

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 January 2007 (11.01.2007)

PCT

(10) International Publication Number
WO 2007/003397 A2

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

(21) International Application Number:
PCT/EP2006/006458

(22) International Filing Date: 3 July 2006 (03.07.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
05090202.2 1 July 2005 (01.07.2005) EP
05090201.4 1 July 2005 (01.07.2005) EP
06090050.3 7 April 2006 (07.04.2006) EP
06090081.8 18 May 2006 (18.05.2006) EP

(71) Applicant (for all designated States except US): **EPIGENOMICS AG** [DE/DE]; Kleine Präsidentenstrasse 1, 10178 Berlin (DE).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **DIETRICH, Dima** [DE/DE]; Greifenhagenerstrasse 39, 10437 Berlin (DE).

(74) Agents: **KRAUSS, Jan, B.** et al.; Boehmert & Boehmert, Pettenkoferstrasse 20-22, 80336 München (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND NUCLEIC ACIDS FOR THE IMPROVED TREATMENT OF CANCERS

(57) Abstract: The present invention provides methods and nucleic acids for determining the prognosis and predicting treatment response of subjects with cancer.



WO 2007/003397 A2

Method and nucleic acids for the improved treatment of cancers.**Field of the Invention**

In American women, breast cancer is the most frequently diagnosed cancer and the second leading cause of cancer death. In women aged 40-55, breast cancer is the leading cause of death (Greenlee *et al.*, 2000). In 2002, there were 204,000 new cases of breast cancer in the US (data from the American Society of Clinical Oncology) and a comparable number in Europe.

Breast cancer is defined as the uncontrolled proliferation of cells within breasts tissues. Breasts are comprised of 15 to 20 lobes joined together by ducts. Cancer arises most commonly in the duct, but is also found in the lobes with the rarest type of cancer termed inflammatory breast cancer. It will be appreciated by those skilled in the art that there exists a continuing need to improve methods of early detection, classification and treatment of breast cancers. In contrast to the detection of some other common cancers such as cervical and dermal there are inherent difficulties in classifying and detecting breast cancers.

Due to current screening programs and the accessibility of this cancer to self-examination, breast cancer is diagnosed comparatively early: in about 93% of all newly diagnosed cases, the cancer has not yet metastasized, and in 65% of cases, even the lymph nodes are not yet affected.

The first step of any treatment is the assessment of the patient's condition comparative to defined classifications of the disease. However the value of such a system is inherently dependent upon the quality of the classification. Breast cancers are staged according to their size, location and occurrence of metastasis. Methods of treatment include the use of surgery, radiation therapy, chemotherapy and endocrine therapy, which are also used as adjuvant therapies to surgery.

Although the vast majority of early cancers are operable, i.e. the tumor can be completely removed by surgery, about one third of the patients with lymph-node negative diseases and

about 50-60% of patients with node-positive disease will develop metastases during follow-up.

Based on this observation, systemic adjuvant treatment has been introduced for both node-positive and node-negative breast cancers. Systemic adjuvant therapy is administered after surgical removal of the tumor, and has been shown to reduce the risk of recurrence significantly (Early Breast Cancer Trialists' Collaborative Group, 1998). Several types of adjuvant treatment are available: endocrine treatment (for hormone receptor positive tumors), different chemotherapy regimens, and novel agents like Herceptin.

The selection of suitable adjuvant systemic therapies is determined according to an assessment of the patient's risk of recurrence. Risk of recurrence is assessed primarily according to node status, histological grade, tumour size, oestrogen receptor (ER) status of the primary tumour and menopausal status. Other factors that may be taken into consideration include *cerbB2* expression, ratio of lymph nodes positive vs number of lymph nodes resected, presence of vascular invasion and age.

According to the risk of recurrence appropriate treatments may be selected that provide a reduction in risk of recurrence or death. Chemotherapy is often prescribed as an adjuvant systematic therapy. The proportional reduction of risk of recurrence and death for any given chemotherapy regimen is fairly constant within defined age and hormone receptor categories but the absolute benefit achieved varies as a function of a patient's risk. Accordingly, in order to determine whether a patient will benefit from chemotherapy it is necessary to accurately determine the risk of recurrence or death.

Current guidelines recommend adjuvant chemotherapy followed by endocrine therapy for most of the women with node-negative, steroid hormone receptor-positive breast cancer. This recommendation is based on a significant reduction of the risk of disease recurrence by chemotherapy in this population, independent of the risk reduction by endocrine therapy. Yet, these patients have a rather good prognosis and in general derive significant benefit from endocrine treatment. Thus, after endocrine treatment, the majority will never experience a recurrence and would be adequately treated by tamoxifen alone. Unfortunately, traditional prognostic factors are not suited to identify patients at low risk in order to avoid over-treatment by chemotherapy in the majority of hormone receptor-positive, node-negative

patients. Moreover, in postmenopausal hormone receptor-positive patients, aromatase inhibitors have become a valid option. Yet, it is still unclear which patients will be sufficiently treated by adjuvant tamoxifen and who will benefit from aromatase inhibitors— an important question given the lack of information on long-term side effects and the increased costs for aromatase inhibitors.

Molecular markers associated with breast cancer prognosis are known, including methylation markers, for the measurement of risk of relapse or survival. PCT/EP2004/014170 discloses a large group of markers the expression, as determined most preferably by methylation status, thereof being indicative of the prognosis of breast cancer patients in terms of overall survival or relapse. The present invention provide novel markers the methylation of which are indicative of the prognosis of a patient with breast cancer in terms of survival and/or relapse.

Widschwendter et al. (Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. *Cancer Res* 2004;64:3807-13.) described an association of ESR1 hypermethylation with favorable outcome in patients receiving adjuvant tamoxifen. Several groups have identified mRNA expression patterns associated with outcome in breast cancer. Van't Veer et al. (*Cancer Res.* 2005 May 15;65(10):4059-66.) identified prognostic signatures in breast cancer patients. Paik et al. (*N Engl J Med.* 2004 Dec 30;351(27):2817-26) developed a recurrence score for tamoxifen-treated patients based on 21 genes. The score assigns 51% of patients into a low-risk group (10-year-MFS 93.2%), 22% of patients into an intermediate-risk group (10-year-MFS 85.7%), and 27% into a high-risk group (10-year-MFS 69.5%).

Anthracyclines are a large group of compounds synthesized by different *Streptomyces* species. They possess antibiotic activity and have cytotoxic effects on eukaryotic cells. All anthracyclines have a tetrahydronaphthacenedione ring structure attached by a glycosidic linkage to a sugar molecule, structural diversity of anthracyclines is generated by modifications of the backbone including a large number of different side chains.

Anthracyclines have excellent antineoplastic activity in metastatic, neoadjuvant, and adjuvant settings and are used in the treatment of various haematopoietic and solid tumours. Commonly used anthracyclines include but are not limited to mitoxantrone, doxorubicin, aclarubicin, daunorubicin, epirubicin and idarubicin. Although their mechanism of

chemotherapeutic action is unclear involves noncovalent DNA intercalation, formation of covalent DNA adducts, topoisomerase II (topo II) poisoning, and free radical effects on cellular membranes and DNA. However, the clinical utility of anthracyclines are limited due to acute and chronic toxicities, particularly cardiotoxicity, myelosuppression, nausea and vomiting, and alopecia.

Heart failure following anthracycline therapy is a major clinical problem in cancer treatment. The establishment of predictors of the anthracycline treatment outcome would allow the identification and exclusion of individuals who would not benefit from said treatment, and thus to increase the safety of anthracycline treatment. Furthermore by determining which patients would benefit from Anthracycline treatment, but wherein said predicted outcome is sub-optimal patients can be recommended for further chemotherapeutic or other treatments. Conversely by determining which patients would be adequately treated by anthracycline treatment alone the over-treatment of patients can be prevented. Accordingly there is a longfelt need in the art for determining which patients will benefit from Anthracycline treatment.

Methylation of the gene Topo II alpha gene was recently observed in the cell line K562/MX2, which displays resistance to the anthracyclines KRN 8602 (MX2), etoposide and doxorubicin (Asano et al. Br J Cancer. 2005 Apr 25;92(8):1486-92.). Sensitivity to the drug was restored by treatment with the demethylating agent 5-Aza-2'-deoxycytidine, thereby implying that Topo Iialpha methylation is a mechanism of drug resistance. The person skilled in the art when considering WO 2004/035803 in light of Asano et al. would not have a reasonable expectation of success that a methylation marker indicative of response to treatment targeting a hormone pathway would be a predictor of response to a treatment with an unrelated mechanism of action.

In humans, DNA methylation occurs in mainly in the context of CpG duncleotides. Bisulfite modification of DNA *is an art-recognized tool used to assess CpG methylation status*. 5-methylcytosine is the most frequent covalent base modification in the DNA of eukaryotic cells. It plays a role, for example, in the regulation of the transcription, in genetic imprinting, and in tumorigenesis. Therefore, the identification of 5-methylcytosine as a component of genetic information is of considerable interest. However, 5-methylcytosine positions cannot be identified by sequencing, because 5-methylcytosine has the same base pairing behavior as

cytosine. Moreover, the epigenetic information carried by 5-methylcytosine is completely lost during, bisulfite amplification.

The most frequently used method for analyzing DNA for the presence of 5-methylcytosine is based upon the specific reaction of bisulfite with cytosine whereby, upon subsequent alkaline hydrolysis, cytosine is converted to uracil which corresponds to thymine in its base pairing behavior. Significantly, however, 5-methylcytosine remains unmodified under these conditions. Consequently, the original DNA is *converted* in such a manner that methylcytosine, which originally could not be distinguished from cytosine by its hybridization behavior, can now be detected as the only remaining cytosine using standard, art-recognized molecular biological techniques, for example, by amplification and hybridization, or by sequencing. All of these techniques are based on differential base pairing properties, which can now be fully exploited.

The prior art, in terms of sensitivity, is defined by a method comprising enclosing the DNA to be analyzed in an agarose matrix, thereby preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and replacing all precipitation and purification steps with fast dialysis (Olek A, et al., A modified and improved method for bisulfite based cytosine methylation analysis, *Nucleic Acids Res.* 24:5064-6, 1996). It is thus possible to analyze individual cells for methylation status, illustrating the utility and sensitivity of the method. An overview of art-recognized methods for detecting 5-methylcytosine is provided by Rein, T., et al., *Nucleic Acids Res.*, 26:2255, 1998.

The bisulfite technique, barring few exceptions (*e.g.*, Zeschnigk M, et al., *Eur J Hum Genet.* 5:94-98, 1997), is currently only used in research. In all instances, short, specific fragments of a known gene are amplified subsequent to a bisulfite treatment, and either completely sequenced (Olek and Walter, *Nat Genet.* 1997 17:275-6, 1997), subjected to one or more primer extension reactions (Gonzalzo and Jones, *Nucleic Acids Res.*, 25:2529-31, 1997; WO 95/00669; U.S. Patent No. 6,251,594) to analyze individual cytosine positions, or treated by enzymatic digestion (Xiong and Laird, *Nucleic Acids Res.*, 25:2532-4, 1997). Detection by hybridization has also been described in the art (Olek et al., WO 99/28498). Additionally, use of the bisulfite technique for methylation detection with respect to individual genes has been described (Grigg and Clark, *Bioessays*, 16:431-6, 1994; Zeschnigk M, et al., *Hum Mol Genet.*, 6:387-95, 1997; Feil R, et al., *Nucleic Acids Res.*, 22:695-, 1994; Martin V, et al., *Gene*,

157:261-4, 1995; WO 9746705 and WO 9515373).

Methylation Assay Procedures. Various methylation assay procedures are known in the art, and can be used in conjunction with the present invention. These assays allow for determination of the methylation state of one or a plurality of CpG dinucleotides (*e.g.*, CpG islands) within a DNA sequence. Such assays involve, among other techniques, DNA sequencing of bisulfite-treated DNA, PCR (for sequence-specific amplification), Southern blot analysis, and use of methylation-sensitive restriction enzymes.

For example, genomic sequencing has been simplified for analysis of DNA methylation patterns and 5-methylcytosine distribution by using bisulfite treatment (Frommer et al., *Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). Additionally, restriction enzyme digestion of PCR products amplified from bisulfite-converted DNA is used, *e.g.*, the method described by Sadri and Hornsby (*Nucl. Acids Res.* 24:5058-5059, 1996), or COBRA (Combined Bisulfite Restriction Analysis) (Xiong and Laird, *Nucleic Acids Res.* 25:2532-2534, 1997).

COBRA. COBRATM analysis is a quantitative methylation assay useful for determining DNA methylation levels at specific gene loci in small amounts of genomic DNA (Xiong and Laird, *Nucleic Acids Res.* 25:2532-2534, 1997). Briefly, restriction enzyme digestion is used to reveal methylation-dependent sequence differences in PCR products of sodium bisulfite-treated DNA. Methylation-dependent sequence differences are first introduced into the genomic DNA by standard bisulfite treatment according to the procedure described by Frommer et al. (*Proc. Natl. Acad. Sci. USA* 89:1827-1831, 1992). PCR amplification of the bisulfite converted DNA is then performed using primers specific for the CpG islands of interest, followed by restriction endonuclease digestion, gel electrophoresis, and detection using specific, labeled hybridization probes. Methylation levels in the original DNA sample are represented by the relative amounts of digested and undigested PCR product in a linearly quantitative fashion across a wide spectrum of DNA methylation levels. In addition, this technique can be reliably applied to DNA obtained from microdissected paraffin-embedded tissue samples.

MethyLightTM. The MethyLightTM assay is a high-throughput quantitative methylation assay that utilizes fluorescence-based real-time PCR (TaqManTM) technology that requires no further manipulations after the PCR step (Eads et al., *Cancer Res.* 59:2302-2306, 1999).

Briefly, the MethyLight™ process begins with a mixed sample of genomic DNA that is converted, in a sodium bisulfite reaction, to a mixed pool of methylation-dependent sequence differences according to standard procedures (the bisulfite process converts unmethylated cytosine residues to uracil). Fluorescence-based PCR is then performed either in an “unbiased” (with primers that do not overlap known CpG methylation sites) PCR reaction, or in a “biased” (with PCR primers that overlap known CpG dinucleotides) reaction. Sequence discrimination can occur either at the level of the amplification process or at the level of the fluorescence detection process, or both.

The MethyLight™ assay may be used as a quantitative test for methylation patterns in the genomic DNA sample, wherein sequence discrimination occurs at the level of probe hybridization. In this quantitative version, the PCR reaction provides for unbiased amplification in the presence of a fluorescent probe that overlaps a particular putative methylation site. An unbiased control for the amount of input DNA is provided by a reaction in which neither the primers, nor the probe overlie any CpG dinucleotides. Alternatively, a qualitative test for genomic methylation is achieved by probing of the biased PCR pool with either control oligonucleotides that do not “cover” known methylation sites (a fluorescence-based version of the “MSP” technique), or with oligonucleotides covering potential methylation sites.

Ms-SNuPE. The Ms-SNuPE™ technique is a quantitative method for assessing methylation differences at specific CpG sites based on bisulfite treatment of DNA, followed by single-nucleotide primer extension (Gonzalzo and Jones, *Nucleic Acids Res.* 25:2529-2531, 1997). Briefly, genomic DNA is reacted with sodium bisulfite to convert unmethylated cytosine to uracil while leaving 5-methylcytosine unchanged. Amplification of the desired target sequence is then performed using PCR primers specific for bisulfite-converted DNA, and the resulting product is isolated and used as a template for methylation analysis at the CpG site(s) of interest. An oligonucleotide is hybridized next to or close to the CpG position of interest, the oligonucleotide is then extended and on the basis of said extension the methylation status of CpG position of interest is determined. Small amounts of DNA can be analyzed (*e.g.*, microdissected pathology sections), and it avoids utilization of restriction enzymes for determining the methylation status at CpG sites.

MSP. MSP (methylation-specific PCR) allows for assessing the methylation status of virtually

any group of CpG sites within a CpG island, independent of the use of methylation-sensitive restriction enzymes (Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996; US Patent No. 5,786,146). Briefly, DNA is modified by sodium bisulfite converting all unmethylated, but not methylated cytosines to uracil, and subsequently amplified with primers specific for methylated versus unmethylated DNA. MSP requires only small quantities of DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed on DNA extracted from paraffin-embedded samples.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides a schematic view of the gene LHX3 the arrows indicate the regions of the gene sequenced in Example 2.

Figure 2 shows the quantified levels of methylation within LHX3 amplicate A of the four sample groups as measured in Example 2. Percentage methylation is shown on the Y-axis.

Figure 3 shows an alternative view of the quantified levels of methylation as measured using Amplicate A of LHX3 in Example 2. Each row of the matrix represents a single CpG site within the fragment and each column represents an individual DNA sample. The bar on the left shows a scale of the percent methylation, with the degree of methylation represented by the shade of each position within the column from black representing 100% methylation to light gray representing 0% methylation. White positions represent a measurement for which no data was available.

Figure 4 shows the correlation of methylation measured at a second region of the gene LHX3 in Example 2, amplicate B as shown in figure 1 within the LHX3 gene show strong co-methylation to amplicon LHX3 A. This indicates that this region (B) is a prognostic biomarker as well as region A.

Figure 5 shows the quantified levels of methylation within PITX3 amplicate A of the four sample groups as measured in Example 2. Percentage methylation is shown on the Y-axis. Group B and C show high PITX3 methylation whereas groups A and D show low methylation. This is in strong correlation to the PITX2 methylation of the same samples.

Figure 6 shows an alternative view of the quantified levels of PITX3 methylation as measured in Example 2. Each row of the matrix represents a single CpG site within the fragment and each column represents an individual DNA sample. The bar on the left shows a scale of the percent methylation, with the degree of methylation represented by the shade of each position within the column from black representing 100% methylation to light gray representing 0% methylation. White positions represented a measurement for which no data was available.

Figure 7 provides a schematic view of the gene PITX3 the arrow indicates the region of the gene sequenced in Example 2.

DESCRIPTION

Characterization of a cancer in terms of prognosis and more specifically in terms of predicting treatment outcome enables the physician to make an informed decision as to a therapeutic regimen with appropriate risk and benefit trade offs to the patient.

In the context of the present invention the terms "estrogen receptor positive" and/or "progesterone receptor positive" when used to describe a cell proliferative disorder are taken to mean that the proliferating cells express said hormone receptor.

In the context of the present invention the term 'aggressiveness' is taken to mean one or more of high likelihood of relapse post surgery; below average or below median patient survival; below average or below median disease free survival; below average or below median relapse-free survival; above average tumor-related complications; fast progression of tumor or metastases. According to the aggressiveness of the disease an appropriate treatment or treatments may be selected from the group consisting of chemotherapy, radiotherapy, surgery, biological therapy, immunotherapy, antibody treatments, treatments involving molecularly targeted drugs, estrogen receptor modulator treatments, estrogen receptor down-regulator treatments, aromatase inhibitors treatments, ovarian ablation, treatments providing LHRH analogues or other centrally acting drugs influencing estrogen production. Wherein a cancer is characterized as 'aggressive' it is particularly preferred that a treatment such as, but not limited to, chemotherapy is provided in addition to or instead of an endocrine targeting therapy.

Indicators of tumor aggressiveness standard in the art include but are not limited to, tumor stage, tumor grade, nodal status and survival.

Unless stated otherwise as used herein the term "survival" shall be taken to include all of the following: survival until mortality, also known as overall survival (wherein said mortality may be either irrespective of cause or tumor related); "recurrence-free survival" (wherein the term recurrence shall include both localized and distant recurrence) ; metastasis free survival; disease free survival (wherein the term disease shall include cancer and diseases associated therewith). The length of said survival may be calculated by reference to a defined start point (e.g. time of diagnosis or start of treatment) and end point (e.g. death, recurrence or metastasis).

As used herein the term "prognostic marker" shall be taken to mean an indicator of the likelihood of progression of the disease, in particular aggressiveness and metastatic potential of a tumor or haematopoietic cell proliferative disorder.

As used herein the term 'predictive marker' shall be taken to mean an indicator of response to therapy, said response is preferably defined according to patient survival. It is preferably used to define patients with high, low and intermediate length of survival or recurrence after treatment, that is the result of the inherent heterogeneity of the disease process.

As defined herein the term 'predictive marker' shall fall within the remit of a herein described 'prognostic marker', as wherein a prognostic marker differentiates between patients with different survival outcomes independent of treatment the markers of the present invention are also predictive of treatment response. Therefore, unless otherwise stated the two terms shall not be taken to be mutually exclusive.

As used herein the term 'expression' shall be taken to mean the transcription and translation of a gene, as well as the genetic or the epigenetic modifications of the genomic DNA associated with the marker gene and/or regulatory or promoter regions thereof. Genetic modifications include SNPs, point mutations, deletions, insertions, repeat length, rearrangements and other polymorphisms. The analysis of either the expression levels of protein, or mRNA or the analysis of the patient's individual genetic or epigenetic

modification of the marker gene are herein summarized as the analysis of expression of the gene.

The level of expression of a gene may be determined by the analysis of any factors associated with or indicative of the level of transcription and translation of a gene including but not limited to methylation analysis, loss of heterozygosity (hereinafter also referred to as LOH), RNA expression levels and protein expression levels.

Furthermore the activity of the transcribed gene may be affected by genetic variations such as but not limited to genetic modifications (including but not limited to SNPs, point mutations, deletions, insertions, repeat length, rearrangements and other polymorphisms).

The terms "endocrine therapy" or "endocrine treatment" are meant to comprise any therapy, treatment or treatments targeting the estrogen receptor pathway or estrogen synthesis pathway or estrogen conversion pathway, which is involved in estrogen metabolism, production or secretion. Said treatments include, but are not limited to estrogen receptor modulators, estrogen receptor down-regulators, aromatase inhibitors, ovarian ablation, LHRH analogues and other centrally acting drugs influencing estrogen production.

The term "monotherapy" shall be taken to mean the use of a single drug or other therapy.

In the context of the present invention the term "chemotherapy" is taken to mean the use of pharmaceutical or chemical substances to treat cancer. This definition excludes radiation therapy (treatment with high energy rays or particles), hormone therapy (treatment with hormones or hormone analogues) and surgical treatment.

In the context of the present invention the term "adjuvant treatment" is taken to mean a therapy of a cancer patient immediately following an initial non chemotherapeutical therapy, e.g. surgery. In general, the purpose of an adjuvant therapy is to decrease the risk of recurrence.

In the context of the present invention the term "determining a suitable treatment regimen for the subject" is taken to mean the determination of a treatment regimen (i.e. a single therapy or a combination of different therapies that are used for the prevention and/or treatment of the

cancer in the patient) for a patient that is started, modified and/or ended based or essentially based or at least partially based on the results of the analysis according to the present invention. One example is starting an adjuvant endocrine therapy after surgery, another would be to modify the dosage of a particular chemotherapy. The determination can, in addition to the results of the analysis according to the present invention, be based on personal characteristics of the subject to be treated. In most cases, the actual determination of the suitable treatment regimen for the subject will be performed by the attending physician or doctor.

In the context of this invention the terms "obtaining a biological sample" or "obtaining a sample from a subject", shall not be taken to include the active retrieval of a sample from an individual, e.g. the performance of a biopsy. Said terms shall be taken to mean the obtainment of a sample previously isolated from an individual. Said samples may be isolated by any means standard in the art, including but not limited to biopsy, surgical removal, body fluids isolated by means of aspiration. Furthermore said samples may be provided by third parties including but not limited to clinicians, couriers, commercial sample providers and sample collections.

In the context of the present invention, the term "CpG island" refers to a contiguous region of genomic DNA that satisfies the criteria of (1) having a frequency of CpG dinucleotides corresponding to an "Observed/Expected Ratio" >0.6 , and (2) having a "GC Content" >0.5 . CpG islands are typically, but not always, between about 0.2 to about 1 kb in length.

In the context of the present invention the term "regulatory region" of a gene is taken to mean nucleotide sequences which affect the expression of a gene. Said regulatory regions may be located within, proximal or distal to said gene. Said regulatory regions include but are not limited to constitutive promoters, tissue-specific promoters, developmental-specific promoters, inducible promoters and the like. Promoter regulatory elements may also include certain enhancer sequence elements that control transcriptional or translational efficiency of the gene.

In the context of the present invention, the term "methylation" refers to the presence or absence of 5-methylcytosine ("5-mCyt") at one or a plurality of CpG dinucleotides within a DNA sequence.

In the context of the present invention the term "methylation state" is taken to mean the degree of methylation present in a nucleic acid of interest, this may be expressed in absolute or relative terms i.e. as a percentage or other numerical value or by comparison to another tissue and therein described as hypermethylated, hypomethylated or as having significantly similar or identical methylation status.

Unless specifically stated the terms "hypermethylated" or "upmethylated" shall be taken to mean a methylation level above that of a specified cut-off point, wherein said cut-off may be a value representing the average or median methylation level for a given population, or is preferably an optimized cut-off level. The "cut-off" is also referred herein as a "threshold". In the context of the present invention the terms "methylated", "hypermethylated" or "upmethylated" shall be taken to include a methylation level above a cut-off between 4% and 0% methylation, most preferably zero (0) % (or equivalents thereof) methylation. The term hypomethylation shall be taken to mean any methylation detected at below said cut-off.

In the context of the present invention, the term "microarray" refers broadly to both "DNA microarrays," and 'DNA chip(s),' as recognized in the art, encompasses all art-recognized solid supports, and encompasses all methods for affixing nucleic acid molecules thereto or synthesis of nucleic acids thereon.

"Genetic parameters" are mutations and polymorphisms of genes and sequences further required for their regulation. To be designated as genetic modifications or mutations are, in particular, insertions, deletions, point mutations, inversions and polymorphisms and, particularly preferred, SNPs (single nucleotide polymorphisms).

"Epigenetic modifications" or "epigenetic parameters" are modifications of DNA bases of genomic DNA and sequences further required for their regulation, in particular, cytosine methylations thereof. Further epigenetic parameters include, for example, the acetylation of histones which, however, cannot be directly analyzed using the described method but which, in turn, correlate with the DNA methylation.

In the context of the present invention, the term "bisulfite reagent" refers to a reagent comprising bisulfite, disulfite, hydrogen sulfite or combinations thereof, useful as disclosed herein to distinguish between methylated and unmethylated CpG dinucleotide sequences.

In the context of the present invention, the term "Methylation assay" refers to any assay for determining the methylation state of one or more CpG dinucleotide sequences within a sequence of DNA.

In the context of the present invention, the term "MS.AP-PCR" (Methylation-Sensitive Arbitrarily-Primed Polymerase Chain Reaction) refers to the art-recognized technology that allows for a global scan of the genome using CG-rich primers to focus on the regions most likely to contain CpG dinucleotides, and described by Gonzalgo et al., *Cancer Research* 57:594-599, 1997.

In the context of the present invention, the term "MethyLight" refers to the art-recognized fluorescence-based real-time PCR technique described by Eads et al., *Cancer Res.* 59:2302-2306, 1999.

In the context of the present invention, the term "HeavyMethyl™" assay, in the embodiment thereof implemented herein, refers to a methylation assay comprising methylation specific *blocking* probes covering CpG positions between the amplification primers.

The term "Ms-SNuPE" (Methylation-sensitive Single Nucleotide Primer Extension) refers to the art-recognized assay described by Gonzalgo and Jones, *Nucleic Acids Res.* 25:2529-2531, 1997.

In the context of the present invention the term "MSP" (Methylation-specific PCR) refers to the art-recognized methylation assay described by Herman et al. *Proc. Natl. Acad. Sci. USA* 93:9821-9826, 1996, and by US Patent No. 5,786,146.

In the context of the present invention the term "COBRA" (Combined Bisulfite Restriction Analysis) refers to the art-recognized methylation assay described by Xiong and Laird, *Nucleic Acids Res.* 25:2532-2534, 1997.

In the context of the present invention the term "hybridization" is to be understood as a bond of an oligonucleotide to a complementary sequence along the lines of the Watson-Crick base pairings in the sample DNA, forming a duplex structure.

"Stringent hybridization conditions," as defined herein, involve hybridizing at 68°C in 5x SSC/5x Denhardt's solution/1.0% SDS, and washing in 0.2x SSC/0.1% SDS at room temperature, or involve the art-recognized equivalent thereof (e.g., conditions in which a hybridization is carried out at 60°C in 2.5 x SSC buffer, followed by several washing steps at 37°C in a low buffer concentration, and remains stable). Moderately stringent conditions, as defined herein, involve including washing in 3x SSC at 42°C, or the art-recognized equivalent thereof. The parameters of salt concentration and temperature can be varied to achieve the optimal level of identity between the probe and the target nucleic acid. Guidance regarding such conditions is available in the art, for example, by Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, *Current Protocols in Molecular Biology*, (John Wiley and Sons, N.Y.) at Unit 2.10.

"Background DNA" as used herein refers to any nucleic acids which originate from sources other than the cancer cells to be analysed.

Using the methods and nucleic acids described herein, statistically significant models of patient relapse, disease free survival, metastasis free survival, overall survival and/or disease progression can be developed and utilized to assist patients and clinicians in determining suitable treatment options to be included in a therapeutic regimen.

In a further aspect of the invention said markers are used as prognostic markers and/or as predictive markers for the outcome of anthracycline and/or endocrine therapy, thereby enabling the physician to determine if said treatments are of benefit to a patient.

Using the methods and nucleic acids as described herein, patient survival can be evaluated before or during treatment for a cell proliferative disorder suitable for treatment with anthracyclines and/or endocrine therapies, in order to provide critical information to the patient and clinician as to the likely progression of the disease, including treatment response. It will be appreciated, therefore, that the methods and nucleic acids exemplified herein can

serve to improve a patient's quality of life and odds of treatment success by allowing both patient and clinician a more accurate assessment of the patient's treatment options.

The present invention makes available a method for the improved treatment of cell proliferative disorders, by enabling the improved prediction of a patient's survival, in particular by predicting the likelihood of relapse post-surgery both with or without anthracycline and/or endocrine therapy.

The method according to the invention may be used for the analysis of a wide variety of cell proliferative disorders, most preferably those suitable for treatment with anthracycline or endocrine therapy including, but not limited to, breast cancer, ovarian cancer, transitional cell bladder cancer, bronchogenic lung cancer, thyroid cancer, pancreatic cancer, prostate cancer, uterine cancer, testicular cancer, gastric cancer, soft tissue and osteogenic sarcomas, neuroblastoma, Wilms' tumor, malignant lymphoma (Hodgkin's and non-Hodgkin's), acute myeloblastic leukemia, acute lymphoblastic leukemia, Kaposi's sarcoma, Ewing's tumor, refractory multiple myeloma, squamous cell carcinomas of the head, neck, cervix, and vagina.

However it is most preferred that the methods of the present invention are applied in the analysis of breast cancer.

It is particularly preferred that said prediction is defined in terms of patient survival and/or relapse. In this embodiment patients survival times and/or relapse are predicted according to their gene expression or genetic or epigenetic modifications thereof. In this aspect of the invention it is particularly preferred that said patients are tested prior to receiving any adjuvant treatment.

It is herein described that aberrant expression of at least one gene selected from the group consisting LHX3 and PITX3, is correlated to outcome of treatment of cell proliferative disorder patients wherein said treatment comprises at least one of an anthracycline and endocrine therapy.

This marker thereby provides a novel means for the characterization of cell proliferative disorders. As described herein determination of the expression of at least one gene selected from the group LHX3 and PITX3 and/or regulatory or promoter regions thereof, enables the

prediction of prognosis and/or treatment response of a patient treated with a therapy comprising at least one of an anthracycline and endocrine therapy.

In a further aspect the invention relates to new methods and sequences, which may be used as tools for the selection of suitable treatments of patients diagnosed with cell proliferative disease based on a prediction of likelihood of relapse, survival or outcome.

One aspect of the invention is the provision of methods for providing a prediction of disease prognosis or a prediction of outcome of a treatment comprising at least one of an anthracycline and endocrine therapy of a patient with a cell proliferative disorder. Preferably said prognosis and/or prediction is provided in terms of likelihood of relapse or the survival of said patient. It is further preferred that said survival is disease free survival or metastasis free survival. It is also preferred that said disease is breast cancer. These methods comprise the analysis of the expression levels of at least one gene selected from the group consisting LHX3 and PITX3 and/or regulatory or promoter regions thereof.

It is preferred that said patients are analyzed prior to receiving any treatment.

The sequence of the genes PITX3 and LHX3 are disclosed in the sequence listing, it is preferred that any transcript thereof or polypeptide transcribed therefrom is analysed and that the prognosis or predictive outcome of anthracycline and/or endocrine treatment in a subject is determined therefrom.

Said method may be enabled by means of any analysis of the expression of a RNA transcribed therefrom or polypeptide or protein translated from said RNA, preferably by means of mRNA expression analysis or polypeptide expression analysis. Accordingly the present invention also provides prognostic assays and methods, both quantitative and qualitative for detecting the expression of at least one gene selected from the group consisting LHX3 and PITX3 in a subject with a cell proliferative disorder and determining therefrom upon the prognosis of pr prediction of outcome of treatment comprising at least one of an anthracycline and endocrine therapy of said subject.

Aberrant expression of mRNA transcribed from the genes LHX3 and PITX3 are associated with prognosis and/or prediction of treatment outcome of cancer.

Overexpression is associated with good prognosis and/or positive prediction of treatment outcome, under expression is associated with bad prognosis and/or negative prediction of treatment outcome.

To detect the presence of mRNA encoding a gene or genomic sequence, a sample is obtained from a patient. The sample may be any suitable sample comprising cellular matter of the tumour most preferably the primary tumour. Suitable sample types include cell lines, histological slides, paraffin embedded tissues, biopsies, tissue embedded in paraffin, bodily fluids (such as but not limited to nipple aspirate and blood) and all possible combinations thereof.

In a particularly preferred embodiment of the method said source is blood. The sample may be treated to extract the RNA contained therein. The resulting nucleic acid from the sample is then analysed. Many techniques are known in the state of the art for determining absolute and relative levels of gene expression, commonly used techniques suitable for use in the present invention include Northern analysis, RNase protection assays (RPA), microarrays and PCR-based techniques, such as quantitative PCR and differential display PCR .

Particularly preferred is the use of the reverse transcription/polymerisation chain reaction technique. The method of reverse transcription /PCR is well known in the art (for example, see Watson and Fleming, *supra*).

The reverse transcription /PCR method can be performed as follows. Total cellular RNA is isolated by, for example, the standard guanidium isothiocyanate method and the total RNA is reverse transcribed. The reverse transcription method involves synthesis of DNA on a template of RNA using a reverse transcriptase enzyme and a 3' end primer. Typically, the primer contains an oligo(dT) sequence. The cDNA thus produced is then amplified by means of PCR. (Belyavsky et al, Nucl Acid Res 17:2919-2932, 1989; Krug and Berger, Methods in Enzymology, Academic Press,N.Y., Vol.152, pp. 316-325, 1987 which are incorporated by reference). Further preferred is the "Real-time" variant of reverse transcription PCR, wherein the PCR product is detect by means of hybridisation probes (E.g TaqMan, Lightcycler, Molecular Beacons and Scorpion) or SYBR green. The detected signal from the probes or SYBR green is then quantitated either by reference to a standard curve or by comparing the Ct

values to that of a calibration standard. Analysis of housekeeping genes is often used to normalize the results

In Northern blot analysis total or poly(A)+ mRNA is run on a denaturing agarose gel and detected by hybridization to a labeled probe in the dried gel itself or on a membrane. The resulting signal is proportional to the amount of target RNA in the RNA population.

Comparing the signals from two or more cell populations or tissues reveals relative differences in gene expression levels. Absolute quantitation can be performed by comparing the signal to a standard curve generated using known amounts of an in vitro transcript corresponding to the target RNA. Analysis of housekeeping genes, genes whose expression levels are expected to remain relatively constant regardless of conditions, is often used to normalize the results, eliminating any apparent differences caused by unequal transfer of RNA to the membrane or unequal loading of RNA on the gel.

The first step in Northern analysis is isolating pure, intact RNA from the cells or tissue of interest. Because Northern blots distinguish RNAs by size, sample integrity influences the degree to which a signal is localized in a single band. Partially degraded RNA samples will result in the signal being smeared or distributed over several bands with an overall loss in sensitivity and possibly an erroneous interpretation of the data. In Northern blot analysis, DNA, RNA and oligonucleotide probes can be used and these probes are preferably labelled (e.g. radioactive labels, massa labels or fluorescent labels). The size of the target RNA, not the probe, will determine the size of the detected band, so methods such as random-primed labeling, which generates probes of variable lengths, are suitable for probe synthesis. The specific activity of the probe will determine the level of sensitivity, so it is preferred that probes with high specific activities ($>10^9$ cpm/ μ g), are used. Labeling methods that produce probes with lower specific activities can be used to detect more abundant RNAs.

In an RNase protection assay, the RNA target and an RNA probe of a defined length are hybridized in solution. Following hybridization, the RNA is digested with RNases specific for single-stranded nucleic acids to remove any unhybridized, single-stranded target RNA and probe. The RNases are inactivated, and the RNA is separated e.g. by denaturing polyacrylamide gel electrophoresis. The amount of intact RNA probe is proportional to the amount of target RNA in the RNA population. RPA can be used for relative and absolute

quantitation of gene expression and also for mapping RNA structure, such as intron/exon boundaries and transcription start sites. The RNase protection assay is preferable to Northern blot analysis as it generally has a lower limit of detection.

The antisense RNA probes used in RPA are generated by *in vitro* transcription of a DNA template with a defined endpoint and are typically in the range of 50–600 nucleotides. The use of RNA probes that include additional sequences not homologous to the target RNA allows the protected fragment to be distinguished from the full-length probe. RNA probes are typically used instead of DNA probes due to the ease of generating single-stranded RNA probes and the reproducibility and reliability of RNA:RNA duplex digestion with RNases (Ausubel *et al.* 2003), particularly preferred are probes with high specific activities.

Particularly preferred is the use of microarrays. The microarray analysis process can be divided into two main parts. First is the immobilization of known gene sequences onto glass slides or other solid support followed by hybridization of the fluorescently labeled cDNA (comprising the sequences to be interrogated) to the known genes immobilized on the glass slide. After hybridization, arrays are scanned using a fluorescent microarray scanner. Analyzing the relative fluorescent intensity of different genes provides a measure of the differences in gene expression.

One method of creating DNA arrays is by immobilizing PCR products onto activated glass surfaces. Typically, probes are first generated by PCR or RT-PCR and cloned into a plasmid vector to create a library of 10,000 or more clones. This plasmid library may be stored in *E. coli*. To generate a new array, the *E. coli* are grown, plasmids are isolated and the cloned genes are amplified with primers common to the plasmid backbone. These amplified products are typically in the range of 100–1,000 bases. Automated means are then used to print the amplified clones on an array of 50–200µm spots on a specially prepared glass slide or other suitable support.

DNA arrays can also be generated by immobilizing presynthesized oligonucleotides onto prepared glass slides. In this case, representative gene sequences are manufactured and prepared using standard oligonucleotide synthesis and purification methods. These synthesized gene sequences are complementary to the genes of interest (in this case LHX3 and PITX3) and tend to be shorter sequences in the range of 25–70 nucleotides. Alternatively,

immobilized oligos can be chemically synthesized in situ on the surface of the slide. In situ oligonucleotide synthesis involves the consecutive addition of the appropriate nucleotides to the spots on the microarray; spots not receiving a nucleotide are protected during each stage of the process using physical or virtual masks.

In expression profiling microarray experiments, the RNA templates used are representative of the transcription profile of the cells or tissues under study. RNA is first isolated from the cell populations or tissues to be compared. Each RNA sample is then used as a template to generate fluorescently labeled cDNA via a reverse transcription reaction. Fluorescent labeling of the cDNA can be accomplished by either direct labeling or indirect labeling methods. During direct labeling, fluorescently modified nucleotides (e.g., Cy[®]3- or Cy[®]5-dCTP) are incorporated directly into the cDNA during the reverse transcription. Alternatively, indirect labeling can be achieved by incorporating aminoallyl-modified nucleotides during cDNA synthesis and then conjugating an N-hydroxysuccinimide (NHS)-ester dye to the aminoallyl-modified cDNA after the reverse transcription reaction is complete. Alternatively, the probe may be unlabeled, but may be detectable by specific binding with a ligand which is labelled, either directly or indirectly. Suitable labels and methods for labelling ligands (and probes) are known in the art, and include, for example, radioactive labels which may be incorporated by known methods (*e.g.*, nick translation or kinasing). Other suitable labels include but are not limited to biotin, fluorescent groups, chemiluminescent groups (*e.g.*, dioxetanes, particularly triggered dioxetanes), enzymes, antibodies, and the like.

To perform differential gene expression analysis, cDNA generated from different RNA samples are labeled in different colors. For example, when comparing RNA from test and reference tissue samples, the cDNA generated from the test RNA can be labeled with Cy[®]3, while the cDNA generated from the reference RNA sample can be labeled with Cy[®]5. The resulting labeled cDNA is purified to remove unincorporated nucleotides, free dye and residual RNA. Following purification, the labeled cDNA samples are combined and then hybridized to the microarray. The stringency of hybridisation is determined by a number of factors during hybridisation and during the washing procedure, including temperature, ionic strength, length of time and concentration of formamide. These factors are outlined in, for example, Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd ed., 1989). The microarray is scanned post-hybridization using a fluorescent microarray scanner. The fluorescent intensity of each spot indicates the level of expression for that gene; bright spots

correspond to strongly expressed genes, while dim spots indicate weak expression. Multiple targets labeled in different dye colors can be analyzed simultaneously to determine which genes are differentially expressed.

Once the images are obtained, the raw data must be analyzed. First, the background fluorescence must be subtracted from the fluorescence of each spot. The data is then normalized to a control sequence, such as an exogenously added RNA, or a housekeeping gene to account for any nonspecific hybridization, array imperfections or variability in the array setup, cDNA labeling, hybridization or washing. Data normalization allows the results of multiple arrays to be compared.

The present invention further provides methods for the detection of the presence of the polypeptide encoded by said genes in a sample obtained from a patient.

Aberrant levels of polypeptide expression of the polypeptides encoded by LHX3 and PITX3 are associated with cell proliferative disorder prognosis and/or treatment outcome.

Accordingly over or under expression of said polypeptides are associable with anthracycline and endocrine treatment outcome. Over expression is associated with positive prognosis and/or treatment outcome and under expression is associated with negative prognosis and/or treatment outcome.

Any method known in the art for detecting polypeptides can be used. Such methods include, but are not limited to immunodiffusion, immunoelectrophoresis, immunochemical methods, binder-ligand assays, immunohistochemical techniques, agglutination and complement assays. (*e.g.*, see Basic and Clinical Immunology, Sites and Terr, eds., Appleton and Lange, Norwalk, Conn. pp 217-262, 1991 which is incorporated by reference). Preferred are binder-ligand immunoassay methods including reacting antibodies with an epitope or epitopes and competitively displacing a labelled polypeptide or derivative thereof.

Certain embodiments of the present invention comprise the use of antibodies specific to the polypeptides encoded by the LHX3 and PITX3 genes. Such antibodies are useful for cancer prognostic and/or predictive applications. In certain embodiments production of monoclonal or polyclonal antibodies can be induced by the use of the coded polypeptide as an antigene.

Such antibodies may in turn be used to detect expressed polypeptides as markers for cell proliferative disorder prognosis. The levels of such polypeptides present may be quantified by conventional methods. Antibody-polypeptide binding may be detected and quantified by a variety of means known in the art, such as labelling with fluorescent or radioactive ligands. The invention further comprises kits for performing the above-mentioned procedures, wherein such kits contain antibodies specific for the investigated polypeptides.

Numerous competitive and non-competitive polypeptide binding immunoassays are well known in the art. Antibodies employed in such assays may be unlabeled, for example as used in agglutination tests, or labelled for use a wide variety of assay methods. Labels that can be used include radionuclides, enzymes, fluorescers, chemiluminescers, enzyme substrates or co-factors, enzyme inhibitors, particles, dyes and the like. Preferred assays include but are not limited to radioimmunoassay (RIA), enzyme immunoassays, e.g., enzyme-linked immunosorbent assay (ELISA), fluorescent immunoassays and the like. Polyclonal or monoclonal antibodies or epitopes thereof can be made for use in immunoassays by any of a number of methods known in the art.

One approach for preparing antibodies to a polypeptide is the selection and preparation of an amino acid sequence of all or part of the polypeptide, chemically synthesising the amino acid sequence and injecting it into an appropriate animal, usually a rabbit or a mouse (Milstein and Kohler *Nature* 256:495-497, 1975; Gutfre and Milstein, *Methods in Enzymology: Immunochemical Techniques* 73:1-46, Langone and Banatis eds., Academic Press, 1981 which are incorporated by reference). Methods for preparation of the polypeptides or epitopes thereof include, but are not limited to chemical synthesis, recombinant DNA techniques or isolation from biological samples.

In the final step of the method the prognosis of the patient is determined, whereby underexpression is indicative of negative prognosis and/or treatment outcome. The term underexpression shall be taken to mean expression at a detected level less than a pre-determined cut off which may be selected from the group consisting of the mean, median or an optimised threshold value.

Another aspect of the invention provides a kit for use in providing a prognosis or prediction of treatment response of a subject with a cell proliferative disorder, comprising: a means for

detecting LHX3 and PITX3 polypeptides. The means for detecting the polypeptides comprise preferably antibodies, antibody derivatives, or antibody fragments. The polypeptides are most preferred detected by means of Western blotting utilizing a labelled antibody. In another embodiment of the invention the kit further comprising means for obtaining a biological sample of the patient. Preferred is a kit which further comprises a container suitable for containing the means for detecting the polypeptides and the biological sample of the patient, and most preferably further comprises instructions for use and interpretation of the kit results.

In a preferred embodiment the kit for use in determining treatment strategy for a patient with a cell proliferative disorder, comprises: (a) a means for detecting polypeptides of at least one gene selected from the group consisting LHX3 and PITX3 ; (b) a container suitable for containing the said means and the biological sample of the patient comprising the polypeptides wherein the means can form complexes with the polypeptides; (c) a means to detect the complexes of (b); and optionally (d) instructions for use and interpretation of the kit results. The kit may also contain other components such as buffers or solutions suitable for blocking, washing or coating, packaged in a separate container.

In one embodiment of the method aberrant expression of at least one gene selected from the group consisting LHX3 and PITX3 may be detected by analysis of loss of heterozygosity of the gene.

In a first step genomic DNA is isolated from a biological sample of the patient's tumor. The isolated DNA is then analyzed for LOH by any means standard in the art including but not limited to amplification of the gene locus or associated microsatellite markers. Said amplification may be carried out by any means standard in the art including polymerase chain reaction (PCR), strand displacement amplification (SDA) and isothermal amplification.

The level of amplificate is then detected by any means known in the art including but not limited to gel electrophoresis and detection by probes (including Real Time PCR). Furthermore the amplicates may be labeled in order to aid said detection. Suitable detectable labels include but are not limited to fluorescence label, radioactive labels and mass labels the suitable use of which shall be described herein.

The detection of a decreased amount of an amplificate corresponding to one of the amplified alleles in a test sample as relative to that of a heterozygous control sample is indicative of LOH.

Another aspect of the invention relates to a kit for use in providing a prognosis of a subject with a cell proliferative disorder, said kit comprising: a means for measuring the level of transcription of at least one gene selected from the group consisting LHX3 and PITX3.

In a preferred embodiment the means for measuring the level of transcription comprise oligonucleotides or polynucleotides able to hybridise under stringent or moderately stringent conditions to the transcription products of at least one gene selected from the group consisting LHX3 and PITX3. In a most preferred embodiment the level of transcription is determined by techniques selected from the group of Northern blot analysis, reverse transcriptase PCR, real-time PCR, RNase protection, and microarray. In another embodiment of the invention the kit further comprises means for obtaining a biological sample of the patient. Preferred is a kit which further comprises a container suitable for containing the means for measuring the level of transcription and the biological sample of the patient, and most preferably further comprises instructions for use and interpretation of the kit results.

In a preferred embodiment the kit for use in determining treatment strategy for a patient with a cell proliferative disorder comprises (a) a plurality of oligonucleotides or polynucleotides able to hybridise under stringent or moderately stringent conditions to the transcription products of at least one gene selected from the group consisting LHX3 and PITX3; (b) a container suitable for containing the oligonucleotides or polynucleotides and a biological sample of the patient comprising the transcription products wherein the oligonucleotides or polynucleotide can hybridise under stringent or moderately stringent conditions to the transcription products, (c) means to detect the hybridisation of (b); and optionally, (d) instructions for use and interpretation of the kit results. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

Most preferably a kit according to the embodiments of the present invention is used for the determination of expression step of the methods according to other aspects of the invention. It is particularly preferred that said kits are utilised in providing a prognosis or a prediction of treatment response of a subject with breast cancer.

In one aspect the invention provides significant improvements over the state of the art in that it provides the first cancer treatment response markers for a treatment comprising an anthracycline, that is also relevant for endocrine therapy.

In the most preferred embodiment of the invention the analysis of expression is carried out by means of methylation analysis of at least one of the genes selected from the group consisting LHX3 and PITX3. It is further preferred that the methylation state of the CpG dinucleotides within the genomic sequence of said genes according to SEQ ID NO: 1 and SEQ ID NO: 2 and sequences complementary thereto are analyzed.

It is further preferred that the methylation state of the CpG dinucleotides within the genomic sequence of said genes according to SEQ ID NO: 3 and SEQ ID NO: 4 and sequences complementary thereto are analyzed. Said sequences provide CpG rich sequences of the genes LHX3 and PITX3 respectively according to Table 1.

The methylation pattern of the genes according to the present invention and their promoter and regulatory elements have heretofore not been analyzed with regard to cancer prognosis or prediction of outcome of anthracycline or endocrine treatment. Due to the degeneracy of the genetic code, the genomic sequences as provided in the sequence listing should be interpreted so as to include all substantially similar and equivalent sequences of a gene which encodes a polypeptide with the biological activity of any of those encoded by the genes of Table 1.

Most preferably the following method is used to detect methylation within the genes LHX3 and PITX3 and/or regulatory or promoter regions thereof.

The method for the analysis of methylation comprises contacting a nucleic acid sample obtained from a subject with at least one reagent or a series of reagents, wherein said reagent or series of reagents, distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid.

Preferably, said method comprises the following steps: In the first *step*, a sample of the tissue to be analyzed is obtained. The source may be any suitable source, preferably, the source of the sample is selected from the group consisting of histological slides, biopsies, paraffin-embedded tissue, bodily fluids, plasma, serum, urine, blood, nipple aspirate and combinations thereof. Preferably, the source is tumor tissue, biopsies, serum, urine, blood or nipple aspirate. The most preferred source is a tumor sample, which may be obtained during surgical removal from the patient or a biopsy sample of said patient.

The DNA is then isolated from the sample. Genomic DNA may be isolated by any means standard in the art, including the use of commercially available kits. Briefly, wherein the DNA of interest is encapsulated in/by a cellular membrane the biological sample must be disrupted and lysed by enzymatic, chemical or mechanical means. The DNA solution may then be cleared of proteins and other contaminants e.g. by digestion with proteinase K. The genomic DNA is then recovered from the solution. This may be carried out by means of a variety of methods including salting out, organic extraction or binding of the DNA to a solid phase support. The choice of method will be affected by several factors including time, expense and required quantity of DNA.

The genomic DNA sample is then treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridization behavior. This will be understood as "treatment" or "pre-treatment" herein.

The above described pre-treatment of genomic DNA is preferably carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis which results in a conversion of non-methylated cytosine nucleobases to uracil or to another base which is dissimilar to cytosine in terms of base pairing behavior. Enclosing the DNA to be analyzed in an agarose matrix, thereby preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and replacing all precipitation and purification steps with fast dialysis (Olek A, et al., A modified and improved method for bisulfite based cytosine methylation analysis, *Nucleic Acids Res.* 24:5064-6, 1996) is one preferred example how to perform said pre-treatment. It is further preferred that the bisulfite treatment is carried out in the presence of a radical scavenger or DNA denaturing agent.

The treated DNA is then analyzed in order to determine the methylation state of at least one of the genes selected from the group consisting LHX3 and PITX3 and/or regulatory regions thereof. This methylation state is then associated with the prognosis of the patient and/or with outcome of a treatment comprising at least one anthracycline and/or endocrine therapy. It is further preferred that the sequences of said genes as described in the accompanying sequence listing (see Table 1) are analyzed.

In the third step of the method, fragments of the pretreated DNA are amplified. Wherein the source of the DNA is free DNA from serum, or DNA extracted from paraffin it is particularly preferred that the size of the amplificate fragment is between 100 and 200 base pairs in length, and wherein said DNA source is extracted from cellular sources (e.g. tissues, biopsies, cell lines) it is preferred that the amplificate is between 100 and 350 base pairs in length. It is particularly preferred that said amplificates comprise at least one 20 base pair sequence comprising at least three CpG dinucleotides. Said amplification is carried out using sets of primer oligonucleotides according to the present invention, and a preferably heat-stable polymerase. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel, in one embodiment of the method preferably six or more fragments are amplified simultaneously. Typically, the amplification is carried out using a polymerase chain reaction (PCR). The set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 18-base-pair long segment of a base sequence selected from the group consisting SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and sequences complementary thereto. In a more preferred embodiment the set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary, identical, or hybridize under stringent or highly stringent conditions to an at least 18-base-pair long segment of a pretreated sequence selected from SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20 and sequences complementary thereto.

In an alternate embodiment of the method, the methylation status of pre-selected CpG positions within the nucleic acid sequences comprising SEQ ID NO: 1 and SEQ ID NO: 2 may be detected by use of methylation-specific primer oligonucleotides. This technique (MSP) has been described in United States Patent No. 6,265,171 to Herman. The use of

methylation status specific primers for the amplification of bisulfite treated DNA allows the differentiation between methylated and unmethylated nucleic acids. MSP primers pairs contain at least one primer which hybridizes to a bisulfite treated CpG dinucleotide. Therefore, the sequence of said primers comprises at least one CpG, TpG or CpA dinucleotide. MSP primers specific for non-methylated DNA contain a "T" at the 3' position of the C position in the CpG. Preferably, therefore, the base sequence of said primers is required to comprise a sequence having a length of at least 18 nucleotides which hybridizes to a pretreated nucleic acid sequence according to SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, tpG or Cpa dinucleotide. It is further preferred that said sequence has a length of at least 18 nucleotides which hybridizes to a pretreated nucleic acid sequence selected from the group consisting SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20.

In this embodiment of the method according to the invention it is particularly preferred that the MSP primers comprise between 2 and 4 CpG, TpG or Cpa dinucleotides. It is further preferred that said dinucleotides are located within the 3' half of the primer e.g. wherein a primer is 18 bases in length the specified dinucleotides are located within the first 9 bases from the 3' end of the molecule. In addition to the CpG, TpG or Cpa dinucleotides it is further preferred that said primers should further comprise several bisulfite converted bases (i.e. cytosine converted to thymine, or on the hybridizing strand, guanine converted to adenosine). In a further preferred embodiment said primers are designed so as to comprise no more than 2 cytosine or guanine bases.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. Preferred are labels in the form of fluorescence labels, radionuclides, or detachable molecule fragments having a typical mass which can be detected in a mass spectrometer. Where said labels are mass labels, it is preferred that the labeled amplicates have a single positive or negative net charge, allowing for better detectability in the mass spectrometer. The detection may be carried out and visualized by means of, e.g., matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

Matrix Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-TOF) is a very

efficient development for the analysis of biomolecules (Karas and Hillenkamp, *Anal Chem.*, 60:2299-301, 1988). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapor phase in an unfragmented manner. The analyte is ionized by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones. MALDI-TOF spectrometry is well suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut and Beck, *Current Innovations and Future Trends*, 1:147-57, 1995). The sensitivity with respect to nucleic acid analysis is approximately 100-times less than for peptides, and decreases disproportionally with increasing fragment size. Moreover, for nucleic acids having a multiply negatively charged backbone, the ionisation process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For the desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallisation. There are now several responsive matrixes for DNA, however, the difference in sensitivity between peptides and nucleic acids has not been reduced. This difference in sensitivity can be reduced, however, by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. For example, phosphorothioate nucleic acids, in which the usual phosphates of the backbone are substituted with thiophosphates, can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut and Beck, *Nucleic Acids Res.* 23: 1367-73, 1995). The coupling of a charge tag to this modified DNA results in an increase in MALDI-TOF sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities, which makes the detection of unmodified substrates considerably more difficult.

In a particularly preferred embodiment of the method the amplification of step three is carried out in the presence of at least one species of blocker oligonucleotides. The use of such blocker oligonucleotides has been described by Yu et al., *BioTechniques* 23:714-720, 1997. The use of blocking oligonucleotides enables the improved specificity of the amplification of a subpopulation of nucleic acids. Blocking probes hybridized to a nucleic acid suppress, or hinder the polymerase mediated amplification of said nucleic acid. In one embodiment of the method blocking oligonucleotides are designed so as to hybridize to background DNA. In a further embodiment of the method said oligonucleotides are designed so as to hinder or suppress the amplification of unmethylated nucleic acids as opposed to methylated nucleic

acids or vice versa.

Blocking probe oligonucleotides are hybridized to the bisulfite treated nucleic acid concurrently with the PCR primers. PCR amplification of the nucleic acid is terminated at the 5' position of the blocking probe, such that amplification of a nucleic acid is suppressed where the complementary sequence to the blocking probe is present. The probes may be designed to hybridize to the bisulfite treated nucleic acid in a methylation status specific manner. For example, for detection of methylated nucleic acids within a population of unmethylated nucleic acids, suppression of the amplification of nucleic acids which are unmethylated at the position in question would be carried out by the use of blocking probes comprising a 'TpG' at the position in question, as opposed to a 'CpG.' In one embodiment of the method the sequence of said blocking oligonucleotides should be identical or complementary to a sequence at least 18 base pairs in length selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 preferably comprising one or more CpG, TpG or CpA dinucleotides. In a preferred embodiment of the method the sequence of said blocking oligonucleotides should be identical or complementary to a sequence at least 18 base pairs in length selected from a pretreated nucleic acid sequence selected from the group consisting SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20; and sequences complementary thereto preferably comprising one or more CpG, TpG or CpA dinucleotides.

For PCR methods using blocker oligonucleotides, efficient disruption of polymerase-mediated amplification requires that blocker oligonucleotides not be elongated by the polymerase. Preferably, this is achieved through the use of blockers that are 3'-deoxyoligonucleotides, or oligonucleotides derivatised at the 3' position with other than a "free" hydroxyl group. For example, 3'-O-acetyl oligonucleotides are representative of a preferred class of blocker molecule.

Additionally, polymerase-mediated decomposition of the blocker oligonucleotides should be precluded. Preferably, such preclusion comprises either use of a polymerase lacking 5'-3' exonuclease activity, or use of modified blocker oligonucleotides having, for example, thioate bridges at the 5'-termini thereof that render the blocker molecule nuclease-resistant. Particular applications may not require such 5' modifications of the blocker. For example, if the blocker- and primer-binding sites overlap, thereby precluding binding of the primer (*e.g.*, with

excess blocker), degradation of the blocker oligonucleotide will be substantially precluded. This is because the polymerase will not extend the primer toward, and through (in the 5'-3' direction) the blocker - a process that normally results in degradation of the hybridized blocker oligonucleotide.

A particularly preferred blocker/PCR embodiment, for purposes of the present invention and as implemented herein, comprises the use of peptide nucleic acid (PNA) oligomers as blocking oligonucleotides. Such PNA blocker oligomers are ideally suited, because they are neither decomposed nor extended by the polymerase.

In one embodiment of the method, the binding site of the blocking oligonucleotide is identical to, or overlaps with that of the primer and thereby hinders the hybridization of the primer to its binding site. In a further preferred embodiment of the method, two or more such blocking oligonucleotides are used. In a particularly preferred embodiment, the hybridization of one of the blocking oligonucleotides hinders the hybridization of a forward primer, and the hybridization of another of the probe (blocker) oligonucleotides hinders the hybridization of a reverse primer that binds to the amplificate product of said forward primer.

In an alternative embodiment of the method, the blocking oligonucleotide hybridizes to a location between the reverse and forward primer positions of the treated background DNA, thereby hindering the elongation of the primer oligonucleotides.

It is particularly preferred that the blocking oligonucleotides are present in at least 5 times the concentration of the primers.

In the fourth step of the method, the amplificates obtained during the third step of the method are analyzed in order to ascertain the methylation status of the CpG dinucleotides prior to the treatment.

In embodiments where the amplificates are obtained by means of MSP amplification and/or blocking oligonucleotides, the presence or absence of an amplificate is in itself indicative of the methylation state of the CpG positions covered by the primers and or blocking oligonucleotide, according to the base sequences thereof. All possible known molecular biological methods may be used for this detection, including, but not limited to gel

electrophoresis, sequencing, liquid chromatography, hybridizations, real time PCR analysis or combinations thereof. This step of the method further acts as a qualitative control of the preceding steps.

In the fourth step of the method amplicates obtained by means of both standard and methylation specific PCR are further analyzed in order to determine the CpG methylation status of the genomic DNA isolated in the first step of the method. This may be carried out by means of hybridization-based methods such as, but not limited to, array technology and probe based technologies as well as by means of techniques such as sequencing and template directed extension.

In one embodiment of the method, the amplicates synthesized in step three are subsequently hybridized to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridization takes place in the following manner: the set of probes used during the hybridization is preferably composed of at least 2 oligonucleotides or PNA-oligomers; in the process, the amplicates serve as probes which hybridize to oligonucleotides previously bonded to a solid phase; the non-hybridized fragments are subsequently removed; said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and the segment comprises at least one CpG, TpG or CpA dinucleotide. More preferably said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the pretreated sequences selected from the group consisting SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20 and the segment comprises at least one CpG, TpG or CpA dinucleotide.

In further embodiments said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 ; and the segment comprises at least one CpG, TpG or CpA dinucleotide.

More preferably said oligonucleotides contain at least one base sequence having a length of at least 9 nucleotides which is reverse complementary or identical to a segment of the pretreated sequences selected from the group consisting SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID

NO: 17 to SEQ ID NO: 20 and the segment comprises at least one CpG, TpG or CpA dinucleotide.

In a preferred embodiment, said dinucleotide is present in the central third of the oligomer. For example, wherein the oligomer comprises one CpG dinucleotide, said dinucleotide is preferably the fifth to ninth nucleotide from the 5'-end of a 13-mer. In a further embodiment one oligonucleotide exists for the analysis of each CpG dinucleotide within the sequence according to of SEQ ID NO: 1 and SEQ ID NO: 2, and the equivalent positions within of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16. One oligonucleotide exists for the analysis of each CpG dinucleotide within the sequence according to SEQ ID NO: 1 and SEQ ID NO: 2, and the equivalent positions within SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16. Said oligonucleotides may also be present in the form of peptide nucleic acids. The non-hybridized amplicates are then removed. The hybridized amplicates are detected. In this context, it is preferred that labels attached to the amplicates are identifiable at each position of the solid phase at which an oligonucleotide sequence is located.

In yet a further embodiment of the method, the genomic methylation status of the CpG positions may be ascertained by means of oligonucleotide probes that are hybridized to the bisulfite treated DNA concurrently with the PCR amplification primers (wherein said primers may either be methylation specific or standard).

A particularly preferred embodiment of this method is the use of fluorescence-based Real Time Quantitative PCR (Heid et al., *Genome Res.* 6:986-994, 1996; *also see* United States Patent No. 6,331,393). There are two preferred embodiments of utilizing this method. One embodiment, known as the TaqMan™ assay employs a dual-labeled fluorescent oligonucleotide probe. The TaqMan™ PCR reaction employs the use of a non-extendible interrogating oligonucleotide, called a TaqMan™ probe, which is designed to hybridize to a CpG-rich sequence located between the forward and reverse amplification primers. The TaqMan™ probe further comprises a fluorescent "reporter moiety" and a "quencher moiety" covalently bound to linker moieties (*e.g.*, phosphoramidites) attached to the nucleotides of the TaqMan™ oligonucleotide. Hybridized probes are displaced and broken down by the polymerase of the amplification reaction thereby leading to an increase in fluorescence. For analysis of methylation within nucleic acids subsequent to bisulfite treatment, it is required

that the probe be methylation specific, as described in United States Patent No. 6,331,393, (hereby incorporated by reference in its entirety) also known as the MethyLight assay. The second preferred embodiment of this MethyLight technology is the use of dual-probe technology (Lightcycler®), each probe carrying donor or recipient fluorescent moieties, hybridization of two probes in proximity to each other is indicated by an increase or fluorescent amplification primers. Both these techniques may be adapted in a manner suitable for use with bisulfite treated DNA, and moreover for methylation analysis within CpG dinucleotides.

Also any combination of these probes or combinations of these probes with other known probes may be used.

In a further preferred embodiment of the method, the fourth step of the method comprises the use of template-directed oligonucleotide extension, such as MS-SNuPE as described by Gonzalgo and Jones, *Nucleic Acids Res.* 25:2529-2531, 1997. In said embodiment it is preferred that the methylation specific single nucleotide extension primer (MS-SNuPE primer) is identical or complementary to a sequence at least nine but preferably no more than twenty five nucleotides in length of one or more of the sequences taken from the group of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16. In a preferred embodiment said MS-SNuPE primer is identical or complementary to a sequence at least nine but preferably no more than twenty five nucleotides in length of one or more of the pretreated sequences selected from the group consisting SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20. It is preferred to use fluorescently labeled nucleotides, instead of radiolabeled nucleotides.

In yet a further embodiment of the method, the fourth step of the method comprises sequencing and subsequent sequence analysis of the amplificate generated in the third step of the method (Sanger F., et al., *Proc Natl Acad Sci USA* 74:5463-5467, 1977).

In the most preferred embodiment of the methylation analysis method the genomic nucleic acids are isolated and treated according to the first three steps of the method outlined above, namely:

- a) obtaining, from a subject, a biological sample having subject genomic DNA;
- b) extracting or otherwise isolating the genomic DNA;

- c) treating the genomic DNA of b), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties; and wherein
- d) amplifying subsequent to treatment in c) is carried out in a methylation specific manner, namely by use of methylation specific primers or blocking oligonucleotides, and further wherein
- e) detecting of the amplicates is carried out by means of a real-time detection probe, as described above.

Preferably, where the subsequent amplification of d) is carried out by means of methylation specific primers, as described above, said methylation specific primers comprise a sequence having a length of at least 9 nucleotides which hybridizes to a treated nucleic acid sequence according to one of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide. It is preferred that said methylation specific primers comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence selected from the group consisting SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

Additionally, further methylation specific primers may also be used for the analysis of a gene panel as described above wherein said primers comprise a sequence having a length of at least 9 nucleotides which hybridizes to a treated nucleic acid sequence according to one of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

In an alternative most preferred embodiment of the method, the subsequent amplification of d) is carried out in the presence of blocking oligonucleotides, as described above. It is particularly preferred that said blocking oligonucleotides comprise a sequence having a length of at least 9 nucleotides which hybridizes to a treated nucleic acid sequence according to one of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG, TpG or CpA dinucleotide.

It is preferred that said blocking oligonucleotides comprise a sequence having a length of at least 9 nucleotides which hybridizes to a pretreated nucleic acid sequence selected from the group consisting SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20 and sequences complementary thereto, wherein the base sequence of said oligomers comprises at least one CpG dinucleotide.

Step e) of the method, namely the detection of the specific amplicates indicative of the methylation status of one or more CpG positions according to SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16, is carried out by means of real-time detection methods as described above.

Additional embodiments of the invention provide a method for the analysis of the methylation status of LHX3 and PITX3 and/or regulatory regions thereof without the need for pre-treatment.

In the *first step* of such additional embodiments, the genomic DNA sample is isolated from tissue or cellular sources. Preferably, such sources include cell lines, histological slides, biopsy tissue, body fluids, or tumor tissue embedded in paraffin. Extraction may be by means that are standard to one skilled in the art, including but not limited to the use of detergent lysates, sonification and vortexing with glass beads. Once the nucleic acids have been extracted, the genomic double-stranded DNA is used in the analysis.

In a preferred embodiment, the DNA may be cleaved prior to the treatment, and this may be by any means standard in the state of the art, but preferably with methylation-sensitive restriction endonucleases.

In the *second step*, the DNA is then digested with one or more methylation sensitive restriction enzymes. The digestion is carried out such that hydrolysis of the DNA at the restriction site is informative of the methylation status of a specific CpG dinucleotide.

In the *third step*, which is optional but a preferred embodiment, the restriction fragments are amplified. This is preferably carried out using a polymerase chain reaction, and said amplicates may carry suitable detectable labels as discussed above, namely fluorophore labels, radionuclides and mass labels.

In the *fourth step* the amplicates are detected. The detection may be by any means standard in the art, for example, but not limited to, gel electrophoresis analysis, hybridization analysis, incorporation of detectable tags within the PCR products, DNA array analysis, MALDI or ESI analysis.

In the final step of the method the prognosis and/or predicting outcome of anthracycline and/or endocrine treatment is determined. Preferably, the correlation of the expression level of the genes with the prognosis and/or predicting outcome of anthracycline and/or endocrine treatment is done substantially without human intervention. Poor prognosis in both general terms (i.e. independent of treatment) and with regard to treatment (i.e. after anthracycline and/or endocrine therapy) is determined by underexpression of mRNA and/or protein, and methylation and hypermethylation of CpG positions. Good prognosis is associated with expression of mRNA and/or protein, and hypomethylation of CpG positions of the genes PITX3 and LHX3. Good prognosis after treatment with anthracycline and/or endocrine therapy is associated with expression of mRNA and/or protein, and hypomethylation of CpG positions of the genes PITX3 and LHX3.

It is particularly preferred that the classification of the sample is carried out by algorithmic means.

In one embodiment machine learning predictors are trained on the methylation patterns at the investigated CpG sites of the samples with known status. A selection of the CpG positions which are discriminative for the machine learning predictor are used in the panel. In a particularly preferred embodiment of the method, both methods are combined; that is, the machine learning classifier is trained only on the selected CpG positions that are significantly differentially methylated between the classes according to the statistical analysis.

The disclosed invention provides treated nucleic acids, derived from genomic SEQ ID NO: 1 and SEQ ID NO: 2, wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization. The genomic sequences in question may comprise one, or more, consecutive or random methylated CpG positions. Said treatment preferably comprises use of a reagent selected from the group consisting of bisulfite,

hydrogen sulfite, disulfite, and combinations thereof. In a preferred embodiment the invention provides a non-naturally occurring modified nucleic acid comprising a sequence of at least 18 contiguous nucleotide bases in length of a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16, wherein said sequence comprises at least one CpG, TpA or CpA dinucleotide and sequences complementary thereto.

In further preferred embodiments of the invention said nucleic acid is at least 50, 100, 150, 200, 250 or 500 base pairs in length of a segment of the nucleic acid sequence disclosed in SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16. Particularly preferred is a nucleic acid molecule that is not identical or complementary to all or a portion of the sequences SEQ ID NO: 1 and SEQ ID NO: 2 or other naturally occurring DNA.

Further preferred is a nucleic acid at least 18, 50, 100, 150, 200, 250 or 500 base pairs in length of a segment of the nucleic acid sequence disclosed in SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20 that is not identical or complementary to all or a portion of the sequences SEQ ID NO: 3 and SEQ ID NO: 4 or other naturally occurring DNA.

The sequences of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 provide non-naturally occurring modified versions of the nucleic acid according to SEQ ID NO: 1 and SEQ ID NO: 2, wherein the modification of each genomic sequence results in the synthesis of a nucleic acid having a sequence that is unique and distinct from said genomic sequence as follows. For each sense strand genomic DNA, e.g., SEQ ID NO: 1, four converted versions are disclosed. A first version wherein "C" to "T," but "CpG" remains "CpG" (i.e., corresponds to case where, for the genomic sequence, all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted); a second version discloses the complement of the disclosed genomic DNA sequence (i.e. antisense strand), wherein "C" to "T," but "CpG" remains "CpG" (i.e., corresponds to case where, for all "C" residues of CpG dinucleotide sequences are methylated and are thus not converted). The 'upmethylated' converted sequences of SEQ ID NO: 1 and SEQ ID NO: 2 correspond to SEQ ID NO: 5 to SEQ ID NO: 8. A third chemically converted version of each genomic sequences is provided, wherein "C" to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (i.e., corresponds to case where, for the genomic sequences, all "C" residues of CpG dinucleotide sequences are unmethylated); a final chemically converted version of each

sequence, discloses the complement of the disclosed genomic DNA sequence (i.e. antisense strand), wherein "C" to "T" for all "C" residues, including those of "CpG" dinucleotide sequences (i.e., corresponds to case where, for the complement (antisense strand) of each genomic sequence, all "C" residues of CpG dinucleotide sequences are unmethylated). The 'downmethylated' converted sequences of SEQ ID NO: 1 and SEQ ID NO: 2 correspond to SEQ ID NO: 13 to SEQ ID NO: 16.

The invention further discloses oligonucleotide or oligomer for detecting the cytosine methylation state within genomic or pre-treated DNA, according to SEQ ID NO: 1 and SEQ ID NO: 2 and SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16. Said oligonucleotide or oligomer comprising a nucleic acid sequence having a length of at least nine (9) nucleotides which hybridizes, under moderately stringent or stringent conditions (as defined herein above), to a treated nucleic acid sequence according to SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and/or sequences complementary thereto, or to a genomic sequence according to SEQ ID NO: 1 and SEQ ID NO: 2 and/or sequences complementary thereto.

Thus, the present invention includes nucleic acid molecules (e.g., oligonucleotides and peptide nucleic acid (PNA) molecules (PNA-oligomers)) that hybridize under moderately stringent and/or stringent hybridization conditions to all or a portion of the sequences SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16, or to the complements thereof. The hybridizing portion of the hybridizing nucleic acids is typically at least 9, 15, 20, 25, 30 or 35 nucleotides in length. However, longer molecules have inventive utility, and are thus within the scope of the present invention.

Preferably, the hybridizing portion of the inventive hybridizing nucleic acids is at least 95%, or at least 98%, or 100% identical to the sequence, or to a portion thereof of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16, or to the complements thereof.

Hybridizing nucleic acids of the type described herein can be used, for example, as a primer (e.g., a PCR primer), or a diagnostic and/or prognostic probe or primer. Preferably, hybridization of the oligonucleotide probe to a nucleic acid sample is performed under stringent conditions and the probe is 100% identical to the target sequence. Nucleic acid duplex or hybrid stability is expressed as the melting temperature or T_m , which is the

temperature at which a probe dissociates from a target DNA. This melting temperature is used to define the required stringency conditions.

For target sequences that are related and substantially identical to the corresponding sequence of SEQ ID NO: 1 and SEQ ID NO: 2 (such as allelic variants and SNPs), rather than identical, it is useful to first establish the lowest temperature at which only homologous hybridization occurs with a particular concentration of salt (e.g., SSC or SSPE). Then, assuming that 1% mismatching results in a 1°C decrease in the T_m , the temperature of the final wash in the hybridisation reaction is reduced accordingly (for example, if sequences having > 95% identity with the probe are sought, the final wash temperature is decreased by 5°C). In practice, the change in T_m can be between 0.5°C and 1.5°C per 1% mismatch.

Examples of inventive oligonucleotides of length X (in nucleotides), as indicated by polynucleotide positions with reference to, e.g., SEQ ID NO:1, include those corresponding to sets (sense and antisense sets) of consecutively overlapping oligonucleotides of length X , where the oligonucleotides within each consecutively overlapping set (corresponding to a given X value) are defined as the finite set of Z oligonucleotides from nucleotide positions:

n to $(n + (X-1))$;

where $n=1, 2, 3, \dots, (Y-(X-1))$;

where Y equals the length (nucleotides or base pairs) of SEQ ID NO: 1 (10858);

where X equals the common length (in nucleotides) of each oligonucleotide in the set (e.g., $X=20$ for a set of consecutively overlapping 20-mers); and

where the number (Z) of consecutively overlapping oligomers of length X for a given SEQ ID NO of length Y is equal to $Y-(X-1)$. For example $Z=10858-19=10839$ for either sense or antisense sets of SEQ ID NO: 1, where $X=20$.

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

Examples of inventive 20-mer oligonucleotides include the following set of oligomers (and the antisense set complementary thereto), indicated by polynucleotide positions with reference to SEQ ID NO: 1: 1 to 20, 2 to 21, 3 to 22, 4 to 23, 5 to 24, ... and 10839 to 10858.

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

Likewise, examples of inventive 25-mer oligonucleotides include the following set of oligomers (and the antisense set complementary thereto), indicated by polynucleotide positions with reference to SEQ ID NO: 1: 1-25, 2-26, 3-27, 4-28, 5-29, ...and 10853 to 10858.

Preferably, the set is limited to those oligomers that comprise at least one CpG, TpG or CpA dinucleotide.

The present invention encompasses, for each of SEQ ID NO: 1 and SEQ ID NO: 2 and SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 (sense and antisense), multiple consecutively overlapping sets of oligonucleotides or modified oligonucleotides of length X, where, e.g., X= 9, 10, 17, 20, 22, 23, 25, 27, 30 or 35 nucleotides.

The oligonucleotides or oligomers according to the present invention constitute effective tools useful to ascertain genetic and epigenetic parameters of the genomic sequence corresponding to SEQ ID NO: 1 and SEQ ID NO: 2 . Preferred sets of such oligonucleotides or modified oligonucleotides of length X are those consecutively overlapping sets of oligomers corresponding to SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 (and to the complements thereof). Preferably, said oligomers comprise at least one CpG, TpG or CpA dinucleotide.

Particularly preferred oligonucleotides or oligomers according to the present invention are those in which the cytosine of the CpG dinucleotide (or of the corresponding converted TpG or CpA dinucleotide) sequences is within the middle third of the oligonucleotide; that is, where the oligonucleotide is, for example, 13 bases in length, the CpG, TpG or CpA dinucleotide is positioned within the fifth to ninth nucleotide from the 5'-end.

The oligonucleotides of the invention can also be modified by chemically linking the oligonucleotide to one or more moieties or conjugates to enhance the activity, stability or detection of the oligonucleotide. Such moieties or conjugates include chromophores, fluorophores, lipids such as cholesterol, cholic acid, thioether, aliphatic chains, phospholipids,

polyamines, polyethylene glycol (PEG), palmityl moieties, and others as disclosed in, for example, United States Patent Numbers 5,514,758, 5,574,142, 5,585,481, 5,587,371, 5,597,696 and 5,958,773. The probes may also exist in the form of a PNA (peptide nucleic acid) which has particularly preferred pairing properties. Thus, the oligonucleotide may include other appended groups such as peptides, and may include hybridization-triggered cleavage agents (Krol et al., *BioTechniques* 6:958-976, 1988) or intercalating agents (Zon, *Pharm. Res.* 5:539-549, 1988). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a chromophore, fluorophore, peptide, hybridization-triggered cross-linking agent, transport agent, hybridisation-triggered cleavage agent, etc.

The oligonucleotide may also comprise at least one art-recognized modified sugar and/or base moiety, or may comprise a modified backbone or non-natural internucleoside linkage.

The oligonucleotides or oligomers according to particular embodiments of the present invention are typically used in 'sets,' which contain at least one oligomer for analysis of each of the CpG dinucleotides of genomic sequences SEQ ID NO: 1 and SEQ ID NO: 2 and sequences complementary thereto, or to the corresponding CpG, TpG or CpA dinucleotide within a sequence of the treated nucleic acids according to SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and sequences complementary thereto. However, it is anticipated that for economic or other factors it may be preferable to analyze a limited selection of the CpG dinucleotides within said sequences, and the content of the set of oligonucleotides is altered accordingly.

Therefore, in particular embodiments, the present invention provides a set of at least two (2) (oligonucleotides and/or PNA-oligomers) useful for detecting the cytosine methylation state of treated genomic DNA (SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16), or in genomic DNA (SEQ ID NO: 1 and SEQ ID NO: 2 and sequences complementary thereto). These probes enable diagnosis, and/or classification of genetic and epigenetic parameters of cell proliferative disorders. The set of oligomers may also be used for detecting single nucleotide polymorphisms (SNPs) in treated genomic DNA (SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16), or in genomic DNA (SEQ ID NO: 1 and SEQ ID NO: 2 and sequences complementary thereto).

In preferred embodiments, at least one, and more preferably all members of a set of

oligonucleotides is bound to a solid phase.

In further embodiments, the present invention provides a set of at least two (2) oligonucleotides that are used as 'primer' oligonucleotides for amplifying DNA sequences of one of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and sequences complementary thereto, or segments thereof.

It is anticipated that the oligonucleotides may constitute all or part of an "array" or "DNA chip" (i.e., an arrangement of different oligonucleotides and/or PNA-oligomers bound to a solid phase). Such an array of different oligonucleotide- and/or PNA-oligomer sequences can be characterized, for example, in that it is arranged on the solid phase in the form of a rectangular or hexagonal lattice. The solid-phase surface may be composed of silicon, glass, polystyrene, aluminium, steel, iron, copper, nickel, silver, or gold. Nitrocellulose as well as plastics such as nylon, which can exist in the form of pellets or also as resin matrices, may also be used. An overview of the prior art in oligomer array manufacturing can be gathered from a special edition of Nature Genetics (Nature Genetics Supplement, Volume 21, January 1999, and from the literature cited therein). Fluorescently labeled probes are often used for the scanning of immobilized DNA arrays. The simple attachment of Cy3 and Cy5 dyes to the 5'-OH of the specific probe are particularly suitable for fluorescence labels. The detection of the fluorescence of the hybridized probes may be carried out, for example, via a confocal microscope. Cy3 and Cy5 dyes, besides many others, are commercially available.

It is also anticipated that the oligonucleotides, or particular sequences thereof, may constitute all or part of a "virtual array" wherein the oligonucleotides, or particular sequences thereof, are used, for example, as 'specifiers' as part of, or in combination with a diverse population of unique labeled probes to analyze a complex mixture of analytes. Such a method, for example is described in US 2003/0013091 (United States serial number 09/898,743, published 16 January 2003). In such methods, enough labels are generated so that each nucleic acid in the complex mixture (i.e., each analyte) can be uniquely bound by a unique label and thus detected (each label is directly counted, resulting in a digital read-out of each molecular species in the mixture).

The described invention further provides a composition of matter useful for providing a prediction of outcome of treatment comprising at least one anthracycline or endocrine therapy

of cancer patients. Said composition comprising at least one nucleic acid 18 base pairs in length of a segment of the nucleic acid sequence disclosed in SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16, and one or more substances taken from the group comprising : magnesium chloride, dNTP, taq polymerase, bovine serum albumen, an oligomer in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridizes under moderately stringent or stringent conditions to a pretreated genomic DNA according to one of the SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and sequences complementary thereto. It is preferred that said composition of matter comprises a buffer solution appropriate for the stabilization of said nucleic acid in an aqueous solution and enabling polymerase based reactions within said solution. Suitable buffers are known in the art and commercially available.

In further preferred embodiments of the invention said at least one nucleic acid is at least 50, 100, 150, 200, 250 or 500 base pairs in length of a segment of the nucleic acid sequence disclosed in SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16.

Kits

Moreover, an additional aspect of the present invention is a kit comprising: a means for determining LHX3 and/or PITX3 methylation. The means for determining LHX3 and/or PITX3 methylation comprise preferably a bisulfite-containing reagent; one or a plurality of oligonucleotides consisting whose sequences in each case are identical, are complementary, or hybridise under stringent or highly stringent conditions to a 9 or more preferably 18 base long segment of a sequence selected from SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 or more preferably SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20; and optionally instructions for carrying out and evaluating the described method of methylation analysis. In one embodiment the base sequence of said oligonucleotides comprises at least one CpG, CpA or TpG dinucleotide.

In a further embodiment, said kit may further comprise standard reagents for performing a CpG position-specific methylation analysis, wherein said analysis comprises one or more of the following techniques: MS-SNuPE, MSP, MethyLightTM, HeavyMethyl, COBRA, and nucleic acid sequencing. However, a kit along the lines of the present invention can also contain only part of the aforementioned components.

In a preferred embodiment the kit may comprise additional bisulfite conversion

reagents selected from the group consisting: DNA denaturation buffer; sulfonation buffer; DNA recovery reagents or kits (*e.g.*, precipitation, ultrafiltration, affinity column); desulfonation buffer; and DNA recovery components.

In a further alternative embodiment, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimised for primer extension mediated by the polymerase, such as PCR. In another embodiment of the invention the kit further comprising means for obtaining a biological sample of the patient. Preferred is a kit, which further comprises a container suitable for containing the means for determining methylation of the genes LHX3 and/or PITX3 in the biological sample of the patient, and most preferably further comprises instructions for use and interpretation of the kit results. In a preferred embodiment the kit comprises: (a) a bisulfite reagent; (b) a container suitable for containing the said bisulfite reagent and the biological sample of the patient; (c) at least one set of primer oligonucleotides containing two oligonucleotides whose sequences in each case are identical, are complementary, or hybridise under stringent or highly stringent conditions to a 9 or more preferably 18 base long segment of a sequence selected from SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 or more preferably SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20; and optionally (d) instructions for use and interpretation of the kit results. In an alternative preferred embodiment the kit comprises: (a) a bisulfite reagent; (b) a container suitable for containing the said bisulfite reagent and the biological sample of the patient; (c) at least one oligonucleotides and/or PNA-oligomer having a length of at least 9 or 16 nucleotides which is identical to or hybridises to a pre-treated nucleic acid sequence according to one of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 or more preferably SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20 and sequences complementary thereto; and optionally (d) instructions for use and interpretation of the kit results.

In an alternative embodiment the kit comprises: (a) a bisulfite reagent; (b) a container suitable for containing the said bisulfite reagent and the biological sample of the patient; (c) at least one set of primer oligonucleotides containing two oligonucleotides whose sequences in each case are identical, are complementary, or hybridise under stringent or highly stringent conditions to a 9 or more preferably 18 base long segment of a sequence selected from SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 or more preferably SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20; (d) at least one oligonucleotides and/or PNA-oligomer having a length of at least 9 or 16 nucleotides which is identical to or hybridises to a pre-treated nucleic acid sequence according to one of SEQ ID

NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 or more preferably SEQ ID NO: 9 to SEQ ID NO: 12 and SEQ ID NO: 17 to SEQ ID NO: 20 and sequences complementary thereto; and optionally (e) instructions for use and interpretation of the kit results.

The kit may also contain other components such as buffers or solutions suitable for blocking, washing or coating, packaged in a separate container.

Another aspect of the invention relates to a kit for use in determining the prognosis and/or treatment response of cell proliferative disorders according to the methods of the present invention, said kit comprising: a means for measuring the level of transcription of the gene LHX3 and/or PITX3 and a means for determining LHX3 and/or PITX3 methylation.

Typical reagents (*e.g.*, as might be found in a typical COBRATM-based kit) for COBRATM analysis may include, but are not limited to: PCR primers for LHX3 and/or PITX3; restriction enzyme and appropriate buffer; gene-hybridization oligo; control hybridization oligo; kinase labeling kit for oligo probe; and labeled nucleotides. Typical reagents (*e.g.*, as might be found in a typical MethyLightTM-based kit) for MethyLightTM analysis may include, but are not limited to: PCR primers for the bisulfite converted sequence of the LHX3 and/or PITX3 gene; bisulfite specific probes (*e.g.* TaqManTM or LightcyclerTM); optimized PCR buffers and deoxynucleotides; and Taq polymerase.

Typical reagents (*e.g.*, as might be found in a typical Ms-SNuPETM-based kit) for Ms-SNuPETM analysis may include, but are not limited to: PCR primers for specific gene (or bisulfite treated DNA sequence or CpG island); optimized PCR buffers and deoxynucleotides; gel extraction kit; positive control primers; Ms-SNuPETM primers for the bisulfite converted sequence of the LHX3 and/or PITX3 gene; reaction buffer (for the Ms-SNuPE reaction); and labelled nucleotides.

Typical reagents (*e.g.*, as might be found in a typical MSP-based kit) for MSP analysis may include, but are not limited to: methylated and unmethylated PCR primers for the bisulfite converted sequence of the LHX3 and/or PITX3 gene, optimized PCR buffers and deoxynucleotides, and specific probes.

Moreover, an additional aspect of the present invention is an alternative kit comprising a means for determining LHX3 and/or PITX3 methylation, wherein said means comprise preferably at least one methylation specific restriction enzyme; one or a plurality of primer oligonucleotides (preferably one or a plurality of primer pairs) suitable for the amplification of a sequence comprising at least one CpG dinucleotide of a sequence selected from SEQ ID NO: 1 and SEQ ID NO: 2 or more preferably SEQ ID NO: 3 and SEQ ID NO: 4; and optionally instructions for carrying out and evaluating the described method of methylation analysis. In one embodiment the base sequence of said oligonucleotides are identical, are complementary, or hybridise under stringent or highly stringent conditions to an at least 18 base long segment of a sequence selected from SEQ ID NO: 1 and SEQ ID NO: 2 or more preferably SEQ ID NO: 3 and SEQ ID NO: 4.

In a further embodiment said kit may comprise one or a plurality of oligonucleotide probes for the analysis of the digest fragments, preferably said oligonucleotides are identical, are complementary, or hybridise under stringent or highly stringent conditions to an at least 16 base long segment of a sequence selected from SEQ ID NO: 1 and SEQ ID NO: 2 or more preferably SEQ ID NO: 3 and SEQ ID NO: 4.

In a preferred embodiment the kit may comprise additional reagents selected from the group consisting: buffer (e.g. restriction enzyme, PCR, storage or washing buffers); DNA recovery reagents or kits (e.g., precipitation, ultrafiltration, affinity column) and DNA recovery components.

In a further alternative embodiment, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimised for primer extension mediated by the polymerase, such as PCR. In another embodiment of the invention the kit further comprising means for obtaining a biological sample of the patient. In a preferred embodiment the kit comprises: (a) a methylation sensitive restriction enzyme reagent; (b) a container suitable for containing the said reagent and the biological sample of the patient; (c) at least one set of oligonucleotides one or a plurality of nucleic acids or peptide nucleic acids which are identical, are complementary, or hybridise under stringent or highly stringent conditions to an at least 9 base long segment of a sequence selected from SEQ ID NO: 1 and SEQ ID NO: 2 or more preferably SEQ ID NO: 3 and SEQ ID NO: 4; and optionally (d) instructions for use and interpretation of the kit results.

In an alternative preferred embodiment the kit comprises: (a) a methylation sensitive restriction enzyme reagent; (b) a container suitable for containing the said reagent and the biological sample of the patient; (c) at least one set of primer oligonucleotides suitable for the amplification of a sequence comprising at least one CpG dinucleotide of a sequence selected from SEQ ID NO: 1 and SEQ ID NO: 2 or more preferably SEQ ID NO: 3 and SEQ ID NO: 4; and optionally (d) instructions for use and interpretation of the kit results.

In an alternative embodiment the kit comprises: (a) a methylation sensitive restriction enzyme reagent; (b) a container suitable for containing the said reagent and the biological sample of the patient; (c) at least one set of primer oligonucleotides suitable for the amplification of a sequence comprising at least one CpG dinucleotide of a sequence selected from SEQ ID NO: 1 and SEQ ID NO: 2 or more preferably SEQ ID NO: 3 and SEQ ID NO: 4; (d) at least one set of oligonucleotides one or a plurality of nucleic acids or peptide nucleic acids which are identical, are complementary, or hybridise under stringent or highly stringent conditions to an at least 9 base long segment of a sequence selected from SEQ ID NO: 1 and SEQ ID NO: 2 or more preferably SEQ ID NO: 3 and SEQ ID NO: 4 and optionally (e) instructions for use and interpretation of the kit results.

The kit may also contain other components such as buffers or solutions suitable for blocking, washing or coating, packaged in a separate container.

The invention further relates to a kit for use in providing a prognosis or prediction of treatment response of a cell proliferative disorder according to the methods of the present invention in a subject by means of methylation-sensitive restriction enzyme analysis. Said kit comprises a container and a DNA microarray component. Said DNA microarray component being a surface upon which a plurality of oligonucleotides are immobilized at designated positions and wherein the oligonucleotide comprises at least one CpG methylation site. At least one of said oligonucleotides is specific for the gene LHX3 and/or PITX3 and comprises a sequence of at least 15 base pairs in length but no more than 200 bp of a sequence according to one of SEQ ID NO: 1 and SEQ ID NO: 2 or more preferably SEQ ID NO: 3 and SEQ ID NO: 4. Preferably said sequence is at least 15 base pairs in length but no more than 80 bp of a sequence according to one of SEQ ID NO: 1 and SEQ ID NO: 2 or more preferably SEQ ID NO: 3 and SEQ ID NO: 4. It is further preferred that said sequence is at least 20 base pairs in

length but no more than 30 bp of a sequence according to one of SEQ ID NO: 1 and SEQ ID NO: 2 or more preferably SEQ ID NO: 3 and SEQ ID NO: 4.

Said test kit preferably further comprises a restriction enzyme component comprising one or a plurality of methylation-sensitive restriction enzymes.

In a further embodiment said test kit is further characterized in that it comprises at least one methylation-specific restriction enzyme, and wherein the oligonucleotides comprise a restriction site of said at least one methylation specific restriction enzymes.

The kit may further comprise one or several of the following components, which are known in the art for DNA enrichment: a protein component, said protein binding selectively to methylated DNA; a triplex-forming nucleic acid component, one or a plurality of linkers, optionally in a suitable solution; substances or solutions for performing a ligation e.g. ligases, buffers; substances or solutions for performing a column chromatography; substances or solutions for performing an immunology based enrichment (e.g. immunoprecipitation); substances or solutions for performing a nucleic acid amplification e.g. PCR; a dye or several dyes, if applicable with a coupling reagent, if applicable in a solution; substances or solutions for performing a hybridization; and/or substances or solutions for performing a washing step.

While the present invention has been described with specificity in accordance with certain of its preferred embodiments, the following examples and figures serve only to illustrate the invention and is not intended to limit the invention within the principles and scope of the broadest interpretations and equivalent configurations thereof.

Table 1: Genes and CpG rich regions thereof according to the present invention

	LHX3	PITX3
Genomic SEQ ID NO:	1	2
Pretreated methylated sequence (sense) SEQ ID NO:	5	7
Pretreated methylated strand (antisense) SEQ ID NO:	6	8
Pretreated unmethylated sequence (sense) SEQ ID NO:	13	15
Pretreated unmethylated sequence (antisense) SEQ ID NO:	14	16
CpG rich region genomic SEQ ID NO:	3	4
CpG rich region pretreated methylated sequence (sense) SEQ ID NO:	9	11
CpG rich region pretreated methylated strand (antisense) SEQ ID NO:	10	12
CpG rich region pretreated unmethylated sequence (sense) SEQ ID NO:	17	19

CpG rich region pretreated unmethylated sequence (antisense) SEQ ID NO:	18	20
---	----	----

Example 1: Microarray Analysis

Microarray analysis

To evaluate marker candidates as predictive markers (based on chemotherapeutic response) a significant number of patient samples were analyzed using the applicant's proprietary methylation sensitive Microarray technology.

Patient Samples

Samples were obtained from 384 lymphnode positive breast cancer patients who had received adjuvant treatment with an anthracycline based treatment regimen. Samples were received from four academic partners in the form of fresh frozen tissues.

Gene selection

64 candidate marker sequences were selected, it has to be noted that not all sequences of the panel currently map to known genes, furthermore not all of said markers are the subject of the present invention.

DNA Extraction

DNA from samples was isolated using commercially available kits.

PCR establishment and Multiplex PCR optimization

To amplify all gene fragments, PCR assays were designed to match bisulfite treated DNA and to allow amplification independent of the methylation status of the respective fragment. A standardized primer design workflow optimized by the applicant for bisulfite treated DNA was employed. Individual PCR assays were considered established when successful amplification on bisulfite treated DNA was reproducible and no background amplification of genomic DNA was detectable, ensuring bisulfite DNA specific amplification. PCR primers were:

LHX3:gggtttggagtttaggga (SEQ ID NO: 21)

LHX3:cctcaatatcctactaaaacttacc (SEQ ID NO: 22)

PITX3: ttagaggtaggtaggggtt (SEQ ID NO: 23)

PITX3: aaacctaataatccacaactaaa (SEQ ID NO: 24)

To allow efficient amplification, individual PCR assays were combined into multiplex PCR (mPCR) assays. Several multiplex PCR sets were calculated based on the primer sequences of the individual PCR amplicates and tested on lymphocyte DNA. Based on ALF express analyses the best performing combination of multiplex PCR sets were chosen.

Bisulfite Treatment and Multiplex PCR

Total genomic DNA of all samples and controls was bisulfite treated converting unmethylated cytosines to uracil. Methylated cytosines are conserved. Bisulfite treatment was performed according to the applicant's optimized proprietary bisulfite treatment procedure as described above. In order to avoid a potential process bias, the samples were randomized into processing batches.

Hybridization

All PCR products from each individual sample were then hybridized to glass slides carrying a pair of immobilized oligonucleotides for each CpG position under analysis. For hybridizations, the samples were grouped into processing batches in order to avoid a potential process-bias. The samples were processed randomized for bisulfite batches. Each detection oligonucleotide was designed to hybridize to the bisulphite converted sequence around one CpG site which was either originally unmethylated (TG) or methylated (CG). Hybridization conditions were selected to allow the detection of the single nucleotide differences between the TG and CG variants.

LHX3:

ggcgggattcggggta(SEQ ID NO: 25)

gggtgggatttggggta(SEQ ID NO: 26)

gggaggagtgggtatgg(SEQ ID NO: 27)

aggagcgggtacgggt(SEQ ID NO: 28)

atcgagcgaggttcgg(SEQ ID NO: 29)

attgagtgaggtttgggggt(SEQ ID NO: 30)

agtatcgcgatagcgt(SEQ ID NO: 31)

tattgtggatagtgttaggttt(SEQ ID NO: 32)

ggggattgggggtgatga(SEQ ID NO: 33)

atcggggcgacgagag(SEQ ID NO: 34)

PITX3:

gggggttgggtatttggt(SEQ ID NO: 35)

aggggtcgggtattcgg(SEQ ID NO: 36)

attcggtcggagtggg(SEQ ID NO: 37)

atttgggtggagtgggg(SEQ ID NO: 38)

agggcgtttttagttcgt(SEQ ID NO: 39)

attagggtgttttagttgt(SEQ ID NO: 40)

gtcgggtattcggtcgg(SEQ ID NO: 41)

gttgggtatttggtggag(SEQ ID NO: 42)

tagttcgtcgcggcga(SEQ ID NO: 43)

ttagttgtgtggtgagt(SEQ ID NO: 44)

Fluorescent signals from each hybridized oligonucleotide were detected using genepix scanner and software. Ratios for the two signals (from the CG oligonucleotide and the TG oligonucleotide used to analyze each CpG position) were calculated based on comparison of intensity of the fluorescent signals.

The samples were processed in randomized batches. For each bisulfite treated DNA sample 2 hybridizations were performed. This means that for each sample a total number of 4 chips were processed.

1820 chips were processed in total. In addition to the 435 patients samples these further included 80 control chips; 40 hybridised to completely methylated control DNA and 40 hybridised to completely unmethylated control DNA. Of these a total of 153 were excluded due to poor quality. The remaining 1667 chips covered all 435 samples. Of these 435 samples only 384 had complete follow ups, the remaining were excluded.

Data Analysis

For the analysis of chip data, the applicants' proprietary software was used. Said software contains a data warehouse that supports queries to sample, genome and laboratory management databases, respectively. It encompasses a variety of statistical tools for analyzing

and visualizing methylation array data. In the following sections we summarize the most important data analysis techniques that were applied for analyzing the data.

From Raw Hybridization Intensities to Methylation Ratios

The log methylation ratio ($\log(\text{CG}/\text{TG})$) at each CpG position was determined according to a standardized preprocessing pipeline that includes the following steps:

- For each spot the median background pixel intensity is subtracted from the median foreground pixel intensity. This gives a good estimate of background corrected hybridization intensities.
- For both CG and TG detection oligonucleotides of each CpG position the background corrected median of the 4 redundant spot intensities is taken.
- For each chip and each CG/TG oligo pair, the $\log(\text{CG}/\text{TG})$ ratio is calculated.
- For each sample the median of $\log(\text{CG}/\text{TG})$ intensities over the redundant chip repetitions is taken.

This log ratio has the property that the hybridization noise has approximately constant variance over the full range of possible methylation rates (see e.g. Huber W, Von Heydebreck A, Sultmann H, Poustka A, Vingron M. 2002. Variance stabilization applied to Microarray data calibration and to the quantification of differential expression. *Bioinformatics*. 18 Suppl 1: S96–S104.)

Principle Component Analysis

Principle component analysis (PCA) projects measurement vectors (e.g. chip data, methylation profiles on several CpG sites etc.) onto a new coordinate system. The new coordinate axes are referred to as principal components. The first principal component spans the direction of largest variance of the data. Subsequent components are ordered by decreasing variance and are orthogonal to each other. Different CpG positions contribute with different weights to the extension of the data cloud along different components. PCA is an unsupervised technique, i.e. it does not take into account any group or label information of the data points (for further details see e.g. Ripley, B. D. 1996. *Pattern Recognition and Neural Networks*, Cambridge, UK, Cambridge University Press.).

PCA is typically used to project high dimensional data (in our case methylation-array data) onto lower dimensional subspaces in order to visualize or extract features with high variance

from the data. In the present report we used 2 dimensional projections for statistical quality control of the data. We investigated the effect of different process parameters on the chip data in order to rule out that changing process parameters caused large alterations in the measurement values.

A robust version of PCA was used to detect single outlier chips and exclude them from further analysis (Model F, Koenig T, Piepenbrock C, Adorjan P. 2002. Statistical process control for large scale Microarray experiments. *Bioinformatics*. 18 Suppl 1:S155-163.).

T² Control Charts

To control the general stability of the chip production process we use methods from the field of multivariate statistical process control (MVSPC). Our major tool is the T² control chart, which is used to detect significant deviations of the chip process from normal working conditions (Model F, Koenig T, Piepenbrock C, Adorjan P. 2002. Statistical process control for large scale Microarray experiments. *Bioinformatics*. 18 Suppl 1:S155-163.).

- Order the chip data with respect to a process parameter (e.g. hybridization data or spotting robot).
- Define a historic data set, which describes the chip process under normal working conditions (e.g. the first 75 hybridized chips). In the chart, data from the historical data set are indicated by a special plot symbol.
- Compute the distance of every new chip to the historic data set. If the distance of several consecutive chips exceeds a given control limit the process has to be regarded as out of control.

Use of T² charts to monitor the chip production process allows us to efficiently detect and eliminate most systematic error sources.

Statistical Methods

Cox Regression

The relation between metastases-free survival times (MFS) and covariates are analyzed using Cox Proportional Hazard models (Cox and Oates 1984; Harrel 2001).

The hazard, i.e. the instantaneous risk of a relapse, is modeled as

$$h(t | x) = h_0(t) \cdot \exp(\beta x) \quad (3)$$

and

$$h(t | x_1, \dots, x_k) = h_0(t) \cdot \exp(\beta_1 x_1 + \dots + \beta_k x_k) \quad (4)$$

for univariate and multiple regression analyses, respectively, where t is the time measured in months after surgery, $h_0(t)$ is the (unspecified) baseline hazard, x_i are the covariates (e.g. measurements of the assays) and β_i are the regression coefficients (parameters of the model). β_i will be estimated by maximizing the partial likelihood of the Cox Proportional Hazard model

Likelihood ratio tests are performed to test whether methylation is related to the hazard. The difference between $-2 \log(\text{Likelihood})$ of the full model and the null-model is approximately χ^2 -distributed with k degrees of freedom under the null hypotheses $\beta_1 = \dots = \beta_k = 0$

Stepwise Regression Analysis

For multiple Cox regression models a stepwise procedure (Venables and Ripley 1999; Harrell 2001) is used in order to find sub-models including only relevant variables. Two effects are usually achieved by these procedures:

- Variables (methylation rates) that are basically unrelated to the dependent variable (DFS/MFS) are excluded as they do not add relevant information to the model.
- Out of a set of highly correlated variables, only the one with the best relation to the dependent variable is retained.

The applied algorithm aims at minimizing the Akaike information criterion (AIC), which is defined as

$$\text{AIC} = -2 \cdot \text{maximized log-likelihood} + 2 \cdot \# \text{parameters}.$$

Kaplan-Meier Survival Curves and Log-Rank Tests

Survival curves are estimated from MFS data using Kaplan-Meier estimator for survival (Kaplan and Meier, 1958). Log-rank tests (Cox and Oates 1984) are used to test for differences of two survival curves, e.g. survival in hyper- vs. hypomethylated groups.

Multiple Test Corrections

Multiple test corrections were carried out by the False Discovery Rate (FDR) at 0.05.

Analysis of Sensitivity and Specificity

The method of calculating sensitivity and specificity using the Bayes-formula is based on the Kaplan-Meier estimates (Heagerty *et al.* 2000) for the survival probabilities in the marker positive and marker negative groups for a given time $T_{Threshold}$. The ROCs were calculated for different reference times $T_{Threshold}$ (36 months, 48 months, 60 months, 72 months) and time dependent AUCs were calculated (see Table 2).

Results

The p-value of all estrogen receptor positive patients (284) in differentiating between the two groups was 0,02600 for LHX3 and 0,00610 for PITX3.

Table 2: AUC of ER+ patient Kaplan-Meier estimates

Gene	36AUC	48AUC	60AUC	72AUC
LHX3	0,59	0,57	0,57	0,58
PITX3	0,54	0,65	0,65	0,64

Example 2: Sequencing of the genes LHX3 and PITX3

Introduction:

Bisulfite sequencing of 94 samples of breast cancer patients was carried out to show the prognostic potential of LHX3 (Fig. 1) and PITX3 (Fig. 7) methylation for predicting clinical outcome after breast cancer surgery. Samples were chosen according to their methylation level in the gene PITX2 – a gene which is known to be a prognostic biomarker (Multicenter study validates PITX2 DNA methylation for risk prediction in tamoxifen-treated, node-negative breast cancer using paraffin-embedded tumor tissue 2005 ASCO Annual Meeting; U.S. patent application number 11/011,332). Four different sample groups were analysed:

- A: 16 samples with low PITX2 methylation and no recurrence of cancer
- B: 31 samples with high PITX2 methylation and no recurrence of cancer
- C: 17 samples with high PITX2 methylation and recurrence of cancer
- D: 30 samples with low PITX2 methylation and recurrence of cancer

PITX2 methylation is known to be a prognostic marker for the recurrence of breast cancer after surgery: a low level of methylation is correlated to a low probability of recurrence

whereas a high probability of recurrence is indicated by a high level of PITX2 methylation. Therefore groups B and D (as described above) were wrongly classified using PITX2 methylation as a marker.

In addition, a second region within the LHX3 gene was analyzed using 35 breast cancer cell lines to show co-methylation of the CpGs within the LHX3 gene.

Figure 1 shows the structure of the gene LHX3 and location of the analyzed regions A and B (both comprised within SEQ ID NO: 1).

Materials and Methods:

Sequencing was carried out according to Lewin J, Schmitt AO, Adorjan P, Hildmann T, Piepenbrock C. (2004). Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplicates. *Bioinformatics*, 22: 3005-3012. PCR was performed in a total volume of 25 µl containing 5 ng template DNA, 1 U Hotstar Taq polymerase (Qiagen), 12,5 pmol of forward and reverse primers, 1x PCR buffer (Qiagen), 0,2 mM of each dNTP (Fermentas). Cycling was performed using a Mastercycler (Eppendorf) under the following conditions: 15 min at 95°C and 40 cycles at 95°C for 1 min, 55°C for 45 s and 72°C for 1 min.

PCR were sequenced using BigDye chemistry (Applied Biosystems) according to the manufacturer's recommendation. Quantitative methylation values were calculated using an algorithm developed by Lewin et. al (2004).

Amplicon sequences and primer sequences:

LHX3 A (genomic sequence) (SEQ ID NO: 45):

```
CAGTGTCTGCTGGGGCTTACCCCGAGTCCCGCCCAAGGTGCAGACGGCGGCGGC
CCCGGGCCTCGCTCGGTCTGCGCTCGAGCCCCGTTTCCAGCAGCATCGCGGCCACC
AGGCCGAGTGGCGCGAGACGCGCTCCTCCTAGGTCAGCGTCCCCTGGAGGGTTTCG
GGGCTCCCAAGTCCCGCCGCGTCGTGCGGGGCAGGGAGCCCGGGAGCCACTGGG
CCTGG
```

LHX3 A (bisulfite sequence) (SEQ ID NO: 46):

TAGTGTTTTGTTGGGGTTTATTTTCGAGTTTCGTTTAAGGTGTAGACGGCGGCGGTT
TCGGGTTTCGTTTCGGTCGCGTTCGAGTTTCGTTTTTAGTAGTATCGCGGTTATTAG
GTCGAGTGGCGCGAGACGCGTTTTTTTTAGGTTAGCGTTTTTTGGAGGGTTCGGG
GTTTTTAAGTTTCGTCGCGTCGTGCGGGGTAGGGAGTTCGGGAGTTATTGGGTTTG
G

Bisulfite specific primers for LHX3 A:

Forward primer: (SEQ ID NO: 47): TAGTGTTTTGTTGGGGTTTATTT

Reverse primer: (SEQ ID NO: 48): CCAAACCCAATAACTCCC

Furthermore, a second region (amplificate B, Figure 1) was sequenced in order to demonstrate co-methylation within the gene. Amplificate B was sequenced on 35 breast cancer cell line samples.

LHX3 B (genomic sequence) (SEQ ID NO: 49):

GGAAAGGCCTGAGGATCTCCTGGTCTCCCCGGTGCAGCCACTCGGCCCGGGCACC
CACCTCGGCGCAGGTCCGCCCTCCGCGCCAGCAGTGCTAGCAGCAGGT

LHX3 B (bisulfite sequence) (SEQ ID NO: 50):

GGAAAGGTTTGAGGATTTTTTGGTTTTTTCGGTGTAGTTATTCGGTTCGGGTATTT
ATTTGCGCGTAGGTTTCGTTTTTCGCGTTAGTAGTGTTAGTAGTAGGT

Bisulfite specific primers for LHX3 B:

Forward primer: (SEQ ID NO: 51): GGAAAGGTTTGAGGATTTTTTG

Reverse primer: (SEQ ID NO: 52): ACCTACTACTAACAATACTAAC

PITX3 (genomic sequence) (SEQ ID NO: 53):

CAGGCAGGGGCCAGGGGCCGGGCACCCGGCCGGAGTGGGGGCCGCCCCCTGCT
CCCGGGCCGCCTCTCCGCTCGGGCGCTCCTGGACTCTCGGAGGGAGTGAGCCTCA
CCGCGTACTGCCACCCCCAGCCGGCGCCCATTCACTTTATGGCAGACCAGGG

PITX3 (bisulfite sequence) (SEQ ID NO: 54):

TAGGTAGGGGTTAGGGGTCGGGTATTCGGTCGGAGTGGGGGTCGTTTTTTGTTTT
CGGGTCGTTTTTTCGTTTCGGGCGTTTTTGGATTTTCGGAGGGAGTGAGTTTTATCG
CGTATTGTTATTTTATGTCGGCGTTTATTTATTTTATGGTAGATTAGGG

Bisulfite specific primers for PITX3:

Forward primer (SEQ ID NO: 55): TAGGTAGGGGTTAGGGGT

Reverse primer (SEQ ID NO 56): CCCTAATCTACCATAAAATAAATAAA

Results:

LHX3:

Results of the sequencing are shown in the box plots of Fig. 2 and the sequencing matrix of Fig. 3 of LHX3 (amplicon A, Fig. 1) show that the methylation of the four sample groups (A, B, C and D) are strong correlated to the methylation of PITX2, namely low methylation of groups A and D and high methylation of groups B and C. This shows that the methylation of LHX3 is a prognostic biomarker for clinical outcome after breast cancer surgery, similar to PITX2.

Figure 2 shows the quantified levels of methylation within LHX3 amplicon A of the four sample groups as measured in Example 3. Percentage methylation is shown in the Y-axis. Group B and C show high LHX3 methylation whereas groups A and D show low methylation. This is in strong correlation to the PITX2 methylation of the same samples.

Figure 4 shows the correlation of methylation measured at a second region, amplicon B as shown in figure 1 within the LHX3 gene show strong co-methylation to amplicon LHX3 A (Fig 5). This indicates that this region (B) is a prognostic biomarker as well as region A.

PITX3:

In contrast to the methylation of PITX2, the methylation of PITX3 is significantly lower in sample group B. This sample group shows high PITX2 methylation (data not shown) although no recurrence occurred resulting in a wrong classification. Analysis of PITX3 therefore significantly lowers the number of patients wrongly classified.

Figure 5 shows the quantified levels of methylation within PITX3 amplicon A of the four sample groups. Percentage methylation is shown in the Y-axis. Group B and C show high

PITX3 methylation whereas groups A and D show low methylation. This is in strong correlation to the PITX2 methylation of the same samples.

Figure 6 shows an alternative view of the quantified levels of PITX3 methylation. Each row of the matrix represents a single CpG site within the fragment and each column represents an individual DNA sample. The bar on the left shows a scale of the percent methylation, with the degree of methylation represented by the shade of each position within the column from black representing 100% methylation to light gray representing 0% methylation. White positions represented a measurement for which no data was available.

E30809PCT
Epigenomics AG

Claims

1. A method for providing a prognosis and/or predicting the outcome of treatment of a subject with a cell proliferative disorder comprising
 - a) obtaining a biological sample from the subject,
 - b) determining the expression of at least one gene and/or genomic sequence selected from the group consisting of LHX3 and PITX3 and/or regulatory sequences thereof within said sample, and
 - c) determining therefrom the prognosis and/or outcome of treatment of said subject.
2. The method according to claim 1 further comprising
 - d) determining a suitable treatment regimen for the subject.
3. The method according to claim 2, wherein said suitable treatment regimen comprises one or more therapies selected from the group consisting of chemotherapy, radiotherapy, surgery, biological therapy, immunotherapy, antibodies, molecularly targeted drugs, estrogen receptor modulators, estrogen receptor down-regulators, aromatase inhibitors, ovarian ablation, LHRH analogues and other centrally acting drugs influencing estrogen production.
4. The method according to any of claims 1 to 3, wherein said cell proliferative disorder is selected from the group consisting breast cancer, ovarian cancer, transitional cell bladder cancer, bronchogenic lung cancer, thyroid cancer, pancreatic cancer, prostate cancer, uterine cancer, testicular cancer, gastric cancer, soft tissue and osteogenic sarcomas, neuroblastoma, Wilms' tumor, malignant lymphoma (Hodgkin's and non-Hodgkin's), acute myeloblastic leukemia, acute lymphoblastic leukemia, Kaposi's sarcoma, Ewing's tumor, refractory multiple myeloma, squamous cell carcinomas of the head, neck, cervix, and vagina.
5. The method according to any of claims 1 to 4 wherein said expression is determined by analysis of at least one of mRNA expression, LOH, protein expression.
6. The method according to any of claims 1 to 4 wherein said expression is determined by determining the methylation status of one or more CpG positions within said genes and/or regulatory regions thereof.

7. A method for providing a prognosis and/or predicting the outcome of treatment of a subject with a cell proliferative disorder comprising,:

- a) extracting or otherwise isolating genomic DNA from a biological sample obtained from the subject,
- b) treating the genomic DNA of a), or a fragment thereof, with one or more reagents to convert cytosine bases that are unmethylated in the 5-position thereof to uracil or to another base that is detectably dissimilar to cytosine in terms of hybridization properties;
- c) contacting the treated genomic DNA, or the treated fragment thereof, with an amplification enzyme and at least one primer comprising, a contiguous sequence of at least 9 nucleotides that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16, and complements thereof, wherein the treated genomic DNA or the fragment thereof is either amplified to produce at least one amplificate, or is not amplified; and
- d) determining, based on the presence or absence of, or on the quantity or on a property of said amplificate, the methylation state of at least one CpG dinucleotide sequence of at least one gene and/or genomic sequence selected from the group consisting of LHX3 and PITX3, or an average, or a value reflecting an average methylation state of a plurality of CpG dinucleotide sequences of at least one gene and/or genomic sequence selected from the group consisting of LHX3 and PITX3, and
- e) determining from said methylation state the prognosis and/or outcome of treatment of said subject.

8. The method of claim 7, wherein treating the genomic DNA, or the fragment thereof in b), comprises use of a reagent selected from the group comprising of bisulfite, hydrogen sulfite, disulfite, and combinations thereof.

9. The method of claim 7, wherein contacting or amplifying in c) comprises use of at least one method selected from the group comprising: use of a heat-resistant DNA polymerase as the amplification enzyme; use of a polymerase lacking 5'-3' exonuclease activity; use of a polymerase chain reaction (PCR); generation of an amplificate nucleic acid molecule carrying a detectable label.

10. The method of any of claims 1 to 9, wherein the biological sample obtained from the subject is selected from the group comprising cell lines, histological slides, paraffin embedded tissues, biopsies, tissue embedded in paraffin, bodily fluids, nipple aspirate and blood and combinations thereof.
11. The method of claim 8, further comprising in step d) the use of at least one nucleic acid molecule or peptide nucleic acid molecule comprising in each case a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16, and complements thereof, wherein said nucleic acid molecule or peptide nucleic acid molecule suppresses amplification of the nucleic acid to which it is hybridized.
12. The method of claim 8, wherein determining in d) comprises hybridization of at least one nucleic acid molecule or peptide nucleic acid molecule in each case comprising a contiguous sequence at least 9 nucleotides in length that is complementary to, or hybridizes under moderately stringent or stringent conditions to a sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16, and complements thereof.
13. The method of claim 12, wherein at least one such hybridizing nucleic acid molecule or peptide nucleic acid molecule is bound to a solid phase.
14. The method of claim 12, further comprising extending at least one such hybridized nucleic acid molecule by at least one nucleotide base.
15. The method of claim 8, wherein determining in d), comprises sequencing of the amplificate.
16. The method of claim 8, wherein contacting or amplifying in c), comprises use of methylation-specific primers.
17. A method for providing a prognosis and/or predicting the outcome of treatment of a subject with a cell proliferative disorder comprising,:

- a) extracting or otherwise isolating genomic DNA from a biological sample obtained from the subject
- b) digesting the genomic DNA of a), or a fragment thereof, with one or more methylation sensitive restriction enzymes;
- c) contacting the DNA restriction enzyme digest of b), with an amplification enzyme and at least two primers suitable for the amplification of a sequence comprising at least one CpG dinucleotide of a sequence selected from SEQ ID NO: 1 and SEQ ID NO: 2,
- d) determining, based on a presence or absence of an amplificate the methylation state or level of at least one CpG dinucleotide of a sequence selected from the group consisting SEQ ID NO: 1 and SEQ ID NO: 2, whereby at least one of detecting and classifying cellular proliferative disorders is, at least in part, afforded.

18. The method according to claim 17 wherein the presence or absence of an amplificate is determined by means of hybridization to at least one nucleic acid or peptide nucleic acid which is identical, complementary, or hybridizes under stringent or highly stringent conditions to an at least 16 base long segment of a sequence selected from SEQ ID NO: 1 and SEQ ID NO: 2.

19. A nucleic acid, comprising at least 18 contiguous nucleotides of a treated genomic DNA sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16, and sequences complementary thereto, wherein the treatment is suitable to convert at least one unmethylated cytosine base of the genomic DNA sequence to uracil or another base that is detectably dissimilar to cytosine in terms of hybridization.

20. A nucleic acid, comprising at least 50 contiguous nucleotides of a DNA sequence selected from the group consisting of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16, and sequences complementary thereto.

21. A kit suitable for performing the method according to claim 1 comprising (a) a bisulfite reagent; (b) a container suitable for containing the said bisulfite reagent and the biological sample of the patient; (c) at least one set of oligonucleotides containing two oligonucleotides whose sequences in each case are identical, are complementary, or hybridize under stringent or highly stringent conditions to a 9 or more preferably 18 base long segment

of a sequence selected from SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16.

22. A composition comprising:

- a) a nucleic acid comprising a sequence at least 18 bases in length of a segment of the chemically pretreated genomic DNA according to one of the sequences taken from the group comprising of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and sequences complementary thereto, and
- b) a buffer comprising at least one of the following substances: magnesium chloride, dNTP, of taq polymerase, an oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which is complementary to, or hybridizes under moderately stringent or stringent conditions to a pre-treated genomic DNA according to one of the of SEQ ID NO: 5 to SEQ ID NO: 8 and SEQ ID NO: 13 to SEQ ID NO: 16 and sequences complementary thereto.

23. The use of a method according to claims 1 to 18, a nucleic acid according to claims 19 or 20, a kit according to claim 21 or a composition of matter according to claim 22 in providing a prognosis and/or predicting the outcome of treatment of a subject with a cell proliferative disorder.

Figure 1

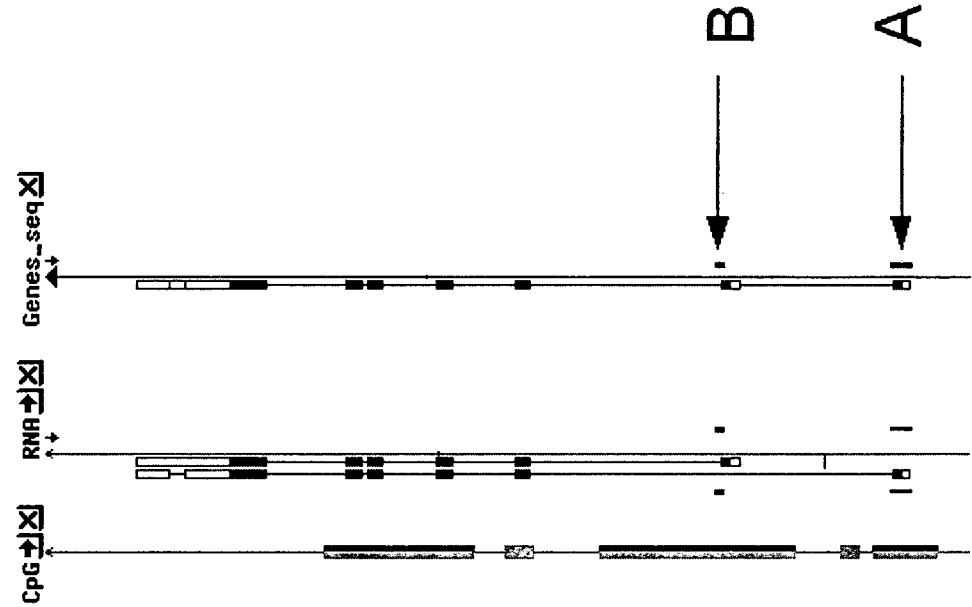


Figure 2

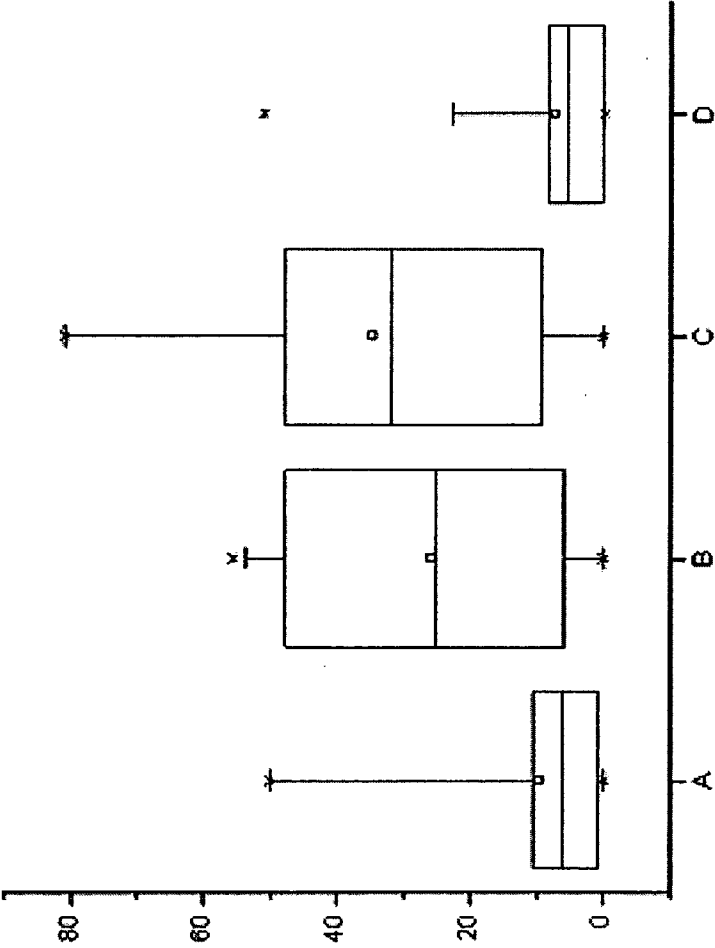


Figure 3

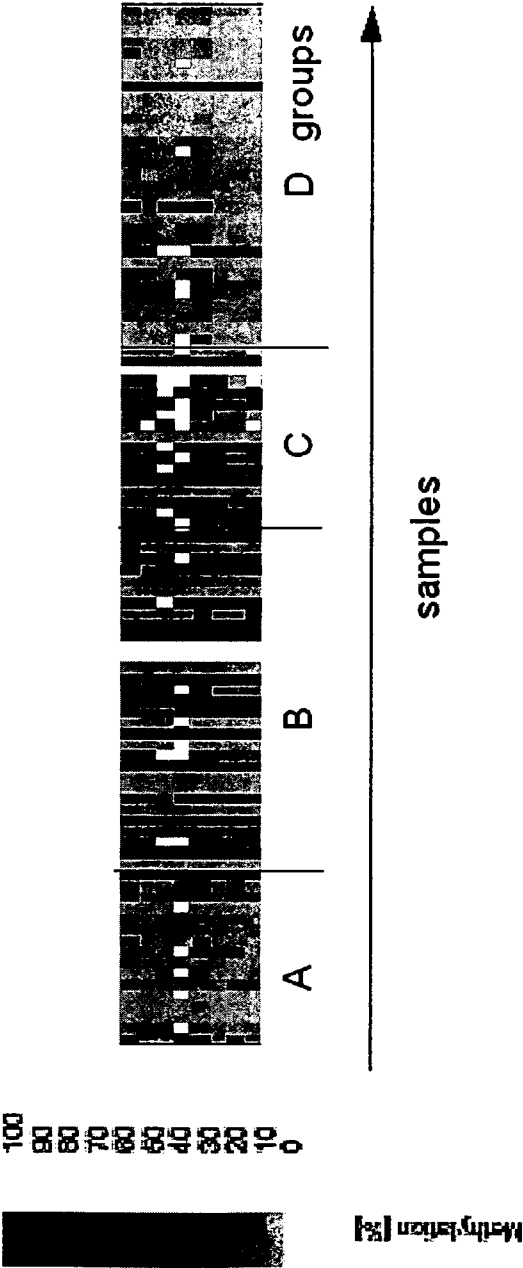


Figure 4

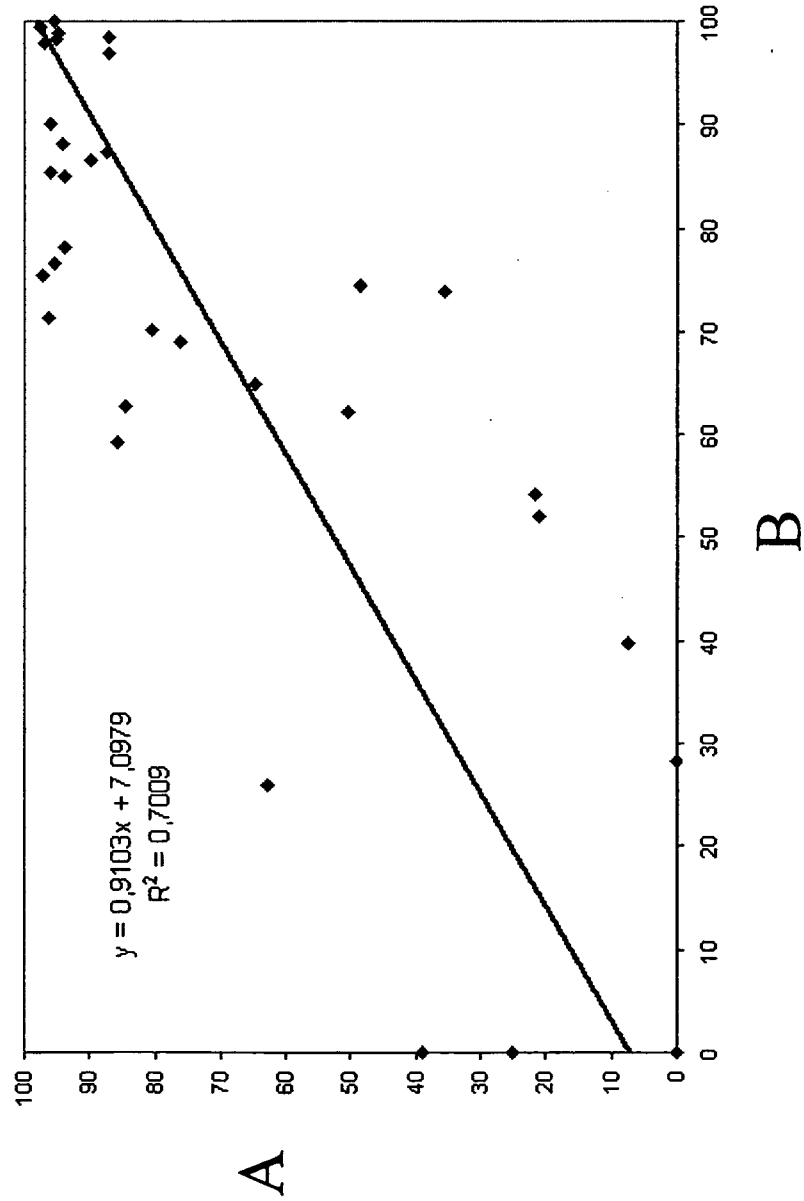


Figure 5

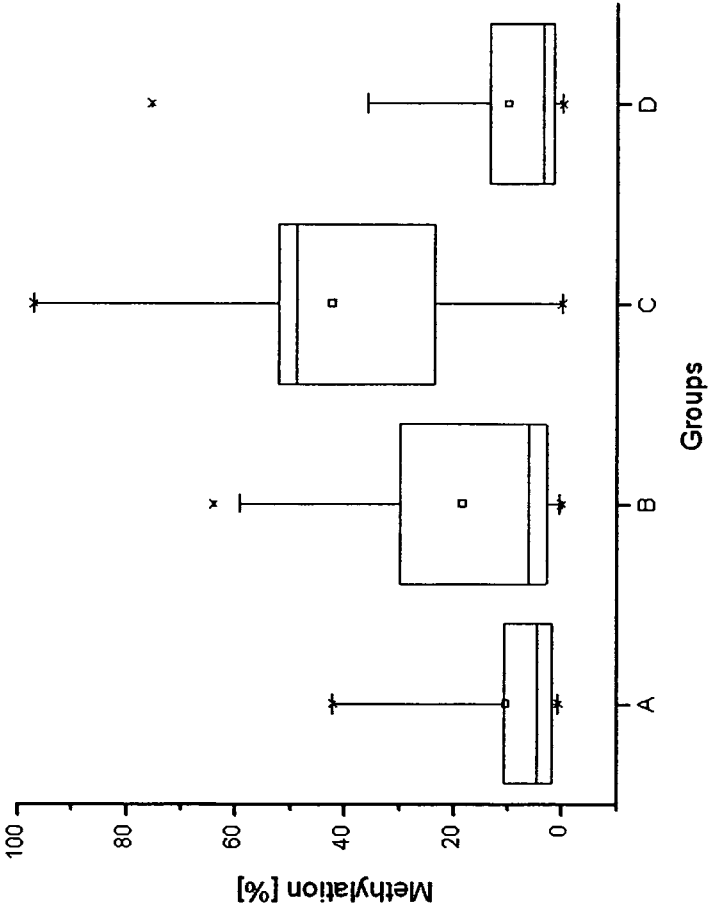


Figure 6

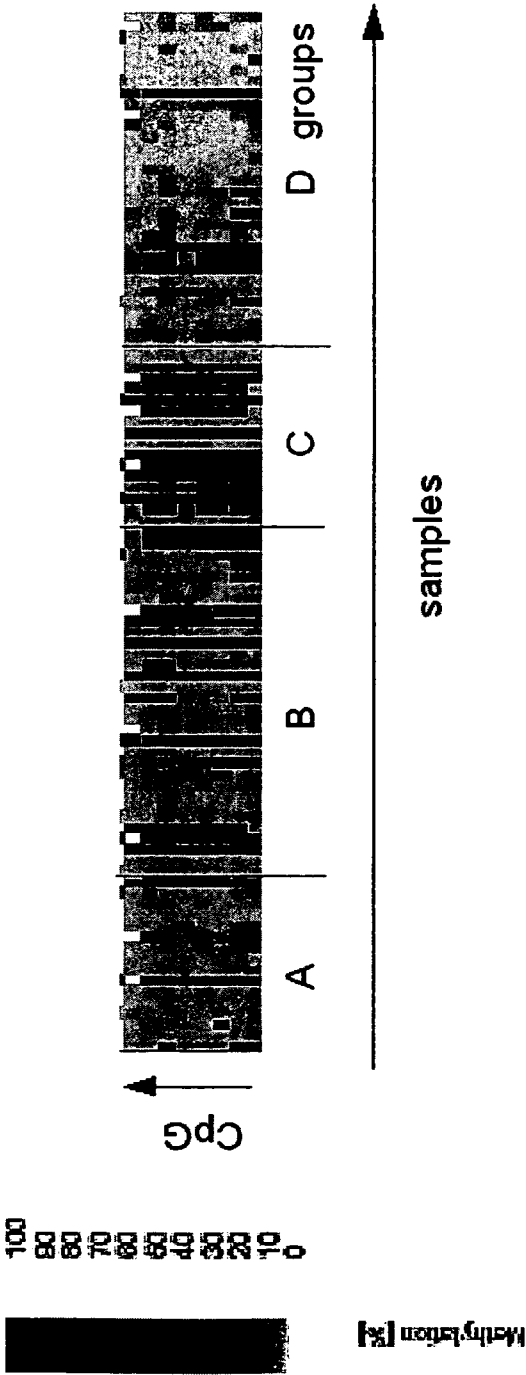


Figure 7

