US 20130281502A1

(19) **United States**

(12) **Patent Application Publication**

Sgroi et al.

(10) **Pub. No.: US 2013/0281502 A1**

(43) **Pub. Date: Oct. 24, 2013**

(54) **POST-TREATMENT BREAST CANCER PROGNOSIS**

(71) Applicants: **Dennis Sgroi**, Cambridge, MA (US); **Mark G. Erlander**, Carlsbad, CA (US); **Yi Zhang**, San Diego, CA (US); **Catherine A. Schnabel**, San Diego, CA (US)

(72) Inventors: **Dennis Sgroi**, Cambridge, MA (US); **Mark G. Erlander**, Carlsbad, CA (US); **Yi Zhang**, San Diego, CA (US); **Catherine A. Schnabel**, San Diego, CA (US)

(73) Assignees: **GENERAL HOSPITAL CORPORATION**, Boston, MA (US); **BIOTHERANOSTICS INC.**, San Diego, CA (US)

(21) Appl. No.: **13/914,617**

(22) Filed: **Jun. 10, 2013**

Related U.S. Application Data

(63) Continuation of application No. PCT/US2011/064290, filed on Dec. 9, 2011.

(60) Provisional application No. 61/421,627, filed on Dec. 9, 2010.

Publication Classification

(51) **Int. Cl.**
C12Q 1/68 (2006.01)

(52) **U.S. Cl.**
CPC ***C12Q 1/6886*** (2013.01)
USPC **514/383**; 435/6.12; 506/7

(57) **ABSTRACT**

The disclosure includes the identification and use of gene expression profiles, or patterns, with clinical relevance to extended treatment and cancer-free survival in a patient. In particular, the disclosure includes the identities of genes that are expressed in correlation with benefit in a switch in endocrine therapy used to treat a patient. The levels of gene expression are disclosed as a molecular index for predicting clinical outcome, and so prognosis, for the patient. The disclosure further includes methods for predicting cancer recurrence, and/or predicting occurrence of metastatic cancer, after initial treatment with an anti-estrogen agent. The disclosure further includes methods for determining or selecting the treatment of a subject based upon the likelihood of life expectancy, cancer recurrence, and/or cancer metastasis.

POST-TREATMENT BREAST CANCER PROGNOSIS

RELATED APPLICATIONS

[0001] This application is a continuation of PCT Application PCT/US2011/064290, filed Dec. 9, 2011 and published as WO 2012/079059 with designation of the U.S., and which claims benefit of priority to U.S. Provisional Patent Application 61/421,627, filed Dec. 9, 2010, both of which is hereby incorporated by reference in their entireties as if fully set forth herein.

[0002] This application is related to International Application No. PCT/US2008/075528, filed on Sep. 6, 2008 (published as WO 2009/108215 A1) with designation of the U.S., and to U.S. patent application Ser. No. 12/718,973, filed Mar. 6, 2010. Both applications are hereby incorporated by reference as if fully set forth herein.

FIELD OF THE DISCLOSURE

[0003] The disclosure relates to the identification and use of gene expression profiles, or patterns, with clinical relevance to breast cancer. In particular, the disclosure is based in part on the identities of genes that are expressed in correlation with the likelihood of cancer recurrence after initial treatment with an aromatase inhibitor or other endocrine therapy. The levels of gene expression form a molecular index that is able to predict clinical outcome, and so prognosis, for a patient after initial treatment with an aromatase inhibitor or other endocrine therapy.

[0004] The gene expression profiles, whether embodied in nucleic acid expression, protein expression, or other expression formats, may be used to predict the post-treatment clinical outcome of subjects afflicted with breast cancer, predict cancer recurrence, and/or predict occurrence of metastatic cancer. The profiles may also be used in the study of a subject's prognosis. When used for prognosis, the profiles are used to determine the treatment of cancer based upon the likelihood of life expectancy, cancer recurrence, and/or cancer metastasis.

BACKGROUND OF THE DISCLOSURE

[0005] The treatment of breast cancer has been a field of intense interest and study. After initial diagnosis of breast cancer by analysis of a sample of breast cancer cells from a subject, treatment methods often begin with surgical removal of the tumor cells. In cases of hormone-dependent breast cancer, such as estrogen receptor positive (ER+) breast cancer, the surgery is followed by antagonizing estrogen to reduce tumor growth or re-growth. In many cases, treatment with the anti-estrogen tamoxifen is used for five years to reduce the risk of disease recurrence and so breast cancer mediated mortality.

[0006] Unfortunately, data from the field indicate that more than half of all breast cancer recurrences occur after five years of treatment with adjuvant tamoxifen.

[0007] Goss et al. (*J. Clin. Oncol.,* 26(12):1948-1955, 2008) report results from a trial examining the use of letrozole started within 3 months after five years of adjuvant tamoxifen in subjects with primary ER+ breast cancer. The results suggested that post-tamoxifen treatment with letrozole improves breast cancer-free survival and distant breast cancer-free survival.

[0008] But Goss et al. provided no means by which to predict which subjects, treated for five years with tamoxifen, would benefit from subsequent letrozole treatment. Therefore, there was no means to direct letrozole treatment only to the subjects for whom a benefit is expected. So letrozole treatment was applied to subjects for whom no benefit would have been expected, resulting in an overtreatment of the population of breast cancer-free subjects treated with for five years with tamoxifen.

[0009] The citation of documents herein is not to be construed as reflecting an admission that any is relevant prior art. Moreover, their citation is not an indication of a search for relevant disclosures. All statements regarding the dates or contents of the documents is based on available information and is not an admission as to their accuracy or correctness.

BRIEF SUMMARY OF THE DISCLOSURE

[0010] The disclosure is based in part on the discovery and determination of gene expression levels in breast cancer tumor cells that are correlated with a beneficial switch in anti-breast cancer chemotherapy. In some cases, the switch is from one form of endocrine therapy to another. The expression levels may be used to provide prognostic information, such as cancer recurrence, and predictive information, such as responsiveness to certain therapies.

[0011] In a first aspect, the disclosure includes a method to identify, or classify, a population of subjects initially treated with an anti-estrogen or anti-aromatase therapy into at least two subpopulations. A first subpopulation would be expected to benefit from a switch in therapy, such as a switch to another anti-estrogen or anti-aromatase therapy. A second subpopulation would not be expected to benefit. In some cases, the initial therapy is with tamoxifen, such as adjuvant tamoxifen therapy for a period of about five years or less. Optionally, the switch is to letrozole, or other anti-aromatase, therapy. The disclosure includes means for a population of subjects treated in this manner, and breast cancer-free during treatment, to be classified into the first, and/or the second, subpopulations.

[0012] The methods of the disclosure are based on the expression levels of certain genes, including the expression level of HoxB13, in breast cancer cells of a subject. In some embodiments, a two-gene ratio of HoxB13 expression to IL17BR expression (or HoxB13:IL17BR ratio) may be used (see Ma et al., *J. Clin. Oncol.,* 24:4611-9 (2006). In alternative embodiments, a two-gene ratio of HoxB13 expression to CHDH expression may be used.

[0013] The HoxB13:IL17BR (H:I) ratio was discovered based upon a study of novel biomarkers predictive of clinical outcome beyond standard prognostic factors. Patients who developed cancer recurrences were matched to those who did not with respect to tumor stage and grade. The simple H:I ratio was found to be suitable for predicting cancer recurrence in patients with estrogen receptor-positive (ER+) breast cancer receiving adjuvant tamoxifen therapy. Subsequent studies (Goetz et al., *Clin Cancer Res.* 12:2080-7 (2006); Jerevall et al., *Breast Cancer Res. Treat* (2007); Jansen et al., *J. Clin. Oncol.* 25:662-8 (2007)) have further shown that the ratio is both prognostic, such as by being an indicator of tumor aggressiveness, and predictive of tamoxifen benefit within both retrospective and randomized clinical trials.

[0014] In further embodiments, the disclosure includes one or more additional genes in combination with HoxB13 expression. The combination may be with any one, two, three, four or all five of the additionally disclosed genes as follows.

[0015] The additional genes of the disclosure encode Bub1B ("budding uninhibited by benzimidazoles 1 beta) or p21 protein-activated kinase 6 (PAK6); CENPA (centromere protein A, isoform a); NEK2 (NIMA-related kinase 2 or "never in mitosis gene a"-related kinase 2); RACGAP1 (Rac GTPase activating protein 1); and RRM2 (ribonucleotide reductase M2). The use of these five genes alone is referred to herein as the Molecular Grade Index (MGI). Aspects of the disclosure include compositions and methods are described for the use of HoxB13 expression, with or without IL17BR expression, in combination with expression level(s) of one or more of the above five genes to study, to provide prognostic information, and/or provide predictions of clinical responsiveness.

[0016] Thus the disclosure is based in part on the discovery that gene expression level(s) are useful for providing prognostic determinations (such as the likelihood of cancer recurrence in the form of breast cancer recurrence either locally or distally or in the form of metastasis) and predictive determinations (such as responsiveness to a course of treatment) for a subject. The use of all seven disclosed genes is referred to as the Breast Cancer Index (BCI).

[0017] When the expression levels of the BCI were analyzed using real-time reverse transcription-polymerase chain reaction (RT-PCR), the combination was found to provide superior stratification of risk of recurrence in subjects treated with five years of tamoxifen therapy. This reflects an unexpected discovery because it identifies for the first time a predictor for beneficial switching of breast cancer therapies.

[0018] In additional aspects, HoxB13 expression, and/or the BCI, may be used to predict late recurrence of cancer in a breast cancer patient. Non-limiting examples of late recurrence include after 5 years of treatment with tamoxifen, but also includes after 4 years, after 3 years, or after 2 years or less time of treatment with tamoxifen. Similarly, HoxB13 expression, and/or the BCI, may be used to predict responsiveness to letrozole or other anti-estrogen or anti-aromatase therapy after the above time periods to inhibit late recurrence.

[0019] Embodiments of the disclosure include an assay method with prognostic value and predictive value for stratifying subjects with original ER+ breast cancer and subsequent breast cancer-free treatment. As a prognostic, the stratification may be based on differential expression levels that correlate with, and so indicate, need for a switch in breast cancer therapies as a non-limiting example. As a non-limiting example, the stratification (based on expression levels) may be used to predict endocrine sensitivity (such as sensitivity to letrozole as a non-limiting example) and/or prediction of benefit from anti-estrogen and/or anti-aromatase inhibitors. The detection of gene expression may of course be in any suitable cell containing sample as described herein. Non-limiting examples of cells for use in the disclosure include those freshly isolated from the subject, those frozen after isolation, and those that are fixed and/or embedded, such as formalin fixed, paraffin embedded (FFPE). In most embodiments, the cells are breast cells, such as breast cancer cells.

[0020] In some embodiments, a method based on the expression levels is advantageously used on a breast cancer cell containing sample from a subject, such as a DCIS sample. As a non-limiting example, the cell may be one from a pre-operative histological sample used to diagnose cancer in the subject. For such a subject, the standard of care is surgery, with breast conserving surgery preferred over a radical mastectomy, to remove the DCIS. This is often followed by post-operative radiotherapy, optionally with endocrine therapy, such as treatment with tamoxifen, a selective estrogen receptor modulator (SERM), a selective estrogen receptor down-regulator (SERD), or an aromatase inhibitor (AI) such as letrozole. In other post-operative cases, endocrine therapy is administered without radiation, and optionally with chemotherapy.

[0021] The instant disclosure is directed to the identification of a subject as expected to benefit from a switch in endocrine therapy, such as from one type of endocrine therapy to another, after breast cancer-free survival during the course of the initial endocrine therapy. In additional embodiments, the switch may be made after an initial course of endocrine therapy for 5 years, 4 years, 3 years, or 2 years.

[0022] The disclosure also includes detecting gene expression where high HoxB13 expression is an indicator of increased likelihood of cancer recurrence in the subject following an initial endocrine therapy, such as adjuvant tamoxifen therapy. The methods may thus include identifying the subject as likely, or unlikely, to experience local cancer recurrence, and further include switching treatment modalities for the subject to address the expected outcome. As a non-limiting example, determination of a likelihood of recurrence in the absence of an extended, post-initial treatment, therapy may be used to confirm the suitability of, or to select, an extended therapy with a switch in the anti-estrogen and/or anti-aromatase modality used.

[0023] In some cases, the disclosed methods may be used to select or eliminate therapies for premenopausal women, or for postmenopausal women, that have undergone treatment with endocrine therapy and remained cancer-free during that time. Premenopausal women include those who are less than about 35 years of age. The method may include assaying a breast cancer cell containing sample from a subject for expression of the disclosed genes. As a non-limiting example, the cell may be one from a pre-operative histological sample used to diagnose cancer in the subject.

[0024] Non-limiting examples of endocrine therapy include treatment with an SERM, such as tamoxifen, or an SERD, or an aromatase inhibitor (AI). Non-limiting examples of an AI include non-steroidal inhibitors such as letrozole and anastrozole and irreversible steroidal inhibitors such as exemestane.

DETAILED DESCRIPTION OF MODES OF PRACTICING THE DISCLOSURE

Definitions of Terms as Used Herein

[0025] A gene expression "pattern" or "profile" or "signature" refers to the relative expression of one or more genes between two or more clinical outcomes, cancer outcomes, cancer recurrence and/or survival outcomes which is correlated with being able to distinguish between said outcomes. In some cases, the outcome is that of breast cancer.

[0026] A "gene" is a polynucleotide that encodes a discrete product, whether RNA or proteinaceous in nature. It is appreciated that more than one polynucleotide may be capable of encoding a discrete product. The term includes alleles and polymorphisms of a gene that encodes the same product, or a functionally associated (including gain, loss, or modulation of function) analog thereof, based upon chromosomal location and ability to recombine during normal mitosis.

[0027] The terms "correlate" or "correlation" or equivalents thereof refer to an association between expression of one

or more genes and a physiologic state of a cell to the exclusion of one or more other state as identified by use of the methods as described herein. A gene may be expressed at a higher or a lower level and still be correlated with one or more cancer state or outcome.

[0028] A "polynucleotide" is a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. This term refers only to the primary structure of the molecule. Thus, this term includes double- and single-stranded DNA and RNA. It also includes known types of modifications including labels known in the art, methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, and internucleotide modifications such as uncharged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), as well as unmodified forms of the polynucleotide.

[0029] The term "amplify" is used in the broad sense to mean creating an amplification product can be made enzymatically with DNA or RNA polymerases. "Amplification," as used herein, generally refers to the process of producing multiple copies of a desired sequence, particularly those of a sample. "Multiple copies" mean at least 2 copies. A "copy" does not necessarily mean perfect sequence complementarity or identity to the template sequence.

[0030] By corresponding is meant that a nucleic acid molecule shares a substantial amount of sequence identity with another nucleic acid molecule. Substantial amount means at least 95%, usually at least 98% and more usually at least 99%, and sequence identity is determined using the BLAST algorithm, as described in Altschul et al., *J. Mol. Biol.* 215:403-410 (1990) (using the published default setting, i.e. parameters w=4, t=17). Methods for amplifying mRNA are generally known in the art, and include reverse transcription PCR (RT-PCR) and those described in U.S. patent application Ser. No. 10/062,857 (filed on Oct. 25, 2001), as well as U.S. Provisional Patent Applications 60/298,847 (filed Jun. 15, 2001) and 60/257,801 (filed Dec. 22, 2000), all of which are hereby incorporated by reference in their entireties as if fully set forth. Another method which may be used is quantitative PCR (or Q-PCR). Alternatively, RNA may be directly labeled as the corresponding cDNA by methods known in the art.

[0031] A "microarray" is a linear or two-dimensional array of preferably discrete regions, each having a defined area, formed on the surface of a solid support such as, but not limited to, glass, plastic, or synthetic membrane. The density of the discrete regions on a microarray is determined by the total numbers of immobilized polynucleotides to be detected on the surface of a single solid phase support, preferably at least about 50/cm^2, more preferably at least about 100/cm^2, even more preferably at least about 500/cm^2, but preferably below about 1,000/cm^2. Preferably, the arrays contain less than about 500, about 1000, about 1500, about 2000, about 2500, or about 3000 immobilized polynucleotides in total. As used herein, a DNA microarray is an array of oligonucleotides or polynucleotides placed on a chip or other surfaces used to hybridize to amplified or cloned polynucleotides from a sample. Since the position of each particular group of primers in the array is known, the identities of a sample polynucleotides can be determined based on their binding to a particular position in the microarray.

[0032] Because the disclosure relies upon the identification of genes that are over- or under-expressed, one embodiment of the disclosure involves determining expression by hybridization of mRNA, or an amplified or cloned version thereof, of a sample cell to a polynucleotide that is unique to a particular gene sequence. Preferred polynucleotides of this type contain at least about 20, at least about 22, at least about 24, at least about 26, at least about 28, at least about 30, or at least about 32 consecutive basepairs of a gene sequence that is not found in other gene sequences. The term "about" as used in the previous sentence refers to an increase or decrease of 1 from the stated numerical value. Even more preferred are polynucleotides of at least or about 50, at least or about 100, at least about or 150, at least or about 200, at least or about 250, at least or about 300, at least or about 350, or at least or about 400 basepairs of a gene sequence that is not found in other gene sequences. The term "about" as used in the preceding sentence refers to an increase or decrease of 10% from the stated numerical value. Such polynucleotides may also be referred to as polynucleotide probes that are capable of hybridizing to sequences of the genes, or unique portions thereof, described herein. Preferably, the sequences are those of mRNA encoded by the genes, the corresponding cDNA to such mRNAs, and/or amplified versions of such sequences. In preferred embodiments of the disclosure, the polynucleotide probes are immobilized on an array, other devices, or in individual spots that localize the probes.

[0033] In another embodiment of the disclosure, all or part of a disclosed sequence may be amplified and detected by methods such as the polymerase chain reaction (PCR) and variations thereof, such as, but not limited to, quantitative PCR (Q-PCR), reverse transcription PCR (RT-PCR), and real-time PCR, optionally real-time RT-PCR. Such methods would utilize one or two primers that are complementary to portions of a disclosed sequence, where the primers are used to prime nucleic acid synthesis. The newly synthesized nucleic acids are optionally labeled and may be detected directly or by hybridization to a polynucleotide of the disclosure. The newly synthesized nucleic acids may be contacted with polynucleotides (containing sequences) of the disclosure under conditions which allow for their hybridization.

[0034] Alternatively, and in another embodiment of the disclosure, gene expression may be determined by analysis of expressed protein in a cell sample of interest by use of one or more antibodies specific for one or more epitopes of individual gene products (proteins) in said cell sample. Such antibodies are preferably labeled to permit their easy detection after binding to the gene product.

[0035] The term "label" refers to a composition capable of producing a detectable signal indicative of the presence of the labeled molecule. Suitable labels include radioisotopes, nucleotide chromophores, enzymes, substrates, fluorescent molecules, chemiluminescent moieties, magnetic particles, bioluminescent moieties, and the like. As such, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means.

[0036] The term "support" refers to conventional supports such as beads, particles, dipsticks, fibers, filters, membranes and silane or silicate supports such as glass slides.

[0037] As used herein, a "cancer tissue sample" or "cancer cell sample" refers to a cell containing sample of tissue isolated from an individual afflicted with the corresponding cancer. The sample may be from material removed via a surgical procedure, such as a biopsy. Such samples are primary isolates (in contrast to cultured cells) and may be collected by any suitable means recognized in the art. In some embodi-

ments, the "sample" may be collected by an non-invasive method, including, but not limited to, abrasion, fine needle aspiration.

[0038] A "breast tissue sample" or "breast cell sample" refers to a sample of breast tissue or fluid isolated from an individual suspected of being afflicted with, or at risk of developing, breast cancer. Such samples are primary isolates (in contrast to cultured cells) and may be collected by any non-invasive means, including, but not limited to, ductal lavage, fine needle aspiration, needle biopsy, the devices and methods described in U.S. Pat. No. 6,328,709, or any other suitable means recognized in the art. Alternatively, the "sample" may be collected by an invasive method, including, but not limited to, surgical biopsy.

[0039] "Expression" and "gene expression" include transcription and/or translation of nucleic acid material. Of course the term may also be limited, if so indicated, as referring only to the transcription of nucleic acids.

[0040] As used herein, the term "comprising" and its cognates are used in their inclusive sense; that is, equivalent to the term "including" and its corresponding cognates.

[0041] Conditions that "allow" an event to occur or conditions that are "suitable" for an event to occur, such as hybridization, strand extension, and the like, or "suitable" conditions are conditions that do not prevent such events from occurring. Thus, these conditions permit, enhance, facilitate, and/or are conducive to the event. Such conditions, known in the art and described herein, depend upon, for example, the nature of the nucleotide sequence, temperature, and buffer conditions. These conditions also depend on what event is desired, such as hybridization, cleavage, strand extension or transcription.

[0042] Sequence "mutation," as used herein, refers to any sequence alteration in the sequence of a gene disclosed herein interest in comparison to a reference sequence. A sequence mutation includes single nucleotide changes, or alterations of more than one nucleotide in a sequence, due to mechanisms such as substitution, deletion or insertion. Single nucleotide polymorphism (SNP) is also a sequence mutation as used herein. Because the present disclosure is based on the relative level of gene expression, mutations in non-coding regions of genes as disclosed herein may also be assayed in the practice of the disclosure.

[0043] "Detection" includes any means of detecting, including direct and indirect detection of gene expression and changes therein. For example, "detectably less" products may be observed directly or indirectly, and the term indicates any reduction (including the absence of detectable signal). Similarly, "detectably more" product means any increase, whether observed directly or indirectly.

[0044] Increases and decreases in expression of the disclosed sequences are defined in the following terms based upon percent or fold changes over expression in normal cells. Increases may be of 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, or 200% relative to expression levels in normal cells. Alternatively, fold increases may be of 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, or fold over expression levels in normal cells. Decreases may be of 10, 20, 30, 40, 50, 55, 60, 65, 70, 75, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 99 or 100% relative to expression levels in normal cells.

[0045] Unless defined otherwise all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs.

General

[0046] The gene expression patterns disclosed herein are predictive factors for therapeutic benefit in a switch in endocrine therapy. In some cases, the prediction is in node-negative breast cancer patients, such as ER+ node-negative patients as a non-limiting example.

[0047] To determine the expression levels of genes in the practice of the present disclosure, any method known in the art may be utilized. In some embodiments, expression based on detection of RNA which hybridizes to the genes identified and disclosed herein is used. This is readily performed by any RNA detection or amplification+detection method known or recognized as equivalent in the art such as, but not limited to, reverse transcription-PCR, the methods disclosed in U.S. patent application Ser. No. 10/062,857 (filed on Oct. 25, 2001) as well as U.S. Provisional Patent Application 60/298,847 (filed Jun. 15, 2001) and 60/257,801 (filed Dec. 22, 2000), and methods to detect the presence, or absence, of RNA stabilizing or destabilizing sequences.

[0048] Alternatively, expression based on detection of DNA status may be used. Detection of the DNA of an identified gene as methylated or deleted may be used for genes that have decreased expression. This may be readily performed by PCR based methods known in the art, including, but not limited to, Q-PCR. Conversely, detection of the DNA of an identified gene as amplified may be used for genes that have increased expression in correlation with a particular breast cancer outcome. This may be readily performed by PCR based, fluorescent in situ hybridization (FISH) and chromosome in situ hybridization (CISH) methods known in the art.

[0049] Expression based on detection of a presence, increase, or decrease in protein levels or activity may also be used. Detection may be performed by any immunohistochemistry (IHC) based, blood based (especially for secreted proteins), antibody (including autoantibodies against the protein) based, exfoliate cell (from the cancer) based, mass spectroscopy based, and image (including used of labeled ligand) based method known in the art and recognized as appropriate for the detection of the protein. Antibody and image based methods are additionally useful for the localization of tumors after determination of cancer by use of cells obtained by a non-invasive procedure (such as ductal lavage or fine needle aspiration), where the source of the cancerous cells is not known. A labeled antibody or ligand may be used to localize the carcinoma(s) within a patient.

[0050] One embodiment using a nucleic acid based assay to determine expression is by immobilization of one or more sequences of the genes identified herein on a solid support, including, but not limited to, a solid substrate as an array or to beads or bead based technology as known in the art. Alternatively, solution based expression assays known in the art may also be used.

[0051] The immobilized gene(s) may be in the form of polynucleotides that are unique or otherwise specific to the gene(s) such that the polynucleotide would be capable of hybridizing to a DNA or RNA corresponding to the gene(s). These polynucleotides may be the full length of the gene(s) or be short sequences of the genes (up to one nucleotide shorter than the full length sequence known in the art by deletion from the 5' or 3' end of the sequence) that are optionally minimally interrupted (such as by mismatches or inserted non-complementary basepairs) such that hybridization with a DNA or RNA corresponding to the gene(s) is not affected. In

some cases, the polynucleotides used are from the 3' end of the gene, such as within about 350, about 300, about 250, about 200, about 150, about 100, or about 50 nucleotides from the polyadenylation signal or polyadenylation site of a gene or expressed sequence. Polynucleotides containing mutations relative to the sequences of the disclosed genes may also be used so long as the presence of the mutations still allows hybridization to produce a detectable signal.

[0052] The immobilized gene(s) may be used to determine the state of nucleic acid samples prepared from sample cancer, or breast, cell(s) for which the outcome of the sample's subject (e.g. patient from whom the sample is obtained) is not known or for confirmation of an outcome that is already assigned to the sample's subject. Without limiting the disclosure, such a cell may be from a patient with ER+ breast cancer. The immobilized polynucleotide(s) need only be sufficient to specifically hybridize to the corresponding nucleic acid molecules derived from the sample under suitable conditions.

[0053] As will be appreciated by those skilled in the art, some of the corresponding sequences noted above include 3' poly A (or poly T on the complementary strand) stretches that do not contribute to the uniqueness of the disclosed sequences. The disclosure may thus be practiced with sequences lacking the 3' poly A (or poly T) stretches. The uniqueness of the disclosed sequences refers to the portions or entireties of the sequences which are found only in the disclosed gene's nucleic acids, including unique sequences found at the 3' untranslated portion of the genes. Preferred unique sequences for the practice of the disclosure are those which contribute to the consensus sequences for each of the three sets such that the unique sequences will be useful in detecting expression in a variety of individuals rather than being specific for a polymorphism present in some individuals. Alternatively, sequences unique to an individual or a subpopulation may be used. The preferred unique sequences are preferably of the lengths of polynucleotides of the disclosure as discussed herein.

[0054] To determine the (increased or decreased) expression levels of the above described sequences in the practice of the disclosure, any method known in the art may be utilized. In one embodiment of the disclosure, expression based on detection of RNA which hybridizes to polynucleotides containing the above described sequences is used. This is readily performed by any RNA detection or amplification+detection method known or recognized as equivalent in the art such as, but not limited to, reverse transcription-PCR (optionally real-time PCR), the methods disclosed in U.S. patent application Ser. No. 10/062,857 entitled "Nucleic Acid Amplification" filed on Oct. 25, 2001 as well as U.S. Provisional Patent Application 60/298,847 (filed Jun. 15, 2001) and 60/257,801 (filed Dec. 22, 2000), the methods disclosed in U.S. Pat. No. 6,291,170, and quantitative PCR. Methods to identify increased RNA stability (resulting in an observation of increased expression) or decreased RNA stability (resulting in an observation of decreased expression) may also be used. These methods include the detection of sequences that increase or decrease the stability of mRNAs containing the genes' sequences. These methods also include the detection of increased mRNA degradation.

[0055] In some embodiments of the disclosure, polynucleotides having sequences present in the 3' untranslated and/or non-coding regions of the above disclosed sequences are used to detect expression levels of the gene sequences in cancer, or breast, cells. Such polynucleotides may optionally contain sequences found in the 3' portions of the coding regions of the above disclosed sequences. Polynucleotides containing a combination of sequences from the coding and 3' non-coding regions preferably have the sequences arranged contiguously, with no intervening heterologous sequences.

[0056] Alternatively, the disclosure may be practiced with polynucleotides having sequences present in the 5' untranslated and/or non-coding regions of the gene sequences in cancer, or breast, cells to detect their levels of expression. Such polynucleotides may optionally contain sequences found in the 5' portions of the coding regions. Polynucleotides containing a combination of sequences from the coding and 5' non-coding regions preferably have the sequences arranged contiguously, with no intervening heterologous sequences. The disclosure may also be practiced with sequences present in the coding regions of the disclosed gene sequences.

[0057] Non-limiting polynucleotides contain sequences from 3' or 5' untranslated and/or non-coding regions of at least about 20, at least about 22, at least about 24, at least about 26, at least about 28, at least about 30, at least about 32, at least about 34, at least about 36, at least about 38, at least about 40, at least about 42, at least about 44, or at least about 46 consecutive nucleotides. The term "about" as used in the previous sentence refers to an increase or decrease of 1 from the stated numerical value. Even more preferred are polynucleotides containing sequences of at least or about 50, at least or about 100, at least about or 150, at least or about 200, at least or about 250, at least or about 300, at least or about 350, or at least or about 400 consecutive nucleotides. The term "about" as used in the preceding sentence refers to an increase or decrease of 10% from the stated numerical value.

[0058] Sequences from the 3' or 5' end of the above described coding regions as found in polynucleotides of the disclosure are of the same lengths as those described above, except that they would naturally be limited by the length of the coding region. The 3' end of a coding region may include sequences up to the 3' half of the coding region. Conversely, the 5' end of a coding region may include sequences up the 5' half of the coding region. Of course the above described sequences, or the coding regions and polynucleotides containing portions thereof, may be used in their entireties.

[0059] Polynucleotides combining the sequences from a 3' untranslated and/or non-coding region and the associated 3' end of the coding region may be at least or about 100, at least about or 150, at least or about 200, at least or about 250, at least or about 300, at least or about 350, or at least or about 400 consecutive nucleotides. Preferably, the polynucleotides used are from the 3' end of the gene, such as within about 350, about 300, about 250, about 200, about 150, about 100, or about 50 nucleotides from the polyadenylation signal or polyadenylation site of a gene or expressed sequence. Polynucleotides containing mutations relative to the sequences of the disclosed genes may also be used so long as the presence of the mutations still allows hybridization to produce a detectable signal.

[0060] In another embodiment of the disclosure, polynucleotides containing deletions of nucleotides from the 5' and/or 3' end of the above disclosed sequences may be used. The deletions are preferably of 1-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-125, 125-150, 150-175, or 175-200 nucleotides from the 5' and/or 3' end, although the extent of the deletions would naturally be limited by the length of the

disclosed sequences and the need to be able to use the polynucleotides for the detection of expression levels.

[0061] Other polynucleotides of the disclosure from the 3' end of the above disclosed sequences include those of primers and optional probes for quantitative PCR. In some embodiments, the primers and probes are those which amplify a region less than about 350, less than about 300, less than about 250, less than about 200, less than about 150, less than about 100, or less than about 50 nucleotides from the from the polyadenylation signal or polyadenylation site of a gene or expressed sequence.

[0062] In yet other embodiments of the disclosure, polynucleotides containing portions of the above disclosed sequences including the 3' end may be used. Such polynucleotides would contain at least or about 50, at least or about 100, at least about or 150, at least or about 200, at least or about 250, at least or about 300, at least or about 350, or at least or about 400 consecutive nucleotides from the 3' end of the disclosed sequences.

[0063] The disclosure also includes polynucleotides used to detect gene expression in breast cells. The polynucleotides may comprise a shorter polynucleotide consisting of sequences found in the above genes in combination with heterologous sequences not naturally found in combination with the sequences. Non-limiting examples include short sequences from cloning vectors or present in restriction fragments used to prepare labeled probes or primers as described herein.

HoxB13 and H/I

[0064] The methods of the disclosure based on the expression levels of HoxB13 in breast cancer cells of a subject may be used as a predictor of benefit in switching endocrine therapy after an initial course of endocrine therapy. In some embodiments, a two-gene ratio of HoxB13 expression to IL17BR expression (or HoxB13:IL17BR ratio) may be used in the manner reported by Ma et al. (*J. Clin. Oncol.*, 24:4611-9 (2006). In alternative embodiments, a two-gene ratio of HoxB13 expression to CHDH expression may be used.

[0065] In cases using HoxB13 expression alone or the HoxB13:IL17BR (H:I) ratio, a cutoff value may be used to define breast cancer cells as having either a "high" and a "low" value corresponding to the expression. In some embodiments, a cutoff may be used to define breast cancer cells as having either a "high H/I" and a "low H/I" value. As a non-limiting example, the value of 0.06 may be used in the manner of Ma et al. In other embodiments, the cutoff may be the average expression of HoxB13 in breast cancer cells from afflicted subjects. In additional possible embodiments, the cutoff may be the average value of H/I in breast cancer cells from afflicted subjects as determined by the average HoxB13 expression/the average IL17BR expression.

MGI

[0066] The genes disclosed below have roles in the cell cycle and reported peak expression as follows:

Gene	Peak of Expression	Role in Cell Cycle
BUB1B	G2/M	mitotic spindle assembly checkpoint
CENPA	G2/M	centromere assembly
NEK2	G2/M	centrosome duplication
RACGAP1	Not Determined	Initiation of cytokinesis
RRM2	S	DNA replication

[0067] The sequences of these genes have been previously reported and characterized in the field. For example, and on Sep. 6, 2007, the human BUB1B (also known as p21 protein-activated kinase 6 or PAK6) gene was identified by Unigene Hs.631699 and was characterized by 273 corresponding sequences. On Mar. 6, 2010, the same gene information was identified by UniGene Hs.513645 and characterized as corresponding to chromosome 15 at position 15q14 and as supported by 23 mRNA sequences and 549 EST sequences.

[0068] Also on Sep. 6, 2007, the human CENPA gene was identified by Hs.1594 (with 129 corresponding sequences). On Mar. 6, 2010, the same gene information was characterized as corresponding to chromosome 2 at 2p24-p21 and as supported by 10 mRNA sequences and 119 EST sequences.

[0069] Also on Sep. 6, 2007, the human NEK2 gene was identified by Hs.153704 (with 221 corresponding sequences). On Mar. 6, 2010, the same gene information was characterized as corresponding to chromosome 1 at 1q32.2-q41 and as supported by 17 mRNA sequences and 205 EST sequences.

[0070] Also on Sep. 6, 2007, the human RACGAP1 gene was identified by Hs.696319 (with 349 corresponding sequences). On Mar. 6, 2010, the same gene information was identified by UniGene Hs.505469 and characterized as corresponding to chromosome 12 at position 12q13.12 and as supported by 15 mRNA sequences and 398 EST sequences.

[0071] Also on Sep. 6, 2007, the human RRM2 gene was identified by Hs.226390 (with 1348 corresponding sequences). On Mar. 6, 2010, the same gene information was characterized as corresponding to chromosome 2 at 2p25-p24 and as supported by 25 mRNA sequences and 1328 EST sequences.

[0072] The mRNA and EST sequences corresponding to each of the above Unigene identifiers are hereby incorporated by reference as if fully set forth and may be used in the practice of the disclosure by the skilled person as deemed appropriate. Representative mRNA sequences for each of BUB1B, CENPA, NEK2, RACGAP1, and RRM2 have been disclosed in U.S. patent application Ser. No. 12/718,973, published as US 2011-0136680 A1 on Jun. 9, 2011. The disclosed sequences are non-limiting for the practice of the disclosed invention and are provided as evidence of the substantial knowledge in the field regarding sequences that are the disclosed genes. Additionally, the skilled person is fully capable of aligning any two or more of the known expressed sequences for each of these genes to identify an area of identity or conserved changes as a region that uniquely identifies each of these genes in comparison to other genes. Furthermore, the skilled person is fully capable of aligning any two or more of the known expressed sequences for each of these genes to identify an area unique to one or more of the of the expressed sequences as a region that uniquely identifies one known expressed sequence relative to at least one other expressed sequence. As a non-limiting example, a unique region may be in a variant of the expressed sequence for one of the known genes such that the region may be used to identify expression of the variant.

[0073] The sequences of the same genes have also been identified and characterized from other animal species. Thus the skilled person in the field is clearly aware of how to identify the disclosed genes relative to other animal genes. The skilled person may also optionally compare the known sequences of the disclosed genes from different animal sources to identify conserved regions and sequences unique to these genes relative to other genes.

Methods

[0074] As described herein, the disclosure includes the identity of genes, the expression of which can be used to provide prognostic information related to cancer. In particular, the expression levels of these genes may be used in relation to breast cancer. In some methods, the gene expression profile correlates with (and so are able to discriminate between) patients expected to benefit from a switch in endocrine therapy following an initial treatment with endocrine therapy for a period of time. In other embodiments, the disclosure includes a method to compare gene expression in a sample of cancer cells from a patient to the gene expression profile to determine the likely clinical or treatment outcome for the patient, or natural biological result, in the absence of a switch.

[0075] These embodiments of the disclosure may be advantageously used to meet an important unmet diagnostic need for the ability to predict whether a patient will likely benefit from a switch in treatment type. For example, a high H:I ratio value is strongly associated with response to a switch from first-line tamoxifen therapy for up to 5 years to letrozole therapy. The switch may occur anytime following the first-line therapy, such as immediately afterward, within three months after termination of first-line therapy, within six months after termination of first-line therapy, within nine months after termination of first-line therapy, within 12 months after termination of first-line therapy, within 18 months after termination of first-line therapy, or within 24 months (or more) after termination of first-line therapy.

[0076] So the disclosure includes a method to identify a patient, from a population of patients with ER+ breast cancer cells treated with a first endocrine therapy and cancer-free for a period of time, as belonging to a subpopulation of patients with a better prognosis if treated with an alternative endocrine therapy. In some cases, the breast cancer in the subject is node negative. The disclosure provides a non-subjective means for the identification of patients in the subpopulation.

[0077] The disclosure also includes a method of determining prognosis and/or survival outcome by assaying for the expression patterns disclosed herein. So where subjective interpretation may have been previously used to determine the prognosis and/or treatment of cancer patients, this disclosure provides objective gene expression patterns, which may used alone or in combination with subjective criteria to provide a more accurate assessment of patient outcomes, including survival and the recurrence of cancer.

[0078] In some embodiments, the disclosure provides a method to determine therapeutic treatment for a cancer patient by determining prognosis for said patient by assaying a sample of cancer cells from said patient for the expression levels described herein, and selecting a treatment for a patient with such gene expression. The assaying may include measuring or detecting or determining the expression level of the genes in any suitable means described herein or known to the skilled person. In many cases, the cancer is breast cancer, and the subject is a human patient. Additionally, the cancer cells may be those of a tumor and/or from a node negative (lymph nodes negative for cancer) or node positive (lymph nodes positive for cancer) subject.

[0079] The requisite level of expression may be that which is identified by the methods described herein for the genes used. Additionally, the assaying may include preparing RNA from the sample, optionally for use in PCR (polymerase chain reaction) or other analytical methodology as described herein. The PCR methodology is optionally RT-PCR (reverse transcription-PCR) or quantitative PCR, such as real-time RT-PCR. Alternatively, the assaying may be conducted by use of an array, such as a microarray as known in the relevant field. Optionally, the sample of cancer cells is dissected from tissue removed or obtained from said subject. As described herein, a variety of sample types may be used, including a formalin fixed paraffin embedded (FFPE) sample as a non-limiting example. And as described herein, the method may include assaying or determining the H:I ratio (ratio of HoxB13 and IL17BR expression levels) in the sample as disclosed herein.

[0080] By way of non-limiting example, all five genes of the MGI may be assayed and used to detect expression levels that correspond to a value that is "high risk" (which is above the cutoff) for MGI, or to detect expression levels that correspond to a value that is "low risk" (which is at or below the cutoff) for MGI, as disclosed herein. In some cases, the MGI cutoff threshold may be 0 (zero), such as where the measurements of expression levels are standardized to 0 (zero) with a standard deviation of 1. In alternative embodiments, the cutoff may be at or about 0.05, at or about 0.10, at or about 0.15, at or about 0.20, at or about 0.25, at or about −0.05, at or about −0.10, at or about −0.15, at or about −0.20, at or about −0.25, at or about −0.30, at or about −0.35, at or about −0.40, at or about −0.45, at or about −0.50, at or about −0.55, at or about −0.60, at or about −0.65, at or about −0.70, at or about −0.75, at or about −0.80, at or about −0.85, at or about −0.90, at or about −0.95, at or about −1.0, at or about −1.1, at or about −1.2, at or about −1.3, at or about −1.4, at or about −1.5, at or about −1.6, at or about −1.7, at or about −1.8, at or about −1.9, at or about −2.0 or lower. With respect to the H:I ratio, its determination maybe made as described in Ma et al., *Cancer Cell,* 5:607-16 (2004) and Ma et al. (2006) as referenced herein. For example, a value of 0.06 may be used to determine whether a sample has a "high risk" (>0.06) or "low risk" (≤0.06) H:I ratio.

[0081] So using a threshold, or cutoff, of 0 (zero) as a non-limiting example for MGI with all five genes, the disclosed methods provide two possible assay outcomes for a given sample: "high risk MGI" corresponding to a value above 0 (zero) and "low risk MGI" corresponding to a value≤0. A "high risk MGI" is indicative of a "high risk" cancer, including breast cancer that is analogous to that of a Grade III tumor as defined by methodologies and standards known in the field. A "low risk MGI" is indicative of a "low risk" cancer, including breast cancer, that is analogous to that of a Grade I tumor as defined by methodologies and standards known in the field.

[0082] In one embodiment of the disclosure, a method is provided for determining the risk or likelihood of cancer recurrence in a subject after treatment for breast cancer, such as removal of the cancer by surgery. The method may comprise i) preparing cDNA from nucleic acids in a sample of ER+ breast cancer cells removed from a breast cancer

afflicted subject; ii) determining the expression levels of the seven genes in the disclosed Breast Cancer Index (BCI) from said cDNA to determine a BCI value; iii) identifying the subject as having been treated with endocrine therapy for a period of time without cancer recurrence; and iv) classifying the cancer as likely to recur due to a high risk BCI value. In some cases, the subject has been treated with endocrine therapy for about 5 years or more, about 4 years or more, about 3 years or more, about 2 years or more, or about 1 year or more.

[0083] In another embodiment of the disclosure, a method is provided for determining the likelihood of a beneficial switch in endocrine therapy as treatment for breast cancer. The method may comprise i) preparing cDNA from nucleic acids in a sample of ER+ breast cancer cells removed from a breast cancer afflicted subject; ii) determining the expression level of the HoxB13 gene from said cDNA; iii) optionally identifying the subject as having undergone surgical removal of the breast cancer; iv) identifying the subject as having been treated with a first endocrine therapy for a period of time without cancer recurrence; and v) classifying the subject as expected to benefit from treatment with a different second endocrine therapy after cessation of the first endocrine therapy, wherein said classifying is based upon an elevated expression level of HoxB13. In some cases, the elevated expression level of HoxB13 is determined as part of the H/I value and the classifying is based upon a high H/I value. In some cases, the subject has been treated with endocrine therapy for about 5 years or more, about 4 years or more, about 3 years or more, about 2 years or more, or about 1 year or more.

[0084] In additional embodiments, the disclosure provides a method to treat a patient that has undergone a first endocrine therapy as described above. The method may comprise the above determining the likelihood of a beneficial switch in endocrine therapy as treatment for breast cancer followed by treating the patient with a second endocrine therapy after ending treatment with the first endocrine therapy.

[0085] As non-limiting examples, the first endocrine therapy may be treatment with an SERM or an SERD and the second endocrine therapy may be treatment with an aromatase inhibitor. Alternatively, the first endocrine therapy may be treatment with an aromatase inhibitor and the second endocrine therapy may be treatment with an SERM or an SERD. Embodiments include tamoxifen as the first endocrine therapy followed by letrozole as the second, or letrozole as the first endocrine therapy followed by taxmoxifen as the second.

[0086] The disclosure further includes a method of determining a prognostic factor or predictor of clinical responsiveness in pre-menopausal women and post-menopausal women. Post-menopausal women may be defined as those that are ≥50 years old while pre-menopausal women may be defined as those who are less than 50 years old.

[0087] The ability to discriminate is conferred by the identification of expression of the individual genes as relevant and not by the form of the assay used to determine the actual level of expression. An assay may utilize any identifying feature of an identified individual gene as disclosed herein as long as the assay reflects, quantitatively or qualitatively, expression of the gene in the "transcriptome" (the transcribed fraction of genes in a genome) or the "proteome" (the translated fraction of expressed genes in a genome). Identifying features include, but are not limited to, unique nucleic acid sequences used to encode (DNA), or express (RNA), said gene or epitopes specific to, or activities of, a protein encoded by said gene. All that is required is the identity of the gene(s) necessary to discriminate between cancer outcomes and an appropriate cell containing sample for use in an expression assay.

[0088] Similarly, the nature of the cell containing sample is not limiting, as fresh tissue, freshly frozen tissue, and fixed tissue, such as formalin-fixed paraffin-embedded (FFPE) tissues, may be used in the disclosed methods.

[0089] In one embodiment, the disclosure provides for the identification of the gene expression patterns by analyzing global, or near global, gene expression from single cells or homogenous cell populations which have been dissected away from, or otherwise isolated or purified from, contaminating cells beyond that possible by a simple biopsy. Because the expression of numerous genes fluctuate between cells from different patients as well as between cells from the same patient sample, the levels of gene expression may be determined in correspondence to one or more "control" or "normalization" genes, the expression(s) of which are relatively constant in the cells of a patient or between patients.

[0090] In another aspect, the disclosure includes physical and methodological means for detecting the expression of gene(s) identified by the models generated by individual expression patterns. These means may be directed to assaying one or more aspect of the DNA template(s) underlying the expression of the gene(s), of the RNA used as an intermediate to express the gene(s), or of the proteinaceous product expressed by the gene(s).

[0091] One advantage provided by the disclosure is that contaminating, non-cancer cells (such as infiltrating lymphocytes or other immune system cells) are not present to possibly affect the genes identified or the subsequent analysis of gene expression to identify the cancer recurrence and/or survival outcomes of patients. Such contamination is present where a biopsy containing many cell types is used to assay gene expression profiles.

[0092] While the present disclosure is described mainly in the context of human cancer, such as breast cancer, it may be practiced in the context of cancer of any animal. Preferred animals for the application of the present disclosure are mammals, particularly those important to agricultural applications (such as, but not limited to, cattle, sheep, horses, and other "farm animals"), animal models of cancer, and animals for human companionship (such as, but not limited to, dogs and cats).

[0093] The methods provided by the disclosure may also be automated in whole or in part.

Kits

[0094] The materials for use in the methods of the present disclosure are ideally suited for preparation of kits produced in accordance with well known procedures. The disclosure thus provides kits comprising agents for the detection of expression of the disclosed genes for grading tumors or determining cancer outcomes. Such kits optionally comprise the agent with an identifying description or label or instructions relating to their use in the methods of the present disclosure. Such a kit may comprise containers, each with one or more of the various reagents (typically in concentrated form) utilized in the methods, including, for example, pre-fabricated microarrays, buffers, the appropriate nucleotide triphosphates (e.g., dATP, dCTP, dGTP and dTTP; or rATP, rCTP, rGTP and UTP), reverse transcriptase, DNA polymerase, RNA polymerase, and one or more primer complexes of the

present disclosure (e.g., appropriate length poly(T) or random primers linked to a promoter reactive with the RNA polymerase). A set of instructions will also typically be included.

[0095] Having now generally provided the disclosure, the same will be more readily understood through reference to the following examples which are provided by way of illustration, and are not intended to be limiting of the disclosure, unless specified.

EXAMPLES

Example I

General

Patients and Tumor Samples

[0096] Samples from the NCIC CTG MA.17 cohort (see Goss et al., *J. Clin. Oncol.,* 26(12):1948-1955, 2008) were used. 100 cases with 200 controls were used. The 100 cases included 61 cases of distant cancer recurrence; 17 cases of local cancer recurrence; 5 cases of regional cancer recurrence; 16 cases of contralateral recurrences; and 1 unknown case. Of these, the contralateral and unknown cases were excluded.

[0097] Clinical follow-up data were available for the samples used, which were formalin-fixed paraffin-embedded (FFPE) tumor blocks from the time of diagnosis. Odds ratios were calculated with analysis of BCI, H:I, HoxB13 and MGI as continuous and categorical variables. Multivariate analysis also included age, tumor grade and treatment in the analysis. Treatment interaction: age and tumor grade were also included in the analysis. P-values were calculated for the interaction term.

[0098] Table 1 summarizes characteristics for the cases and controls (N=249).

Real-Time RT-PCR Assays for H/I and MGI

[0099] Primer and probe sequences for HOXB13 and IL17BR, as well as control genes ESR1, PGR, CHDH, ACTB, HMBS, SDHA and UBC, were used as described previously (Ma et al., supra). Primer and probe sequences for the five molecular grade genes (BUB1B, CENPA, NEK2, RACGAP1 and RRM2) as well as ERBB2 (HER2) were prepared using Primer Express (ABI).

[0100] Sections of each FFPE sample were used for RNA extraction. Gross macro-dissection was used to enrich for tumor content. RNA extraction, reverse transcription, and TaqMan RT-PCR using the ABI 7900HT instrument (Applied Biosystem, Inc) were performed as described before (Ma et al., id.). The cycling threshold numbers (CTs) were normalized to the mean CT of four reference genes (ACTB, HMBS, SDHA and UBC). The use of these genes is supported by the previous reports regarding these genes and representative sequences of each of these genes known to the skilled person. Normalized CTs were taken to represent relative gene expression levels.

Calculation of H/I and MGI

[0101] Generally, and with respect to MGI, it is preferred that the expression levels of the disclosed genes are combined to form a single index that serves as a strong prognostic factor and predictor of clinical outcome(s). The index is a summation of the expression levels of the genes used and uses coefficients determined from principle component analysis to combine cases of more than one disclosed gene into a single index. The coefficients are determined by factors such as the standard deviation of each gene's expression levels across a representative dataset, and the expression value for each gene in each sample. The representative dataset is quality controlled based upon the average expression values for reference gene(s) as disclosed herein.

TABLE 1

Patient and tumor characteristics

Factor Description		MA-17 overall (n = 5157)	Case-control study (n = 249)	Cases (n = 83)	Controls (n = 166)	P-value
Age	<50		9 (4%)	4 (5%)	5 (3%)	0.64
—	>=50, <60		83 (33%)	27 (32%)	56 (34%)	
	>=60, <70		82 (33%)	24 (29%)	58 (35%)	
	>=70		75 (30%)	28 (34%)	47 (28%)	
Tumor Grade	1		26 (10%)	6 (7%)	20 (12%)	0.28
	2		166 (67%)	54 (65%)	112 (67.5%)	
—	3		57 (23%)	23 (27%)	34 (20.5%)	
Tumor Type	Ductal		218 (88%)	71 (86%)	147 (88.6%)	0.63
—	Lobular		31 (12%)	12 (14%)	19 (11.4%)	
N Stage	N0		94 (38%)	31 (37%)	63 (38%)	0.45
	N1		138 (55%)	44 (53%)	94 (57%)	
	N2, N3, NX		17 (7%)	8 (10%)	9 (5%)	
T Stage	T1		110 (44%)	37 (45%)	73 (44%)	0.58
	T2		111 (45%)	35 (42%)	76 (46%)	
	T3		21 (8%)	7 (8%)	14 (8%)	
	T4, TX		7 (3%)	4 (5%)	3 (2%)	
Prior Chemo Treatment	No		148 (59%)	49 (59%)	99 (60%)	0.96
	Yes		101 (41%)	34 (41%)	67 (40%)	
Prior Radiation	No		150 (60%)	49 (59%)	101 (61%)	0.89
Treatment	Yes		99 (40%)	34 (41%)	65 (39%)	
Treatment Arm	Letrozole		122 (49%)	31 (37%)	91 (55%)	0.01
	Placebo		127 (51%)	52 (63%)	75 (45%)	

[0102] Stated differently, and with respect to MGI, normalized expression levels for the five genes from microarrays or RT-PCR were standardized to mean of 0 and standard deviation of 1 across samples within each dataset and then combined into a single index per sample via principle component analysis (PCA) using the first principle component. Standardization of the primary expression data within each dataset was necessary to account for the different platforms (microarrays and RT-PCR) and sample types (frozen and FFPE). As a result, and following scaling parameters, a formula for the summation of expression values that defines the index is generated. The precision of the scaling parameters can then be tested based on the means, standard errors, and standard deviations (with confidence intervals) of the expression levels of the genes across the data set. Therefore, generation of the formula for the index is dependent upon the dataset, reference gene, and genes of the MGI.

[0103] The HOXB13:IL17BR ratio was calculated as the difference in standardized expression levels between HOXB13 and IL17BR as described previously (Ma et al., id.). The means and standard deviations for HOXB13 and IL17BR used for standardizing the Table 1 cohort may be derived from an analysis of 190 FFPE tissue sections from a separate population-based cohort of estrogen receptor-positive lymph node-negative breast cancer patients.

[0104] For MGI, obviously abnormal raw C_T values were removed prior to averaging the values over duplicates for each gene and each sample. The averaged raw C_T value for each gene was then normalized by the averaged C_T value of four reference genes (ACTB, HMBS, SDHA, and UBC). The normalized expression levels (ΔC_T) for the five genes were combined into a single index per sample, which can be compared to a pre-determined cutoff value, such as 0, where high MGI is above the cutoff and low MGI is below the cutoff.

Continuos BCI

[0105] A continuous risk model was built by combining H:I and MGI as continuous variables. The linearity of these two variables were checked by fitting a Cox proportional hazard regression model with restricted cubic splines, and H:I demonstrated significant non-linearity. A polynomial function of H:I was used to approximate the restricted models using Akaike Information Criterion. The resulting predictor from the final Cox regression model was then re-scaled into the range of 0 to 10, which is referred to as the BCI.

[0106] The BCI is further categorized into three levels: low risk, BCI<5; intermediate risk, 5≤BCI<6.4; high risk, BCI>6.4. These cut-offs were chosen such that the resulting proportions of low, intermediate, and high risk groups were similar to those formed by the three categorical combination groups of H:I and MGI.

Cut-Points and Statistical Analyes

[0107] H/I CUT-POINT: The cutpoint of 0.06 for the HOXB13:IL17BR ratio, previously defined to stratify patients treated with adjuvant tamoxifen into low and high risk of recurrence, may be used in this study.

[0108] MGI CUT-POINT: The calculation and the cutpoint for MGI were defined without using any clinical outcome data and instead was a natural cutpoint. Initial analysis of MGI in the Uppsala cohort indicated good discrimination of grade 1 and grade 3 tumors using the mean (O) as cutpoint, and model-based clustering of MGI also indicated a bimodal distribution with a natural cutpoint around 0. This cutpoint was further supported by receiver operating characteristic (ROC) analysis.

[0109] STATISTICAL ANALYSES: Kaplan-Meier analysis with logrank test and Cox proportional hazards regression were performed to assess the association of gene expression indexes with clinical outcome. Multivariate Cox regression models were performed to assess the prognostic capacity of gene expression indexes after adjusting for known prognostic factors.

[0110] Proportional hazards (PH) assumption was checked by scaled Schoenfeld residuals; variables violating PH assumption were adjusted for in the model through stratification. To account for the case-cohort design of the Table 1 cohort, we used weighted Kaplan-Meier analysis and Cox regression models with modifications to handle case-cohort designs (see [19,20] as implemented in the survey package in R (www.r-project.org). To test for interaction between dichotomized MGI and the H:I ratio in Cox regression models, the Wald statistic was used in the Table 1 cohort and likelihood ratio test was used in the last cohort.

[0111] Correlations of continuous variables with categorical factors were examined using non-parametric two-sample Wilcoxon test or Kruskal-Wallis test for factors with more than two levels.

[0112] All statistical analyses were performed in the R statistical environment. All significance test were two-sided, and $p<0.05$ was considered significant.

Example II

Prognostic Performance

[0113] Table 2 shows the distribution of the cases and controls to the continuous BCI risk groups.

TABLE 2

BCI group, (%)	Cases (n = 83)	Controls (n = 166)
Low	43.4%	57.8%
Intermediate	22.9%	18.1%
High	33.7%	24.1%

[0114] Table 3 shows the univariate analysis in relation to cancer recurrence in MA.17 subjects.

TABLE 3

Univariate analysis in relation to cancer recurrence

	Odds Ratio (95% CI)	P-value
Treatment (Placebo vs Letrozole)	2.02 (1.17-3.47)	0.01
Tumor Grade		0.28
II vs. I	1.73 (0.61-4.88)	0.30
III vs. I	2.53 (0.81-7.90)	0.11
Analysis of BCI		
BCI	2.38 (1.21-4.69)	0.01
BCI, High vs Low	1.87 (1.00-3.50)	0.05
Analysis of components of BCI		
HoxB13	1.34 (1.05-1.70)	0.02
HoxB13, High vs Low	2.17 (1.27-3.69)	0.004
H:I	2.52 (1.08-5.85)	0.03
H:I, High vs Low	1.68 (1.00-2.81)	0.049

TABLE 3-continued

Univariate analysis in relation to cancer recurrence		
	Odds Ratio (95% CI)	P-value
MGI	1.83 (0.93-3.58)	0.08
MGI, High vs Low	1.49 (0.87-2.56)	0.15

[0115] Table 4 shows multivariate analysis in relation to cancer recurrence

TABLE 4

Multivariate analysis in relation to cancer recurrence		
	Odds Ratio (95% CI)	P-value
Analysis with BCI		
BCI	2.37 (1.08-5.22)	0.03
BCI, High vs Low	1.87 (0.88-3.95)	0.10
Analysis with components of BCI		
HoxB13	1.35 (1.05-1.74)	0.02
HoxB13, High vs Low	2.32 (1.32-4.10)	0.004
H:I	2.55(1.03-6.32)	0.04
H:I, High vs Low	1.71 (0.98-2.97)	0.06
MGI	1.61 (0.73-3.54)	0.24
MGI, High vs Low	1.37 (0.73-2.55)	0.33

[0116] As shown by the above, BCI is prognostic of late cancer recurrences in ER+ patients following 5 years of tamoxifen treatment. HoxB13 expression is also prognostic of late cancer recurrences in ER+ patients following 5 years of tamoxifen treatment.

Example III

Biomarker and Treatment Interaction

[0117] Table 5 shows interactions between the gene expression analyzed and treatment.

TABLE 5

Biomarker and Treatment Interaction	
	P-value
HoxB13	0.047
H:I	0.97
MGI	0.06
BCI	0.42
BCI High vs Intermediate + Low	0.08

[0118] Table 6 shows the distribution of HoxB13 gene expression in relation to treatments used. HoxB13 expression at diagnosis predicts patient benefit from extended endocrine therapy with letrozole after 5 years of adjuvant tamoxifen therapy.

TABLE 6

	Letrozole		Placebo		
	Controls	Cases	Controls	Cases	P-value
Low HoxB13	48	14 (23%)	46	16 (26%)	0.83
High HoxB13	43	17 (28%)	29	36 (55%)	0.004

BIBLIOGRAPHY

[0119] 1. Ma et al., *Cancer Cell,* 5:607-16 (2004)
[0120] 2. Ma et al., *J. Clin. Oncol.,* 24:4611-9 (2006)
[0121] 3. Goetz et al., *Clin. Cancer Res.,* 12:2080-7 (2006)
[0122] 4. Jerevall et al., *Breast Cancer Res. Treat* (2007)
[0123] 5. Jansen et al., *J. Clin. Oncol.* 25:662-8 (2007)
[0124] 6. Cianfrocca et al., *Oncologist,* 9:606-16 (2004)
[0125] 7. Sotiriou et al., *J. Natl. Cancer Inst.,* 98:262-72 (2006)
[0126] 8. van 't Veer et al., *Nature,* 415:530-6 (2002)
[0127] 9. Paik et al., *N. Engl. J. Med.,* 351:2817-26 (2004)
[0128] 10. Desmedt et al., *Cell Cycle,* 5:2198-202 (2006)
[0129] 11. Loi et al. *J. Clin. Oncol.,* 25:1239-46 (2007)
[0130] 12. Sotiriou et al., *Nat. Rev. Cancer,* 7:545-53 (2007)
[0131] 13. Miller et al., *Proc. Natl. Acad. Sci. USA,* 102: 13550-5 (2005)
[0132] 14. Pawitan et al., *Breast Cancer Res.* 7:R953-64 (2005)
[0133] 15. Rundle et al., *Cancer Epidemiol Biomarkers Prev.,* 14:1899-907 (2005)
[0134] 16. Ma et al., *Proc. Natl. Acad. Sci. USA,* 100:5974-9 (2003)
[0135] 17. Whitfield et al., *Mol. Biol. Cell,* 13:1977-2000 (2002)
[0136] 18. Hirose et al., *J. Biol. Chem.,* 276:5821-5828 (2001)
[0137] 19. Goldhirsch et al., *Ann. Oncol.,* 16:1569-83 (2005)

[0138] All references cited herein, including patents, patent applications, and publications, are hereby incorporated by reference in their entireties, whether previously specifically incorporated or not.

[0139] Having now fully described the inventive subject matter, it will be appreciated by those skilled in the art that the same can be performed within a wide range of equivalent parameters, concentrations, and conditions without departing from the spirit and scope of the disclosure and without undue experimentation.

[0140] While this disclosure has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications. This application is intended to cover any variations, uses, or adaptations of the disclosure following, in general, the principles of the disclosure and including such departures from the present disclosure as come within known or customary practice within the art to which the disclosure pertains and as may be applied to the essential features hereinbefore set forth.

What is claimed is:

1. A method comprising:

preparing cDNA from nucleic acids in a sample of ER+ breast cancer cells from a breast cancer afflicted subject,

determining the expression level of the HoxB13 gene from said cDNA,

identifying the subject as having undergone removal of the breast cancer and as having been treated with a first endocrine therapy for a period of time without cancer recurrence, and

classifying the subject as expected to benefit from treatment with the same first endocrine therapy or a different second endocrine therapy after cessation of the first endocrine therapy, wherein said classifying is based upon an elevated expression level of HoxB13.

2. A method comprising:
preparing cDNA from nucleic acids in a sample of ER+ breast cancer cells from a breast cancer afflicted subject,
determining the expression levels of the seven genes in the disclosed Breast Cancer Index (BCI) from said cDNA, wherein said genes are HoxB13, IL17BR, Bub1B, CENPA, NEK2, RACGAP1, and RRM2, to determine a BCI value,
identifying the subject as having undergone removal of the breast cancer and as having been treated with endocrine therapy for a period of time without cancer recurrence, and
classifying the cancer as likely to recur due to a high risk BCI value.

3. The method of claim **1** wherein said nucleic acids are mRNA from said sample.

4. The method of claim **3** wherein said RNA is used for PCR (polymerase chain reaction).

5. The method of claim **1** wherein said determining comprises using an array.

6. The method of claim **1** wherein said sample is dissected from tissue removed from said subject.

7. The method of claim **4** wherein said PCR is RT-PCR (reverse transcription-PCR), optionally real time RT-PCR.

8. The method of claim **1** wherein said sample is a formalin fixed paraffin embedded (FFPE) sample.

9. The method of claim **1**, further comprising treating said subject with letrozole after cessation of tamoxifen treatment.

10. The method of claim **1** further comprising assaying for the H:I ratio in said sample and wherein said ratio is used to indicate an elevated expression level of HoxB13.

11. The method of claim **1** wherein said cancer is ductal carcinoma in situ (DCIS) and said cancer recurrence comprises local recurrence.

12. The method of claim **1** wherein the first endocrine therapy is treatment with tamoxifen and the second endocrine therapy is treatment with letrozole.

13. A method to treat a cancer patient, said method comprising
preparing cDNA from nucleic acids in a sample of ER+ breast cancer cells from a breast cancer afflicted subject,
determining the expression level of the HoxB13 gene from said cDNA,
identifying the subject as having undergone removal of the breast cancer and as having been treated with a first endocrine therapy for a period of time without cancer recurrence,
classifying the subject as expected to benefit from treatment with the same first endocrine therapy or a different second endocrine therapy after cessation of the first endocrine therapy, wherein said classifying is based upon an elevated expression level of HoxB13, and
treating the patient with the second endocrine therapy after ending treatment with the first endocrine therapy.

14. The method of claim **13**, further comprising assaying for the H:I ratio in said sample, wherein said ratio is used to indicate an elevated expression level of HoxB13.

15. The method of claim **13** wherein the first endocrine therapy is treatment with tamoxifen and the second endocrine therapy is treatment with letrozole.

* * * * *