HU000030164T2

(19) **HU** (11) Lajstromszám: **E 030 164** (13) **T2**

MAGYARORSZÁG
Szellemi Tulajdon Nemzeti Hivatala

EURÓPAI SZABADALOM
SZÖVEGÉNEK FORDÍTÁSA

(21) Magyar ügyszám: **E 14 188791**
(22) A bejelentés napja: **2011. 03. 29.**

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

(96) Az európai bejelentés bejelentési száma:
EP 20110188791
(97) Az európai bejelentés közzétételi adatai:
EP 2845911 A1 **2015. 03. 11.**
(97) Az európai szabadalom megadásának meghirdetési adatai:
EP 2845911 B1 **2016. 05. 18.**

(30) Elsőbbségi adatok:
10158561 **2010. 03. 31.** **EP**

(72) Feltaláló(k):
Petry, Christoph, 50829 Cologne (DE)
Feder, Inke Sabine, 50829 Cologne (DE)
Gehrmann, Mathias, 50829 Cologne (DE)
Henning, Guido, 50829 Cologne (DE)
Weber, Karsten, 50829 Cologne (DE)
Von Törne, Christian, 50829 Cologne (DE)
Kronenwett, Ralf, 50829 Cologne (DE)
Dartmann, Mareike, 50829 Cologne (DE)

(73) Jogosult(ak):
Sividon Diagnostics GmbH, 50829 Cologne (DE)

(74) Képviselő:
SBGK Szabadalmi Ügyvivői Iroda, Budapest

(54) **Eljárás emlőrák kiújulásának előrejelzésére endokrin kezelés alatt**

Az európai szabadalom ellen, megadásának az Európai Szabadalmi Közlönyben való meghirdetésétől számított kilenc hónapon belül, felszólalást lehet benyújtani az Európai Szabadalmi Hivatalnál. (Európai Szabadalmi Egyezmény 99. cikk(1))

A fordítást a szabadalmas az 1995. évi XXXIII. törvény 84/H. §-a szerint nyújtotta be. A fordítás tartalmi helyességét a Szellemi Tulajdon Nemzeti Hivatala nem vizsgálta.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 845 911 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
18.05.2016 Bulletin 2016/20

(51) Int Cl.:
C12Q 1/68 $^{(2006.01)}$

(21) Application number: **14188791.9**

(22) Date of filing: **29.03.2011**

(54) Method for breast cancer recurrence prediction under endocrine treatment

Verfahren zur Vorhersage des Rückkehr von Brustkrebs unter Endocrinbehandlung

Procédé de prédiction de la récurrence du cancer du sein sous traitement endocrinien

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **31.03.2010 EP 10158561**

(43) Date of publication of application:
11.03.2015 Bulletin 2015/11

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
11710526.2 / 2 553 118

(73) Proprietor: **Sividon Diagnostics GmbH**
50829 Cologne (DE)

(72) Inventors:
- **Petry, Christoph**
  50829 Cologne (DE)
- **Feder, Inke Sabine**
  50829 Cologne (DE)
- **Gehrmann, Mathias**
  50829 Cologne (DE)
- **Henning, Guido**
  50829 Cologne (DE)
- **Weber, Karsten**
  50829 Cologne (DE)
- **Von Törne, Christian**
  50829 Cologne (DE)
- **Kronenwett, Ralf**
  50829 Cologne (DE)
- **Dartmann, Mareike**
  50829 Cologne (DE)

(74) Representative: **von Kreisler Selting Werner -**
Partnerschaft
von Patentanwälten und Rechtsanwälten mbB
Deichmannhaus am Dom
Bahnhofsvorplatz 1
50667 Köln (DE)

(56) References cited:

EP-A1- 2 163 649	**WO-A1-2006/119593**
WO-A1-2009/158143	**WO-A1-2010/003773**
WO-A2-2008/006517	**WO-A2-2008/079269**
WO-A2-2009/132928	

- **SORLIE T ET AL: "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES (PNAS), NATIONAL ACADEMY OF SCIENCE, US, vol. 98, no. 19, 11 September 2001 (2001-09-11), pages 10869-10874, XP002215483, ISSN: 0027-8424, DOI: DOI:10.1073/PNAS.191367098 & Sorlie T et al.: "Supporting figure 7", PNAS, vol. 98, no. 19 11 September 2001 (2001-09-11), Retrieved from the Internet: URL:http://www.pnas.org/content/suppl/2001/09/05/98.19.10869.DC1/3670Fig7.pdf [retrieved on 2011-04-26]**
- **PAIK S ET AL: "A MULTIGENE ASSAY TO PREDICT RECURRENCE OF TAMOXIFEN-TREATED, NODE-NEGATIVE BREAST CANCER", NEW ENGLAND JOURNAL OF MEDICINE, MASSACHUSETTS MEDICAL SOCIETY, BOSTON, MA, US, vol. 351, no. 27, 30 December 2004 (2004-12-30), pages 2817-2826, XP008043033, ISSN: 1533-4406, DOI: DOI:10.1056/NEJMOA041588**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Printed by Jouve, 75001 PARIS (FR)

(Cont. next page)

- CHANRION MAÏA ET AL: "A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer.", CLINICAL CANCER RESEARCH : AN OFFICIAL JOURNAL OF THE AMERICAN ASSOCIATION FOR CANCER RESEARCH 15 MAR 2008 LNKD- PUBMED:18347175, vol. 14, no. 6, 15 March 2008 (2008-03-15) , pages 1744-1752, XP002634156, ISSN: 1078-0432
- LOUSSOUARN D ET AL: "Validation of UBE2C protein as a prognostic marker in node-positive breast cancer.", BRITISH JOURNAL OF CANCER 7 JUL 2009 LNKD- PUBMED:19513072, vol. 101, no. 1, 7 July 2009 (2009-07-07), pages 166-173, XP002634157, ISSN: 1532-1827
- J. Wray ET AL: "PARP1 is required for chromosomal translocations", Blood, vol. 121, no. 21, 8 April 2013 (2013-04-08), pages 4359-4365, XP055166409, ISSN: 0006-4971, DOI: 10.1182/blood-2012-10-460527
- Davide F. Robbiani ET AL: "AID Is Required for the Chromosomal Breaks in c-myc that Lead to c-myc/IgH Translocations", Cell, vol. 135, no. 6, 1 December 2008 (2008-12-01), pages 1028-1038, XP055084813, ISSN: 0092-8674, DOI: 10.1016/j.cell.2008.09.062

Description

Technical Field

[0001] The present invention relates to methods, kits and systems for the prognosis of the disease outcome of breast cancer. More specific, the present invention relates to the prognosis of breast cancer based on measurements of the expression levels of marker genes in tumor samples of breast cancer patients.

Background of the Invention

[0002] 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 systemic therapy has led to markedly improved survival in early breast cancer. This clinical experience has led to consensus recommendations offering adjuvant systemic therapy for the vast majority of breast cancer patients (EBCAG). In breast cancer a 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 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 trials 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 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. However, the IBCSG IX study comparing tamoxifen alone against tamoxifen plus chemotherapy failed to show any significant benefit for the addition of cytotoxic agents. 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. Since the benefit of chemotherapy is relatively large in HER2/neu positive and tumors characterized by absence of HER2/neu and estrogen receptor expression (basal type), compared to HER2/neu negative and estrogen receptor positive tumors (luminal type), the most challenging treatment decision concerns luminal tumors for which classical clinical factors like grading, tumor size or lymph node involvement do not provide a clear answer to the question whether to use chemotherapy or not. Newer molecular tools like a 21 gene assay, a genomic grade index assay and others have been developed to address this medical need.

[0003] Treatment guidelines are usually developed by renowned experts in the field. In Europe the St Gallen guidelines from the year 2009 recommend chemotherapy to patients with HER2 positive breast cancer as well as to patients with HER2 negative and ER negative disease. Uncertainty about the usefulness of chemotherapy arises in patients with HER2 negative and ER positive disease. In order to make a balanced treatment decision for the individual the likelihood of cancer recurrence is used as the most useful criteria. Clinical criteria like lypmph node status, tumor grading, tumor size and others are helpful since they provide information about the risk of recurrence. More recently, multigene assays have been shown to provide information superior or additional to the standard clinical risk factors. It is generally recognized, that proliferation markers seem to provide the dominant prognostic information. Prominent examples of those predictors are the Mammaprint test from Agendia, the Relapse Score from Veridex and the Genomic Grade Index, developed at the institute Jules Bordet and licensed to Ipsogen. All of these assays are based on determination of the expression levels of at least 70 genes and all have been developed for RNA not heavily degraded by formalin fixation and paraffin embedding, but isolated from fresh tissue (shipped in RNALaterTM). Another prominent multigene assay is the Recurrence Score test of Genomic Health Inc. The test determines the expression level of 16 cancer related genes and 5 reference genes after RNA extraction from formalin fixed and paraffin embedded tissue samples.

[0004] However, the current tools suffer from a lack of clinical validity and utility in the most important clinical risk group, i.e. those breast cancer patients of intermediate risk of recurrence based on standard clinical parameter. Therefore, better tools are needed to optimize treatment decisions based on patient prognosis. For the clinical utility of avoiding

chemotherapy, a test with a high sensitivity and high negative predictive value is needed, in order not to undertreat a patient that eventually develops a distant metastasis after surgery.
In regard to the continuing need for materials and methods useful in making clinical decisions on adjuvant therapy, the 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

[0005] 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.

[0006] The term "cancer" is not limited to any stage, grade, histomorphological feature, aggressivity, or malignancy of an affected tissue or cell aggregation.

[0007] The term "predicting an outcome" of a disease, as used herein, is meant to include both a prediction of an outcome of a patient undergoing a given therapy and a prognosis of a patient who is not treated. The term "predicting an outcome" may, in particular, relate to the risk of a patient developing metastasis, local recurrence or death.

[0008] The term "prediction", as used herein, relates to an individual assessment of the malignancy of a tumor, or to the expected survival rate (OAS, overall survival or DFS, disease free survival) of a patient, if the tumor is treated with a given therapy. In contrast thereto, the term "prognosis" relates to an individual assessment of the malignancy of a tumor, or to the expected survival rate (OAS, overall survival or DFS, disease free survival) of a patient, if the tumor remains untreated.

[0009] An "outcome" within the meaning of the present invention is a defined condition attained in the course of the disease. This disease outcome may e.g. be a clinical condition such as "recurrence of disease", "development of metastasis", "development of nodal metastasis", development of distant metastasis", "survival", "death", "tumor remission rate", a disease stage or grade or the like.

[0010] A "risk" is understood to be a number related to the probability of a subject or a patient to develop or arrive at a certain disease outcome. The term "risk" in the context of the present invention is not meant to carry any positive or negative connotation with regard to a patient's wellbeing but merely refers to a probability or likelihood of an occurrence or development of a given condition.

[0011] The term "clinical data" relates to the entirety of available data and information concerning the health status of a patient including, but not limited to, age, sex, weight, menopausal/hormonal status, etiopathology data, anamnesis data, data obtained by in vitro diagnostic methods such as histopathology, blood or urine tests, data obtained by imaging methods, such as x-ray, computed tomography, MRI, PET, spect, ultrasound, electrophysiological data, genetic analysis, gene expression analysis, biopsy evaluation, intraoperative findings.

[0012] The term "node positive", "diagnosed as node positive", "node involvement" or "lymph node involvement" means a patient having previously been diagnosed with lymph node metastasis. It shall encompass both draining lymph node, near lymph node, and distant lymph node metastasis. This previous diagnosis itself shall not form part of the inventive method. Rather it is a precondition for selecting patients whose samples may be used for one embodiment of the present invention. This previous diagnosis may have been arrived at by any suitable method known in the art, including, but not limited to lymph node removal and pathological analysis, biopsy analysis, in-vitro analysis of biomarkers indicative for metastasis, imaging methods (e.g. computed tomography, X-ray, magnetic resonance imaging, ultrasound), and intraoperative findings.

[0013] In the context of the present invention a "biological sample" is a sample which is derived from or has been in contact with a biological organism. Examples for biological samples are: cells, tissue, body fluids, lavage fluid, smear samples, biopsy specimens, blood, urine, saliva, sputum, plasma, serum, cell culture supernatant, and others.

[0014] A "tumor sample" is a biological sample containing tumor cells, whether intact or degraded. 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 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.

[0015] A "gene" is a set of segments of nucleic acid that contains the information necessary to produce a functional RNA product. A "gene product" is a biological molecule produced through transcription or expression of a gene, e.g. an mRNA, cDNA or the translated protein.

[0016] An "mRNA" is the transcribed product of a gene and shall have the ordinary meaning understood by a person skilled in the art. A "molecule derived from an mRNA" is a molecule which is chemically or enzymatically obtained from an mRNA template, such as cDNA.

The term "expression level" refers to a determined level of gene expression. This may be a determined level of gene expression as an absolute value or compared to a reference gene (e.g. a housekeeping gene), to the average of two or more reference genes, or to a computed average expression value (e.g. in DNA chip analysis) or to another informative gene without the use of a reference sample. The expression level of a gene may be measured directly, e.g. by obtaining a signal wherein the signal strength is correlated to the amount of mRNA transcripts of that gene or it may be obtained indirectly at a protein level, e.g. by immunohistochemistry, CISH, ELISA or RIA methods. The expression level may also be obtained by way of a competitive reaction to a reference sample. An expression value which is determined by measuring some physical parameter in an assay, e.g. fluorescence emission, may be assigned a numerical value which may be used for further processing of information.

[0017] 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 average pattern of expression levels observed in a group of healthy individuals, diseased individuals, or diseased individuals having received a particular type of therapy, serving as a reference group, or individuals with good or bad outcome.

[0018] The term "mathematically combining expression levels", within the meaning of the invention shall be understood as deriving a numeric value from a determined expression level of a gene and applying an algorithm to one or more of such numeric values to obtain a combined numerical value or combined score.

[0019] An "algorithm" is a process that performs some sequence of operations to produce information.

[0020] A "score" is a numeric value that was derived by mathematically combining expression levels using an algorithm. It may also be derived from expression levels and other information, e.g. clinical data. A score may be related to the outcome of a patient's disease.

[0021] A "discriminant function" is a function of a set of variables used to classify an object or event. A discriminant function thus allows classification of a patient, sample or event into a category or a plurality of categories according to data or parameters available from said patient, sample or event. Such classification is a standard instrument of statistical analysis well known to the skilled person. E.g. a patient may be classified as "high risk" or "low risk", "high probability of metastasis" or "low probability of metastasis", "in need of treatment" or "not in need of treatment" according to data obtained from said patient, sample or event. Classification is not limited to "high vs. low", but may be performed into a plurality of categories, grading or the like. Classification shall also be understood in a wider sense as a discriminating score, where e.g. a higher score represents a higher likelihood of distant metastasis, e.g. the (overall) risk of a distant metastasis. Examples for discriminant functions which allow a classification include, but are not limited to functions defined by support vector machines (SVM), k-nearest neighbors (kNN), (naive) Bayes models, linear regression models or piecewise defined functions such as, for example, in subgroup discovery, in decision trees, in logical analysis of data (LAD) and the like. In a wider sense, continuous score values of mathematical methods or algorithms, such as correlation coefficients, projections, support vector machine scores, other similarity-based methods, combinations of these and the like are examples for illustrative purpose.

[0022] The term "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 cell proliferative agents, and/or radiation therapy, and/or hyperthermia, and/or hypothermia 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 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.

[0023] The term "cytotoxic chemotherapy" 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 antitumor agents, including monoclonal antibodies 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 stability of microtubules, preventing the separation of chromosomes during anaphase.

[0024] The term "endocrine treatment" or "hormonal treatment" (sometimes also referred to as "anti-hormonal treatment") 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 are tamoxifene therapy which modulates signalling of the estrogen receptor, or aromatase treatment which interferes with steroid hormone production.

[0025] Tamoxifen is an orally active selective estrogen receptor modulator (SERM) that is used in the treatment of breast 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 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 RNA synthesis and inhibits estrogen effects.

[0026] 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 receptor, progesterone receptor; Glucocorticoid receptor, mineralocorticoid receptor; and type II Receptors, e.g. vitamin A receptor, vitamin D receptor, retinoid receptor, thyroid hormone receptor.

[0027] 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 used 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.

[0028] An oligonucleotide capable of specifically binding sequences a gene or fragments thereof relates to an oligonucleotide which specifically hybridizes to a gene or gene product, such as the gene's mRNA or cDNA or to a fragment thereof. To specifically detect the gene or gene product, it is not necessary to detect the entire gene sequence. A fragment of about 20-150 bases will contain enough sequence specific information to allow specific hybridization.

[0029] The term "a PCR based method" as used herein refers to methods comprising a polymerase chain reaction (PCR). This is a method of exponentially amplifying nucleic acids, e.g. DNA by enzymatic replication in vitro. 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).

[0030] Moreover, PCR-based methods comprise e.g. real time PCR, and, particularly suited for the analysis of expression levels, kinetic or quantitative PCR (qPCR).

[0031] The term "Quantitative PCR" (qPCR)" refers to any type of a PCR method which allows the quantification of the template in a sample. Quantitative real-time PCR comprise different techniques of performance or product detection as for example the TaqMan technique or the LightCycler technique. The TaqMan technique, for examples, uses a dual-labelled fluorogenic probe. The TaqMan real-time PCR measures accumulation of a product via the fluorophore during the exponential stages of the PCR, rather than at the end point as in conventional PCR. The exponential increase of the product is used to determine the threshold cycle, CT, i.e. the number of PCR cycles at which a significant exponential increase in fluorescence is detected, and which is directly correlated with the number of copies of DNA template present in the reaction. The set up of the reaction is very similar to a conventional PCR, but is carried out in a real-time thermal cycler that allows measurement of fluorescent molecules in the PCR tubes. Different from regular PCR, in TaqMan real-time PCR a probe is added to the reaction, i.e., a single-stranded oligonucleotide complementary to a segment of 20-60 nucleotides within the DNA template and located between the two primers. A fluorescent reporter or fluorophore (e.g., 6-carboxyfluorescein, acronym: FAM, or tetrachlorofluorescin, acronym: TET) and quencher (e.g., tetramethylrhodamine, acronym: TAMRA, of dihydrocyclopyrroloindole tripeptide 'black hole quencher', acronym: BHQ) are covalently attached to the 5' and 3' ends of the probe, respectively[2]. The close proximity between fluorophore and quencher attached to the probe inhibits fluorescence from the fluorophore. During PCR, as DNA synthesis commences, the 5' to 3' exonuclease activity of the Taq polymerase degrades that proportion of the probe that has annealed to the template. Degradation of the probe releases the fluorophore from it and breaks the close proximity to the quencher, thus relieving the quenching effect and allowing fluorescence of the fluorophore. Hence, fluorescence detected in the real-time PCR thermal cycler is directly proportional to the fluorophore released and the amount of DNA template present in the PCR.

[0032] By "array" or "matrix" an arrangement of addressable locations or "addresses" on a device is meant. The locations can be arranged in two dimensional arrays, three dimensional arrays, or other matrix formats. The number of locations can range from several to at least hundreds of thousands. Most importantly, each location represents a totally 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, nucleotide analogues, polynucleotides, polymers of nucleotide analogues, morpholinos or larger portions of genes. The nucleic acid and/or analogue on the array is preferably single stranded. Arrays wherein the probes are oligonucleotides are referred to as "oligo-nucleotide arrays" or "oligonucleotide chips." A "microarray," herein also refers to a "biochip" or "biological chip", an array of regions having a density of discrete regions of at least about 100/cm2, and preferably at least about 1000/cm2.

[0033] "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 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 technologies used for kinetic or real time PCR applications could be e.g. TaqMan® systems obtainable at Applied Biosystems, extension probes such as Scorpion® Primers, Dual Hybridisation Probes, Amplifluor® obtainable at Chemicon International, Inc, or Minor Groove Binders.

[0034] "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 quantification of the probe. Preferred labels are fluorescent molecules, luminescent molecules, radioactive molecules, enzymatic molecules and/or quenching molecules.

[0035] "Arrayed probes", within the meaning of the invention, shall 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".
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.

Summary of the Invention

[0036] In general terms, the present invention provides a method to assess the risk of recurrence of a node negative or positive, estrogen receptor positive and HER2/NEU negative breast cancer patient, in particular patients receiving endocrine therapy, for example when treated with tamoxifen. Estrogen receptor status is generally determined using immunohistochemistry, HER2/NEU (ERBB2) status is generally determined using immunohistochemistry and fluorescence in situ hybridization. However, estrogen receptor status and HER2/NEU (ERBB2) status may, for the purposes of the invention, be determined by any suitable method, e.g. immunohistochemistry, fluorescence in situ hybridization (FISH), or RNA expression analysis.

[0037] The present invention relates to a method for predicting an outcome of breast cancer in an estrogen receptor positive and HER2 negative tumor of a breast cancer patient, said method comprising:

(a) determining in a tumor sample from said patient the RNA expression levels of at least 2 of the following 9 genes: UBE2C, BIRC5, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP

(b) mathematically combining expression level values for the genes of the said set which values were determined in the tumor sample to yield a combined score, wherein said combined score is indicative of a prognosis of said patient. In one embodiment at least 3, 4, 5 or 6 genes are selected.

[0038] In a further embodiment of the invention the method comprises:

(a) determining in a tumor sample from said patient the RNA expression levels of the following 8 genes: UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP
(b) mathematically combining expression level values for the genes of the said set which values were determined in the tumor sample to yield a combined score, wherein said combined score is indicative of a prognosis of said patient.

[0039] In a further embodiment the method of the invention comprises:

(a) determining in a tumor sample from said patient the RNA expression levels of the following 8 genes: UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP;

(b) mathematically combining expression level values for the genes of the said set which values were determined in the tumor sample to yield a combined score, wherein said combined score is indicative of a prognosis of said patient.

In yet another embodiment of the invention

[0040] **BIRC5** may be replaced by UBE2C or TOP2A or RACGAP1 or AURKA or NEK2 or E2F8 or PCNA or CYBRD1 or DCN or ADRA2A or SQLE or CXCL12 or EPHX2 or ASPH or PRSS16 or EGFR or CCND1 or TRIM29 or DHCR7 or PIP or TFAP2B or WNT5A or APOD or PTPRT with the proviso that after a replacement 8 different genes are selected; and

[0041] **UBE2C** may be replaced by BIRC5 or RACGAP1 or TOP2A or AURKA or NEK2 or E2F8 or PCNA or CYBRD1 or ADRA2A or DCN or SQLE or CCND1 or ASPH or CXCL12 or PIP or PRSS16 or EGFR or DHCR7 or EPHX2 or TRIM29 with the proviso that after a replacement 8 different genes are selected; and **DHCR7** may be replaced by AURKA, BIRC5, UBE2C or by any other gene that may replace BIRC5 or UBE2C with the proviso that after a replacement 8 different genes are selected; and

[0042] **STC2** may be replaced by INPP4B or IL6ST or SEC14L2 or MAPT or CHPT1 or ABAT or SCUBE2 or ESR1 or RBBP8 or PGR or PTPRT or HSPA2 or PTGER3 with the proviso that after a replacement 8 different genes are selected; and **AZGP1** may be replaced by PIP or EPHX2 or PLAT or SEC14L2 or SCUBE2 or PGR with the proviso that after a replacement 8 different genes are selected; and

[0043] **RBBP8** may be replaced by CELSR2 or PGR or STC2 or ABAT or IL6ST with the proviso that after a replacement 8 different genes are selected; and **IL6ST** may be replaced by INPP4B or STC2 or MAPT or SCUBE2 or ABAT or PGR or SEC14L2 or ESR1 or GJA1 or MGP or EPHX2 or RBBP8 or PTPRT or PLAT with the proviso that after a replacement 8 different genes are selected; and

[0044] **MGP** may be replaced by APOD or IL6ST or EGFR with the proviso that after a replacement 8 different genes are selected.

[0045] According to an aspect of the invention there is provided a method as described above, wherein said combined score is indicative of benefit from cytotoxic chemotherapy.

[0046] Using the method of the invention before a patient receives endocrine therapy allows a prediction of the efficacy of endocrine therapy.

[0047] Table 2 below shows whether the overexpression of each of the above marker genes is indicative of a good outcome or a bad outcome in a patient receiving endocrine therapy. The skilled person can thus construct a mathematical combination i.e. an algorithm taking into account the effect of a given genes. For example a summation or weighted summation of genes whose overexpression is indicative of a good outcome results in an algorithm wherein a high risk score is indicative of a good outcome. The validity of the algorithm may be examined by analyzing tumor samples of patients with a clinical record, wherein e.g. the score for good outcome patients and bad outcome patients may be determined separately and compared. The skilled person, a biostatistician, will know to apply further mathematical methods, such as discriminate functions to obtain optimized algorithms. Algorithms may be optimized e.g. for sensitivity or specificity. Algorithms may be adapted to the particular analytical platform used to measure gene expression of marker genes, such as quantitiative PCR.

[0048] According to an aspect of the invention there is provided a method as described above, wherein said endocrine therapy comprises tamoxifen or an aromatase inhibitor.

[0049] According to an aspect of the invention there is provided a method as described above, wherein a risk of developing recurrence is predicted.

[0050] According to an aspect of the invention there is provided a method as described above, wherein said expression level is determined as a non-protein expression level.

[0051] According to an aspect of the invention there is provided a method as described above, wherein said expression level is determined as an RNA expression level.

[0052] According to an aspect of the invention there is provided a method as described above, wherein said expression level is determined by at least one of

a PCR based method,
a micorarray based method, and
a hybridization based method.

[0053] According to an aspect of the invention there is provided a method as described above, wherein said determination of expression levels is in a formalin-fixed paraffin embedded tumor sample or in a fresh-frozen tumor sample.

[0054] According to an aspect of the invention there is provided a method as described above, wherein the expression level of said at least on marker gene is determined as a pattern of expression relative to at least one reference gene or to a computed average expression value.

[0055] According to an aspect of the invention there is provided a method as described above, wherein said step of mathematically combining comprises a step of applying an algorithm to values representative of an expression level of a given gene.

[0056] According to an aspect of the invention there is provided a method as described above, wherein said algorithm is a linear combination of said values representative of an expression level of a given gene.

[0057] According to an aspect of the invention there is provided a method as described above, wherein a value for a representative of an expression level of a given gene is multiplied with a coefficient.

[0058] According to an aspect of the invention there is provided a method as described above, wherein one, two or more thresholds are determined for said combined score and discriminated into high and low risk, high, intermediate and low risk, or more risk groups by applying the threshold on the combined score.

[0059] According to an aspect of the invention there is provided a method as described above, wherein a high combined score is indicative of benefit from a more aggressive therapy, e.g. cytotoxic chemotherapy. The skilled person understands that a "high score" in this regard relates to a reference value or cutoff value. The skilled person further understands that depending on the particular algorithm used to obtain the combined score, also a "low" score below a cut off or reference value can be indicative of benefit from a more aggressive therapy, e.g. cytotoxic chemotherapy. This is the case when genes having a positive correlation with high risk of metastasis factor into the algorithm with a positive coefficient, such that an overall high score indicates high expression of genes having a positive correlation with high risk.

[0060] According to an aspect of the invention there is provided a method as described above, wherein an information regarding nodal status of the patient is processed in the step of mathematically combining expression level values for the genes to yield a combined score.

[0061] According to an aspect of the invention there is provided a method as described above, wherein said information regarding nodal status is a numerical value ≤ 0 if said nodal status is negative and said information is a numerical value > 0 if said nodal status positive or unknown. In exemplary embodiments of the invention a negative nodal status is assigned the value 0, an unknown nodal status is assigned the value 0.5 and a positive nodal status is assigned the value 1. Other values may be chosen to reflect a different weighting of the nodal status within an algorithm.

[0062] The invention further relates to a kit for performing a method as described above, said kit comprising a set of oligonucleotides capable of specifically binding sequences or to seqences of fragments of the genes in a combination of genes, wherein

(i) said combination comprises at least the 8 genes UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP; or
(ii) said combination comprises at least the 10 genes BIRC5, AURKA, PVALB, NMU, STC2, RBBP8, PTGER3, CXCL12, CDH1, and PIP; or
(iii) said combination comprises at least the 9 genes BIRC5, DHCR7, RACGAP1, PVALB, STC2, IL6ST, PTGER3, CXCL12, and ABAT; or
(iv) said combination comprises at least the 9 genes DHCR7, RACGAP1, NMU, AZGP1, RBBP8, IL6ST, and MGP;

[0063] The invention further relates to the use of a kit for performing a method of any of claims 1 to 17, said kit comprising a set of oligonucleotides capable of specifically binding sequences or to sequences of fragments of the genes in a combination of genes, wherein

(i) said combination comprises at least the 8 genes UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP; or
(ii) said combination comprises at least the 10 genes BIRC5, AURKA, PVALB, NMU, STC2, RBBP8, PTGER3, CXCL12, CDH1, and PIP; or
(iii) said combination comprises at least the 9 genes BIRC5, DHCR7, RACGAP1, PVALB, STC2, IL6ST, PTGER3, CXCL12, and ABAT; or
(iv) said combination comprises at least the 9 genes DHCR7, RACGAP1, NMU, AZGP1, RBBP8, IL6ST, and MGP;19. A computer program product capable of processing values representative of an expression level of the genes AKR1C3, MAP4 and SPP1 by mathematically combining said values to yield a combined score, wherein said combined score is indicative of benefit from cytotoxic chemotherapy of said patient.

[0064] The invention further relates to a computer program product capable of processing values representative of an expression level of a combination of genes mathematically combining said values to yield a combined score, wherein said combined score is indicative of efficacy or benefit from endocrine therapy of said patient, according to the above methods.

[0065] Said computer program product may be stored on a data carrier or implemented on a diagnostic system capable of outputting values representative of an expression level of a given gene, such as a real time PCR system.

[0066] If the computer program product is stored on a data carrier or running on a computer, operating personal can input the expression values obtained for the expression level of the respective genes. The computer program product can then apply an algorithm to produce a combined score indicative of benefit from cytotoxic chemotherapy for a given patient.

[0067] The methods of the present invention have the advantage of providing a reliable prediction of an outcome of disease based on the use of only a small number of genes. The methods of the present invention have been found to be especially suited for analyzing the response to endocrine treatment, e.g. by tamoxifen, of patients with tumors classified as ESR1 positive and ERBB2 negative.

Detailed description of the invention

[0068] The invention is explained in conjunction with exemplary embodiments and the attached figures:

Figure 1 shows a Forrest Plot of the adjusted hazard unit ratio with 95% confidence intervall of the T5 score in the combined cohort, as well as the individual treatment arms of the ABCSG06 and 08 studies, using distant metastasis as endpoint.

Figure 2 shows a Kaplan Meier Analysis of ER+, HER-, N0-3 patients from the combined ABCSG06 and 08 cohorts, stratified as high or low risk according to T5 Score value.

[0069] Herein disclosed are unique combinations of marker genes which can be combined into an algorithm for the here presented new predictive test. Technically, the method of the invention can be practiced using two technologies: 1.) Isolation of total RNA from fresh or fixed tumor tissue and 2.) Kinetic RT-PCR of the isolated nucleic acids. Alternatively, it is contemplated to measure expression levels using alternative technologies, e.g by microarray or by measurement at a protein level.

[0070] The methods of the invention are 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. RNA species might be isolated from any type of tumor sample, e.g. biopsy samples, smear samples, resected tumor material, fresh frozen tumor tissue or from paraffin embedded and formalin fixed tumor tissue. First, RNA levels of genes coding for specific combinations of the genes UBE2C, BIRC5, DHCR7, RACGAP1, AURKA, PVALB, NMU, STC2, AZGP1, RBBP8, IL6ST, MGP, PTGER3, CXCL12, ABAT, CDH1, and PIP or specific combinations thereof, as indicated, are determined. Based on these expression values a prognostic score is calculated by a mathematical combination, e.g. according to formulas T5 T1, T4, or T5b (see below). A high score value indicates a high risk for development of distant metastasis, a low score value indicates a low risk of distant metastasis. Consequently, a high score also indicates that the patient is a high risk patient who will benefit from a more aggressive therapy, e.g. cytotoxic chemotherapy.

[0071] The present examples are based on identification of prognostic genes using tumors of patients homogeneously treated in the adjuvant setting with tamoxifen. Furthermore, identification of relevant genes has been restricted to tumors classified as ESR1 positive and ERBB2 negative based on RNA expression levels. In addition, genes allowing separation of intermediate risk, e.g. grade 2 tumors were considered for algorithm development. Finally, a platform transfer from Affymetrix HG_U133a arrays to quantitative real time PCR, as well as a sample type transfer from fresh frozen tissue to FFPE tissue was performed to ensure robust algorithm performance, independent from platform and tissue type. As a result, determination of the expression level of RNA species from the primary tumor and the subsequent complex and multivariate analysis as described above provides a superior method for prediction of the likelihood of disease recurrence in patients diagnosed with lymph node negative or positive early breast cancer, when treated with tamoxifen only in the adjuvant setting. Thus the test relies on fewer genes than those of the competitors but provides superior information regarding high sensitivity and negative predictive value, in particular for tumors considered to exhibit an intermediate risk of recurrence based on standard clinical factors.

[0072] The total RNA was extracted with a Siemens, silica bead-based and fully automated isolation method for RNA from one 10 μm whole FFPE tissue section on a Hamilton MICROLAB STARlet liquid handling robot (17). The robot, buffers and chemicals were part of a Siemens VERSANT® kPCR Molecular System (Siemens Healthcare Diagnostics, Tarrytown, NY; not commercially available in the USA). Briefly, 150 μl FFPE buffer (Buffer FFPE, research reagent, Siemens Healthcare Diagnostics) were added to each section and incubated for 30 minutes at 80°C with shaking to melt the paraffin. After cooling down, proteinase K was added and incubated for 30 minutes at 65°C. After lysis, residual tissue debris was removed from the lysis fluid by a 15 minutes incubation step at 65°C with 40 μl silica-coated iron oxide beads. The beads with surface-bound tissue debris were separated with a magnet and the lysates were transferred to a standard 2 ml deep well-plate (96 wells). There, the total RNA and DNA was bound to 40 μl unused beads and incubated at room temperature. Chaotropic conditions were produced by the addition of 600 μl lysis buffer. Then, the beads were magnetically separated and the supernatants were discarded. Afterwards, the surface-bound nucleic acids were washed three times followed by magnetization, aspiration and disposal of supernatants. Afterwards, the nucleic acids were eluted by incubation of the beads with 100 μl elution buffer for 10 minutes at 70°C with shaking. Finally, the beads were separated and the supernatant incubated with 12 μl DNase I Mix (2 μL DNase I (RNase free); 10 μl 10x DNase I buffer; Ambion/Applied Biosystems, Darmstadt, Germany) to remove contaminating DNA. After incubation for 30 minutes at 37°C, the DNA-free total RNA solution was aliquoted and stored at -80°C or directly used for mRNA expression analysis by reverse transcription kinetic PCR (RTkPCR). All the samples were analyzed with one-step RT-kPCR for the gene expression of up to three reference genes, (RPL37A, CALM2, OAZ1) and up to 16 target genes in an ABI PRISM® 7900HT (Applied Biosystems, Darmstadt, Germany). The superscript® III Platinum® One-Step Quantitative RT-PCR System with ROX (6- carboxy-X-rhodamine) (Invitrogen, Karlsruhe, Germany) was used according to the manufacturer's instructions. Respective probes and primers are shown in table 1. The PCR conditions were as follows: 30 minutes at 50°C, 2 minutes at 95°C followed by 40 cycles of 15 seconds at 95°C and 30 seconds at 60°C. All the PCR assays were performed in triplicate. As surrogate marker for RNA yield, the housekeeper gene, RPL37A cycle threshold (Ct) value was used as described elsewhere (17). The relative gene expression levels of the target genes were calculated by the delta-Ct method using the formula:

$$20 - (Ct(target) - mean(Ct(reference\ genes))).$$

[0073] A platform transfer from Affymetrix HG_U133a arrays (fresh frozen tissue) to quantitative real time PCR (FFPE tissue) was calculated as follows. Material from 158 patients was measured using both platforms to yield paired samples. Delta-Ct values were calculated from the PCR data. Log2-Expressions were calculated from the Affymetrix data by applying a lower bound (setting all values below the lower bound to the lower bound) and then calculating the logarithm of base 2. The application of a lower bound reduces the effect of increased relative measurement noise for low expressed genes/samples; a lower bound of 20 was used, lower bounds between 0.1 and 200 also work well. A HG_U133a probe set was selected for each PCR-measured gene by maximizing the Pearson correlation coefficient between the delta-Ct value (from PCR) and the log2-expression (from Affymetrix). Other correlation measures will also work well, e.g. the Spearman correlation coefficient. In most cases the best-correlating probe set belonged to the intended gene, for the remaining cases the PCR-gene was removed for further processing. Those genes showing a bad correlation between platforms were also removed, where a threshold on the Pearson correlation coefficient of 0.7 was used (values of between 0.5 and 0.8) also work well. The platform transformation was finalized by calculating unsupervised z-transformations for both platforms and combining them; a single PCR-delta-Ct value then is transformed to the Affymetrix scale by the following steps: (i) apply affine linear transformation where coefficients were determined by z-transformation of PCR data, (ii) apply inverse affine linear transformation where coefficients were determined by z-transformation of Affymetrix data, (iii) invert log2, i.e. calculate exponential with respect to base 2. Alternatives to the two-fold z-transformations are linear or higher order regression, robust regression or principal component based methods, which will also work well.

[0074] The sequences of the primers and probes were as follows:

Table 1 Primer and probe sequences for the respective genes:

gene	probe	Seq ID	forward primer	Seq ID	reverse primer	Seq ID
ABAT	TCGCCCTAAGAGGCTCTTCCTC	1	GGCAACTTGAGGTCTGACTTTTG	2	GGTCAGCTCACAAGTGGTGTGA	3
ADRA2A	TTGTCCTTTCCCCCCTCCGTGC	4	CCCCAAGAGCTGTTAGGTATCAA	5	TCAATGACATGATCTCAACCAGAA	6
APOD	CATCAGCTCTCAACTCCTGGTTTAACA	7	ACTCACTAATGGAAAACGGAAAGATC	8	TCACCTTCGATTTGATTCACAGTT	9
ASPH	TGGGAGGAAGGCAAGGTGCTCATC	10	TGTGCCAACGAGACCAAGAC	11	TCGTGCTCAAAGGAGTCATCA	12
AURKA	CCGTCAGCCTGTGCTAGGCAT	13	AATCTGGAGGCAAGGTTCGA	14	TCTGGATTTGCCTCCTGTGAA	15
BIRC5	AGCCAGATGACGACCCCATAGAGGAACA	16	CCCAGTGTTTCTTCTGCTTCAAG	17	CAACCGGACGAATGCTTTTT	18
CELSR2	ACTGACTTTCCTTCTGGAGCAGGTGGC	19	TCCAAGCATGTATTCCAGACTTGT	20	TGCCCACAGCCTCTTTTTCT	21
CHPT1	CCACGGCCACCGAAGAGGCAC	22	CGCTCGTGCTCATCTCCTACT	23	CCCAGTGCACATAAAAGGTATGTC	24
CXCL12	CCACAGCAGGGTTTCAGGTTCC	25	GCCACTACCCCCTCCTGAA	26	TCACCTTGCCAACAGTTCTGAT	27
CYBRD1	AGGGCATCGCCATCATCGTC	28	GTCACCGGCTTCGTCTTCA	29	CAGGTCCACGGCAGTCTGT	30
DCN	TCTTTTCAGCAACCCGGTCCA	31	AAGGCTTCTTATTCGGGTGTGA	32	TGGATGGCTGTATCTCCCAGTA	33
DHCR7	TGAGCGCCCACCCTCTCGA	34	GGGCTCTGCTTCCCGATT	35	AGTCATAGGGCAAGCAGAAAATTC	36
E2F8	CAGGATACCTAATCCCTCTCACGCAG	37	AAATGTCTCCGCAACCTTGTTC	38	CTGCCCCCAGGGATGAG	39
EPHX2	TGAAGCGGGAGGACTTTTTGTAAA	40	CGATGAGAGTGTTTTATCCATGCA	41	GCTGAGGCTGGGCTCTTCT	42
ESR1	ATGCCCTTTTGCCGATGCA	43	GCCAAATTGTGTTTGATGGATTAA	44	GACAAAACCGAGTCACATCAGTAATAG	45
GJA1	TGCACAGCCTTTTGATTTCCCCGAT	46	CGGGAAGCACCATCTCTAACTC	47	TTCATGTCCAGCAGCTAGTTTTTT	48
HSPA2	CAAGTCAGCAAACACGCAAAA	49	CATGCACGAACTAATCAAAAATGC	50	ACATTATTCGAGGTTTCTCTTTAATGC	51

(continued)

gene	probe	Seq ID	forward primer	Seq ID	reverse primer	Seq ID
IL6ST	CAAGCTCCACCTTCCAAAGGACCT	52	CCCTGAATCCATAAAGGCATACC	53	CAGCTTCGTTTTTCCCTACTTTTT	54
INPP4B	TCCGAGCGCTGGATTGCATGAG	55	GCACCAGTTACACAAGGACTTCTTT	56	TCTCTATGCGGCATCCTTCTC	57
MAPT	AGACTATTTGCACACTGCCGCCT	58	GTGGCTCAAAGGATAATATCAAACAC	59	ACCTTGCTCAGGTCAACTGGTT	60
MGP	CCTTCATATCCCCTCAGCAGAGATGG	61	CCTTCATTAACAGGAGAAATGCAA	62	ATTGAGCTCGTGGACAGGCTTA	63
NEK2	TCCTGAACAAATGAATCGCATGTCCTACAA	64	ATTTGTTGGCACACCTTATTACATGT	65	AAGCAGCCCAATGACCAGATa	66
PCNA	AAATACTAAAATGCGCCGGCAATGA	67	GGGCGTGAACCTCACCAGTA	68	CTTCGGCCCTTAGTGTAATGATATC	69
PGR	TTGATAGAAACGCTGTGAGCTCGA	70	AGCTCATCAAGGCAATTGGTTT	71	ACAAGATCATGCAAGTTATCAAGAAGTT	72
PIP	TGCATGGTGGTTAAAACTTACCTCA	73	TGCTTGCAGTTCAAACAGAATTG	74	CACCTTGTAGAGGGATGCTGCTA	75
PLAT	CAGAAAGTGGCCATGCCACCCTG	76	TGGGAAGACATGAATGCACACTA	77	GGAGGTTGGGCTTTAGCTGAA	78
PRSS16	CACTGCCGGTCACCCACACCA	79	CTGAGGAGCACAGAACCTCAACT	80	CGAACTCGGTACATGTCTGATACAA	81
PTGER3	TCGGTCTGCTGGTCTCCGCTCC	82	CTGATTGAAGATCATTTTCAACATCA	83	GACGGCCATTCAGCTTATGG	84
PTPRT	TTGGCTTCTGGACACCCTCACA	85	GAGTTGTGGCCTCTACCATTGC	86	GAGCGGGAACCTTGGGATAG	87
RACGAP1	ACTGAGAATCTCCACCCGGCGCA	88	TCGCCAACTGGATAAATTGGA	89	GAATGTGCGGAATCTGTTTGAG	90
RBBP8	ACCGATTCCGCTACATTCCACCCAAC	91	AGAAATTGGCTTCCTGCTCAAG	92	AAAACCAACTTCCCAAAAATTCTCT	93
SCUBE2	CTAGAGGGTTCCAGGTCCCATACGTGACATA	94	TGTGGATTCAGTTCAAGTCCAATG	95	CCATCTCGAACTATGTCTTCAATGAGT	96
SEC14L2	TGGGAGGCATGCAACGCGTG	97	AGGTCTTACTAAGCAGTCCCATCTCT	98	CGACCGGCACCTGAACTC	99
SQLE	TATGCGTCTCCCAAAAGAAGAACACCTCG	100	GCAAGCTTCCTTCCTCCTTCA	101	CCTTTAGCAGTTTTCTCCATAGTTTTATATC	102

(continued)

gene	probe	Seq ID	forward primer	Seq ID	reverse primer	Seq ID
TFAP2B	CAACACCACCACTAACAGGCACACGTC	103	GGCATGGACAAGATGTTCTTGA	104	CCTCCTTGTCGCCAGTTTTACT	105
TOP2A	CAGATCAGGACCAAGATGGTTCCCACAT	106	CATTGAAGACGCTTCGTTATGG	107	CCAGTTGTGATGGATAAAATTAATCAG	108
TRIM29	TGCTGTCTCACTACCGGCCATTCTACG	109	TGGAAATCTGGCAAGCAGACT	110	CAATCCCGTTGCCTTTGTTG	111
UBE2C	TGAACACACATGCTGCCGAGCTCTG	112	CTTCTAGGAGAACCCAACATTGATAGT	113	GTTTCTTGCAGGTACTTCTTAAAAGCT	114
WNT5A	TATTCACATCCCCTCAGTTGCAGTGAATTG	115	CTGTGGCTCTTAATTTATTGCATAATG	116	TTAGTGCTTTTTGCTTTCAAGATCTT	117
STC2	TCTCACCTTGACCCTCAGCCAAG	118	ACATTTGACAAATTTCCCTTAGGATT	119	CCAGGACGCAGCTTTACCAA	120
AZGP1	*CACCAGCCACCAGGCCCCAG*	121	*TCCTGGACCGGCAAGATC*	122	*TAGGCCAGGCACTTCAGTTTC*	*123*
CALM2	*TCGCGTCTCGGAAACCGGTAGC*	124	*GAGCGAGCTGAGTGGTTGTG*	125	*AGTCAGTTGGTCAGCCATGCT*	*126*
CDH1	*CCTGCCAATCCCGATGAAATTGGAAAT*	127	*TGAGTGTCCCCCGGTATCTTC*	*128*	*TCAGCCGCTTTCAGATTTTCA*	129
NMU	ACCCTGCTGACCTTCTTCCATTCCGT	130	AGAAATTGGCTTCCTGCTCAAG	131	AAAACCAACTTCCCAAAAATTCTCT	132
OAZ1	*TGCTTCCACAAGAACCGCGAGGA*	*133*	*CGAGCCGACCATGTCTTCAT*	*134*	*AAGCCCAAAAAGCTGAAGGTT*	*135*
PVALB	*AAGTTCTTCCAAATGGTCGGCC*	*136*	*CCGACTCCTTCGACCACAA*	137	*CATCATCCGCACTCTTTTTCTTC*	*138*
RPL37A	*TGGCTGGCGGTGCCTGGA*	*139*	*TGTGGTTCCTGCATGAAGACA*	*140*	*GTGACAGCGGAAGTGGTATTGTAC*	141

[0075] Table 2, below, lists the genes used in the methods of the invention and in the particular embodiments T5, T1, T4, and T5b. Table 2 also shows whether overexpression of a given gene is indicative of good or bad outcome under Tamoxifen therapy., Table 2 lists the function of the gene, the compartment localization within the cell and the cellular processes it is involved in.

Table 2 List of genes of algorithms T5, T1, T4, and T5b:

Gene	Name	High Expression	Function	Component	Process
UBE2C	ubiquitin-conjugating enzyme E2C	Bad Outcome	ATP binding	cytosol	cell division
BIRC5	IAP repeat-containing 5	Bad Outcome	Ran GTPase binding	cytosol	cell cycle
DHCR7	7-dehydrochol esterol reductase	Bad Outcome	7-dehydrocholesterol reductase activity	endoplasmatic reticulum membrane	regulation of cell proliferation
RACGAP1	Rac GTPase activating protein 1	Bad Outcome	GTPase activator activity	cytoplasm	cell cycle
AURKA	aurora kinase A	Bad Outcome	ATP binding	centrosome	mitotic cell cycle
PVALB	parvalbumin	Bad Outcome	calcium ion binding		
NMU	neuromedin U	Bad Outcome	receptor binding	extracellular region	signal transduction
STC2	stanniocalcin 2	Good Outcome	hormone activity	extracellular region	cell surface receptor linked signal transduction
AZGP1	alpha-2-glycoprotein 1	Good Outcome	protein transmembrane transporter activity	extracellular region	negative regulation of cell proliferat ion
RBBP8	retinoblast oma binding protein 8	Good Outcome	protein binding	nucleus	cell cycle checkpoint
IL6ST	interleukin 6 signal transducer	Good Outcome	receptor activity	extracellular region	signal transduction
MGP	matrix Gla protein	Good Outcome	extracellular matrix structural constituent	extracellular region	cell differentiation
PTGER3	prostagland in E receptor 3	Good Outcome	ligand-dependent receptor activity	plasma membrane	signal transduction
CXCL12	chemokine (C-XC motif) ligand 12	Good Outcome	chemokine activity	extracellular region	signal transduction
ABAT	4-aminobutyrate aminotransf erase	Good Outcome	transferase activity	mitochondrion	gamma-aminobutyric acid catabolic process
CDH1	cadherin 1	Good Outcome	cell adhesion molecule binding	plasma membrane	homophilic cell adhesion
PIP	prolactin-induced protein	Good Outcome	actin bindin	extracellular region	
CALM2					Reference Gene
OAZ1					Reference Gene

(continued)

Gene	Name	High Expression	Function	Component	Process
RPL37A					Reference Gene

[0076] Table 3, below, shows the combinations of genes used for each algorithm.

Table 3: Combination of genes for the respective algorithms:

Gene	Algo_T1	Algo_T4	Algo_T5	Algo_T5b
UBE2C			X	
BIRC5	X	X	X	
DHCR7		X	X	X
RACGAP1		X		X
AURKA	X			
PVALB	X	X		
NMU	X			X
STC2	X	X	X	
AZGP1			X	X
RBBP8	X		X	X
IL6ST		X	X	X
MGP			X	X
PTGER3	X	X		
CXCL12	X	X		
ABAT		X		
CDH1	X			
PIP	X			

[0077] Table 4, below, shows Affy probeset ID and TaqMan design ID mapping of the marker genes of the present invention.

Table 4: Gene symbol, Affy probeset ID and TaqMan design ID mapping:

Gene	Design ID	Probeset ID
UBE2C	R65	202954_at
BIRC5	SC089	202095_s_at
DHCR7	CAGMC334	201791_s_at
RACGAP1	R125-2	222077_s_at
AURKA	CAGMC336	204092_s_at
PVALB	CAGMC339	205336_at
NMU	CAGMC331	206023_at
STC2	R52	203438_at
AZGP1	CAGMC372	209309_at
RBBP8	CAGMC397	203344_s_at

(continued)

Gene	Design ID	Probeset ID
IL6ST	CAGMC312	212196_at
MGP	CAGMC383	202291_s_at
PTGER3	CAGMC315	213933_at
CXCL12	CAGMC342	209687_at
ABAT	CAGMC338	209460_at
CDH1	CAGMC335	201131_s_at

[0078] Table 5, below, shows full names, Entrez GeneID, gene bank accession number and chromosomal location of the marker genes of the present invention

Official Symbol	Official Full Name	Entrez GeneID	Accesion Number	Location
UBE2C	ubiquitin-conjugating enzyme E2C	11065	U73379	20q13.12
BIRC5	baculoviral IAP repeat-containing 5	332	U75285	17q25
DHCR7	7-dehydrocholesterol reductase	1717	AF034544	11q13.4
STC2	staniocalcin 2	8614	AB012664	5q35.2
RBBP8	retinoblastoma binding protein 8	5932	AF043431	18q11. 2
IL6ST	interleukin 6 signal transducer	3572	M57230	5q11
MGP	matrix Gla protein	4256	M58549	12p12.3
AZGP1	alpha-2-glycoprotein 1, zinc-binding	563	BC005306	11q22.1
RACGAP1	Rac GTPase activating protein 1	29127	NM_013277	12q13
AURKA	aurora kinase A	6790	BC001280	20q13
PVALB	parvalbumin	5816	NM_002854	22q13.1
NMU	neuromedin U	10874	X76029	4q12
PTGER3	prostaglandin E receptor 3 (subtype EP3)	5733	X83863	1p31.2
CXCL12	chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)	6387	L36033	10q11.1
ABAT	4-aminobutyrat aminotransferase	18	L32961	16p13.2
CDH1	cadherin 1, type 1, E-cadherin (epithelial)	999	L08599	16q22.1
PIP	prolactin-induced protein	5304	NMM_002652	7q32-qter

[0079] Example algorithm T5:

Algorithm T5 is a committee of four members where each member is a linear combination of two genes. The mathematical formulas for T5 are shown below; the notation is the same as for T1. T5 can be calculated from gene expression data only.

$$\begin{aligned}\text{riskMember1} = {} & 0.434039\ [0.301..0.567] * (0.939 * \text{BIRC5} -3.831) \\ & -0.491845\ [-0.714..-0.270] * (0.707 * \text{RBBP8} -0.934)\end{aligned}$$

$$\begin{aligned}\text{riskMember2} = {} & 0.488785\ [0.302..0.675] * (0.794 * \text{UBE2C} -1.416) \\ & -0.374702\ [-0.570..-0.179] * (0.814 * \text{IL6ST} -5.034)\end{aligned}$$

```
riskMember3 = -0.39169 [-0.541..-0.242] * (0.674 * AZGP1 -0.777)
              +0.44229 [0.256..0.628] * (0.891 * DHCR7 -4.378)

riskMember4 = -0.377752 [-0.543..-0.212] * (0.485 * MGP +4.330)
              -0.177669 [-0.267..-0.088] * (0.826 * STC2 -3.630)

risk = riskMember1 + riskMember2 + riskMember3 + riskMember4
```

[0080] Coefficients on the left of each line were calculated as COX proportional hazards regression coefficients, the numbers in squared brackets denote 95% confidence bounds for these coefficients. In other words, instead of multiplying the term (0.939 * BIRC5 -3.831) with 0.434039, it may be multiplied with any coefficient between 0.301 and 0.567 and still give a predictive result with in the 95% confidence bounds. Terms in round brackets on the right of each line denote a platform transfer from PCR to Affymetrix: The variables PVALB, CDH1, ... denote PCR-based expressions normalized by the reference genes (delta-Ct values), the whole term within round brackets corresponds to the logarithm (base 2) of Affymetrix microarray expression values of corresponding probe sets.

[0081] Performance of the algorithm T5 was tested in Tamoxifen or Anastrozole treated patients with no more than 3 positive lymph nodes and ER+, HER2-tumors, who participated in the randomized clinical trials ABCSG06 (n=332) or ABCSG08(n=1244). As shown in figure 1, Cox regression analysis reveals, that the T5 score has a significant association with the development of distant metastasis in all cohorts tested.

[0082] Kaplan Meier analysis was performed, after classifying the patients of the combined ABCSG cohorts using a predefined cut off for T5 score. Patients with a low risk of development of a distant metastasis had a T5 score ≤ -9.3, while patients with a high risk of development of a distant metastasis had a T5 score above -9.3. As shown in figure 2, a highly significant separation of both risk groups is observed.

[0083] Importantly, the T5 score was evaluated and compared against "Adjuvant! Online", an online tool to aid in therapy selection based on entry of clinical parameters such as tumor size, tumor grade and nodal status. When the T5 score was tested by bivariate Cox regression against the Adjuvant!Online Relapse Risk score, both scores remained a significant association with the development of distant metastasis. Bivariate Cox regression using dichotomized data, which were stratified according to T5 (cut off -9.3) respectively to Adjuvant!Online (cut off 8), again yielded highly significant and independent associations with time to metastasis as clinical endpoint.

Table 6: Bivariate Cox regression von T5 und Adjuvant!Online

Variable	Hazard ratio	95% CI*	*P*
Adjuvant!Online	2.36	1.58-3.54	<0.0001
Gene-expression signature (risk group)	2.62	1.71-4.01	<0.0001
Adjuvant!Online (score)	1.04	1.02-1.06	<0.0001
Gene-expression signature (risk group)	1.35	1.21-1.49	<0.0001
with HR = Hazard Ratio, 95%CI = 95% Confidence Interval, p = P value.			

[0084] Exemplary Kaplan Meyer Curves are shown in Figs. 1 wherein High = High Risk Group, Low = Low Risk Group according to a predefined cut off

[0085] A high value of the T5 score indicates an increased risk of occurrence of distant metastasis in a given time period.

[0086] This has been shown to be the case for patients having been treated with tamoxifen and also for patients having been treated with aromatase inhibitors.

[0087] Example algorithm T1:

Algorithm T1 is a committee of three members where each member is a linear combination of up to four variables. In general variables may be gene expressions or clinical variables. In T1 the only non-gene variable is the nodal status coded 0, if patient is lymph-node negative and 1, if patient is lymph-node-positive. The mathematical formulas for T1 are shown below.

```
riskMember1 = +0.193935 [0.108..0.280] * (0.792 * PVALB -2.189)
              -0.240252 [-0.400..-0.080] * (0.859 * CDH1 -2.900)
              -0.270069 [-0.385..-0.155] * (0.821 * STC2 -3.529)
              +1.2053 [0.534..1.877] * nodalStatus
```

```
riskMember2 = -0.25051 [-0.437..-0.064] * (0.558 * CXCL12 +0.324)
              -0.421992 [-0.687..-0.157] * (0.715 * RBBP8 -1.063)
              +0.148497 [0.029..0.268] * (1.823 * NMU -12.563)
              +0.293563 [0.108..0.479] * (0.989 * BIRC5 -4.536)
```

```
riskMember3 = +0.308391 [0.074..0.543] * (0.812 * AURKA -2.656)
              -0.225358 [-0.395..-0.055] * (0.637 * PTGER3 + 0.492)
              -0.116312 [-0.202..-0.031] * (0.724 * PIP + 0.985)
```

```
risk = + riskMember1 + riskMember2 + riskMember3
```

[0088] Coefficients on the left of each line were calculated as COX proportional hazards regression coefficients, the numbers in squared brackets denote 95% confidence bounds for these coefficients. Terms in round brackets on the right of each line denote a platform transfer from PCR to Affymetrix: The variables PVALB, CDH1, ... denote PCR-based expressions normalized by the reference genes, the whole term within round brackets corresponds to the logarithm (base 2) of Affymetrix microarray expression values of corresponding probe sets.

[0089] Example algorithm T4:

> Algorithm T4 is a linear combination of motifs. The top 10 genes of several analyses of Affymetrix datasets and PCR data were clustered to motifs. Genes not belonging to a cluster were used as single gene-motifs. COX proportional hazards regression coefficients were found in a multivariate analysis.

[0090] In general motifs may be single gene expressions or mean gene expressions of correlated genes. The mathematical formulas for T4 are shown below.

```
prolif = ((0.84  [0.697..0.977] * RACGAP1 -2.174) + (0.85 [0.713..0.988]
*DHCR7 -3.808)+ (0.94 [0.786..1.089] * BIRC5 -3.734)) / 3
```

```
motiv2 = ((0.83  [0.693..0.96] * IL6ST -5.295) + (1.11 [0.930..1.288] *
ABAT -7.019) + (0.84 [0.701..0.972] * STC2 -3.857)) / 3
```

```
ptger3 =  (PTGER3 * 0.57 [0.475..0.659] + 1.436)
```

```
cxcl12 =  (CXCL12 * 0.53 [0.446..0.618] + 0.847)
```

$$\mathrm{pvalb} = (\mathrm{PVALB} * 0.67\ [0.558..0.774]\ -0.466)$$

[0091] Factors and offsets for each gene denote a platform transfer from PCR to Affymetrix: The variables RACGAP1, DHCR7, ... denote PCR-based expressions normalized by CALM2 and PPIA, the whole term within round brackets corresponds to the logarithm (base 2) of Affymetrix microarray expression values of corresponding probe sets.
The numbers in squared brackets denote 95% confidence bounds for these factors.

[0092] As the algorithm performed even better in combination with a clinical variable the nodal status was added. In T4 the nodal status is coded 0, if patient is lymph-node negative and 1, if patient is lymph-node-positive. With this, algorithm T4 is:

$$\begin{aligned}\mathrm{risk} = {} & -0.32\ [-0.510..-0.137] * \mathrm{motiv2}\\ & + 0.65\ [0.411..0.886] * \mathrm{prolif}\\ & - 0.24\ [-0.398..-0.08] * \mathrm{ptger3}\\ & - 0.05\ [-0.225..0.131] * \mathrm{cxcl12}\\ & + 0.09\ [0.019..0.154] * \mathrm{pvalb}\\ & + \mathrm{nodalStatus}\end{aligned}$$

[0093] Coefficients of the risk were calculated as COX proportional hazards regression coefficients, the numbers in squared brackets denote 95% confidence bounds for these coefficients.

[0094] Algorithm T5b is a committee of two members where each member is a linear combination of four genes. The mathematical formulas for T5b are shown below, the notation is the same as for T1 and T5. In T5b a non-gene variable is the nodal status coded 0, if patient is lymph-node negative and 1, if patient is lymph-node-positive and 0.5 if the lymph-node status is unknown. T5b is defined by:

$$\begin{aligned}\mathrm{riskMember1} = {} & 0.359536\ [0.153..0.566] * (0.891 * \mathrm{DHCR7}\ -4.378)\\ & -0.288119\ [-0.463..-0.113] * (0.485 * \mathrm{MGP} + 4.330)\\ & +0.257341\ [0.112..0.403] * (1.118 * \mathrm{NMU}\ -5.128)\\ & -0.337663\ [-0.499..-0.176] * (0.674 * \mathrm{AZGP1}\ -0.777)\end{aligned}$$

$$\begin{aligned}\mathrm{riskMember2} = {} & -0.374940\ [-0.611..-0.139] * (0.707 * \mathrm{RBBP8}\ -0.934)\\ & -0.387371\ [-0.597..-0.178] * (0.814 * \mathrm{IL6ST}\ -5.034)\\ & +0.800745\ [0.551..1.051] * (0.860 * \mathrm{RACGAP1}\ -2.518)\\ & +0.770650\ [0.323..1.219] * \mathrm{Nodalstatus}\end{aligned}$$

$$\mathrm{risk} = \mathrm{riskMember1} + \mathrm{riskMember2}$$

[0095] The skilled person understands that these algorithms represent particular examples and that based on the information regarding association of gene expression with outcome as given in table 2 alternative algorithms can be established using routine skills.

Algorithm Simplification by employing Subsets of Genes

[0096] "Example algorithm T5" is a committee predictor consisting of 4 members with 2 genes of interest each. Each member is an independent and self-contained predictor of distant recurrence, each additional member contributes to robustness and predictive power of the algorithm to predict time to metastasis, time to death or likelihood of survival for a breast cancer patient. The equation below shows the "Example Algorithm T5"; for ease of reading the number of digits after the decimal point has been truncated to 2; the range in square brackets lists the estimated range of the coefficients (mean +/- 3 standard deviations).

[0097] T5 Algorithm:

```
+0.41 [0.21..0.61] * BIRC5 -0.33 [-0.57..-0.09] * RBBP8

+0.38 [0.15..0.61] * UBE2C -0.30 [-0.55..-0.06] * IL6ST

-0.28 [-0.43..-0.12] * AZGP1 +0.42 [0.16..0.68] * DHCR7

-0.18 [-0.31..-0.06] * MGP -0.13 [-0.25..-0.02] * STC2
```

c-indices: trainSet=0.724,

[0098] Gene names in the algorithm denote the difference of the mRNA expression of the gene compared to one or more housekeeping genes as described above.

[0099] Analysing a cohort different from the finding cohort (234 tumor samples) it was surprising to learn that some simplifications of the "original T5 Algorithm" still yielded a diagnostic performance not significantly inferior to the original T5 algorithm. The most straightforward simplification was reducing the committee predictor to one member only. Examples for the performance of the "one-member committees" are shown below:

member 1 only:

```
+0.41 [0.21..0.61] * BIRC5 -0.33 [-0.57..-0.09] * RBBP8
```

c-indices: trainSet=0.653, independentCohort=0.681

member 2 only:

```
+0.38 [0.15..0.61] * UBE2C -0.30 [-0.55..-0.06] * IL6ST
```

c-indices: trainSet=0.664, independentCohort=0.696

member 3 only:

```
-0.28 [-0.43..-0.12] * AZGP1 +0.42 [0.16..0.68] * DHCR7
```

c-indices: trainSet=0.666, independentCohort=0.601

member 4 only:

```
-0.18 [-0.31..-0.06] * MGP -0.13 [-0.25..-0.02] * STC2
```

c-indices: trainSet=0.668, independentCohort=0.593

[0100] The performance of the one member committees as shown in an independent cohort of 234 samples is notably reduced compared to the performance of the full algorithm. Still, using a committee consisting of fewer members allows for a simpler, less costly estimate of the risk of breast cancer recurrence or breast cancer death that might be acceptable for certain diagnostic purposes.

[0101] Gradually combining more than one but less than four members to a new prognostic committee predictor algorithm, frequently leads to a small but significant increase in the diagnostic performance compared to a one-member committee. It was surprising to learn that there were marked improvements by some combination of committee members while other combinations yielded next to no improvement. Initially, the hypothesis was that a combination of members representing similar biological motives as reflected by the employed genes yielded a smaller improvement than combining members reflecting distinctly different biological motives. Still, this was not the case. No rule could be identified to foretell the combination of some genes to generate an algorithm exibiting more prognostic power than another combination of genes. Promising combinations could only be selected based on experimental data. Identified combinations of combined committee members to yield simplified yet powerful algorithms are shown below.

members 1 and 2 only:

```
+0.41 [0.21..0.61] * BIRC5 -0.33 [-0.57..-0.09] * RBBP8

+0.38 [0.15..0.61] * UBE2C -0.30 [-0.55..-0.06] * IL6ST
```

c-indices: trainSet=0.675, independentCohort=0.712

members 1 and 3 only:

```
+0.41 [0.21..0.61] * BIRC5 -0.33 [-0.57..-0.09] * RBBP8

-0.28 [-0.43..-0.12] * AZGP1 +0.42 [0.16..0.68] * DHCR7
```

c-indices: trainSet=0.697, independentCohort=0.688

members 1 and 4 only:

```
+0.41 [0.21..0.61] * BIRC5 -0.33 [-0.57..-0.09] * RBBP8

-0.18 [-0.31..-0.06] * MGP -0.13 [-0.25..-0.02] * STC2
```

c-indices: trainSet=0.705, independentCohort=0.679

members 2 and 3 only:

```
+0.38 [0.15..0.61] * UBE2C -0.30 [-0.55..-0.06] * IL6ST

-0.28 [-0.43..-0.12] * AZGP1 +0.42 [0.16..0.68] * DHCR7
```

c-indices: trainSet=0.698, independentCohort=0.670

members 1, 2 and 3 only:

```
+0.41 [0.21..0.61] * BIRC5 -0.33 [-0.57..-0.09] * RBBP8

+0.38 [0.15..0.61] * UBE2C -0.30 [-0.55..-0.06] * IL6ST

-0.28 [-0.43..-0.12] * AZGP1 +0.42 [0.16..0.68] * DHCR7
```

`c-indices: trainSet=0.701, independentCohort=0.715`

[0102] Not omitting complete committee members but a single gene or genes from different committee members is also possible but requires a retraining of the entire algorithm. Still, it can also be advantageous to perform. The performance of simplified algorithms generated by omitting entire members or individual genes is largely identical.

Algorithm Variants by Gene Replacement

[0103] Described algorithms, such as "Example algorithm T5", above can be also be modified by replacing one or more genes by one or more other genes. The purpose of such modifications is to replace genes difficult to measure on a specific platform by a gene more straightforward to assay on this platform. While such transfer may not necessarily yield an improved performance compared to a starting algorithm, it can yield the clue to implanting the prognostic algorithm to a particular diagnostic platform. In general, replacing one gene by another gene while preserving the diagnostic power of the predictive algorithm can be best accomplished by replacing one gene by a co-expressed gene with a high correlation (shown e.g. by the Pearson correlation coefficient). Still, one has to keep in mind that the mRNA expression of two genes highly correlative on one platform may appear quite independent from each other when assessed on another platform. Accordingly, such an apparently easy replacement when reduced to practice experimentally, may yield disappointingly poor results as well as surprising strong results, always depending on the imponderabilia of the platform employed. By repeating this procedure one can replace several genes.

[0104] The efficiency of such an approach can be demonstrated by evaluating the predictive performance of the T5 algorithm score and its variants on the validation cohorts. The following table shows the c-index with respect to endpoint distant recurrence in two validation cohorts.

Table 7

Variant	**Validation Study A**	**Validation Study B**
original algorithm T5	c-index = 0.718	c-index = 0.686
omission of BIRC5 (setting expression to some constant)	c-index = 0.672	c-index = 0.643
replacing BIRC5 by UBE2C (no adjustment of the coefficient)	c-index = 0.707	c-index = 0.678

[0105] One can see that omission of one of the T5 genes, here shown for BIRC5 for example, notably reduces the predictive performance. Replacing it with another gene yields about the same performance.

[0106] A better method of replacing a gene is to re-train the algorithm. Since T5 consists of four independent committee members one has to re-train only the member that contains the replaced gene. The following equations demonstrate replacements of genes of the T5 algorithm shown above trained in a cohort of 234 breast cancer patients. Only one member is shown below, for c-index calculation the remaining members were used unchanged from the original T5 Algorithm. The range in square brackets lists the estimated range of the coefficients: mean +/- 3 standard deviations.

Member 1 of T5:

[0107] Original member 1:

`+0.41 [0.21..0.61] * BIRC5 -0.33 [-0.57..-0.09] * RBBP8`

c-indices: trainSet=0.724, independentCohort=0.705
replace BIRC5 by TOP2A in member 1:

`+0.47 [0.24..0.69] * TOP2A -0.34 [-0.58..-0.10] * RBBP8`

c-indices: trainSet=0.734, independentCohort=0.694
replace BIRC5 by RACGAP1 in member 1:

+0.69 [0.37..1.00] * RACGAP1 -0.33 [-0.57..-0.09] * RBBP8

c-indices: trainSet=0.736, independentCohort=0.743
replace RBBP8 by CELSR2 in member 1:

+0.38 [0.19..0.57] * BIRC5 -0.18 [-0.41..0.05] * CELSR2

c-indices: trainSet=0.726, independentCohort=0.680

replace RBBP8 by PGR in member 1:

+0.35 [0.15..0.54] * BIRC5 -0.09 [-0.23..0.05] * PGR

c-indices: trainSet=0.727, independentCohort=0.731

Member 2 of T5:

[0108] Original member 2:

+0.38 [0.15..0.61] * UBE2C -0.30 [-0.55..-0.06] * IL6ST

c-indices: trainSet=0.724, independentCohort=0.725
replace UBE2C by RACGAP1 in member 2:

+0.65 [0.33..0.96] * RACGAP1 -0.38 [-0.62..-0.13] * IL6ST

c-indices: trainSet=0.735, independentCohort=0.718
replace UBE2C by TOP2A in member 2:

+0.42 [0.20..0.65] * TOP2A -0.38 [-0.62..-0.13] * IL6ST

c-indices: trainSet=0.734, independentCohort=0.700
replace IL6ST by INPP4B in member 2:

+0.40 [0.17..0.62] * UBE2C -0.25 [-0.55..0.05] * INPP4B

c-indices: trainSet=0.725, independentCohort=0.686
replace IL6ST by MAPT in member 2:

+0.45 [0.22..0.69] * UBE2C -0.14 [-0.28..0.01] * MAPT

c-indices: trainSet=0.727, independentCohort=0.711

Member 3 of T5:

[0109] Original member 3:

-0.28 [-0.43..-0.12] * AZGP1 +0.42 [0.16..0.68] * DHCR7

c-indices: trainSet=0.724, independentCohort=0.705
replace AZGP1 by PIP in member 3:

-0.10 [-0.18..-0.02] * PIP +0.43 [0.16..0.70] * DHCR7

c-indices: trainSet=0.725, independentCohort=0.692
replace AZGP1 by EPHX2 in member 3:

-0.23 [-0.43..-0.02] * EPHX2 +0.37 [0.10..0.64] * DHCR7

c-indices: trainSet=0.719, independentCohort=0.698
replace AZGP1 by PLAT in member 3:

-0.23 [-0.40..-0.06] * PLAT +0.43 [0.18..0.68] * DHCR7

c-indices: trainSet=0.712, independentCohort=0.715
replace DHCR7 by AURKA in member 3:

-0.23 [-0.39..-0.06] * AZGP1 +0.34 [0.10..0.58] * AURKA

c-indices: trainSet=0.716, independentCohort=0.733

Member 4 of T5:

[0110] Original member 4:

-0.18 [-0.31..-0.06] * MGP -0.13 [-0.25..-0.02] * STC2

c-indices: trainSet=0.724, independentCohort=0.705
replace MGP by APOD in member 4:

-0.16 [-0.30..-0.03] * APOD -0.14 [-0.26..-0.03] * STC2

c-indices: trainSet=0.717, independentCohort=0.679
replace MGP by EGFR in member 4:

-0.21 [-0.37..-0.05] * EGFR -0.14 [-0.26..-0.03] * STC2

c-indices: trainSet=0.715, independentCohort=0.708
replace STC2 by INPP4B in member 4:

-0.18 [-0.30..-0.05] * MGP -0.22 [-0.53..0.08] * INPP4B

c-indices: trainSet=0.719, independentCohort=0.693
replace STC2 by SEC14L2 in member 4:

$$-0.18\ [-0.31..-0.06] * \mathrm{MGP}\ -0.27\ [-0.49..-0.06] * \mathrm{SEC14L2}$$

c-indices: trainSet=0.718, independentCohort=0.681

[0111] One can see that replacements of single genes experimentally identified for a quantification with kinetic PCR normally affect the predictive performance of the T5 algorithm, assessed by the c-index only insignificantly.

[0112] The following table (Tab. 8) shows potential replacement gene candidates for the genes of T5 algorithm. Each gene candidate is shown in one table cell: The gene name is followed by the bracketed absolute Pearson correlation coefficient of the expression of the original gene in the T5 Algorithm and the replacement candidate, and the HG-U133A probe set ID.

Table 8

BIRC5	RBBP8	UBE2C	IL6ST	AZGP1	DHCR7	MGP	STC2
UBE2C (0.775), 202954_at	CELSR2(0.548), 204029_at	BIRC5 (0.775), 202095_s_at	INPP4B (0.477), 205376_at	PIP (0.530), 206509_at	AURKA (0.345), 204092_s_at	APOD (0.368), 201525_at	INPP4B (0.500), 205376_at
TOP2A (0.757), 201292_at	PGR (0.392), 208305_at	RACGAP1 (0.756), 222077_s_at	STC2 (0.450), 203438_at	EPHX2 (0.369), 209368_at	BIRC5 (0.323), 202095_s_at	IL6ST (0.327), 212196_at	IL6ST (0.450), 212196_at
RACGAP1 (0.704), 222077_s_at	STC2 (0.361), 203438_at_	TOP2A (0.753), 201292_at	MAPT (0.440), 206401_s_at	PLAT (0.366), 201860_s_at	UBE2C (0.315), 202954_at	EGFR (0.308), 201983_s_at	SEC14L2 (0.417), 204541_at
AURKA (0.681), 204092_s_at	ABAT (0.317), 209459_s_at	AURKA (0.694), 204092_s_at	SCUBE2 (0.418), 219197_s_at	SEC14L2 (0.351), 204541_at			MAPT (0.414), 206401_s_at
NEK2 (0.680), 204026_s_at	IL6ST (0.311), 212196_at	NEK2 (0.684), 204026_s_at	ABAT (0.389), 209459_s_at	SCUBE2 (0.331), 219197 s_at			CHPT1 (0.410), 221675_s_at
E2F8 (0.640), 219990_at		E2F8 (0.652), 219990_at	PGR (0.377), 208305_at	PGR (0.302), 208305_at			ABAT (0.409), 209459_at
PCNA (0.544), 201202_at		PCNA (0.589), 201202_at	SEC14L2 (0.356), 204541_at				SCUBE2 (0.406), 219197_s_at
CYBRD1 (0.462), 217889_s_at		CYBRD1 (0.486), 217889_at	ESR1 (0.353), 205225_at				ESR1 (0.394), 205225_at
DCN (0.439), 209335_at		ADRA2A (0.391), 209869_at	GJA1 (0.335), 201667_at				RBBP8 (0.361), 203344_s_at
ADRA2A (0.416), 209869_at		DCN (0.384), 209335_at	MGP (0.327), 202291_s_at				PGR (0.347), 208305_at
SQLE (0.415), 209218_at		SQLE (0.369), 209218_at	EPHX2 (0.313), 209368_at				PTPRT (0.343), 205948_at
CXCL12 (0.388), 209687_at		CCND1 (0.347), 208712_at	RBBP8 (0.311), 203344_s_at				HSPA2 (0.317), 211538_s_at
EPHX2 (0.362), 209368_at		ASPH (0.344), 210896_s_at	PTPRT (0.303), 205948_at				PTGER3 (0.314), 210832_x_at
ASPH (0.352), 210896_s_at		CXCL12 (0.342), 209687_at	PLAT (0.301), 201860_s_at				

(continued)

BIRC5	RBBP8	UBE2C	IL6ST	AZGP1	DHCR7	MGP	STC2
PRSS16 (0.352), 208165_s_at		PIP (0.328), 206509_at					
EGFR (0.346), 201983_s_at		PRSS16 (0.326), 208165_s_at					
CCND1 (0.331), 208712_at		EGFR (0.320), 201983_s_at					
TRIM29 (0.325), 202504_at		DHCR7 (0.315), 201791_s_at					
DHCR7 (0.323), 201791_s_at		EPHX2 (0.315), 209368_at					
PIP (0.308), 206509_at		TRIM29 (0.311), 202504_at					
TFAP2B (0.306), 214451_at							
WNT5A (0.303), 205990_s_at							
APOD (0.301), 201525_at							
PTPRT (0.301), 205949_at							

[0113] The following table (Tab. 9) lists qRT-PCR primer and probe sequences used for the table above.

Table 9

gene	probe	forward primer	reverse primer
ABAT	TCGCCCTAAGAGGCTCTTCCTC	GGCAACTTGAGGTCTGACTTTTG	GGTCAGCTCACAAGTGGTGTGA
ADRA2A	TTGTCCTTTCCCCCCTCCGTGC	CCCCAAGAGCTGTTAGGTATCAA	TCAATGACATGATCTCAACCAGAA
APOD	CATCAGCTCTCAACTCCTGGTTTAACA	ACTCACTAATGGAAAACGGAAAGATC	TCACCTTCGATTTGATTCACAGTT
ASPH	TGGGAGGAAGGCAAGGTGCTCATC	TGTGCCAACGAGACCAAGAC	TCGTGCTCAAAGGAGTCATCA
AURKA	CCGTCAGCCTGTGCTAGGCAT	AATCTGGAGGCAAGGTTCGA	TCTGGATTTGCCTCCTGTGAA
BIRC5	AGCCAGATGACGACCCCATAGAGGAACA	CCCAGTGTTTCTTCTGCTTCAAG	CAACCGGACGAATGCTTTTT
CCND1			
CELSR2	ACTGACTTTCCTTCTGGAGCAGGTGGC	TCCAAGCATGTATTCCAG ACTTGT	TGCCCACAGCCTTTTTCT
CHPT1	CCACGGCCACCGAAGAGGCAC	CGCTCGTGCTCATCTCCTACT	CCCAGTGCACATAAAAGGTATGTC
CXCL12	CCACAGCAGGGTTTCAGGTTCC	GCCACTACCCCCTCCTGAA	TCACCTTGCCAACAGTTCTGAT
CYBRD1	AGGGCATCGCCATCATCGTC	GTCACCGGCTTCGTCTTCA	CAGGTCCACGGCAGTCTGT
DCN	TCTTTTCAGCAACCCGGTCCA	AAGGCTTCTTATTCGGGTGTGA	TGGATGGCTGTATCTCCCAGTA
DHCR7	TGAGCGCCCACCCTCTCGA	GGGCTCTGCTTCCCGATT	AGTCATAGGGCAAGCAGAAAATTC
E2F8	CAGGATACCTAATCCCTCTCACGCAG	AAATGTCTCCGCAACCTTGTTC	CTGCCCCCAGGGATGAG
EGFR			
EPHX2	TGAAGCGGGAGGACTTTTTGTAAA	CGATGAGAGTGTTTTATCCATGCA	GCTGAGGCTGGGCTCTTCT
ESR1	ATGCCCTTTTGCCGATGCA	GCCAAATTGTGTTTGATGGATTAA	GACAAAACCGAGTCACATCAGTAATAG
GJA1	TGCACAGCCTTTTGATTTCCCCGAT	CGGGAAGCACCATCTCTAACTC	TTCATGTCCAGCAGCTAGTTTTTT
HSPA2	CAAGTCAGCAAACACGCAAAA	CATGCACGAACTAATCAAAAATGC	ACATTATTCGAGGT1TCTCTTTAATGC
IL6ST	CAAGCTCCACCTTCCAAAGGACCT	CCCTGAATCCATAAAGGCATACC	CAGCTFCGTTTTTCCCTACTTTTT
INPP4B	TCCGAGCGCTGGATTGCATGAG	GCACCAGTTACACAAGGACTTCTTT	TCTCTATGCGGCATCCTTCTC
MAPT	AGACTATTTGCACACTGCCGCCT	GTGGCTCAAAGGATAATATCAAACAC	ACCTTGCTCAGGTCAACTGGTT
MGP	CCTTCATATCCCCTCAGCAGAGATGG	CCTTCATTAACAGGAGAAATGCAA	ATTGAGCTCGTGGACAGGCTTA
NEK2	TCCTGAACAAATGAATCGCATGTCCTACAA	ATTTGTTGGCACACCTTATTACATGT	AAGCAGCCCAATGACCAGATa
PCNA	AAATACTAAAATGCGCCGGCAATGA	GGGCGTGAACCTCACCAGTA	CTTCGGCCCTTAGTGTAATGATATC

(continued)

gene	probe	forward primer	reverse primer
PGR	TTGATAGAAACGCTGTGAGCTCGA	AGCTCATCAAGGCAATTGGTTT	ACAAGATCATGCAAGTTATCAAGAAGTT
PIP	TGCATGGTGGTTAAAACTTACCTCA	TGCTTGCAGTTCAAACAGAATTG	CACCTTGTAGAGGGATGCTGCTA
PLAT	CAGAAAGTGGCCATGCCACCCTG	TGGGAAGACATGAATGCACACTA	GGAGGTTGGGCTTTAGCTGAA
PRSS16	CACTGCCGGTCACCCACACCA	CTGAGGAGCACAGAACCTCAACT	CGAACTCGGTACATGTCTGATACAA
PTGER3	TCGGTCTGCTGGTCTCCGCTCC	CTGATTGAAGATCATTTfCAACATCA	GACGGCCATTCAGCTTATGG
PTPRT	TTGGCTTCTGGACACCCTCACA	GAGTTGTGGCCTCTACCATTGC	GAGCGGGAACCTTGGGATAG
RACGAP1	ACTGAGAATCTCCACCCGGCGCA	TCGCCAACTGGATAAATTGGA	GAATGTGCGGAATCTGTTTGAG
RBBP8	ACCGATTCCGCTACATTCCACCCAAC	AGAAATTGGCTTCCTGCTCAAG	AAAACCAACTTCCCAAAAATTCTCT
SCUBE2	CTAGAGGGTTCCAGGTCCCATACGTGACATA	TGTGGATTCAGTTCAAGTCCAATG	CCATCTCGAACTATGTCTTCAATGAGT
SEC14L2	TGGGAGGCATGCAACGCGTG	AGGTCTTACTAAGCAGTCCCATCTCT	CGACCGGCACCTGAACTC
SQLE	TATGCGTCTCCCAAAAGAAGAACACCTCG	GCAAGCTTCCTTCCTCCTTCA	CCTTTAGCAGTTTTTCTCCATATTTATATC
STC2	TCTCACCTTGACCCTCAGCCAAG	ACATTTGACAAATTTCCCTTAGGATT	CCAGGACGCAGCTTTACCAA
TFAP2B	CAACACCACCACTAACAGGCACACGTC	GGCATGGACAAGATGTTCTTGA	CCTCCTTGTCGCCAGTTTTACT
TOP2A	CAGATCAGGACCAAGATGGTTCCCACAT	CATTGAAGACGCTTCGTTATGG	CCAGTTGTGATGGATAAAATTAATCAG
TRIM29	TGCTGTCTCACTACCGGCCATTCTACG	TGGAAATCTGGCAAGCAGACT	CAATCCCGTTGCCTTTGTTG
UBE2C	TGAACACACATGCTGCCGAGCTCTG	CTTCTAGGAGAACCCAACATTGATAGT	GTTTCTTGCAGGTACTTCTTAAAAGCT
WNT5A	TATTCACATCCCCTCAGTTGCAGTGAATTG	CTGTGGCTCTTAATTTATTGCATAATG	TTAGTGCTTTTTGCTTTCAAGATCTT

[0114] A second alternative for unsupervised selection of possible gene replacement candidates is based on Affymetrix data only. This has the advantage that it can be done solely based on already published data (e.g. from www.ncbi.nlm.nih.gov/geo/). The following table (Tab. 10) lists HG-U133a probe set replacement candidates for the probe sets used in algorithms T1 - T5. This is based on training data of these algorithms. The column header contains the gene name and the probe set ID in bold. Then, the 10 best-correlated probe sets are listed, where each table cell contains the probe set ID, the correlation coefficient in brackets and the gene name.

Table 10

UBE2C	**BIRC5**	**DHCR7**	**RACGAP1**	**AURKA**	**PVALB**	**NMU**	**STC2**
202954_at	**202095_s_at**	**201791_s_at**	**222077_s_at**	**204092_s_at**	**205336_at**	**206023_at**	**203438_at**
210052_s_at (0.82) TPX2	202954_at (0.82) UBE2C	201790_s_at (0.66) DHCR7	218039_at (0.79) NUSAP1	208079_s_at (0.89) STK6	208683_at (-0.33) CAPN2	205347_s_at (0.45) TMSL8	203439_s_at (0.88) STC2
202095_s_at (0.82) BIRC5	218039_at (0.81) NUSAP1	202218_s_at (0.48) FADS2	214710_s_at (0.78) CCNB1	202954_at(0.80) UBE2C	219682_s_at (0.30) TBX3	203764_at (0.45) DLG7	212496_s_at (0.52) JMJD2B
218009_s_at (0.82) PRC1	218009_s_at (0.79) PRC1	202580_x_at (0.47) FOXM1	203764_at (0.77) DLG7	210052_s_at (0.77) TPX2	218704_at (0.30) FU20315	203554_x_at (0.44) PTTG1	219440_at (0.52) RAI2
203554_x_at (0.82) PTTG1	202705_at (10.78) CCNB2	208944_at (-0.46) TGFBR2	204026_s_at (0.77) ZWINT	202095_s_at (0.77) BIRC5		204962_s_at (0.44) CFNPA	215867_x_at (0.51) CA12
208079_s_at (0.81) STK6	204962_s_at (0.78) CENPA	202954_at (0.46) UBE2C	218009_s_at (0.76) PRC1	203554_x_at (0.76) PTTG1		204825_at (0.43) MELK	214164_x_at (0.50) CA12
202705_at (0.81) CCNB2	203554_x_at (0.78) PTTG1	209541_at (-0.45) IGF1	204641_at (0.76) NEK2	218009_s_at (0.75) PRC1		209714_s_at (0:41) CDKN3	204541_at (0.50) SEC14L2
218039_at (0.81) NUSAP1	208079_s_at (0.78) STK6	201059_at (0.45) CTTN	204444_at (0.75) KIF11	201292_at(0.73) TOP2A		219918_s_at (0.41) ASPM	203963_at (0.50) CA12
202870_s_at (0.80) CDC20	210052_s_at (0.77) TPX2	200795_at (-0.45) SPARCL1	202705_at (0.75) CCNB2	214710_s_at (0.73) CCNB1		207828_s_at (0.41) CENPF	212495_at (0.50) JMJ$_{D2B}$
204092_s_at (0.80) STK6	202580_x_at (0.77) FOXM1	218009_s_at (0.45) PRC1	203362_s_at (0.75) MAD2L1	204962_s_at (0.73) CENPA		202705_at (0.41) CCNB2	208614_s_at (0.49) FLNB
209408_at (0.80) KIF2C	204092_s_at (0.77) STK6	218542_at (0.45) C10orf3	202954_at (0.75) UBE2C	218039_at(0.73) NUSAP1		219787_s_at (0.40) ECT2	213933_at (0.49) PTGER3
AZGP1	**RBBP8**	**IL6ST**	**MGP**	**PTGER3**	**CXCL12**	**ABAT**	**CDH1**
209309_at	**203344_s_at**	**212196_at**	**202291_s_at**	**213933_at**	**209687_at**	**209460_at**	**201131_s_at**
217014_s_at (0.92) AZGP1	36499_at(0.49) CELSR2	212195_at (0.85) IL6ST	201288_at (0.46) ARHGDIB	210375_at(0.74) PTGER3	204955_at (0.81) SRPX	209459_s_at (0.92) ABAT	201130_s_at (0.57) CDH1
206509_at (0.52) PIP	204029_at (0.45) CELSR2	204864_s_at (0.75) IL6ST	219768_at (0.42) VTCN1	210831_s_at (0,74)PTGER3	209335_at (0.81) DCN	206527_at (0.63) ABAT	221597_s_at (0.40) HSPC171

(continued)

AZGP1	RBBP8	IL6ST	MGP	PTGER3	CXCL12	ABAT	CDH1
204541_at (0.46) SEC14L2	208305_at (0.45) PGR	211000_s_at (0.68)IL6ST	202849_x_at (-0.41) GRK6	210374_x_at (0.73) PTGER3	211896_s_at (0.81) DCN	213392_at (0.54) MGC35048	203350_at (0.38) AP1G1
200670_at (0.45) XBP1	205380_at (0.43)PDZK1	214077_x_at (0.61) MEIS4	205382_s_at (0.40) DF	210832_x_at (0.73) PTGER3	201893_x_at (0.81) DCN	221666_s_at (0.49) PYCARD	209163_at (0.36) CYB561
209368_at (0:45) EPHX2	203303_at (0.41) ICTE1L	204863_s_at (0,58) IL6ST	200099_s_at (0.39) RPS3A	210834_s_at (0.55) PTGER3	203666_at (0.80) CXCL12	218016_s_at (0.48) POLR3E	210239_at (0.35) IRX5
218627_at (-0.43) FU11259	205280_at (0.38) GLRB	202089_s_at (0.57) SLC39A6	221591_s_at (-0.37) FAM64A	210833_at(0.55) PTGER3	211813_x_at (0.80) DCN	214440_at (0.46) NAT1	200942_s_at (0.34) HSBP1
202286_s_at (0.43) TACSTD2	205279_s_at (0.38) GLRB	210735_s_at (0.56) CA12	214629_x_at (0.37) RTN4	203438_at 0.49) STC2	208747_s_at (0.79) C1S	204981_at (0.45) SLC22A18	209157_at (0.34) DNAJA2
213832_at (0.42)	203685_at (0.38) BCL2	200648_s_at (0.52) GLUL	200748_s_at (0.37) FTH1	203439_s_at (0.46) STC2	203131_at (0.78) PDGFRA	212195_at (0.45) IL6ST	210715_s_at (0.33) SPINT2
204288_s_at (0.41) SORBS2	203304_at (-0.38) BAMBI	214552_s_at (0.52) RABEP1	209408_at (-0.37) KIF2C	212195_at(0.41) IL6ST	202994_s_at (0.78) FBLN1	204497_at (0.45) ADCY9	203219_s_at (0.33) APRT
202376_at (0.41) SERPINA3	205862_at (0.36) GRFB1	219197_s_at (0.51) SCUBE2	218726_at (-0.36) DKFZp762E1312	217764_s_at (0.40) RAB31	208944_at (0,78) TGFBR2	215867_x_at (0.45) CA12	218074_at (0.33) FAM96B

[0115] After selection of a gene or a probe set one has to define a mathematical mapping between the expression values of the gene to replace and those of the new gene. There are several alternatives which are discussed here based on the example "replace delta-Ct values of BIRC5 by RACGAP1". In the training data the joint distribution of expressions looks like in Figure 3.

[0116] The Pearson correlation coefficient is 0.73.

One approach is to create a mapping function from RACGAP1 to BIRC5 by regression. Linear regression is the first choice and yields in this example

$$\mathrm{BIRC5} = 1.22 * \mathrm{RACGAP1} - 2.85.$$

[0117] Using this equation one can easily replace the BIRC5 variable in e.g. algorithm T5 by the right hand side. In other examples robust regression, polynomial regression or univariate nonlinear pre-transformations may be adequate. The regression method assumes measurement noise on BIRC5, but no noise on RACGAP1. Therefore the mapping is not symmetric with respect to exchangeability of the two variables. A symmetric mapping approach would be based on two univariate z-transformations.

$$z = (\mathrm{BIRC5} - \mathrm{mean}(\mathrm{BIRC5})) \;/\; \mathrm{std}(\mathrm{BIRC5}) \text{ and}$$

$$z = (\mathrm{RACGAP1} - \mathrm{mean}(\mathrm{RACGAP1})) \;/\; \mathrm{std}(\mathrm{RACGAP1})$$

$$z = (\mathrm{BIRC5} - 8.09) \;/\; 1.29 = (\mathrm{RACGAP1} - 8.95) \;/\; 0.77$$

$$\mathrm{BIRC5} = 1.67 * \mathrm{RACGAP1} + -6.89$$

Again, in other examples, other transformations may be adequate: normalization by median and/or mad, nonlinear mappings, or others.

[0118] The invention is summarized according to the following items:

1. Method for predicting an outcome of breast cancer in an estrogen receptor positive and HER2 negative tumor of a breast cancer patient, said method comprising:

   (a) determining in a tumor sample from said patient the RNA expression levels of at least 2 of the following 9 genes: UBE2C, BIRC5, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP

   (b) mathematically combining expression level values for the genes of the said set which values were determined in the tumor sample to yield a combined score, wherein said combined score is indicative of a prognosis of said patient.

2. Method of item 1 comprising:

   determining in a tumor sample from said patient the RNA expression levels of at least 3, 4, 5 or 6 of the following 9 genes: UBE2C, BIRC5, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP

3. Method of item 1 or 2 comprising:

   (a) determining in a tumor sample from said patient the RNA expression levels of the following 8 genes: UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP
   (b) mathematically combining expression level values for the genes of the said set which values were determined in the tumor sample to yield a combined score, wherein said combined score is indicative of a prognosis of said patient.

4. Method of item 1 or 2 comprising:

(a) determining in a tumor sample from said patient the RNA expression levels of the following 8 genes: UBE2C, BIRC5, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP;

(b) mathematically combining expression level values for the genes of the said set which values were determined in the tumor sample to yield a combined score, wherein said combined score is indicative of a prognosis of said patient.

5. Method according to item 4 wherein
BIRC5 may be replaced by UBE2C or TOP2A or RACGAP1 or AURKA or NEK2 or E2F8 or PCNA or CYBRD1 or DCN or ADRA2A or SQLE or CXCL12 or EPHX2 or ASPH or PRSS16 or EGFR or CCND1 or TRIM29 or DHCR7 or PIP or TFAP2B or WNT5A or APOD or PTPRT with the proviso that after a replacement 8 different genes are selected; and
UBE2C may be replaced by BIRC5 or RACGAP1 or TOP2A or AURKA or NEK2 or E2F8 or PCNA or CYBRD1 or ADRA2A or DCN or SQLE or CCND1 or ASPH or CXCL12 or PIP or PRSS16 or EGFR or DHCR7 or EPHX2 or TRIM29 with the proviso that after a replacement 8 different genes are selected; and **DHCR7** may be replaced by AURKA, BIRC5, UBE2C or by any other gene that may replace BIRC5 or UBE2C with the proviso that after a replacement 8 different genes are selected; and
while STC2 may be replaced by INPP4B or IL6ST or SEC14L2 or MAPT or CHPT1 or ABAT or SCUBE2 or ESR1 or RBBP8 or PGR or PTPRT or HSPA2 or PTGER3 with the proviso that after a replacement 8 different genes are selected; and
AZGP1 may be replaced by PIP or EPHX2 or PLAT or SEC14L2 or SCUBE2 or PGR with the proviso that after a replacement 8 different genes are selected; and
RBBP8 may be replaced by CELSR2 or PGR or STC2 or ABAT or IL6ST with the proviso that after a replacement 8 different genes are selected; and
IL6ST may be replaced by INPP4B or STC2 or MAPT or SCUBE2 or ABAT or PGR or SEC14L2 or ESR1 or GJA1 or MGP or EPHX2 or RBBP8 or PTPRT or PLAT with the proviso that after a replacement 8 different genes are selected; and
MGP may be replaced by APOD or IL6ST or EGFR with the proviso that after a replacement 8 different genes are selected.

6. Method according to item 1 to 5, wherein said patient has received endocrine therapy or is contemplated to receive endocrine treatment.

7. Method of item 6, wherein said endocrine therapy comprises tamoxifen or an aromatase inhibitor.

8. Method according to any of items 1 to 7 wherein a risk of developing breast cancer recurrence or cancer related death is predicted.

9. Method according to any of items 1 to 8, wherein said expression level is determined as a Messenger-RNA expression level.

10. Method according to item 8, wherein said expression level is determined by at least one of

a PCR based method,
a microrarry based method, and
a hybridization based method.

11. Method of any one of the preceding items, wherein said determination of expression levels is in a formalin-fixed paraffin embedded tumor sample or in a fresh-frozen tumor sample.

12. A method of any of the above items, wherein the expression level of at least one marker gene is determined as a pattern of expression relative to at least one reference gene or to a computed average expression value.

13. Method of any one of the preceding items, wherein said step of mathematically combining comprises a step of applying an algorithm to values representative of expression levels of given genes.

14. Method of item 13, wherein said algorithm is a linear combination of said values representative of expression levels of given genes.

15. Method of item 14 wherein a value for a representative value of an expression level of a given gene is multiplied with a coefficient.

16. Method of any one of the preceding items, wherein one, two or more thresholds are determined for said combined score, that discriminate into high and low risk, high, intermediate and low risk, or more risk groups by applying the threshold on the combined score.

17. Method of any of items 1 to 16, wherein a high combined score is indicative of benefit from cytotoxic chemotherapy.

18. Method of any one of the preceding items, wherein information regarding nodal status of the patient is processed in the step of mathematically combining expression level values for the genes to yield a combined score.

19. Method of items 17 and 18, wherein said information regarding nodal status is a numerical value if said nodal status is negative and said information is a different numerical value if said nodal status positive and a different or identical number if said nodal status is unknown.

20. A kit for performing a method of any of items 1 to 19, said kit comprising a set of oligonucleotides capable of specifically binding sequences or to sequences of fragments of the genes in a combination of genes, wherein said combination comprises at least the two of the 9 genes UBE2C, BIRC5, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP;

21. A computer program product capable of processing values representative of expression levels of a set of genes, mathematically combining said values to yield a combined score, wherein said combined score is indicative of efficacy from endocrine therapy of said patient, according to the methods of any of items 1 to 17.

SEQUENCE LISTING

[0119]

<110> Sividon Diagnostics GmbH

<120> Method for breast cancer recurrence prediction under endocrine
treatment

<130> 141469EP

<150> 11710526.2
<151> 2011-03-29

<160> 141

<170> PatentIn version 3.3

<210> 1
<211> 22
<212> DNA
<213> Homo sapiens

<400> 1
tcgccctaag aggctcttcc tc 22

<210> 2
<211> 23
<212> DNA
<213> Homo sapiens

<400> 2
ggcaacttga ggtctgactt ttg 23

<210> 3
<211> 22
<212> DNA
<213> Homo sapiens

<400> 3
ggtcagctca caagtggtgt ga 22

<210> 4
<211> 22
<212> DNA
<213> Homo sapiens

<400> 4
ttgtcctttc ccccctccgt gc 22

<210> 5
<211> 23
<212> DNA
<213> Homo sapiens

<400> 5
ccccaagagc tgttaggtat caa 23

<210> 6
<211> 24
<212> DNA
<213> Homo sapiens

<400> 6
tcaatgacat gatctcaacc agaa 24

<210> 7
<211> 27
<212> DNA
<213> Homo sapiens

<400> 7
catcagctct caactcctgg tttaaca 27

<210> 8
<211> 26
<212> DNA
<213> Homo sapiens

<400> 8
actcactaat ggaaaacgga aagatc 26

<210> 9
<211> 24
<212> DNA
<213> Homo sapiens

<400> 9
tcaccttcga tttgattcac agtt 24

<210> 10
<211> 24

<212> DNA
<213> Homo sapiens

<400> 10
tgggaggaag gcaaggtgct catc 24

<210> 11
<211> 20
<212> DNA
<213> Homo sapiens

<400> 11
tgtgccaacg agaccaagac 20

<210> 12
<211> 21
<212> DNA
<213> Homo sapiens

<400> 12
tcgtgctcaa aggagtcatc a 21

<210> 13
<211> 21
<212> DNA
<213> Homo sapiens

<400> 13
ccgtcagcct gtgctaggca t 21

<210> 14
<211> 20
<212> DNA
<213> Homo sapiens

<400> 14
aatctggagg caaggttcga 20

<210> 15
<211> 21
<212> DNA
<213> Homo sapiens

<400> 15
tctggatttg cctcctgtga a 21

<210> 16
<211> 28
<212> DNA
<213> Homo sapiens

<400> 16
agccagatga cgaccccata gaggaaca 28

<210> 17
<211> 23
<212> DNA
<213> Homo sapiens

<400> 17
cccagtgttt cttctgcttc aag 23

<210> 18
<211> 20
<212> DNA
<213> Homo sapiens

<400> 18
caaccggacg aatgcttttt 20

<210> 19
<211> 27
<212> DNA
<213> Homo sapiens

<400> 19
actgactttc cttctggagc aggtggc 27

<210> 20
<211> 24
<212> DNA
<213> Homo sapiens

<400> 20
tccaagcatg tattccagac ttgt 24

<210> 21
<211> 20
<212> DNA
<213> Homo sapiens

<400> 21
tgcccacagc ctctttttct 20

<210> 22
<211> 21
<212> DNA
<213> Homo sapiens

<400> 22
ccacggccac cgaagaggca c 21

<210> 23
<211> 21
<212> DNA
<213> Homo sapiens

<400> 23
cgctcgtgct catctcctac t 21

<210> 24
<211> 24
<212> DNA
<213> Homo sapiens

<400> 24
cccagtgcac ataaaaggta tgtc 24

<210> 25
<211> 22
<212> DNA
<213> Homo sapiens

<400> 25
ccacagcagg gtttcaggtt cc 22

<210> 26
<211> 19
<212> DNA
<213> Homo sapiens

<400> 26
gccactaccc cctcctgaa 19

<210> 27
<211> 22
<212> DNA
<213> Homo sapiens

<400> 27
tcaccttgcc aacagttctg at 22

<210> 28
<211> 20
<212> DNA
<213> Homo sapiens

<400> 28
agggcatcgc catcatcgtc 20

<210> 29
<211> 19
<212> DNA
<213> Homo sapiens

<400> 29
gtcaccggct tcgtcttca 19

<210> 30
<211> 19
<212> DNA
<213> Homo sapiens

<400> 30
caggtccacg gcagtctgt 19

<210> 31
<211> 21
<212> DNA
<213> Homo sapiens

<400> 31
tcttttcagc aacccggtcc a 21

<210> 32
<211> 22

<212> DNA
<213> Homo sapiens

<400> 32
aaggcttctt attcgggtgt ga 22

<210> 33
<211> 22
<212> DNA
<213> Homo sapiens

<400> 33
tggatggctg tatctcccag ta 22

<210> 34
<211> 19
<212> DNA
<213> Homo sapiens

<400> 34
tgagcgccca ccctctcga 19

<210> 35
<211> 18
<212> DNA
<213> Homo sapiens

<400> 35
gggctctgct tcccgatt 18

<210> 36
<211> 24
<212> DNA
<213> Homo sapiens

<400> 36
agtcataggg caagcagaaa attc 24

<210> 37
<211> 26
<212> DNA
<213> Homo sapiens

<400> 37
caggatacct aatccctctc acgcag 26

<210> 38
<211> 22
<212> DNA
<213> Homo sapiens

<400> 38
aaatgtctcc gcaaccttgt tc 22

<210> 39
<211> 17
<212> DNA
<213> Homo sapiens

<400> 39
ctgcccccag ggatgag 17

<210> 40
<211> 24
<212> DNA
<213> Homo sapiens

<400> 40
tgaagcggga ggactttttg taaa 24

<210> 41
<211> 24
<212> DNA
<213> Homo sapiens

<400> 41
cgatgagagt gttttatcca tgca 24

<210> 42
<211> 19
<212> DNA
<213> Homo sapiens

<400> 42
gctgaggctg ggctcttct 19

<210> 43
<211> 19
<212> DNA
<213> Homo sapiens

<400> 43
atgccctttt gccgatgca 19

<210> 44
<211> 24
<212> DNA
<213> Homo sapiens

<400> 44
gccaaattgt gtttgatgga ttaa 24

<210> 45
<211> 27
<212> DNA
<213> Homo sapiens

<400> 45
gacaaaaccg agtcacatca gtaatag 27

<210> 46
<211> 25
<212> DNA
<213> Homo sapiens

<400> 46
tgcacagcct tttgatttcc ccgat 25

<210> 47
<211> 22
<212> DNA
<213> Homo sapiens

<400> 47
cgggaagcac catctctaac tc 22

<210> 48
<211> 24
<212> DNA
<213> Homo sapiens

<400> 48
ttcatgtcca gcagctagtt tttt 24

<210> 49
<211> 21
<212> DNA
<213> Homo sapiens

<400> 49
caagtcagca aacacgcaaa a 21

<210> 50
<211> 24
<212> DNA
<213> Homo sapiens

<400> 50
catgcacgaa ctaatcaaaa atgc 24

<210> 51
<211> 27
<212> DNA
<213> Homo sapiens

<400> 51
acattattcg aggtttctct ttaatgc 27

<210> 52
<211> 24
<212> DNA
<213> Homo sapiens

<400> 52
caagctccac cttccaaagg acct 24

<210> 53
<211> 23
<212> DNA
<213> Homo sapiens

<400> 53
ccctgaatcc ataaaggcat acc 23

<210> 54
<211> 24

<212> DNA
<213> Homo sapiens

<400> 54
cagcttcgtt tttccctact tttt 24

<210> 55
<211> 22
<212> DNA
<213> Homo sapiens

<400> 55
tccgagcgct ggattgcatg ag 22

<210> 56
<211> 25
<212> DNA
<213> Homo sapiens

<400> 56
gcaccagtta cacaaggact tcttt 25

<210> 57
<211> 21
<212> DNA
<213> Homo sapiens

<400> 57
tctctatgcg gcatccttct c 21

<210> 58
<211> 23
<212> DNA
<213> Homo sapiens

<400> 58
agactatttg cacactgccg cct 23

<210> 59
<211> 26
<212> DNA
<213> Homo sapiens

<400> 59
gtggctcaaa ggataatatc aaacac 26

<210> 60
<211> 22
<212> DNA
<213> Homo sapiens

<400> 60
accttgctca ggtcaactgg tt 22

<210> 61
<211> 26
<212> DNA
<213> Homo sapiens

<400> 61
ccttcatatc ccctcagcag agatgg 26

<210> 62
<211> 24
<212> DNA
<213> Homo sapiens

<400> 62
ccttcattaa caggagaaat gcaa 24

<210> 63
<211> 22
<212> DNA
<213> Homo sapiens

<400> 63
attgagctcg tggacaggct ta 22

<210> 64
<211> 30
<212> DNA
<213> Homo sapiens

<400> 64
tcctgaacaa atgaatcgca tgtcctacaa 30

<210> 65
<211> 26
<212> DNA
<213> Homo sapiens

<400> 65
atttgttggc acaccttatt acatgt 26

<210> 66
<211> 21
<212> DNA
<213> Homo sapiens

<400> 66
aagcagccca atgaccagat a 21

<210> 67
<211> 25
<212> DNA
<213> Homo sapiens

<400> 67
aaatactaaa atgcgccggc aatga 25

<210> 68
<211> 20
<212> DNA
<213> Homo sapiens

<400> 68
gggcgtgaac ctcaccagta 20

<210> 69
<211> 25
<212> DNA
<213> Homo sapiens

<400> 69
cttcggccct tagtgtaatg atatc 25

<210> 70
<211> 24
<212> DNA
<213> Homo sapiens

<400> 70
ttgatagaaa cgctgtgagc tcga 24

<210> 71
<211> 22
<212> DNA
<213> Homo sapiens

<400> 71
agctcatcaa ggcaattggt tt 22

<210> 72
<211> 28
<212> DNA
<213> Homo sapiens

<400> 72
acaagatcat gcaagttatc aagaagtt 28

<210> 73
<211> 25
<212> DNA
<213> Homo sapiens

<400> 73
tgcatggtgg ttaaaactta cctca 25

<210> 74
<211> 23
<212> DNA
<213> Homo sapiens

<400> 74
tgcttgcagt tcaaacagaa ttg 23

<210> 75
<211> 23
<212> DNA
<213> Homo sapiens

<400> 75
caccttgtag agggatgctg cta 23

<210> 76
<211> 23

<212> DNA
<213> Homo sapiens

<400> 76
cagaaagtgg ccatgccacc ctg 23

<210> 77
<211> 23
<212> DNA
<213> Homo sapiens

<400> 77
tgggaagaca tgaatgcaca cta 23

<210> 78
<211> 21
<212> DNA
<213> Homo sapiens

<400> 78
ggaggttggg ctttagctga a 21

<210> 79
<211> 21
<212> DNA
<213> Homo sapiens

<400> 79
cactgccggt cacccacacc a 21

<210> 80
<211> 23
<212> DNA
<213> Homo sapiens

<400> 80
ctgaggagca cagaacctca act 23

<210> 81
<211> 25
<212> DNA
<213> Homo sapiens

<400> 81
cgaactcggt acatgtctga tacaa 25

<210> 82
<211> 22
<212> DNA
<213> Homo sapiens

<400> 82
tcggtctgct ggtctccgct cc 22

<210> 83
<211> 26
<212> DNA
<213> Homo sapiens

<400> 83
ctgattgaag atcattttca acatca 26

<210> 84
<211> 20
<212> DNA
<213> Homo sapiens

<400> 84
gacggccatt cagcttatgg 20

<210> 85
<211> 22
<212> DNA
<213> Homo sapiens

<400> 85
ttggcttctg gacaccctca ca 22

<210> 86
<211> 22
<212> DNA
<213> Homo sapiens

<400> 86
gagttgtggc ctctaccatt gc 22

<210> 87
<211> 20
<212> DNA
<213> Homo sapiens

<400> 87
gagcgggaac cttgggatag 20

<210> 88
<211> 23
<212> DNA
<213> Homo sapiens

<400> 88
actgagaatc tccacccggc gca 23

<210> 89
<211> 21
<212> DNA
<213> Homo sapiens

<400> 89
tcgccaactg gataaattgg a 21

<210> 90
<211> 22
<212> DNA
<213> Homo sapiens

<400> 90
gaatgtgcgg aatctgtttg ag 22

<210> 91
<211> 26
<212> DNA
<213> Homo sapiens

<400> 91
accgattccg ctacattcca cccaac 26

<210> 92
<211> 22
<212> DNA
<213> Homo sapiens

<400> 92
agaaattggc ttcctgctca ag 22

<210> 93
<211> 25
<212> DNA
<213> Homo sapiens

<400> 93
aaaaccaact tcccaaaaat tctct 25

<210> 94
<211> 31
<212> DNA
<213> Homo sapiens

<400> 94
ctagagggtt ccaggtccca tacgtgacat a 31

<210> 95
<211> 24
<212> DNA
<213> Homo sapiens

<400> 95
tgtggattca gttcaagtcc aatg 24

<210> 96
<211> 27
<212> DNA
<213> Homo sapiens

<400> 96
ccatctcgaa ctatgtcttc aatgagt 27

<210> 97
<211> 20
<212> DNA
<213> Homo sapiens

<400> 97
tgggaggcat gcaacgcgtg 20

<210> 98
<211> 26

<212> DNA
<213> Homo sapiens

<400> 98
aggtcttact aagcagtccc atctct 26

<210> 99
<211> 18
<212> DNA
<213> Homo sapiens

<400> 99
cgaccggcac ctgaactc 18

<210> 100
<211> 29
<212> DNA
<213> Homo sapiens

<400> 100
tatgcgtctc ccaaaagaag aacacctcg 29

<210> 101
<211> 21
<212> DNA
<213> Homo sapiens

<400> 101
gcaagcttcc ttcctccttc a 21

<210> 102
<211> 31
<212> DNA
<213> Homo sapiens

<400> 102
cctttagcag ttttctccat agttttatat c 31

<210> 103
<211> 27
<212> DNA
<213> Homo sapiens

<400> 103
caacaccacc actaacaggc acacgtc 27

<210> 104
<211> 22
<212> DNA
<213> Homo sapiens

<400> 104
ggcatggaca agatgttctt ga 22

<210> 105
<211> 22
<212> DNA
<213> Homo sapiens

<400> 105
cctccttgtc gccagtttta ct 22

<210> 106
<211> 28
<212> DNA
<213> Homo sapiens

<400> 106
cagatcagga ccaagatggt tcccacat 28

<210> 107
<211> 22
<212> DNA
<213> Homo sapiens

<400> 107
cattgaagac gcttcgttat gg 22

<210> 108
<211> 27
<212> DNA
<213> Homo sapiens

<400> 108
ccagttgtga tggataaaat taatcag 27

<210> 109
<211> 27
<212> DNA
<213> Homo sapiens

<400> 109
tgctgtctca ctaccggcca ttctacg 27

<210> 110
<211> 21
<212> DNA
<213> Homo sapiens

<400> 110
tggaaatctg gcaagcagac t 21

<210> 111
<211> 20
<212> DNA
<213> Homo sapiens

<400> 111
caatcccgtt gcctttgttg 20

<210> 112
<211> 25
<212> DNA
<213> Homo sapiens

<400> 112
tgaacacaca tgctgccgag ctctg 25

<210> 113
<211> 27
<212> DNA
<213> Homo sapiens

<400> 113
cttctaggag aacccaacat tgatagt 27

<210> 114
<211> 27
<212> DNA
<213> Homo sapiens

<400> 114
gtttcttgca ggtacttctt aaaagct 27

<210> 115
<211> 30
<212> DNA
<213> Homo sapiens

<400> 115
tattcacatc ccctcagttg cagtgaattg 30

<210> 116
<211> 27
<212> DNA
<213> Homo sapiens

<400> 116
ctgtggctct taatttattg cataatg 27

<210> 117
<211> 26
<212> DNA
<213> Homo sapiens

<400> 117
ttagtgcttt ttgctttcaa gatctt 26

<210> 118
<211> 23
<212> DNA
<213> Homo sapiens

<400> 118
tctcaccttg accctcagcc aag 23

<210> 119
<211> 26
<212> DNA
<213> Homo sapiens

<400> 119
acatttgaca aatttccctt aggatt 26

<210> 120
<211> 20

<212> DNA
<213> Homo sapiens

<400> 120
ccaggacgca gctttaccaa 20

<210> 121
<211> 20
<212> DNA
<213> Homo sapiens

<400> 121
caccagccac caggccccag 20

<210> 122
<211> 18
<212> DNA
<213> Homo sapiens

<400> 122
tcctggaccg gcaagatc 18

<210> 123
<211> 21
<212> DNA
<213> Homo sapiens

<400> 123
taggccaggc acttcagttt c 21

<210> 124
<211> 22
<212> DNA
<213> Homo sapiens

<400> 124
tcgcgtctcg gaaaccggta gc 22

<210> 125
<211> 20
<212> DNA
<213> Homo sapiens

<400> 125
gagcgagctg agtggttgtg 20

<210> 126
<211> 21
<212> DNA
<213> Homo sapiens

<400> 126
agtcagttgg tcagccatgc t 21

<210> 127
<211> 27
<212> DNA
<213> Homo sapiens

<400> 127
cctgccaatc ccgatgaaat tggaaat 27

<210> 128
<211> 21
<212> DNA
<213> Homo sapiens

<400> 128
tgagtgtccc ccggtatctt c 21

<210> 129
<211> 21
<212> DNA
<213> Homo sapiens

<400> 129
tcagccgctt tcagattttc a 21

<210> 130
<211> 26
<212> DNA
<213> Homo sapiens

<400> 130
accctgctga ccttcttcca ttccgt 26

<210> 131
<211> 22
<212> DNA
<213> Homo sapiens

<400> 131
agaaattggc ttcctgctca ag 22

<210> 132
<211> 25
<212> DNA
<213> Homo sapiens

<400> 132
aaaaccaact tcccaaaaat tctct 25

<210> 133
<211> 23
<212> DNA
<213> Homo sapiens

<400> 133
tgcttccaca agaaccgcga gga 23

<210> 134
<211> 20
<212> DNA
<213> Homo sapiens

<400> 134
cgagccgacc atgtcttcat 20

<210> 135
<211> 21
<212> DNA
<213> Homo sapiens

<400> 135
aagcccaaaa agctgaaggt t 21

<210> 136
<211> 22
<212> DNA
<213> Homo sapiens

<400> 136
aagttcttcc aaatggtcgg cc 22

<210> 137
<211> 19
<212> DNA
<213> Homo sapiens

<400> 137
ccgactcctt cgaccacaa 19

<210> 138
<211> 23
<212> DNA
<213> Homo sapiens

<400> 138
catcatccgc actctttttc ttc 23

<210> 139
<211> 18
<212> DNA
<213> Homo sapiens

<400> 139 18
tggctggcgg tgcctgga 18

<210> 140
<211> 21
<212> DNA
<213> Homo sapiens

<400> 140 21
tgtggttcct gcatgaagac a 21

<210> 141
<211> 24
<212> DNA
<213> Homo sapiens

<400> 141 24
gtgacagcgg aagtggtatt gtac 24

Claims

1. Method for predicting an outcome of breast cancer in an estrogen receptor positive and HER2 negative tumor of a breast cancer patient, said method comprising:

   (a) determining in a tumor sample from said patient the RNA expression levels of the following 8 genes: UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP
   (b) mathematically combining expression level values for the genes of the said set which values were determined in the tumor sample to yield a combined score, wherein said combined score is indicative of a prognosis of said patient.

2. Method according to claim 1 wherein
   UBE2C may be replaced by BIRC5 or TOP2A or AURKA or NEK2 or E2F8 or PCNA or CYBRD1 or ADRA2A or DCN or SQLE or CCND1 or ASPH or CXCL12 or PIP or PRSS16 or EGFR or DHCR7 or EPHX2 or TRIM29 and
   DHCR7 may be replaced by AURKA, BIRC5, UBE2C or by any other gene that may replace BIRC5 or UBE2C and
   while STC2 may be replaced by INPP4B or IL6ST or SEC14L2 or MAPT or CHPT1 or ABAT or SCUBE2 or ESR1 or RBBP8 or PGR or PTPRT or HSPA2 or PTGER3 and
   AZGP1 may be replaced by PIP or EPHX2 or PLAT or SEC14L2 or SCUBE2 or PGR and
   RBBP8 may be replaced by CELSR2 or PGR or STC2 or ABAT or IL6ST and
   IL6ST may be replaced by INPP4B or STC2 or MAPT or SCUBE2 or ABAT or PGR or SEC14L2 or ESR1 or GJA1 or MGP or EPHX2 or RBBP8 or PTPRT or PLAT and
   MGP may be replaced by APOD or IL6ST or EGFR.

3. The method according to claim 1 or 2, wherein said patient has received endocrine therapy or is contemplated to receive endocrine treatment.

4. The method of claim 3, wherein said endocrine therapy comprises tamoxifen or an aromatase inhibitor.

5. The method according to any of claims 1 to 4 wherein a risk of developing breast cancer recurrence or cancer related death is predicted.

6. The method according to claim 1, wherein said expression level is determined by at least one of

   a PCR based method,
   a microrarry based method, and
   a hybridization based method.

7. The method of any one of the preceding claims, wherein said determination of expression levels is in a formalin-fixed paraffin embedded tumor sample or in a fresh-frozen tumor sample.

8. The method of any of the preceding claims, wherein the expression level of at least one marker gene is determined as a pattern of expression relative to at least one reference gene or to a computed average expression value.

9. The method of any one of the preceding claims, wherein said step of mathematically combining comprises a step of applying an algorithm to values representative of expression levels of given genes.

10. The method of claim 9, wherein said algorithm is a linear combination of said values representative of expression levels of given genes.

11. The method of claim 10 wherein a value for a representative value of an expression level of a given gene is multiplied with a coefficient.

12. The method of any one of the preceding claims, wherein one, two or more thresholds are determined for said combined score, that discriminate into high and low risk, high, intermediate and low risk, or more risk groups by applying the threshold on the combined score, in paricular wherein a high combined score is indicative of benefit from cytotoxic chemotherapy.

13. The method of any one of the preceding claims, wherein information regarding nodal status of the patient is processed

in the step of mathematically combining expression level values for the genes to yield a combined score.

14. The method of claims 12 or 13, wherein said information regarding nodal status is a numerical value if said nodal status is negative and said information is a different numerical value if said nodal status positive and a different or identical number if said nodal status is unknown.

15. Use of a kit for performing a method of any of claims 1 to 14, said kit comprising a set of oligonucleotides capable of specifically binding sequences or to sequences of fragments of the genes in a combination of genes, wherein said combination comprises the 8 genes UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, and MGP.

Patentansprüche

1. Verfahren zur Vorhersage eines Ergebnisses bezüglich Brustkrebs in einem Estrogenrezeptor-positiven und HER2-negativen Tumor bei einem Brustkrebspatienten, wobei das Verfahren umfasst:

(a) Bestimmen der RNA-Expressionsniveaus der folgenden 8 Gene in einer Tumorprobe von dem Patienten: UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST und MGP;
(b) mathematisches Kombinieren der Expressionsniveauwerte für die Gene der genannten Menge, wobei die Werte in der Tumorprobe bestimmt wurden, unter Erhalt eines kombinierten Score, wobei der kombinierte Score auf eine Prognose für den Patienten hinweist.

2. Verfahren gemäß Anspruch 1, wobei
UBE2C durch BIRC5 oder TOP2A oder AURKA oder NEK2 oder E2F8 oder PCNA oder CYBRD1 oder ADRA2A oder DCN oder SQLE oder CCND1 oder ASPH oder CXCL12 oder PIP oder PRSS16 oder EGFR oder DHCR7 oder EPHX2 oder TRIM29 ersetzt sein kann; und
DHCR7 durch AURKA, BIRC5, UBE2C oder durch jedes andere Gen, das BIRC5 oder UBE2C ersetzen kann, ersetzt sein kann; und
während STC2 durch INPP4B oder IL6ST oder SEC14L2 oder MAPT oder CHPT1 oder ABAT oder SCUBE2 oder ESR1 oder RBBP8 oder PGR oder PTPRT oder HSPA2 oder PTGER3 ersetzt sein kann; und
AZGP1 durch PIP oder EPHX2 oder PLAT oder SEC14L2 oder SCUBE2 oder PGR ersetzt sein kann; und
RBBP8 durch CELSR2 oder PGR oder STC2 oder ABAT oder IL6ST ersetzt sein kann; und
IL6ST durch INPP4B oder STC2 oder MAPT oder SCUBE2 oder ABAT oder PGR oder SEC14L2 oder ESR1 oder GJA1 oder MGP oder EPHX2 oder RBBP8 oder PTPRT oder PLAT ersetzt sein kann; und
MGP durch APOD oder IL6ST oder EGFR ersetzt sein kann.

3. Verfahren gemäß Anspruch 1 oder 2, wobei der Patient eine endokrine Therapie erhalten hat oder bei ihm eine endokrine Therapie in Betracht gezogen wurde.

4. Verfahren gemäß Anspruch 3, wobei die endokrine Therapie Tamoxifen oder einen Aromatase-Inhibitor umfasst.

5. Verfahren gemäß einem der Ansprüche 1 bis 4, wobei ein Risiko der Entwicklung eines Brustkrebsrezidivs oder eines mit Krebs zusammenhängenden Todes vorhergesagt wird.

6. Verfahren gemäß Anspruch 1, wobei das Expressionsniveau durch wenigstens eines der folgenden Verfahren bestimmt wird:

ein Verfahren auf PCR-Basis;
ein Verfahren auf Mikroarray-Basis; und
ein Verfahren auf Hybridisierungsbasis.

7. Verfahren gemäß einem der vorstehenden Ansprüche, wobei die Bestimmung der Expressionsniveaus in einer mit Formalin fixierten und in Paraffin eingebetteten Tumorprobe oder in einer frisch eingefrorenen Tumorprobe erfolgt.

8. Verfahren gemäß einem der vorstehenden Ansprüche, wobei das Expressionsniveau von wenigstens einem Marker-Gen als Expressionsmuster relativ zu wenigstens einem Referenz-Gen oder zu einem berechneten mittleren Expressionswert bestimmt wird.

9. Verfahren gemäß einem der vorstehenden Ansprüche, wobei der Schritt des mathematischen Kombinierens einen Schritt des Anwendens eines Algorithmus auf Werte, die repräsentativ für Expressionsniveaus gegebener Gene sind, umfasst.

10. Verfahren gemäß Anspruch 9, wobei der Algorithmus eine Linearkombination der Werte ist, die repräsentativ für Expressionsniveaus gegebener Gene sind.

11. Verfahren gemäß Anspruch 10, wobei ein Wert, der repräsentativ für ein Expressionsniveau eines gegebenen Gens ist, mit einem Koeffizienten multipliziert wird.

12. Verfahren gemäß einem der vorstehenden Ansprüche, wobei eine, zwei oder mehr Schwellen für den kombinierten Score bestimmt werden, die zwischen hohen und geringem Risiko, hohem, mittlerem und geringem Risiko oder mehr Risikogruppen diskriminieren, indem man die Schwelle auf den kombinierten Score anwendet, wobei ein hoher kombinierter Score insbesondere den Nutzen einer cytotoxischen Chemotherapie anzeigt.

13. Verfahren gemäß einem der vorstehenden Ansprüche, wobei in dem Schritt des mathematischen Kombinierens der Expressionsniveauwerte für die Gene Informationen bezüglich des Nodalstatus des Patienten verarbeitet werden, wobei man einen kombinierten Score erhält.

14. Verfahren gemäß Anspruch 12 oder 13, wobei es sich bei den Informationen bezüglich des Nodalstatus um einen numerischen Wert handelt, wenn der Nodalstatus negativ ist, und es sich bei den Informationen um einen anderen numerischen Wert handelt, wenn der Nodalstatus positiv ist, und es sich um eine andere oder identische Zahl handelt, wenn der Nodalstatus unbekannt ist.

15. Verwendung eines Kits zur Durchführung eines Verfahrens gemäß einem der Ansprüche 1 bis 14, wobei der Kit eine Menge von Oligonucleotiden umfasst, die spezifisch Sequenzen oder an Sequenzen von Fragmenten der Gene in einer Kombination von Genen binden können, wobei die Kombination die folgenden 8 Gene umfasst: UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST und MGP.

Revendications

1. Procédé pour prédire une évolution d'un cancer du sein dans une tumeur positive au récepteur des oestrogènes et négative pour HER2 chez un patient atteint de cancer du sein, ledit procédé comprenant :

   (a) la détermination, dans un échantillon tumoral provenant dudit patient, du niveau d'expression de l'ARN des 8 gènes suivants : UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST et MGP;
   (b) la combinaison mathématique des valeurs de niveau d'expression par les gènes dudit ensemble, lesdites valeurs ayant été déterminées dans l'échantillon tumoral, pour donner un score combiné, dans lequel ledit score combiné est indicatif d'un pronostic dudit patient.

2. Procédé selon la revendication 1, dans lequel **UBE2C** peut être remplacé par BIRC5 ou TOP2A ou AURKA ou NEK2 ou E2F8 ou PCNA ou CYBRD1 ou ADRA2A ou DCN ou SOLE ou CCND1 ou ASPH ou CXCL12 ou PIP ou PRSS16 ou EGFR ou DHCR7 ou EPHX2 ou TRIM29 et
   DHCR7 peut être remplacé par AURKA, BIRC5, UBE2C ou par tout autre gène pouvant remplacer BIRC5 ou UBE2C et **alors que STC2** peut être remplacé par INPP4B ou IL6ST ou SEC14L2 ou MAPT ou CHPT1 ou ABAT ou SCUBE2 ou ESR1 ou RBBP8 ou PGR ou PTPRT ou HSPA2 ou PTGER3 et
   AZGP1 peut être remplacé par PIP ou EPHX2 ou PLAT ou SEC14L2 ou SCUBE2 ou PGR et
   RBBP8 peut être remplacé par CELSR2 ou PGR ou STC2 ou ABAT ou IL6ST et
   IL6ST peut être remplacé par INPP4B ou STC2 ou MAPT ou SCUBE2 ou ABAT ou PGR ou SEC14L2 ou ESR1 ou GJA1 ou MGP ou EPHX2 ou RBBP8 ou PTPRT ou PLAT et
   MGP peut être remplacé par APOD ou IL6ST ou EGFR.

3. Procédé selon la revendication 1 ou 2, dans lequel ledit patient a reçu une thérapie endocrinienne ou est envisagé pour recevoir un traitement endocrinien.

4. Procédé selon la revendication 3, dans lequel ladite thérapie endocrinienne comprend du tamoxifène ou un inhibiteur de l'aromatase.

5. Procédé selon l'une quelconque des revendications 1 à 4, dans lequel un risque de récidive de cancer du sein ou de mort liée au cancer est prédit.

6. Procédé selon la revendication 1, dans lequel ledit niveau d'expression est déterminé par au moins un des procédés suivants :

- procédé basé sur une PCR,
- procédé basé sur un microréseau, et
- procédé basé sur une hybridation.

7. Procédé selon l'une quelconque des revendications précédentes, dans lequel ladite détermination des niveaux d'expression s'effectue dans un échantillon tumoral fixé de formaline enrobé dans la paraffine ou dans un échantillon tumoral frais congelé.

8. Procédé selon l'une quelconque des revendications précédentes, dans lequel le niveau d'expression d'au moins un gène marqueur est déterminé sous la forme d'un profil d'expression par rapport à au moins un gène de référence ou à une valeur d'expression moyenne calculée.

9. Procédé selon l'une quelconque des revendications précédentes, dans lequel ladite étape de combinaison mathématique comprend une étape consistant à appliquer un algorithme à des valeurs représentatives du niveau d'expression de gènes donnés.

10. Procédé selon la revendication 9, dans lequel ledit algorithme est une combinaison linéaire desdites valeurs représentatives du niveau d'expression de gènes donnés.

11. Procédé selon la revendication 10, dans lequel une valeur d'une valeur représentative d'un niveau d'expression d'un gène donné est multipliée par un coefficient.

12. Procédé selon l'une quelconque des revendications précédentes, dans lequel un, deux ou plus de deux seuils sont déterminés pour ledit score combiné, seuils qui discriminent des groupes à risque élevé et faible, des groupes à risque élevé, intermédiaire et faible, ou davantage de groupes de risque, en appliquant le seuil au score combiné, en particulier dans lequel un score combiné élevé est indicatif d'un bénéfice d'une chimiothérapie cytotoxique.

13. Procédé selon l'une quelconque des revendications précédentes, dans lequel des informations sur le statut ganglionnaire du patient sont traitées à l'étape de combinaison mathématique de valeurs de niveau d'expression des gènes pour donner un score combiné.

14. Procédé selon les revendications 12 ou 13, dans lequel lesdites informations sur le statut ganglionnaire sont une valeur numérique si ledit statut ganglionnaire est négatif et lesdites informations sont une valeur numérique différente si ledit statut ganglionnaire est positif et un nombre identique ou différent si ledit statut ganglionnaire est inconnu.

15. Utilisation d'un kit pour réaliser un procédé selon l'une quelconque des revendications 1 à 14, ledit kit comprenant un ensemble d'oligonucléotides capables de lier spécifiquement des séquences ou de se lier spécifiquement à des séquences de fragments des gènes dans une combinaison de gènes, dans laquelle ladite combinaison comprend les 8 gènes UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST et MGP.

Time to distant metastasis

Trial	Events / Patients	Adjusted hazard ratio (95% CI)
ABCSG06 Tam	26 / 332	1.26 (1.03-1.55)
ABCSG08 Tam	46 / 600	1.30 (1.10-1.54)
ABCSG08 Tam+Ari	36 / 644	1.45 (1.22-1.71)
Total	108 / 1576	1.34 (1.21-1.49)

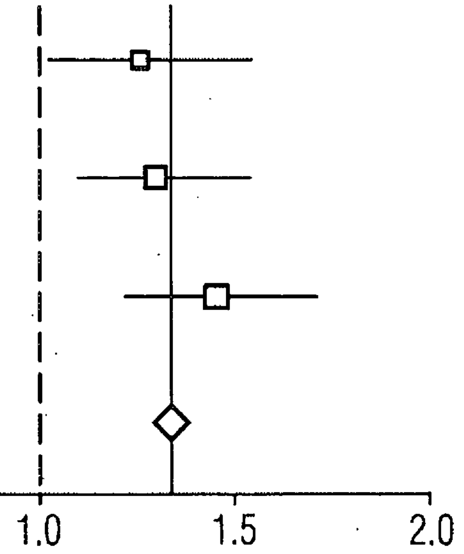

Fig.1

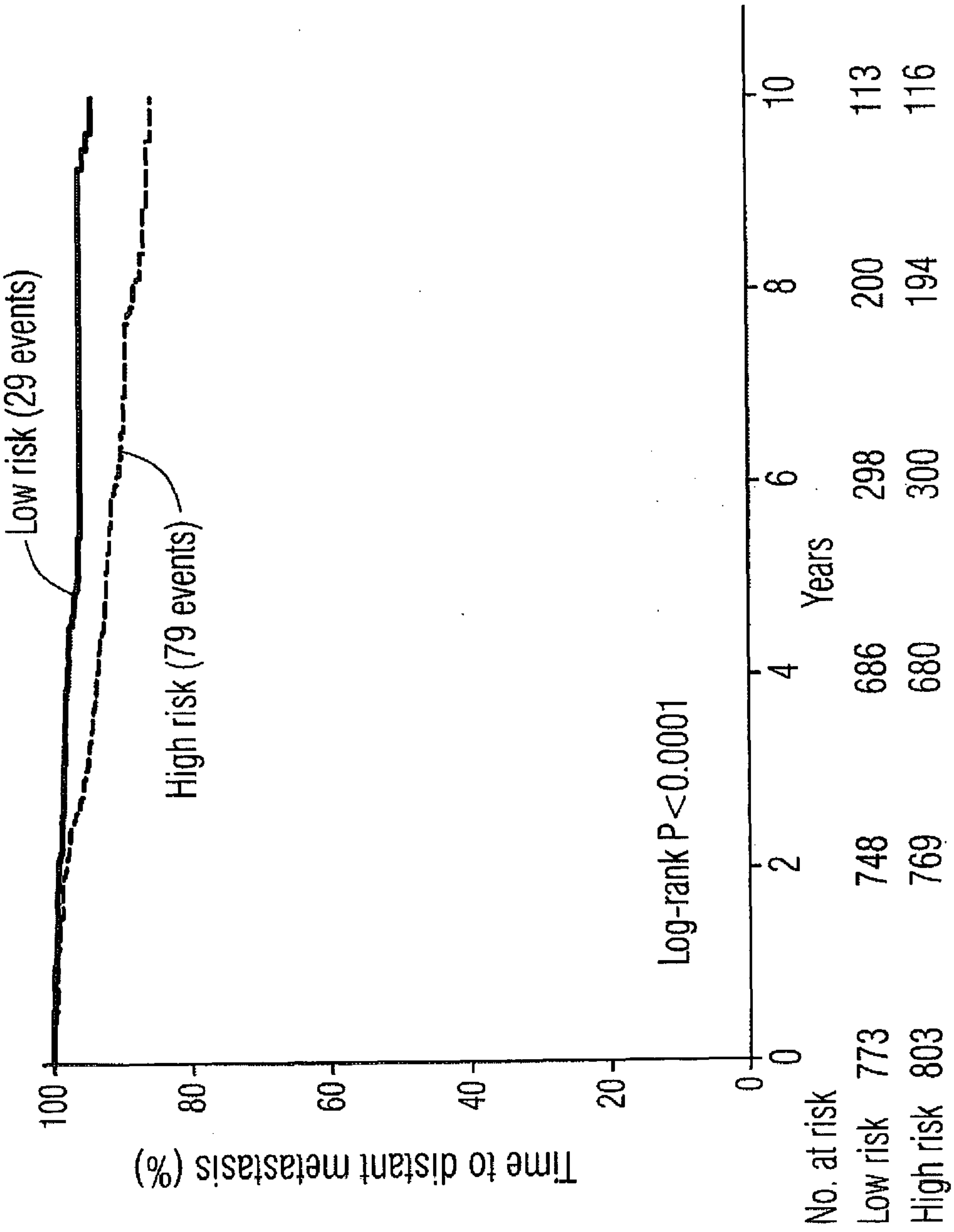
Time to distant metastasis (%)
100
80
60
40
20
0
0
2
4
6
8
10
Years
Low risk (29 events)
High risk (79 events)
Log-rank P<0.0001
No. at risk
Low risk 773 748 686 298 200 113
High risk 803 769 680 300 194 116

Fig.2

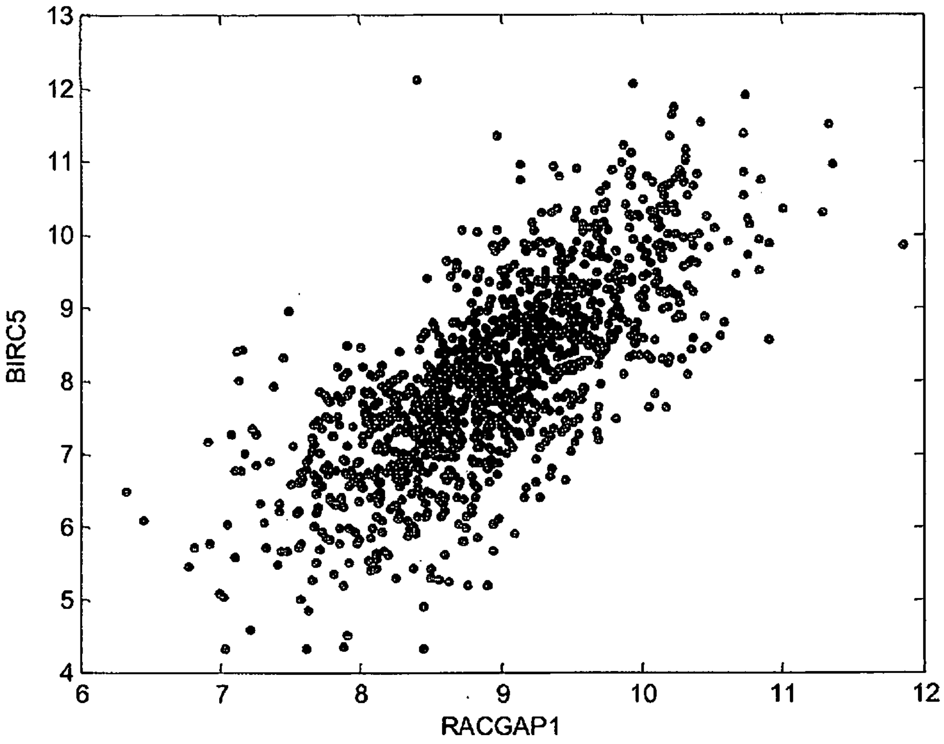

Fig.3

REFERENCES CITED IN THE DESCRIPTION

Patent documents cited in the description

- EP 11710526 A **[0119]**

SBGK
Szabadalmi
Ügyvivői Iroda

715.209/SZE

Eljárás emlőrák kiújulásának előrejelzésére endokrin kezelés alatt

SZABADALMI IGÉNYPONTOK

1. Eljárás emlőrák kimenetelének előrejelzésére egy emlőrákos beteg ösztrogén receptor pozitív és HER2 negatív tumora esetében, azzal jellemezve, hogy az eljárás a következő lépéseket tartalmazza:

(a) egy, a szóban forgó betegből származó tumormintában meghatározzuk az alábbi 8 gén RNS-expressziós szintjét: UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, és MGP

(b) a szóban forgó csoport génjeinek egy tumormintában meghatározott expressziós szintjeinek értékeit matematikailag kombináljuk, hogy egy kombinált pontszámot kapjunk, ahol a szóban forgó kombinált pontszám a szóban forgó beteg prognózisára jellemző.

2. Az 1. igénypont szerinti eljárás, azzal jellemezve, hogy

az **UBE2C** helyettesíthető BIRC5-tel vagy TOP2A-val vagy AURKA-val vagy NEK2-vel vagy E2F8-cal vagy PCNA-val vagy CYBRD1-gyel vagy ADRA2A-val vagy DCN-nel vagy SQLE-vel vagy CCND1-gyel vagy ASPH-val vagy CXCL12-vel vagy PIP-pel vagy PRSS16-tal vagy EGFR-rel vagy DHCR7-tel vagy EPHX2-vel vagy TRIM29-cel és

a DHCR7 helyettesíthető AURKA-val, BIRC5-tel, UBE2C-vel vagy bármilyen más génnel, amely képes helyettesíteni BIRC5-öt vagy UBE2C-t és

míg az STC2 helyettesíthető INPP4B-vel vagy IL6ST-vel vagy SEC14L2-vel vagy MAPT-vel vagy CHPT1-gyel vagy ABAT-tal vagy SCUBE2-vel vagy ESR1-gyel vagy RBBP8-cal vagy PGR-rel vagy PTPRT-vel vagy HSPA2-vel vagy PTGER3-mal, és

az **AZGP1** helyettesíthető PIP-pel vagy EPHX2-vel vagy PLAT-tal vagy SEC14L2-vel vagy SCUBE2-vel vagy PGR-rel és

az **RBBP8** helyettesíthető CELSR2-vel vagy PGR-rel vagy STC2-vel vagy ABAT-tal vagy IL6ST-vel és

az **IL6ST** helyettesíthető INPP4B-vel vagy STC2-vel vagy MAPT-vel vagy SCUBE2-vel vagy ABAT-tal vagy PGR-rel vagy SEC14L2-vel vagy ESR1-gyel vagy GJA1-gyel vagy MGP-vel vagy EPHX2-vel vagy RBBP8-cal vagy PTPRT-vel vagy PLAT-tal és

az **MGP** helyettesíthető APOD-dal vagy IL6ST-vel vagy EGFR-rel.

3. Az 1. vagy 2. igénypont szerinti eljárás, azzal jellemezve, hogy a szóban forgó beteg endokrin terápiát kapott, vagy tervezik, hogy endokrin kezelést kap.

4. A 3. igénypont szerinti eljárás, azzal jellemezve, hogy a szóban forgó endokrin terápia tamoxifent vagy egy aromatáz inhibitort tartalmaz.

5. Az 1-4. igénypontok bármelyike szerinti eljárás, azzal jellemezve, hogy az emlőrák kiújulásának vagy a rákhoz köthető halálnak a kockázatát előre jelezzük.

6. Az 1. igénypont szerinti eljárás, azzal jellemezve, hogy a szóban forgó expressziós szintet az alábbi eljárások közül legalább egynek a használatával határozzuk meg

egy PCR-alapú eljárás

egy mikrotömb alapú eljárás, és

egy hibridizáción alapuló eljárás.

7. Az előző igénypontok bármelyike szerinti eljárás, azzal jellemezve, hogy az expressziós szintek szóban forgó meghatározását egy formalinnal rögzített, beágyazott tumormintából, vagy egy frissen lefagyasztott tumormintából hajtjuk végre.

8. Az előző igénypontok bármelyike szerinti eljárás, azzal jellemezve, hogy legalább egy markergén expressziós szintjét meghatározzuk egy expressziós mintázatként, legalább egy referencia génhez, vagy egy számított átlagos expressziós értékhez viszonyítva.

9. Az előző igénypontok bármelyike szerinti eljárás, azzal jellemezve, hogy a matematikai kombinálás szóban forgó lépése tartalmaz egy lépést, amelyben egy algoritmust alkalmazunk az adott gének expressziós szintjeit reprezentáló értékekre.

10. A 9. igénypont szerinti eljárás, azzal jellemezve, hogy a szóban forgó algoritmus az adott gének expressziós szintjét reprezentáló szóban forgó értékek lineáris kombinációja.

11. A 10. igénypont szerinti eljárás, azzal jellemezve, hogy egy adott gén expressziós szintje reprezentációs értékének az értékét egy koefficienssel szorozzuk.

12. Az előző igénypontok bármelyike szerinti eljárás, azzal jellemezve, hogy egy, kettő, vagy több küszöbértéket határozunk meg a szóban forgó kombinált pontszámra, amely magas és alacsony kockázatú, magas, közepes és alacsony kockázatú, vagy több kockázati csoportot különböztet meg a küszöbértéket alkalmazva a kombinált pontszámra, különösen, ha egy magas kombinált pontszám a citotoxikus kemoterápiából származó jótékony hatást jelzi.

13. Az előző igénypontok bármelyike szerinti eljárás, azzal jellemezve, hogy a beteg nyirokcsomó státuszára vonatkozó információt abban a lépésben dolgozzuk fel, amelyben a gének expressziós szintjeit matematikailag kombináljuk, hogy kombinált pontszámot kapjunk.

14. A 12. vagy 13. igénypont szerinti eljárás, azzal jellemezve, hogy a nyirokcsomó státuszára vonatkozó információ egy numerikus érték, ha a szóban forgó nyirokcsomó státusz negatív, és a szóban forgó információ egy eltérő számérték, ha a szóban forgó nyirokcsomó státusz pozitív, és különböző vagy azonos szám, ha a szóban forgó nyirokcsomó státusz ismeretlen.

15. Egy készlet alkalmazása az 1-14. igénypontok bármelyike szerinti eljárás végrehajtására, a szóban forgó készlet oligonukleotidok egy csoportját tartalmazza, amelyek képesek specifikusan megkötni szekvenciákat, vagy a gének fragmenseinek a szekvenciáit gének kombinációiban, ahol a szóban forgó kombináció a következő 8 gént tartalmazza: UBE2C, RACGAP1, DHCR7, STC2, AZGP1, RBBP8, IL6ST, és MGP.

Ifj. Szentpéteri Ádám
szabadalmi ügyvivő
SBGK Szabadalmi Ügyvivői Iroda
H-1062 [illegible] Andrássy út 113.
Telefon: [illegible] Fax: 461-1099
Email: [illegible]sbgk.hu