The instant application provides methods and related compositions pertaining to the identification of chemo sensitivity in a patient. In a particular embodiment, the invention provides biomarkers for the identification of chemo sensitivity in a breast cancer patient, wherein a reduced expression of an X-linked gene, such as XIST, in the breast cancer cells of the patient indicates that the breast cancer cells in the patient may be successfully treated with a platinum-based compound, such as cisplatin.
TITLE OF THE INVENTION
METHODS AND COMPOSITIONS FOR PREDICTING CHEMOTHERAPY SENSITIVITY

CROSS-REFERENCE TO RELATED APPLICATIONS
This application claims the benefit of priority of U.S. Provisional Application Serial No. 61/428,674, filed December 30, 2010 and U.S. Provisional Application Serial No. 61/429,364, filed January 3, 2011, which are incorporated herein by reference in their entirety.

FIELD OF THE INVENTION
The invention relates to the field of methods and related compositions for predicting chemotherapy sensitivity. In certain embodiments, the invention relates to the field of methods and related compositions for predicting chemotherapy sensitivity in a cancer patient by detecting a reduced expression level of an X-linked gene, such as XIST, in one or more cancer cells of the patient.

BACKGROUND OF THE INVENTION
Lack of markers that predict chemotherapy response is a handicap in cancer treatment. Genome-wide gene expression analyses of human cell lines and breast cancer samples have not produced markers that predict sensitivity to cytotoxic chemotherapy.

Most forms of cytotoxic cancer chemotherapy also hit normal tissues. This is acceptable when the tumor responds, but frustrating when the tumor is intrinsically resistant and the patient only suffers from the side effects of an unsuccessful treatment. A major goal of molecular oncology is therefore to identify biomarkers that predict the response of tumors before treatment is started. Such predictive markers have been found for some targeted therapies in which the target and its interaction with drugs are well defined. Examples are the presence of a functional estrogen receptor predicting response to anti-estrogens (McGuire, 1980; Early Breast Cancer Trialists' Collaborative Group, 1998); the requirement of substantial HER2 levels for a therapeutic effect of trastuzumab (Slamon et al., 2001); defects in homologous DNA recombination making the tumor susceptible to inhibitors of Poly (ADP-ribose) polymerase (Tutt et al, 2010; Fong et al, 2010); and several examples in which alterations in drug targets or oncogenes prevent drug action (Gorre et al, 2001; Turke et al, 2010; Heinrich et al, 2003; Berns et al,
For classical cytotoxic chemotherapy with DNA damaging drugs or antimitotics, however, predictive biomarkers have been harder to find.

In an attempt to find new biomarkers many investigators have turned to the analysis of genome-wide gene expression profiles. These profiles have been successful for predicting prognosis, i.e., whether patients will require adjuvant chemotherapy after tumor removal (Van 't Veer et al., 2002; van de Vijver et al., 2002). Prognostic and predictive biomarkers are fundamentally different, however (Borst and Wessels, 2010). To detect predictive markers, considerable effort and money has been invested in the analysis of human breast cancer samples (Nuyten and van de Vijver, 2008). In particular the neoadjuvant setting seemed attractive to correlate gene expression profiles with therapy outcome. No clear response profile was obtained, however (Hannemann et al., 2005). Other studies have gathered a number of unrelated signatures (Borst and Wessels, 2010). These profiles either still await validation in an independent study, or the sensitivity and specificity was inadequate for clinical decision making.

Predictive gene expression signature profiles for individual chemotherapy drugs have been reported by Potti et al. (2006). These were derived by screening the NCI-60 panel of diverse human cancer cell lines (Potti et al., 2006; Bonnefoi et al., 2007; Hsu et al, 2007). However, these results are controversial, as they are based on flawed data and are not reproducible (Coombes et al, 2007; Baggerly and Coombes, 2009; Borst and Wessels, 2010; Liedtke et al, 2010).

Progress in defining useful biomarkers using human tumor material has been limited.

Citation or identification of any document in this application is not an admission that such document is available as prior art to the present invention.

**SUMMARY OF THE INVENTION**

The instant invention provides a method of evaluating chemotherapy sensitivity in a patient in need thereof, comprising (a) measuring expression levels of one or more X-linked genes in the patient; and (b) comparing the expression levels of the one or more X-linked genes in (a) with the expression levels of one or more reference X-linked genes, wherein the one or more reference X-linked genes are from a control sample, wherein a reduction in the expression of the one or more X-linked genes in comparison to the one or more reference X-linked genes is indicative of chemotherapy sensitivity in the patient.
In certain embodiments, the patient has breast cancer, ovarian cancer, cervical cancer, testicular cancer, head and neck cancer, lung cancer, colorectal cancer, or bladder cancer.

In certain embodiments, the chemotherapy sensitivity is sensitivity to treatment with a platinum-based chemotherapeutic agent. Examples of platinum-based chemotherapeutic agents include cisplatin, carboplatin, oxaliplatin, satraplatin, picoplatin, and triplatin tetranitrate.

In certain embodiments, the chemotherapy sensitivity is sensitivity to treatment with a PARP inhibitor. Examples of PARP inhibitors include ABT-888, olaparib, CEP-9722, AGO14699, BSI-201, MK4827, LT-673, GPI21016, and INO-1001.

In some embodiments, the X-linked gene is XIST. In certain embodiments, the expression levels of the XIST gene in (a) are measured in a tumor sample from the patient. In certain embodiments, the patient tumor sample is a breast tumor sample.

In other embodiments, the X-linked gene is selected from the group consisting of: Utx, Jarid1c, and Eif2s3x.

In some embodiments, the X-linked gene escapes XIST-mediated inactivation. In certain embodiments, the expression levels of the X-linked gene in (a) are measured in a tumor sample from the patient. In certain embodiments, the patient tumor sample is a breast tumor sample.

In certain embodiments, the method provides that further to a reduction in the one or more X-linked gene expression levels, there is a reduction in the expression level of one or more genes selected from the group consisting of: ATM, Rad50, and Ube3b. In certain embodiments, the reduction in the expression level of ATM, Rad50, and/or Ube3b is measured by determining a reduction in the amount of ATM, Rad50, and/or Ube3b protein. In other embodiments, the reduction in the expression level of ATM, Rad50, and/or Ube3b is measured by determining a reduction in the activity of ATM, Rad50, and/or Ube3b protein. In some embodiments, the reduction in the expression level of ATM, Rad50, and/or Ube3b is due to one or more inactivating mutations in the ATM, Rad50, and/or Ube3b gene. In other embodiments, the reduction in the expression level of ATM, Rad50, and/or Ube3b is due to protein dysfunction of the ATM, Rad50, and/or Ube3b protein.

In some embodiments, patient has breast cancer. In certain embodiments, the breast cancer is HER-2 negative, advanced breast cancer. In some embodiments, the
cancer is BRCA-1 deficient. In other embodiments, the cancer is non-BRCA-1 deficient. In certain further embodiments, the cancer is breast cancer or ovarian cancer.

In some embodiments, the method further comprises conducting comparative genomic hybridization (CGH) with DNA from the patient. In certain embodiments, the patient has breast cancer. In further embodiments, the CGH is a BRCA-CGH. In a particular embodiment, the BRCA-CGH is BRCA-1-CGH. In some embodiments, the cancer is BRCA-1 deficient. In further embodiments, the X-linked gene is XIST. In yet further embodiments, the patient is selected for treatment with high-dose alkylating therapy.

In certain embodiments, the chemotherapeutic agent is a DNA-cross-linking agent.

In some embodiments, the level of expression of said genes is measured by determination of their level of transcription, using a DNA array. In other embodiments, the level of expression of said genes is measured by determination of their level of transcription, using quantitative RT-PCR. In yet other embodiments, the level of expression of said genes is measured in a tumor sample from the patient. In certain embodiments, the tumor sample is a breast tumor sample. In some embodiments, the level of expression of said genes is measured by determination of their level of translation. In certain embodiments, the level of translation is measured by determination of the amount of translated protein. In other embodiments, the level of translation is measured by determination of the activity of translated protein. In certain embodiments, the translated protein is selected from the group consisting of: Utx, Jaridl2, and Eif2s3x. In some embodiments, a reduction in the expression of the one or more X-linked genes is due to one or more inactivating mutations in the X-linked gene.

In yet other embodiments, the instant invention provides a microarray comprising a plurality of polynucleotide probes each complementary and hybridizable to a sequence in a different gene that is an X-linked gene that is a marker for chemosensitivity in a patient that has cancer.

In certain embodiments, the plurality of probes is at least 70 %, at least 80 %, at least 90 %, at least 95 %, or at least 98 % of the probes on the microarray.

In certain embodiments, the X-linked gene that is a marker for chemosensitivity is selected from the group consisting of XIST, Utx, Jaridl2, and Eif2s3x.

In some embodiments, the microarray further comprises a plurality of polynucleotide probes each complementary and hybridizable to a sequence in a different
gene that is a marker for chemosensitivity in a patient that has cancer, wherein the gene is not an X-linked gene. In certain embodiments, the gene that is not an X-linked gene is selected from the group consisting of: ATM, Rad50, and Ube3b.

In yet other embodiments, the invention provides a kit, comprising at least one pair of primers specific for an X-linked gene that is a marker for chemosensitivity in a patient that has cancer, at least one reagent for amplification of the X-linked gene, and instructions for use.

In some embodiments, the primers are specific for an X-linked gene selected from the group consisting of XIST, Utx, Jaridlc, and Eif2s3x. In certain embodiments, the primers are specific for XIST.

In certain embodiments, the marker for chemosensitivity is a marker for sensitivity to a platinum-based drug. Examples of platinum-based drugs include cisplatin, carboplatin, oxaliplatin, satraplatin, picoplatin, and triplatin tetranitrate.

In certain embodiments, the marker for chemosensitivity is a marker for sensitivity to a PARP inhibitor. Examples of PARP inhibitors include ABT-888, olaparib, CEP-9722, AGO14699, BSI-201, MK4827, LT-673, GPI21016, and INO-1001.

In certain embodiments, the kit is a PCR kit. In other embodiments, the kit is an MLPA kit. In yet other embodiments, the kit is an RT-MLPA kit.

It is noted that in this disclosure and particularly in the claims, terms such as "comprises", "comprised", "comprising" and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean "includes", "included", "including", and the like; and that terms such as "consisting essentially of" and "consists essentially of" have the meaning ascribed to them in U.S. Patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention.

These and other embodiments are disclosed or are obvious from and encompassed by, the following Detailed Description.

**BRIEF DESCRIPTION OF THE DRAWINGS**

FIGURE 1 depicts the reproducibility of docetaxel or cisplatin sensitivity of individual KBIP tumors using orthotopic transplantation. Small tumor fragments of spontaneous mammary tumors were transplanted orthotopically into syngeneic wild-type
female mice as shown in (A) for docetaxel and (B) for cisplatin. After a mean latency of about 4 weeks, when tumors reached a size of 150-250mm3 \( (V = \text{length} \times \text{width}^2 \times 0.5) \), animals were treated with 25 mg of docetaxel per kg i.v. on days 0, 7 and 14 (A, filled squares) or 6 mg cisplatin per kg i.v. (B, filled squares). Treatment of tumors was resumed once the tumor relapsed to its original size (100%). See also Figure 2 and Figure 3.

FIGURE 2 depicts the response of BRCA1; p53-deficient tumors to the maximum tolerable dose of docetaxel, doxorubicin or cisplatin. A, 6 individual K14cre,BrcalF/F,p53F/F animals carrying a spontaneous mammary tumor of -200 mm3 (T1-T6) were treated with 25 mg docetaxel per kg mouse i.v. on days 0, 7 and 14 (arrows). When tumors relapsed back to -200 mm3 or showed progressive growth (tumor size \( \geq 50\% \)) after a recovery time of 7 days following the day 14 injection, treatment was resumed as indicated by the filled squares. Animals carrying T2 or T4 had to be sacrificed before full docetaxel resistance developed due to the presence of a squamous cell carcinoma of the lip (T2) or ear (T4). B, animals with 37 individual orthotopically transplanted BRCA1;p5 3-deficient mammary tumors (volume -200 mm3) were left untreated (black line) or treated with 25 mg docetaxel per kg i.v. on days 0, 7 and 14 (gray line), 5 mg doxorubicin per kg i.v. on day 0 (dark gray line), or 6 mg cisplatin per kg i.v. on day 0 (light gray line). When tumors relapsed or showed progressive growth (tumor size \( \geq 50\% \)) after a recovery time of 7 days (docetaxel, doxorubicin) or 14 days (cisplatin), treatment was resumed as indicated by the filled squares. The responses of T1, T3, T7, T10, T11 and T14 have already been published in Rottenberg et al. PNAS 104:12117-22, 2007.

FIGURE 3 is a table depicting an overview of treatment responses of KBP1 tumors.

FIGURE 4 depicts analysis of intrinsic docetaxel resistance of KB1P tumors using gene expression profiling of samples before treatment. (A) 43 individual tumors were treated with 25mg of docetaxel per kg i.v. on days 0, 7 and 14. The time until relapse is the time after the last treatment on day 14 required for the tumors to grow back to 100\%. Tumors that did not shrink by at least 50\% were scored as 0. (B) Significance analysis of microarrays (SAM) of good vs. poor docetaxel responders \( (\Delta=0.8) \) using the MEEBO or Illumina platform. (C) The average of median-normalized cycle threshold
(CT) values determined by quantitative TLDA of 46 genes encoding ABC proteins is shown for good versus poor docetaxel responders. For Abcal4, Abcal5, Abcal6, Abcal7, Abcb5, Abcbl1 and Abcg8 no expression was detected in the tumors.

FIGURE 5 depicts correlation of gene expression with the response to platinum drugs. (A) 35 mice with KB1P tumors were treated with 6mg of cisplatin per kg i.v. on day 0. The time until relapse is the time required for the tumors to grow back to 100%. (B) SAM of highly vs. moderately cisplatin-sensitive KB1P tumors using the MEEBO (Δ=1.5; FDR=0) or Illumina (Δ=0.85; FDR=0) platform. (C) Kaplan Meier survival curves according to XIST gene expression levels of patients who had been randomized between conventional (CONV, gray) and intensive platinum-based chemotherapy (IPB, black). P values were calculated using the logrank test. See also Figure 6 and Figure 7.

FIGURE 6 depicts gene expression analysis of mouse or human Xist/XIST. (A) Heatmap of the top 50 genes that correlate with the Xist expression pattern using the Illumina platform (r > 0.53). (B) Quantification of the expression of the human Xist gene using two independent probes.

RT-MLPA analyses of a pool of RNA isolated from FFPE normal breast tissue of 8 different patients. The XIST gene expression levels were normalized to the internal reference genes present in the RT-MLPA mix (LDHA, 2 probes for GAPDH, B2M, ARHGDIA, FAU, OAZ1 and BIRC2). Depicted is the average gene expression of 8 independent measurements (one measurement is the average of 3 independent MLPA reactions); the error bars indicate the standard deviation of these 8 measurements; the dashed line represent two times the standard deviation and was used as the cut-off for further analyses. (C) Association of XIST gene expression with BRCA1 status. Tumors of patients present in Applicants’ study were previously screened for the presence of a BRCA1 mutation using a method that identifies about 73% of the types of BRCA1 -mutations prevalent in the Netherlands (Vollebergh et al.; Annals of Oncology, in press). BRCA1 -mutated tumors and non-mutated tumors had equally low and high expressions of XIST when analyzed with XIST probe exon 2-3 and XIST probe exon 4-5. BRCA1 promoter methylation data acquired by MS-MLPA analyses were also available (Vollebergh et al.; Annals of Oncology, in press). BRCA1 -methylated tumors were highly associated with low XIST expression. Using a CGH-classifier derived from BRCA1-mutated tumors (Vollebergh et al; Annals of Oncology, in press) tumors could
be classified as BRCAl-likeCGH or Non-BRCAl-likeCGH based on their specific copy number aberrations present in the tumors. Tumors with a BRCAl-like CGH tumor frequently had a low XIST expression (26/32 BRCAl-likeCGH tumors versus 13/28 Non-BRCAl-likeCGH tumor for the exon 2-3 probe; respectively 25/32 versus 17/28 for the exon 4-5 probe).

FIGURE 7 is a table of (A) sequences of oligonucleotides used for RT-MLPA reactions to quantify human XIST gene expression; (B) patient characteristics distributed by XIST expression; and (C) multivariate Cox proportional-hazard analysis of the risk of recurrence (RFS) and XIST expression.

FIGURE 8 depicts distribution of gene expression levels of untreated KB IP mouse tumors using the probes for Xist or Abcblb present on the MEEBO and Illumina platforms. P values were determined by the Kruskal-Wallis and post-hoc Mann-Whitney U tests (Xist probes, highest P value shown) or the Mann-Whitney U test only (Abcblb).

FIGURE 9 depicts quantification of the expression of the human ATM and RAD50 genes using RT-MLPA analyses of a pool of RNA isolated from FFPE normal breast tissue of 8 different patients. ATM gene expression levels are depicted. These gene expression levels were normalized to the internal reference genes present in the RT-MLPA mix (LDHA, 2 probes for GAPDH, B2M, ARHGDIA, FAU, OAZ1 and BIRC2). Depicted is the average gene expression of 8 independent measurements (one measurement is the average of 3 independent MLPA reactions); the error bars indicate the standard deviation of these 8 measurements; the dashed line represents two times the standard deviation and was used as the cut-off for further analyses.

FIGURE 10 depicts the association of ATM gene expression with recurrence-free survival (RFS) after intensified PB-chemotherapy and conventional chemotherapy. Kaplan Meier survival curves according to ATM gene expression levels of patients who had been randomized between conventional (gray) and intensive platinum-based chemotherapy (black). P values were calculated using the logrank test.

FIGURE 11 Left Panel is a graph depicting survival of patients with low $X_{\text{ist}}$ expression on high dose (gray) versus conventional (black) chemotherapy. Right panel is a graph depicting the combination of BRCAl-like CGH patients with low $X_{\text{ist}}$ expression.
FIGURE 12 Left Panel is a graph depicting survival of patients with a BRCA-like CGH tumor on high dose (gray) versus conventional (black) chemotherapy. Right Panel is a graph depicting patients with a BRCA-like CGH tumor and low Xist expression.

FIGURE 13 depicts the recurrence free survival data of the investigated population. These measurements are sorted by the respective marker population (indicated in the table underneath), survival time and treatment. Dark gray (DG) bars represent conventional treatment, light gray (LG) bars represent patients that received high dose. Crosses mark censored patients. The gradient is percentage of the sum of the levels of prognostic factors divided by the maximum sum of the levels of these factors. The deeper the shade of dark gray or the deeper the shade of light gray, the worse the prognosis.

DETAILED DESCRIPTION

The instant invention provides methods and related compositions pertaining to the identification of a tumor that will be sensitive to treatment by a certain compound or class of compounds. In certain embodiments, the invention provides one or more markers for chemosensitivity in a patient. In some embodiments, the marker is an X-linked gene. In further embodiments, the X-linked gene is one that is known to escape X inactivation (Xi). In a particular embodiment, the X-linked gene is XIST. Examples of other suitable X-linked genes that may serve as a marker for chemosensitivity in a patient as described herein include Utx (ubiquitously transcribed tetratricopeptide repeat, X chromosome), Jaridlc (Jumonji, AT rich interactive domain 1C), and Eif2s3x (eukaryotic translation initiation factor 2, subunit 3, structural gene X-linked).

In some embodiments, the invention provides methods whereby measurement of reduced expression of an X-linked gene, such as XIST, in one or more cancer cells of a patient identifies these cancer cells as cells that may be sensitive to treatment by one or more DNA alkylating/cross-linking agents, such as platinum-based compounds, cyclophosphamide, melphalan, thiotepa, mitomycin C, and nimustine.

In certain embodiments, identification of a reduced expression of an X-linked gene, such as XIST, in one or more cancer cells of a patient is indicative that the one or more cancer cells will be sensitive to treatment by a compound or class of compounds, such as one or more platinum-based compounds. Examples of platinum-based
compounds that cells expressing a reduced level of an X-linked gene, such as XIST, may be sensitive to include cisplatin, carboplatin, oxaliplatin, satraplatin, picoplatin, and triplatin tetrinitrate.

In yet other embodiments, the invention provides methods whereby measurement of reduced expression of an X-linked gene, such as XIST, in one or more cancer cells of a patient identifies these cancer cells as cells that may be sensitive to treatment by one or more PARP (poly (ADP-ribose) polymerase) inhibitors. Examples of PARP inhibitors include ABT-888, olaparib (AZD-2281), CEP-9722, AG014699, BSI-201, MK4827, LT-673, GPI21016, and INO-1001.

In certain embodiments, the methods and compositions of the invention identify one or more cancer cells that may be sensitive to treatment by one or more DNA cross-linking agents.

In certain embodiments, the prognostic methods and compositions of the instant invention predict chemosensitivity to a combination of at least two chemotherapeutic agents, wherein the at least two chemotherapeutic agents are administered at the same time and/or sequentially. In further embodiments, the invention provides methods wherein a measurement of reduced expression of an X-linked gene, such as XIST, in or one or more cancer cells of a tumor of a patient identifies the tumor as one that may be sensitive to treatment by a combination of at least one platinum-based drug and at least one PARP inhibitor.

In some embodiments, the markers of the instant invention enable the detection of chemosensitivity in a patient in combination with one or more known markers of hypersensitivity to a chemotherapeutic agent or class of agents. For example, in certain embodiments, the invention provides a method of detecting chemosensitivity in a patient, comprising measuring the expression level of an X-linked gene, such as XIST, Utx, Jaridlc, and/or Eif2s3x, in combination with the expression level of a known marker for sensitivity to one or more chemotherapeutic agents, such as cisplatin hypersensitivity marker ATM (Ataxia telangiectasia mutated), Rad50, or Ube3b (ubiquitin protein ligase E3B).

In other embodiments, expression levels of one or more X-linked genes (e.g., XIST, Utx, Jaridlc, and/or Eif2s3x) are measured in one or more cancer cells of a patient in combination with an array profile, such as a BRCA1-CGH (comparative genomic hybridization) array analysis. For example, in certain embodiments, XIST levels will be detected in a patient that is determined to have a BRCA-deficient tumor based on the use
of a CGH classifier. If XIST levels are low, the patient will be selected for treatment with high-dose alkylating therapy. In other embodiments, if XIST levels are high, the patient with the BRCA-deficient tumor will not be selected for high dose alkylating therapy. In yet other embodiments, a patient is determined to be BRCA-proficient based on the use of a CGH classifier. In certain further embodiments, if XIST levels are low, the patient will be selected for treatment with high-dose alkylating therapy. For examples of BRCA-CGH arrays, see e.g., International Application Publication Nos. WO 201 1/048499 and WO 201 1/048495, which are incorporated herein by reference. "Comparative genomic hybridization" refers generally to molecular-cytogenetic techniques for the analysis of copy number changes, gains and/or losses, in the DNA content of a given subject's DNA. A BRCA-deficient tumor refers to a tumor comprising one or more cells containing a mutation of the BRCA1 and/or BRCA2 locus or a deficiency in the homologous recombination-dependent double strand break DNA repair pathway that alters BRCA1 and/or BRCA2 activity or function, either directly or indirectly.

In certain embodiments, the invention provides methods and compositions for identifying a cancer patient who would benefit from a certain chemotherapeutic treatment. For example, an aspect of the invention is a method of screening cancer patients to determine those cancer patients more likely to benefit from a particular chemotherapy, such as platinum-based chemotherapy, comprising obtaining a sample of genetic material from a tumor of the patient, and assaying for the presence of a genotype in the patient that is associated with sensitivity to the particular chemotherapy, such as a genotype characterized by a reduced expression in one or more X-linked genes. In some embodiments, the genotype is further characterized by a reduced expression in one or more known markers for chemotherapeutic sensitivity, such as, in the case of platinum-based chemotherapy, ATM, Rad50, and/or Ube3b.

In a particular embodiment, the instant invention provides methods and compositions for the identification of a breast cancer patient who would benefit from platinum-based chemotherapy (e.g., the patient will be recurrence-free for a period of time greater than a patient undergoing conventional chemotherapy). In further embodiments, the breast cancer patient that receives platinum-based chemotherapy exhibits an increased recurrence-free survival (RFS) in a 5-year period compared to breast cancer patients who receive conventional chemotherapy. In yet other embodiments, the instant invention provides methods and compositions for the identification of an ovarian cancer patient who would benefit from platinum-based
chemotherapy (e.g., the patient will be recurrence-free for a period of time greater than a patient undergoing conventional chemotherapy). In preferred embodiments, the methods of the instant invention predict whether a chemotherapeutic agent or other compound is likely to be cytotoxic to one or more cancer cells.

In certain embodiments, cancers for which the prognostic methods and compositions of the instant invention provide predictive results for chemosensitivity include cancers such as breast cancer (e.g., BRCA-1 deficient, stage-III HER2-negative), ovarian cancer (e.g., BRCA-1 deficient, epithelial ovarian cancer), testicular cancer (e.g., germ cell tumors, sex cord/gonadal stromal tumors, and miscellaneous non-specific stromal tumors), lung cancer (e.g., non-small-cell lung cancer or small-cell lung cancer), head and neck cancer (e.g., laryngeal cancer, hypopharyngeal cancer, oropharyngeal cancer, and oral cavity cancer), bladder cancer (e.g., transitional cell carcinoma of the bladder), and colorectal cancer (e.g., advanced (non-resectable locally advanced or metastatic) colorectal cancer). In yet other embodiments, cancers for which the methods and compositions of the invention provide predictive results for chemosensitivity include cervical cancer (e.g., recurrent and stage IVB), thymoma, malignant pleural mesothelioma, esophageal carcinoma (e.g., squamous cell cancer and adenocarcinoma), rectal cancer, and anal squamous neoplasms (e.g., anal canal squamous cell carcinoma, anal margin squamous cell carcinoma, and anal intraepithelial neoplasia), and gastric cancer.

In some embodiments, the cancer is one in which one or more platinum-based drugs are employed either alone or in combination with other chemotherapeutic agents as a part of a chemotherapy treatment regimen. In other embodiments, the cancer is one in which one or more platinum-based drugs are employed either alone or in combination with additional treatment regimens, such as surgical procedures, radiation, and/or other chemotherapy treatments. In certain embodiments, the cancer is one in which one or more platinum-based agents are used as a first-line form of treatment.

In certain embodiments, the instant invention relates to methods and compositions encompassing the detection of expression levels of an X-linked gene in one or more cells of a subject. Typically, the subject is a human patient who has or is suspected of having at least one type of cancer, and the expression levels of the X-linked gene are detected in a sample of one or more cells, typically one or more tumor cells, from the human patient, which are then compared with the expression levels of the X-linked gene in a control sample. A control sample will generally be one in which the X-linked gene expression
levels are known and correlated with chemosensitivity to a certain drug or group of
drugs. In some embodiments, the control sample is one in which the X-linked gene
expression levels are known and correlated with a lack of chemosensitivity to a certain
drug or group of drugs. In some embodiments, more than one control sample is used for
comparative purposes with the test sample from the subject.

In certain embodiments, the invention relates to a method for predicting a breast
cancer patient's response to platinum-based drug chemotherapy, such as cisplatin
treatment. Typically, the breast cancer patient has not yet received platinum-based drug
chemotherapy. In further embodiments, a sample of the breast cancer cells from the
patient is analyzed for the levels of expression of an X-linked gene, such as XIST. If
expression levels of the X-linked gene (e.g., XIST) are low compared to expression levels
in normal breast tissue, then the breast cancer cells in the patient are sensitive to
platinum-based chemotherapy.

In certain embodiments, the expression level of the X-linked gene, such as XIST,
in cancer tissue is lower than the expression level of the gene in normal tissue. In
predicting chemotherapy sensitivity of a tumor, cut-off levels of expression may be
determined empirically for the subject cancer for which chemotherapy sensitivity is being
assessed. For example, in some embodiments, the cut-off to determine low XIST
expression in breast cancer tissue is defined as 2 times the standard deviation below the
average expression level of XIST in normal breast tissue. Accordingly, in some
embodiments, breast cancer tissue exhibiting XIST expression levels 2 times the standard
deviation below the average in normal breast tissue will be identified as sensitive to
chemotherapy, such as platinum-based drug treatment.

The terms "marker" and "biomarker" are used interchangeably herein and refer to
a gene, protein, or fragment thereof, the expression or level of which changes between
certain conditions. Where the expression or level of the gene, protein, or fragment
thereof correlates with a certain condition, the gene, protein, or fragment thereof is a
marker for that condition.

By "inactivating mutation" is meant a mutation in, for example, a nucleic acid
that encodes a protein that is inactive. This includes, for example, mutations that result in
the loss of protein expression and/or activity and includes genetic mutations such as point
mutations, translocations, amplifications, deletions (including whole gene deletions), and
hypomorphic mutations (e.g., where an altered gene product possesses a reduced level of
activity or where the wild-type gene product is expressed at a reduced level).
"Inactivating mutation" also includes biomarker dysfunctions due to post-translational protein regulation, for example, where a protein biomarker is inactive or exhibits impaired activity due to, for example, one or more posttranslational modifications, such as phosphorylation that results in protein inactivity.

The term "biomarker dysfunction" with regard to a protein or protein fragment refers to dysfunction of the protein or fragment thereof as a result of improper regulation at the posttranslational level, such as, for example, phosphorylation that results in protein inactivity.

"Sensitive," "sensitivity," or "chemosensitivity" in the context of treatment of a cancer cell with a chemotherapeutic agent or other compound means that the chemotherapeutic agent or other compound is likely to have an effect on the cancer cell. In preferred embodiments, the effect on the cancer cell is cytotoxicity.

When referring to the X-linked gene, XIST, both "XIST" and "Xist" are used interchangeably and refer to an XIST gene or the encoded XIST RNA regardless of the origin of the sequence (e.g., human, mouse, etc.).

By "X-linked gene" is meant any gene located on an X chromosome.

By "reference X-linked gene" is meant an X-linked gene in a control sample, e.g., a normal sample such as a non-cancerous tissue sample. Typically, the expression levels of a reference X-linked gene are known (e.g., quantified) and serve as a reference for comparative purposes with the levels of expression of the same X-linked gene in a different sample, typically a test sample, such as a breast tumor sample.

As used herein, the terms "drug," "agent," and "compound," either alone or together with "chemotherapeutic" or "chemotherapy," encompass any composition of matter or mixture which provides some pharmacologic effect that can be demonstrated in-vivo or in vitro. This includes small molecules, antibodies, microbiologicals, vaccines, vitamins, and other beneficial agents. As used herein, the terms further include any physiologically or pharmacologically active substance that produces a localized or systemic effect in a patient.

The term "nucleic acid" encompasses DNA, RNA (e.g., mRNA, tRNA), heteroduplexes, and synthetic molecules capable of encoding a polypeptide and includes all analogs and backbone substitutes such as PNA that one of ordinary skill in the art would recognize as capable of substituting for naturally occurring nucleotides and backbones thereof. Nucleic acids may be single stranded or double stranded, and may be chemical modifications. The terms "nucleic acid" and "polynucleotide" are used
interchangeably. Because the genetic code is degenerate, more than one codon may be used to encode a particular amino acid, and the present compositions and methods encompass nucleotide sequences which encode a particular amino acid sequence.

Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

As used herein, the term "amino acid sequence" is synonymous with the terms "polypeptide," "protein," and "peptide," and are used interchangeably. Where such amino acid sequences exhibit activity, they may be referred to as an "enzyme." The conventional one-letter or three-letter code for amino acid residues are used herein.

As used herein, a "synthetic" molecule is produced by \textit{in vitro} chemical or enzymatic synthesis rather than by an organism.

As used herein, the term "expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation. The term "expression" also includes the protein product of a translated mRNA, such as ATM protein that is translated from an ATM mRNA. The term "expression" as it refers to protein includes both protein levels (e.g., amounts) and protein activity (e.g., protein binding, enzymatic activity, etc.). The term "expression" also refers to the transcription of non-translated nucleic acid (e.g., non-coding mRNA), such as XIST mRNA.

A "gene" refers to the DNA segment encoding a polypeptide or RNA.

By "homolog" is meant an entity having a certain degree of identity with the subject amino acid sequences and the subject nucleotide sequences. As used herein, the term "homolog" covers identity with respect to structure and/or function, for example, the expression product of the resultant nucleotide sequence has the enzymatic activity of a subject amino acid sequence. With respect to sequence identity, preferably there is at least 70\%, 75\%, 80\%, 81\%, 82\%, 83\%, 84\%, 85\%, 86\%, 87\%, 88\%, 89\%, 90\%, 91\%, 92\%, 93\%, 94\%, 95\%, 96\%, 97\%, 98\%, or even 99\% sequence identity. These terms also encompass allelic variations of the sequences. The term, homolog, may apply to the relationship between genes separated by the event of speciation or to the relationship between genes separated by the event of genetic duplication.

Relative sequence identity can be determined by commercially available computer programs that can calculate \% identity between two or more sequences using any suitable
algorithm for determining identity, using, for example, default parameters. A typical example of such a computer program is CLUSTAL. Advantageously, the BLAST algorithm is employed, with parameters set to default values. The BLAST algorithm is described in detail on the National Center for Biotechnology Information (NCBI) website.

Examples of homologs according to the invention include XIST homologs, such as nucleotides with at least 70%, at least 80%, at least 90%, at least 95%, at least 98% sequence identity to the nucleotide sequence depicted in GenBank Accession No. NR_001564.1. See also, for example, Gene ID 7503 available through the NCBI website.

Examples of homologs according to the invention also include Utx homologs, such as nucleotides with at least 70%, at least 80%, at least 90%, at least 95%, at least 98% sequence identity to the nucleotide sequence depicted in GenBank Accession No. NM_021140.2. See also, for example, Gene ID 7403 available through the NCBI website.

Examples of homologs according to the invention also include Jaridlc homologs, such as nucleotides with at least 70%, at least 80%, at least 90%, at least 95%, at least 98% sequence identity to the nucleotide sequence depicted in GenBank Accession No. NM_001146702.1. See also, for example, Gene ID 8242 available through the NCBI website.

Examples of homologs according to the invention also include Eif2s3x homologs, such as nucleotides with at least 70%, at least 80%, at least 90%, at least 95%, at least 98% sequence identity to the nucleotide sequence depicted in GenBank Accession No. NM_001100542.1. See also, for example, Gene ID 299027 available through the NCBI website.

Examples of homologs according to the invention further include ATM homologs, such as nucleotides with at least 70%, at least 80%, at least 90%, at least 95%, at least 98% sequence identity to the nucleotide sequence depicted in GenBank Accession No. NM_000051.3. See also, for example, Gene ID 472 available through the NCBI website.

Examples of homologs according to the invention further include Rad50 homologs, such as nucleotides with at least 70%, at least 80%, at least 90%, at least 95%, at least 98% sequence identity to the nucleotide sequence depicted in GenBank Accession No. NM_005732.3. See also, for example, Gene ID 1011 available through the NCBI website.

Examples of homologs according to the invention further include Ube3b homologs, such as nucleotides with at least 70%, at least 80%, at least 90%, at least 95%, at least 98% sequence identity to the nucleotide sequence depicted in GenBank Accession No. NM_021564.1. See also, for example, Gene ID 7503 available through the NCBI website.
sequence identity to the nucleotide sequence depicted in GenBank Accession No. NM_130466.2. See also, for example, Gene ID 89910 available through the NCBI website.

The homologs of the peptides as provided herein typically have structural similarity with such peptides. A homolog of a polypeptide includes one or more conservative amino acid substitutions, which may be selected from the same or different members of the class to which the amino acid belongs.

In one embodiment, the sequences may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent substance. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the secondary binding activity of the substance is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine, glycine, alanine, asparagine, glutamine, serine, threonine, phenylalanine, and tyrosine.

The present invention also encompasses conservative substitution (substitution and replacement are both used herein to mean the interchange of an existing amino acid residue with an alternative residue) that may occur e.g., like-for-like substitution such as basic for basic, acidic for acidic, polar for polar, etc. Non-conservative substitution may also occur e.g., from one class of residue to another or alternatively involving the inclusion of unnatural amino acids such as ornithine (hereinafter referred to as Z), diaminobutyric acid ornithine (hereinafter referred to as B), norleucine ornithine (hereinafter referred to as O), pyriylalanine, thienylalanine, naphthylalanine and phenylglycine. Conservative substitutions that may be made are, for example, within the groups of basic amino acids (Arginine, Lysine and Histidine), acidic amino acids (glutamic acid and aspartic acid), aliphatic amino acids (Alanine, Valine, Leucine, Isoleucine), polar amino acids (Glutamine, Asparagine, Serine, Threonine), aromatic amino acids (Phenylalanine, Tryptophan and Tyrosine), hydroxyl amino acids (Serine, Threonine), large amino acids (Phenylalanine and Tryptophan) and small amino acids (Glycine, Alanine).

The present invention employs, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA and immunology, which are within the capabilities of a person of ordinary skill in the art.

**METHODS OF DETECTING EXPRESSION LEVELS**

There are many methods known in the art for determining the genotype of a patient and for identifying or analyzing whether a given DNA sample contains a particular expression level of a gene. Any method for determining genotype can be used for determining genotypes in the present invention. Such methods include, but are not limited to, amplimer sequencing, DNA sequencing, fluorescence spectroscopy, fluorescence resonance energy transfer (or "FRET")-based hybridization analysis, high throughput screening, mass spectroscopy, nucleic acid hybridization, polymerase chain reaction (PCR), RFLP analysis and size chromatography (e.g., capillary or gel chromatography), all of which are well known to one of ordinary skill in the art.

Many methods of sequencing genomic DNA are known in the art, and any such method can be used, see for example Sambrook et al., *Molecular Cloning: A Laboratory Manual* 2d ed. (1989). For example, a DNA fragment of interest can be amplified using the polymerase chain reaction or some other cyclic polymerase mediated amplification reaction. The amplified region of DNA can then be sequenced using any method known in the art. Advantageously, the nucleic acid sequencing is by automated methods (reviewed by Meldrum, Genome Res. September 2000;10(9):1288-303, the disclosure of which is incorporated by reference in its entirety), for example using a Beckman CEQ 8000 Genetic Analysis System (Beckman Coulter Instruments, Inc.). Methods for sequencing nucleic acids include, but are not limited to, automated fluorescent DNA sequencing (see, e.g., Watts & MacBeath, Methods Mol Biol. 2001;167:153-70 and MacBeath et al., Methods Mol Biol. 2001;167:1 19-52), capillary electrophoresis (see,
e.g., Bosserhoff et al, Comb Chem High Throughput Screen. December 2000;3(6):455-
66), DNA sequencing chips (see, e.g., Jain, Pharmacogenomics. August 2000;1(3):289-
307), mass spectrometry (see, e.g., Yates, Trends Genet. January 2000;16(1):5-8),
pyrosequencing (see, e.g., Ronaghi, Genome Res. January 2001;11(1):3-11), and
ultrathin-layer gel electrophoresis (see, e.g., Guttman & Ronai, Electrophoresis.
December 2000; 21(18):3952-64), the disclosures of which are hereby incorporated by
reference in their entireties. The sequencing can also be done by any commercial
company. Examples of such companies include, but are not limited to, the University of
Georgia Molecular Genetics Instrumentation Facility (Athens, Ga.) or SeqWright DNA
Technologies Services (Houston, Tex.).

Any one of the methods known in the art for amplification of DNA may be used,
such as for example, the polymerase chain reaction (PCR), the ligase chain reaction
displacement assay (SDA), or the oligonucleotide ligation assay ("OLA") (Landegren, U.
et al, Science 241:1077-1080 (1988)). Nickerson, D. A. et al. have described a nucleic
acid detection assay that combines attributes of PCR and OLA (Nickerson, D. A. et al.,
Proc. Natl. Acad. Sci. (U.S.A.) 87:8923-8927 (1990)). Other known nucleic acid
amplification procedures, such as transcription-based amplification systems (Malek, L. T.
Schuster et al, U.S. Pat. No. 5,169,766; Miller, H. I. et al, PCT Application
T. R. et al., PCT Application W088/10315)), or isothermal amplification methods
used.

To perform a cyclic polymerase mediated amplification reaction according to the
present invention, the primers are hybridized or annealed to opposite strands of the target
DNA, the temperature is then raised to permit the thermostable DNA polymerase to
extend the primers and thus replicate the specific segment of DNA spanning the region
between the two primers. Then the reaction is thermocycled so that at each cycle the
amount of DNA representing the sequences between the two primers is doubled, and
specific amplification of gene DNA sequences, if present, results.

Any of a variety of polymerases can be used in the present invention. For
thermocyclic reactions, the polymerases are thermostable polymerases such as Taq,
KlenTaq, Stoffel Fragment, Deep Vent, Tth, Pfu, Vent, and UlTma, each of which are
readily available from commercial sources. For non-thermocyclic reactions, and in certain thermocyclic reactions, the polymerase will often be one of many polymerases commonly used in the field, and commercially available, such as DNA pol 1, Klenow fragment, T7 DNA polymerase, and T4 DNA polymerase. Guidance for the use of such polymerases can readily be found in product literature and in general molecular biology guides.

Typically, the annealing of the primers to the target DNA sequence is carried out for about 2 minutes at about 37-55° C, extension of the primer sequence by the polymerase enzyme (such as Taq polymerase) in the presence of nucleoside triphosphates is carried out for about 3 minutes at about 70-75° C, and the denaturing step to release the extended primer is carried out for about 1 minute at about 90-95° C. However, these parameters can be varied, and one of skill in the art would readily know how to adjust the temperature and time parameters of the reaction to achieve the desired results. For example, cycles may be as short as 10, 8, 6, 5, 4.5, 4, 2, 1, 0.5 minutes or less.

Also, "two temperature" techniques can be used where the annealing and extension steps may both be carried out at the same temperature, typically between about 60-65° C, thus reducing the length of each amplification cycle and resulting in a shorter assay time.

Typically, the reactions described herein are repeated until a detectable amount of product is generated. Often, such detectable amounts of product are between about 10 ng and about 100 ng, although larger quantities, e.g. 200 ng, 500 ng, 1 mg or more can also, of course, be detected. In terms of concentration, the amount of detectable product can be from about 0.01 pmol, 0.1 pmol, 1 pmol, 10 pmol, or more. Thus, the number of cycles of the reaction that are performed can be varied, the more cycles are performed, the more amplified product is produced. In certain embodiments, the reaction comprises 2, 5, 10, 15, 20, 30, 40, 50, or more cycles.

For example, the PCR reaction may be carried out using about 25-50 μl samples containing about 0.01 to 1.0 ng of template amplification sequence, about 10 to 100 pmol of each generic primer, about 1.5 units of Taq DNA polymerase (Promega Corp.), about 0.2 mM dDATP, about 0.2 mM dCTP, about 0.2 mM dGTP, about 0.2 mM dTTP, about 15 mM MgCl₂, about 10 mM Tris-HCl (pH 9.0), about 50 mM KC1, about 1 μg/ml gelatin, and about 10 μl/ml Triton X-100 (Saiki, 1988).
Those of ordinary skill in the art are aware of the variety of nucleotides available for use in the cyclic polymerase mediated reactions. Typically, the nucleotides will consist at least in part of deoxynucleotide triphosphates (dNTPs), which are readily commercially available. Parameters for optimal use of dNTPs are also known to those of skill, and are described in the literature. In addition, a large number of nucleotide derivatives are known to those of skill and can be used in the present reaction. Such derivatives include fluorescently labeled nucleotides, allowing the detection of the product including such labeled nucleotides, as described below. Also included in this group are nucleotides that allow the sequencing of nucleic acids including such nucleotides, such as chain-terminating nucleotides, dideoxynucleotides and boronated nuclease-resistant nucleotides. Commercial kits containing the reagents most typically used for these methods of DNA sequencing are available and widely used. Other nucleotide analogs include nucleotides with bromo-, iodo-, or other modifying groups, which affect numerous properties of resulting nucleic acids including their antigenicity, their replicatability, their melting temperatures, their binding properties, etc. In addition, certain nucleotides include reactive side groups, such as sulphydryl groups, amino groups, N-hydroxysuccinimidyl groups, that allow the further modification of nucleic acids comprising them.

In certain embodiments, oligonucleotides that can be used as primers to amplify specific nucleic acid sequences of a gene in cyclic polymerase-mediated amplification reactions, such as PCR reactions, consist of oligonucleotide fragments. Such fragments should be of sufficient length to enable specific annealing or hybridization to the nucleic acid sample. The sequences typically will be about 8 to about 44 nucleotides in length, but may be longer. Longer sequences, e.g., from about 14 to about 50, are advantageous for certain embodiments.

In embodiments where it is desired to amplify a fragment of DNA, primers having contiguous stretches of about 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 nucleotides from a gene sequence are contemplated.

As used herein, "hybridization" refers to the process by which one strand of nucleic acid base pairs with a complementary strand, as occurs during blot hybridization techniques and PCR techniques.

Whichever probe sequences and hybridization methods are used, one ordinarily skilled in the art can readily determine suitable hybridization conditions, such as temperature and chemical conditions. Such hybridization methods are well known in the
art. For example, for applications requiring high selectivity, one will typically desire to employ relatively stringent conditions for the hybridization reactions, e.g., one will select relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.10 M NaCl at temperatures of about 50 °C to about 70 °C. Such high stringency conditions tolerate little, if any, mismatch between the probe and the template or target strand. It is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide. Other variations in hybridization reaction conditions are well known in the art (see for example, Sambrook et al., Molecular Cloning: A Laboratory Manual 2d ed. (1989)).

Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex, as taught, e.g., in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA), and confer a defined "stringency" as explained below.

Maximum stringency typically occurs at about Tm-5 °C (5 °C below the Tm of the probe); high stringency at about 5 °C to 10 °C below Tm; intermediate stringency at about 10 °C to 20 °C below Tm; and low stringency at about 20 °C to 25 °C below Tm. As will be understood by those of ordinary skill in the art, a maximum stringency hybridization can be used to identify or detect identical nucleotide sequences while an intermediate (or low) stringency hybridization can be used to identify or detect similar or related polynucleotide sequences.

In one aspect, the present invention employs nucleotide sequences that can hybridize to another nucleotide sequence under stringent conditions (e.g., 65 °C and O.1xSSC {1xSSC = 0.15 M NaCl, 0.015 M Na3 Citrate pH 7.0). Where the nucleotide sequence is double-stranded, both strands of the duplex, either individually or in combination, may be employed by the present invention. Where the nucleotide sequence is single-stranded, it is to be understood that the complementary sequence of that nucleotide sequence is also included within the scope of the present invention.

Stringency of hybridization refers to conditions under which polynucleic acid hybrids are stable. Such conditions are evident to those of ordinary skill in the field. As known to those of ordinary skill in the art, the stability of hybrids is reflected in the melting temperature (Tm) of the hybrid which decreases approximately 1 to 1.5 °C with every 1 % decrease in sequence homology. In general, the stability of a hybrid is a function of sodium ion concentration and temperature. Typically, the hybridization
reaction is performed under conditions of higher stringency, followed by washes of varying stringency.

As used herein, high stringency includes conditions that permit hybridization of only those nucleic acid sequences that form stable hybrids in 1 M Na+ at 65-68 °C. High stringency conditions can be provided, for example, by hybridization in an aqueous solution containing 6x SSC, 5x Denhardt's, 1 % SDS (sodium dodecyl sulphate), 0.1 Na+ pyrophosphate and 0.1 mg/ml denatured salmon sperm DNA as non-specific competitor. Following hybridization, high stringency washing may be done in several steps, with a final wash (about 30 minutes) at the hybridization temperature in 0.2 - 0.1x SSC, 0.1 % SDS.

It is understood that these conditions may be adapted and duplicated using a variety of buffers, e.g., formamide-based buffers, and temperatures. Denhardt's solution and SSC are well known to those of ordinary skill in the art as are other suitable hybridization buffers (see, e.g., Sambrook, et al., eds. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York or Ausubel, et al., eds. (1990) Current Protocols in Molecular Biology, John Wiley & Sons, Inc.). Optimal hybridization conditions are typically determined empirically, as the length and the GC content of the hybridizing pair also play a role.

Nucleic acid molecules that differ from the sequences of the primers and probes disclosed herein, are intended to be within the scope of the invention. Nucleic acid sequences that are complementary to these sequences, or that are hybridizable to the sequences described herein under conditions of standard or stringent hybridization, and also analogs and derivatives are also intended to be within the scope of the invention. Advantageously, such variations will differ from the sequences described herein by only a small number of nucleotides, for example by 1, 2, or 3 nucleotides.

Nucleic acid molecules corresponding to natural allelic variants, homologues (i.e., nucleic acids derived from other species), or other related sequences (e.g., paralogs) of the sequences described herein can be isolated based on their homology to the nucleic acids disclosed herein, for example by performing standard or stringent hybridization reactions using all or a portion of the known sequences as probes. Such methods for nucleic acid hybridization and cloning are well known in the art.

Similarly, a nucleic acid molecule detected in the methods of the invention may include only a fragment of the specific sequences described. Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids, a length sufficient to
allow for specific hybridization of nucleic acid primers or probes, and are at most some portion less than a full-length sequence. Fragments may be derived from any contiguous portion of a nucleic acid sequence of choice. Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below.

Derivatives, analogs, homologues, and variants of the nucleic acids of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids of the invention, in various embodiments, by at least about 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or even 99% identity over a nucleic acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art.

For the purposes of the present invention, sequence identity or homology is determined by comparing the sequences when aligned so as to maximize overlap and identity while minimizing sequence gaps. In particular, sequence identity may be determined using any of a number of mathematical algorithms. A nonlimiting example of a mathematical algorithm used for comparison of two sequences is the algorithm of Karlin & Altschul, Proc. Natl. Acad. Sci. USA 1990;87: 2264-2268, modified as in Karlin & Altschul, Proc. Natl. Acad. Sci. USA 1993;90: 5873-5877.

Another example of a mathematical algorithm used for comparison of sequences is the algorithm of Myers & Miller, CABIOS 1988;4: 11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Yet another useful algorithm for identifying regions of local sequence similarity and alignment is the FASTA algorithm as described in Pearson & Lipman, Proc. Natl. Acad. Sci. USA 1988;85: 2444-2448.

Advantageous for use according to the present invention is the WU-BLAST (Washington University BLAST) version 2.0 software. WU-BLAST version 2.0 executable programs for several UNIX platforms can be downloaded from ftp://blast.wustl.edu/blast/executables. This program is based on WU-BLAST version 1.4, which in turn is based on the public domain NCBI-BLAST version 1.4 (Altschul & Gish, 1996, Local alignment statistics, Doolittle ed., Methods in Enzymology 266: 460-480; Altschul et al, Journal of Molecular Biology 1990;215: 403-410; Gish & States,
In all search programs in the suite the gapped alignment routines are integral to the database search itself. Gapping can be turned off if desired. The default penalty (Q) for a gap of length one is Q=9 for proteins and BLASTP, and Q=10 for BLASTN, but may be changed to any integer. The default per-residue penalty for extending a gap (R) is R=2 for proteins and BLASTP, and R=10 for BLASTN, but may be changed to any integer. Any combination of values for Q and R can be used in order to align sequences so as to maximize overlap and identity while minimizing sequence gaps. The default amino acid comparison matrix is BLOSUM62, but other amino acid comparison matrices such as PAM can be utilized.

Alternatively or additionally, the term "homology" or "identity", for instance, with respect to a nucleotide or amino acid sequence, can indicate a quantitative measure of homology between two sequences. The percent sequence homology can be calculated as (N_{ref} - N_{dif}) * 100 / N_{ref}, wherein N_{dif} is the total number of non-identical residues in the two sequences when aligned and wherein N_{ref} is the number of residues in one of the sequences. Hence, the DNA sequence AGTCAGTC will have a sequence identity of 75% with the sequence AATCAATC (N_{ref}=8; N_{dif}=2). "Homology" or "identity" can refer to the number of positions with identical nucleotides or amino acids divided by the number of nucleotides or amino acids in the shorter of the two sequences wherein alignment of the two sequences can be determined in accordance with the Wilbur and Lipman algorithm (Wilbur & Lipman, Proc Natl Acad Sci USA 1983;80:726, incorporated herein by reference), for instance, using a window size of 20 nucleotides, a word length of 4 nucleotides, and a gap penalty of 4, and computer-assisted analysis and interpretation of the sequence data including alignment can be conveniently performed using commercially available programs (e.g., Intelligenetics.TM. Suite, Intelligenetics Inc. CA). When RNA sequences are said to be similar, or have a degree of sequence identity or homology with DNA sequences, thymidine (T) in the DNA sequence is considered equal to uracil (U) in the RNA sequence. Thus, RNA sequences are within the scope of the invention and can be derived from DNA sequences, by thymidine (T) in the DNA sequence being considered equal to uracil (U) in RNA sequences. Without undue experimentation, the skilled artisan can consult with many other programs or references for determining percent homology.
In embodiments where expression of a particular gene (e.g., a translated X-linked gene) is assessed by determining the expression of the protein product of the gene, any suitable assay for detecting protein levels and/or activity may be employed. For example, the protein products of the translated X-linked genes, Utx and Jarid1C, are demethylases. Utx is an H3K27me3 demethylase, and Jarid1C is a histone H3K4 demethylase. Accordingly, in certain embodiments, Utx and/or Jarid1C expression is assessed by determining the activity of Utx and/or Jarid1C protein in an assay that detects Utx and/or Jarid1C-associated histone demethylation. See, for example, Agger, K., et al. "UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development" Nature 449:731-734 (2007) and Tahiliani M., et al. "The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation" Nature 447:601-605 (2007). Likewise, the protein product of the translated X-linked gene, Eif2s3x, is a translation initiation factor and possesses GTP-binding and GTPase activity. Accordingly, in certain embodiments, Eif2s3x expression is assessed by determining the activity of Eif2s3 protein in an assay that detects Eif2s3x-associated translation initiation, Eif2s3x-associated GTP-binding and/or GTPase activity.

Other suitable protein activity assays include ubiquitination assays, kinase assays, protein-binding assays, DNA-binding and unwinding assays, and any other suitable assay for assessing the activity of the protein product of a translated gene according to the invention. For example, Ube3b is an E3 ubiquitin protein ligase. Accordingly, in certain embodiments, Ube3b gene expression is assessed by determining the ubiquitination activity of Ube3b protein. Likewise, the protein product of Rad50 is part of the MRN complex, which recruits ATM to sites of DNA damage. In certain embodiments, determination of Rad50 gene expression is determined by assessing Rad50 protein activity, for example, by assessing the ability of Rad50 protein to bind DNA. In some embodiments, determination of ATM gene expression is determined by assessing ATM protein activity, for example, by assessing the phosphorylation state of ATM in a kinase assay or by the ability of antibodies to bind ATM that are specific for ATM phosphorylation. See, for example, Lamarche, BJ, et al. "The MRN complex in double-strand break repair and telomere maintenance" FEBS Lett. 584(17):3682-95 (2010) and Ji-Hoon Lee and Tanya T. Paull. "ATM Activation by DNA Double-Strand Breaks Through the Mrel 1-Rad50-Nbsl Complex" Science 308(5721):551-554 (2005).
SAMPLING

In order to determine the genotype or expression level of a particular X-linked or other gene of a patient according to the methods of the present invention, it may be necessary to obtain a sample of genomic DNA or RNA or protein from that patient. That sample of genomic DNA or RNA or protein may be obtained from a sample of tissue or cells taken from that patient.

A sample may comprise any clinically relevant tissue sample, such as a tumor biopsy or fine needle aspirate, hair (including roots), skin, buccal swabs, saliva, or a sample of bodily fluid, such as blood, plasma, serum, lymph, ascitic fluid, cystic fluid, urine or nipple exudate. The sample may be taken from a human, or, in a veterinary context, from non-human animals such as ruminants, horses, swine or sheep, or from domestic companion animals such as felines and canines.

The tissue sample may be marked with an identifying number or other indicia that relates the sample to the individual patient from which the sample was taken. The identity of the sample advantageously remains constant throughout the methods of the invention thereby guaranteeing the integrity and continuity of the sample during extraction and analysis. Alternatively, the indicia may be changed in a regular fashion that ensures that the data, and any other associated data, can be related back to the patient from whom the data was obtained. The amount/size of sample required is known to those ordinarily skilled in the art.

Generally, the tissue sample may be placed in a container that is labeled using a numbering system bearing a code corresponding to the patient. Accordingly, the genotype of a particular patient is easily traceable.

In one embodiment of the invention, a sampling device and/or container may be supplied to the physician. The sampling device advantageously takes a consistent and reproducible sample from individual patients while simultaneously avoiding any cross-contamination of tissue. Accordingly, the size and volume of sample tissues derived from individual patients would be consistent.

According to certain embodiments of the present invention, a sample of genomic DNA or RNA or protein is obtained from the tissue sample of the patient of interest. Whatever source of cells or tissue is used, a sufficient amount of cells must be obtained to provide a sufficient amount of DNA or RNA or protein for analysis. This amount will be known or readily determinable by those ordinarily skilled in the art.
DNA or RNA is isolated from the tissue/cells by techniques known to those ordinarily skilled in the art (see, e.g., U.S. Pat. Nos. 6,548,256 and 5,989,431, Hirota et al, Jinrui Idengaku Zasshi. September 1989; 34(3):217-23 and John et al, Nucleic Acids Res. Jan. 25, 1991;19(2):408; the disclosures of which are incorporated by reference in their entireties). For example, high molecular weight DNA may be purified from cells or tissue using proteinase K extraction and ethanol precipitation. DNA may be extracted from a patient specimen using any other suitable methods known in the art.

In certain embodiments, target polynucleotide molecules are extracted from a sample taken from an individual afflicted with breast cancer. The sample may be collected in any clinically acceptable manner, but must be collected such that marker-derived polynucleotides (e.g., RNA) are preserved. mRNA or nucleic acids derived therefrom (e.g., cDNA or amplified DNA) are preferably labeled distinguishably from standard or control polynucleotide molecules, and both are simultaneously or independently hybridized to a microarray comprising one or more markers of chemosensitivity as described above. Alternatively, mRNA or nucleic acids derived therefrom may be labeled with the same label as the standard or control polynucleotide molecules, wherein the intensity of hybridization of each at a particular probe is compared.


RNA may be isolated from eukaryotic cells by procedures that involve lysis of the cells and denaturation of the proteins contained therein. Cells of interest include wild-type cells (i.e., non-cancerous), drug-exposed wild-type cells, tumor- or tumor-derived cells, modified cells, normal or tumor cell line cells, and drug-exposed modified cells.

Additional steps may be employed to remove DNA. Cell lysis may be accomplished with a nonionic detergent, followed by microcentrifugation to remove the nuclei and hence the bulk of the cellular DNA. In one embodiment, RNA is extracted from cells of the various types of interest using guanidinium thiocyanate lysis followed by CsCl centrifugation to separate the RNA from DNA (Chirgwin et al., Biochemistry 18:5294-5299 (1979)). Poly(A)+ RNA is selected by selection with oligo-dT cellulose
Alternatively, separation of RNA from DNA can be accomplished by organic extraction, for example, with hot phenol or phenol/chloroform/isoamyl alcohol.

If desired, RNase inhibitors may be added to the lysis buffer. Likewise, for certain cell types, it may be desirable to add a protein denaturation/digestion step to the protocol.

In certain embodiments, it is desirable to preferentially enrich mRNA with respect to other cellular RNAs, such as transfer RNA (tRNA) and ribosomal RNA (rRNA). Most mRNAs contain a poly(A) tail at their 3' end. This allows them to be enriched by affinity chromatography, for example, using oligo(dT) or poly(U) coupled to a solid support, such as cellulose or Sephadex™ (see Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, vol. 2, Current Protocols Publishing, New York (1994). Once bound, poly(A)+ mRNA is eluted from the affinity column using 2 mM EDTA/0.1% SDS.

The sample of RNA can comprise a plurality of different mRNA molecules, each different mRNA molecule having a different nucleotide sequence. In a specific embodiment, the RNA sample is a mammalian RNA sample.

In a specific embodiment, total RNA or mRNA from cells are used in the methods of the invention. The source of the RNA can be cells of any animal, human, mammal, primate, non-human animal, dog, cat, mouse, rat, bird, yeast, eukaryote, etc. In specific embodiments, the method of the invention is used with a sample containing total mRNA or total RNA from 1×10⁶ cells or less. In another embodiment, proteins can be isolated from the foregoing sources, by methods known in the art, for use in expression analysis at the protein level.

In certain embodiments, expression of a biomarker according to the invention is measured using multiplex ligation-dependent probe amplification (MLPA) (see, e.g., WO 01/61033 and Schouten, JP et al. (2002) "Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification" Nucleic Acids Res 30, e57) or reverse transcriptase MLPA (RT-MLPA) (see, e.g., Eldering, E et al. (2003) "Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signaling pathways" Nucleic Acids Res 31, e153). In RT-MLPA,
mRNA is converted to cDNA by reverse transcriptase, followed by a normal MLPA reaction. In other embodiments, methylation-specific MLPA is employed to detect expression of a biomarker according to the instant invention (see, e.g., Nygren, AO et al. (2005) "Methylation-specific MLPA (MS-MPLA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences" Nucleic Acids Res 33, 14:el28).

ARRAYS

As defined herein, a "nucleic acid array" refers to a plurality of unique nucleic acids (or "nucleic acid members") attached to a support where each of the nucleic acid members is attached to a support in a unique pre-selected region.

In one embodiment, the nucleic acid member attached to the surface of the support is DNA. In another embodiment, the nucleic acid member attached to the surface of the support is either cDNA or oligonucleotides. In another embodiment, the nucleic acid member attached to the surface of the support is cDNA synthesized by polymerase chain reaction (PCR). In another embodiment, sequences bound to the array can be an isolated oligonucleotide, cDNA, EST or PCR product corresponding to any biomarker of the invention total cellular RNA is applied to the array.

Thus in one aspect, the present invention relates to an array comprising a nucleic acid which binds to at least one of the markers selected from the group consisting of XIST, Utx, Jaridlc and Eif2s3x for the determination of chemosensitivity to a certain compound, such as a platinum-based compound (e.g., cisplatin).


In general, any library may be arranged in an orderly manner into an array, by spatially separating the members of the library. Examples of suitable libraries for arraying include nucleic acid libraries (including DNA, cDNA, oligonucleotide, etc. libraries), peptide, polypeptide and protein libraries, as well as libraries comprising any molecules, such as ligand libraries, among others.

The samples (e.g., members of a library) are generally fixed or immobilized onto a solid phase, preferably a solid substrate, to limit diffusion and admixing of the samples. In particular, the libraries may be immobilized to a substantially planar solid phase, including membranes and non-porous substrates such as plastic and glass. Furthermore, the samples are preferably arranged in such a way that indexing (i.e., reference or access to a particular sample) is facilitated. Typically the samples are applied as spots in a grid formation. Common assay systems may be adapted for this purpose. For example, an array may be immobilized on the surface of a microplate, either with multiple samples in a well, or with a single sample in each well. Furthermore, the solid substrate may be a membrane, such as a nitrocellulose or nylon membrane (for example, membranes used in blotting experiments). Alternative substrates include glass, or silica-based substrates. Thus, the samples are immobilized by any suitable method known in the art, for example, by charge interactions, or by chemical coupling to the walls or bottom of the wells, or the
surface of the membrane. Other means of arranging and fixing may be used, for example, pipetting, drop-touch, piezoelectric means, ink-jet and bubblejet technology, electrostatic application, etc. In the case of silicon-based chips, photolithography may be utilized to arrange and fix the samples on the chip.

The samples may be arranged by being "spotted" onto the solid substrate; this may be done by hand or by making use of robotics to deposit the sample. In general, arrays may be described as macroarrays or microarrays, the difference being the size of the sample spots. Macroarrays typically contain sample spot sizes of about 300 microns or larger and may be easily imaged by existing gel and blot scanners. The sample spot sizes in microarrays are typically less than 200 microns in diameter and these arrays usually contain thousands of spots. Thus, microarrays may require specialized robotics and imaging equipment, which may need to be custom made. Instrumentation is described generally in a review by Cortese, 2000, The Scientist 14[11]:26.

Techniques for producing immobilized libraries of DNA molecules have been described in the art. Generally, most prior art methods described how to synthesize single-stranded nucleic acid molecule libraries, using for example masking techniques to build up various permutations of sequences at the various discrete positions on the solid substrate. U.S. Patent No. 5,837,832 describes an improved method for producing DNA arrays immobilized to silicon substrates based on very large scale integration technology. In particular, U.S. Patent No. 5,837,832 describes a strategy called "tiling" to synthesize specific sets of probes at spatially-defined locations on a substrate which may be used to produced the immobilized DNA libraries of the present invention. U.S. Patent No. 5,837,832 also provides references for earlier techniques that may also be used. Arrays may also be built using photo deposition chemistry.

To aid detection, labels are typically used - such as any readily detectable reporter, for example, a fluorescent, bioluminescent, phosphorescent, radioactive, etc. reporter. Labelling of probes and targets is also disclosed in Shalon et al, 1996, Genome Res 6(7):639-45.

Examples of DNA arrays include where probe cDNA (500-5,000 bases long) is immobilized to a solid surface such as glass using robot spotting and exposed to a set of targets either separately or in a mixture. This method is widely considered as having been developed at Stanford University (Ekins and Chu, 1999, Trends in Biotechnology, 1999, 17, 217-218).
Another example of a DNA array is where an array of oligonucleotides (20-25-mer oligos, preferably, 40-60 mer oligos) or peptide nucleic acid (PNA) probes are synthesized either in situ (on-chip) or by conventional synthesis followed by on-chip immobilization. The array is exposed to labelled sample DNA, hybridized, and the identity/abundance of complementary sequences are determined. Such a DNA chip is sold by Affymetrix, Inc., under the GeneChip® trademark. Agilent and Nimblegen also provide suitable arrays (e.g., genomic tiling arrays).

In other embodiments, high throughput DNA sequencing promises to become an affordable and more quantitative alternative for microarrays to analyze large collections of DNA sequences. Examples of high-throughput sequencing approaches are listed in E.Y. Chan, Mutation Research 573 (2005) 13-40 and include, but are not limited to, near-term sequencing approaches such as cycle-extension approaches, polymerase reading approaches and exonuclease sequencing, revolutionary sequencing approaches such as DNA scanning and nanopore sequencing and direct linear analysis. Examples of current high-throughput sequencing methods are 454 (pyro)sequencing, Solexa Genome Analysis System, Agencourt SOLiD sequencing method (Applied Biosystems), MS-PET sequencing (Ng et al., 2006, http://nar.oxfordjournals.org/cgi/content/full/34/12/e84).

PROBES

As used herein, the term "probe" refers to a molecule (e.g., an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification), that is capable of hybridizing to another molecule of interest (e.g., another oligonucleotide). When probes are oligonucleotides they may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular targets (e.g., gene sequences). As described herein, it is contemplated that probes used in the present invention may be labelled with a label so that is detectable in any detection system, including, but not limited to enzyme (e.g., ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems.

With respect to arrays and microarrays, the term "probe" is used to refer to any hybridizable material that is affixed to the array for the purpose of detecting a nucleotide sequence that has hybridized to said probe. Preferably, these probes are 25-60 mers or longer.
The present invention further encompasses probes according to the present invention that are immobilized on a solid or flexible support, such as paper, nylon or other type of membrane, filter, chip, glass slide, microchips, microbeads, or any other such matrix, all of which are within the scope of this invention.


In order to generate data from array-based assays a signal is detected that signifies the presence of or absence of hybridization between a probe and a nucleotide sequence.
The present invention further contemplates direct and indirect labelling techniques. For example, direct labelling incorporates fluorescent dyes directly into the nucleotide sequences that hybridize to the array-associated probes (e.g., dyes are incorporated into nucleotide sequence by enzymatic synthesis in the presence of labelled nucleotides or PCR primers). Direct labelling schemes yield strong hybridization signals, typically using families of fluorescent dyes with similar chemical structures and characteristics, and are simple to implement. In some embodiments comprising direct labelling of nucleic acids, cyanine or alexa analogs are utilized in multiple-fluor comparative array analyses. In other embodiments, indirect labelling schemes can be utilized to incorporate epitopes into the nucleic acids either prior to or after hybridization to the microarray probes. One or more staining procedures and reagents are used to label the hybridized complex (e.g., a fluorescent molecule that binds to the epitopes, thereby providing a fluorescent signal by virtue of the conjugation of dye molecule to the epitope of the hybridized species).

Oligonucleotide sequences used as probes according to the present invention may be labeled with a detectable moiety. Various labeling moieties are known in the art. Said moiety may be, for example, a radiolabel (e.g., 3H, 125I, 35S, 14C, 32P, etc.), detectable enzyme (e.g. horse radish peroxidase (HRP), alkaline phosphatase etc.), a fluorescent dye (e.g., fluorescein isothiocyanate, Texas red, rhodamine, Cy3, Cy5, Bodipy, Bodipy Far Red, Lucifer Yellow, Bodipy 630/650-X, Bodipy R6G-X and 5-CR 6G, and the like), a colorimetric label such as colloidal gold or colored glass or plastic (e.g. polystyrene, polypropylene, latex, etc.), beads, or any other moiety capable of generating a detectable signal such as a colorimetric, fluorescent, chemiluminescent or electrochemiluminescent (ECL) signal.

Probes may be labeled directly or indirectly with a detectable moiety, or synthesized to incorporate the detectable moiety. In one embodiment, a detectable label is incorporated into a nucleic acid during at least one cycle of a cyclic polymerase-mediated amplification reaction. For example, polymerases can be used to incorporate fluorescent nucleotides during the course of polymerase-mediated amplification reactions. Alternatively, fluorescent nucleotides may be incorporated during synthesis of nucleic acid primers or probes. To label an oligonucleotide with the fluorescent dye, one of conventionally-known labeling methods can be used (Nature Biotechnology, 14, 303-308, 1996; Applied and Environmental Microbiology, 63, 1143-147, 1997; Nucleic Acids Research, 24, 4532-4535, 1996). An advantageous probe is one labeled with a
fluorescent dye at the 3' or 5' end and containing G or C as the base at the labeled end. If the 5' end is labeled and the 3' end is not labeled, the OH group on the C atom at the 3'-position of the 3' end ribose or deoxyribose may be modified with a phosphate group or the like although no limitation is imposed in this respect.

Spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means can be used to detect such labels. The detection device and method may include, but is not limited to, optical imaging, electronic imaging, imaging with a CCD camera, integrated optical imaging, and mass spectrometry. Further, the amount of labeled or unlabeled probe bound to the target may be quantified. Such quantification may include statistical analysis. In other embodiments the detection may be via conductivity differences between concordant and discordant sites, by quenching, by fluorescence perturbation analysis, or by electron transport between donor and acceptor molecules.

In yet another embodiment, detection may be via energy transfer between molecules in the hybridization complexes in PCR or hybridization reactions, such as by fluorescence energy transfer (FET) or fluorescence resonance energy transfer (FRET). In FET and FRET methods, one or more nucleic acid probes are labeled with fluorescent molecules, one of which is able to act as an energy donor and the other of which is an energy acceptor molecule. These are sometimes known as a reporter molecule and a quencher molecule respectively. The donor molecule is excited with a specific wavelength of light for which it will normally exhibit a fluorescence emission wavelength. The acceptor molecule is also excited at this wavelength such that it can accept the emission energy of the donor molecule by a variety of distance-dependent energy transfer mechanisms. Generally the acceptor molecule accepts the emission energy of the donor molecule when they are in close proximity (e.g., on the same, or a neighboring molecule). FET and FRET techniques are well known in the art. See for example U.S. Pat. Nos. 5,668,648, 5,707,804, 5,728,528, 5,853,992, and 5,869,255 (for a description of FRET dyes), Tyagi et al. Nature Biotech. vol. 14, p 303-8 (1996), and Tyagi et al, Nature Biotech, vol 16, p 49-53 (1998) (for a description of molecular beacons for FET), and Mergny et al. Nucleic Acid Res. vol 22, p 920-928, (1994) and Wolf et al. PNAS vol 85, p 8790-94 (1988) (for general descriptions and methods for FET and FRET), each of which is hereby incorporated by reference.

The probes for use in an array of the invention may be greater than 40 nucleotides in length and may be isothermal.
In some embodiments, the probes, array of probes or set of probes will be immobilized on a support. Supports (e.g., solid supports) can be made of a variety of materials, such as glass, silica, plastic, nylon or nitrocellulose. Supports are preferably rigid and have a planar surface. Supports typically have from about 1-10,000,000 discrete spatially addressable regions, or cells. Supports having about 10-1,000,000 or about 100-100,000 or about 1000-100,000 cells are common. The density of cells is typically at least about 1000, 10,000, 100,000 or 1,000,000 cells within a square centimeter. In some supports, all cells are occupied by pooled mixtures of probes or a set of probes. In other supports, some cells are occupied by pooled mixtures of probes or a set of probes, and other cells are occupied, at least to the degree of purity obtainable by synthesis methods, by a single type of oligonucleotide.

Arrays of probes or sets of probes may be synthesized in a step-by-step manner on a support or can be attached in presynthesized form. One method of synthesis is VLSIPS™ (as described in U.S. 5,143,854 and EP 476,014), which entails the use of light to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays. Algorithms for design of masks to reduce the number of synthesis cycles are described in U.S. 5,571,639 and U.S. 5,593,839. Arrays can also be synthesized in a combinatorial fashion by delivering monomers to cells of a support by mechanically constrained flowpaths, as described in EP 624,059. Arrays can also be synthesized by spotting reagents on to a support using an ink jet printer (see, for example, EP 728,520).

DATA ANALYSIS

Data analysis is also an important part of an experiment involving arrays. The raw data from an array experiment typically are images, which need to be transformed into matrices - tables where rows represent, for example, genes, columns represent, for example, various samples such as tissues or experimental conditions, and numbers in each cell for example characterize the expression of a particular sequence (for example, a second sequence that has ligated to the first (target) nucleotide sequence) in the particular sample. These matrices have to be analyzed further, if any knowledge about the underlying biological processes is to be extracted. Methods of data analysis (including supervised and unsupervised data analysis as well as bioinformatics approaches) are disclosed in Brazma and Vilo J (2000) FEBS Lett 480(1): 17-24.

KITS
The materials for use in the methods of the present invention are ideally suited for preparation of kits. Oligonucleotides may be provided in containers that can be in any form, e.g., lyophilized, or in solution (e.g., a distilled water or buffered solution), etc. In one aspect of the present invention, there is provided a kit comprising a set of probes as described herein, an array and optionally one or more labels. In another aspect, there is provided an RT-MLPA kit comprising a set of reverse transcriptase primers as described herein, and appropriate ligases, buffers, and PCR primers. In the kits of the invention, a set of instructions will also typically be included.

The oligonucleotide primers and probes of the present invention have commercial applications in prognostic kits for the detection of the expression level of a gene, such as an X-linked gene such as XIST, in the tumor cells of a patient. A test kit according to the invention may comprise any of the oligonucleotide primers or probes according to the invention. Such a test kit may additionally comprise one or more reagents for use in cyclic polymerase mediated amplification reactions, such as DNA polymerases, nucleotides (dNTPs), buffers, and the like. A kit according to the invention may also include, for example, a lysing buffer for lysing cells contained in the specimen.

A test kit according to the invention may comprise a pair of oligonucleotide primers according to the invention and a probe comprising an oligonucleotide according to the invention. Advantageously, the kit further comprises additional means, such as reagents, for detecting or measuring the binding of the primers and probes of the present invention, and also ideally a positive and negative control.

The invention will now be further described by way of the following non-limiting examples.

EXAMPLE 1

Applicants have searched for gene expression patterns that correlate with cisplatin response using mammary tumors generated in a mouse model for BRCA1-associated breast cancer. Low expression of the Xist gene was found in most mouse tumors with cisplatin hypersensitivity, and this marker also predicts long recurrence-free survival of HER2-negative, stage-III breast cancer patients treated with platinum-based chemotherapy. In contrast, array-based expression profiling did not identify a single marker gene predicting docetaxel response, despite increased Abcb1 (P-glycoprotein) gene expression in 5/22 poor responders. Applicants’ results demonstrate that
identification of predictive markers typically requires high marker prevalence within the
tumor population, or quantitative analyses of known resistance mechanisms.

Using a mouse model for human breast cancer, Applicants show that relevant
markers, such as upregulation of the transporter gene Abcbl, will be missed if only
present in a sub-set of tumors. Applicants' results demonstrate why it is difficult to
identify predictive signatures. Typically, only if a marker correlates with response in a
substantial fraction of tumors, like Xist in Applicants' model, will it be detected. In
certain embodiments, Applicants' findings will help to select patients with high risk
breast cancer who may benefit from platinum-based therapy.

Applicants have employed a mouse model. In recent years, chemotherapy
responses have been investigated in a new generation of genetically engineered mouse
models (GEMMs) (Rottenberg and Jonkers, 2008). These models employ conditional,
tissue-specific activation of oncogenes and/or deletion of tumor suppressor genes in a
stochastic fashion (Jonkers and Berns, 2002). The resulting tumors closely mimic the
epithelial cancers in humans. Using the K14cre;Brca1F/F;p53F/F model for hereditary
breast cancer (Liu et al., 2007), it has been shown that these tumors acquire resistance to
classical and novel targeted anti-cancer drugs such as the topoisomerase I-targeting
camptothecin topotecan, the topoisomerase II-inhibiting anthracycline doxorubicin, and
the PARP inhibitor olaparib (Rottenberg et al., 2007; Rottenberg et al., 2008; Pajic et al.,
2009; Zander et al., 2010). Applicants have shown that the initial response of these
tumors is variable, as in human tumors, thus allowing them to correlate drug response
with gene expression. As the tumors arise in a genetically homogeneous background, the
genetic variability present in human tumors is absent in the mouse tumors. The tumors
are all very similar, as they start out with the ablation of the Brca1 and the p53 genes.
Differences between tumors are minimal (Rottenberg et al., 2007), and these small
differences make it comparatively easy to sort out which genes determine whether a
tumor responds to drug or not. An advantage of this model is that tumors can be
orthotopically transplanted into syngeneic, immunocompetent animals without losing
their molecular fingerprint, morphologic phenotype or drug sensitivity (Rottenberg et al.,
2007). For the identification of underlying resistance mechanisms, gene expression
analysis of matched samples of untreated and resistant tumors derived from the same
individual tumor was employed successfully (Rottenberg et al., 2007; Rottenberg et al,
2008; Zander et al., 2010; Pajic et al, 2009).
Applicants' analysis of initial drug response in this tumor model has yielded the result that Applicants have found that a low Xist expression correlates with high cisplatin sensitivity. In some embodiments, this preclinical finding is potentially relevant to the treatment of human breast cancer patients, as Applicants found that low XIST expression also correlates with benefit of intensive platinum-based chemotherapy. Applicants' results illustrate the use of GEMMs for the identification of predictive markers.

Applicants investigated whether predictive markers for chemotherapy benefit can be identified in a GEMM using genome-wide expression profiling. GEMMs should be ideal for this purpose, as they lack the profound genetic heterogeneity of tumors from human patients. The mice are inbred, all tumors originate from the targeted deletion of Brca1 and p53, and all differences between tumors originate from a limited number of random mutations in the period between the initiating deletions of Brca1 and p53 and the development of a mammary tumor. These additional mutations are responsible for the marked and stable differences in sensitivity to docetaxel and cisplatin that Applicants find in individual tumors.

Even in this genetically homogenous tumor system, Applicants did not find a signature predicting docetaxel response, using genome-wide expression profiling. This negative result is instructive, however, because it allowed Applicants to delineate what is required to get useful predictive signatures. In a collection of 22 poor docetaxel responders, 5 tumors contained a substantial increase in Abcbl RNA, known to be sufficient to cause drug resistance (Pajic et al, 2009). Nevertheless, this increase in Abcbl RNA was completely missed by 2 independent platforms measuring gene expression profiles. The Abcbl a RNA was missed, because the oligos in the platforms were poorly detecting this transcript. The Abcblb transcript was readily detected in the 5 tumors with elevated transcript levels, as long as these tumors were analyzed as a group. However, when the results were pooled with those from only 5 tumors without elevated Abcbl RNA, the positive result was completely lost. This shows why it is difficult to develop predictive markers, based on genome-wide expression arrays: generally, only if the response to a drug is primarily determined by the expression level of a gene in most tumors, can one expect that gene to show up in the array-based gene expression analyses.

Applicants found such a gene in analyzing the response of the murine tumors to cisplatin. The low Xist expression associated with tumors hypersensitive to cisplatin was present in 10 out of 12 tumors and therefore detectable in Applicants' array analysis.
Applicants’ results show that standard statistical tests will only detect predictive markers if they can explain therapy response in a high proportion of the samples analyzed. Special algorithms will be required to investigate genes that are altered in sample subsets, but these will probably yield a higher frequency of false positive hits as well. The problem remains that probes on the arrays are not sensitive enough to detect all relevant expression differences of genes causing therapy resistance. Gillet and co-workers found in a panel of cancer cell lines that the expression of the 380 "resistance-relevant" genes could only be reliably measured by quantitative PCR. For many genes the results obtained by microarrays were useless because of low sensitivity (Orina et al., 2009; Gillet and Gottesman, 2010).

Given all these hurdles in finding predictive markers for chemotherapy, it is encouraging that the low expression of XIST predicting high sensitivity to cisplatin in drug-naïve mouse tumors, also predicted an increased recurrence-free survival of high risk, primary breast cancer patients treated with intensive platinum-based chemotherapy. Although detected in a rather small group of 60 patients, the effect found is considerable. Intensive chemotherapy has largely been abandoned for the treatment of breast cancer, because for many patients the therapeutic benefit is limited (Crown, 2004). Nevertheless, several studies suggest that there are subgroups of patients that do benefit from this therapy, but the predictive tests to identify them are lacking (Cheng and Ueno, 2010; Rodenhuis, 2005; Zander and Kroger, 2005). Hence, in one embodiment, the analysis of XIST gene expression is a useful tool to decide whether intensive platinum-based chemotherapy should be considered as alternative therapy for patients with HER2-negative, high risk breast cancer. Not all patients with a low XIST expression that Applicants investigated benefited from the platinum-based therapy. Accordingly, an optimized cut-off for the level of XIST expression may be employed to increase the positive predictive value. Moreover, in some embodiments, a combination with other classifiers, such as CGH profiles (Vollebergh et al., 2010) may further optimize the predictive power of XIST gene expression as biomarker.

In 2002 a low XIST expression was reported in a recurrent tumor of an ovarian cancer patient (Huang et al., 2002). Subsequent work with a few cell lines suggested that a low XIST expression predicts paclitaxel resistance of ovarian cancer. However, this result was never validated in clinical trials and in Applicants’ mouse model, they do not find it. Intriguingly, Applicants observed a positive correlation between Xist expression
and the expression of X-linked genes that often escape X inactivation. Without being bound to theory, loss of Xi is the cause of low Xist expression. Of these X-linked genes, Jaridlc and Utx are histone H3 demethylases. Histone H3 methylation by Jaridla or Jaridlb is a feature of drug tolerant cells (Sharma et al., 2010; Roesch et al., 2010), and Jaridlc or Utx may also contribute to chromatin remodeling of platinum drug-tolerant cells.

The precise mechanism of XIST-mediated X inactivation is still under debate (Wutz, 2007). It was suggested that BRCA1 supports the localization of XIST RNA to the Xi, as the BRCA1-deficient cells or tumors examined had lost localized XIST RNA (Richardson et al., 2006; Silver et al., 2007; Ganesan et al., 2002). However, this hypothesis was challenged by others (Xiao et al., 2007). In several human breast cancer cell lines, a heterogenous picture of XIST expression and localization was found, which could not be directly linked to the BRCA1 status (Pageau et al., 2007). Despite the large Brcal deletion present in the mammary tumors of Applicants’ mouse model, Xist gene expression varies considerably. Variability of XIST expression was also present in those human breast cancers in which a BRCA1 mutation was found, or which were classified as BRCA1-like by comparative genomic hybridization (CGH) (Vollebergh et al., 2010) (Figure 6C). Intriguingly, all 12 tumors with BRCA1 promoter methylation showed a low XIST expression (Figure 6C).

As described herein, Applicants’ study shows that GEMMs that resemble breast cancer in humans are useful to investigate chemotherapy response prediction. Tools to identify predictive markers can be tested under controlled conditions, and targeted ablation of genes helps to dissect mechanisms of resistance. Ultimately, predictive markers identified in GEMMs may improve the clinical success rate for chemotherapy in humans.

1. Materials and Methods

   Mice

   KB1P mammary tumors were generated and genotyped as described (Liu et al., 2007). To produce KB1PM mammary tumors on a FVB/N genetic background, FVB.129P2-AbcblatmlBorAbcblbtmlBor, FVB-Tg(KRT14-cre)8Brn, FVB.129P2-Trp53tmlBmr, or FVB.129P2-BrcaltmlBmr mice were backcrossed on FVB/N animals for at least 8 generations (the first 5 generations using marker-assisted breeding) and eventually crossed to generate FVB.Cg-AbcblatmlBorAbcblbtmlBor Trp53tmlBmr
Brcaltm1Brn Tg(KRT14-cre)8Brn/A compound mice. Orthotopic transplantations, mammary tumor measurements and sampling were performed as explained previously (Rottenberg et al., 2007). In addition, KB1PM mammary tumors were transplanted into FVB/N animals. Deletion of Brca1 and p53 of orthotopically transplanted tumors was confirmed by PCR (absence of BrcaF5-13/F5-13;p53F2-10/F2-10 alleles and presence of BrcaA5-13/A5-13;p53A2-10/A2-10 alleles) as described (Liu et al., 2007). All experimental procedures on animals were approved by the Animal Ethics Committee of the Netherlands Cancer Institute.

Drugs and treatment of tumor-bearing animals

Docetaxel (Taxotere, 10 mg/ml in Tween80/ethanol/saline 20:13:67 vol/vol/vol; Aventis, Antony Cedex, France) was diluted with saline to 5 mg/ml before injection. Cisplatin (1 mg/ml in saline-mannitol) originated from Mayne Pharma (Brussels, Belgium). Doxorubicin (Adriblastina; Amersham Pharmacia Netherlands, Woerden, The Netherlands) was diluted to 1 mg/ml in saline (Braun, Emmer-Compascuum, The Netherlands). When mammary tumors reached a volume of 150-250mm3 (0.5 × length × width2) 25mg docetaxel per kg (days 0, 7, 14), 6mg cisplatin per kg (day 0) or 5mg doxorubicin per kg (day 0) were injected i.v. To avoid accumulating toxicity of repeated injections, an additional treatment was not given during the recovery time of 7 days in case the tumor responded to the treatment (tumor size <50% of the original volume, partial response). Treatment was continued once the tumor relapsed to its original size (100%). For tumors with a volume >50% after the recovery time, an additional treatment with the same dose as mentioned above was given.

Genome-wide expression profiling

RNA extraction, amplification, microarray hybridization using MEEBO arrays (Illumina BV, Eindhoven), data processing and statistical analyses were performed as described (Rottenberg et al., 2007; Zander et al., 2010; Pajic et al., 2010). For the gene expression analysis using 45K MouseWG-6 v2.0 BeadChips (Illumina, Eindhoven, The Netherlands) total RNA was processed according to the manufacture's instructions (www(dot)illumina(dot)com/products/mousewg_6_expression_beadchip_kits_v2(dot)ilm n). All probes with significant changes in expression (log2ratio > 1 or <-1) in <10% of the samples were filtered out. Also probes with missing data points in more than 10% of the hybridizations were excluded.
Reverse Transcription-Multiplex ligation-dependent probe amplification (RT-MLPA)

From snap frozen mouse tumors total RNA was isolated with Trizol (Invitrogen) according to the manufacturer's protocol. RNA from FFPE human breast cancers was isolated using the High Pure RNA Paraffin Kit (Roche, Woerden, The Netherlands) as described in the instruction manual. Reverse transcription, hybridization, ligation, PCR amplification and fragment analysis by capillary electrophoresis were carried out as reported previously (Pajic et al, 2009; Rottenberg et al, 2008). A list of the human-specific RT-MLPA probes is provided in the table in Figure 7A. For normalization of XIST gene expression, Applicants used the mean of 8 reference probes detecting B2M, GAPDH (2x), LDH, FAU, OAZ1, BIRC2 and ARHGDIA gene expression.

TaqMan low density arrays (TLDA)

Synthesis of cDNA from 1 µg total RNA in a 20 µl reaction volume was carried out using the High Capacity cDNA kit with RNAse inhibitor (Applied Biosystems, Foster City, CA, USA) as per the manufacturer's instructions. The reverse transcription conditions were as follows: 10 minutes at 25 °C, 120 minutes at 37 °C, 5 seconds at 85 °C. Following reverse transcription, cDNA was stored at 4 °C. Expression levels of 49 murine Abe transporter genes were measured using custom-made Taqman Low Density Arrays (Applied Biosystems, Foster City, CA, USA). cDNA was mixed with 2X Taqman Universal PCR Master Mix (Applied Biosystems, Foster City, CA, USA), loaded on the TLDA card (125ng per port), and run on an ABI Prism 7900 HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA) as per the manufacturer's instructions.

Histology

Tissues were fixed in 4% formaldehyde overnight, embedded in paraffin, and cut in 4 µm sections. After deparaffinization and rehydration, sections were stained with hematoxylin and eosin according to standard procedures.

Patients

In a previous study stage-III HER2-negative breast cancer patients were randomly selected from a large randomized controlled trial (RCT) performed in the Netherlands between 1993 and 1999 (Rodenhuis et al, 2003) and analyzed for aCGH (array CGH) classification (Vollebergh et al., 2010). Patients were randomized between conventional
chemotherapy (5*FEC: 5-fluorouracil 500 mg/m2, epirubicin 90 mg/m2, cyclophosphamide 500 mg/m2) and intensive platinum-based chemotherapy (4*FEC, followed by 1*CTC: cyclophosphamide 6000mg/m2, thiotepa 480mg/m2 and carboplatin 1600mg/m2). All trials described in this manuscript were approved by the Institutional Review Board of the Netherlands Cancer Institute. From this aCGH study Applicants selected those patients of whom BRCA1 -mutation status was available (n=60 from the previous study (Vollebergh et al, 2010) and 13 additionally analyzed). Of 13 patients no fragments could be visualized upon electrophoresis due to poor FFPE RNA quality, resulting in a subset of 60 patients of whom RT-MLPA data was available. For statistical analyses recurrence-free survival (RFS) was calculated from randomization to appearance of local or regional recurrence, metastases or to death from any cause. All other events were censored. Differences between groups of interest were tested using Fisher's exact tests. Survival curves were generated using the Kaplan-Meier method and compared using log-rank tests. Hazard ratios (HR) were calculated using Cox-proportional hazards regression.

Accession Numbers

The microarray data reported herein have been deposited in the Array Express database, www(dot)ebi(dot)ac(dot)uk/arrayexpress (accession no. E-MTAB-413 [Illumina] and E-MTAB-415 [MEEBO]).

2. Brcal-/-;p53-/- (KB IP) mammary tumors show individual and reproducible differences in docetaxel or cisplatin sensitivity.

Applicants have previously shown that individual KB IP mammary tumors differ substantially in their response to docetaxel (Rottenberg et al., 2007). The response to cisplatin varied as well: although all tumors were sensitive to cisplatin, the time until relapse differed between tumors (Rottenberg et al., 2007). To exploit these inter-tumoral differences, Applicants analyzed docetaxel or cisplatin responses of 43 individual tumors. Six tumors were analyzed when they first occurred, the others after orthotopic transplantation into syngeneic mice, as described earlier (Rottenberg et al., 2007). The advantage of orthotopic grafting is that tumors derived from the same original tumor can be treated with different drugs. This allowed Applicants to collect response data for docetaxel, cisplatin and doxorubicin for 36 individual tumors (Figure 2).
The correlation of drug sensitivities with characteristics of a particular tumor is only possible if the responses are reproducible. Applicants therefore explored the heterogeneity within a single tumor by orthotopic transplantation of multiple tumor fragments (Figure 1). For this purpose, 3 animals carrying orthotopically transplanted fragments of the same spontaneous tumor were treated with the maximum tolerable dose of docetaxel on days 0, 7 and 14.

Also for cisplatin, Applicants confirmed that the time to relapse is reproducible (Fig. IB). Fragments from tumor T9 produced tumors that relapsed early, whereas tumors arising from tumor T13 fragments relapsed late. Hence, this tumor model can be used to correlate initial docetaxel or cisplatin responses with other tumor characteristics, such as gene expression levels.

3. Low expression of the Xist gene correlates with high cisplatin sensitivity of KB IP tumors and predicts benefit of platinum-based chemotherapy in patients with high risk primary breast cancer.

Applicants found a bimodal distribution of the time to relapse, not only for docetaxel, but also for cisplatin (Figure 5A). All KB IP tumors were cisplatin sensitive, but 23 tumors relapsed already within 39 days, whereas 12 tumors only grew back to 100% after 49 days. When Applicants stratified the gene expression profiles of the untreated tumors based on their cisplatin sensitivity, they found a low expression of the Xist gene to correlate significantly with cisplatin hypersensitivity on 2 independent gene expression platforms (Figure 5B).

The physiological role of the non-coding RNA Xist is to coat one X chromosome of female cells in cis and subsequently trigger chromatin remodeling to form the heterochromatic Barr body (condensed inactivated X chromosome [Xi]). XIST is transcribed exclusively from the Xi to achieve equal X-linked gene dosage between the sexes. The analysis of genes correlating with a low Xist expression in Applicants' tumor model revealed a reduced expression of 3 other X-linked genes: Utx, Jaridlc, and Eif2s3x (Figure 6A). As all of these genes are known to escape X inactivation (Lopes et al., 2010), they are independent markers for the loss of the Xi.

Given the high frequency of reduced Xist expression in cisplatin hypersensitive mouse tumors, Applicants tested whether XIST expression could serve as a biomarker to predict response to platinum-based chemotherapy in human breast cancer. For this
purpose, Applicants took tumor samples of 60 stage-III, HER2-negative breast cancer patients who had been randomized between two treatment arms: intensive platinum-based chemotherapy, or a standard anthracycline-based regimen (conventional chemotherapy) (Rodenhuis et al., 2003). The patient information is summarized in the tables Figure 7B+C. To quantify XIST gene expression levels of FFPE material (>60% tumor cells), Applicants used RT-MLPA including 2 independent probes hybridizing to the exon 2-3 or exon 4-5 boundary of XIST cDNA (Figure 7A). The cut-off to determine low XIST expression was defined as 2 times the SD below the average expression of normal breast tissue (Figure 5B). Analysis of the recurrence-free survival (RFS) showed that patients with a low XIST expression significantly benefited from the intensive platinum-based therapy compared to conventional chemotherapy: the 5-year RFS increased from 37% to 75% (Figure 5C, adjusted hazard-ratio: 0.30, 95%CI: 0.1-0.82 for the probe of exon 4-5, Figure 7C). In patients with XIST gene expression above the cut-off, no significant survival benefit was observed of platinum-based chemotherapy (5-year RFS 33% both treatment arms, Figure 5C; adjusted hazard-ratio: 0.81, 95%CI: 0.23-2.89 for the exon 4-5 probe, Figure 7C). The effect of treatment on RFS was not significantly different between the XIST subgroups based on homogeneity of the hazard ratios (p-interaction: 0.24 exon 4-5, Figure 7C), which is likely due to the relative small numbers in Applicants’ series. Analyses with the exon 2-3 probe confirmed those obtained with the exon 4-5 probe (Figure 5C and Figure 7B+C).

4. High prevalence of a predictive marker is required for its detection.

Xist was readily identified as predictive marker for cisplatin sensitivity in Applicants’ mouse model by SAM, but Abcblb was not identified as a predictive marker for docetaxel response (Figure 4). Although sensitivity of Applicants’ ordered arrays for Abcbla is low, Abcblb transcripts are readily detected. When Applicants analyzed only the tumors with an intrinsically high Abcbl expression (T8*con, T9*con, T15*con, T26*con and T41*con) versus the 21 docetaxel-sensitive tumors as defined in Figure 4A, Abcblb was one of the most significantly increased genes on both the MEEBO and Illumina gene expression platforms. Also the TLDA expression data showed a significant difference for Abcbla and Abcblb when only the 5 poor responders T8*con, T9*con, T15*con, T26*con and T41*con were compared with the docetaxel-sensitive tumors (Abcbla: P < 0.0064; Abcblb: P < 0.0043; Mann-Whitney U test). However,
since increased expression of the Abcbl genes is only found in a subgroup of the poor
docetaxel responders, this significance is lost when samples with other docetaxel
resistance mechanisms are added (Figure 8). In fact, addition of 5 samples without
Abcbl upregulation suffices to dilute the Abcbl signal below significance.

In contrast to Abcbl in the case of docetaxel treatment, the prevalence of low
Xist expression was high in cisplatin hypersensitive tumors: 11 (MEEBO platform) or 10
(Illumina platform) out of the 12 showed Xist gene expression below the median (Figure
8). It is therefore not surprising that Xist was picked up by SAM.

Reference List

lines: Forensic bioinformatics and reproducible research in high-throughput biology.

Berns, K., Horlings, H.M., Hennessy, B.T., Madiredjo, M., Hijmans, E.M.,
Beelen, K., Linn, S.C, Gonzalez-Angulo, A.M., Stemke-Hale, K., Hauptmann, M.,
 genetic approach identifies the PI3K pathway as a major determinant of trastuzumab

Bonnefoi, H., Potti, A., Delorenzi, M., Mauriac, L., Campone, M., Tubiana-
Hulin, M., Petit, T., Rouanet, P., Jassem, J., Blot, E., Becette, V., Farmer, P., Andre, S.,
Validation of gene signatures that predict the response of breast cancer to neoadjuvant
chemotherapy: a substudy of the EORTC 10994/BIG 00-01 clinical trial. Lancet Oncol.
8, 1071-1078.


Borst, P. and Wessels, L. Do predictive signatures really predict response to

Ref Type: In Press

BRCA1 downregulation leads to premature inactivation of spindle checkpoint and


Pageau,G.J., Hall,L.L., and LawrenceJ.B. (2007). BRCA1 does not paint the inactive X to localize XIST RNA but may contribute to broad changes in cancer that impact XIST and Xi heterochromatin. J. Cell Biochem. 100, 835-850.


EXAMPLE 2

Markers Previously Correlated with Cisplatin Hypersensitivity

Patients and Methods

Patients:

In a previous study, stage-Ill HER2-negative breast cancer patients were randomly selected from a large randomized controlled trial (RCT) performed in the Netherlands between 1993 and 1999 and analyzed for aCGH classification. Patients were randomized between conventional chemotherapy (5*FEC: 5-fluorouracil 500mg/m2, epirubicin 90mg/m2, cyclophosphamide 500mg/m2) and intensive platinum-based chemotherapy (PB-chemotherapy; 4*FEC, followed by 1*CTC: cyclophosphamide 6000mg/m2, thiotepa 480mg/m2 and carboplatin 1600mg/m2). All trials described in this manuscript were approved by the Institutional Review Board of the Netherlands Cancer Institute. From this aCGH study, Applicants selected those patients of whom BRCA1-mutation status was available (n=60 were already analyzed for the previous study and from this study, Applicants additionally analyzed 13 patients for BRCA1-mutation status). Of 13 patients, no fragments could be visualized upon electrophoresis due to poor FFPE RNA quality, resulting in a subset of 60 patients of whom RT-MLPA data was available.

Reverse Transcription-Multiplex ligation-dependent probe amplification (RT-MLPA)

RNA from FFPE human breast cancers was isolated using the High Pure RNA Paraffin Kit (Roche, Woerden, The Netherlands) as described in the instruction manual. Reverse transcription, hybridization, ligation, PCR amplification and fragment analysis by capillary electrophoresis were carried out as reported previously. In short, probes were designed in such manner that they spanned exon borders to avoid detection of contaminating DNA in the samples. Total RNA isolated from FFPE tissue was converted to cDNA using MMLV reverse transcriptase and probe-specific primers that are less than 20 nucleotides downstream of their probes target. A mixture of hemi-probes was hybridized to the generated cDNA. In a subsequent step, the two hemi-probes were annealed by a heat stable ligase. The annealed probes, which have identical 5'- and 3'-prime ends, were amplified by PCR using a single fluorescent primer pair. The PCR reaction could then be quantified by capillary sequencer, 3730 DNA Analyzer (AB,
USA). A list of the human-specific RT-MLPA probes is provided in the table depicted in Figure 7A.

Normalization and Statistical Analyses

For normalization of gene expression of genes of interest, Applicants normalized to the internal reference genes present in the RT-MLPA mix (LDHA, 2 probes for GAPDH, β2M, ARHGDIA, FAU, OAZ1 and BIRC2). RT-MLPA analyses of a pool of RNA isolated from FFPE normal breast tissue of 8 different patients were performed in 8 different measurements (one measurement is the average of 3 independent MLPA reactions). To divide the gene expression levels of genes of interest into high and low, the average gene expression plus two times the standard deviation of the normal breast tissue pool was used as cut-off (Figure 9).

For statistical analyses, recurrence-free survival (RFS) was calculated from randomization to appearance of local or regional recurrence, metastases or to death from any cause. All other events were censored. Differences between groups of interest were tested using Fisher's exact tests. Survival curves were generated using the Kaplan-Meier method and compared using log-rank tests. Hazard ratios (HR) were calculated using Cox-proportional hazards regression.

Results

Applicants investigated whether gene expression (high versus low) was associated with the presence of a BRCA1-like CGH pattern or Xist low expression. Both features have previously been associated with an improved outcome on intensified platinum-based chemotherapy215.

Applicants found a significant association of ATM and RAD50 with both the BRCA1-like CGH pattern and a low Xist expression (Table 1). In this example, the number of patients with a tumor expressing high RAD50 levels was too low to study treatment effects on RFS (only n=4 and n=1 patients received intensified PB-chemotherapy).

Analysis of the recurrence-free survival (RFS) showed that patients with a low ATM expression significantly benefited from the platinum-containing therapy compared
to conventional chemotherapy: 5-year RFS 87% compared to 36% (Figure 10A, adjusted hazard-ratio: 0.14, 95%CI: 0.0-0.64, Table 2). In patients with ATM gene expression above the cut-off, no survival benefit was observed of platinum-containing chemotherapy compared to conventional chemotherapy (5-year RFS 33% and 36% for conventional, respectively intensified PB-chemotherapy, Figure 10B; adjusted hazard-ratio: 0.79, 95%CI: 0.25-2.55, Table 2). However, the effect of treatment on RFS was not statistically significantly different between the ATM subgroups (p-interaction: 0.08, Table 2); this is likely due to the relative small numbers in this series. Patient characteristics of high versus low ATM expression are shown in Table 3.
Table 1. Association of RAD50 and ATM with the BRCA1-likeCGH pattern and Xist low expression.

<table>
<thead>
<tr>
<th></th>
<th>Non-BRCA1-likeCGH</th>
<th>BRCA1-likeCGH</th>
<th>p-value</th>
<th>High Xist1</th>
<th>Low Xist1</th>
<th>p-value</th>
<th>High Xist2</th>
<th>Low Xist2</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>High ATM</td>
<td>16</td>
<td>7</td>
<td>&lt;0.01</td>
<td>15</td>
<td>8</td>
<td></td>
<td>13</td>
<td>10</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Low ATM</td>
<td>12</td>
<td>25</td>
<td></td>
<td>6</td>
<td>31</td>
<td></td>
<td>5</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>High RAD50_1</td>
<td>9</td>
<td>4</td>
<td>0.12</td>
<td>9</td>
<td>4</td>
<td>&lt;0.01</td>
<td>8</td>
<td>5</td>
<td>0.01</td>
</tr>
<tr>
<td>Low RAD50_1</td>
<td>19</td>
<td>28</td>
<td></td>
<td>12</td>
<td>35</td>
<td></td>
<td>10</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>High RAD50_2</td>
<td>6</td>
<td>1</td>
<td>0.04</td>
<td>7</td>
<td>0</td>
<td>&lt;0.01</td>
<td>7</td>
<td>0</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Low RAD50_2</td>
<td>22</td>
<td>31</td>
<td></td>
<td>14</td>
<td>39</td>
<td></td>
<td>11</td>
<td>42</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Multivariate Cox proportional-hazard analysis of the risk of recurrence (RFS) and ATM expression

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. Events / No. patients</th>
<th>Hazard Ratio</th>
<th>95% CI</th>
<th>p values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ATM expression</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>18 / 37</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>15 / 23</td>
<td>0.87</td>
<td>0.30 - 2.46</td>
<td>0.79</td>
</tr>
<tr>
<td><strong>ATM Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional chemotherapy</td>
<td>16 / 22</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensified PB-chemotherapy</td>
<td>2 / 15</td>
<td>0.14*</td>
<td>0.03 - 0.64</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>ATM High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional chemotherapy</td>
<td>6 / 9</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intensified PB-chemotherapy</td>
<td>9 / 14</td>
<td>0.79*</td>
<td>0.25 - 2.55</td>
<td>0.70</td>
</tr>
</tbody>
</table>

All analyses shown were adjusted for: number of positive lymph nodes (4-9 vs. >10); pathological T-stage (1 or 2 vs. 3); histologic grade (I or II vs. III). Homogeneity of both hazard ratios was not rejected based on an interaction term with *p value = 0.08; Number of events is not equal for all variables, since some patients have missing data; maximum missing variables (i.e. events) is 1/33.

**Abbreviations:** CI, confidence interval; PB, platinum-based.
Table 3. Patient characteristics distributed by ATM expression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total</th>
<th>ATM expression</th>
<th>p values*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n  %</td>
<td>Low n %</td>
<td>High n %</td>
</tr>
<tr>
<td>Total</td>
<td>60 100.0</td>
<td>37 61.7</td>
<td>23 38.3</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional Chemotherapy</td>
<td>31 51.7</td>
<td>22 59.5</td>
<td>9 39.1</td>
</tr>
<tr>
<td>Intensified PB-chemotherapy</td>
<td>29 48.3</td>
<td>15 40.5</td>
<td>14 60.9</td>
</tr>
<tr>
<td>Age in categories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 40 years</td>
<td>26 43.3</td>
<td>20 54.1</td>
<td>6 26.1</td>
</tr>
<tr>
<td>&gt; 40 years</td>
<td>34 56.7</td>
<td>17 45.9</td>
<td>17 73.9</td>
</tr>
<tr>
<td>Type of surgery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast conserving therapy</td>
<td>16 26.7</td>
<td>8 21.6</td>
<td>8 34.8</td>
</tr>
<tr>
<td>Mastectomy</td>
<td>44 73.3</td>
<td>29 78.4</td>
<td>15 65.2</td>
</tr>
<tr>
<td>Tumour classification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 or T2</td>
<td>51 85.0</td>
<td>31 83.8</td>
<td>20 87.0</td>
</tr>
<tr>
<td>T3</td>
<td>9 15.0</td>
<td>6 16.2</td>
<td>3 13.0</td>
</tr>
<tr>
<td>No. of positive lymph nodes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-9</td>
<td>37 61.7</td>
<td>24 64.9</td>
<td>13 56.5</td>
</tr>
<tr>
<td>≥ 10</td>
<td>23 38.3</td>
<td>13 35.1</td>
<td>10 43.5</td>
</tr>
<tr>
<td>Histologic grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1+ II</td>
<td>15 25.0</td>
<td>9 24.3</td>
<td>6 26.1</td>
</tr>
<tr>
<td>III</td>
<td>41 68.3</td>
<td>26 70.3</td>
<td>15 65.2</td>
</tr>
<tr>
<td>Not determined</td>
<td>4 6.7</td>
<td>2 5.4</td>
<td>2 8.7</td>
</tr>
<tr>
<td>Hormone receptor status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER and PR negative (&lt;10)</td>
<td>46 76.7</td>
<td>27 73.0</td>
<td>19 82.6</td>
</tr>
<tr>
<td>ER or PR positive (&gt;10)</td>
<td>12 20.0</td>
<td>8 21.6</td>
<td>4 17.4</td>
</tr>
<tr>
<td>Unknown</td>
<td>2 3.3</td>
<td>2 5.4</td>
<td>0 0.0</td>
</tr>
<tr>
<td>P53 status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative (&lt;10)</td>
<td>32 53.3</td>
<td>21 56.8</td>
<td>11 47.8</td>
</tr>
<tr>
<td>Positive (&gt;10)</td>
<td>28 46.7</td>
<td>16 43.2</td>
<td>12 52.2</td>
</tr>
</tbody>
</table>

* p values were calculated using the Fisher's exact test; patients with unknown values were omitted from analyses. Abbreviations: PB, platinum-based; ER, estrogen receptor; PR, progesterone receptor.

Reference List


EXAMPLE 3

The following data were obtained using the XIST 1 probe.

Combination of Xist and BRCA-like CGH status as a predictive marker for platinum-based chemotherapy

Kaplan-Meier curves of the combination of BRCA-like CGH status and Xist status.

The combination of low Xist expression and BRCA-like CGH profiles separates the groups of patients better both in univariate Kaplan-Meier curves as well as in the Cox proportional hazards model. The combination marker is defined in this embodiment as patients that have both a BRCA-like CGH profile and low Xist expression. Figures 11 and 12 display the better separation of the Kaplan-Meier curves when Xist and BRCA-like status are considered together.

Cox proportional hazards model of the combination of BRCA-like CGH status and low Xist expression

The univariate analyses shows improvement of the Xist marker when combined with BRCA-like CGH status. Applicants tested the marker in a multivariate model that corrects for T stage, histological grade, and number of lymph nodes. Low Xist patients have a hazard rate of 0.27 (95% confidence interval: 0.10-0.73, p=0.01) of recurrence on high dose chemotherapy, whereas high Xist patients have a hazard rate of 0.82 (95% confidence interval: 0.31-2.18, p=0.69) on high dose chemotherapy. This means that patients with low Xist expression benefit greatly from high dose chemotherapy, whereas high dose chemotherapy does not confer a survival benefit in Xist high patients. To further investigate a difference in effect size between both groups, Applicants did an interaction test.

For BRCA-like CGH patients in this series the hazard rate of recurrence after high dose chemotherapy is 0.19 (95% confidence interval: 0.05-0.65, p=0.01), versus 0.77 (0.32-1.82, p=0.55) in non-BRCA-like patients that received high dose chemotherapy. This means that also BRCA-like CGH patients benefit from high dose chemotherapy and patients with a
non-BRCA-like tumor do not. To further investigate a differential effect in the marker negative and marker positive group, Applicants did an interaction test.

When Applicants combine low Xist expression and BRCA-like status and compare the effect of high dose chemotherapy with patients that do not have this combination of markers, Applicants find that marker positive patients have a benefit of high dose chemotherapy. The hazard rate of recurrence is 0.12 (95% confidence interval: 0.03-0.55, p=0.01) in the marker positive group. The marker negative group does not benefit from high dose chemotherapy with a hazard rate of recurrence of 0.84 (95% confidence interval: 0.37-1.93, p=0.69). The differential effect is statistically significant, determined by a test of interaction (p=0.02).

Combining these models indicates that a combination of BRCA-like status and Xist expression is at least as good as using the BRCA-like CGH profile alone, but may have a biologically/clinically relevant population.

Picture of survival of all patients, split per group

Some of the groups of patients are small, therefore Applicants present a plot of the survival data within all the different marker populations (Fig. 13). In this graph, a shift towards long survival on high dose within a group of patients selected by a certain marker represents benefit of high dose chemotherapy within that group. Applicants conclude that there are no obvious outliers in prognosis in these groups (also indicated by the cox models). Visually, only the BRCA-like, Xist low group demonstrates such a shift. The BRCA-like, Xist high group is small, but all patients treated with high dose chemotherapy had an early recurrence. Furthermore, survival times were similar as the conventionally treated patients of this group. In the Xist low, non-BRCA like group, no shift is noted between the two different regimens.

Combined with the animal data, these data suggest that when employing BRCA-like CGH status to select patients that benefit from high dose chemotherapy, Xist expression should be used in combination with BRCA-like CGH status to select the group of patients that benefit from high dose chemotherapy.

***

Having thus described in detail embodiments of the present invention, it is to be understood that the invention defined by the above paragraphs is not to be limited to particular details set forth in the above description as many apparent variations thereof are possible without departing from the spirit or scope of the present invention.
Each patent, patent application, and publication cited or described in the present application is hereby incorporated by reference in its entirety as if each individual patent, patent application, or publication was specifically and individually indicated to be incorporated by reference.
WHAT IS CLAIMED IS:

1. A method of evaluating chemotherapy sensitivity in a patient in need thereof, comprising:
   (a) measuring expression levels of one or more X-linked genes in the patient; and
   (b) comparing the expression levels of the one or more X-linked genes in (a) with the expression levels of one or more reference X-linked genes, wherein the one or more reference X-linked genes are from a control sample,

wherein a reduction in the expression of the one or more X-linked genes in comparison to the one or more reference X-linked genes is indicative of chemotherapy sensitivity in the patient.

2. The method of claim 1, wherein the patient has breast cancer, ovarian cancer, cervical cancer, testicular cancer, head and neck cancer, lung cancer, colorectal cancer, or bladder cancer.

3. The method of any preceding claim, wherein the chemotherapy sensitivity is sensitivity to treatment with a platinum-based chemotherapeutic agent.

4. The method of any preceding claim, wherein the platinum-based chemotherapeutic agent is selected from the group consisting of: cisplatin, carboplatin, oxaliplatin, satraplatin, picoplatin, and triplatin tetranitrate.

5. The method of claim 1 or claim 2, wherein the chemotherapy sensitivity is sensitivity to treatment with a PARP inhibitor.

6. The method of claim 5, wherein the PARP inhibitor is selected from the group consisting of: ABT-888, olaparib, CEP-9722, AG014699, BSI-201, MK4827, LT-673, GPI21016, and INO-1001.

7. The method of any preceding claim, wherein the X-linked gene is XIST.

8. The method of any of claims 1-6, wherein the X-linked gene is selected from the group consisting of: Utx, Jarid1c, and Eif2s3x.
9. The method of any preceding claim, wherein further to a reduction in the one or more X-linked gene expression levels, there is a reduction in the expression level of one or more genes selected from the group consisting of: ATM, Rad50, and Ube3b.

10. The method of any preceding claim, wherein the patient has breast cancer.

11. The method of claim 10, wherein the breast cancer is HER-2 negative, advanced breast cancer.

12. The method of any preceding claim, wherein the cancer is BRCA-1 deficient.

13. The method of any of claims 1-11, wherein the cancer is non-BRCA-1 deficient.

14. The method of claim 12 or claim 13, wherein the cancer is breast cancer or ovarian cancer.

15. The method of any preceding claim, further comprising conducting comparative genomic hybridization (CGH) with DNA from the patient.

16. The method of claims 1-2 and 7-15, wherein the chemotherapy sensitivity is sensitivity to treatment with a DNA-cross-linking agent.

17. A microarray comprising a plurality of polynucleotide probes each complementary and hybridizable to a sequence in a different gene that is an X-linked gene that is a marker for chemosensitivity in a patient that has cancer.

18. The microarray of claim 17, wherein the plurality of probes is at least 70 %, at least 80 %, at least 90 %, at least 95 %, or at least 98 % of the probes on the microarray.

19. The microarray of claim 17 or claim 18, wherein the X-linked gene that is a marker for chemosensitivity is selected from the group consisting of XIST, Utx, Jarid1c, and Eif2s3x.
20. The microarray of any of claims 17-19, further comprising a plurality of polynucleotide probes each complementary and hybridizable to a sequence in a different gene that is a marker for chemosensitivity in a patient that has cancer, wherein the gene is not an X-linked gene.

21. The microarray of claim 20, wherein the gene that is not an X-linked gene is selected from the group consisting of: ATM, Rad50, and Ube3b.

22. A kit, comprising at least one pair of primers specific for an X-linked gene that is a marker for chemosensitivity in a patient that has cancer, at least one reagent for amplification of the X-linked gene, and instructions for use.

23. The kit of claim 22, wherein the primers are specific for an X-linked gene selected from the group consisting of XIST, Utx, Jarid1c, and Eif2s3x.

24. The kit of claim 23, wherein the primers are specific for XIST.

25. The kit of any of claims 22-24, wherein the marker for chemosensitivity is a marker for sensitivity to a platinum-based drug.

26. The kit of claim 25, wherein the platinum-based drug is selected from the group consisting of: cisplatin, carboplatin, oxaliplatin, satraplatin, picoplatin, and triplatin tetranitrate.

27. The kit of any of claims 22-24, wherein the marker for chemosensitivity is a marker for sensitivity to a PARP inhibitor.

28. The kit of claim 27, wherein the PARP inhibitor is selected from the group consisting of: ABT-888, olaparib, CEP-9722, AG014699, BSI-201, MK4827, LT-673, GPI21016, and INO-1001.

29. The kit of any of claims 22-28, wherein the kit is a PCR kit.
30. The kit of any of claims 22-28, wherein the kit is an MLPA kit.

31. The kit of any of claims 22-28, wherein the kit is an RT-MLPA kit.

32. The method of any of claims 1-16, wherein the level of expression of said genes is measured by determination of their level of transcription, using a DNA array.

33. The method of any of claims 1-16, wherein the level of expression of said genes is measured by determination of their level of transcription, using quantitative RT-PCR.

34. The method of any of claims 1-16, wherein the level of expression of said genes is measured in a tumor sample from the patient.

35. The method of claim 34, wherein the tumor sample is a breast tumor sample.

36. The method of claim 7, wherein the expression levels of the XIST gene in (a) are measured in a tumor sample from the patient.

37. The method of claim 36, wherein the patient tumor sample is a breast tumor sample.

38. The method of any of claims 1-16, wherein the X-linked gene escapes XIST-mediated inactivation.

39. The method of claim 38, wherein the expression levels of the X-linked gene in (a) are measured in a tumor sample from the patient.

40. The method of claim 39, wherein the patient tumor sample is a breast tumor sample.

41. The method of any of claims 1-6 and 8-16, wherein the level of expression of said genes is measured by determination of their level of translation.

42. The method of claim 41, wherein the level of translation is measured by determination of the amount of translated protein.
43. The method of claim 41, wherein the level of translation is measured by determination of the activity of translated protein.

44. The method of claim 42, wherein the translated protein is selected from the group consisting of: Utx, Jaridlc, and Eif2s3x.

45. The method of claim 43, wherein the translated protein is selected from the group consisting of: Utx, Jaridlc, and Eif2s3x.

46. The method of claim 9, wherein the reduction in the expression level of ATM, Rad50, and/or Ube3b is measured by determining a reduction in the amount of ATM, Rad50, and/or Ube3b protein.

47. The method of claim 9, wherein the reduction in the expression level of ATM, Rad50, and/or Ube3b is measured by determining a reduction in the activity of ATM, Rad50, and/or Ube3b protein.

48. The method of claim 9, wherein the reduction in the expression level of ATM, Rad50, and/or Ube3b is due to one or more inactivating mutations in the ATM, Rad50, and/or Ube3b gene.

49. The method of claim 9, wherein the reduction in the expression level of ATM, Rad50, and/or Ube3b is due to protein dysfunction of the ATM, Rad50, and/or Ube3b protein.

50. The method of claim 15, wherein the patient has breast cancer.

51. The method of claim 50, wherein the CGH is a BRCA-CGH.

52. The method of claim 51, wherein the BRCA-CGH is BRCA-1-CGH.

53. The method of claim 52, wherein the cancer is BRCA-1 deficient.

54. The method of claim 53, wherein the X-linked gene is XIST.
55. The method of claim 54, wherein the patient is selected for treatment with high-dose alkylating therapy.

56. The method of any of claims 1-16, wherein a reduction in the expression of the one or more X-linked genes is due to one or more inactivating mutations in the X-linked gene.
Figure 1

A

Orthotopic transplantation of tumor fragments (~1mm) into wt animals

Tumor size ~200mm³, therapy start (MTD)

B

Orthotopic transplantation of tumor fragments (~1mm) into wt animals

Tumor size ~200mm³, therapy start (MTD)

Graphs T26, T27, T28, T9, T13, T13
**T30** control  
**T30** docetaxel  
**T30** doxorubicin  
**T30** cisplatin

relative tumor size

0 20 40 60 80 100 120  
days

---

**T31** control  
**T31** docetaxel

relative tumor size

0 20 40 60 80 100 120  
days

---

SUBSTITUTE SHEET (RULE 26)
### FIGURE 3

Overview of treatment responses of KB1P tumors.

<table>
<thead>
<tr>
<th>no</th>
<th>docetaxel (d)</th>
<th>doxorubicin (d)</th>
<th>cisplatin (d)</th>
<th>classification drug sensitivity</th>
<th>morphology</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>poor</td>
<td>N/A</td>
</tr>
<tr>
<td>T2</td>
<td>29</td>
<td>N/A</td>
<td>N/A</td>
<td>good</td>
<td>N/A</td>
</tr>
<tr>
<td>T3</td>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>poor</td>
<td>N/A</td>
</tr>
<tr>
<td>T4</td>
<td>17</td>
<td>N/A</td>
<td>N/A</td>
<td>good</td>
<td>N/A</td>
</tr>
<tr>
<td>T5</td>
<td>23</td>
<td>N/A</td>
<td>N/A</td>
<td>good</td>
<td>N/A</td>
</tr>
<tr>
<td>T6</td>
<td>25</td>
<td>N/A</td>
<td>N/A</td>
<td>good</td>
<td>N/A</td>
</tr>
<tr>
<td>T7con</td>
<td>0</td>
<td>39</td>
<td>71</td>
<td>poor</td>
<td>high</td>
</tr>
<tr>
<td>T8con</td>
<td>0</td>
<td>16</td>
<td>49</td>
<td>poor</td>
<td>high</td>
</tr>
<tr>
<td>T9con</td>
<td>0</td>
<td>12</td>
<td>29</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T10con</td>
<td>0</td>
<td>40</td>
<td>32</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T11con</td>
<td>0</td>
<td>24</td>
<td>31</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T12con</td>
<td>27</td>
<td>47</td>
<td>28</td>
<td>good</td>
<td>moderate</td>
</tr>
<tr>
<td>T13con</td>
<td>0</td>
<td>28</td>
<td>72</td>
<td>poor</td>
<td>high</td>
</tr>
<tr>
<td>T14con</td>
<td>0</td>
<td>31</td>
<td>30</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T15con</td>
<td>0</td>
<td>0</td>
<td>84</td>
<td>poor</td>
<td>high</td>
</tr>
<tr>
<td>T16con</td>
<td>0</td>
<td>45</td>
<td>32</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T17con</td>
<td>30</td>
<td>57</td>
<td>35</td>
<td>good</td>
<td>moderate</td>
</tr>
<tr>
<td>T18con</td>
<td>32</td>
<td>41</td>
<td>74</td>
<td>good</td>
<td>high</td>
</tr>
<tr>
<td>T19con</td>
<td>0</td>
<td>28</td>
<td>32</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T20con</td>
<td>22</td>
<td>29</td>
<td>23</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T21con</td>
<td>21</td>
<td>55</td>
<td>&gt;100</td>
<td>good</td>
<td>high</td>
</tr>
<tr>
<td>T22con</td>
<td>39</td>
<td>&gt;100</td>
<td>39</td>
<td>good</td>
<td>moderate</td>
</tr>
<tr>
<td>T23con</td>
<td>0</td>
<td>28</td>
<td>51</td>
<td>poor</td>
<td>high</td>
</tr>
<tr>
<td>T24con</td>
<td>42</td>
<td>24</td>
<td>18</td>
<td>good</td>
<td>moderate</td>
</tr>
<tr>
<td>T25con</td>
<td>0</td>
<td>35</td>
<td>33</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T26con</td>
<td>0</td>
<td>14</td>
<td>35</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T27con</td>
<td>0</td>
<td>23</td>
<td>35</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T28con</td>
<td>33</td>
<td>40</td>
<td>24</td>
<td>good</td>
<td>moderate</td>
</tr>
<tr>
<td>T29con</td>
<td>11</td>
<td>31</td>
<td>N/A</td>
<td>good</td>
<td>N/A</td>
</tr>
<tr>
<td>T30con</td>
<td>30</td>
<td>23</td>
<td>13</td>
<td>good</td>
<td>moderate</td>
</tr>
<tr>
<td>T31con</td>
<td>52</td>
<td>N/A</td>
<td>N/A</td>
<td>good</td>
<td>N/A</td>
</tr>
<tr>
<td>T32con</td>
<td>16</td>
<td>42</td>
<td>53</td>
<td>good</td>
<td>high</td>
</tr>
<tr>
<td>T33con</td>
<td>0</td>
<td>26</td>
<td>20</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T34con</td>
<td>29</td>
<td>33</td>
<td>26</td>
<td>good</td>
<td>moderate</td>
</tr>
<tr>
<td>T35con</td>
<td>21</td>
<td>32</td>
<td>&gt;100</td>
<td>poor</td>
<td>high</td>
</tr>
<tr>
<td>T36con</td>
<td>0</td>
<td>33</td>
<td>&gt;100</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T37con</td>
<td>0</td>
<td>39</td>
<td>26</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T38con</td>
<td>36</td>
<td>0</td>
<td>52</td>
<td>poor</td>
<td>high</td>
</tr>
<tr>
<td>T39con</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T40con</td>
<td>0</td>
<td>24</td>
<td>35</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T41con</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>poor</td>
<td>moderate</td>
</tr>
<tr>
<td>T42con</td>
<td>26</td>
<td>28</td>
<td>77</td>
<td>good</td>
<td>high</td>
</tr>
<tr>
<td>T43con</td>
<td>13</td>
<td>18</td>
<td>73</td>
<td>good</td>
<td>high</td>
</tr>
</tbody>
</table>
Figure 4

A

B

C

Figure 4
### Figure 7

**A. Sequence of oligonucleotides used for the RT-MLPA reactions to quantify human XIST gene expression**

<table>
<thead>
<tr>
<th>length</th>
<th>number</th>
<th>gene</th>
<th>mspview</th>
<th>chr. pos.</th>
<th>LPO</th>
<th>SP</th>
<th>RPO</th>
<th>start</th>
<th>end</th>
<th>RT primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>121</td>
<td>11</td>
<td>LOHA</td>
<td>11:018,374,441</td>
<td>11p15.1</td>
<td>TCCGAGCTGACATTCCCGATT</td>
<td>TOCTGGAAGGCACCTGACACCTGACATGCTGACCTGCACCC</td>
<td>CCGAGACGCTGAGCTGAGCGCCCTCCCTGATGAA</td>
<td>18376941</td>
<td>18379495</td>
<td>TATATACTACGCTGATGCTTAG</td>
</tr>
<tr>
<td>148</td>
<td>30224</td>
<td>GARCH</td>
<td>12:406,517,036</td>
<td>12p13</td>
<td>GGCGCGCGGACTGACAGTGGACAATTG</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>CCGAGACGCTGAGCTGAGCGCCCTCCCTGATGAA</td>
<td>6517093</td>
<td>6517096</td>
<td>GCCGAGCGTACGCTGATGCTTGT</td>
</tr>
<tr>
<td>163</td>
<td>100544</td>
<td>2DM</td>
<td>15:542,795,157</td>
<td>15q21.1</td>
<td>TGGGGTGTGACATTCCCGATT</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>CCGAGACGCTGAGCTGAGCGCCCTCCCTGATGAA</td>
<td>6527851</td>
<td>4527854</td>
<td>CAGACGACGGAGCGATGCTTCA</td>
</tr>
<tr>
<td>260</td>
<td>18097</td>
<td>264368</td>
<td>17:477,435,956</td>
<td>17q25.3</td>
<td>GGGGAGGAGGACTGACAGTGGACAATTG</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>CCGAGACGCTGAGCTGAGCGCCCTCCCTGATGAA</td>
<td>77420356</td>
<td>77420358</td>
<td>ACCATGACGCTGATGCTTCA</td>
</tr>
<tr>
<td>344</td>
<td>180566</td>
<td>217309</td>
<td>18:984,950,601</td>
<td>18q11.1</td>
<td>CAGAGACGCTGACATTCCCGATT</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>CCGAGACGCTGAGCTGAGCGCCCTCCCTGATGAA</td>
<td>64651705</td>
<td>64651706</td>
<td>GGCGGCGGCGGACCGCCCTGAG</td>
</tr>
<tr>
<td>333</td>
<td>140381</td>
<td>193605</td>
<td>Xq12</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>CCGAGACGCTGAGCTGAGCGCCCTCCCTGATGAA</td>
<td>72965661</td>
<td>72965664</td>
<td>GTATGACTGACGCTGAGGCC</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>196655</td>
<td>101235</td>
<td>Xq12</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>CCGAGACGCTGAGCTGAGCGCCCTCCCTGATGAA</td>
<td>6518114</td>
<td>6518116</td>
<td>CCGGCGTACGCTGATGCTTGA</td>
<td></td>
</tr>
<tr>
<td>445</td>
<td>141701</td>
<td>163996</td>
<td>Xq12</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>CCGAGACGCTGAGCTGAGCGCCCTCCCTGATGAA</td>
<td>72965735</td>
<td>72965736</td>
<td>GTGATGACTGACGCTGAGGCC</td>
<td></td>
</tr>
<tr>
<td>454</td>
<td>145789</td>
<td>592328</td>
<td>15:222,22792</td>
<td>19q13.3</td>
<td>GGGGAGGAGGACTGACAGTGGACAATTG</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>CCGAGACGCTGAGCTGAGCGCCCTCCCTGATGAA</td>
<td>2222793</td>
<td>2222796</td>
<td>CTCGCTGACGCTGATGCTTCA</td>
</tr>
<tr>
<td>472</td>
<td>148664</td>
<td>40075</td>
<td>11:183,744,981</td>
<td>11q22</td>
<td>GAAAGCGAAGCGAAGCAGTGGACAATTG</td>
<td>TCTGCGGAGCTGACACTGCTGACATGCTGACCTGCACCC</td>
<td>CCGAGACGCTGAGCTGAGCGCCCTCCCTGATGAA</td>
<td>101744491</td>
<td>101744501</td>
<td>GTGATGACTGACGCTGAGGCC</td>
</tr>
</tbody>
</table>
### B. Patient characteristics distributed by XIST expression

<table>
<thead>
<tr>
<th>Variable</th>
<th>XIST expression (exon 2-3)</th>
<th></th>
<th>XIST expression (exon 4-5)</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>p values*</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Total</td>
<td>39</td>
<td>65.0</td>
<td>21</td>
<td>35.0</td>
<td>42</td>
</tr>
<tr>
<td>Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional Chemotherapy</td>
<td>19</td>
<td>45.7</td>
<td>12</td>
<td>57.1</td>
<td>21</td>
</tr>
<tr>
<td>IPB chemotherapy</td>
<td>20</td>
<td>51.3</td>
<td>9</td>
<td>42.9</td>
<td>21</td>
</tr>
<tr>
<td>Age in categories</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 40 years</td>
<td>19</td>
<td>48.7</td>
<td>7</td>
<td>33.3</td>
<td>19</td>
</tr>
<tr>
<td>&gt; 40 years</td>
<td>20</td>
<td>51.3</td>
<td>14</td>
<td>66.7</td>
<td>23</td>
</tr>
<tr>
<td>Type of surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast conserving therapy</td>
<td>10</td>
<td>25.6</td>
<td>6</td>
<td>28.6</td>
<td>11</td>
</tr>
<tr>
<td>Mastectomy</td>
<td>29</td>
<td>74.4</td>
<td>15</td>
<td>71.4</td>
<td>31</td>
</tr>
<tr>
<td>Tumour classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1 or T2</td>
<td>33</td>
<td>84.6</td>
<td>18</td>
<td>85.7</td>
<td>36</td>
</tr>
<tr>
<td>T3</td>
<td>6</td>
<td>15.4</td>
<td>3</td>
<td>14.3</td>
<td>6</td>
</tr>
<tr>
<td>No. of positive lymph nodes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-9</td>
<td>26</td>
<td>60.7</td>
<td>11</td>
<td>52.4</td>
<td>26</td>
</tr>
<tr>
<td>≥ 10</td>
<td>13</td>
<td>33.3</td>
<td>10</td>
<td>47.6</td>
<td>16</td>
</tr>
<tr>
<td>Histologic grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I + II</td>
<td>6</td>
<td>15.4</td>
<td>9</td>
<td>42.9</td>
<td>9</td>
</tr>
<tr>
<td>III</td>
<td>31</td>
<td>79.5</td>
<td>10</td>
<td>47.6</td>
<td>31</td>
</tr>
<tr>
<td>Not determined</td>
<td>2</td>
<td>5.1</td>
<td>2</td>
<td>9.5</td>
<td>2</td>
</tr>
<tr>
<td>Hormone receptor status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER and PR negative (&lt;10%)</td>
<td>33</td>
<td>84.6</td>
<td>13</td>
<td>61.0</td>
<td>34</td>
</tr>
<tr>
<td>ER or PR positive (&gt;10%)</td>
<td>4</td>
<td>10.3</td>
<td>6</td>
<td>38.1</td>
<td>6</td>
</tr>
<tr>
<td>Unknown</td>
<td>2</td>
<td>5.1</td>
<td>0</td>
<td>0.0</td>
<td>2</td>
</tr>
<tr>
<td>PS3 status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative (&lt;10%)</td>
<td>21</td>
<td>53.8</td>
<td>11</td>
<td>52.4</td>
<td>23</td>
</tr>
<tr>
<td>Positive (&gt;10%)</td>
<td>18</td>
<td>46.2</td>
<td>10</td>
<td>47.6</td>
<td>19</td>
</tr>
</tbody>
</table>

* p values were calculated using the Fisher's exact test; patients with unknown values were omitted from analyses. Abbreviations: IPB-chemotherapy, intensive, platinum-based chemotherapy; ER, estrogen receptor; PR, progesterone receptor.
C. Multivariate Cox proportional-hazard analysis of the risk of recurrence (RFS) and XIST expression

<table>
<thead>
<tr>
<th>Variable</th>
<th>No. Events / No. patients</th>
<th>Hazard Ratio</th>
<th>95% CI</th>
<th>p values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>XIST expression (exon 2-3)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>18 / 39</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>15 / 21</td>
<td>1.56</td>
<td>0.60 – 4.04</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>XIST (exon 2-3) Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional chemotherapy</td>
<td>13 / 19</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPB chemotherapy</td>
<td>5 / 20</td>
<td>0.31*</td>
<td>0.11 – 0.88</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>XIST (exon 2-3) High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional chemotherapy</td>
<td>9 / 12</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPB chemotherapy</td>
<td>6 / 9</td>
<td>0.63*</td>
<td>0.20 – 1.91</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>XIST expression (exon 4-5)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>21 / 42</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>12 / 18</td>
<td>1.07</td>
<td>0.41 – 2.82</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>XIST (exon 4-5) Low</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional chemotherapy</td>
<td>15 / 21</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPB chemotherapy</td>
<td>6 / 21</td>
<td>0.30*</td>
<td>0.11 – 0.82</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>XIST (exon 4-5) High</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional chemotherapy</td>
<td>7 / 10</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPB chemotherapy</td>
<td>5 / 8</td>
<td>0.81*</td>
<td>0.23 – 2.89</td>
<td>0.74</td>
</tr>
</tbody>
</table>

All analyses shown were adjusted for: number of positive lymph nodes (4-9 vs. >10); pathological T-stage (1 or 2 vs. 3); histologic grade (I or II vs. III). Homogeneity of both hazard ratios was not rejected based on an interaction term with * p value = 0.36 for XIST expression exon 2-3 and * p value = 0.24 for XIST expression exon 4-5. Number of events is not equal for all variables, since some patients have missing data; maximum missing variables (i.e. events) is 1/33. Abbreviations: CI, confidence interval.
**Figure 8**

**Xist**
- high cisplatin sensitivity
- moderate cisplatin sensitivity

- $P < 0.0001$

**Abcb1b**
- poor docetaxel sensitivity
- good docetaxel sensitivity

- $P < 0.34$

**MEEBO**
- high cisplatin sensitivity
- moderate cisplatin sensitivity

**Illumina**
- high cisplatin sensitivity
- moderate cisplatin sensitivity

- $P < 0.006$

- $P < 0.1$
Figure 9

![Graph showing average normalized gene expression with a range indicated by a bar and error bars. The y-axis ranges from 0.00 to 1.20, and the x-axis is labeled 'Average (normalized) gene expression ATM'.]
Figure 10

A. Patients with Low ATM expression

B. Patients with High ATM expression

Recurrence Free Survival

Years after randomization

P = 0.001

P = 0.56

Chemotherapy
Conventional chemoradiation
Intensive PB-chemotherapy
Fig. 12

---

Upper graph:
- CONV
- HD
- Logrank test: \( p = 0.00235 \)

Lower graph:
- CONV
- HD
- Logrank test: \( p = 0.00198 \)