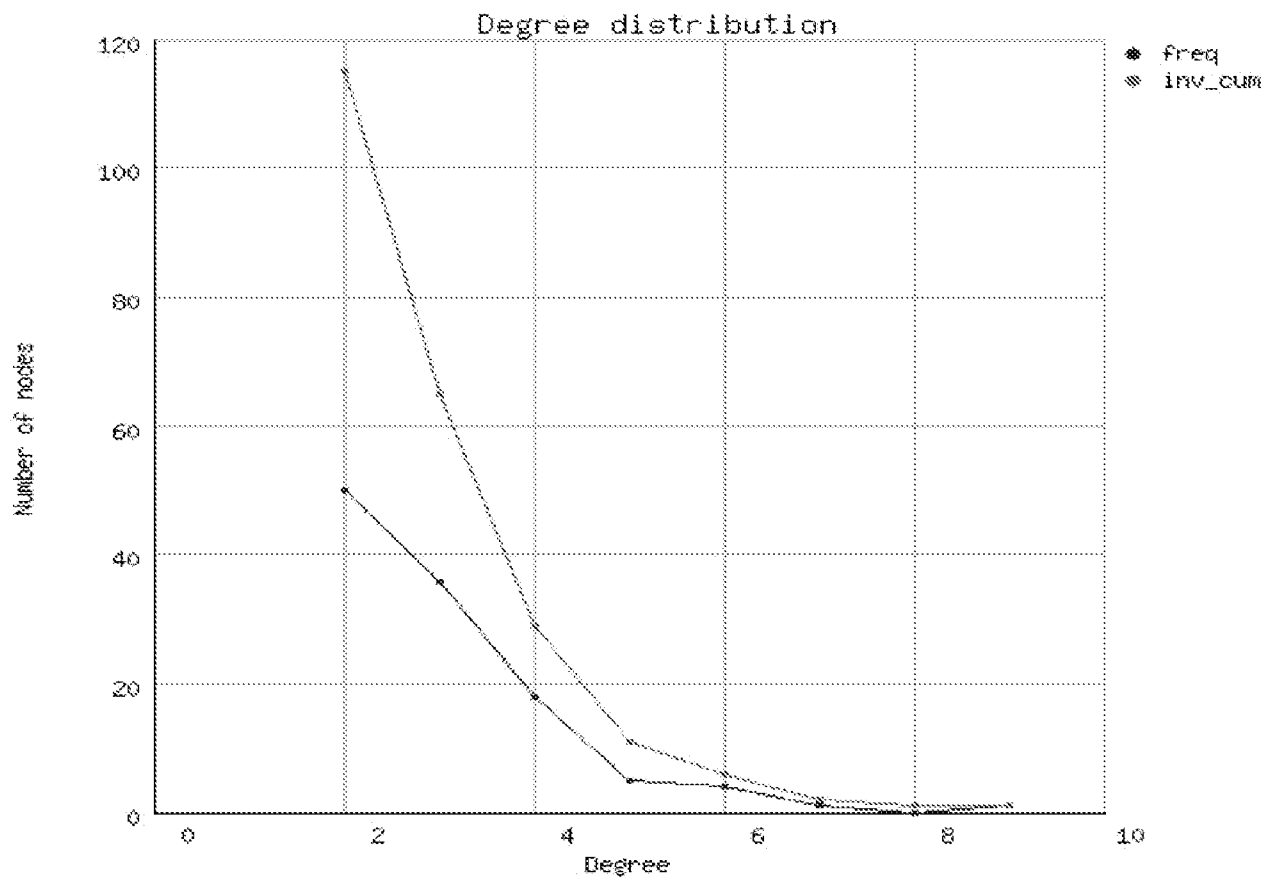
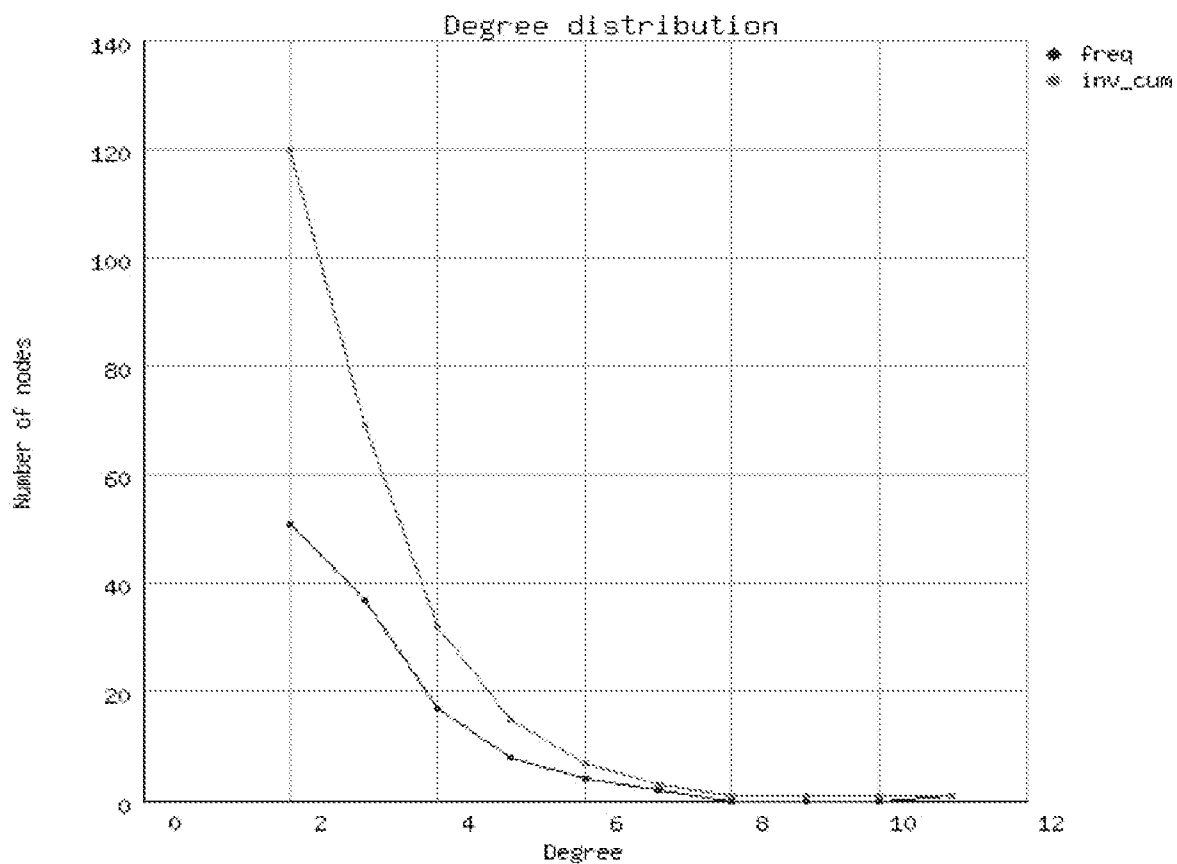
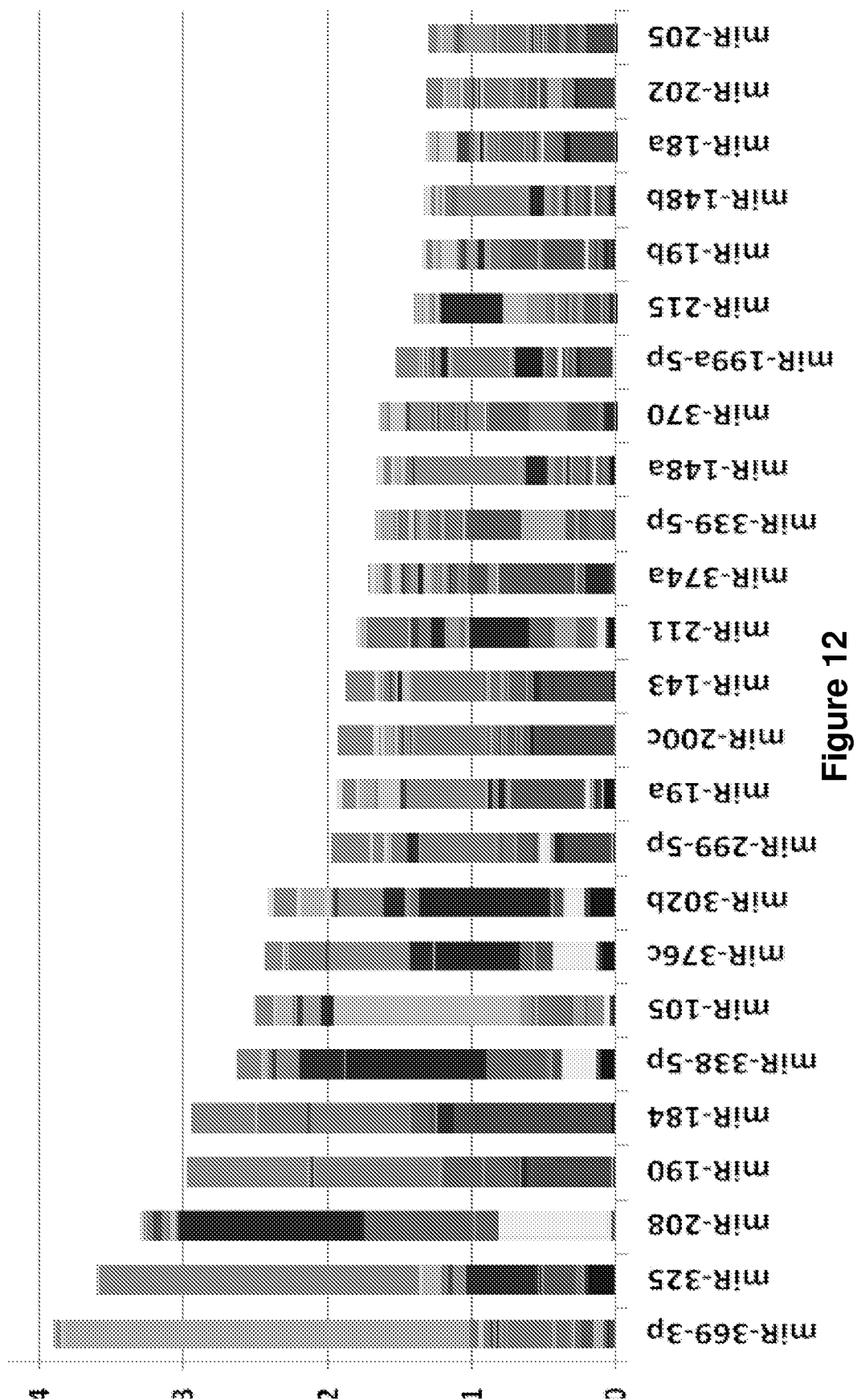


Figure 10

**Figure 11A****Figure 11B**



1: ACUTE LYMPHOBLASTIC LEUKEMIA	■	2: ACUTE MONOCYTIC LEUKEMIA	■	3: ACUTE MYELOID LEUKEMIA	■
4: ACUTE PROMYELOCYTIC LEUKEMIA	■	5: ANAPLASTIC LARGE CELL LYMPHOMA	■	6: ANGIOIMMUNOBLASTIC CELL LYMPHOMA	■
7: BLADDER TRANSITIONAL CELL CARCINOMA	■	8: BREAST DUCTAL CARCINOMA	■	9: BREAST LOBEULAR CARCINOMA	■
10: BREAST MALE CARCINOMA	■	11: BURKITT'S LYMPHOMA ENDEMIC	■	12: BURKITT'S LYMPHOMA SPORADIC	■
13: CERVIX ADENOCARCINOMA	■	14: CERVIX CARCINOMA	■	15: CHRONIC LYMPHOCYTIC LEUKEMIA	■
16: CHRONIC MYELOGENOUS LEUKEMIA	■	17: COLON ADENOMA	■	18: COLORECTAL ADENOCARCINOMA	■
19: DLBCL ANAPLASTIC	■	20: DLBCL CENTROBLASTIC	■	21: DLBCL IMMUNOBLASTIC	■
22: DLBCL NOS	■	23: EPIDERMIS BASAL CELL CARCINOMA	■	24: EPIDERMIS SQUAMOUS CELL CARCINOMA	■
25: ESOPHAGUS ADENOCARCINOMA	■	26: FOLLICULAR ADENOMA	■	27: FOLLICULAR LYMPHOMA	■
28: GASTRIC CANCER	■	29: GLIOMA	■	30: HEPATOBLASTOMA	■
31: HEPATOCELLULAR CARCINOMA	■	32: HGSL	■	33: HODGKIN LYMPHOMA	■
34: INSULINOMA	■	35: KIDNEY CARCINOMA	■	36: LIVER CHOLANGIOCELLULAR CARCINOMA	■
37: LUNG ADENOCARCINOMA	■	38: LUNG SQUAMOUS CELL CARCINOMA	■	39: LYMPHOMA CUTANEOUS	■
40: MANTLE CELL LYMPHOMA	■	41: MELANOMA	■	42: MULTIPLE MYELOMA	■
43: MYELODYSPLASTIC SYNDROMES	■	44: NEUROBLASTOMA	■	45: NSCLC	■
46: OVARIAN ADENOCARCINOMA	■	47: OVARIAN EPITHELIAL CARCINOMA	■	48: PANCREAS CARCINOMA	■
49: PANCREAS NONFUNCTIONING ENDOCRINE CANCER	■	50: PROSTATE ADENOCARCINOMA	■	51: THYROID PAPILLARY CANCER	■

Figure 12 cont.

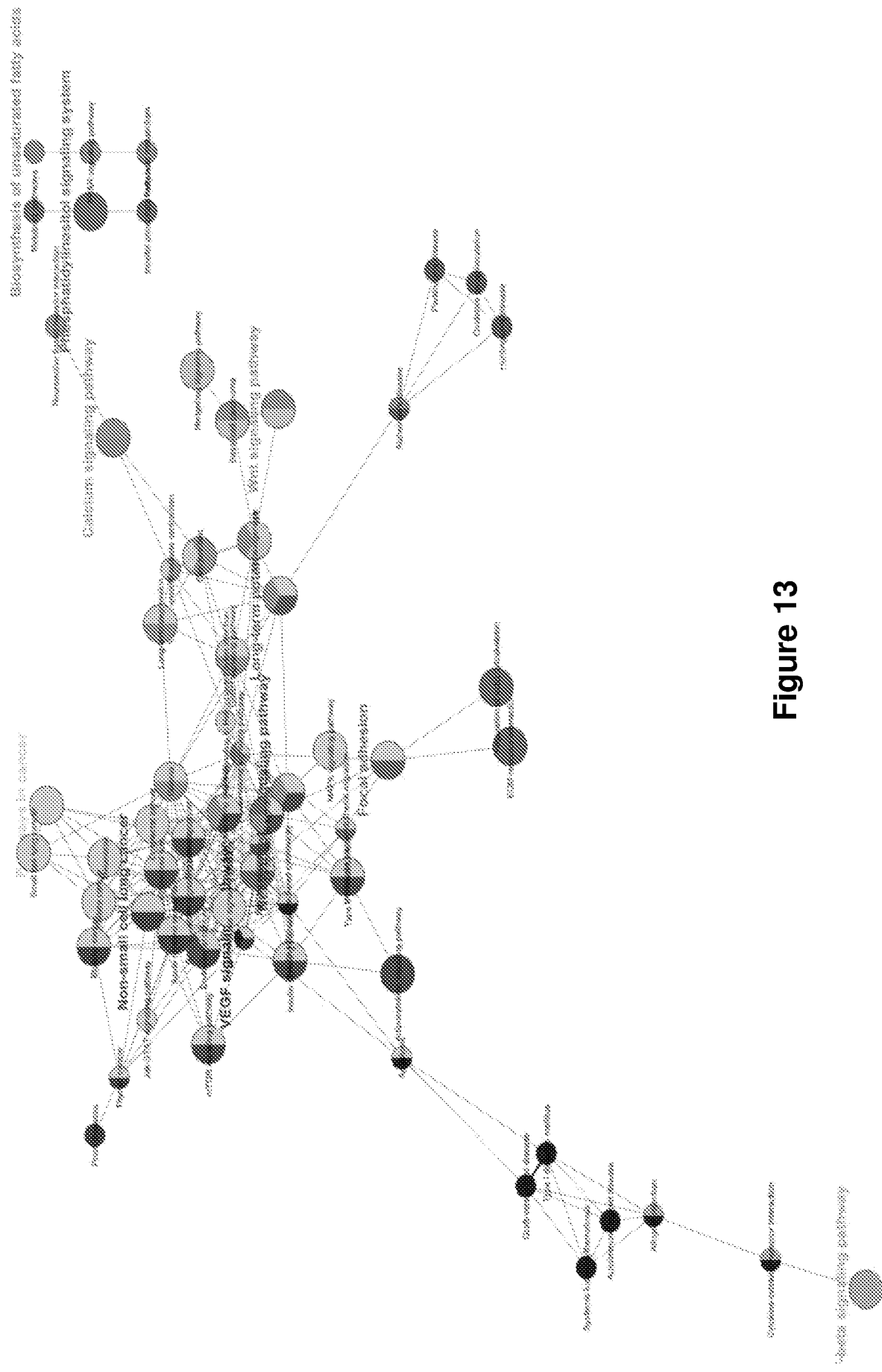


Figure 13

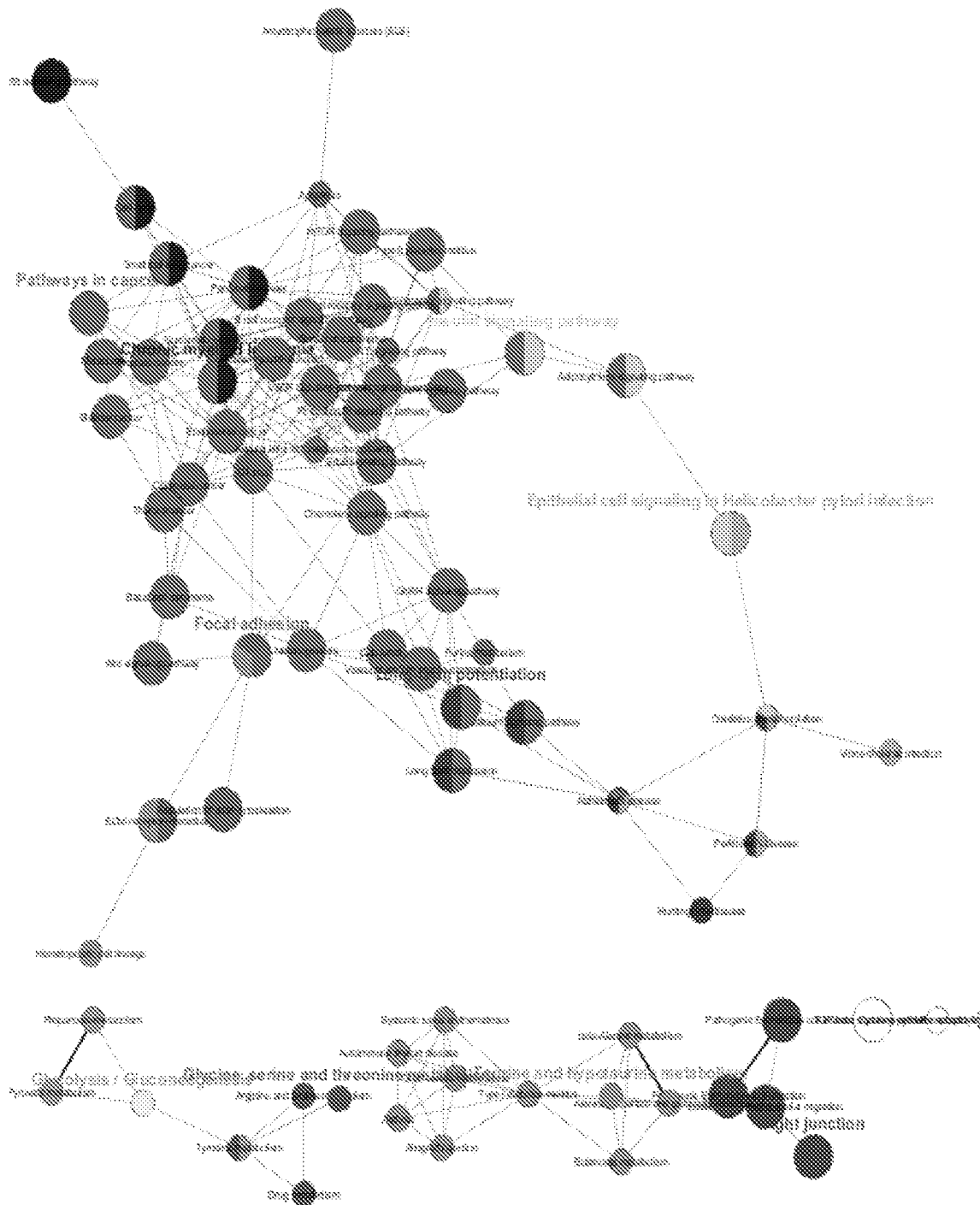
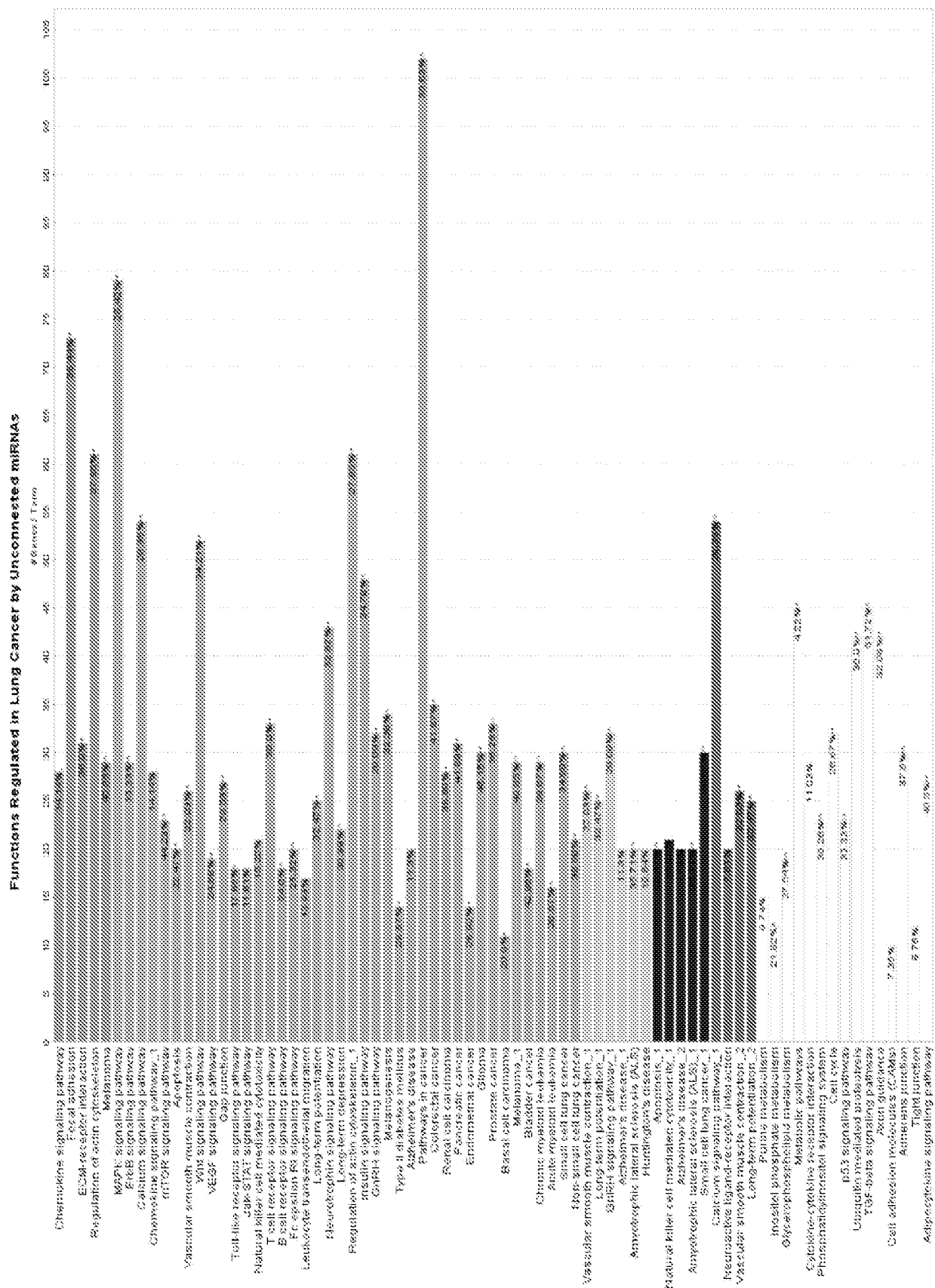


Figure 14



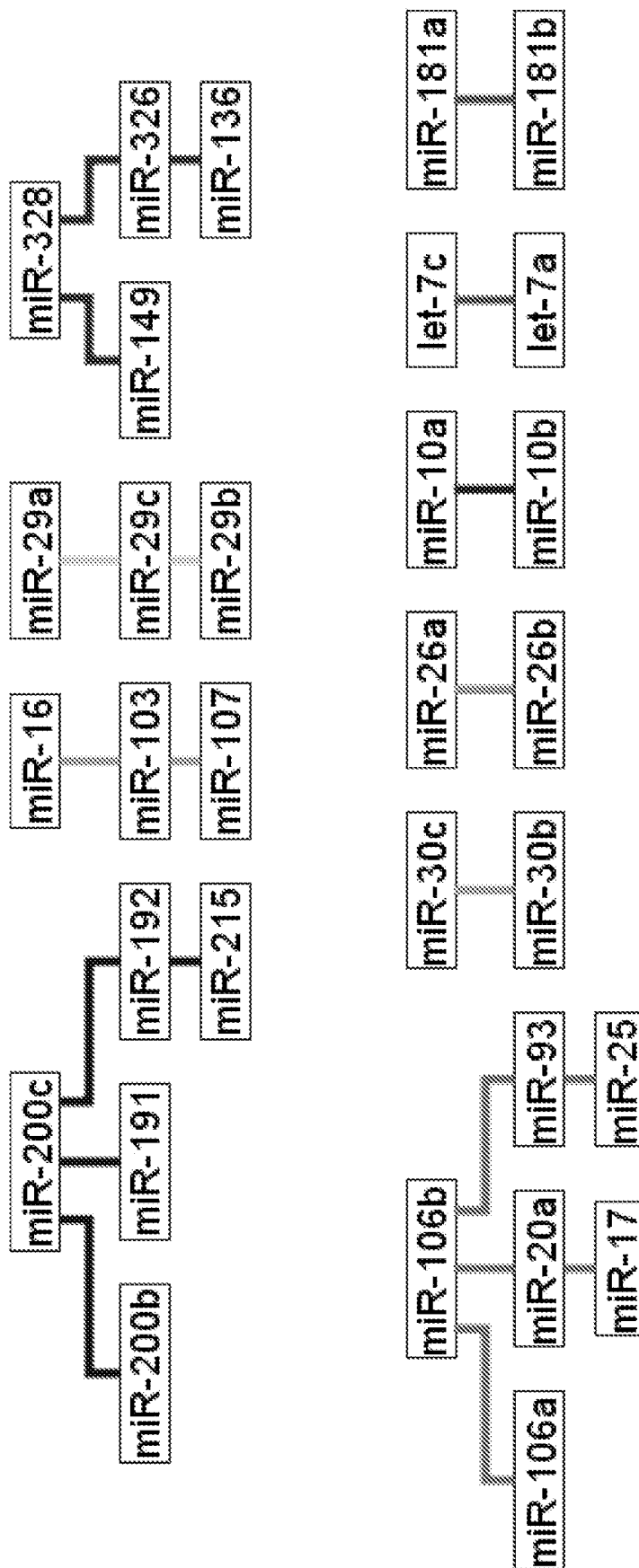


Figure 16

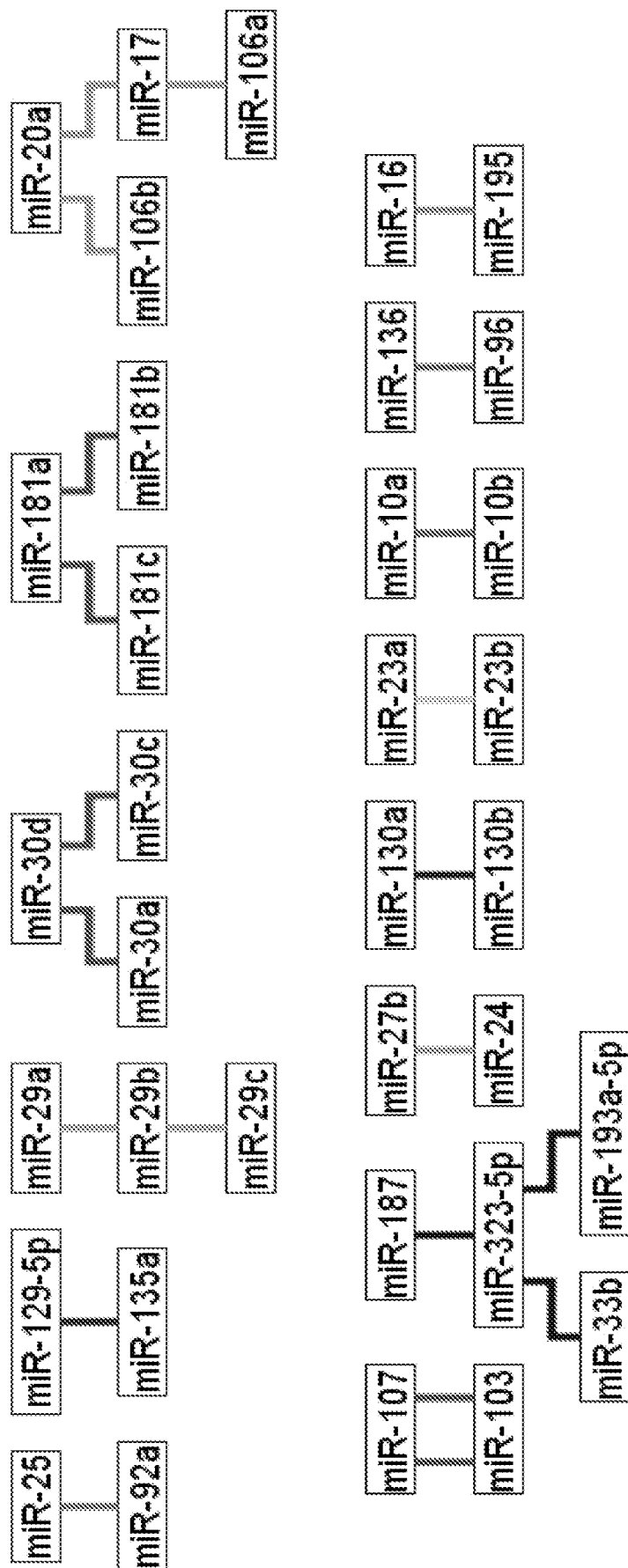


Figure 17

