

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 June 2009 (11.06.2009)

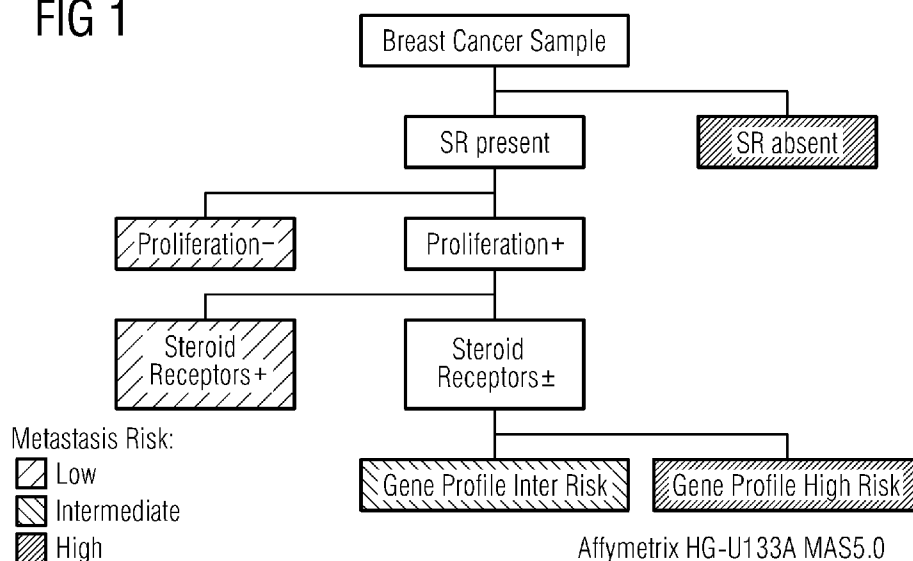
PCT

(10) International Publication Number
WO 2009/071655 A2

- (51) International Patent Classification: **Not classified**
- (21) International Application Number: PCT/EP2008/066868
- (22) International Filing Date: 5 December 2008 (05.12.2008)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/992,735 6 December 2007 (06.12.2007) US
- (71) Applicant (for all designated States except US): **SIEMENS HEALTHCARE DIAGNOSTICS INC.** [US/US]; 511 Benedict Avenue, Tarrytown, NY 10591-5098 (US).
- (72) Inventor; and
- (75) Inventor/Applicant (for US only): **GEHRMANN, Mathias** [DE/DE]; Schlebuscher Heide 13, 51375 Leverkusen (DE).
- (74) Agent: **MAIER, Daniel**; Postfach 22 16 34, 80506 München (DE).
- (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

(54) Title: METHODS FOR BREAST CANCER PROGNOSIS

FIG 1



(57) Abstract: The present invention provides a method to assess the risk of recurrence of a node positive and hormone receptor positive breast cancer patient when treated with tamoxifen. The method is based on quantitative determination of RNA species isolated from the tumor in order to obtain expression values and subsequent bioinformatic analysis of said determined expression values. First, steroid hormone receptor (SR) status is determined. Then, one or more RNA species coding for proliferation associated genes are determined. Then a combination of genes indicating the presence of T-lymphocyte infiltration is used to classify a tumor as low, intermediate or high risk tumor.

METHODS FOR BREAST CANCER PROGNOSIS

Technical Field

5 The present invention relates to methods, kits and systems for the prognosis of the disease outcome of breast cancer in node positive breast cancer patients treated with tamoxifen. More specific, the present invention relates to the prognosis of breast cancer based on measurements of the expression
10 levels of marker genes in tumor samples of breast cancer patients.

Background of the Invention

15 Breast cancer is one of the leading causes of cancer death in women in western countries. More specifically breast cancer claims the lives of approximately 40,000 women and is diagnosed in approximately 200,000 women annually in the United States alone. Over the last few decades, adjuvant
20 systemic therapy has led to markedly improved survival in early breast cancer (EBCTCG, 1998 a+b). This clinical experience has led to consensus recommendations offering adjuvant systemic therapy for the vast majority of breast cancer patients (Goldhirsch et al., 2003). In breast cancer a
25 multitude of treatment options are available which can be applied in addition to the routinely performed surgical removal of the tumor and subsequent radiation of the tumor bed. Three main and conceptually different strategies are endocrine treatment, chemotherapy and treatment with targeted
30 therapies. Prerequisite for treatment with endocrine agents is expression of hormone receptors in the tumor tissue i.e. either estrogen receptor, progesterone receptor or both. Several endocrine agents with different mode of action and differences in disease outcome when tested in large patient

cohorts are available. Tamoxifen has been the mainstay of endocrine treatment for the last three decades. Large clinical trial showed that tamoxifen significantly reduced the risk of tumor recurrence. An additional treatment option is based on aromatase inhibitors which belong to a new endocrine drug class. In contrast to tamoxifen which is a competitive inhibitor of estrogen binding aromatase inhibitors block the production of estrogen itself thereby reducing the growth stimulus for estrogen receptor positive tumor cells. Still, some patients experience a relapse despite endocrine or anti-hormonal treatment and in particular these patients might benefit from additional therapeutic drugs. Chemotherapy with anthracyclines, taxanes and other agents have been shown to be efficient in reducing disease recurrence in estrogen receptor positive as well as estrogen receptor negative patients. The NSABP-20 study compared tamoxifen alone against tamoxifen plus chemotherapy in node negative estrogen receptor positive patients and showed that the combined treatment was more effective than tamoxifen alone. Recently, a systemically administered antibody directed against the HER2/neu antigen on the surface of tumor cells have been shown to reduce the risk of recurrence several fold in a patients with Her2neu over expressing tumors.

Yet, most if not all of the different drug treatments have numerous potential adverse effects which can severely impair patients' quality of life (Shapiro and Recht, 2001; Ganz et al., 2002). This makes it mandatory to select the treatment strategy on the basis of a careful risk assessment for the individual patient to avoid over- as well as under treatment.

Positive lymph node status is a strong adverse prognostic factor in breast cancer and women with node positive disease are generally considered for chemotherapy. However, treatment

guidelines for women with node positive and hormone receptor positive breast cancer are not well defined by St Gallen or NCI criteria. Neither is chemotherapy definitely recommended nor is a clear advice given regarding the preferred type of
5 endocrine treatment.

Gene expression profiling has greatly extended the possibility to analyze the underlying biology of the heterogeneous nature of breast cancer.

10

Another potential prognostic factor, which was largely unattended in gene expression studies, is the immune system. Tumor infiltration by lymphocytes has long been suggested to influence clinical outcome. In particular, medullary breast
15 cancer (MBC), which is characterized by prominent lymphocytic infiltrates, is linked with relatively good outcome despite estrogen receptor negativity and poor histological grade.

US 2004/0229297-A1, filed 27 January 2004, discloses a method
20 for the prognosis of the breast cancer in a patient said method comprising detecting in human tumor tissues the infiltration of certain immune cells. High infiltration of the tumor with immune cells was associated with poor cancer prognosis. The method, however, does not use information on
25 the nodal status and does not rely on information on the rate of proliferation of the tumor.

In regard to the continuing need for materials and methods useful in making clinical decisions on adjuvant therapy, the
30 present invention fulfills the need for advanced methods for the prognosis of breast cancer on the basis of readily accessible clinical and experimental data.

Definitions

5

Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

10 The term "prediction" generally relates to the prediction of an event occurring in the future, e.g. prediction of survival, of prediction of development of metastasis or the like.

15 The term "prognosis" generally relates to the course of a disease under a given treatment.

The term "prediction of prognosis of a patient receiving a treatment" as used herein relates to the likelihood that a
20 patient will have either favourable or unfavourable clinical course under a given therapy. Especially, the term "prediction of prognosis", as used herein, relates to an individual assessment of the malignancy of a tumor, or to the expected survival rate (DFS, disease free survival) of a
25 patient, if the tumor is treated with a given therapy. In contrast thereto, the term "prognosis" relates to an individual assesment of the malignancy of a tumor, or to the expected survival rate (DFS, disease free survival) of a patient, if the tumor remains untreated.

30

"Classification", within the meaning of the invention, shall be understood as being the process of classifying objects into one of multiple classes, said classes being defined by certain properties of the objects therein. In methods of the

invention, tumors or patients are classified into one of several "risk classes", said risk classes being defined e.g. by the likelihood of recurrence of cancer in said patients after surgery and treatment with a given drug. A "risk" is the likelihood of the occurrence of a certain event, usually within a given period of time.

"Sequential classification steps", within the meaning of the invention, shall be understood as being multiple classification steps in a sequence of classification steps, wherein the subsequent classification step further classifies a class which was previously classified by a preceding classification step. Sequential classification steps can be sequential classification steps in a "branched" classification scheme, such as the decision tree shown in Figure 1-3. Classification by sequential classification steps is effectively be implemented as a combination of the individual comparison steps, combined by the logical AND operator.

The terms "cancer" and "cancerous" refer to or describe the physiological condition in mammals that is typically characterized by unregulated cell growth. The term "cancer" as used herein includes carcinomas, (e.g., carcinoma in situ, invasive carcinoma, metastatic carcinoma) and pre-malignant conditions, adenomas, blood cell neoplasms and neomorphic changes independent of their histological origin. The term "breast cancer" relates to any form of cancer of the male or female breast, regardless of its histological origin. The term "cancer" is not limited to any stage, grade, histomorphological feature, invasiveness, aggressiveness or malignancy of an affected tissue or cell aggregation. In particular stage 0 cancer, stage I cancer, stage II cancer, stage III cancer, stage IV cancer, grade I cancer, grade II cancer, grade III cancer, malignant cancer, primary

carcinomas, and all other types of cancers, malignancies and transformations specially associated with gynecologic cancer are included. The terms "neoplastic disease" or "cancer" are not limited to any tissue or cell type they also include
5 primary, secondary or metastatic lesions of cancer patients, and also comprise lymph nodes affected by cancer cells or minimal residual disease cells either locally deposited or freely floating throughout the patients body.

10 The term "tumor" as used herein, refers to all abnormal masses of tissue preferably exhibiting neoplastic cell growth and proliferation or impaired cell death mechanisms, whether malignant or benign, and all pre-cancerous and cancerous cells and tissues.

15 The term "treatment", "therapy modality", "therapy mode", "regimen" as well as "therapy regimen" refers to a timely sequential or simultaneous administration of anti-tumor, and/or anti vascular, and/or immune stimulating, and/or blood
20 cell proliferative agents, and/or radiation therapy, and/or hyperthermia, and/or hypothermia and/or hormonal treatment for cancer therapy. The administration of these can be performed in an adjuvant and/or neoadjuvant mode. The composition of such "protocol" may vary in the dose of the
25 single agent, timeframe of application and frequency of administration within a defined therapy window. Currently various combinations of various drugs and/or physical methods, and various schedules are under investigation.

30 The term "cytotoxic treatment" refers to various treatment modalities affecting cell proliferation and/or survival. The treatment may include administration of alkylating agents, antimetabolites, anthracyclines, plant alkaloids, topoisomerase inhibitors, and other antitumour agents,
35 including monoclonal antibodies, inhibitors of repair mechanisms and kinase inhibitors. In particular, the cytotoxic treatment may relate to a taxane treatment. Taxanes are plant alkaloids which block cell division by preventing

microtubule function. The prototype taxane is the natural product paclitaxel, originally known as Taxol and first derived from the bark of the Pacific Yew tree. Docetaxel is a semi-synthetic analogue of paclitaxel. Taxanes enhance
5 stability of microtubules, preventing the separation of chromosomes during anaphase.

The term "endocrine treatment" or "hormonal treatment" (sometimes also referred to as "anti-hormonal treatment")
10 denotes a treatment which targets hormone signalling, e.g. hormone inhibition, hormone receptor inhibition, use of hormone receptor agonists or antagonists, use of scavenger- or orphan receptors, use of hormone derivatives and interference with hormone production. Particular examples
15 are tamoxifene therapy which modulates signalling of the estrogen receptor, or aromatase treatment which interferes with steroid hormone production.

Tamoxifen is an orally active selective estrogen receptor modulator (SERM) that is used in the treatment of breast
20 cancer and is currently the world's largest selling drug for that purpose. Tamoxifen is sold under the trade names Nolvadex, Istubal, and Valodex. However, the drug, even before its patent expiration, was and still is widely
25 referred to by its generic name "tamoxifen." Tamoxifen and Tamoxifen derivatives competitively bind to estrogen receptors on tumors and other tissue targets, producing a nuclear complex that decreases DNA synthesis and inhibits estrogen effects.

30 Steroid receptors are intracellular receptors (typically cytoplasmic) that perform signal transduction for steroid hormones. Examples include type I Receptors, in particular sex hormone receptors, e.g. androgen receptor, estrogen
35 receptor, progesterone receptor; Glucocorticoid receptor, mineralocorticoid receptor; and type II Receptors, e.g. vitamin A receptor, vitamin D receptor, retinoid receptor, thyroid hormone receptor.

The term "expression level" refers, e.g., to a determined level of gene expression. In its simplest form this can mean "determining if there is a detectable expression of a gene" (qualitative determination). This can be done e.g. immunohistochemically by qualitative determination by a pathologist. It can also be done by a quantitative determination wherein the gene is present, if the expression is above a threshold level.

The term "pattern of expression levels" refers to a determined level of gene expression compared either to a reference gene, e.g. housekeeper, or inversely regulated genes, or to a computed average expression value, e.g. in DNA-chip analyses. A pattern is not limited to the comparison of two genes but is more related to multiple comparisons of genes to reference genes or samples. A certain "pattern of expression levels" may also result and be determined by comparison and measurement of several genes disclosed hereafter and display the relative abundance of these transcripts to each other.

The term "determining the expression level of a gene on a non protein basis" relates to methods which are not focussed on the secondary gene translation products, i.e. proteins, but on other levels of the gene expression, based on RNA and DNA analysis. In one embodiment of this invention the analysis uses mRNA including its precursor forms. An exemplary determinable property is the amount of the estrogen receptor or progesterone receptor RNA, i.e. of the ESR1, ESR2 and/or PGR gene.

The term "RNA expression level" refers to a determined level of the converted DNA gene sequence information into transcribed RNA, the initial unspliced RNA transcript or the mature mRNA. RNA expression can be monitored by measuring the levels of either the entire RNA of the gene or subsequences.

The term "pattern of RNA expression" refers to a determined level of RNA expression compared either to a reference RNA or to a computed average expression value. A pattern is not limited to the comparison of two RNAs but is more related to multiple comparisons of RNAs to reference RNAs or samples. A certain "pattern of expression levels" may also result and be determined by comparison and measurement of several RNAs and display the relative abundance of these transcripts to each other.

A "reference pattern of expression levels", within the meaning of the invention shall be understood as being any pattern of expression levels that can be used for the comparison to another pattern of expression levels. In a preferred embodiment of the invention, a reference pattern of expression levels is, e.g., an expression level of at least one reference gene, e.g. a housekeeping gene or a mixture of housekeeping genes. In a preferred embodiment of the invention, a reference pattern of expression levels is, e.g., an average pattern of expression levels observed in a group of healthy or diseased individuals, serving as a reference group.

The terms "tumor sample" or "tumor containing sample", as used herein, refer to a sample obtained from a patient which sample contains tumor cells. The sample may be of any biological tissue or fluid. Such samples include, but are not limited to, sputum, blood, serum, plasma, blood cells (e.g., white cells), tissue, core or fine needle biopsy samples, cell-containing body fluids, urine, peritoneal fluid, and pleural fluid, liquor cerebrospinalis, tear fluid, or cells isolated therefrom. This may also include sections of tissues such as frozen or fixed sections taken for histological purposes or microdissected cells or extracellular parts thereof. A tumor sample to be analyzed can be tissue material from a neoplastic lesion taken by aspiration or punctuation, excision or by any other surgical method leading to biopsy or resected cellular material. Such comprises tumor cells or

tumor cell fragments obtained from the patient. The cells may be found in a cell "smear" collected, for example, by a nipple aspiration, ductal lavage, fine needle biopsy or from provoked or spontaneous nipple discharge. In another
5 embodiment, the sample is a body fluid. Such fluids include, for example, blood fluids, serum, plasma, lymph, ascitic fluids, gynecologic fluids, or urine but not limited to these fluids.

10 By "array" or "microarray" is meant an arrangement of addressable locations or "addresses" on a device. The locations can be arranged in two dimensional arrays, three dimensional arrays, or other matrix formats. The number of
15 locations can range from several to at least hundreds of thousands. Most importantly, each location represents an independent reaction site. Arrays include but are not limited to nucleic acid arrays, protein arrays and antibody arrays. A "nucleic acid array" refers to an array containing nucleic acid probes, such as oligonucleotides, polynucleotides or
20 larger portions of genes. The nucleic acid on the array is preferably single stranded. Arrays wherein the probes are oligonucleotides are referred to as "oligonucleotide arrays" or "oligonucleotide chips." A "microarray," herein also refers to a "biochip" or "biological chip", an array of
25 regions having a density of discrete regions of at least about 100/cm², and preferably at least about 1000/cm². The regions in a microarray have typical dimensions, e.g., diameters, in the range of between about 10-250 μm, and are separated from other regions in the array by about the same
30 distance.

The term "oligonucleotide" refers to a relatively short polynucleotide, including, without limitation, single-stranded deoxyribonucleotides, single- or double-stranded
35 ribonucleotides, RNA:DNA hybrids and double-stranded DNAs. Oligonucleotides are preferably single-stranded DNA probe oligonucleotides. Moreover, in context of applicable detection methodologies, the term "oligonucleotide" also

refers to nucleotide analogues such as PNAs and morpholinos.

The terms "modulated" or "modulation" or "regulated" or "regulation" and "differentially regulated" as used herein
5 refer to both upregulation, i.e., activation or stimulation, e.g., by agonizing or potentiating, and down regulation, i.e., inhibition or suppression, e.g., by antagonizing, decreasing or inhibiting.

10 "Primer pairs" and "probes", within the meaning of the invention, shall have the ordinary meaning of this term which is well known to the person skilled in the art of molecular biology. In a preferred embodiment of the invention "primer pairs" and "probes", shall be understood as being
15 polynucleotide molecules having a sequence identical, complementary, homologous, or homologous to the complement of regions of a target polynucleotide which is to be detected or quantified. In yet another embodiment, nucleotide analogues are also comprised for usage as primers and/or probes. Probe
20 technologies used for kinetic or real time PCR applications could be e.g. TaqMan® systems obtainable at Roche Molecular Diagnostics, extension probes such as Scorpion® Primers, Dual Hybridisation Probes, Amplifluor® obtainable at Chemicon International, Inc, or Minor Groove Binders.

25 "Individually labeled probes", within the meaning of the invention, shall be understood as being molecular probes comprising a polynucleotide, oligonucleotide or nucleotide analogue and a label, helpful in the detection or
30 quantification of the probe. Preferred labels are fluorescent molecules, luminescent molecules, radioactive molecules, enzymatic molecules and/or quenching molecules.

"Arrayed probes", within the meaning of the invention, shall
35 be understood as being a collection of immobilized probes, preferably in an orderly arrangement. In a preferred embodiment of the invention, the individual "arrayed probes"

can be identified by their respective position on the solid support, e.g., on a "chip".

The phrase "response", "therapeutic success", or "response to therapy" refers in the neoadjuvant, adjuvant and palliative chemotherapeutic setting to the observation of a defined tumor free or recurrence free or progression free survival time (e.g. 2 years, 4 years, 5 years, 10 years). This time period of disease free -, recurrence free - or progression free survival may vary among the different tumor entities but is sufficiently longer than the average time period in which most of the recurrences appear in the absence of the given therapy. In a neoadjuvant and palliative therapy modality, response may additionally be monitored by measurement of tumor shrinkage and regression due to apoptosis and necrosis of the tumor mass or reduced blood supply due to altered angiogenic events.

The term "recurrence" or "recurrent disease" includes distant metastasis that can appear even many years after the initial diagnosis and therapy of a tumor, or local events such as infiltration of tumor cells into regional lymph nodes, or occurrence of tumor cells at the same site and organ of origin within an appropriate time.

"Prediction of recurrence" or "prediction of death after recurrence" does refer to the methods described in this invention, wherein a tumor specimen is analyzed for e.g. its gene expression, genomic status and/or histopathological parameters (such as TNM and Grade) and/or imaging data and furthermore classified based on correlation of the expression pattern to known ones from reference samples. This classification may either result in the statement that such given tumor will develop recurrence and therefore is considered as a "not sufficiently responding" tumor to the given therapy, or may result in a classification as a tumor with a prolonged disease free post therapy time.

The term "marker" or "biomarker" refers to a biological molecule, e.g., a nucleic acid, peptide, protein, hormone, etc., whose presence or concentration can be detected and correlated with a known condition, such as a disease state or a combination of these, e.g. by a mathematical algorithm.

The term "marker gene" as used herein, refers to a differentially expressed gene whose expression pattern may be utilized as part of a predictive, prognostic or diagnostic process in malignant neoplasia or cancer evaluation, or which, alternatively, may be used in methods for identifying compounds useful for the treatment or prevention of malignant neoplasia and gynecological cancer in particular. A marker gene may also have the characteristics of a target gene.

The term "signalling pathway" is related to any intra- or intercellular process by which cells converts one kind of signal or stimulus into another, most often involving ordered sequences of biochemical reactions out- and inside the cell, that are carried out by enzymes and linked through hormones and growth factors (intercellular), as well as second messengers (intracellular), the latter resulting in what is thought of as a "second messenger pathway". In many signalling pathways, the number of proteins and other molecules participating in these events increases as the process emanates from the initial stimulus, resulting in a "signal cascade" and often results in a relatively small stimulus eliciting a large response.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe that can hybridize (i.e., it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described above.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids.

The term "hybridization based method", as used herein, refers to methods imparting a process of combining complementary, single-stranded nucleic acids or nucleotide analogues into a single double stranded molecule. Nucleotides or nucleotide analogues will bind to their complement under normal conditions, so two perfectly complementary strands will bind to each other readily. In bioanalytics, very often labeled, single stranded probes are in order to find complementary target sequences. If such sequences exist in the sample, the probes will hybridize to said sequences which can then be detected due to the label. Other hybridization based methods comprise microarray and/or biochip methods. Therein, probes are immobilized on a solid phase, which is then exposed to a sample. If complementary nucleic acids exist in the sample, these will hybridize to the probes and can thus be detected. These approaches are also known as "array based methods". Yet another hybridization based method is PCR, which is described below. When it comes to the determination of expression levels, hybridization based methods may for example be used to determine the amount of mRNA for a given gene.

The term "a PCR based method" as used herein refers to methods comprising a polymerase chain reaction (PCR). This is an approach for exponentially amplifying nucleic acids, like DNA or RNA, via enzymatic replication, without using a living organism. As PCR is an in vitro technique, it can be performed without restrictions on the form of DNA, and it can be extensively modified to perform a wide array of genetic manipulations. When it comes to the determination of expression levels, a PCR based method may for example be used to detect the presence of a given mRNA by (1) reverse transcription of the complete mRNA pool (the so called transcriptome) into cDNA with help of a reverse transcriptase enzyme, and (2) detecting the presence of a given cDNA with help of respective primers. This approach is commonly known as reverse transcriptase PCR (rtPCR). The term "PCR based method" comprises both end-point PCR applications as well as kinetic/real time PCR techniques applying special fluorophors

or intercalating dyes which emit fluorescent signals as a function of amplified target and allow monitoring and quantification of the target. Quantification methods could be either absolute by external standard curves or relative to a comparative internal standard.

The term "nucleic acid molecule" is intended to indicate any single- or double stranded nucleic acid and/or analogous molecules comprising DNA, cDNA and/or genomic DNA, RNA, preferably mRNA, peptide nucleic acid (PNA), locked nucleic acid (LNA) and/or Morpholino.

The term "stringent conditions" relates to conditions under which a probe will hybridize to its target subsequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5° C. lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. (As the target sequences are generally present in excess, at T_m , 50% of the probes are occupied at equilibrium). Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C. for short probes (e.g. 10 to 50 nucleotides) and at least about 60° C. for longer probes. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide and the like.

The term "hybridizing counterparts" as used herein, refers to a nucleic acid molecule that is capable of hybridizing to a nucleic acid molecules under stringent conditions.

Summary of the Invention

The present invention is based on the surprising finding that
5 the outcome of cancer, preferably breast cancer in patients
which do receive tamoxifen but not chemotherapy, can be
accurately predicted from the expression levels of a small
number of marker genes. Accordingly, the present invention
relates to classification methods for the determination of
10 the outcome of breast cancer in tamoxifen-treated breast
cancer patients, using information on the expression of a
small number of highly informative marker genes. The method
is based on multiple comparisons of determined expression
levels with predetermined threshold levels. Depending on the
15 outcome of a first comparison, further comparisons of
expression levels with threshold values are performed, as
shown in the decision tree in Figure 1. The outcome of the
prognostic method is a classification of the tumor under
investigation into one of several "risk classes". Risk
20 classes of the invention are preferably a "low risk of
recurrence" class, an "intermediate risk of recurrence class"
and a "high risk of recurrence class". The outcome of the
prognostic methods of the invention provides useful
information for the selection of the most suitable treatment
25 regimen for the patient. Accordingly, chemotherapy may be
avoided if a patient has a low risk of recurrence, or
aggressiveness of chemotherapy may be adjusted to the
likelihood of recurrence of the tumor, i.e. a tumor with a
intermediate risk might justify a less aggressive treatment
30 compared to a tumor with a high risk of recurrence.

Methods of the invention are multi-step classification
methods in which consecutive classification steps are

performed, the classifiers used in each step (i.e. the marker genes and threshold levels applied) being dependent on the outcome of the previous classification step(s). It has been found that this classification strategy is in many cases superior to other methods, using a single multi-variate classification step. Without being bound by theory, it is assumed that the superior performance of sequential classification schemes over single-step classification schemes is due to the fact that marker gene expression can be differently correlated with the disease outcome in different subgroups of patients. For example, high expression of a marker gene A may be positively correlated with good disease outcome in patients having high expression of marker gene B, but may be negatively correlated with good disease outcome in patients having low expression of marker gene B. In multi-step classification methods, this can be taken into account by e.g. first classifying on the basis of the expression level of marker gene B, and then classifying on the basis of the correct correlation of marker gene A expression with the disease outcome.

The method of the invention is a method of prediction of a prognosis of cancer in a patient treated with Tamoxifen using the classification scheme generally shown in Figures 1 to 3.

The present invention provides a method to assess the risk of recurrence of a node positive and hormone receptor positive breast cancer patient when treated with tamoxifen. The method is based on quantitative determination of RNA species isolated from the tumor in order to obtain expression values and subsequent bioinformatic analysis of said determined expression values, as generally depicted in Figs.1 and 2. First, RNA coding for ESR1 and PGR are determined. Based on these expression values the tumor is classified as steroid

hormone receptor (SR) present or SR absent. Alternatively, this classification may also be performed on a protein-basis, for example determining the presence or absence of steroid receptor expression by immunohistological staining, as is known in the art. Then, one or more RNA species coding for proliferation associated genes are determined. E.g. when expression values of said proliferation associated genes are below a predetermined cut off value said tumor is classified as low proliferation tumor. Suitable genes for making this determination are listed in table 1. Next, expression values of ESR1, PGR or expression values of genes co-regulated with ESR1 or PGR are used to determine whether these values are higher than a predetermined cut off value. Tumors with expression values above the cut off value are classified as SR high. In a preferred embodiment, expression values of proliferation associated genes are taken into account when determining the SR high class, thereby classifying only those tumors as SR high that do not exhibit a high expression of proliferation associated genes. A tumor not classified as SR absent, proliferation low or SR high is further stratified using expression values of RNA species suitable to classify the tumor as intermediate or high risk. Suitable RNA species are those indicating the presence of tumor infiltrating T-lymphocytes, in particular RNA molecules coding for genes listed in table 2. Further suitable RNA species are listed in table 2. In a preferred embodiment of the invention a combination of genes indicating the presence of T-lymphocyte infiltration and one or more from the genes listed in table 2 is used to classify a tumor as intermediate or high risk tumor.

Detailed description of the Invention

In particular, the invention relates to:

- 5 1. A Method of classifying a tumor sample for the prediction of prognosis of a breast cancer patient receiving anti-hormonal treatment, said method comprising
- a) determining in a tumor sample whether steroid receptor expression in a patient sample is present or absent
- 10 by determining if there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene;
- (b) determining the expression level of at least one first marker gene in said tumor sample from said patient,
- 15 said at least one first marker gene providing information on whether said tumor is fast proliferating or slow proliferating;
- (c) determining whether the expression level of said at least one steroid receptor gene or at least one gene co-
- 20 regulated with a steroid receptor gene is above or below a predetermined first threshold level;
- (d) determining whether the expression level of at least one second marker gene in a tumor sample of said patient is above a predetermined second threshold level, wherein said
- 25 second marker gene is specifically expressed in T cells;

Wherein said sample is classified as low risk if

- (i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with
- 30 a steroid receptor gene and
- (ii) the expression level of said first marker gene indicates that said tumor is slow proliferating;

Wherein said sample is classified as low risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene,

(ii) the expression level of said first marker gene indicates that said tumor is fast proliferating; and

(iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is above the predetermined first threshold level;

10

Wherein said sample is classified as intermediate risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene,

(ii) the expression level of said first marker gene indicates that said tumor is fast proliferating,

(iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is below the predetermined first threshold level, and

(iv) the expression level of the at least one second marker gene in a tumor sample of said patient is above the predetermined second threshold level;

25

Wherein said sample is classified as high risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene,

(ii) the expression level of said first marker gene indicates that said tumor is fast proliferating,

(iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a

30

steroid receptor gene is below the predetermined first threshold level, and

(iv) the expression level of the at least one second marker gene in a tumor sample of said patient is below the
5 predetermined second threshold level;

and wherein said sample is classified as high risk if

(v) there is no detectable expression of at least one steroid receptor gene or at least one gene co-regulated with
10 a steroid receptor gene.

2. Method of the above numbered paragraph(s) 1, wherein in step (b) an expression level of said at least one first marker gene above a third threshold value is indicative of a
15 fast proliferating tumor.

3. Method of above numbered paragraph(s) 1 or 2, wherein said risk is the risk of developing metastasis in a predetermined period of time.
20

4. Method of any one of the preceding above numbered paragraphs, wherein said at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is selected from the group consisting of estrogen
25 receptor 1 (ESR1), progesterone receptor (PGR) or a combination thereof.

5. Method of any one of the preceding above numbered paragraphs, wherein said first marker gene is selected from
30 Table 1.

6. Method of any one of the preceding above numbered paragraphs, wherein said first marker gene is TOP2A or a gene co-regulated with TOP2A.

7. Method of any one of the preceding above numbered paragraphs, wherein said second marker gene is selected from Table 2.

5

8. Method of any of the preceding above numbered paragraphs, wherein the determination of expression levels is performed using a gene chip.

10 9. Method of any of the preceding above numbered paragraphs, wherein the determination of expression levels is performed using a PCR-based method.

15 10. Method of any of the preceding above numbered paragraphs, wherein said anti-hormonal treatment comprises a tamoxifen treatment.

20 11. Method of any of the preceding above numbered paragraphs, wherein the expression level of at least one first marker gene is determined as a proliferation metagene expression value which is constructed by mathematically combining expression values of a plurality of first marker genes using 2, 3, 4, 5, 10, 20, 50, or all of the genes listed in Table 1.

25

30 12. Method of any of the preceding above numbered paragraphs, wherein the expression level of at least one second marker gene is determined as a T-cell metagene expression value which is constructed by mathematically combining expression values of a plurality of second marker genes using 2, 3, 4, 5, 6, 7, or all of the genes listed in Table 2.

14. Method of any of the preceding numbered paragraphs, wherein the patient is node positive.

15. A system for classifying a tumor sample of a breast cancer patient to be selected for receiving anti-hormonal treatment, said system comprising

(a) means for storing information on the SR status of said patient;

(b) means for determining the expression level of at least one first marker gene indicative of whether a tumor is fast proliferating;

(c) means for comparing said expression level of said first marker gene with a predetermined first threshold value;

(d) means for determining the expression level of at least one second marker gene wherein said second marker gene is specifically expressed in T cells; and

(e) computing means programmed to give a classification,

Wherein said sample is classified as low risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene,

(ii) the expression level of said first marker gene indicates that said tumor is fast proliferating; and

(iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is above the predetermined second threshold level;

Wherein said sample is classified as intermediate risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene,

(ii) the expression level of said first marker gene indicates that said tumor is fast proliferating,

(iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is below the predetermined first threshold level, and

(iv) the expression level of the at least one second marker gene in a tumor sample of said patient is above the predetermined second threshold level;

Wherein said sample is classified as high risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene,

(ii) the expression level of said first marker gene indicates that said tumor is fast proliferating,

(iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is below the predetermined first threshold level, and

(iv) the expression level of the at least one second marker gene in a tumor sample of said patient is below the predetermined second threshold level;

and wherein said sample is classified as high risk if

(v) there is no detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene.

For a prognostic method to "be based" on a multiple pieces of information (as is the case in the present invention) all

individual pieces of information must be taken into consideration for arriving at the prognosis or classification. This means that all individual pieces of information can influence the outcome of the prognosis or
5 classification. It is well understood that a piece of information, such as e.g. the steroid receptor status of a patient, can influence the outcome of the prognosis or classification in that the is only applied when said steroid receptor status is e.g. positive. Likewise, it is understood
10 that a method can "be based" on information relating to the proliferation rate of the tumor, e.g. if fast proliferation is a conditional criterion applied in the course of the method.

15 In the methods of the invention, said classification or prognosis is entirely based on the information that said steroid receptor status is positive and that said tumor is a fast proliferating tumor and on information on the expression level of said second marker gene in said tumor sample.

20 A "fast proliferating tumor" within the meaning of the present invention is a tumor with a gene expression pattern indicative of fast proliferation and not taking into account the actual proliferation rate or rate of tumor growth in
25 terms of mass or size.

In preferred methods of the invention, said classification is an estimation of the likelihood of metastasis fee survival of said patient over a predetermined period of time, e.g. over a
30 period of 5 years.

In further preferred methods of the invention, said classification is an estimation of the likelihood of death of

disease of said patient over a predetermined period of time, e.g. over a period of 5 years.

"Death of disease", within the meaning of the invention, shall be understood to be the death of a breast cancer patient after recurrence of the disease.

"Recurrence", within the meaning of the invention, shall be understood to be the recurrence of breast cancer in form of metastatic spread of tumor cells, local recurrence, contralateral recurrence or recurrence of breast cancer at any site of the body of the patient.

In preferred methods of the invention, the expression of said first marker gene indicative of fast proliferation of the tumor is a gene selected from Table 1.

In specific embodiments of the invention, a single, or 2 to 5, 2 to 10, 2 to 20, 2 to 50 or 2 to 100 first marker genes are used.

Table 1:

Probe Set	Classification	Gene Symbol	Location
222039_at	Proliferation	LOC146909	Chr:17q21.31
218662_s_at	Proliferation	HCAP-G	Chr:4p16-p15
221520_s_at	Proliferation	FLJ10468	Chr:1p34.2
218755_at	Proliferation	KIF20A	Chr:5q31
204825_at	Proliferation	MELK	Chr:9p13.1
218542_at	Proliferation	C10orf3	Chr:10q23.33
204444_at	Proliferation	KIF11	Chr:10q24.1
218039_at	Proliferation	ANKT	Chr:15q14
202705_at	Proliferation	CCNB2	Chr:15q21.2
218009_s_at	Proliferation	PRC1	Chr:15q26.1
210052_s_at	Proliferation	C20orf1	Chr:20q11.2

202954_at	Proliferation	UBE2C	Chr:20q13.11
202095_s_at	Proliferation	BIRC5	Chr:17q25
			Chr:20q13.2-
208079_s_at	Proliferation	STK6	q13.3
			Chr:20q13.2-
204092_s_at	Proliferation	STK6	q13.3
209642_at	Proliferation	BUB1	Chr:2q14
204962_s_at	Proliferation	CENPA	Chr:2p24-p21
218355_at	Proliferation	KIF4A	Chr:Xq13.1
209408_at	Proliferation	KIF2C	Chr:1p34.1
202870_s_at	Proliferation	CDC20	Chr:1p34.1
202580_x_at	Proliferation	FOX M1	Chr:12p13
209714_s_at	Proliferation	CDKN3	Chr:14q22
203764_at	Proliferation	DLG7	Chr:14q22.1
203554_x_at	Proliferation	PTTG1	Chr:5q35.1
214710_s_at	Proliferation	CCNB1	Chr:5q12
210559_s_at	Proliferation	CDC2	Chr:10q21.1
203214_x_at	Proliferation	CDC2	Chr:10q21.1
203213_at	Proliferation	CDC2	Chr:10q21.1
206102_at	Proliferation	KIAA0186	Chr:20p11.1
218726_at	Proliferation	DKFZp762E1312	Chr:2q37.1
213226_at	Proliferation	PMSCL1	Chr:4q27
203362_s_at	Proliferation	MAD2L1	Chr:4q27
203418_at	Proliferation	CCNA2	Chr:4q25-q31
219918_s_at	Proliferation	ASPM	Chr:1q31
204641_at	Proliferation	NEK2	Chr:1q32.2-q41
207828_s_at	Proliferation	CENPF	Chr:1q32-q41
			Chr:1pter-
206364_at	Proliferation	KIF14	q31.3
204822_at	Proliferation	TTK	Chr:6q13-q21
204162_at	Proliferation	HEC	Chr:18p11.31
204033_at	Proliferation	TRIP13	Chr:5p15.33
212022_s_at	Proliferation	MKI67	Chr:10q25-qter
205046_at	Proliferation	CENPE	Chr:4q24-q25
219148_at	Proliferation	TOPK	Chr:8p21.2
219978_s_at	Proliferation	ANKT	Chr:15q14
218883_s_at	Proliferation	FLJ23468	Chr:4q35.1
209773_s_at	Proliferation	RRM2	Chr:2p25-p24

201890_at	Proliferation	RRM2	Chr:2p25-p24
204026_s_at	Proliferation	ZWINT	Chr:10q21-q22
202503_s_at	Proliferation	KIAA0101	Chr:15q22.1
203145_at	Proliferation	SPAG5	Chr:17q11.1
201292_at	Proliferation	TOP2A	Chr:17q21-q22
201291_s_at	Proliferation	TOP2A	Chr:17q21-q22
			Chr:5q33.2-
207165_at	Proliferation	HMMR	qter
218663_at	Proliferation	HCAP-G	Chr:4p16-p15
209464_at	Proliferation	STK12	Chr:17p13.1
221436_s_at	Proliferation	GRCC8	Chr:12p13
202779_s_at	Proliferation	E2-EPF	Chr:19q13.43
220651_s_at	Proliferation	MCM10	Chr:10p13
205394_at	Proliferation	CHEK1	Chr:11q24-q24
205393_s_at	Proliferation	CHEK1	Chr:11q24-q24
212949_at	Proliferation	BRRN1	Chr:2q11.2
			Chr:12p13.2-
204146_at	Proliferation	PIR51	p13.1
204023_at	Proliferation	RFC4	Chr:3q27
202107_s_at	Proliferation	MCM2	Chr:3q21
202589_at	Proliferation	TYMS	Chr:18p11.32
219555_s_at	Proliferation	BM039	Chr:16q23.1
202094_at	Proliferation	BIRC5	Chr:17q25
204603_at	Proliferation	EXO1	Chr:1q42-q43
204170_s_at	Proliferation	CKS2	Chr:9q22
203358_s_at	Proliferation	EZH2	Chr:7q35-q36
			Chr:5q23.3-
203276_at	Proliferation	LMNB1	q31.1
201710_at	Proliferation	MYBL2	Chr:20q13.1
218585_s_at	Proliferation	RAMP	---
218308_at	Proliferation	TACC3	Chr:4p16.3
211814_s_at	Proliferation	CCNE2	Chr:8q22.1
205034_at	Proliferation	CCNE2	Chr:8q22.1
219000_s_at	Proliferation	MGC5528	Chr:8q24.12
203046_s_at	Proliferation	TIMELESS	Chr:12q12-q13
			Chr:17q23.2-
202338_at	Proliferation	TK1	q25.3
220295_x_at	Proliferation	FLJ20354	Chr:1p31.2

206632_s_at	Proliferation	APOBEC3B	Chr:22q13.1-q13.2
204318_s_at	Proliferation	GTSE1	Chr:22q13.2-q13.3
213008_at	Proliferation	FLJ10719	Chr:15q25-q26
202240_at	Proliferation	PLK	Chr:16p12.3
219493_at	Proliferation	SHCBP1	Chr:16q11.2
219105_x_at	Proliferation	ORC6L	Chr:16q12
221521_s_at	Proliferation	LOC51659	Chr:16q24.1
203968_s_at	Proliferation	CDC6	Chr:17q21.3
203967_at	Proliferation	CDC6	Chr:17q21.3
209916_at	Proliferation	KIAA1630	Chr:10p14
205436_s_at	Proliferation	H2AFX	Chr:11q23.2-q23.3
221922_at	Proliferation	LGN	Chr:1p13.2
205240_at	Proliferation	LGN	Chr:1p13.2
218741_at	Proliferation	MGC861	Chr:22q13.2
216237_s_at	Proliferation	MCM5	Chr:22q13.1
201755_at	Proliferation	MCM5	Chr:22q13.1
209832_s_at	Proliferation	CDT1	Chr:16q24.3
222077_s_at	Proliferation	RACGAP1	Chr:12q13.13

In a preferred embodiment of the invention, said first marker gene is TOP2A, UBE2C or a combination of TOP2A and UBE2C. A combination of 2 markers preferably comprises the mathematical combination of the individual signal strengths of each single marker. This may be a direct sum or multiplication product or a weighted sum or product. In another specific embodiment of the invention said first marker gene is a gene co-regulated with TOP2A. Co-regulation of two genes, according to the invention, is preferably exemplified by a correlation coefficient between expression levels of said two genes in multiple tissue samples of greater than 0.5, 0.6, 0.7, 0.8, 0.9, or, most preferably greater than 0.9. The statistical accuracy of the

determination of said correlation coefficient is preferably +/- 0.1 (absolute standard deviation).

5 In a preferred embodiment of the invention, a proliferation metagene expression value is constructed by mathematical combination using at least 2, 3, 4, 5, 10, 20, 50, or all of the genes listed in Table 1.

10 In a preferred embodiment of the invention, a proliferation metagene expression value is constructed using 2, 3, 4, 5, 6, or all genes from the list of TOP2A, UBE2C, RACGAP1, CENPE, PTTG1, MKI67, or CCNB1.

15 "Proliferation metagene expression value", within the meaning of the invention, shall be understood to be a calculated gene expression value representing the proliferative activity of a tumor.

20 A metagene expression value, in this context, is to be understood as being the median of the normalized expression of multiple marker genes. Normalization of the expression of multiple marker genes is preferably achieved by dividing the expression level of the individual marker genes to be normalized by the respective individual median expression of these marker genes (per gene normalization), wherein said median expression is preferably calculated from multiple measurements of the respective gene in a sufficiently large cohort of test individuals. The test cohort preferably comprises at least 3, 10, 100, or 200 individuals.

30

Preferably, the calculation of the proliferation metagene expression value is performed by:

- i) determining the gene expression value of at least two, preferably more genes from the list of table 1
- ii) "normalizing" the gene expression value of each individual gene by dividing the expression value with a coefficient which is approximately the median expression value of the respective gene in a representative node negative breast cancer cohort
- iii) calculating the median of the group of normalized gene expression values

The present invention further relates to a method as defined above, wherein said second marker gene is a gene expressed in T cells which is specifically expressed in T-cells and not in tumor cells. A gene shall be understood to be specifically expressed in a certain cell type, within the meaning of the invention, if the expression level of said gene in said cell type is at least 2-fold, 5-fold, 10-fold, 100-fold, 1000-fold, or 10000-fold higher than in a reference cell type, or in a mixture of reference cell types. Preferred reference cell types are muscle cells, smooth muscle cells, fibroblast, fat cells, B-cells, macrophages or non-cancerous breast tissue cells.

Alternatively, a gene expressed in T cells shall be understood as being a gene selected from Table 2. In preferred methods of the invention said second marker gene is selected from Table 2.

Table 2: List of Preferred T-cell Genes

Probeset ID	Gene Symbol	Title	Location	Motif

222062_at	WSX1	Class I cytokine receptor	Chr: 19p13.11	T-cell
215346_at	TNFRSF5	Tumor necrosis factor receptor superfamily, member 5	Chr: 20q12- q13.2	T-cell
211902_x_at	TRA@	T cell receptor alpha locus	Chr: 14q11.2	T-cell
211796_s_at	TRB@		Chr: 7q34	T-cell
209671_x_at	TRA@	T cell receptor alpha locus	Chr: 14q11.2	T-cell
209670_at	TRA@	T cell receptor alpha locus	Chr: 14q11.2	T-cell
205758_at	CD8A	CD8 antigen, alpha polypeptide	Chr: 2p12	T-cell
205456_at	CD3E	CD3E antigen, epsilon polypeptide	Chr: 11q23	T-cell

Because of the great variability in the primary sequence of immune genes it is conceived that the concept of using metagenes is particularly useful when determining the T cell gene status in methods of the invention. Thus, in a preferred embodiment of the invention, the claimed methods use the information on the expression of at least one proliferation

marker gene (preferably selected from Table 1), and information on the expression of multiple T cell specific genes (preferably selected from Table 2), e.g., an T cell metagene expression is applied.

5

In further preferred embodiments of the invention, the expression level of multiple first and second marker genes are determined in steps (b), (c) and/or (d), and a comparison step between the multiple first and the multiple second
10 marker genes is performed by a "majority voting algorithm".

The expression status of a plurality of first marker genes and second marker genes, respectively can also independently be determined by a majority voting algorithm.

15

In a majority voting algorithm, according to the invention, a suitable threshold level is first determined for each individual first (or each individual second) marker gene used in the method. The suitable threshold level can be
20 determined from measurements of the marker gene expression in multiple individuals from a test cohort. For example, the median expression of the first said marker gene in said multiple expression measurements can be taken as the suitable threshold value for the first said marker gene. And, for
25 example, the third quartile expression of the second said marker gene in said multiple expression measurements is taken as the suitable threshold value for the second said marker gene.

30 In a majority voting algorithm, the comparison of multiple marker genes with a threshold level is performed as follows:

1. The individual marker genes are compared to their respective threshold levels.

2. The number of marker genes, the expression level of which is above their respective threshold level, is determined.

5 3. If a sufficiently large number of marker genes is expressed above their respective threshold level, then the expression level of the multiple marker genes is taken to be "above the threshold level".

10 "A sufficiently large number", in this context, means preferably 30%, 50%, 80%, 90%, or 95% of the marker genes used.

Because of the great variability in the primary sequence of immune genes it is conceived that the concept "majority voting" is particularly useful when determining the immune gene status in methods of the invention. Thus, in a preferred embodiment of the invention, the claimed methods use the information on the expression of one or two proliferation marker genes (preferably selected from Table 1), but information on the expression of multiple T-cell genes (preferably selected from Table 2) is compared to a threshold level using a majority voting algorithm.

25 In specific embodiments of the invention, a single, or at least 2, 5, 10, 20, 50 or 100 second marker genes are used.

Preferably, the calculation of an T cell metagene is done by

- 30 1. determining the gene expression value of at least two, preferably more genes from the list of table 2
2. "normalizing" the gene expression value of each individual gene by dividing the expression value with a coefficient which is approximately the median expression

value of the respective gene in a representative node
negative breast cancer cohort

3. calculating the median of the group of normalized gene
expression values

5

The invention will be described in exemplary fashion with
reference to the figures which show:

10 Fig. 1: Diagram of a decision tree depicting an embodiment
of the inventive method;

Fig. 2: Diagram of a decision tree depicting an embodiment
of the inventive method;

15

Fig. 3: Diagram of a decision tree depicting an embodiment
of the inventive method;

Fig. 4: Kaplan Meyer Diagram showing the results of a
20 classification according to an embodiment of the inventive
method; and

Fig. 5: Kaplan Meyer Diagram showing the results of a
classification according to an embodiment of the inventive
25 method.

A preferred method of the invention is a method of prognosis
of cancer in a patient using the classification scheme shown
in Figures 1 to 3

30

The embodiment shown in Figure 3 can also be summarized by
the following set of formulas:

SR absent = 202225_at Signal < 2000 and 208305_at Signal < 100

SR present = 202225_at Signal $\geq$ 2000 or 208305_at Signal $\geq$ 100

TOP2A_UBE2C Score = (201291_s_at Signal / 350 + 202954_at Signal / 900) / 2

10 ESR1_PGR Score = (202225_at Signal / 5000 + 208305_at Signal / 150) / 2

ProSR = (1/TOP2A_UBE2C Score + 2 * ESR1_PGR Score)

15

T-cell score = Median of :

205456_at Signal / 226

205758_at Signal / 317

209670_at Signal / 401

20 209671_x_at Signal / 432

211796_s_at Signal / 448

211902_x_at Signal / 355

215346_at Signal / 159

222062_at Signal / 132

25

Wherein the term "Signal" refers to the signal strength obtained for the respective Affymetrix probe.

The Affymetrix (Santa Clara, CA, USA) HG-U133A array and
 30 GeneChip SystemTM was used to quantify the relative transcript abundance in the breast cancer tissues. Starting from 5 μ g total RNA labelled cRNA was prepared using the Roche Microarray cDNA Synthesis, Microarray RNA Target Synthesis (T7) and Microarray Target Purification Kit

according to the manufacturer's instruction. In brief, synthesis of first strand cDNA was done by a T7-linked oligo-dT primer, followed by second strand synthesis. Double-stranded cDNA product was purified and then used as template
5 for an in vitro transcription reaction (IVT) in the presence of biotinylated UTP. Labelled cRNA was hybridized to HG-U133A arrays at 45°C for 16 h in a hybridization oven at a constant rotation (60 r.p.m.) and then washed and stained with a streptavidin-phycoerythrin conjugate using the GeneChip
10 fluidic station. We scanned the arrays at 560 nm using the GeneArray Scanner G2500A from Hewlett Packard. The readings from the quantitative scanning were analysed using the Microarray Analysis Suit 5.0 from Affymetrix. In the analysis settings the global scaling procedure was chosen which
15 multiplied the output signal intensities of each array to a mean target intensity of 500. Samples with suboptimal average signal intensities (i.e., scaling factors > 25) or GAPDH 3'/5'ratios > 5 were relabeled and rehybridized on new arrays. Routinely we obtained over 40 percent present calls
20 per chip as calculated by MAS 5.0.

Previously published microarray datasets

A breast cancer Affymetrix HG-U133A microarray dataset
25 including patient outcome information was downloaded from the NCBI GEO data repository (<http://www.ncbi.nlm.nih.gov/geo/>). The data set (GSE2034) represents 180 lymph-node negative relapse free patients and 106 lymph-node negative patients that developed a distant metastasis. None of the patients did
30 receive systemic neoadjuvant or adjuvant therapy.

Analysis of microarray data

For our unpublished dataset selection of "informative" genes was done using the quality control criteria "absent" or "present" as provided by the Affymetrix software, the absolute median signal intensity and the coefficient of variation of a gene within our dataset. Genes passing the quality control filter of having a "present" call in at least 10 samples, median signal intensity above 75 and a coefficient of variation above 60 % within our dataset were considered to be informative and used for subsequent analysis. For unsupervised analysis we performed average linkage hierarchical clustering on all informative genes and samples using Pearson correlation as implemented in GeneSpring 7.0 software (Agilent Technologies, USA). Principle component analysis was performed using GeneSpring 7.0. Clinical information was visualized as categorical or continues variable and relative gene expression was visualized on a relative scale from red, indicating high expression, to blue, indicating low expression. Gene groups were defined after manual selection of nodes of the gene dendrogram as suggested by the occurrence of cluster regions within the heatmap. A metagene was calculated as representative of all genes contained within one gene cluster based on the normalized expression values within the respective dataset. The genes contained within the proliferation cluster are listed in Table 1 and the genes contained within the T-cell gene clusters are listed in Table 2.

Primary tumor tissues from 105 patients with invasive breast carcinoma were analyzed by gene expression profiling using HG-U133A oligonucleotide arrays. All patients were node positive and did receive tamoxifen therapy after surgery. An additional Affymetrix HG-U133A microarray datasets was representing 73 node positive patients treated with tamoxifen

was downloaded from the NCBI GEO data repository (<http://www.ncbi.nlm.nih.gov/geo/>). Tumors with negative immunohistochemical staining of estrogen and progesterone receptor were excluded from the analysis. Genes and algorithm structure were designed prospectively, cut offs were defined in the Multicenter cohort. Kaplan-Meier analysis as shown in Figures 4 and 5 was performed with distant metastasis as endpoint and survival curves were compared with the logrank test. MFI = metastasis free interval in months.

References

1. Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 1998; 352:930-942
2. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998; 351:1451-1467
3. Goldhirsch A, et al. Progress and promise: highlights of the international expert consensus on the primary therapy of early breast cancer 2007. *Anal Oncol* 2007; 18: 1133-1144
4. Shapiro CL and Recht A. Side effects of adjuvant treatment of breast cancer. *N Engl J Med* 2001; 344: 1997-2008
5. Ganz PA, Desmond KA, Leedham B, Rowland JH, Meyerowitz BE, Belin TR. Quality of life in long-term, disease-free survivors of breast cancer: a follow-up study. *J Natl Cancer Inst* 2002; 94: 39-49

Claims

1. Method of classifying a tumor sample for prediction of prognosis of a breast cancer patient receiving anti-hormonal treatment, said method comprising

5 a) determining in a tumor sample whether steroid receptor expression in a patient sample is present or absent by determining if there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene;

10 (b) determining the expression level of at least one first marker gene in said tumor sample from said patient, said at least one first marker gene providing information on whether said tumor is fast proliferating or slow proliferating;

(c) determining whether the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is above or below a predetermined first threshold level;

20 (d) determining whether the expression level of at least one second marker gene in a tumor sample of said patient is above a predetermined second threshold level, wherein said second marker gene is specifically expressed in T cells;

25 Wherein said sample is classified as low risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene and

30 (ii) the expression level of said first marker gene indicates that said tumor is slow proliferating;

Wherein said sample is classified as low risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene,

(ii) the expression level of said first marker gene indicates that said tumor is fast proliferating; and

(iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is above the predetermined first threshold level;

10

Wherein said sample is classified as intermediate risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene,

(ii) the expression level of said first marker gene indicates that said tumor is fast proliferating,

(iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is below the predetermined first threshold level, and

(iv) the expression level of the at least one second marker gene in a tumor sample of said patient is above the predetermined second threshold level;

25

Wherein said sample is classified as high risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene,

(ii) the expression level of said first marker gene indicates that said tumor is fast proliferating,

(iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a

30

steroid receptor gene is below the predetermined first threshold level, and

(iv) the expression level of the at least one second marker gene in a tumor sample of said patient is below the
5 predetermined second threshold level;

and wherein said sample is classified as high risk if

(v) there is no detectable expression of at least one steroid receptor gene or at least one gene co-regulated with
10 a steroid receptor gene.

2. Method of claim 1, wherein in step (b) an expression level of said at least one first marker gene above a third threshold value is indicative of a fast proliferating tumor.
15

3. Method of claim 1 or 2, wherein said risk is the risk of developing metastasis in a predetermined period of time.

4. Method of any one of the preceding claims, wherein said
20 at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is selected from the group consisting of estrogen receptor 1 (ESR1), progesterone receptor (PGR) or a combination thereof.

25 5. Method of any one of the preceding claims, wherein said first marker gene is selected from Table 1.

6. Method of any one of the preceding claims, wherein said
30 first marker gene is TOP2A or a gene co-regulated with TOP2A.

7. Method of any one of the preceding claims, wherein said second marker gene is selected from Table 2.

8. Method of any of the preceding claims, wherein the determination of expression levels is performed using a gene chip.

5 9. Method of any of the preceding claims, wherein the determination of expression levels is performed using a PCR-based method.

10 10. Method of any of the preceding claims, wherein said anti-hormonal treatment comprises a tamoxifen treatment.

11. Method of any of the preceding claims, wherein the expression level of at least one first marker gene is determined as a proliferation metagene expression value which is constructed by mathematically combining expression values of a plurality of first marker genes using 2, 3, 4, 5, 10, 20, 50, or all of the genes listed in Table 1.

12. Method of any of the preceding claims, wherein the expression level of at least one second marker gene is determined as a T-cell metagene expression value which is constructed by mathematically combining expression values of a plurality of second marker genes using 2, 3, 4, 5, 6, 7, or all of the genes listed in Table 2.

13. Method of any of the preceding claims, wherein the patient is node positive.

14. A system for classifying a tumor sample of a breast cancer patient to be selected for receiving anti-hormonal treatment, said system comprising

(a) means for storing information on the SR status of said patient;

(b) means for determining the expression level of at least one first marker gene indicative of whether a tumor is fast proliferating;

(c) means for comparing said expression level of said
5 first marker gene with a predetermined first threshold value;

(d) means for determining the expression level of at least one second marker gene wherein said second marker gene is specifically expressed in T cells; and

(e) computing means programmed to give a
10 classification,

Wherein said sample is classified as low risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with
15 a steroid receptor gene,

(ii) the expression level of said first marker gene indicates that said tumor is fast proliferating; and

(iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a
20 steroid receptor gene is above the predetermined second threshold level;

Wherein said sample is classified as intermediate risk if

25 (i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene,

(ii) the expression level of said first marker gene indicates that said tumor is fast proliferating,

30 (iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is below the predetermined first threshold level, and

(iv) the expression level of the at least one second marker gene in a tumor sample of said patient is above the predetermined second threshold level;

5 Wherein said sample is classified as high risk if

(i) there is a detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene,

10 (ii) the expression level of said first marker gene indicates that said tumor is fast proliferating,

(iii) the expression level of said at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene is below the predetermined first threshold level, and

15 (iv) the expression level of the at least one second marker gene in a tumor sample of said patient is below the predetermined second threshold level;

and wherein said sample is classified as high risk if

20 (v) there is no detectable expression of at least one steroid receptor gene or at least one gene co-regulated with a steroid receptor gene.

1/3

FIG 1

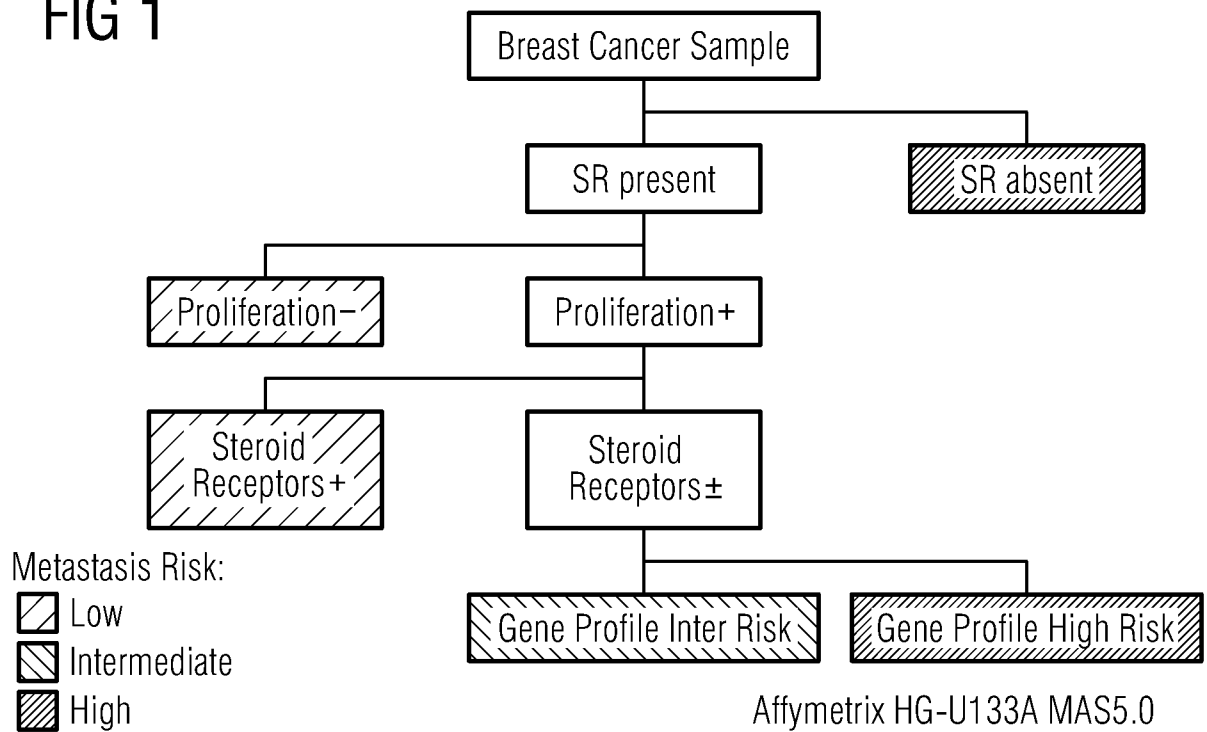


FIG 2

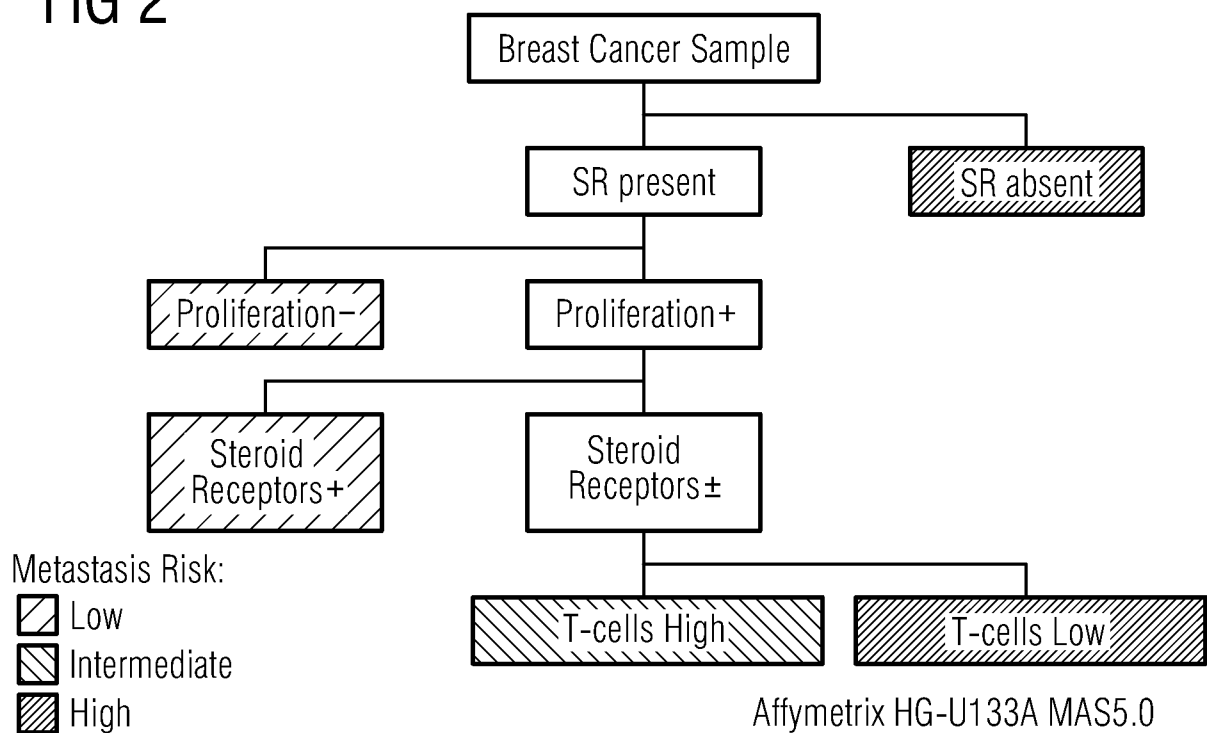
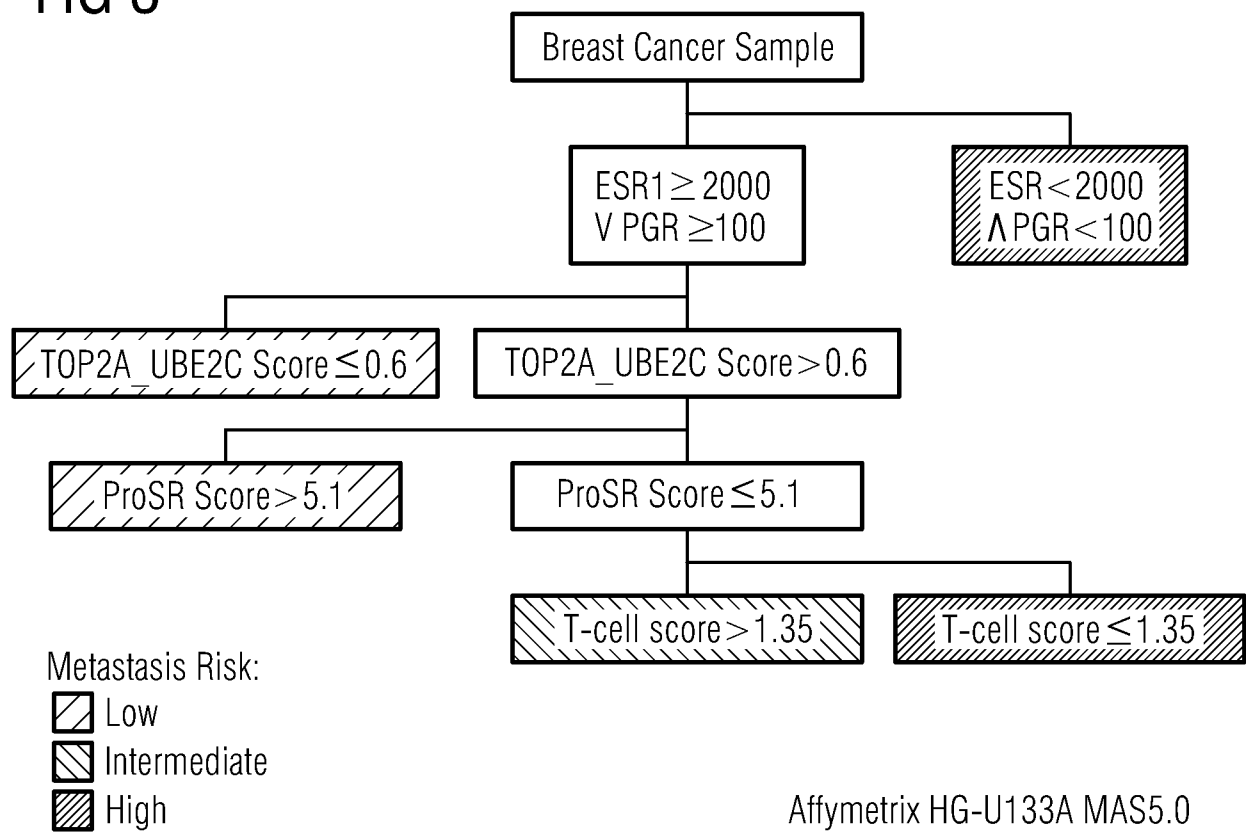


FIG 3



3/3

FIG 4

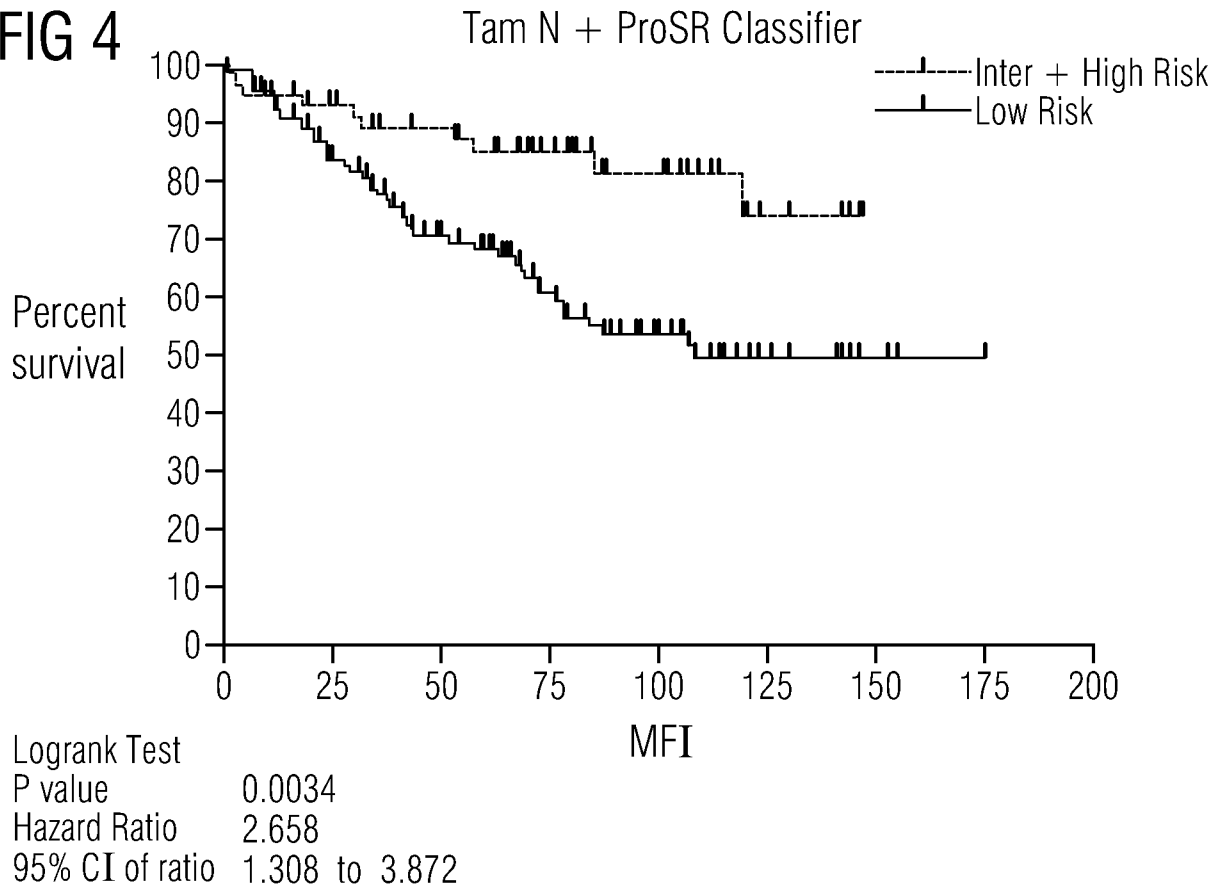


FIG 5

