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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 July 2009 (02.07.2009)

PCT

(10) International Publication Number
WO 2009/082744 A2

(51) International Patent Classification:
A61K 48/00 (2006.01)

(21) International Application Number:
PCT/US2008/088065

(22) International Filing Date:
22 December 2008 (22.12.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/008,643 22 December 2007 (22.12.2007) US
61/010,757 10 January 2008 (10.01.2008) US

(71) Applicant *(for all designated States except US)*: **SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH** [US/US]; Office Of Industrial Affairs, 1275 York Avenue, New York, NY 10065 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **MASSAGUE, Joan** [US/US]; c/o Memorial Sloan Kettering Cancer Center, Office Of Industrial Affairs, 1275 York Avenue, New York, NY 10065 (US). **TAVAZOIE, Sohail, F.** [US/US]; 325 E. 84th Street, Apt 6B, New York, NY 10028 (US).

(74) Agent: **LARSON, Marina, T.**; Larson & Anderson, LLC, P.O. Box 4928, Dillon, CO 80435-4928 (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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

WO 2009/082744 A2

(54) **Title:** PROGNOSIS AND INTERFERENCE-MEDIATED TREATMENT OF BREAST CANCER

(57) **Abstract:** Six genes that are overexpressed in metastatic breast cancer, and in particular in bone and lung metastases of breast cancer are identified, and whose overexpression also correlates with underexpression of miR-335, miR-206 and/or miR-126 have been identified. Testing for levels of any one or more of the gene products or miRNA provides a prognostic/diagnostic indicator of the likelihood or presence of breast cancer metastases. In addition, agents that decrease the amount of protein gene product and/or increase the amount of one of the relevant miRNA can be used in the treatment of metastatic breast cancer.

PROGNOSIS AND INTERFERENCE-MEDIATED TREATMENT OF BREAST CANCER

DESCRIPTION

This application claims the benefit of the filing date of U.S. Provisional Application No. 61/008,643, filed on 12/22/2007, and U.S. Provisional Application No. 61/010,757, filed on 01/10/2008 both of which are incorporated herein by reference.

Background of the Invention

Although metastasis is the overwhelming cause of mortality in patients with solid tumours, the understanding of its molecular and cellular determinants is limited. Fidler, I. J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nature Rev. Cancer 3, 453 458 (2003); Weigelt, B., Peterse, J. L. & van't Veer, L. J. Breast cancer metastasis: markers and models. Nature Rev. Cancer 5, 591 602 (2005); Gupta, G. P. & Massague, J. Cancer metastasis: building a framework. Cell 127, 679 695 (2006). Transcriptional profiling has revealed sets of genes, or 'signatures', for which expression in primary tumours correlates with metastatic relapse or poor survival. Nevins, J. R. & Potti, A. Mining gene expression profiles: expression signatures as cancer phenotypes. Nature Rev. Genet. 8, 601 609 (2007). Some of these genes endow cancer cells with a more invasive phenotype, enhanced angiogenic and intravasation activity, the ability to exit from the circulation, or an ability to modify the metastasis microenvironment. Nguyen, D. X. & Massague, J. Genetic determinants of cancer metastasis. Nature Rev. Genet. 8, 341 352 (2007); Gupta, G. P. et al. Mediators of vascular remodelling co opted for sequential steps in lung metastasis. Nature 446, 765 770 (2007). Such gene sets are thus providing numerous candidate mediators of metastasis to be validated through functional and clinical studies. Much less insight, however, has been gained into the regulatory networks that establish such altered gene expression states. Seligson, D. B. et al. Global histone modification patterns predict risk of prostate cancer recurrence. Nature 435, 1262-1266 (2005).

MicroRNAs (miRNAs) are attractive candidates as upstream regulators of metastatic progression because miRNAs can post-transcriptionally regulate entire sets of genes. Lim, L.

P. et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433, 769-773 (2005); Chan, C. S., Elemento, O. & Tavazoie, S. Revealing posttranscriptional regulatory elements through network-level conservation. PLoS Comput. Biol. 1, e69 (2005). Recent work has revealed important roles for miRNAs and miRNA processing in tumorigenesis. Voorhoeve, P. M. et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumours. Cell 124, 1169-1181 (2006); Kumar, M. S., Lu, J., Mercer, K. L., Golub, T. R. & Jacks, T. Impaired microRNA processing enhances cellular transformation and tumorigenesis. Nature Genet. 39, 673-677 (2007); Ma, L., Teruya-Feldstein, J. & Weinberg, R. A. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. Nature 449, 682-688 (2007). Large sets of miRNAs are underexpressed in human tumours compared to normal tissues (Lu, J. et al. MicroRNA expression profiles classify human cancers. Nature 435, 834-838 (2005), and interfering with miRNA processing enhances experimental tumorigenesis. Kumar, Nature Genet. These findings suggest that this class of regulators contains suppressors of tumour progression and possibly metastasis.

US Patent Publication No. 2008/0306006 and US 2006/0094061, which are incorporated herein by reference for the diagnostic and therapeutic techniques they disclose, both provide examples of the use of miRNAs in diagnosis and therapy for different types of cancer. These disclosures make it apparent that there are multiple miRNAs found in cancer tissues, and that different miRNAs are associated with different cancers.

Summary of the Invention

Applicants have investigated expression patterns of proteins and miRNAs in bone and lung metastases of breast cancer and identified an interrelated group of proteins which are overexpressed in these metastases and for which the overexpression correlates with underexpression of miR-335, miR-126 and/or miR-206. In accordance with the invention, these overexpressed proteins or underexpressed miRNA, individually or in combinations can serve as diagnostic markers or as a basis for treatment. Testing for levels of any one or more of the gene products or miRNA provides a prognostic/diagnostic indicator of the likelihood or presence of breast cancer metastases. In addition, agents that decrease the amount of protein gene product and/or increase the amount of the relevant miRNA can be used in the treatment of metastatic breast cancer.

Thus, the invention provides a method for treating metastatic breast cancer comprising administering to a patient diagnosed with breast cancer an agent, said agent

(a) comprising an oligonucleotide that reduces the expression of a gene in cancerous breast cells exposed to the oligonucleotide wherein the gene is overexpressed in both bone and lung metastases of breast cancer, and is down-regulated by miR-335, miR-126 and/or miR-206 and wherein the oligonucleotide has a sequence that is complementary to a portion of said gene or a transcript of said gene; and/or

(b) comprising an oligonucleotide that increases the amount of miR-335, miR-126 and/or miR-206, thereby treating the metastatic breast cancer.

Also disclosed is a method of diagnosing metastatic breast cancer in a patient comprising obtaining a diagnostic signature from breast cancer cells. The diagnostic signature may be obtained by measuring levels of RNA or protein that are the product of one or more genes selected from the group consisting of SRY-box containing transcription factor 4 (SOX4), Tenascin C (TNC), receptor-type tyrosine protein phosphatase (PTPRN), collagen, type I, alpha 1 (COL1A1), Mer receptor tyrosine kinase (MERTK), and phospholipase C beta 1 (PLCB1); comparing said diagnostic signature to a control signature; and based on the comparison, giving a prognosis of high risk for metastasis if the difference between the diagnostic signature and the control signature is greater than a threshold. Observed levels for miRNA selected from the from among miR-335, miR-126, and/or miR-206 can also be used for diagnostic purposes.

The invention also provides for the use of the diagnostic method in conjunction with the treatment method. This permits more aggressive treatment where a breast cancer is one that has a greater risk of metastases.

Also disclosed is the use of an oligonucleotide in the manufacture of a medicament for the treatment of metastatic breast cancer; wherein said oligonucleotide has a sequence that reduces the expression of a gene in cancerous breast cells exposed to the oligonucleotide; wherein the gene is overexpressed in both bone and lung metastases of a source cell; and wherein the oligonucleotide has a sequence that is complementary to a portion of said gene or a transcript of said gene.

Also disclosed is a composition comprising a pharmaceutically acceptable carrier, and an oligonucleotide that reduces the expression of human Tenascin C (Seq. ID No. 10). In a specific embodiment of this, the agent is an oligonucleotide comprising a sequence selected from the group consisting of Seq. ID No. 4, Seq. ID No. 5, Seq. ID No. 6, and Seq. ID No. 7.

Also disclosed is a kit for detection and characterization of breast cancer comprising reagents in packaged combination specifically for detecting or quantifying two or more markers selected from the group consisting of SRY-box containing transcription factor 4 (SOX4), Tenascin C (TNC), receptor-type tyrosine protein phosphatase (PTPRN), collagen, type I, alpha 1 (COL1A1), Mer receptor tyrosine kinase (MERTK), phospholipase C beta 1 (PLCB1), miR-335, miR-126 and miR-206.

Brief Description of the Drawings

Figs. 1A-D show systematic identification of miRNAs that suppress lung and bone metastasis in multiple human breast cancer cell derivatives.

Figs. 2A-C show miR-126 suppresses overall tumour growth and proliferation whereas miR-335 and miR-206 regulate migration and morphology.

Figs. 3A-B show clinical association of miR-335 and miR-126 with metastasis-free survival.

Figs. 4A-E show a miR-335-regulated gene set includes SOX4 as a miR-335 direct target. Figs. 5A-C show miR-335 regulates metastasis and invasion through suppression of SOX4 and TNC.

Fig. 6 shows stem-loop quantitative RT-PCR for the mature forms of the downregulated miRNAs identified through hybridization profiling.

Fig. 7 shows quantification of miRNA expression in LM2 cells with restored expression of miRNAs.

Fig. 8 shows bioluminescence quantification of lung colonization by additional miRNAs that are downregulated in metastatic cells.

Figs. 9A-C show loss of metastasis suppressor miRNA expression from primary human cancer cells as well as exogenously introduced vector subsequent to metastatic progression.

Fig. 10 shows miR-126 suppression of primary tumour growth.

Figs. 11A-B show immunohistochemical analysis for apoptotic and proliferation markers in mammary tumours.

Fig. 12 shows restoration of miR-335 and miR-206 expression in metastatic cells reduces the fraction of elongated cells.

Fig. 13 shows immunohistochemical analysis of the mesenchymal marker vimentin in LM2 cells expressing miR-335, miR-126, or miR-206.

Fig. 14 shows quantification of miRNA levels in human breast tumour samples. Each vertical bar represents expression levels of metastasis suppressor miRNAs miR-335, miR-206, and miR-126 in addition to miR-122a and miR-199a.

Fig. 15 shows the association between the loss of miR-206 expression in primary breast tumours and metastatic relapse.

Figs. 16A-B show the miR-335 regulated genes SOX4 and TNC alter the morphology of lung metastatic cells and suppress migration.

Figs. 17A-B show the miR-335 regulated six-gene signature predicts overall metastasis-free survival in the MSK-82 and EMC-286 datasets.

Fig. 18 shows miR-335 RNA mimetic delivery to breast cancer cells specifically suppresses the expression of metastasis promoter gene SOX4 similarly as virally expressed miR-335.

Fig. 19 shows miR-126 RNA mimetic delivery to breast cancer cells suppresses proliferation of metastatic cells similarly as virally expressed miR-126.

Detailed Description of the Invention

Based on an extensive evaluation of the genes that are overexpressed in bone and lung metastases of breast cancer, identification of the miRNA that are underexpressed in cells in a manner that correlates with the observation of overexpression of particular proteins, and then further testing to determine the miRNAs that actually have have the ability, when present at normal levels to prevent the overexpression of the proteins, Applicants have identified six genes that are overexpressed and three miRNA that are underexpressed in metastatic breast cancer, and in particular in bone and lung metastases of breast cancer. Testing for levels of any one or more of the products (RNA or protein) of the six genes, and/or levels of one or more of the three miRNA provides a prognostic/diagnostic indicator of the likelihood or presence of breast cancer metastases. In addition, agents that decrease the amount of protein gene product and/or decrease the amount of the relevant miRNA can be used in the treatment of metastatic breast cancer.

Definitions

For the purposes of this specification and claims, the following terms are to have the specified meaning. All other terms are to be defined as used in the specification or used by a person having ordinary skill in the art.

As used in this application "oligonucleotides" means a short sequence of nucleic acids, no more than 200 bases long. An oligonucleotide may be made of DNA, RNA or a combination, and may be chemically modified to increase stability.

As used in this application, "complementary" refers to the base pairing relationshiop between strands of nucleic acids arranged in an antiparallel manner, such that the strands will hybridize with sufficient strength under intracellular conditions to change expression of a gene.

As used in this application, the terms 'treatment' or 'treating' refer to the application of an agent with the intent to produce a benefit to a patient being treated. Such a benefit need not be a complete or permanent cure, but may be only a lessening of the rate at which tumorigenesis is occurring, thereby delaying onset or progression of metastases, particularly to bone and lung. In addition, such treatment need not in actuality produce the beneficial, as long is it is given with the intention of obtaining a result.

As used in this application, the term "gene product" refers to the product of transcription or translation of a gene sequence. This may include RNA or protein products, including or regardless of later processing or conformational changes.

As used in this application, the term "3'-UTR" or "3'-untranslated region" is a particular section of messenger RNA (mRNA) that follows the coding region. This region contains transcription and translation regulating sequences.

As used in this application, the term "genetic signature" refers to the numerical value of the level of expression of one or more genes in a tissue sample. A genetic signature can be based on a single gene/gene product, or it may be a composite of values for multiple genes/gene products.

As used in this application, the term "control signature" means an average genetic signature obtained from measurements of multiple samples from non-metastasizing breast cells (normal or cancerous). The sample size should be large enough to be statistically relevant. It will be apparent to one skilled in the art that a control signature and the genetic signature to which it is compared are obtained using the same test conditions, although they need not be done at the same time.

As used in this application, the term "reduce the expression of a gene" refers to the reduction of the amount of a product of the gene. This includes RNA and proteins in all forms.

As used in this application, "shRNA" means short hairpin RNA. It is a sequence of RNA that makes a tight hairpin turn that can be used to silence gene expression via RNA interference. The shRNA hairpin structure is cleaved by the cellular machinery and are precursors to miRNA.

As used in this application, "microRNAs" (miRNA) are single-stranded RNA molecules generally of less than 24 nucleotides in length excised from 60- to 110-nucleotide shRNA precursors involved in the regulation of expression of protein-coding genes (PCGs), which regulate gene expression. miRNAs are encoded by genes from whose DNA they are transcribed but miRNAs are not translated into protein (non-coding RNA); instead each primary transcript (a pri-miRNA) is processed into a short stem-loop structure called a pre-miRNA and finally into a functional miRNA. Mature miRNA molecules are partially complementary to one or more messenger RNA (mRNA) molecules, and their main function is to down-regulate gene expression. miR-335 has the sequence set forth in Seq. ID No. 11. miR-126 has the sequence set forth in Seq. ID No. 12. miR-206 has the sequence set forth in Seq. ID No. 13.

In developing this invention, Applicants tested numerous samples from bone and lung metastases of breast cancer to determine whether any genes or proteins were overexpressed. When compared to nonmetastasizing breast cancer samples, six genes were overexpressed and 3 miRNAs were underexpressed in both. The genes were SRY-box containing transcription factor 4 (SOX4), Tenascin C (TNC), receptor-type tyrosine protein phosphatase (PTPRN), collagen, type I, alpha 1 (COL1A1), Mer receptor tyrosine kinase (MERTK), and holipase C beta 1 (PLCB1). The three miRNA were miR-335, miR-126, and miR-206.

Method of Diagnosing Metastatic Potential/Determination of a Genetic Signature

One embodiment of the present invention is a method for diagnosing metastatic potential of breast cancer for a patient, through the determination of a genetic signature and comparison of that signature to a control signature. In accordance with the invention, this can be accomplished through a determination of the amount of one or more markers, wherein the markers are gene products (protein or RNA) selected from the group consisting of SOX4, TNC, PTPRN, COL1A1), MERTK and PLCB1, and/or one or more micro RNAs selected from the gorup consisting of miR-335, miR-126, and miR-206. In specific embodiments, the marker(s) include SOX4. In specific embodiments, the marker(s) include TNC. In one specific embodiments, al six proteins are used as markers. In specific embodiments, miR-335 and/or miR-126 are used as markers.

The genetic signature is obtained on a sample of breast cancer cells obtained from a patient. These can be metastatic cells or cells that have not yet undergone metastases. In either case, a characterization of the metastatic potential of the cells is obtained which can be used in counseling and selection of treatment options, including the selection or very aggressive treatment or no treatment where the cancer has a high metastatic potential.

The genetic signature is compared to a control signature obtained from non-cancerous breast cells. When the comparison is made, if a protein marker is present at higher than a threshold level, or an miRNA marker is present at lower than a threshold level, then the breast cancer from the patient is considered to be one with high metastatic potential. It will be appreciated that the determination of specific numerical values for the threshold is dependent on the particular tests that are included in the formation of the signatures, and the level of risk that is to be assigned as high risk. By way of example, however, when all six proteins are tested using the procedures described below, and the sum of the six expression levels is the genetic signature, the threshold level is suitably the sum of the six expression levels of the control signature plus one standard deviation for the control signature.

Detection of the gene products or miRNAs may be done using any of various known methods from detection or proteins or RNA, including without limitation: a) polymerase chain reaction (PCR) technology, for example quantitative PCR of one or more nucleic acids, preceded by retro-transcription in case the sample to be analyzed is RNA, another example is stem-loop quantitative RT-PCR which is a modification of RT-PCR specially adapted to the amplification and quantification of microRNAs, b) nucleic acid hybridization technologies, for example nucleic acid gel electrophoresis followed by transfer to a membrane and

posterior hybridization (Northern/Southern blot), or hybridization to membranes where the nucleic acid is deposited (purified or not or as a fraction) directly to a membrane (slot or dot blot), or direct hybridization into tissue or cells mounted on a slide; c) microarray technology. Although those technologies all have the aim of quantifying nucleic acids contained in a sample, a profile can also be obtained by determining the presence, absence or quantity of a specific protein or proteins, and such profiling can be done using methods well known in the art like ELISA, immunohistochemistry or Western blot, that use polyclonal or monoclonal antibodies able to recognize specific proteins.

There are several methods to obtain breast cancer cells for analysis according to the methods of the present invention. One such method is biopsy, removing a sample of breast tissue containing cancer. Further the tissue can be separated into cancer cells and non-cancer cells and the cancer cells processed for further analysis of DNA, RNA or protein. Another method is obtaining cancer cells from bodily fluids for example ductal lavages or blood according to methods known in the art.

An embodiment of the present invention provides a method to provide to a breast cancer patient a prognosis of metastasis-free survival consisting of obtaining breast tumor cells from that breast cancer patient, determining the RNA or protein levels of at least one of the following: COL1A gene, PTPRN gene, MERKT gene, PLCB1 gene, TNC gene and SOX4 gene, determining an aggregate expression score based on the expression of at least those genes or proteins, the miR-335-signature, and considering those patients whose expression scores exceed one standard deviation from the mean to be miR-335-signature-positive and therefore to have worse metastasis-free survival prognosis than those patients that are not miR-335- signature-positive.

Another embodiment of the present invention is a method to determine whether a breast cancer patient should have aggressive treatment and follow up after initial diagnosis of breast cancer, the method consisting of determining that patient's signature status. If the patient is positive, that patient is considered to be at high risk for metastasis and therefore needing more aggressive treatment and follow up, to prevent or identify metastasis. If the patient signature-negative that patient is considered to be at a low risk for metastasis. The signature can include one or more of the following: miR-335, -206 and -126 and the genes collagen, type I, alpha 1 (COL1A1), receptor-type tyrosine protein phosphatase (PTPRN), c-Mer tyrosine kinase (MERTK), phospholipase C beta 1 (PLCB1), tenascin C (TNC) and

SRY-box containing transcription factor (SOX4). Preferred targets for the diagnosis include SOX4, TNC, and miR-335.

Another embodiment of the present invention is a method to provide to a breast cancer patient a prognosis of metastasis-free survival consisting of obtaining breast tumor cells, or breast tumor tissue, from that breast cancer patient, determining the RNA or gene expression levels on those breast tumor cells, or breast tumor tissue, determining an expression score based on the expression of these targets and considering patients whose expression scores are below the mean expression level determined from a random sample of primary breast cancer cells or breast cancer tissue to have worse metastasis-free survival prognosis than those patients that are above such mean expression level. The targets can include one or more of the following: miR-335, -206 and -126 and the genes collagen, type I, alpha 1 (COL1A1), receptor-type tyrosine protein phosphatase (PTPRN), c-Mer tyrosine kinase (MERTK), phospholipase C beta 1 (PLCB1), tenascin C (TNC) and SRY-box containing transcription factor (SOX4). Preferred targets for the diagnosis include SOX4, TNC, and miR-335

Treatment

The present invention also teaches a method of treating metastatic breast cancer. The metastatic breast cancer may be determined using the above techniques, or any other technique. The treatment is done by administering oligonucleotides that will increase the levels of underexpressed miRNA or decrease the level of over expressed genes. There are numerous types of such oligonucleotides, including: antisense oligonucleotides, ribozymes, shRNA, miRNA, siNA, and dsRNA. The amount given is sufficient to cause a change in the expression of the target.

The terms RNA, RNA molecule(s), RNA segment(s) and RNA fragment(s) may be used interchangeably to refer to RNA that mediates RNA interference. These terms include double-stranded RNA, single-stranded RNA, isolated RNA (partially purified RNA, essentially pure RNA, synthetic RNA, recombinantly produced RNA), as well as altered RNA that differs from naturally occurring RNA by the addition, deletion, substitution and/or alteration of one or more nucleotides. Such alterations can include addition of non-nucleotide material, such as to the end(s) of the RNA or internally (at one or more nucleotides of the RNA). Nucleotides in the RNA molecules of the present invention can also comprise non-standard nucleotides, including non-naturally occurring nucleotides or deoxyribonucleotides.

Collectively, all such altered RNAi compounds are referred to as analogs or analogs of naturally-occurring RNA. RNA of the present invention need only be sufficiently similar to natural RNA that it has the ability to mediate RNAi. As used herein the phrase "mediate RNAi" refers to and indicates the ability to distinguish which mRNA are to be affected by the RNAi machinery or process. RNA that mediates RNAi interacts with the RNAi machinery such that it directs the machinery to degrade particular mRNAs or to otherwise reduce the expression of the target protein. In one embodiment, the present invention relates to RNA molecules that direct cleavage of specific mRNA to which their sequence corresponds. It is not necessary that there be perfect correspondence of the sequences, but the correspondence must be sufficient to enable the RNA to direct RNAi inhibition by cleavage or lack of expression of the target mRNA.

RNA interference or "RNAi" is a term initially coined by Fire and co-workers to describe the observation that double-stranded RNA (dsRNA) can block gene expression when it is introduced into worms (Fire et al. (1998) Nature 391, 806-811, incorporated herein by reference). dsRNA directs gene-specific, post-transcriptional silencing in many organisms, including vertebrates, and has provided a new tool for studying gene function. RNAi involves mRNA degradation, but many of the biochemical mechanisms underlying this interference are unknown. The use of RNAi has been further described in Carthew et al. (2001) Current Opinions in Cell Biology 13, 244-248, and Elbashir et al. (2001) Nature 411, 494-498, both of which are incorporated herein by reference. The RNAi molecules of the invention are double-stranded or single-stranded RNA of from about 21 to about 23 nucleotides which mediate RNA inhibition. That is, the isolated RNAi of the present invention mediate degradation of mRNA of the hsp27 gene.

siRNA can be introduced directly into a cell to mediate RNAi (Elbashir et al., 2001, Nature 411:494 498). Many methods have been developed to make siRNA, e.g, chemical synthesis or in vitro transcription. Once made, the siRNAs are introduced into cells via transient transfection. A number of expression vectors have also been developed to continually express siRNAs in transiently and stably transfected mammalian cells (Brummelkamp et al., 2002 Science 296:550 553; Sui et al., 2002, PNAS 99(6):5515 5520; Paul et al., 2002, Nature Biotechnol. 20:505 508). Some of these vectors have been engineered to express shRNAs, which get processed in vivo into siRNA-like molecules capable of carrying out gene-specific silencing. Another type of siRNA expression vector encodes the sense and antisense siRNA strands under control of separate pol III promoters

(Miyagishi and Taira, 2002, Nature Biotechnol. 20:497 500). The siRNA strands from this vector, like the shRNAs of the other vectors, have 3' thymidine termination signals. Silencing efficacy by both types of expression vectors was comparable to that induced by transiently transfecting siRNA.

Antisense technology is also useful for reducing expression of a gene. When the genetic sequence of a particular gene is known to be causative of a particular disease, it is possible to synthesize a strand of nucleic acid (DNA, RNA or a chemical analogue) that will bind to the messenger RNA (mRNA) produced by that gene and inactivate it, effectively turning that gene "off". This is because mRNA has to be single stranded for it to be translated. Alternatively, the strand might be targeted to bind a splicing site on pre-mRNA and modify the exon content of an mRNA. Antisense oligonucleotides act by blocking translation or causing degradation of mRNA. This synthesized nucleic acid is termed an "anti-sense" oligonucleotide because its base sequence is complementary to the gene's messenger RNA (mRNA), which is called the "sense" sequence (so that a sense segment of mRNA 5'-AAGGUC-3' would be blocked by the anti-sense RNA segment 3'-UUCCAG-5').

The present invention may be used to introduce oligonucleotides into a cell for the treatment or prevention of disease. For example, dsRNA may be introduced into a cancerous cell or tumor and thereby inhibit gene expression of a gene required for maintenance of the carcinogenic/ tumorigenic/ metastatic phenotype. To prevent a disease or other pathology, a target gene may be selected which is required for initiation or maintenance of the disease/pathology. Treatment would include amelioration of any symptom associated with the disease or clinical indication associated with the pathology.

One method to introduce such constructs in a cell is by electroporation, another method is by transfection, yet another method is by viral mediated transfection. Yet other methods to introduce such DNA constructs into an organism, like a mammal, is by using tools available for gene therapy. For example, a vector encoding a nucleic acid molecule of the invention (e.g., a viral vector) can be administered to a cell. Viral vectors include, without limitation, a lentivirus, an adenovirus, an adeno-associated virus, a retrovirus, a vaccinia virus, a herpes viruses, and a bovine papilloma virus. . In addition, a nucleic acid molecule of the invention or a vector encoding such a nucleic acid can be encapsulated in, for example, a nanoparticle or a liposome, and administered to a cell.

The present invention also concerns methods and reagents useful in reducing metastatic characteristics of cancer cells therefore providing a therapy useful in the treatment

of cancer. The present invention introduces RNA or DNA able to express RNA into a cell for the treatment or prevention of breast cancer. For example, dsRNA may be introduced into a cancerous cell or tumor and thereby inhibit gene expression of a gene required for maintenance of the carcinogenic/tumorigenic phenotype. To prevent a disease or other pathology, a target gene may be selected which is required for initiation or maintenance of the disease/pathology. Treatment would include amelioration of any symptom associated with the disease or clinical indication associated with the pathology.

The present invention also relates to genes that are overexpressed in certain breast cancers, making the breast cancer more prone to growth and metastatic behavior. Such genes can be targeted individually or in groups for higher therapeutic effect.

Another embodiment of the present invention is a method of treating breast cancer consisting of providing to a patient in need of such treatment small nucleic acid molecules, or a plasmid able to drive the expression of such small nucleic acid molecules, such as siNA, siRNA, dsRNA, miRNA, or shRNA capable of mediating RNAi that contain the sequence Seq. ID No. 1.

Another embodiment of the present invention is a method of treating breast cancer consisting of providing to a patient in need of such treatment small nucleic acid molecules, or a plasmid able to drive the expression of such small nucleic acid molecules, such as siNA, siRNA, dsRNA, miRNA, shRNA capable of mediating RNAi that contains the sequence Seq. ID No. 2.

A method of treating breast cancer is disclosed consisting of providing to a patient in need of such treatment small nucleic acid molecules, or a plasmid able to drive the expression of such small nucleic acid molecules, such as siNA, siRNA, dsRNA, miRNA, shRNA molecules capable of mediating RNAi that contain the sequence Seq. ID No. 3.

Another embodiment of the present invention is a method of treating breast cancer consisting of providing to a patient in need of such treatment an oligonucleotide comprising Seq. ID No. 21 or Seq. ID No. 22.

Pharmaceutical composition

The present invention also discloses compositions that are useful in treating metastasizing breast cancer. This includes oligonucleotides capable of regulating expression of genes and miRNA that are abnormally expressed in metastases. These oligonucleotides may be: antisense oligonucleotides, ribozymes, shRNA, miRNA, siNA, and dsRNA.

It is the aim of the present invention to provide molecules that can be useful in the treatment of cancer, specifically breast cancer. Moreover, the molecules provided in this invention are also useful in limiting and/or eliminating the metastatic potential of breast cancer cells. Molecules useful in the practice of the present invention are molecules that can mediate RNAi of one or more of the following genes: collagen, type I, alpha 1 (COL1A1), receptor-type tyrosine protein phosphatase (PTPRN), c-Mer receptor tyrosine kinase (MERTK), phospholipase C beta 1 (PLCB1), tenascin C (TNC) and SRY-box containing transcription factor (SOX4). Such molecules contain a sequence of at least 20 nucleotides that is at least 90% identically, but ideally 100% identical, to the complementary sequence of the RNA that is being targeted, ie the RNA of one or more of those genes. Once such an oligonucleotide sequence has been identified, it can be used to create a variety of siNA, siRNA, dsRNA, miRNA, shRNA molecules capable of mediating RNAi. All such molecules contain the short oligonucleotide sequence or a sequence at least 90% identical to it.

The identification of nucleotide molecules, and the corresponding protein product in the case of genes, that are either increased or decreased in breast cancer provides an opportunity for therapy of such cancer. Specifically genes that are overexpressed in such cancers with respect to normal cells or genes that are overexpressed in certain breast cancer types that have bad prognosis with respect to breast cancer types that have a good prognosis, provide an avenue for targeted therapy. Such targeted therapy can be implemented using small molecules in the case of protein products that have enzymatic functions that can be targeted through the use of such small molecules.

Moreover, targeted inhibition of gene expression may also be implemented via the use of polynucleotide compounds, such as but not limited to, antisense polynucleotides, ribozymes, RNA interference molecules, triple helix polynucleotides and the like, where the nucleotide sequence of such compounds is related to the RNA sequence that is intended to be targeted. Antisense technology has been the most commonly described approach in protocols to achieve sequence-specific interference. For antisense strategies, stoichiometric amounts of single-stranded nucleic acid perfectly or partially complementary to the RNA of interest are introduced into the cell. Another method for engineered interference is based on a triple helical nucleic acid structure. This approach relies on the ability of certain nucleic acid populations to adopt a triple-stranded structure.

The present invention relates to types of molecules identified for targeted therapy of breast cancer. One is a series of RNA molecules that are naturally found in a human cell and

that control the expression, through RNAi, of a series of genes. Such molecules have been found to be at low levels in certain breast cancers. It is the object of the present invention to provide a therapy that can increase the level of those molecules in breast cancer cells so that the expression of the genes they control is lowered so that the phenotype of such breast cancer cell becomes less malignant and its metastatic ability is either eliminated or reduced. Such RNA molecules are of the type known as microRNAs. They are able to induce RNAi of a group of genes through a core sequence in their short hairpin RNA (shRNA) that is homologous to the RNA sequence of the target gene of less than 15 nucleotides. This invention describes those microRNAs and their shRNA sequence. However the same RNAi can be accomplished through the use of other RNA types as long as a core sequence of around 21 nucleotides is present in a double stranded form, with perfect or partially imperfect sequence complementarity in that double strand.

The oligonucleotide may comprise one or more strands of polymerized ribonucleotide. It may include modifications to either the phosphate-sugar backbone or the nucleoside. For example, the phosphodiester linkages of natural RNA may be modified to include at least one of a nitrogen or sulfur heteroatom. Modifications in oligonucleotide structure may be tailored to allow specific genetic inhibition while avoiding a general panic response in some organisms which is generated by dsRNA. Likewise, bases may be modified to block the activity of adenosine deaminase. oligonucleotides may be produced enzymatically or by partial/total organic synthesis. Any modified ribonucleotide can be introduced by in vitro enzymatic or organic synthesis.

Oligonucleotides containing a nucleotide sequences identical to a portion of the target gene are preferred for inhibition as this provides greater specificity and controls for non-desirable targeting. However, oligonucleotides sequences with insertions, deletions, and single point mutations relative to the target sequence have also been found to be effective for inhibition. Thus, sequence identity may be optimized by sequence comparison and alignment algorithms known in the art (see http://www.genelink.com/sirna/RNAicustomorder.asp; Gribskov and Devereux, Sequence Analysis Primer, Stockton Press, 1991, and references cited therein) and calculating the percent difference between the nucleotide sequences by, for example, the Smith-Waterman algorithm as implemented in the BESTFIT software program using default parameters (e.g., University of Wisconsin Genetic Computing Group). Greater than 90% sequence identity, or even 100% sequence identity, between the inhibitory oligonucleotides and the portion of the target gene is preferred. Alternatively, the duplex

region of the RNA may be defined functionally as a nucleotide sequence that is capable of hybridizing with a portion of the target gene transcript (e.g., 400 mM NaCl, 40 mM PIPES pH 6.4, 1 mM EDTA, 50.degree. C. or 70.degree. C. hybridization for 12 16 hours; followed by washing). The length of the identical nucleotide sequences may be at least 25, 50, 100, 200, 300 or 400 bases.

One hundred percent sequence identity between the oligonucleotides and the target gene is not required to practice the present invention. Thus the invention has the advantage of being able to tolerate sequence variations that might be expected due to genetic mutation, strain polymorphism, or evolutionary divergence.

Oligonucleotides may be synthesized either in vivo or in vitro. Endogenous polymerase of the cell may mediate transcription in vivo, or cloned polymerase can be used for transcription in vivo or in vitro. For transcription from a transgene in vivo or an expression construct, a regulatory region (e.g., promoter, enhancer, silencer, splice donor and acceptor, polyadenylation) may be used to transcribe the oligonucleotides strand (or strands). Inhibition may be targeted by specific transcription in an organ, tissue, or cell type; stimulation of an environmental condition (e.g., infection, stress, temperature, chemical inducers); and/or engineering transcription at a developmental stage or age. The oligonucleotide strands may or may not be polyadenylated; the oligonucleotide strands may or may not be capable of being translated into a polypeptide by a cell's translational apparatus. Oligonucleotides may be chemically or enzymatically synthesized by manual or automated reactions. The oligonucleotides may be synthesized by a cellular RNA polymerase or a bacteriophage RNA polymerase (e.g., T3, T7, SP6). The use and production of an expression construct are known in the art (see also WO 97/32016; U.S. Pat. Nos. 5,593,874, 5,698,425, 5,712,135, 5,789,214, and 5,804,693; and the references cited therein). If synthesized chemically or by in vitro enzymatic synthesis, the oligonucleotides may be purified prior to introduction into the cell. For example, RNA can be purified from a mixture by extraction with a solvent or resin, precipitation, electrophoresis, chromatography, or a combination thereof. Alternatively, the oligonucleotides may be used with no or a minimum of purification to avoid losses due to sample processing. The oligonucleotides may be dried for storage or dissolved in an aqueous solution. The solution may contain buffers or salts to promote annealing, and/or stabilization of the duplex strands.

Oligonucleotides may be directly introduced into the cell (i.e., intracellularly); or introduced extracellularly into a cavity, interstitial space, into the circulation of an organism,

introduced orally, or may be introduced by bathing an organism in a solution containing the oligonucleotides. Physical methods of introducing nucleic acids, for example, injection directly into the cell or extracellular injection into the organism, may also be used. Vascular or extravascular circulation, the blood or lymph system, and the cerebrospinal fluid are sites where the oligonucleotides may be introduced.

In preferred embodiments, an RNAi molecule is used to decrease gene expression of genes related to metastasis. Since the only RNA molecules normally found in the cytoplasm of a cell are molecules of single-stranded mRNA, the cell has enzymes that recognize and cut dsRNA into fragments containing 21-25 base pairs (approximately two turns of a double helix and which are referred to as siRNA). The antisense strand of the fragment separates enough from the sense strand so that it hybridizes with the complementary sense sequence on a molecule of endogenous cellular mRNA. This hybridization triggers cutting of the mRNA in the double-stranded region, thus destroying its ability to be translated into a polypeptide. Introducing dsRNA corresponding to a particular gene thus knocks out the cell's own expression of that gene in particular tissues and/or at a chosen time.

The siRNAs of the present invention can be generated by any method known in the art, for example, by in vitro transcription, recombinantly, or by synthetic means (e.g., having either a TT or a UU overhang at the 3' end). siRNAs can be generated in vitro by using a recombinant enzyme, such as T7 RNA polymerase, and DNA oligonucleotide templates, or can be prepared in vivo, for example, in cultured cells (*See, e.g.*, Elbashir et al., *supra*; Brummelkamp et al., *supra*; and Lee et al., 2002, Nat. Biotech., 20:500).

Moreover, DNA constructs or plasmids containing DNA versions of such siRNA, dsRNA, miRNA or shRNA downstream of adequate promoters can be made so that, once such constructs are inside the cell, such RNAi molecules are continuously produced. One method to produce such constructs is described in Harper SQ, Davidson BL. Methods Mol Biol. 2005;309:219-35, but there are many other methods that are known in the art. In addition, strategies have been described for producing a hairpin siRNA from vectors containing a RNA polymerase III promoter. Various vectors have been constructed for generating hairpin siRNAs in host cells using either an H1-RNA or an snU6 RNA promoter. An RNA molecule as described above (e.g., a first portion, a linking sequence, and a second portion) can be operably linked to such a promoter. When transcribed by RNA polymerase III, the first and second portions form a duplexed stem of a hairpin and the linking sequence

forms a loop. The pSuper vector (OligoEngines Ltd., Seattle, Wash.) also can be used to generate siRNA.

A TTTTT penta-nucleotide usually is attached to the end of the second portion (i.e., the antisense strand) in a vector to serve as a terminator for RNA polymerase III transcription. For that reason, siRNA candidates that contain more than three consecutive Ts should be avoided since four or more consecutive Ts in the template nucleic acid triggers termination of RNA polymerase III transcription.

There are well-established criteria for designing siRNAs (*See, e.g.,* Elbashire et al., 2001, Nature, 411:494 8; Amarzguioui et al., 2004, Biochem. Biophys. Res. Commun., 316(4):1050 8; Reynolds et al., 2004, Nat. Biotech., 22(3):326-30). Details can be found in the websites of several commercial vendors such as Ambion, Dharmacon, GenScript, and OligoEngine. Other such algorithm can be found at http://www.genelink.com/sirna/RNAicustomorder.asp. The sequence of any potential siRNA candidate generally is checked for any possible matches to other nucleic acid sequences or polymorphisms of nucleic acid sequence using the BLAST alignment program (*See* ncbi.nlm.nih.gov on the World Wide Web). Typically, a few iRNAs have to be generated and screened in order to compare their effectiveness.

Double-stranded (ds) RNA can be used to interfere with gene expression in mammals. dsRNA is used as RNAi of the function of a nucleic acid molecule of the invention to produce a phenotype that is the same as that of a null mutant of a nucleic acid molecule of the invention (Wianny & Zernicka-Goetz, 2000, Nature Cell Biology 2: 70 75). The double-stranded structure may be formed by a single self-complementary RNA strand or two complementary RNA strands. RNA duplex formation may be initiated either inside or outside the cell. The RNA may be introduced in an amount which allows delivery of at least one copy per cell. Higher doses (e.g., at least 5, 10, 100, 500 or 1000 copies per cell) of double-stranded material may yield more effective inhibition; lower doses may also be useful for specific applications. Inhibition is sequence-specific in that nucleotide sequences corresponding to the duplex region of the RNA are targeted for genetic inhibition. The RNA molecule may be at least 10, 12, 15, 20, 21, 22, 23, 24, 25, 30, nucleotides in length.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. For example, the therapeutic agent may be introduced directly into the cancer of interest via direct injection. Additionally, examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g.,

ingestion or inhalation), transdermal (topical), transmucosal, and rectal administration. siRNA molecules of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and anti-fungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art.

Solutions or suspensions can include the following components: a sterile diluent such as water for injection, saline solution (e.g., phosphate buffered saline (PBS)), fixed oils, a polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), glycerine, or other synthetic solvents; antibacterial and antifungal agents such as parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol or sorbitol, and sodium chloride in the composition. Prolonged administration of the injectable compositions can be brought about by including an agent that delays absorption. Such agents include, for example, aluminum monostearate and gelatin. The parenteral preparation can be enclosed in ampules, disposable syringes, or multiple dose vials made of glass or plastic.

Oral compositions generally include an inert diluent or an edible carrier. Oral compositions can be liquid, or can be enclosed in gelatin capsules or compressed into tablets. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of an oral composition. Tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose; a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring. Transmucosal administration can be accomplished through the use of nasal sprays or

suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

It is especially advantageous to formulate compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for an individual to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The dosage unit forms of the invention are dependent upon the amount of a compound necessary to inhibit proliferation of the cancer cells. The amount of a compound necessary to inhibit proliferation of the cancer cells can be formulated in a single dose, or can be formulated in multiple dosage units. Treatment may require a one-time dose, or may require repeated doses.

Specific species of oligonucleotides that can be used in the present invention are short interfering nucleic acid (siNA), siRNA, double-stranded RNA (dsRNA), micro-RNA (miRNA), and short hairpin RNA (shRNA) molecules capable of mediating RNAi that contain the sequence of Seq. ID Nos. 1, 2, or 3. Other specific sequences are plasmid constructs that drive the expression of a small nucleic acid molecule, such as short interfering nucleic acid (siNA), siRNA, double-stranded RNA (dsRNA), micro-RNA (miRNA), and short hairpin RNA (shRNA) capable of mediating RNAi that contain the sequence: Seq. ID No. 1, Seq. ID No. 2, or Seq. ID No. 3.

The present the invention also relates to small nucleic acid molecules, such assiNA, siRNA, dsRNA, miRNA, or shRNA capable of mediating RNAi against any of COL1A, PTPRN, MERKT, PLCB1, TNC and SOX4 RNAs useful for the treatment of breast cancer.

Specific examples of an shRNA molecule capable of mediating RNAi against Tenascin C, are given by Seq. ID No. 4 or Seq. ID No. 5. Other examples

Another embodiment of the present invention is a plasmid construct containing a DNA sequence of the shRNA molecule capable of mediating RNAi against Tenascin C, the shRNA containing the sequence: Seq. ID No. 4 or Seq. ID No. 5. Antisense has been shown to regulate Tenascin C in pulmonary hypertension using Seq. ID No. 21 and Seq. ID No. 22. Cowan, K. et al, Elastase and matrix metalloproteinase inhibitors induce regression, and tenascin-C antisense prevents progression, of vascular disease, J. Clin. Invest. 2000 January 1; 105(1): 21–34. RNAi interference with sequences target the first 164 bases starting from the 5'-end of TNC has been demonstrated by Zukiel et al, *Cancer biology & therapy* 2006:

Vol. 5, pp. 1002-1007; "Suppression of human brain tumor with interference RNA specific for tenascin-C" which is incorporated herein by reference.

Another embodiment of the present invention is a siNA molecule, or a plasmid construct able to drive the expression of such siNA, capable of mediating RNAi against Tenascin C, the siRNA containing the sequence: Seq. ID No. 6.

Another embodiment of the present invention is a siNA molecule, or a plasmid construct able to drive the expression of such siNA, capable of mediating RNAi against Tenascin C, the siRNA containing the sequence: Seq. ID No. 7.

Additional oligonucleotides capable of reducing expression of COL1A1 are known from Millington-Ward et al., *Eur. J Human Genetics* (2004) 12, 864-866, and Wang et al, *Experimental dermatology* 2007-Jul; vol 16 (issue 7) : pp 611-7, which are incorporated herien by reference.

Additional oligonucleotides capable of reducing expression of PLCB1 are known from A. Diaz Añel, Biochem J. 2007 August 15; 406(Pt 1): 157-165, which is incorporated herein by reference.

Additional oligoncleotides capable of reducing expression of SOX4 are known from Pramoonjago et al, . *Oncogene* (2006) 25, 5626-5639, which is incorporated herein by reference.

Additional oligonucleotides capable of reducing expression of PTPRN are known from US patent Publication No. 2006/0094061 which is incorporated herein by reference.

Kit

The diagnostic assay described above can be suitably practiced using just one of the markers (genes or miRNA) and provide an acceptable result. However, it may be desirable to perform tests on a sample for more than one of the markers. This is facilitated by a kit that includes at least the sequence specific reagents for the detection and quantification of two or more of the markers, i.e. two or more of SRY-box containing transcription factor 4 (SOX4), Tenascin C (TNC), receptor-type tyrosine protein phosphatase (PTPRN), collagen, type I, alpha 1 (COL1A1), Mer receptor tyrosine kinase (MERTK), phospholipase C beta 1 (PLCB1), miR-335, miR-126, and miR-206.

As discussed above, suitable reagents include antibodies, oligonucleotide probes, polymerases and reverse transcriptase. The kit may also include buffers, enzymes, membranes and gels suitable for the assays.

Preferred kits include reagents for detection of SOX4, TNC or miR335 as one of the reagent types. Other embodiments of kits include reagents for detection of SOX4 and TNC, or SOX4, TNC and or miR335.

Standard values for the control signature (mean and standard deviation) for a detected/quantified material in non-metastatic breast cells may be included in the kit, for example as a printed insert.

The present invention is illustrated by the following examples. These examples are not meant in any way to limit the scope of this invention.

Example 1

Identification of miRNAs that suppress metastasis

We performed array-based miRNA profiling (Calin, G. A. et al. MicroRNA profiling reveals distinct signatures in B cell chronic lymphocytic leukemias. Proc. Natl Acad. Sci. USA 101, 11755-11760 (2004)) of MDA-MB-231 human breast cancer cell derivatives that are highly metastatic to bone (1833 or 2287; denoted BoM1 lines) or lung (4175, 4173, 4180 and 4142; LM2 lines), as well as the parental, unselected MDA-MB-231 cell population. Out of 453 human miRNAs assayed, 179 miRNAs were expressed above background levels in one or more of the highly metastatic cells (data not shown). Notably, hierarchical clustering based on the expression of these miRNAs correctly classified the MDA-MB-231 derivatives into three groups comprising the BoM1 lines, the LM2 lines and the parental lines, respectively. Hierarchical clustering on the basis of the expression of the 20 miRNAs for which expression was most altered across various cell lines again correctly classified the MDA-MB-231 derivatives into these three groups. The most salient finding was a set of eight miRNAs for which expression was decreased across all metastatic sub-lines compared with their expression in the parental line. Through quantitative stem-loop polymerase chain reaction (qRT-PCR), we were able to validate the differential expression of seven of these miRNAs in MDA-MB-231 and representative LM2 and BoM1 sub-lines (Fig. 6).

We focused on the six miRNAs (miR-335, miR-126, miR-206, miR-122a, miR-199a*, miR-489) whose expression was most decreased in metastatic cells on the basis of the combination weight of fold change in hybridization and PCR-based detection methods. Restoring the expression of miR-335, miR-126 or miR-206 in LM2 cells through retroviral transduction (Voorhoeve, Cell 124, 1169-1181; Chen, C. Z., Li, L., Lodish, H. F. & Bartel, D. P. MicroRNAs modulate hematopoietic lineage differentiation. Science 303, 83-86 (2004)) (Fig. 7) decreased the lung colonizing activity of these cells by more than fivefold (Fig. 1A). Restoration of miR-122a, miR-199a* or miR-489 expression decreased lung colonization at early time points but did not result in a significant decrease in lung colonization at the end point (Fig. 8) Histological assessment of lungs extracted at 8 weeks post cell inoculation revealed a marked decrease in the number of metastatic foci where bioluminescence imaging revealed a significant reduction. The decreased expression of these candidate miRNAs in both lung and bone metastatic cells suggested a role for these molecules in general metastatic activity of breast cancer cells. Indeed, expression of miR-335, miR-206 or miR-126 significantly decreased bone metastasis formation as assessed by bioluminescence imaging and histological analysis of hindlimbs subsequent to intracardiac injection of BoM1 cells into the arterial circulation (Fig. 1B).

Figs. 1A-D show systematic identification of miRNAs that suppress lung and bone metastasis in multiple human breast cancer cell derivatives. Fig. 1A shows Bioluminescence imaging of lung metastasis by lung metastatic breast cancer cells with restored expression of specific miRNAs. 1 ′ 104 LM2 cells expressing individual miRNAs or the control hairpin, as well as the parental MDA-MB-231 cells, were inoculated intravenously into immunodeficient mice. Lung colonization was measured by bioluminescence and quantified. n = 5; error bars represent s.e.m.; asterisk, $P < 0.05$. Fig. 1B shows 2 ′ 104 BM2 cells expressing individual miRNAs or the control hairpin were inoculated into the arterial circulation via intracardiac injection of immunodeficient mice. Bone metastasis was measured by bioluminescence and quantified as the normalized hindlimb photon flux. n=6–10; horizontal line represents median signal for each cohort; P-values based on a one-tailed rank-sum test. Fig. 1C shows 2′105 primary human cancer derivative CN34-LM1 cells expressing individual miRNAs or the control vector were inoculated intravenously into immunodeficient mice. Lung colonization was measured by bioluminescence, quantified and

normalized. n = 8; error bars represent s.e.m.; P-values based on a one-sided rank-sum test. Fig. 1D shows 2 ′ 105 primary human cancer derivative CN34-BoM1 cells expressing individual miRNAs or the control vector were inoculated into the arterial circulation via intracardiac injection of immunodeficient mice. Bone metastasis was measured by bioluminescence, quantified and normalized. n = 9–10; error bars represent s.e.m.; P-values based on a one-tailed rank-sum test.

Figs. 2A-C shows miR-126 suppresses overall tumour growth and proliferation whereas miR-335 and miR-206 regulate migration and morphology. Fig. 2A - 5 ′ 105 LM2 cells expressing individual miRNAs or the control hairpin, as well as the parental MDA-MB-231 cells, were injected into the mammary fat pads of immunodeficient mice and tumour volumes were measured over time. n = 5 (MDA-MB-231, LM2/miR-126, LM2/miR-206) and n = 10 (LM2, LM2/miR-335); error bars indicate s.e.m.; P-values based on a one-sided student's t-test at day 32. Fig.2B - 5 ′ 104 LM2 or primary breast cancer line CN34-BoM1 expressing miR-126, miR-335 or miR-206 and control cells were seeded in triplicate and viable cells were counted at 5 days after seeding. n = 3; error bars represent s.e.m.; P-values obtained using a one-sided student's t-test. Fig. 2C - 2.5 ′ 104 LM2 and CN34-BoM1 cells were transduced with the indicated miRNAs or a control hairpin, and trans-well migration was assessed. Images of cells that had migrated through trans-well inserts were obtained and analyzed in automated fashion using Metamorph software. n = 3; error bars represent s.e.m.; P-values obtained using a one-sided student's t-test.

Fig. 6 shows stem-loop quantitative RT-PCR for the mature forms of the downregulated miRNAs identified through hybridization profiling. Shown are levels (log10(ratio)) of each miRNA in the lung metastatic 4175 LM2 line (black) and bone metastatic 1833 BM2 line (brown) relative to the parental MDA-MB-231 cell line from which LM2 and BM2 cells were in vivo selected. n=3; error bars indicate s.e.m.

Fig. 7 shows quantification of miRNA expression in LM2 cells with restored expression of miRNAs. Stem-loop quantitative (q)RT-PCR for mature forms of miRNAs in LM2 cells with restored expression of miRNAs. Shown are levels (log10(ratio)) of each miRNA in the LM2/miR cell lines relative to the LM2 cell line that has lost expression of these miRs.

Fig. 8 shows bioluminescence quantification of lung colonization by additional miRNAs that are downregulated in metastatic cells. LM2 cells were transduced with a

retrovirus encoding a control hairpin or precursors of endogenous human miRNAs. 1x104 LM2 cells expressing a control hairpin or the indicated miRNAs, as well as the parental MDA-MB-231 cells were inoculated intravenously into immunodeficient mice. Lung colonization was subsequently measured by bioluminesence and quantified. Plot shows effect of restoring miR-122a, miR-199a*, or miR-489 expression. n=5; error bars represent s.e.m.

Example 2

Selective pressure for specific miRNA loss

Consistent with a selective pressure against these regulators during the metastatic process, quantitative PCR of rare metastatic foci revealed a reduction in expression of all three metastasis suppressor miRNAs in cells that had metastasized relative to the inoculated population (Fig. 9A). Restoration of miR-335 or miR-206 expression in LM2 cells also significantly reduced their metastatic dissemination from the primary mammary tumour site (Fig. 9B). miR-126 was not included in this assay because of its inhibitory effect on mammary tumour growth (see below).

To determine whether the expression of these miRNAs is lost in other human breast cancer cells with enhanced metastatic propensity, we inoculated mice with a purified population of malignant cells (CN34) obtained from the pleural fluid of a patient with metastatic breast cancer who was treated at our institution. Gupta, Nature 446, 765. We then isolated human tumour cells from metastatic lesions that formed in the lungs (CN34-LM1 cells) and bones (CN34-BoM1 cells) of the mice. Both the CN34-LM1 and CN34-BoM1 derivatives displayed a loss of miR-335, miR-126 and miR-206 expression relative to the parental population from which they were isolated (Fig. 9C). miR-335, miR-206 and miR-126 significantly reduced the ability of CN34-LM1 and CN34-BoM1 cells to metastasize to lung (Fig. 1C) and bone (Fig. 1D), respectively. The expression of these miRNAs, therefore, is lost in multiple, independently derived, breast cancer samples and their restoration suppresses lung and bone metastasis in these distinct metastatic cell populations.

Fig. 9A-C shows loss of metastasis suppressor miRNA expression from primary human cancer cells as well as exogenously introduced vector subsequent to metastatic

progression. Fig. 9A - Quantification of the loss of metastasis miRNA expression from transduced vector upon in vivo metastatic selection. 1X104 LM2 cells expressing metastasis suppressor miRNAs (miR-335, miR-206, and miR-126) were inoculated intravenously into immunodeficient mice. The lungs of mice that emitted a bioluminescence signal were extracted at 10 weeks and rare solitary nodules were identified through ex-vivo bioluminescence and extracted. These metastatic derivatives of transduced parental lines were expanded and miRNA expression was assayed through qRT-PCR. Shown are normalized expression levels of metastasis suppressor miRNA expression in miR-transduced LM2 lines and their progeny that had undergone metastatic selection. Error bars indicate s.e.m. of triplicate realtime PCR measurements. Fig.9B - metastasis suppressor miRNAs miR-335 and miR-126 suppress orthotopic metastasis from the mammary fat pad. 5x105 LM2 cells expressing control hairpin or individual miRNAs were injected into the mammary fat pads of immunodeficient mice. Orthotopic lung metastasis was quantified as bioluminescence at 32 days post inoculation. n=4-6; Error bars indicate s.e.m. p-values based on a one-sided rank-sum test. Fig. 9C - qRT-PCR for metastasis suppressor miRNA expression in a primary tumour cell population from the pleural of a patient with metastatic breast cancer, as well as its in vivo selected metastatic derivatives. Shown are normalized expression levels in the bone metastatic line CN34-BoM1 and the lung metastatic line CN34-LM1 compared to the parental CN34 sample. n=3; Error bars indicate s.e.m. p-values determined using a one-tailed student's t-test.

Example 3

Distinct mechanisms of metastasis suppression

Cell proliferation, survival and migration are among the common functions required by tumour cells for metastatic progression in target microenvironments. Of the three miRNAs that suppressed metastasis, only miR-126 significantly suppressed overall tumour growth, as assessed by tumour volume (Fig. 2A and Fig. 10). Immunohistochemistry on these mammary tumours revealed a decrease in proliferation (Fig. 11A), but not apoptosis (Fig. 11B). Consistent with this, restoration of miR-126 reduced the proliferation rate of LM2 cells as well as CN34-BoM1 cells in vitro (Fig. 2B). miR-126, therefore, suppresses tumorigenesis and metastasis, in part, through an inhibition of cancer cell proliferation.

In contrast to miR-126, restoring the expression of miR-335 or miR-206 did not alter the proliferation or apoptotic rates of LM2 cells in vivo or in vitro but it induced an altered morphology. LM2 cells and CN34-BoM1 cells expressing miR-335 or miR-206 as a population contained a significantly lower fraction of elongated cells than did control cells (Fig. 12). Despite these morphological changes, LM2 cells expressing these miRNAs continued to express the mesenchymal marker vimentin, in vivo, suggesting that the suppression of metastasis in these cells was not due to classical mesenchymal-epithelial transition. $5X10^5$ LM2 cells were transduced with individual miRNAs or the control hairpin, cells were inoculated into mammary fat pads of immunodeficient mice. The mammary fat pad tumours were subsequently resected, and sections were immunohistochemically stained in automated fashion for the mesenchymal marker vimentin. We postulated that such an alteration in shape may be associated with a decrease in cell motility, which would limit metastatic migration. Indeed, in trans-well migration assays, LM2 cells and CN34-BoM1 cells expressing miR-335 or miR-206 displayed a significant reduction in migration compared with controls (Fig. 2C). miR-335 and miR-206 also caused a significant reduction in invasive capacity, as assessed through matrigel invasion assays (Fig. 13). These findings highlight cell-autonomous mechanisms of cell proliferation, migration and invasion through which specific miRNAs may suppress metastasis.

Fig. 10 shows miR-126 suppression of primary tumour growth. Scatter-plot of individual tumour volumes at day 32 post-orthotopic mammary fat pad injection; p-values based on a one-sided student's t-test. Tumour numbers: n=10 (MDA-MB-231, LM2/miR-126, LM2/miR-206); n=20 (LM2, LM2/miR-335); p-values based on a one-sided student's t-test at day 32.

Figs. 11A-B show immunohistochemical analysis for apoptotic and proliferation markers in mammary tumours. Fig. 11A - Quantification of phospho-histone 3 through immunohistochemical staining of mammary fat pad tumours; n=6; error bars indicate s.e.m.; p-values were calculated using a one-sided rank-sum test. Fig. 11B - Quantification of apoptosis through immunohistochemical staining of mammary fat pad tumours for cleaved caspase-3; n=6; error bars indicate s.e.m.; p-values were calculated using a one-sided rank-sum test.

Fig. 12 shows restoration of miR-335 and miR-206 expression in metastatic cells reduces the fraction of elongated cells. Morphometric quantification of cell elongation. Indicated cell lines were plated on glass chamber slides and phase-contrast images were obtained at 20X magnification. Morphometric analysis on >1600 cells was performed using Metamorph software. Cell elongation (the aspect ratio) was calculated as the ratio of the maximum length:maximum width of each cell. The fraction of cells with an aspect ratio greater than 3:2 is plotted for each MDA-MB-231 derivative cell line. The CN-BoM1 and its derivatives were in general more elongated than the MDA-MB-231 derivatives, therefore cells with a length:width ratio greater than 5:2 were considered elongated and are plotted as a fraction of the total cells for each CNBoM1 line. n=195-300 cells for MDA-MB-231 derivative lines and 171-174 cells for CN34 lines; Error bars represent s.e.m. p-values obtained using a one-sided rank sum test comparing fraction of cells per field.

Fig. 13 shows miR-335 and miR-206 suppress invasion by lung metastatic cells. 5.0X104 LM2 cells were transduced with a control vector short hairpin, or vectors encoding miR-335 or miR206, and invasion through a matrigel coated trans-well membrane insert after 22 hours was quantified. Images of cells that had migrated through trans-well inserts were obtained and cell counts were quantified. n=6; p-values based on a one-sided student's t-test.

Example 4

MicroRNA expression in clinical metastasis

To determine whether these miRNAs are associated with human metastasis, we determined by qRT-PCR the expression levels of miR-335 and miR-126 in 20 archived primary breast tumours.

The histology, staging, clinical characteristics and results of quantitative miRNA expression analysis (using qRT-PCR) for those 20 primary breast tumours are summarized in the table below.

Table 1

Met event	Tumour #	Size	LN Pos/All	ER	PR	HER2	MFS(YR)
1	1	1.3	2/22	P	P	N	5.39
1	2	3.8	3/19	N	N	3+	1.84
1	3	1.5	3/22	P	P	N	2.57
1	4	4	1/18	N	N	N	0.67
1	5	6	2/32	N	N	N	1.38
1	6	2.5	5/10	P	P		1.30
1	7	2.6	1/29	P	P	N	6.54
0	8	2.5	1/21	P	P	N	7.28
0	9	4	1/20	N	P	N	6.98
1	10	3.9	24/28	P	P	N	4.59
0	11	2.3	3/22	P	P	2+	6.50
0	12	3.5	2/26	N	N	3+	10.74
0	13	2.2	3/37	P	P	2+	7.11
0	14	3	3/30	N	N	N	8.74
0	15	2.9	0/16	P	P	N	7.63
1	16	3	0/2	N	N	N	0.77
0	17	2.1	0/19	P	P	N	7.08
0	18	3	0/19	P	P	N	9.15
1	19	4.5	0/26	N	N	1+	3.47
1	20	3.5	20/21	P	P	N	1.30

These tumours comprised large (>2.5 cm) and small (<2.5 cm) oestrogen receptor (ER)-positive and -negative subtypes that were surgically resected before administration of chemotherapy. This set consisted of primary tumours resected from 11 patients that ultimately relapsed to lung, bone, or brain as well as tumours resected from nine patients who did not suffer metastatic relapse (Table 1). qRT-PCR (Fig. 14) revealed that patients whose primary tumours displayed low expression of miR-335, miR-126 (Fig. 3A) or miR-206 (Fig. 15) had a shorter median time to metastatic relapse. Notably, the low expression levels of miR-335 ($P = 0.0022$; median survival of 1.84 yr; hazard ratio (ratio of predicted hazard (metastasis) for member of control group to that for member of test group) of 8.95) or miR-126 ($P = 0.0156$; median survival of 2.57 yr; hazard ratio of 5.08) were associated with very poor overall metastasis-free survival compared to the group whose tumours expressed a high level of these miRNAs. The median expression values of these miRNAs was more than eightfold lower in tumours from patients that ultimately relapsed compared to tumours from patients that did not relapse (Fig. 14). The expression levels of these miRNAs were not significantly correlated with ER status or HER-2 (also called ERBB2) amplification status (Table 1). The group of patients whose primary tumours had a low level of miR-122a and miR-199a expression, two miRNAs that do not suppress metastasis, showed no difference in

metastasis-free survival compared to those with high expression (Fig. 3A and Fig. 14). These findings uncover a significant association between the loss of metastasis suppressor miRNA expression in primary human breast tumours and the likelihood of future distal metastatic recurrence.

Fig. 14 shows quantification of miRNA levels in human breast tumour samples. Each vertical bar represents expression levels of metastasis suppressor miRNAs miR-335, miR-206, and miR-126 in addition to miR-122a and miR-199a. The expression level of a given miRNA in each sample was normalized to the median expression value for that miRNA across the 20 tumour samples. The Y-axis is represented in log2 scale. Vertical black bars correspond to tumours extracted from patients that subsequently suffered metastatic relapse.

Fig. 15 shows the association between the loss of miR-206 expression in primary breast tumours and metastatic relapse. a, miR-206 expression was assessed in a set of twenty primary breast tumour specimens through qRT-PCR. Kaplan-Meier curve depict metastasis-free survival of patients whose primary tumours contained low or high levels of the miR-206. n=20; p-value obtained using a log-rank test.

Example 5

miR-335 regulates a set of metastasis genes

Given the strong association between the loss of miR-335 expression and clinical relapse, we wondered whether miR-335 loss could promote metastasis. To test this, we transfected the poorly metastatic MDA-MB-231 cells with an anti-miRNA antagomir (Meister, G., Landthaler, M., Dorsett, Y. & Tuschl, T. Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. RNA 10, 544-550 (2004)) targeting either miR-335, miR-199a* or a control sequence. Indeed, inhibition of miR-335 enhanced the lung-colonizing ability of MDA-MB-231 cells compared with control cells (Fig. 3B). To identify putative metastasis genes that miR-335 suppresses, we transcriptionally profiled LM2 cells with restored miR-335 expression and arrived at the set of 756 genes for which expression is decreased (using a permissive threshold; see Methods) compared with control LM2 cells. We separately identified genes for which expression levels were increased across both bone and lung metastatic cells. On the basis of available gene expression data sets (Kang, Y. et al. A multigenic program mediating breast cancer metastasis to bone. Cancer

Cell 3, 537-549 (2003); Minn, A. J. et al. Genes that mediate breast cancer metastasis to lung. Nature 436, 518-524 (2005)) we identified 116 genes for which expression was increased by at least twofold in both bone and lung metastatic MDA-MB-231 derivatives compared with the parental line (Table 2). The overlap between these two lists yielded a set of six genes for which expression is high in metastatic cells and suppressed by miR-335 (Fig. 4A). This list of miR-335-regulated metastasis genes included genes previously implicated in extracellular matrix and cytoskeleton control, such as the type 1 collagen COL1A1 (Egeblad, M. et al. Type I collagen is a genetic modifier of matrix metalloproteinase 2 in murine skeletal development. Dev. Dyn. 236, 1683-1693 (2007)); in signal transduction, such as the receptor-type tyrosine protein phosphatase PTPRN2 (Varadi, A., Tsuboi, T. & Rutter, G. A. Myosin Va transports dense core secretory vesicles in pancreatic MIN6 beta-cells. Mol. Biol. Cell 16, 2670-2680 (2005)), the c-Mer tyrosine kinase (MERTK) (Graham, D. K. et al. Cloning and developmental expression analysis of the murine c-mer tyrosine kinase. Oncogene 10, 2349-2359 (1995)) and the phospholipase PLCB1 (Lyu, M. S., Park, D. J., Rhee, S. G. & Kozak, C. A. Genetic mapping of the human and mouse phospholipase C genes. Mamm. Genome 7, 501-504 (1996)); as well as in cell migration, such as tenascin C (TNC) (Ilunga, K. et al. Co-stimulation of human breast cancer cells with transforming growth factor- and tenascin-C enhances matrix metalloproteinase-9 expression and cancer cell invasion. Int. J. Exp. Pathol. 85, 373-379 (2004) and the SRY-box containing transcription factor SOX4 (van de Wetering, M., Oosterwegel, M., van Norren, K. & Clevers, H. Sox-4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. EMBO J. 12, 3847-3854 (1993); Hoser, M. et al. Prolonged glial expression of Sox4 in the CNS leads to architectural cerebellar defects and ataxia. J. Neurosci. 27, 5495-5505 (2007)).

Bioinformatic analysis of the 3'-UTRs of these genes revealed them all to have at least six nucleotides of sequence complementarity to the miR-335 seed region (data not shown). To determine whether the altered expression of these genes in metastatic cells is, in part, mediated through their 3'-UTRs, we cloned the 3'-UTR of five of these genes downstream of a luciferase gene as a reporter, and assayed their expression in LM2 cells (low miR-335) and MDA-MB-231 cells (high miR-335). The expression from UTR reporters corresponding to SOX4, PTPRN2, TNC and MERTK, but not that of the control gene UBE2F lacking the miR-335 target sequence, was significantly lower in LM2 cells relative to MDA-MB-231 cells (Fig. 4B). Furthermore, downregulation of miR-335 in MDA-MB-231 cells by means of an antagomir was sufficient to enhance the expression of SOX4, PTPRN2

and MERTK, but not the expression of a control UBE2F reporter (Fig. 4C). These results suggest that SOX4, PTPRN2, MERTK and possibly TNC are direct targets of endogenous miR-335.

SOX transcription factors are known to regulate progenitor cell development and migration. Hoser, J. Neurosci. 27, 5495-5505. In reporter assays, mutation of the miR-335 seed sequence in the SOX4 UTR as well as miR-335 inhibition with an antagomir significantly increased reporter expression (Fig. 4D). Restoration of miR-335 expression in LM2 cells reduced endogenous SOX4 messenger RNA expression, whereas miR-335 inhibition increased SOX4 expression in MDA-MB-231 cells (Fig. 4E). The knockdown of SOX4 in LM2 cells with either of two unique short hairpin RNAs (shRNAs) reduced the overall fraction of elongated cells, similar to the phenotype observed with miR-335 restoration in these cells (Fig. 16A). The morphological change resulting from SOX4 knockdown was also associated with a decrease in cell migration in a trans-well assay (Fig. 16B). Another miR-335-regulated gene, TNC, also caught our attention, as its expression occurs in the invasive front of human carcinomas. Orend, G. & Chiquet-Ehrismann, R. Tenascin-C induced signaling in cancer. Cancer Lett. 244, 143-163 (2006). Knockdown of TNC in LM2 cells using either of two RNA interference molecules also reduced migration in a trans-well assay (Fig. 16B). Notably, knockdown of SOX4 or TNC significantly diminished the invasive ability of LM2 cells (Fig. 5A). Moreover, short-hairpin inhibition of SOX4 or TNC significantly abolished metastasis by LM2 cells in lung colonization assays (Fig. 5B). Thus, miR-335 regulates metastasis through suppression of transcription factor SOX4 and extracellular matrix component tenascin C.

To determine whether the expression of the genes that miR-335 regulates is associated with human breast cancer metastasis, we examined primary breast tumour gene expression data sets with corresponding disease outcome annotation. Minn, Nature 436, 518-524; Wang, Y. et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 365, 671-679 (2005). Tumours in these cohorts for which aggregate six-gene expression scores exceeded one standard deviation from the mean were considered miR-335-signature-positive (equivalent to loss of miR-335 function). Patients whose primary breast tumours were positive for the miR-335 signature had a significantly worse metastasis-free survival both in the combined cohort of 368 patients (Fig. 5C) as well as in each cohort separately classified (Fig. 17A). Consistent with the results of

our functional studies, the miR-335 signature performed better as a predictor of overall metastasis rather than lung- or bone-specific metastasis (Fig. 17B).

Figs. 3A-B shows clinical association of miR-335 and miR-126 with metastasis-free survival. Fig. 3A - miR-335, miR-126, miR-122a and miR-199a expression was assessed in a set of 20 primary breast tumour samples through qRT–PCR. Kaplan–Meier curves depict metastasis-free survival of patients whose primary tumours contained low or high levels of the indicated miRNAs. P-values were obtained using a log-rank test. Fig. 3B - The parental MDA-MB-231 cells were transfected with antagomirs targeting endogenous miR-335, miR-199a* or a control antagomir. Four days after transfection, 1 ´ 104 LM2 cells from each cohort were inoculated intravenously into immunodeficient mice. Lung colonization was measured by bioluminescence at day 35 and quantified. n = 5; error bars represent s.e.m.; P-values based on a one-sided rank-sum test.

Figs. 4A-E show a miR-335-regulated gene set includes SOX4 as a miR-335 direct target. Fig. 4A - The miR-335 metastasis signature consists of genes downregulated by miR-335 and are also overexpressed in bone as well as lung metastatic MDA-MB-231 derivatives. The heatmap depicts the variance-normalized expression values for each gene averaged across two MDA-MB-231 samples, four LM2 samples and two LM2/miR-335 samples. The scale bar depicts standard deviation change from the mean for the expression value of each gene. Fig. 4B - UTR reporter assays of miR-335 metastasis genes in LM2 and MDA-MB-231 cells. Reporter constructs consisting of the luciferase sequence fused to the 3? UTRs of the miR-335 metastasis genes as well as the control gene UBE2F were transfected into the LM2 and parental MDA-MB-231 cell lines. Luciferase activity of cells was assayed at 32 h after transfection and the values were normalized to the LM2 cell line. n = 3; error bars represent s.e.m.; P-values were obtained using a one-sided student's t-test. Fig. 4C - Luciferase activity of MDA-MB-231 cells co-transfected with 3? UTR reporter constructs with, or without, miR-335 antagomir was assayed at 32 h after transfection and normalized to control cells. n = 3; error bars represent s.e.m.; P-values obtained using a one-sided student's t-test. Fig. 4D - Schematic diagram depicts seed sequence in SOX4 UTR. The 60 bp containing this sequence or the miR-335 seed target site mutant were subjected to UTR reporter assays in LM2 cells with restored miR-335 expression. Luciferase activity of inecells was also assayed in the presence or absence of the miR-335 antagomir. n = 3; error bars represent s.e.m.; P-values

obtained using a one-sided student's t-test. Fig. 4E - SOX4 expression in LM2 and MDA-MB-231 cells was obtained through real time qRT–PCR in cells expressing miR-335 or transfected with the miR-335 antagomir. n = 3; error bars represent s.e.m.; P-values derived using a one-sided student's t-test.

Figs. 5A-C show miR-335 regulates metastasis and invasion through suppression of SOX4 and TNC. Fig. 5A - 5.0 ´ 104 LM2 cells were transduced with short hairpin control vector, either of two shRNAs targeting SOX4, a hairpin targeting TNC, or an siRNA targeting TNC, and invasion of cells through a trans-well membrane insert was quantified. n = 6; error bars represent s.e.m.; P-values based on a one-sided student's t-test. Fig. 5B - 2 ´ 105 LM2 cells transduced with either a control short hairpin, a short hairpin targeting SOX4 or one targeting TNC were inoculated intravenously into immunodeficient mice. Lung colonization was measured by bioluminescence and quantified. n = 7; error bars represent s.e.m.; P-values based on a one-tailed rank-sum test. Fig. 5D - Kaplan–Meier curves for the combined Memorial Sloan Kettering and Erasmus Medical Center breast tumour cohorts (368 tumours) depicting metastasis-free survival of patients whose primary tumours expressed the miR-335 six-gene signature (positive) and those that did not (negative). n = 368; P-value based on the Mantel–Cox log-rank test.

Figs. 16A-B show the miR-335 regulated genes SOX4 and TNC alter the morphology of lung metastatic cells and suppress migration. Fig. 16A - Cell elongation (the aspect ratio) for the indicated cell lines was calculated as the ratio of the maximum length:maximum width of each cell obtained through morphometric analysis of phase-contrast images obtained using Metamorph software at 20X magnification. The fraction of cells with an aspect ratio greater than 3:2 are plotted for indicated cell lines. n=159-227 cells; Error bars represent s.e.m. p-values obtained using a one-sided rank sum test comparing fraction of cells per field. Fig. 16B - 2.5X104 LM2 cells expressing either of two short-hairpin RNAs targeting SOX4, a hairpin targeting TNC, or an siRNA targeting TNC, or a short hairpin control vector, were seeded onto trans-well inserts and migration through a trans-well membrane insert was quantified. Images of cells that had migrated through the inserts were obtained using Metamorph software and cells were counted in automated fashion using Metamorph software. n=6; p-values based on a one-sided student's t-test.

Figs. 17A-B show the miR-335 regulated six-gene signature predicts overall metastasis-free survival in the MSK-82 and EMC-286 datasets. Fig. 17A - The MSK-82

cohort, which consists of 82 patients with primary breast cancer 18, and the EMC-286 cohort, consisting of 286 patients with primary breast cancer 27, were classified into two groups based on the expression of the miR-335 regulated signature in primary tumours. Kaplan-Meier curves representing metastasis-free survival of patients whose original tumours expressed the miR-335 signature (positive) and those whose original tumours did not (negative). n=82 for MSK cohort and n=286 for EMC cohort. p-values based on a Mantel-Cox Log-rank test. Fig. 17B - Kaplan-Meier curves representing lung-metastasis-free survival and bone-metastasis-free survival of MSK-82 patients whose original tumours expressed the miR-335 signature. p-values based on a Mantel-Cox Log-rank test.

Example 6

MicroRNAs as suppressors of metastasis

Multiple lines of evidence provided here argue for the involvement of specific miRNAs in suppressing breast cancer metastasis. miR-335, miR-206 and miR-126 are selectively downregulated across a number of highly metastatic human cell lines compared to the general tumour cell population and have demonstrated abilities to suppress metastasis of breast cancer cells to different organ sites. The expression of miR-335 and miR-126 in human mammary tumours is inversely associated with metastatic relapse of these tumours to distant organs, and the expression of miR-206 also shows a trend in the same direction. The expression of a set of genes regulated by one of these miRNAs, miR-335, is directly associated with relapse. We also establish two of these genes, SOX4 and TNC, as in vivo mediators of metastasis. The role of SOX4 in haematopoietic progenitor development (Schilham, M. W., Moerer, P., Cumano, A. & Clevers, H. C. Sox-4 facilitates thymocyte differentiation. Eur. J. Immunol. 27, 1292-1295 (1997)) suggests that its transcriptional programme may be re-used not only for cancer cell invasion in cooperation with tenascin C but also for tumour initiation in the metastatic niche. miR-335 and miR-126 are expressed in normal human breast tissue. Lu, Nature 435, 834-838. Our findings on their roles in the pathogenesis of human breast cancer argue for an important function for these regulators in maintaining normal tissue integrity. Recently, miR-10b was identified as a miRNA whose overexpression in breast cancer cells promoted tumour growth and lung micrometastasis. Ma, Nature 449, 682-688. Our work expands on this by identifying miRNAs as clinically

meaningful suppressors of metastasis. MicroRNAs are thus uncovered as another class of molecules, along with metastasis suppressor genes (Steeg, P. S. Metastasis suppressors alter the signal transduction of cancer cells. Nature Rev. Cancer 3, 55-63 (2003)), that negatively regulate tumour progression.

The recent implication of miRNAs and miRNA processing in tumorigenesis (Kumar, Nature Genet. 39, 673-677: Lu, Nature 435, 834-838) has raised interest in identifying the specific miRNAs for which loss of expression enhances tumorigenesis, the mechanisms by which they act, and the phenotypic advantages afforded to cells that lose miRNA expression. Our findings suggest that within a tumour, the loss of specific miRNAs provides a selective advantage for cells destined for metastatic colonization. A global downregulation of miRNAs in cancer may serve metastasis by reducing the threshold for loss of specific miRNAs in a subset of cells that will ultimately metastasize. That one such miRNA (miR-335) regulates a set of putative, and a subset of validated, metastasis genes argues that the multi-gene regulatory capacity of a miRNA can function as a barrier to tumour progression in humans. The strong association of the loss of miR-335 and miR-126 expression with metastatic relapse suggests the potential for the use of these molecules in prognostic stratification of breast cancer patients in addition to conventional clinical and pathological staging markers.

Gene expression analysis was performed of the LM2 cells and bone-metastatic BoM2 cells, using as reference the parental, non-metastatic, MDA-MB-231 line. The gene-list below shows genes with more than two-fold increase expression in the metastatic cell lines compared to the parental, non-metastatic cell line, and was obtained through an analysis of available gene-expression datasets for MDA-MB-231 derivative lung metastatic cell lines 4175, 4180, 4142, 4173, and bone metastatic cell lines 2287 and 1833. Kang, Cancer Cell 3, 537-549; Minn, Nature 436, 518-524.

Table 2

Gene Symbol	Affy ID	fold increase LM2(s)	fold increase BoM2(s)
MMP1	204475_at	13.90871447	28.31526445
CSF3	207442_at	47.10144928	6.632850242
TMEM16A	218804_at	22.70212766	12.63829787
PTPRN2	203029_s_at	105.5351351	2.313513514

SERPINB2	204614 at	21.69727891	8.034013605
SSX1	206627 s at	5.918918919	27.45945946
DZIP1	204557 s at	18.38709677	8.64516129
MAGEA3	209942 x at	3.831185567	34.87628866
COL1A1	202311 s at	13.76136364	8.113636364
KLRC2	206785 s at	27.23579545	3.823863636
IL13RA2	206172 at	34.83654559	2.979243884
SPANXC	220217 x at	7.365867472	11.10760808
FOXA2	210103 s at	6.550373134	11.93843284
UTY	211149 at	10.69444444	6.944444444
PLP1	210198 s at	6.324561404	10.48245614
NLGN4Y	207703 at	6.458677686	10
GARNL4	213280 at	11.77433628	5.247787611
SPANXB2	220922 s at	7.773363967	7.905284221
KYNU	217388 s at	12.15726902	4.788722826
SPANXB2	220921 at	9.805483947	5.511835013
DSCR3	220374 at	5.279005525	9.320441989
KISS1	205563 at	3.947058824	12.35294118
CDC14A	210743 s at	6.162393162	6.598290598
IL12RB1	216459 x at	5.811403509	6.964912281
USP9Y	206624 at	5.24015748	7.322834646
CXCL12	209687 at	4.176470588	8.823529412
PDGFRA	203131 at	4.704545455	7.772727273
LAMA4	202202 s at	12.7341629	2.830316742
MTMR7	215793 at	3.875	8.6125
LOC401168	217698 at	6.618421053	5
SIAT1	214970 s at	4.929054054	6.547297297
FLJ23259	220622 at	6.129411765	5.182352941
KRTHB1	213711 at	8.828861708	3.559873742
LOC91120	216468 s at	4.888888889	6.238095238
MAGED4	221261 x at	11.67437722	2.535587189
C14orf143	210525 x at	5	5.855421687
CRP	205753 at	4.527237354	6.455252918
HPGD	203913 s at	7.647727273	3.818181818
SLC24A5	222153 at	10.65853659	2.731707317
MYOM1	205610 at	5.753846154	4.923076923
AMPH	205257 s at	2.714536928	10.33763189
VCL	200930 s at	5.631178707	4.828897338
FCHO1	213669 at	3.661691542	7.383084577
KIAA1181	220728 at	5.747058824	4.647058824
UBCE7IP5	215615 x at	4.01843318	6.580645161
MTAP	204956 at	5.183098592	4.957746479
CA8	220234 at	3.83255814	6.697674419
EAF2	219551 at	2.649659864	9.659863946
KIAA1411	221233 s at	5.5	4.65060241
LOC51212	220697 at	2.835443038	8.708860759
GNRH1	207987 s at	5.0375	4.9
BTEB1	203543 s at	6.102167183	4.037151703
F9	207218 at	7.502427184	3.257281553

ABCB9	207321 s at	5.7375	4.173684211
KIAA0087	207161 at	6.185185185	3.87037037
CNTNAP2	219301 s at	4.634920635	5.142857143
ARL1	214932 at	4.243093923	5.508287293
TTC9	213172 at	5.197368421	4.388157895
C6orf148	220513 at	3.810185185	5.814814815
SOX4	201418 s at	6.133084577	3.611940299
GATM	203178 at	5.079710145	4.275362319
HIP14	217486 s at	3.083769634	7.031413613
PTGS2	204748 at	6.082887701	3.559714795
LOC401284	214747 at	4.003968254	5.317460317
KLHL11	220657 at	4.651877133	4.505119454
ROBO1	213194 at	4.323113208	4.657429245
EREG	205767 at	8.124174917	2.437706271
CHIA	220630 s at	4.475504323	4.368876081
PLN	204939 s at	4.444444444	4.364197531
FLJ10178	219355 at	3.400414938	5.663900415
AKAP9	207870 at	5.046255507	3.797356828
ARFD1	210994 x at	3.632272228	5.254665203
SLC4A7	207603 at	3.088607595	6.056962025
ABCG5	220383 at	4.205521472	4.349693252
PLCB1	213222 at	8.325757576	2.19047619
DPYSL4	205493 s at	4.18	4.36
CYP3A43	211442 x at	2.623809524	6.885714286
KYNU	210663 s at	5.57995283	3.230660377
P2RX2	221372 s at	3.489361702	5.154255319
FLJ13841	219995 s at	2.3984375	7.453125
FLJ23514	220389 at	5.139705882	3.463235294
HOXA7	206847 s at	3.683035714	4.75
SULT1C1	211470 s at	4.147331787	4.187935035
FCN2	207804 s at	4.97254902	3.478431373
SIGLEC8	208253 at	5.413265306	3.193877551
TM4SF12	219274 at	4.069333333	4.242666667
C21orf5	215163 at	3.182741117	5.360406091
PARK2	207058 s at	4.272093023	3.990697674
KCNJ15	210119 at	3.229166667	5.229166667
LOC128387	219857 at	3.411392405	4.911392405
CD24	266 s at	4.420529801	3.781456954
XYLB	214776 x at	6.75	2.458333333
ZNF257	210697 at	2.85	5.791666667
GPR17	206190 at	2.458874459	6.70995671
ASMT	210551 s at	3.416666667	4.767676768
RGS2	202388 at	4.888722555	3.331337325
SLC9A7	214863 at	3.406896552	4.751724138
STX11	210190 at	4.521929825	3.570175439
ALK	208211 s at	4.495738636	3.565340909
TNC	201645 at	6.720522388	2.335074627
HAPLN1	205523 at	5.794871795	2.692307692
KIAA0033	207481 at	4.584615385	3.338461538

FLJ20668	219574 at	3.237179487	4.71474359
HIST1H2AG	207156 at	6.654320988	2.271604938
C11orf1	216189 at	7.429347826	2.02173913
TP53AP1	210241 s at	4.067901235	3.617283951
FLJ12650	219438 at	3.093049327	4.730941704
MDM2	211832 s at	3.170658683	4.556886228
CD24	209772 s at	2.08496732	6.901960784
SGKL	207479 at	4.679487179	3.051282051
IFNA8	207932 at	3.092307692	4.615384615
DKFZp761O01 13	207797_s_at	3.167844523	4.441696113
GDF11	216860 s at	3.963338534	3.514820593
IGFBP5	211959 at	2.979591837	4.673469388
CNTNAP2	219300 s at	3.357407407	4.125925926
SSX2	216471 x at	2.084398977	6.615089514
MERTK	211912 at	2.549618321	5.305343511
IL28RA	221071 at	3.037974684	4.417721519
RARA	211605 s at	3.676470588	3.612456747
RPL34	217155 at	4.299270073	3.058394161
LOC401581	220859 at	4.069078947	3.157894737
DJ462O23.2	214579 at	2.956470588	4.339607843
MEOX2	206201 s at	2.094594595	6.108108108
FRS2	221308 at	4.036231884	3.166666667
DLG2	206253 at	3.682291667	3.458333333
THAP9	220360 at	3.52238806	3.582089552

Example 7

Identification of genes that suppress metastatic phenotype and behaviour.

Of the six genes that we have found to be downregulated by miR-335 we studied two further, SOX4 and TNC. As Fig. 16A-B show, RNAi of SOX 4 via a SOX 4 specific shRNA, or indirectly through overexpression of miR-335, induced a phenotype of more elongated LM2 cells compared to the LM2 cells that were used as control, indicating that SOX4 constitutes a target gene for the treatment of breast cancer and prevention or reduction of metastatic potential of that cancer. Likewise, RNAi against SOX4 and TNC was performed using either specific shRNA or specific siRNA, Fig 16B shows that reduction of SOX4 or TNC via RNAi significantly reduced the ability of the LM2 cells to migrate in a trans-well migration assay, which is a proxy for metastatic ability. These results confirm SOX4 as a target to limit or eliminate metastatic potential of a cancer cell and further indicate that targeting of TNC, with RNAi for example, is also useful in eliminating or controlling

metastatic behaviour in cancer cells. Inhibition of SOX4 or /and TNC via RNAi or via small moleculer or through any other inhibitor is a therapy useful in the treatment of cancer, and specifically in the treatment of breast cancer.

Methods

The MDA-MB-231 cell line and its metastatic derivatives (Kang, Cancer Cell 3, 537-549; Minn, Nature 436, 518-524; Minn, A. J. et al. Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. J. Clin. Invest. 115, 44-55 (2005)) as well as the pleural effusion-derived CN34 cancer cells have been described previously. Gomis, R. R., Alarcon, C., Nadal, C., Van Poznak, C. & Massague, J. C/EBP at the core of the TGF cytostatic response and its evasion in metastatic breast cancer cells. Cancer Cell 10, 203-214 (2006). All metastatic derivatives were obtained through in vivo selection in immunodeficient mice. Kang, Cancer Cell 3, 537-549; Minn, Nature 436, 518-524. MicroRNA microarray profiling was performed using LC Sciences technology (LC Sciences, LLC). Mature miRNA expression was assayed using Taqman MicroRNA assays as previously described. Chen, C. et al. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. 33, e179 (2005). MicroRNA expression was achieved through retroviral transduction of metastatic cells using miR-Vec technology. Chen, Science 303, 83-86. The miRNA vectors were provided by R. Agami. Voorhoeve, Cell 124, 1169-1181. In brief, they consist of approximately 500-bp fragments that span a given miRNA genomic region under the control of a CMV promoter. Voorhoeve, Cell 124, 1169-1181; Chen, Science 303, 83-86. All animal work was done in accordance with the MSKCC Institutional Animal Care and Use Committee. Statistical analyses were performed using the Graphpad Prism 5 software package.

Cell culture. The MDA-MB-231 line and its metastatic derivatives have been described previously. Minn, Nature 436, 518-524; Minn, J. Clin. Invest. 115, 44-55 (2005). All cell lines were cultured in DMEM supplemented with 10% FBS, glutamine, pyruvate, penicillin, streptomycin, and fungizone. CN34 cancer cells were isolated from the malignant pleural effusion of a patient with metastatic breast cancer treated at our institution upon obtaining written consent following IRB regulations as previously described. Gomis, Cancer Cell 10, 203-214 (2006). The GPG29 retroviral packaging cell line was maintained in DMEM containing 10% FBS supplemented with G418, puromycin, penicillin, streptomycin,

doxycycline, and fungizone, as previously described. Minn, Nature 436, 518-524. All transfections were done using Lipofectamine 2000 (Invitrogen). Subsequent to transfection, GPG29 cells were maintained in DMEM supplemented with 10% FBS and pyruvate.

Analysis of miRNA and mRNA expression. Total RNA from MDA-MB-231 cells, its derivatives, or CN34 cells and its derivatives was extracted and purified using the MiRvana kit (Ambion). Tissues from primary breast cancers were obtained from therapeutic procedures performed as part of routine clinical management at our institution. Methods for primary tumour sample processing were described previously. Minn, Nature 436, 518-524. In brief, RNA from these samples was extracted from frozen tissues by homogenization in TRIzol reagent (Gibco/BRL), as described previously. Minn, Nature 436, 518-524. Taqman MicroRNA assays (Applied Biosystems) were used to quantify mature miRNA expression as previously described. Chen, Nucleic Acids Res. 33, e179 (2005). RNU6 (Applied Biosystems) was used as endogenous control for miRNA expression studies. For cell line miRNA expression quantification, each reverse-transcriptase (RT) reaction consisted of 50 ng of purified total RNA, 1X RT buffer, dNTPs (each at 0.375 mM), 5 Uμl-1 MultiScribe reverse transcriptase, 50nM stem-loop RT primer, and 0.38 Uμl-1 RNase inhibitor (Applied Biosystems). RT reactions were incubated at 16 °C for 30 min, 42 °C for 30 min, and 85 °C for 5 min. Real-time PCR reactions for miRNAs from cell lines were performed in quadruplicate in 10μl volumes. The Real-time reaction mix consisted of 0.67 μl of RT product, 0.5 μl of 20X TaqMan microRNA assay mix, 5 μl TaqMan 2X universal PCR Master Mix. Quantitative miRNA expression data were acquired and analyzed using an ABI Prism 7900HT Sequence Detection System (Applied Biosystems).

For primary tumour miRNA expression quantification, each RT reaction contained 10 ng of purified total RNA. RT-PCR reactions from archived tumour RNA were performed in replicates of six. For mRNA expression quantification, four-hundred nanograms of total purified RNA was reverse transcribed using the SuperScript III first-strand synthesis system (Invitrogen). CDNA corresponding to approximately 5 ng of initial RNA was used for each quadruplicate quantitative PCR reaction. Human SOX4 and ß-glucuronidase (as endogenous control) were amplified using commercially designed Taqman gene expression assays (Applied Biosystems) and the Taqman Universal PCR master mix (Applied Biosystems). Quantitative mRNA Expression data were acquired and analyzed in 384 well plate format using an ABI Prism 7900HT Sequence Detection System (Applied Biosystems).

miRNA microarray hybridization. 20 μg total RNA obtained from MDA-MB-231 and its derivatives was labeled and hybridized on miRNA microarrays by LC sciences (Houston, TX). The arrays were designed to detect miRNA transcripts corresponding to 453 miRNAs contained in the Sanger miRBase Release 8.2. Of all the probes assayed, those corresponding to 179 miRNAs revealed a signal above a threshold above background in at least one of the MDA-MB-231 cell lines. Raw signal intensities representing hybridization to probes were mean-normalized across across cell lines. Hierarchical clustering of variance-normalized expression values was performed using Matlab software. Hierarchical clustering was performed using euclidean distance as a metric and utilizing average linkage. Clustering of cell types along organotropic and metastatic phenotypes was achieved using both the expression of 179 miRNAs as well as the 20 miRNAs that displayed the highest coefficients of variation across cell lines.

RNA microarray hybridization and miR-335 signature. The gene expression datasets from MDA-MB-231 cell derivatives as well as the primary breast tumour gene expression datasets from the MSK and EMC cohorts have been published previously. Minn, Nature 436, 518-524; Wang, Lancet 365, 671-679 (2005). Transcriptional profiling of LM2 cells expressing miR-335 was performed as previously described (Minn, Nature 436, 518-524) through the use of the Affymetrix HG-U133 plus 2.0 array by the genomics core laboratory at MSKCC. 874 probes corresponding to 756 genes displayed a greater than 0.2 log10 (ratio) fold decrease in expression relative to LM2 cells expressing a short hairpin control. This low permissive threshold was chosen so as to include genes whose fold reduction at the transcript level as detected by hybridization may be much less than their potential reduction at the protein level. The number of transcripts downregulated at the permissive threshold as well as the number of transcripts downregulated at the less-permissive threshold of two-fold (458 probes) is comparable to profiling analyses with other miRNAs. He, L. et al. A microRNA component of the p53 tumour suppressor network. Nature 447, 1130-1134 (2007). The miR-335 regulated 6-gene signature consists of the set of genes corresponding to the 756 miR-335 downregulated genes that 1) were also upregulated by more than two-fold in both lung and bone metastatic MDA-MB-231 derivatives relative to the parental line (available gene expression data-set Minn, Nature 436, 518-524) and 2) were expressed by multiple MDA-MB-231 cell lines.

Generation of retrovirus, miRNA expressing cells, and knockdown cells. The plasmid vectors encoding the miRNAs were kindly provided by R. Agami (Voorhoeve, Cell 124, 1169-1181), and have been previously described. To produce retrovirus for restoration of miRNA expression in cells, these PMSCV-based vectors were individually transfected into the GPG29 amphotropic packaging cell line. Virus particles were collected at 48 and 72 h after transfection, filtered, and concentrated through ultracentrifugation, as previously described. Gupta, Nature 446, 765. These concentrated retroviral particles were then used to transduce cells in the presence of 8 μg ml-1 polybrene. Vector-transduced cells were selected with blasticidin. miRNA expression was confirmed in all cases using quantitative RT-PCR for the mature forms of individual miRNAs, suggesting that the precursor transcripts were endogenously processed to the mature forms. Importantly, miRNA expression did not interfere with endogenous processing of miRNAs as revealed by qRT-PCR for the expression of other miRNAs in these cell lines. For in vitro miRNA inhibition studies, MDA-MB-231 or LM2 cells were transfected with antagomirs (Meister, RNA 10, 544-550) (25 nM; Ambion) based on manufacturer's recommendations. For in vivo miR-inhibition metastasis studies, MDA-MB-231 cells were transiently transfected with the control anti-miR Locked Nucleic Acid (LNA)-159, LNA-335, or LNA-199a* (20nM Exiqon) based on manufacturers recommendations. Transfected cells were inoculated into immunodeficient mice via tail-vein injection four days post transfection. SOX4 knockdown cells were generated through infection with Mission TRC lentiviral particles (Sigma-Aldrich) containing short hairpins targeting the human SOX4 gene. The TRC numbers and the shRNA sequences that yielded the greatest knockdown were TRCN0000018216: Seq. ID No. 8 and TRCN0000018217: Seq. ID No. 9. These sequences contain complementary portions set forth in seq. ID Nos. 19 and 20, respectively. TNC knockdown cells were generated either by infection with a pRetrosuper vector expressing the forward sequence Seq. ID No. 4 or by transfecting cells with the following ON-TARGET siRNA with sense sequence: Seq. ID No. 6 (Dharmacon).

UTR reporter assays. The full-length 3' UTRs of the miR-335 regulated human genes COL1A1, SOX4, PTPRN2, TNC, MERTK, as well as the control gene UBE2F were amplified from human genomic DNA (Novagen) and individually cloned into the Psicheck 2 dual luciferase reporter vector (Promega). MDA-MB-231 or LM2 cells were then transfected with each reporter construct and MDA-MB-231 cells were also transfected with a miR-335 antagomir (25 nM) targeting endogenous miR-335 (ambion). Cells were lysed at 30 hours

after transfection and the ratio of Renilla and firefly luciferase was measured with the dual luciferase assay (Promega). Normalized Renilla to Firefly ratios were determined in the presence or absence of miR-335 inhibition. The segment (base pairs 449-509) of the SOX4 3' UTR containing the wildtype or mutated miR-335 target sequence at base pair 483 (tcttgg to tgaggg) were also cloned into the psicheck 2 dual luciferase vector.

Animal studies. All animal work was done in accordance with an approved protocol by the MSKCC Institutional Animal Care and Use Committee. NOD/SCID female mice (NCI) age-matched between 6-8 weeks were used for lung colonization, tumour growth, and orthotopic metastasis xenograft assays. Athymic female mice (NCI) age-matched between 6-8 weeks were used for bone metastasis xenograft assays. For primary tumour growth assays and orthotopic metastasis assays, 5X105 viable cells were re-suspended in a 1:1 mixture of PBS and growth-factor-reduced Matrigel (BD Biosciences) and injected orthotopically into the fourth mammary gland in a total volume of 25µl. Primary tumour growth rates were analyzed by measuring tumour length (L) and width (W), and calculating volume through the use of the formula $\pi LW2/6$, as previously described. Minn, Nature 436, 518-524. Orthotopic lung metastasis was quantified using non-invasive bioluminescence at 32 days post xenografting. For experimental metastasis assays, 1X104 viable cells were re-suspended in 0.1 ml PBS and injected into the lateral tail vein. Lung metastatic colonization was monitored and quantified using non-invasive bioluminescence as previously described. Minn, Nature 436, 518-524. For bone metastasis assays, 2X104 viable cells were re-suspended in 0.1 ml PBS and introduced into the arterial circulation via intra-cardiac injection, as previously described. Kang, Cancer Cell 3, 537-549. Bone metastatic colonization was monitored and quantified using non-invasive bioluminescence of normalized hind-limb photon flux at Day 35.

Trans-well migration and invasion assays. Cancer cells were conditioned overnight in 0.2% FBS ECM media without growth factors. The next day, cells were pulsed with 5 µM cell tracker green (Invitrogen) for 30 minutes, trypsinized, and re-suspended in 0.2% FBS ECM media without growth factors. Tumour cells were then seeded at 25000 cells per well into trans-well inserts (3µ pore size, BD Falcon) for migration assays or at 50000 cells per well into matrigel-coated, growth factor reduced, invasion chambers (8µ pore size, BD Biosciences). The wells were washed with PBS and fixed with 4% paraformaldehyde after 6 hours for migration assays or after 22 hours for invasion assays.

The cells on the apical side of each insert were then scraped off. The number of cells that had migrated to the basal side of the membrane was visualized with a Zeiss Axiovert 200M microscope at 10X magnification. Pictures of 3 random fields from three or six replicate wells were obtained and the number of cells that had migrated quantified in automated fashion using Metamorph analysis software.

Lung and tumour immunostaining. Mice were sacrificed and perfused with PBS followed by 4% paraformaldehyde. Tumours and lungs were extracted and fixed in 4% paraformaldehyde in PBS. Paraffin-embedding, sectioning, and Hematoxylin and Eosin staining of tumours and lungs were performed by Histoserv technologies. Immunohistochemical staining for cleaved caspase-3 (Cell Signaling) and phospho-histone-3 (Upstate) were performed on paraffin-embedded sections by the MSKCC Molecular Cytology Core Facility. Images were obtained at 10X magnification using Axioplan2 microscopy system (Zeiss) and apoptosis (cleaved caspase-3) and cell proliferation (phospho-histone-3) were quantified as number of cells per HPF averaged across 18 images from 3 tumours per cell line. Immunohistochemical staining for human vimentin (Novocastra) was performed on paraffin-embedded sections by the MSKCC Molecular Cytology Core Facility.

Statistical analysis and clinical validation of miR-335 signature. The Kaplan-Meier method was utilized for survival curve analysis, and the log-rank (Mantel-Cox) test was used to determine the statistical significance of difference between survival curves using Graphpad Prism 5 software. To determine the association of miRNA expression with metastatic relapse, the raw miRNA expression values for each tumour sample were normalized to the median expression value across the twenty tumour samples. Tumours whose expression values for each miRNA fell in the lowest third (7/20) of the entire set of tumours were deemed the low expression group for each miRNA. The Mantel-Cox log-rank test was then used to determine the statistical significance for differences between survival curves of patients from whom the tumours were resected. The Mann-Whitney test was used to determine the significance of difference between the expression levels for miRNAs in tumours that ultimately relapsed relative to those that did not. The correlation of miR-335 or miR-126 expression status with ER histologic or Her-2/Neu amplification status was determined by calculating the significance of the Spearman correlation coefficients in a two-tailed test using Graphpad Prism 5 software. Primary tumour gene expression data from the MSK (Minn, Nature 436, 518-524) and EMC (Wang, Lancet 365, 671-679 (2005)) cohorts

were used to classify each tumour as miR-335 signature positive if the sum of Z-scores for the expression values of the 6 miR-335 regulated genes was greater than one standard deviation from the mean. Previously published metastasis-free survival data corresponding to these tumours was then plotted in Kaplan-Meier form for patients whose tumours were miR-335 signature positive versus negative.

All of the patents and publications referenced herein are incorporated by reference in their entirety. All of the compositions and/or methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the composition and methods have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and/or methods and in the steps or sequence of steps of the methods described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents that are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope, and concept of the invention as defined by the appended claims. It is to be understood that all matter disclosed herein is to be interpreted merely as illustrative, and not in a limiting sense.

CLAIMS

1. A method for treating metastatic breast cancer comprising

administering to a patient diagnosed with breast cancer an agent, said agent

(a) comprising an oligonucleotide that reduces the expression of a gene in cancerous breast cells exposed to the oligonucleotide wherein the gene is overexpressed in both bone and lung metastases of breast cancer, and is down-regulated by miR-335, miR-126 and/or miR-206 and wherein the oligonucleotide has a sequence that is complementary to a portion of said gene or a transcript of said gene; and/or

(b) comprising an oligonucleotide that increases the amount of miR-335, miR-126 and/or miR-206,

thereby treating the breast cancer.

2. The method of claim 1, wherein the agent reduces expression of a gene selected from the group consisting of human SRY-box containing transcription factor 4 (SOX4)(Seq. ID No. 14), human Tenascin C (TCN) (Seq. ID No. 10), human receptor-type tyrosine protein phosphatase (PTPRN) (Seq. ID No. 16), human collagen, type I, alpha 1 (COL1A1) (Seq. ID No. 15), human Mer receptor tyrosine kinase (MERTK) (Seq. ID No. 17), and human phospholipase C beta 1 (PLCB1) (Seq. ID No. 18).

3. The method of claim 2, wherein the oligonucleotide comprises RNA residues that mediate RNAi to reduce expression of said gene.

4. The method of claim 2, wherein the agent reduce expression of SOX4 or TCN.

5. The method of claim 2, wherein the oligonucleotide is an shRNA.

6. The method of claim 4, wherein the oligonucleotide comprises a sequence selected from the group consisting of Seq. ID Nos. 4-9, 11-13, 19 and 20.

7. The method of claim 1 wherein the agent comprises Seq. ID No. 21 or 22.

8. The method of claim 1, wherein the agent increases the amount of miR-335, miR-126 and/or miR-206.

9. The method of claim 8, wherein the agent comprises an oligonucleotide comprising Seq. ID No. 1, 2 or 3 or the complement thereof

10. The method of claim 8, wherein the agent comprises an oligonucleotide comprising Seq. ID No. 11, 12 or 13 or the complement thereof.

11. The method of claim 1, further comprising the step of determining the metastaic potential of the breast cancer comprising obtaining a diagnostic signature from breast cancer cells indicative of the metastatic potential of the breast cancer cells;

wherein said diagnostic signature is obtained by measuring levels of one or more markers selected from the group consisting of SRY-box containing transcription factor 4 (SOX4), Tenascin C (TNC), receptor-type tyrosine protein phosphatase (PTPRN), collagen, type I, alpha 1 (COL1A1), Mer receptor tyrosine kinase (MERTK), phospholipase C beta 1 (PLCB1), miR-335, miR-206 and miR-126;

comparing said diagnostic signature to a control signature; and

based on the comparison, giving a prognosis of high risk for metastasis if the diagnostic signature is greater than the control signature by a threshold amount.

12. A method of diagnosing metastatic potential of breast cancer in a patient comprising

obtaining a diagnostic signature from breast cancer cells indicative of the metastatic potential of the breast cancer cells;

wherein said diagnostic signature is obtained by measuring levels in breast cells from the patient of one or more markers selected from the group consisting of SRY-box containing transcription factor 4 (SOX4), Tenascin C (TNC), receptor-type tyrosine protein phosphatase

(PTPRN), collagen, type I, alpha 1 (COL1A1), Mer receptor tyrosine kinase (MERTK), phospholipase C beta 1 (PLCB1), miR-335, miR-206 and miR-126;

comparing said diagnostic signature to a control signature; and

based on the comparison, giving a prognosis of high risk for metastasis if the diagnostic signature is greater than the control signature by a threshold amount.

13. Use of an oligonucleotide in the manufacture of a medicament for the treatment of breast cancer, wherein the oligonucleotide reduces the expression of a gene in cancerous breast cells exposed to the oligonucleotide wherein the gene is overexpressed in both bone and lung metastases of breast cancer, and is down-regulated by miR-335, miR-126 and/or miR-206 and wherein the oligonucleotide has a sequence that is complementary to a portion of said gene or a transcript of said gene; or increases the amount of miR-335, miR-126 and/or miR-206.

14. Use of claim 13, wherein the oligonucleotide comprises a sequence selected from the group consisting of Seq. ID No. 1-9, 11-13 and 19-22.

15. A kit for diagnosis of breast cancer comprising

reagents in packaged combination specifically for detecting or quantifying two or more materials selected from the group consisting of SRY-box containing transcription factor 4 (SOX4), Tenascin C (TNC), receptor-type tyrosine protein phosphatase (PTPRN), collagen, type I, alpha 1 (COL1A1), Mer receptor tyrosine kinase (MERTK), phospholipase C beta 1 (PLCB1), miR-335, miR-126, and miR-206.

16. The kit of claim 15 comprising reagents for detecting or quantifying SRY-box containing transcription factor 4 (SOX4).

17. The kit of claim 15 comprising reagents for detecting or quantifying Tenascin C (TNC).

18. The kit of claim 15 comprising reagents detecting or quantifying SRY-box containing transcription factor 4 (SOX4) and Tenascin C (TNC).

19. The kit of claim 15 comprising reagents for detecting or quantifying miR-335.

20. The kit of claim 15 comprising reagents for detecting or quantifying SRY-box containing transcription factor 4 (SOX4), Tenascin C (TNC), and miR-335.

21. A purified oligonucleotide comprising Seq. ID No. 5.

22. The oligonucleotide of claim 21 consisting of the sequence of Seq. ID No. 5

23. A purified oligonucleotide comprising Seq. ID No. 6.

24. The oligonucleotide of claim 23 consisting of the sequence of Seq. ID No. 6

25. A purified oligonucleotide comprising Seq. ID No. 4.

26. The oligonucleotide of claim 25 consisting of the sequence of Seq. ID No. 4.

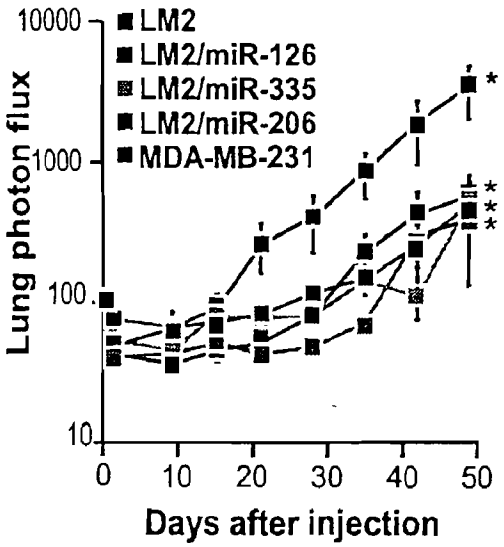

10000
1000
100
10
Lung photon flux
LM2
LM2/miR-126
LM2/miR-335
LM2/miR-206
MDA-MB-231
0
10
20
30
40
50
Days after injection

Fig. 1A

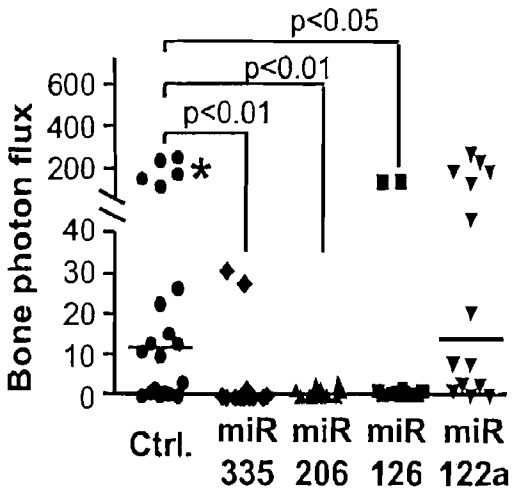

p<0.05
p<0.01
p<0.01
600
400
200
40
30
20
10
0
Bone photon flux
Ctrl.
miR 335
miR 206
miR 126
miR 122a

Fig. 1B

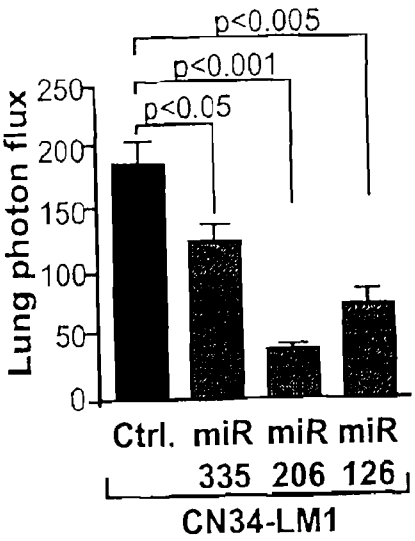

p<0.005
p<0.001
p<0.05
Lung photon flux
250
200
150
100
50
0
Ctrl.
miR 335
miR 206
miR 126
CN34-LM1

Fig. 1C

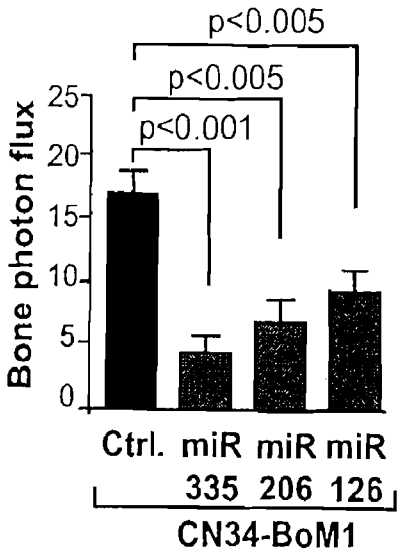

p<0.005
p<0.005
p<0.001
Bone photon flux
25
20
15
10
5
0
Ctrl.
miR 335
miR 206
miR 126
CN34-BoM1

Fig. 1D

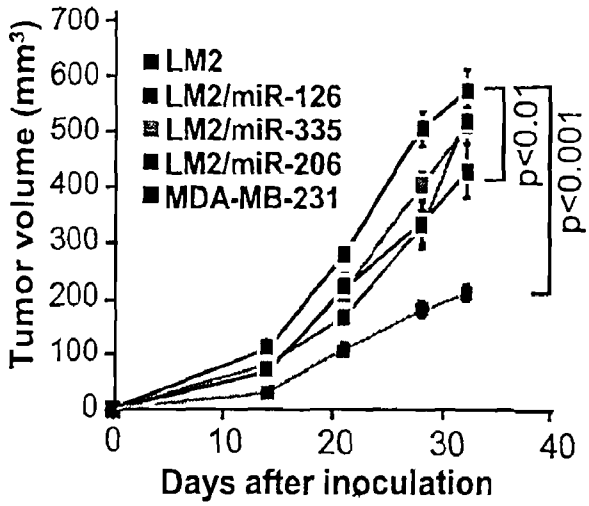

Tumor volume (mm3)
700
600
500
400
300
200
100
0
LM2
LM2/miR-126
LM2/miR-335
LM2/miR-206
MDA-MB-231
p<0.01
p<0.001
0
10
20
30
40
Days after inoculation

Fig. 2A

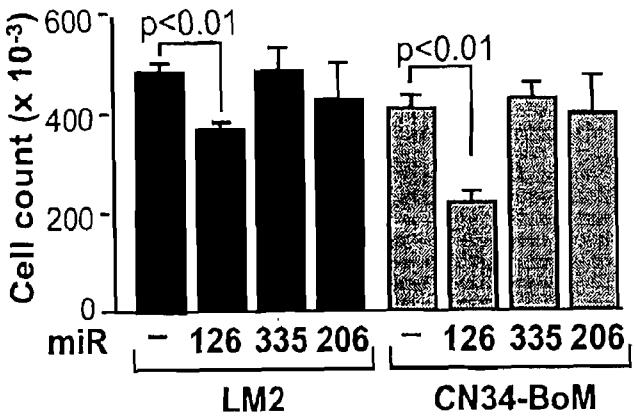

Cell count (x 10-3)
600
400
200
0
p<0.01
p<0.01
miR
–
126
335
206
–
126
335
206
LM2
CN34-BoM

Fig. 2B

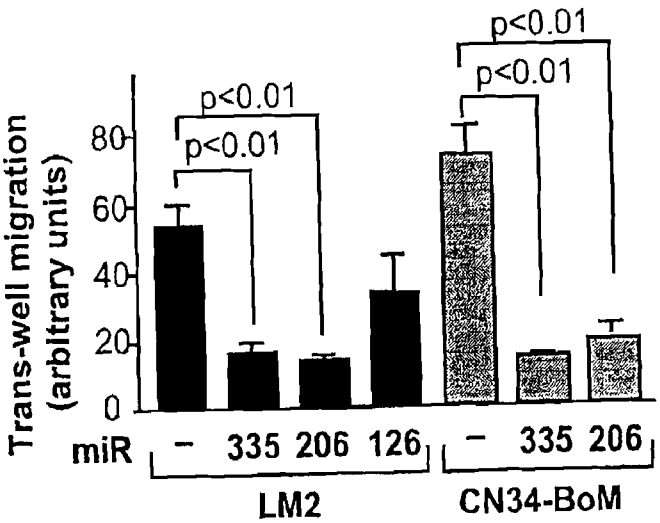

p<0.01
p<0.01
p<0.01
p<0.01
Trans-well migration (arbitrary units)
80
60
40
20
0
miR
-
335
206
126
-
335
206
LM2
CN34-BoM

Fig. 2C

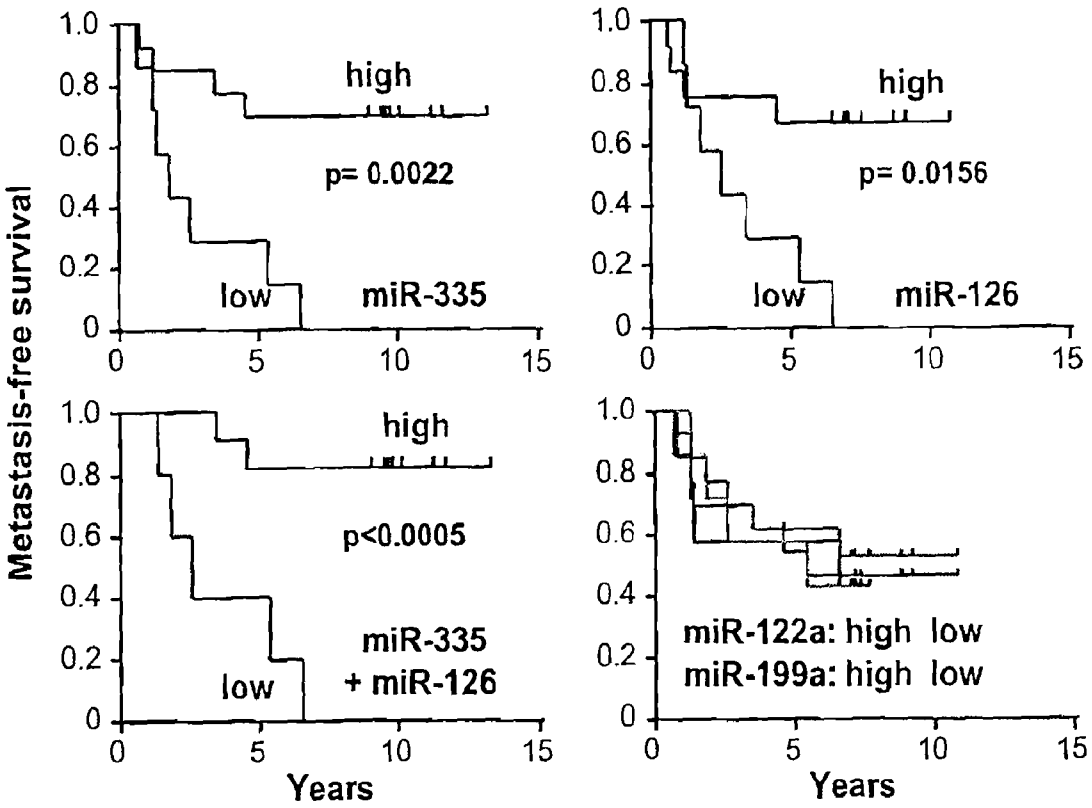

Metastasis-free survival
1.0
0.8
0.6
0.4
0.2
0
high
p= 0.0022
low
miR-335
0
5
10
15
high
p= 0.0156
low
miR-126
high
p<0.0005
low
miR-335
+ miR-126
Years
miR-122a: high low
miR-199a: high low
Years

Fig. 3A

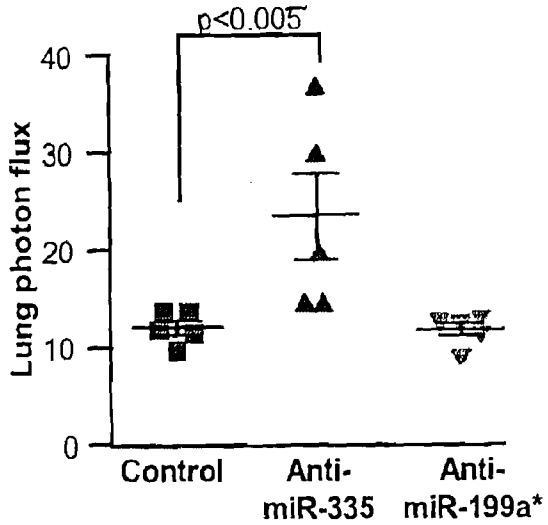

p<0.005
Lung photon flux
40
30
20
10
0
Control
Anti-
miR-335
Anti-
miR-199a*

Fig. 3B

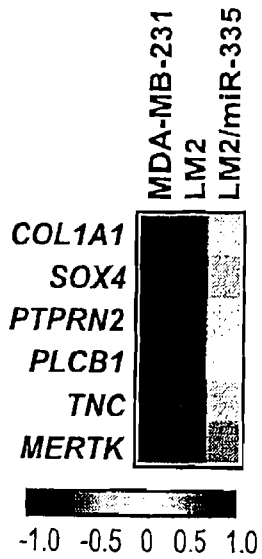

MDA-MB-231
LM2
LM2/miR-335
COL1A1
SOX4
PTPRN2
PLCB1
TNC
MERTK
-1.0 -0.5 0 0.5 1.0

Fig. 4A

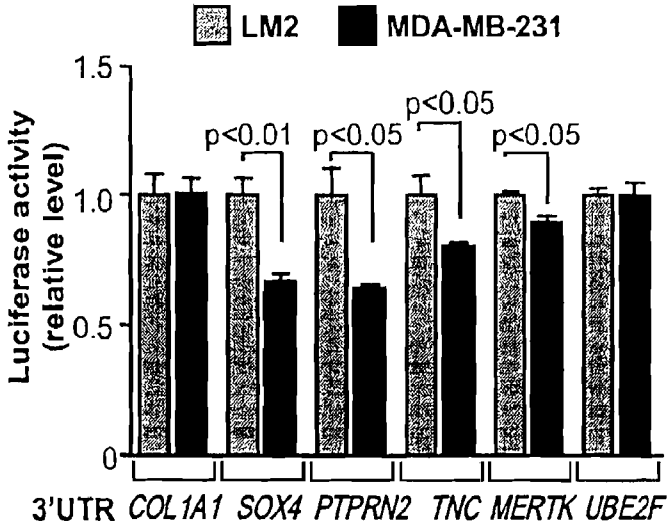

LM2
MDA-MB-231
Luciferase activity (relative level)
1.5
1.0
0.5
0
p<0.01
p<0.05
p<0.05
p<0.05
3'UTR COL1A1 SOX4 PTPRN2 TNC MERTK UBE2F

Fig. 4B

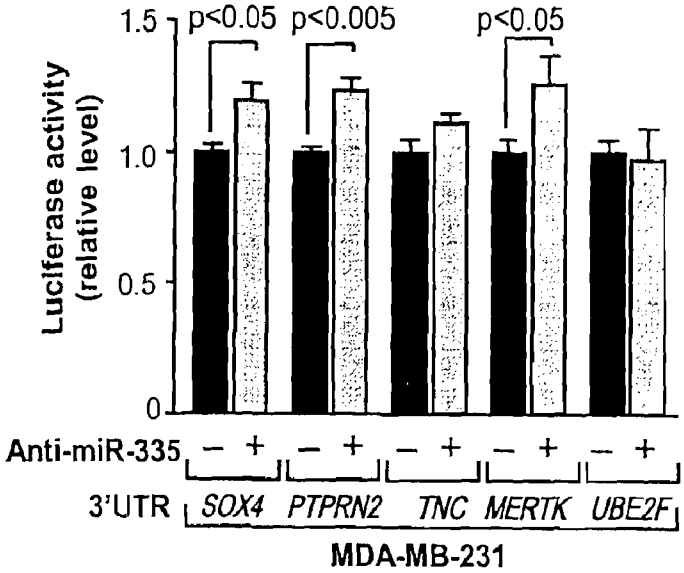

1.5
1.0
0.5
0
p<0.05
p<0.005
p<0.05
Luciferase activity (relative level)
Anti-miR-335 – + – + – + – + – +
3'UTR SOX4 PTPRN2 TNC MERTK UBE2F
MDA-MB-231

Fig. 4C

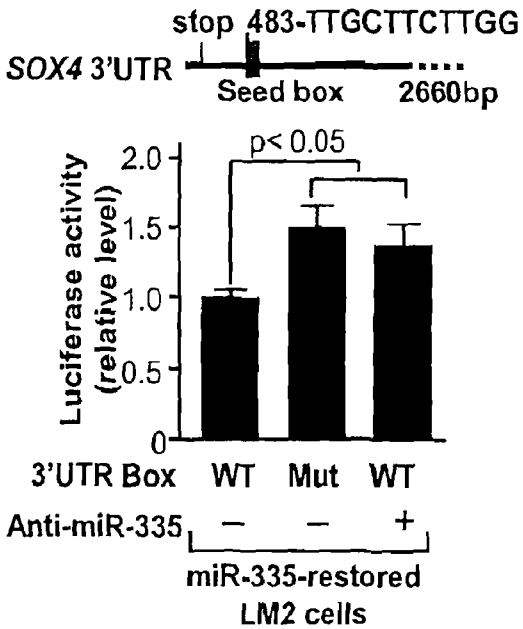

stop 483-TTGCTTCTTGG
SOX4 3'UTR
Seed box
2660bp
p< 0.05
2.0
1.5
1.0
0.5
0
Luciferase activity (relative level)
3'UTR Box WT Mut WT
Anti-miR-335 – – +
miR-335-restored LM2 cells

Fig. 4D

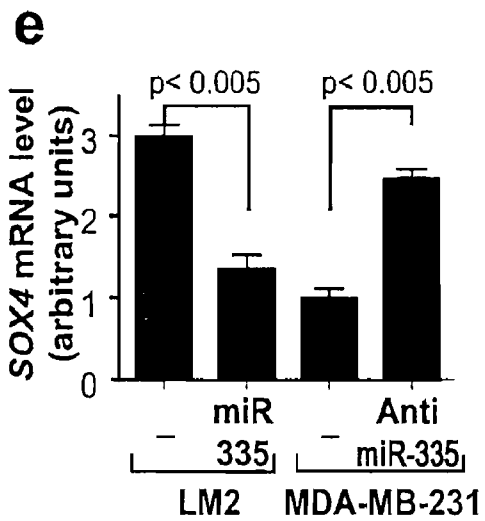

e
p< 0.005
p< 0.005
SOX4 mRNA level (arbitrary units)
0
1
2
3
–
miR 335
–
Anti miR-335
LM2
MDA-MB-231

Fig. 4E

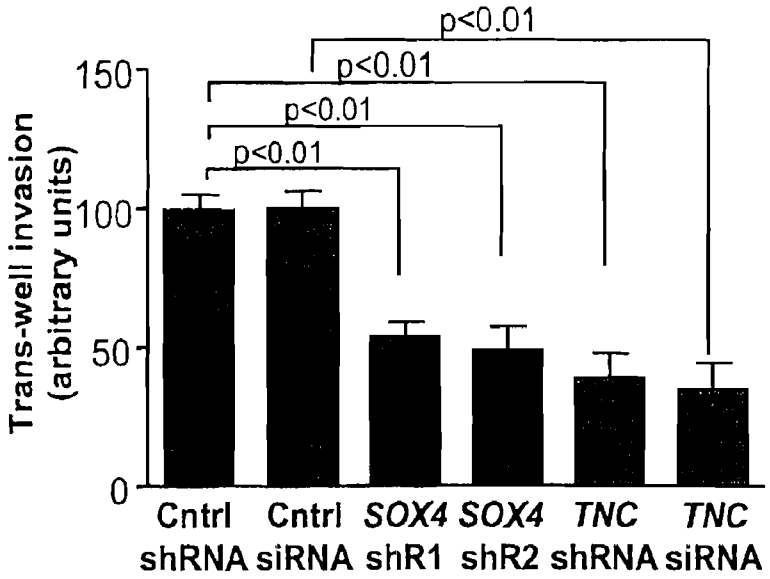

p<0.01
p<0.01
p<0.01
p<0.01
150
100
50
0
Trans-well invasion (arbitrary units)
Cntrl shRNA
Cntrl siRNA
SOX4 shR1
SOX4 shR2
TNC shRNA
TNC siRNA

Fig. 5A

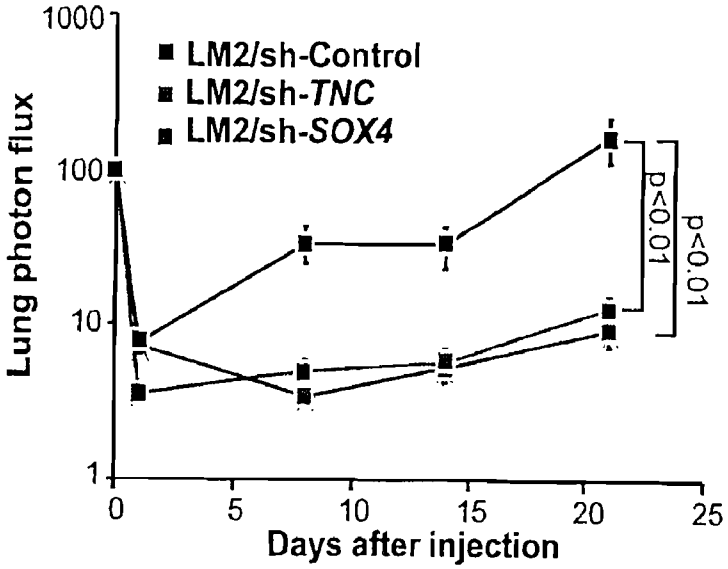

1000
100
10
1
Lung photon flux
LM2/sh-Control
LM2/sh-TNC
LM2/sh-SOX4
p<0.01
p<0.01
0
5
10
15
20
25
Days after injection

Fig. 5B

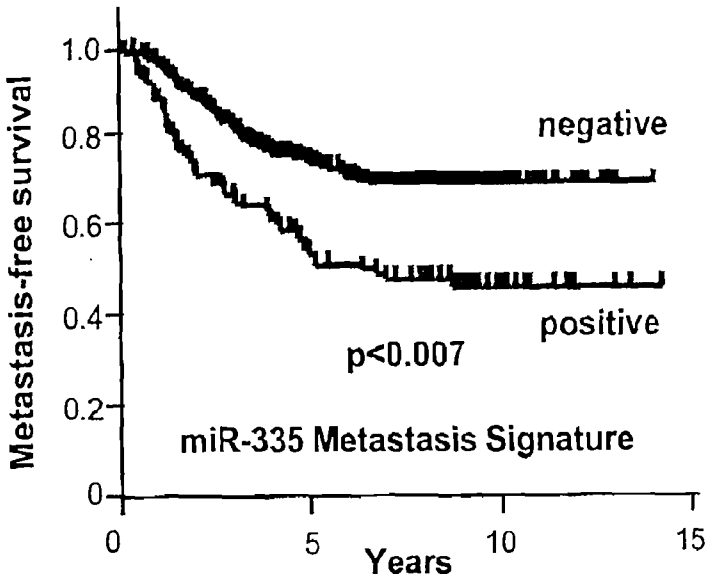

Metastasis-free survival
1.0
0.8
0.6
0.4
0.2
0
0
5
10
15
Years
negative
positive
p<0.007
miR-335 Metastasis Signature

Fig. 5C

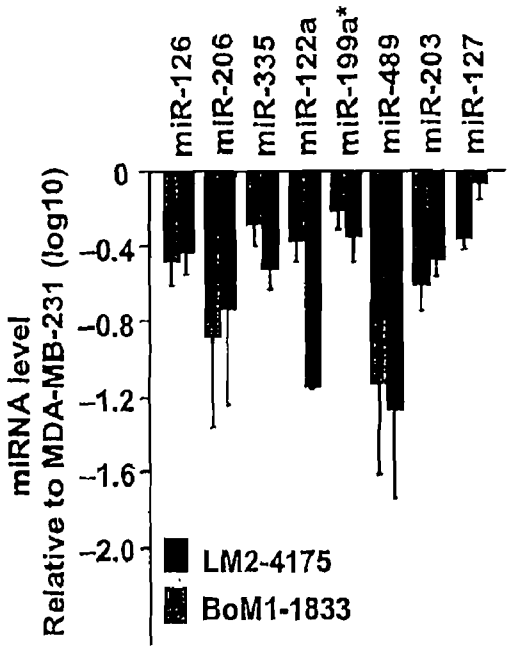
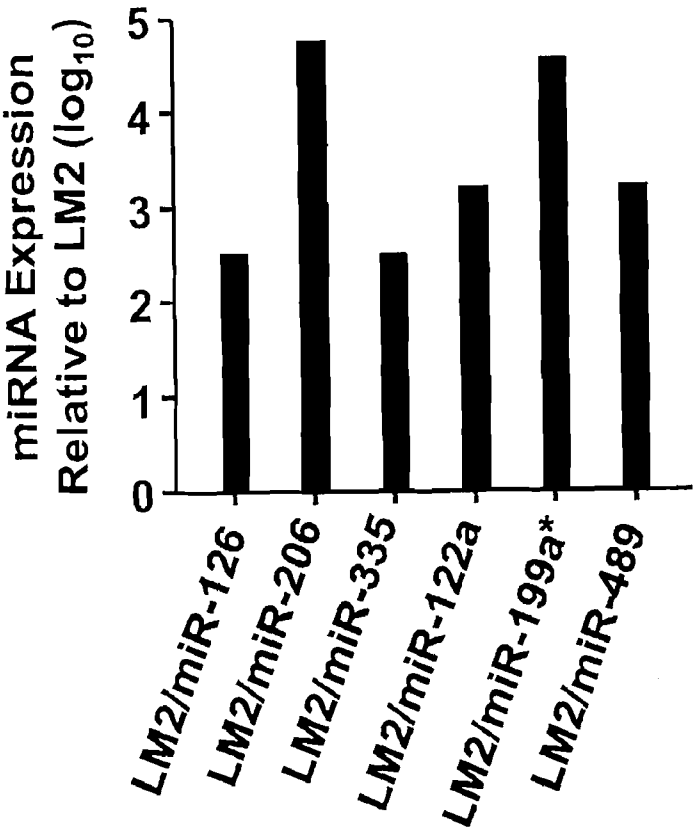

miR-126
miR-206
miR-335
miR-122a
miR-199a*
miR-489
miR-203
miR-127
0
-0.4
-0.8
-1.2
-1.6
-2.0
miRNA level
Relative to MDA-MB-231 (log10)
LM2-4175
BoM1-1833

Fig. 6


miRNA Expression
Relative to LM2 (log10)
5
4
3
2
1
0
LM2/miR-126
LM2/miR-206
LM2/miR-335
LM2/miR-122a
LM2/miR-199a*
LM2/miR-489

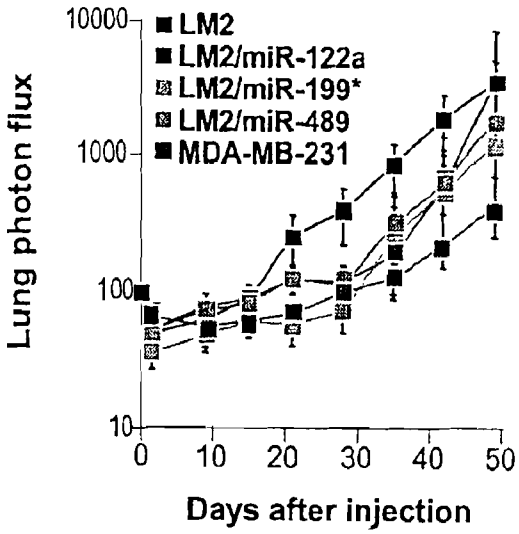

10000
1000
100
10
Lung photon flux
LM2
LM2/miR-122a
LM2/miR-199*
LM2/miR-489
MDA-MB-231
0
10
20
30
40
50
Days after injection

Fig. 8

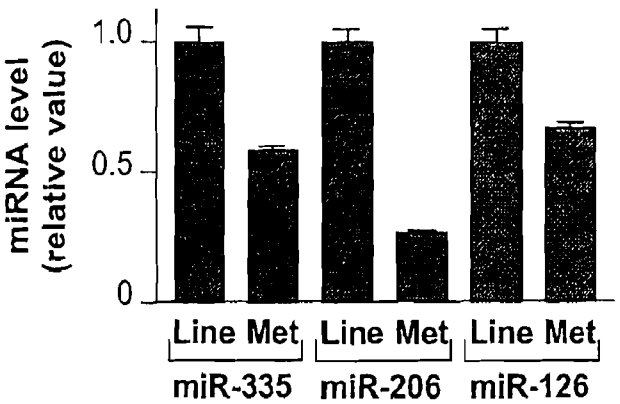

1.0
0.5
0
miRNA level
(relative value)
Line
Met
Line
Met
Line
Met
miR-335
miR-206
miR-126

Fig. 9A

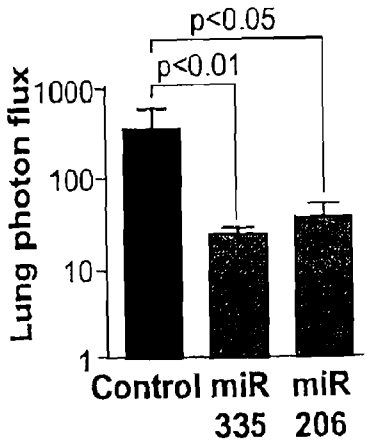

p<0.05
p<0.01
Lung photon flux
1000
100
10
1
Control
miR 335
miR 206

Fig. 9B

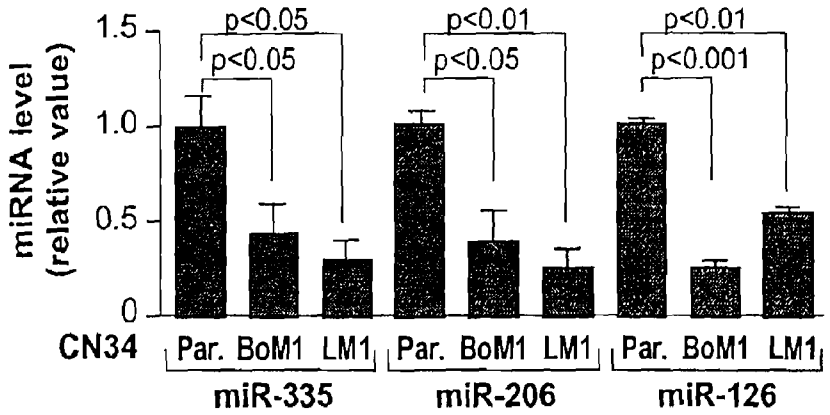

miRNA level (relative value)
1.5
1.0
0.5
0
p<0.05
p<0.05
p<0.01
p<0.05
p<0.01
p<0.001
CN34
Par. BoM1 LM1
Par. BoM1 LM1
Par. BoM1 LM1
miR-335
miR-206
miR-126

Fig. 9C

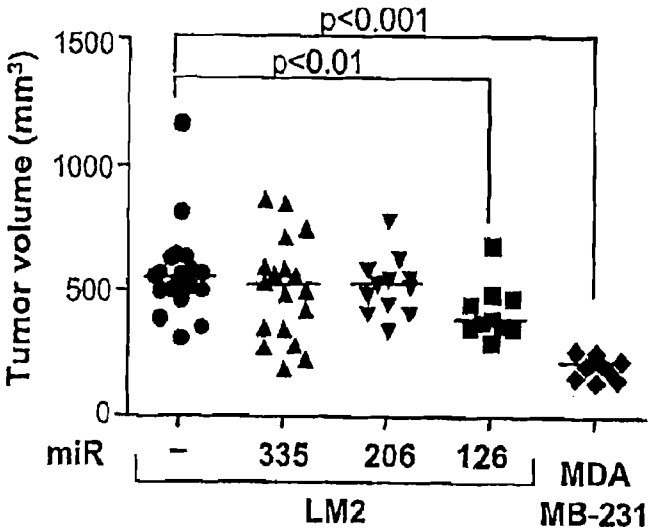

p<0.001
p<0.01
1500
1000
500
0
Tumor volume (mm³)
miR
–
335
206
126
MDA
MB-231
LM2

Fig. 10

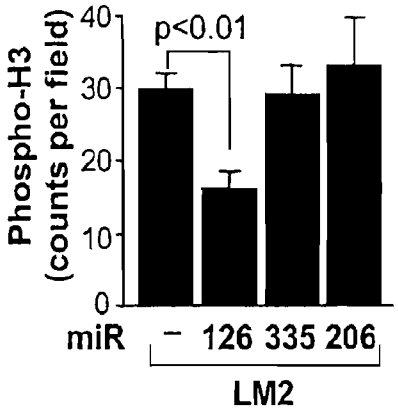


Fig. 11A

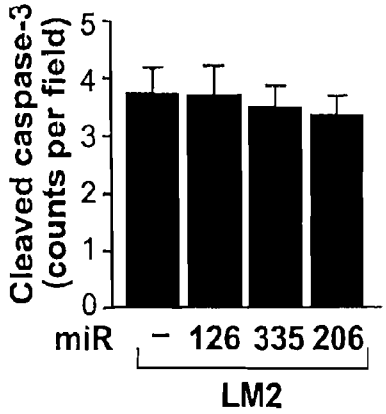


Fig. 11B

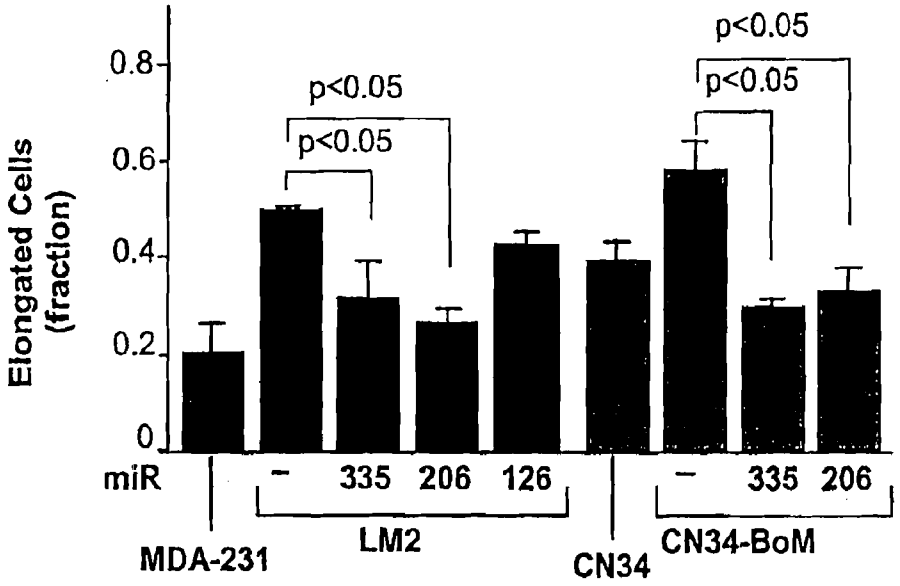

Elongated Cells (fraction)
0.8
0.6
0.4
0.2
0
p<0.05
p<0.05
p<0.05
p<0.05
miR
–
335
206
126
–
335
206
MDA-231
LM2
CN34
CN34-BoM

Fig. 12

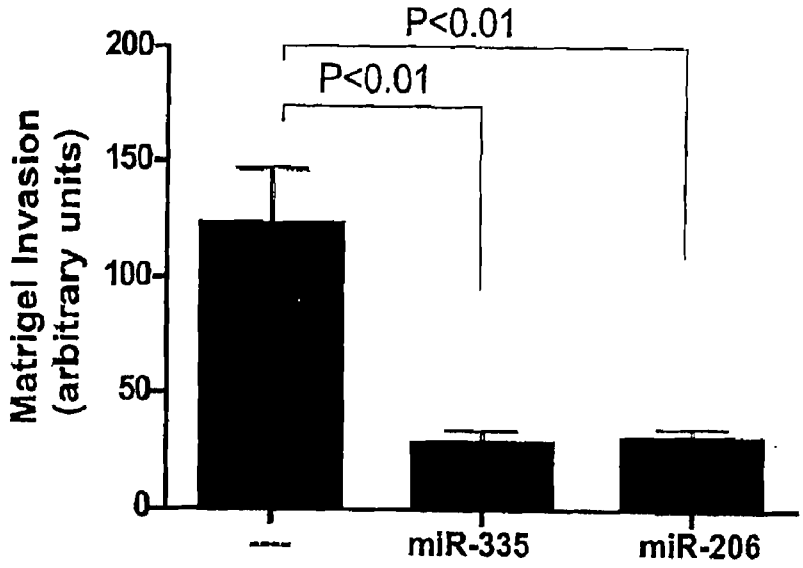

Matrigel Invasion (arbitrary units)
200
150
100
50
0
P<0.01
P<0.01
---
miR-335
miR-206

Fig. 13

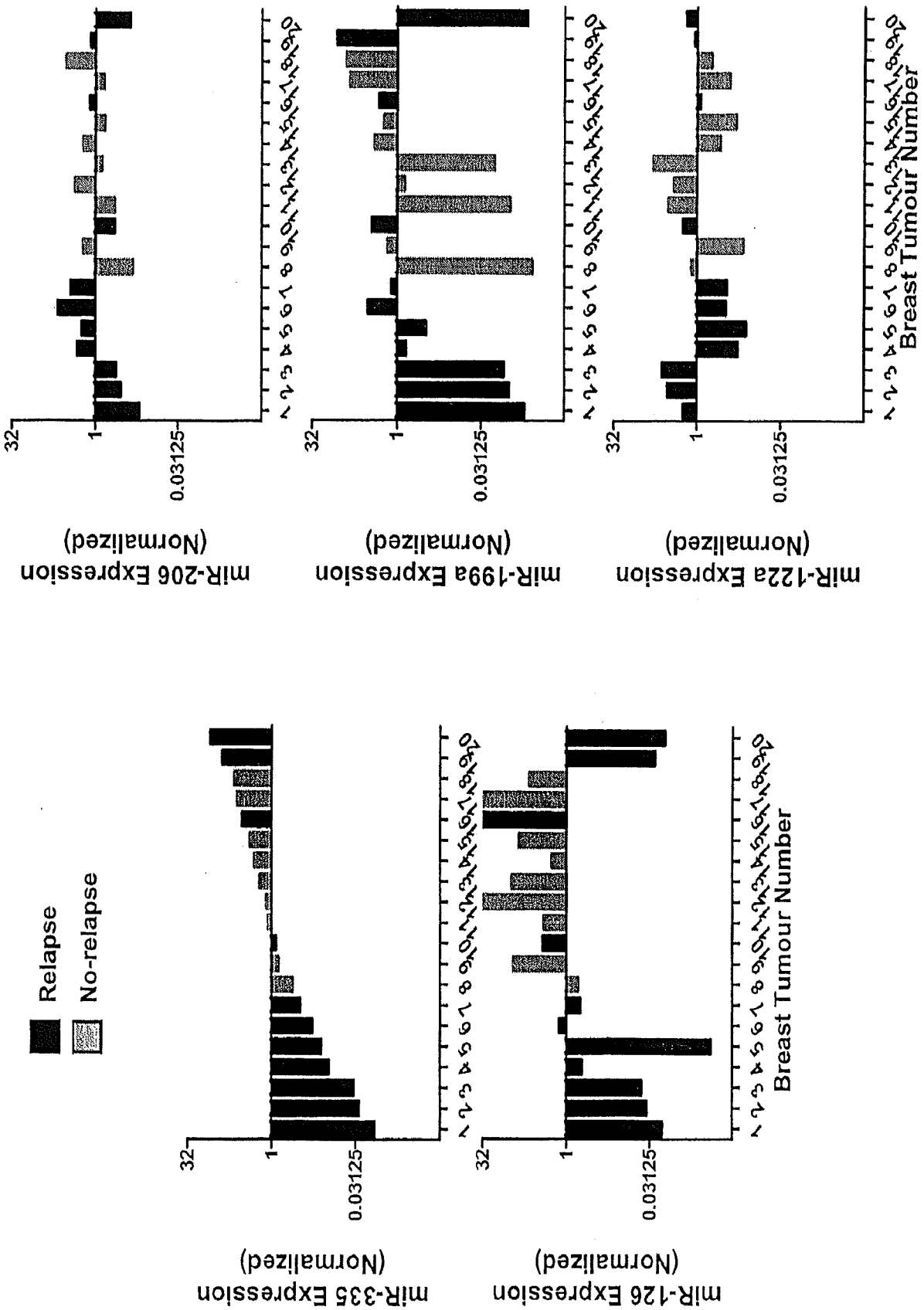

Relapse
No-relapse
miR-126 Expression (Normalized)
miR-335 Expression (Normalized)
miR-122a Expression (Normalized)
miR-199a Expression (Normalized)
miR-206 Expression (Normalized)
32
1
0.03125
Breast Tumour Number

Fig. 14

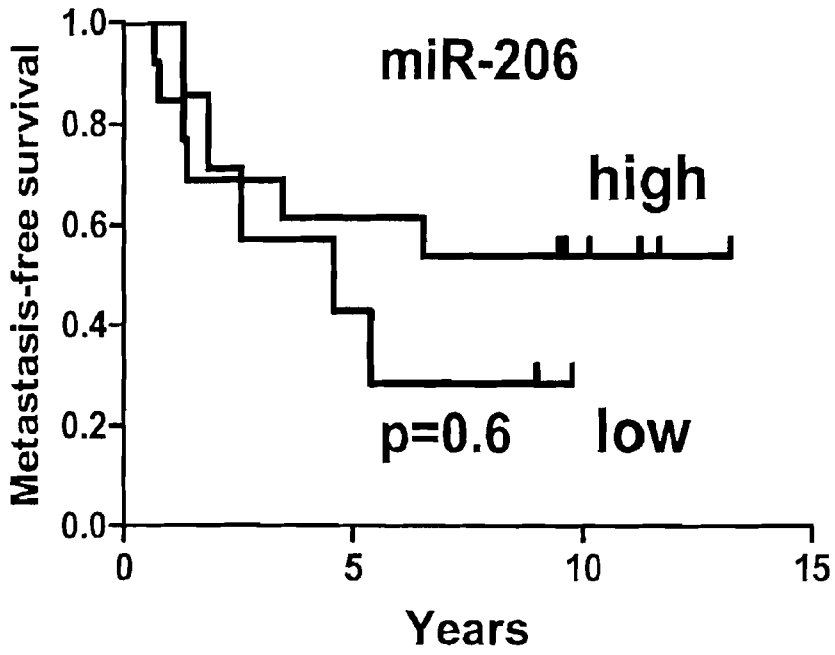

miR-206
high
low
p=0.6
Metastasis-free survival
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
Years

Fig. 15

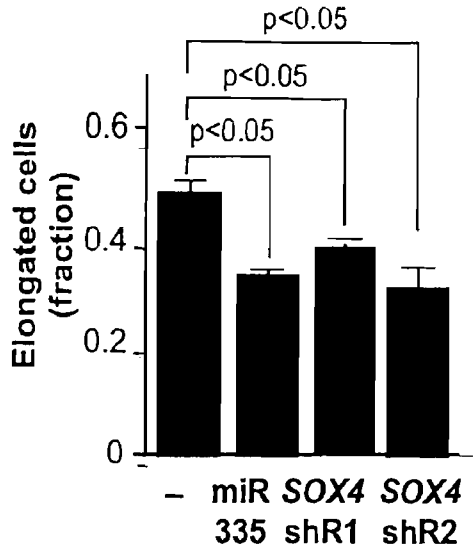


Fig. 16A

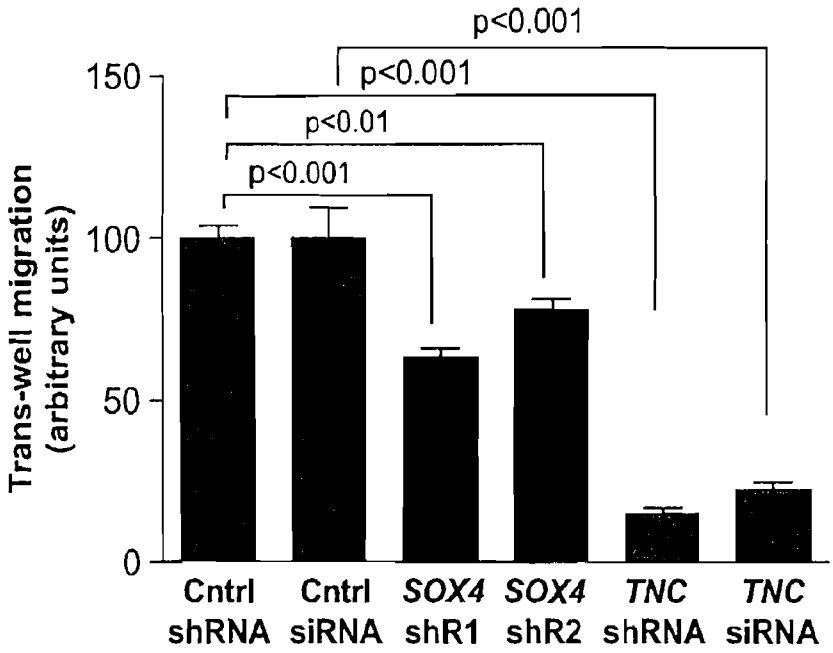


Fig. 16B

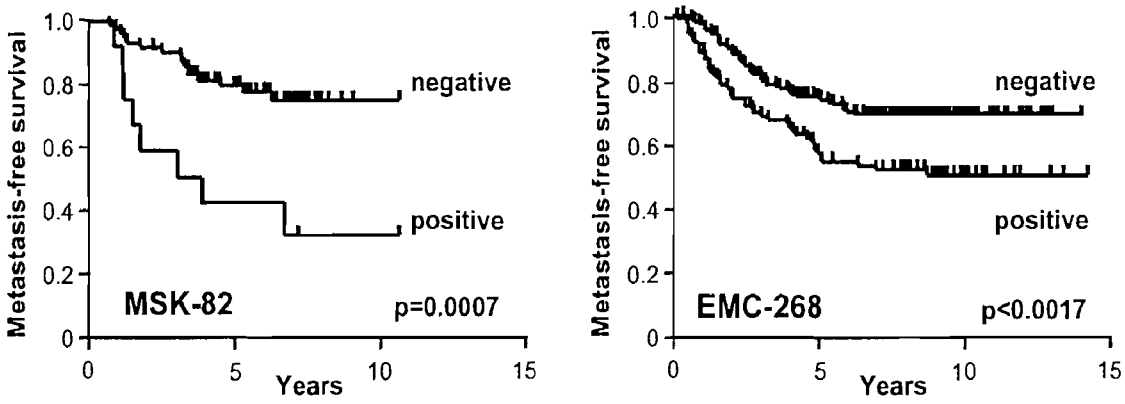


Fig. 17A

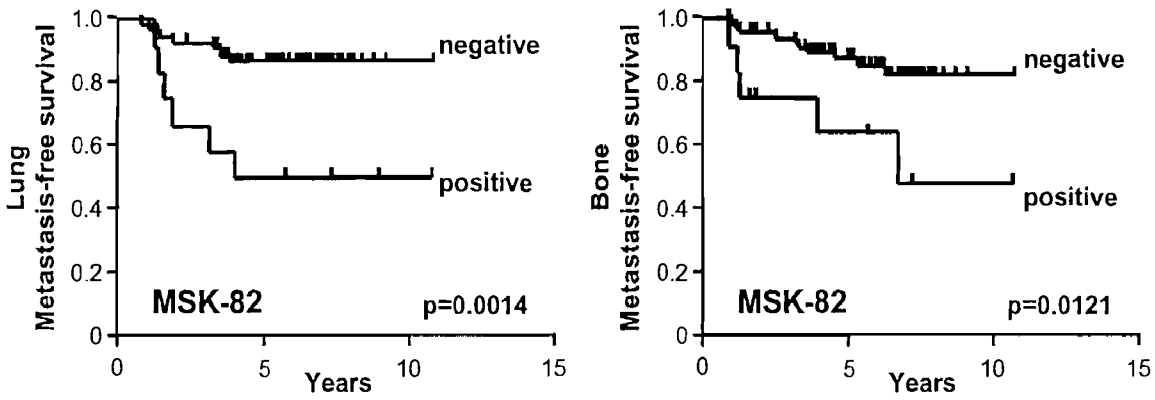


Fig. 17B

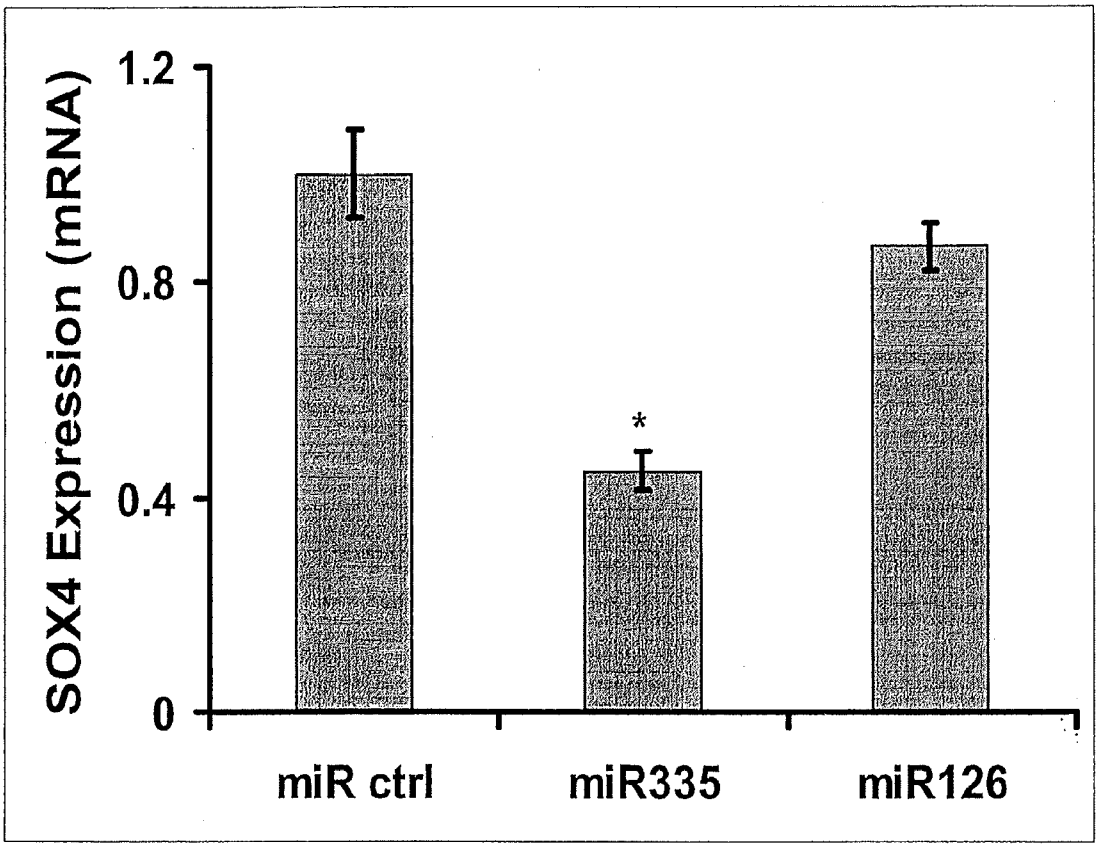


Fig. 18

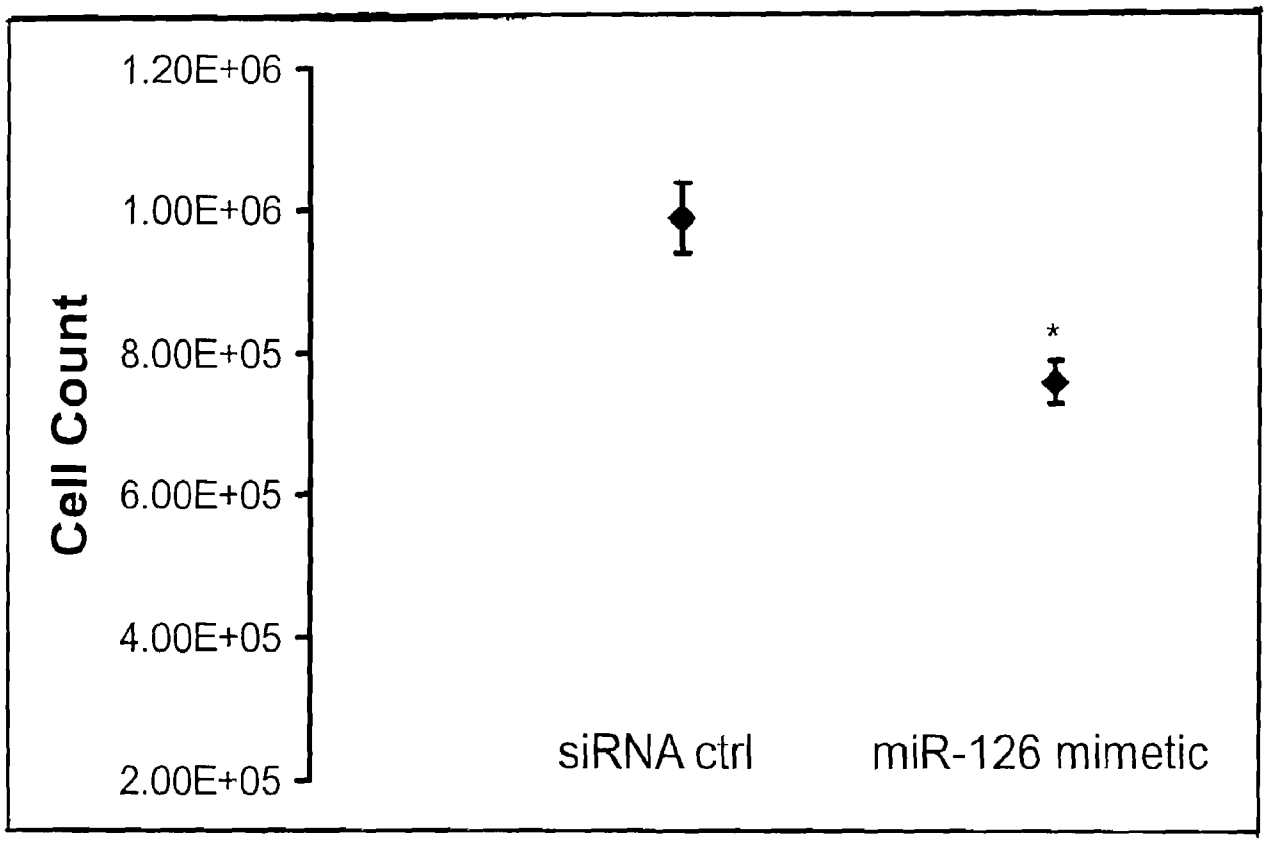

1.20E+06
1.00E+06
8.00E+05
6.00E+05
4.00E+05
2.00E+05
Cell Count
*
siRNA ctrl
miR-126 mimetic

Fig. 19