METHODS FOR DIAGNOSIS AND/OR PROGNOSIS OF OVARIAN CANCER

FIELD OF THE INVENTION

The present invention relates to method(s) for diagnosis and/or prognosis of ovarian cancer in a subject and in particular but not exclusively by analyzing MDS1 and EVI1 complex locus (also known as MECOM locus).

BACKGROUND TO THE INVENTION

Ovarian cancer is a very common cancer that affects women worldwide. Out of all the cancers that affect women, epithelial ovarian cancer (EOC) ranks fifth when considering woman cancer mortality in the world and fourth for the age group 40 to 59 years old, with overall a 5-year survival rate of women inflicted by EOC of only 46%. This is despite improvements in surgical techniques and the advent of more targeted therapeutics such as bevacizimab. The low survival rate is explained by the lack of diagnosis of EOC at an early stage, acquired resistance to chemotherapy in a significant number of patients over the years and the lack of effective therapies for advanced refractory disease. The 5-year survival rate is up to 95% if EOC diagnosis is made in the early stage of the disease.

EOC comprises three major histological subtypes; serous, mucinous and endometrioid. Serous EOC includes serous cystomas, serous benign cystadenomas, serous cystadenomas with proliferating activity of the epithelial cells and nuclear abnormalities but with no infiltrative destructive growth (low potential or borderline malignancy), and serous cystadenocarcinomas. Mucinous EOC includes mucinous cystomas, mucinous benign cystadenomas, mucinous cystadenomas with proliferating activity of the epithelial cells and nuclear abnormalities but with no infiltrative destructive growth (low potential or borderline malignancy), and mucinous cystadenocarcinomas. Endometrioid EOC includes endometrioid tumors (similar to adenocarcinomas in the endometrium), endometrioid benign cysts, endometrioid tumors with proliferating activity of the epithelial cells and nuclear abnormalities but with no infiltrative destructive growth (low malignant potential or borderline malignancy), and endometrioid adenocarcinomas. Two further, less-prevalent histological subtypes also exist, clear cell and undifferentiated.

EOC may also be categorized by "stages", depending upon how far they have spread beyond the ovary. Thus, Stage I is defined as ovarian cancer that is confined to one or both

ovaries. Stage II is defined as ovarian cancer that has spread to pelvic organs (e.g., uterus, fallopian tubes), but has not spread to abdominal organs. Stage III is defined as ovarian cancer that has spread to abdominal organs or the lymphatic system (e.g., pelvic or abdominal lymph nodes, on the liver, on the bowel). Finally, Stage IV is defined as ovarian cancer that has spread to distant sites (e.g., lung, inside the liver, brain, lymph nodes in the neck).

EOCs may be graded according to the appearance of the cancer cells. Low-grade (or Grade 1) means that the cancer cells look very like the normal cells of the ovary; they usually grow slowly and are less likely to spread. Moderate-grade (or Grade 2) means that the cells look more abnormal than low-grade cells. High-grade (or Grade 3) means that the cells look very abnormal. They are likely to grow more quickly and are more likely to spread.

EOC, like most other cancers, is thus a complex heterogeneous disease, influenced and controlled by multiple genetic and epigenetic alterations leading to an increasingly aggressive phenotype. It is now well recognised that the characteristics of an individual tumor and its life course results from multiple somatic mutations acquired over time (e.g. TP53, PTEN, RAS) and continual evolution of the host responses to environmental factors. From a therapeutic standpoint EOC is best considered a collection of complex inter-related diseases represented by an immense natural heterogeneity in tumor phenotypes, disease outcomes, and response to treatment.

A major challenge is thus to identify and thoroughly validate diagnostic and prognostic biomarkers that can accurately describe the heterogeneity ascribed to EOC. In addition, accurate predictive biomarkers are required to guide current treatment protocols, as well as to guide the development and application of new targeted therapies. CA-125 (MUC16, Cancer antigen 125) protein is currently considered the best diagnostic marker of EOC. However, the true positive rate of MUC16 test is only about 50% of stage I EOC patients, while it returns more than 80% of true positives for patients at stages I-IV. About 25% of EOCs especially at the early stages, do not produce reliably-detectable CA-125 and therefore its application in clinical settings is limited. It has been reported that CA-125 in combination with human epididymis protein 4 (HE4) is elevated in greater than 90% of EOC. Such poor prognostic statistics indicate that there is an urgent need to improve understanding of the molecular mechanisms underlying EOC, so as to develop better prognostic and predictive assays and identify new therapeutic targets.

Currently there are about 15 oncogenes considered particularly important for ovarian cancers and 11 of them, including EVI1 have been shown to be amplified in the genome of cancer cells. Proto-oncogene EVI1 (ectopic viral integration site 1), encoded by the MECOM locus (Entrez GeneID: 2122) located at the 3q26 chromosome region, is amplified in many cancers. EVI1 protein was identified as an evolutionary conserved transcription factor sharing 94% amino acid sequence homology between human and mice. In the adult human tissues it is highly expressed in kidney, lung, pancreas, brain and ovaries. In mouse embryos it is highly expressed in the urinary system, lungs and heart and its activity is vital for the embryonic development. The majority of research of EVI1 describes its significance in pathology. If over-expressed in blood cells, EVI1 has been shown to produce a number of alternatively spliced transcripts and causes various hematopoietic disorders, including myeloid leukemias. EVI1 was found to be overexpressed in the blood of up to 21% patients with acute myeloid leukemia (AML). In 4% of AML cases chromosome region 3q is aberrated. High expression of EVI1, regardless the amplification of MECOM locus alone was recently found to be a significant survival factor for ovarian cancer patients. Chromosome region 3q25-27 is amplified in cancers in a various organs: ovary, cervix, lung, oesophagus, colon, head and neck and prostate. Amplification of MECOM is also associated with resistance to chemotherapy in EOC.

However, the information about the role of MECOM locus in ovarian cancer development has been contradictory. In a recent research, although remarkably high expression of EVI1 was detected in serous ovarian cancer tissues, fallopian tube fimbria and benign neoplasms, in comparison with normal tissues of different types, no influence of EVI1 transcript expression on cancer cell proliferation, induced apoptosis or double-strand DNA breaks was revealed in the tests of OVCAR8 cell culture (Jazaeri AA, 2010). In another research it was reported that amplification of MECOM locus can be associated with favorable patient prognosis in ovarian cancer.[Nanjundan et al, 2007]

Accordingly, there is still a need in determining if EVI1 can be used as a biomarker in diagnosis of EOC and to find improved methods for the diagnosis and prognosis of EOC.

SUMMARY OF THE INVENTION

The markers currently being used for detection of EOC lack adequate sensitivity and specificity to be applicable in large populations particularly during the early stages of ovarian cancer. The present invention attempts to fill this gap by proposing MECOM locus and its genes including non-limiting examples, EVI1 and/or MDS1 as more sensitive and

specific diagnostic markers than MUC16, as well as other widely accepted biomarkers. In particular, the present invention relates to identification of clinically distinct sub-groups of EOC patients differentially characterized in one of the following aspects:

- ovarian tumor primary origin (primary EOC versus secondary metastasis),
- presence of ovarian cancer,
- tumor malignancy potential,
- patient's survival prognosis, and
- effectiveness of treatment.

The identification of patient groups is achieved in the integrative analysis of DNA copy number variations and/or gene expression level of at least one gene from the MDS1 and/or EVI1 complex locus (also known as MECOM locus) in the tumor samples. Cutoff values for DNA copy number values of MECOM locus and expression levels of EVI1 and MDS1 genes belonging to this locus maximally separating the patient groups by the chosen criteria are obtained separately for MECOM locus genes: EVI1 and MDS1.

In particular, the diagnostic procedure for individual patients may comprise the steps of:

- obtaining tumor sample material;
- measurement of DNA copy number value of EVI1 and/or MDS1 gene;
- measurement of gene expression level (using a non-limiting example of mRNA) of EVI1 and/or MDS1 genes;
- comparison of the values obtained in the measurements of the above steps against the DNA copy number and/or expression cutoff values respectively.

According to one aspect, the present invention relates to an *in vitro* method for diagnosing EOC and/or predisposition to epithelial ovarian cancer in a subject and/or determining survival prognosis of a subject with epithelial ovarian cancer, and/or determining effectiveness of treatment of epithelial ovarian cancer, and/or determining if an epithelial ovarian cancer in a subject is of primary origin or secondary origin, the method comprising determining in a sample of the subject:

(i) gene expression level (using a non-limiting example of mRNA) of at least one gene in MECOM locus; and/or

(ii) DNA copy number value of at least one gene in MECOM locus;

wherein the level against at least one expression cutoff value and/or DNA copy number against at least one DNA copy number cutoff value are indicative of the subject having EOC, or predisposition to EOC, or the survival prognosis of the subject or the effectiveness of the treatment on the subject.

According to other aspects, the present invention provides kits, computer programs, and computer systems using the method according to any aspect of the present invention.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 is a graph showing that the expression of EVI1 gene of MECOM locus gene was strongly increased in ovarian cancer cells, in comparison with normal epithelium. Absciss axis – EOC tissue samples of individual patients at EOC stages I-IV and normal ovarian tissue from patients with non-cancerous gynecological diseases. Ordinate axis – MDS1 and EVI1 expression microarray signal.

Figure 2 are four graphs (A-D) with A and B showing that MECOM locus genes expression microarray signals in combination with each other and with MUC16 discriminates between normal and cancerous ovarian epithelia better than the current best clinically used biomarker combination MUC16-WFDC2 (C). D is a comparison of ROC curves of MECOM locus gene EVI1 with the current best clinical diagnostic markers WFDC2 and MUC16.

Figure 3 are graphs (A-B) showing that EVI1 and MDS1 genes of MECOM locus are more sensitive and specific than many conventional biomarkers in discriminating diagnosis of primary EOC tumors vs. breast cancer metastases in the ovaries (A) and LMP vs. malignant tumors (B). Cumulative distribution curves of microarray expression values are given.

Figure 4 are two graphs showing that a combination of expression levels of MECOM locus genes EVI1 and MDS1 is important for diagnostics of primary EOC tumors vs. breast cancer metastases in the ovaries (A) and LMP vs. malignant tumors (B). Microarray expression values are given.

Figure 5 are graphs showing that MECOM locus gene EVI1 is a more sensitive and specific biomarker in of primary EOC tumors vs. breast cancer metastases in the ovaries (A) and LMP vs. malignant tumors (B) than the most sensitive and specific clinically used biomarkers WFDC2 and MUC16 which are amongst the many conventional biomarkers. ROC curves of microarray expression values are given.

Figure 6 are graphs showing that a combination of EVI1 and ERBB family genes expression are one of the most sensitive biomarkers of primary EOC. Microarray expression values are given.

Figure 7 is an illustration of DNA copy number variations on chromosome 3 that shows that the 3' end of MECOM locus is the strongest gene amplification hot spot in the genomes of EOC tumor cells.

Figure 8 are graphs (A-D) showing A) that the amplification of MECOM locus together with the expression of its genes EVI1 and MDS1 are stronger indicators of EOC patients survival prognosis than many conventional biomarkers, B) the distribution of EVI1 and MDS1 genes copy number values, C) the optimal values P_{min} of survival prognosis predicted by the combination of DNA copy number and expression values of EVI1 and MDS1, D) the optimal values giving the strongest survival prognosis are different for subgroups of patients with EVI1 and MDS1 gene amplification, the prognostic value of both genes being stronger within these patient subgroups.

Figure 9 is a graph showing that EVI1 is a powerful marker for discriminating between poor and good prognosis patients at a short prognostic time period (up to 300 days after surgery treatment). Left to right: DNA copy number, microarray expression and survival curves of patients separated by the expression of EVI1 and MDS1 genes are presented.

Figure 10 are graphs that show that EVI1 is a marginally significant marker for discriminating between poor and good prognosis patients at medium prognostic time period (up to 1500 days after surgery treatment). Left to right: DNA copy number, microarray expression and survival curves of patients separated by the expression of EVI1 and MDS1 genes are presented.

Figure 11 are nomograms showing the dependency of the primary therapy response outcome patient group composition by response cohort (patient cohort fraction) on EVI1 and MDS1 copy number quantile values Q. The nomograms are used for therapy outcome prediction. Linear equations are chosen to approximate the dependency, but other types of equation are also possible.

Figure 12 are graphs (A-B) showing a possibility of the secondary therapy outcome prediction based on EVI1 and MDS1 genes expression for A) patients without secondary chemotherapy B) with secondary chemotherapy. Copy number and microarray expression values are given.

Figure 13 is an illustration of a method for robust survival prognosis using a combination of MECOM locus copy number and EVI1 and MDS1 expression values with primary therapy response outcome.

Figure 14 is a graph showing the dependence of P-value of the difference between MECOM copy number distributions for patients with good and poor prognoses for a given limited time period on the time limit with a strong positive correlation.

Figure 15 are graphs (A-C) showing that primary therapy response strongly correlates with patient last follow up time (A), which, in its turn, correlates with the P-values of EVI1 (B) and MDS1 (C) genes expression differentiating between patients with good and poor prognosis.

Figure 16 are graphs (A-C) showing that the EVI1 and MDS1 transcription is a significant prognostic factor when used with patients with complete response to the primary therapy ("Complete response" group) and compared with a combined group containing patients demonstrating partial response, stable or progressive disease ("Incomplete response"). A) The cumulative distribution functions of EVI1 and MDS1 gene expression values for "Complete response" (black dots) and "Incomplete response" (grey dots). B) That EVI1 and MDS1 copy number and expression values as survival prognostic markers for patients with complete response to primary therapy. C) EVI1 and MDS1 copy number (MECOM locus) and expression values as survival prognostic markers for patients with partial response to primary therapy, stable or progressive disease.

Figure 17 are graphs showing the algorithm for primary therapy outcome prediction and associated patient survival prognosis based on MECOM locus copy number data. 1) Identify the quantile (Q) of EVI1 gene copy number value, 2,3) determine the expected patient cohort composition based on Q, 4,5,6) determine the expected follow-up time for each patient cohort, 7) estimate the most probable survival time and the responses for each therapy outcome scenario (complete, partial or no response to the therapy).

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Bibliographic references mentioned in the present specification are for convenience listed in the form of a list of references and added at the end of the examples. The whole content of such bibliographic references is herein incorporated by reference.

Definitions

For convenience, certain terms employed in the specification, examples and appended claims are collected here.

The term "aptamer" is herein defined to be oligonucleic acid or peptide molecule that binds to a specific target molecule. In particular, an aptamer used in the present invention may be generated using different technologies known in the art which include but is not limited to systematic evolution of ligands by exponential enrichment (SELEX) and the like.

The term "comprising" is herein defined to be that where the various components, ingredients, or steps, can be conjointly employed in practicing the present invention. Accordingly, the term "comprising" encompasses the more restrictive terms "consisting essentially of" and "consisting of." With the term "consisting essentially of" it is understood that the method according to any aspect of the present invention "substantially" comprises the indicated step as "essential" element. Additional steps may be included.

The term "difference" between two groups of patients is herein defined to be the statistical significance (p-value) of a partitioning of the patients within the two groups. Thus, achieving a "maximum difference" means finding a partition of maximal statistical significance (i.e. minimal p-value).

The term "label" or "label containing moiety" refers in a moiety capable of detection, such as a radioactive isotope or group containing same and nonisotopic labels, such as enzymes, biotin, avidin, streptavidin, digoxygenin, luminescent agents, dyes, haptens, and the like. Luminescent agents, depending upon the source of exciting energy, can be classified as radio luminescent, chemiluminescent, bio luminescent, and photo luminescent (including fluorescent and phosphorescent). A probe described herein can be bound, for example, chemically bound to label-containing moieties or can be suitable to be so bound. The probe can be directly or indirectly labelled.

The term "locus" is herein defined to be a specific location of a gene or DNA sequence on a chromosome. A variant of the DNA sequence at a given locus is called an allele. The ordered list of loci known for a particular genome is called a genetic map. Gene mapping is the process of determining the locus for a particular biological trait. For example, the MECOM locus comprises at least two genes MDS1 and EVI1, which expression results in transcription of, at least two corresponding transcripts, i.e. mRNA variants. The two transcripts may be the longer transcript of MDS1 and the shorter transcript of EVI1.

The term "MECOM locus" is herein defined according to the definition provided in the RefSeq NCBI database as "MDS1 and EVI1 complex locus (MECOM)" or "MDS1 and EVI1

complex locus" (Unigene Hs.659873) and may be essentially characterized by its genomic coordinates hg18.chr3:170,283,981-170,864,257 and by its non-limiting longest transcripts NM_004991.3 and NM_001205194.1.

The term "copy number (CN) value " or "DNA copy number value" is herein defined to refer to the number of copies of at least one DNA segment (locus) in the genome. The genome comprises DNA segments that may range from a small segment, the size of a single base pair to a large chromosome segment covering more than one gene. This number may be used to measure DNA structural variations, such as insertions, deletions and inversions occurring in a given genomic segment in a cell or a group of cells. In particular, the CN value may be determined in a cell or a group of cells by several methods known in the art including but not limited to comparative genomic hybridisation (CGH) microarray, qPCR, electrophoretic separation and the like. CN value may be used as a measure of the copy number of a given DNA segment in a genome. In a single cell, the CN value may be defined by discrete values (0, 1, 2, 3 etc.). In a group of cells it may be a continuous variable, for example, a measure of DNA fragment CN ranging around 2 plus/minus increment d (theoretically or empirically defined variations). This number may be larger than 2+d or smaller than 2-d in the cells with a gain or loss of the nucleotides in a given locus, respectively.

With respect to associations between disease and CN value, a level of variation (deviation) in a DNA segment CN might be important. A level of positive or negative increment of the CN from normal dynamical range in a DNA sample of a given cell group or a single cell may be called CN variation.

The term "diagnosing" and "diagnosis" is herein defined to include the act or process of identifying the existence and/or type of cancer from which an individual may be suffering. Thus, in one embodiment, diagnosis includes the differentiation of a particular cancer type, namely EOC, from one or more other cancers. In an alternative embodiment, binding moieties of the invention are for use in classifying EOC patients into clinically relevant groups based on overall survival and/or cancer-specific survival.

The term "prognosis" is herein defined to include the act or process of predicting the probable course and outcome of a cancer, e.g. determining survival probability and/or recurrence-free survival (RFS) probability.

The term "binding moiety" is herein defined to be a molecule or entity that is capable of binding to target genomic DNA, a target protein or mRNA encoding the same. For example, a binding moiety can be a probe such as a single stranded oligonucleotide at the time of hybridization to a target protein. Probes include but are not limited to primers, i.e., oligonucleotides that can be used to prime a reaction, for example at least in a PCR reaction. In particular, the probe may be capable of binding to ENV1 or MDS1 protein or mRNA encoding the same and be used to quantify the gene expression level of ENV1 or MDS1. In another embodiment, the probe may capable of binding to the genomic DNA of MECOM locus (i.e. ENV1 or MDS1 genomic DNA) and may be capable of determining the copy number of ENV1 or MDS1.

It will be appreciated by persons skilled in the art that the binding moieties of the invention may be used for the diagnosis or prognosis of EOC of any histological subtype (for example, serous, mucinous, endometrioid, clear cell, undifferentiated or unclassifiable).

The term "sample" is herein defined to include but is not limited to be blood, sputum, saliva, mucosal scraping, tissue biopsy and the like. The sample may be an isolated cell sample which may refer to a single cell, multiple cells, more than one type of cell, cells from tissues, cells from organs and/or cells from tumors.

A person skilled in the art will appreciate that the present invention may be practiced without undue experimentation according to the method given herein. The methods, techniques and chemicals are as described in the references given or from protocols in standard biotechnology and molecular biology text books.

According to one aspect, there is provided at least one method for diagnosing epithelial ovarian cancer (EOC), predisposition to EOC in a subject, determining survival prognosis of a subject with EOC, and/or determining the effectiveness of treatment of EOC, and/or determining if an epithelial ovarian cancer in a subject is of primary origin or secondary origin, the method comprising determining in a sample of the subject:

(i) gene expression level of at least one gene in the MECOM locus; and/or

(ii) copy number of at least one gene in the MECOM locus

wherein the gene expression level and/or copy number against at least one expression and copy number cutoff value respectively is indicative of the subject having EOC or predisposition to EOC and/or survival prognosis of a subject with EOC, and/or determining the effectiveness of treatment of EOC.

EVI1 expression as a clinical biomarker may have the largest sensitivity and specificity among conventional biomarkers in the diagnostics of ovarian tumor metastasis. The EVI1 gene expression level is a good indicator of whether the tumors are primary or secondary. EVI1 gene expression may also be useful as a clinical biomarker with high specificity but lower sensitivity in the diagnostics of ovarian tumor malignancy potential.

In particular, EVI1 expression may be a clinical biomarker with large sensitivity and specificity among conventional biomarkers in the diagnostics of ovarian tumor primary origin and as a clinical biomarker with high specificity but lower sensitivity in the diagnostics of ovarian tumor malignancy potential.

The method according to any aspect of the present invention may be *in vitro*, or *in vivo*. In particular, the method may be *in vitro*, where the steps are carried out on a sample isolated from the subject. The sample may be taken from a subject by any method known in the art. For example but not limiting, ovarian tumor material may be extracted from ovaries, fallopian tubes, uterus, vagina and the like. Metastatic tumor may be extracted from peritoneal cavity, other body organs, tissues and the like. Cancer cells may be extracted from non limiting examples such as biological fluids, which include but are not limited to peritoneal liquid, blood, lymph, urine, products of body secretion and the like.

Quantifying of expression ecotropic virus integration site 1 protein homolog (EVI1), Myelodysplasia syndrome-1 (MDS1) and other gene transcripts used according to any aspect of the present invention may be done using any technique of gene expression quantification. Such techniques include, but are not limited to quantitative PCR, semi-quantitative PCR, gene expression microarray, next generation RNA sequencing and the like.

The copy number of MECOM, EVI1, MDS1 and/or other genes used according to any aspect of the present invention may be determined using any technique of gene copy number quantification. Such techniques include, but are not limited to quantitative PCR, semi-quantitative PCR, SNP microarrays, next-generation sequencing, cytogenetic techniques (such as in-situ hybridization, comparative genomic hybridization, comparative genomic hybridization), Southern blotting, multiplex ligation-dependent probe amplification (MLPA) and Quantitative Multiplex PCR of Short Fluorescent Fragments (QMPSF) and the like.

In particular, the method according to any aspect of the present invention comprises, consists of or consists essentially of determining the level of gene expression and/or copy number of the MECOM locus.

In particular, the method according to any aspect of the present invention comprises, consists of or consists essentially of determining the level of gene expression of EV1 and/or MDS1 of the MECOM locus. More in particular, the method further comprises, consists of or consists essentially of the step of determining the copy number of EV1 and/or MDS1 of the MECOM locus.

According to one embodiment, the method comprises, consists of or consists essentially of determining the level of gene expression and copy number of EV1 or MDS1. According to another embodiment, the method comprises, consists of or consists essentially of determining the level of gene expression of EV1 and copy number of MDS1. According to a further embodiment, the method comprises, consists of or consists essentially of determining the level of gene expression of MDS1 and the copy number of EV1. According to one embodiment, the method comprises, consists of or consists essentially of determining the level of gene expression of EV1 and MDS1 and the copy number of EV1. According to a further embodiment, the method comprises, consists of or consists essentially of determining the level of gene expression of MDS1 and the copy number of EV1 and MDS1. According to an even further embodiment, the method comprises, consists of or consists essentially of determining the level of gene expression and copy number of both EV1 and MDS1.

The method according to any aspect of the present invention may include a further step of determining the gene expression level of at least one further gene selected from the group consisting of MUC16, WFDC2, P53, KRAS, ERBB1, ERBB2, ERBB3, EGF, NGR1, TGFA, and MYC in the sample.

MUC16 and WFDC2 are existing clinically-approved markers for ovarian cancer. The use of these genes with EVI1 gene and/or the MDS1 may result in more sensitive and specific diagnosis and/or prgonosis. In particular, better results may be obtained with a combination of the EVI1 gene expression level and the WFDC2 gene expression level (i.e. double combination). This combination can be used to obtain more accurate information about a subject relating to ovarian cancer (for example, whether the ovarian epithelia of the subject is normal or cancerous). Alternatively, the EVI1 gene expression level may also be combined with both the WFDC2 gene expression level and the MUC16 gene expression

level to achieve even more accurate results. This triple combination achieves a more specific gene signature (i.e. a more specific marker) for ovarian cancer as compared to the double combination or to using the EVI1 gene expression level alone.

The treatment may be selected from the group consisting of chemotherapy, surgery and post-surgery chemotherapy.

The combination of expression data of MDS1 with EVI1 could be used also in pair for increasing the discrimination ability of tumor subtypes in ovarian cancer for example but not limited to low malignancy potential tumors, breast cancer metastases in the ovary, EOC and the like.

The expression cutoff values and copy number cutoff values may be obtained during a training stage using a plurality of training subjects with known diagnosis relating EOC whereby the training subjects comprise two sets of training subjects, each set associated with a different diagnosis. This may be done using a variety of ways.

In one example, the expression cutoff value is obtained for each gene with the following steps:

(1) extract the expression level of the gene for each training subject;

(2) obtain the cumulative distribution function of the gene expression levels of each set of training subjects (each set being associated with a different diagnosis as mentioned above); and

(3) select an expression cutoff value which divides the training subjects into two groups based on whether the gene expression level of each training subject is higher or lower than the expression cutoff value.

Any method known in the art for obtaining the cumulative distribution factor in step (2) may be used. In particular, the method may assume that there are two training sets: one with negative diagnosis/prognosis/prediction of size N_1, and the other with positive diagnosis/prognosis/prediction of size N_2, X_1 - the value of parameter X for the negatively diagnosed patients, X_2 - the value of parameter Y for the positively diagnosed patients, C - the chosen cutoff level for parameter X.

Definitions:

$$CDF_X(x) = count(X \leq x)/N$$

$$\text{Observed negatives} = count(X \leq C)$$

$$\text{Observed positives} = count(X > C)$$

Then,

$$\begin{cases} TN = \text{True negatives} = count(X_1 \leq C) = N_1 CDF_{X_1}(C) \\ TP = \text{True positives} = count(X_2 \geq C) = N_2 CDF_{X_2}(C) \\ FN = \text{False negatives} = count(X_1 > C) = N_1(1 - CDF_{X_1}(C)) \\ FP = \text{False positives} = count(X_2 < C) = N_2(1 - CDF_{X_2}(C)) \end{cases}$$

In the simplest 1-dimensional case, the maximal difference is obtained with

$$C_{opt} = \underset{C}{argmin}(FP+FN) = \underset{C}{argmin}\left(N_1(1 - CDF_{X_1}(C)) + N_2(1 - CDF_{X_2}(C))\right)$$

In a more complex, 2-dimentional case (e.g. 2 genes A and B), the CDF becomes a 2-D function CDF_X (X_A, X_B) and the maximal difference is obtained with a 2-D minimization.

In particular, the division of the training subjects into the two groups based on whether the gene expression level of each training subject is higher or lower than the expression cutoff value may be identical to a division of the training subjects into the two sets with different diagnoses. However, it may be difficult to achieve this ideal situation.

Therefore, in one example, the expression cutoff value may be chosen such that it achieves a maximum difference between the two divisions. In this example, the expression cutoff value may be chosen to minimize the number of false negatives. Depending on the diagnosis to be determined using the expression cutoff value, the false negatives can refer to the training subjects wrongly classified by the expression cutoff value into the group corresponding to the set of training subjects not having cancer, having secondary tumors, having low malignancy potential tumors and having a survival time beyond a certain value.

In another example, the diagnosis relates to the survival time of the subject and the cutoff expression value is chosen in step (3) to separate the training subjects into the two groups such that the two groups have maximally different (Wald statistics) survival curves (Cox-

proportional model). This single 1-D optimization to achieve maximally different survival curves was used to obtain the expression cutoff values for the curves in Figure 8A and the curves for the "Combined" cohort in Figure 8D. The expression cutoff values for these figures are in log2 scale and are denoted by the letter "C" above the graphs.

In this study, data regarding the test subjects and training subjects are obtained from the same database with the test subjects having known diagnosis as well (though, this known diagnosis is not required for obtaining information about the test subject). Note that it cannot be guaranteed that an expression cutoff value obtained using a group of subjects (training subjects) can definitely be used for evaluating another group of subjects (test subjects). Neither can it be guaranteed that the expression cutoff value can be used in a general diagnostic kit for all test subjects. It has been demonstrated through years of clinical practice that there is almost always a specific cohort having clinical values (obtained for a specific drug or marker) significantly different from the clinical values obtained during clinical trials. Many existing and certified diagnostic/prognostic kits (e.g. MammaPrint, OncoMine) face the same problem. To determine a more optimal cutoff expression value, either more specific to a given cohort, or more powerful across several cohorts, specialized clinical studies may be required.

In one example, a copy number cutoff value (for primary patient stratification) and two expression cutoff values are obtained for each gene using a two-step optimization algorithm as follows:

(1) extract the expression level and copy number of the gene for each training subject;

(2) estimate a copy number value and divide the training subjects into two cohorts according to whether their copy numbers of the gene is above or below the copy number;

(3) for each cohort obtained in step (2), select an expression value that divides the training subjects into two groups based on whether the gene expression level of each training subject is higher or lower than the expression value.

In another example, the diagnosis relates to the survival time of the subject and the expression value is chosen to separate the training subjects in the cohort into two groups such that the two groups have maximally different (Wald statistics) survival curves (Cox-proportional model). The training subjects in each cohort also comprise two sets of training subjects, each set associated with a different diagnosis. In a second example, the

expression value is selected such that it achieves a maximum difference between the division of the training subjects in the cohort into the two groups based on the expression value and a division of the training subjects in the cohort into the two sets with different diagnoses. For example, the expression value may be chosen to minimize the number of false negatives. This way of selecting the expression value can be used for any type of diagnosis, include a diagnosis relating to the survival time of the subject.

(4) repeat steps (2) – (3) for a range of copy numbers;

(5) select the copy number cutoff value that achieves the maximum difference between the survival curves in a cohort (as in the first example) or the maximum difference between the divisions in a cohort (as in the second example). In one example, only the cohorts with the copy number of the gene above the respective estimated copy numbers are considered for selecting the copy number cutoff value. The expression cutoff values are then obtained as the expression values associated with the copy number cutoff value.

The cutoff values for obtaining the graphs (corresponding to both the "Amplified" and "Not amplified" cohorts) in Figure 8D are obtained using the two-step optimization algorithm above.

The copy number of the gene of the subject may be determined from a tumor of the subject. The method may be for determining survival prognosis of the subject, wherein the method further comprises the steps of:

(i) parametrizing a dependence between a patient cohort fraction and the copy number of the gene of the set training subjects;

(ii) parametrizing a dependence between the patient cohort fraction and the survival time of the set of training subjects; and

(iii) using the copy number of the subject to determine the patient cohort fraction from the dependence of (i) and using the patient cohort fraction of the subject to determine an estimated survival time of the subject using dependence of (ii).

The survival time of the training subject may be based on a last follow-up time for the training subject. According to any aspect of the present invention, last follow-up time may

refer to the last visit by the subject to a medical practitioner to be examined. It may be 50, 100, 300, 500, 1000, 1500, 2000, 2500 and the like from prognosis of EOC.

The method may be for determining effectiveness of treatment further comprises the steps of:

(i) parametrizing a dependence between the copy number of the gene of the training subject and a possible treatment response of the training subject selected from a group consisting of:

- a complete response;

- an incomplete response; and

- a null response; and

(ii) using the copy number of the subject to determine an estimated possible treatment response of the subject.

According to one aspect, there is provided at least one kit for diagnosing epithelial ovarian cancer and/or predisposition to epithelial ovarian cancer in a subject, the kit comprising at least one probe that can identify the level of expression and/or copy number of at least one gene in the MECOM locus. The probe may be an oligonucleotide, aptamer, antibody and/or drug that may bind to MDS1and/or EVI1 gene/protein suitably.

In particular, the probe may be a drug. A non-limiting example of the drug may be sinefungin (Sigma-Aldrich), deazaneplanocin (Moravek Biochemicals) other inhibitors of histone-methytransferases and their derivatives and the like.

According to a further aspect, there is provided at least one computer system having a processor arranged to perform a method according to any aspect of the present invention.

According to another aspect, there is provided a computer program product such as a tangible data storage device, readable by a computer and containing instructions operable by a processor of a computer system to cause the processor to perform a method according to any aspect of the present invention.

Having now generally described the invention, the same will be more readily understood through reference to the following examples which are provided by way of illustration, and are not intended to be limiting of the present invention.

A person skilled in the art will appreciate that the present invention may be practised without undue experimentation according to the method given herein. The methods, techniques and chemicals are as described in the references given or from protocols in standard biotechnology and molecular biology text books.

EXAMPLES

Standard molecular biology techniques known in the art and not specifically described were generally followed as described in Sambrook and Russel, Molecular Cloning: A Laboratory Manual, Cold Springs Harbor Laboratory, New York (2001).

Clinical data description

The National Institute of Health (NIH) Cancer Genome Atlas (TCGA) database (McLendon, 2008) was used for copy number analysis of patients with ovarian cancer. The data on 514 EOC patients were downloaded from TCGA website. Among them 449 (87%) patients were initially classified to stages 3 and 4 of EOC; 448 (87%) patients received chemotherapy; 44 (8.5%) patients were younger than 45 years old, 21 (4%) were between 45 and 65 years old and 290 (56%) were older than 65. For 484 (94%) patients the time of the last follow-up and for 264 (51%) the time of death were defined.

For copy number analysis the data on 504 patients (of 514 in total) in TCGA set were available. Among them for 337 samples gene expression data was also available. This group of samples was used for survival analysis.

The predictive power of MECOM transcripts expression and gene copy number on the efficiency of treatment on 524 patients from TCGA database was tested. The patients were stratified by the type of response to primary therapeutic treatment into 2 classes: 1) 288 patients with complete response, 2) 121 patients with "incomplete response" (including partial response, stable or progressive disease), and 3) 115 patients with no response registered ("null response"), largely due to their deaths earlier than the due time of primary therapy response assessment.

For gene expression analysis the following publicly available datasets were obtained from Gene Expression Omnibus (GEO) website (Edgar R. 2002), GSE12172 including 90

samples (Anglesio MS, 2008), GSE20656 including 172 samples (Meyniel JP, 2010), GSE14407 including 24 patients (Bowen NJ, 2009). Among the 72 patients of GSE12172 dataset, which passed the quality assessment, all the tumors were characterized with serous phenotype, 56 (77%) were classified to stages 3 and 4 of EOC, 50 (72%) tumors were characterized as malignant, 22 (28%) were characterized as LMP (low malignancy potential).

Among the 116 patients of GSE20656 dataset, which passed the quality assessment, 59 tumors (78% of of the 76 tumors with available information) were characterized with serous phenotype, 55 (68% of the 81 tumors with available information) were classified to stages 3 and 4 of EOC, 83 (72%) tumors were identified as primary EOC, 28 (72%) were identified as breast cancer metastases in the ovaries. Among the 24 patients of GSE14407 12 patients were diagnosed with EOC at stages Ic (2 patients), II and IIb (2 patients), III (1 patient), IIIc (5 patients), IV (2 patients). The rest 12 samples were taken from normal ovarian epithelium of patients with subjected to surgical treatment if diseases different from epithelial cancer.

Detailed information on the studied clinical cohorts can be found in Tables 1 and 2 below.

Parameter	Value	No. of Patients
GEO	**GSE12172**	**72**
Primary.Site	**NA**	**72**
Source Name	**OVARIAN CARCINOMA**	**50**
	OVARIAN TUMOR(LMP)	**22**
Grade	2	11
	3	37
	HIGH	10
	LOW	8
	NA	6
stage1	1	11
	2	5
	3	47
	4	9
stage2	A	14
	B	6
	C	42
	NA	10
histotype	SEROUS	72
tumor/metastasis	**METASTASIS**	**14**
	PRIMARY	**58**
LMP/malignant	**LMP**	**22**
	MALIGNANT	**50**
subtype 2	MUTANT BRAF	9
	MUTANT ERBB2	2
	MUTANT KRAS	10
	NOT MUTATION TESTED	18
	WILD TYPE (BRAF/KRAS)	27
	WILD TYPE (BRAF/KRAS/ERBB2)	6
Micropapillary feature	NA	54
	NEGATIVE	9
	POSITIVE	9
(LMP) implants	MICRO-INVASIVE	3
	NA	54
	NEGATIVE	6
	NON-INVASIVE	9
Co-exhistant histology	ADJACENT BRENNER TUMOR	1
	ADJACENT INTRAEPITHELIAL CARCINOMA	1
	ADJACENT LMP	2
	ADJACENT LMP/MPSC	1
	ADJACENT SEROUS CYSTADENOMA	1
	NA	56
	NO NOTED	10
Cluster info	INV CLUSTER	48
	LMP CLUSTER	24
arrayed site	OV	58
	PE	14

Table 1. Ovarian carcinoma vs LMP

	Value	No. of Patients
GEO	**GSE20565**	**116**
Primary.Site	**OVARY**	**116**
Source Name	**BREAST METASTASIS IN THE OVARY**	**27**
	OVARIAN CARCINOMA	**76**
	PLAUSIBLE BREAST METASTASIS IN TH	**7**
	PLAUSIBLE OVARIAN CARCINOMA	**6**
Grade	1	6
	2	24
	3	53
	NA	33
stage1	1	18
	2	8
	3	44
	4	11
	NA	35
stage2	A	14
	B	9
	C	47
	NA	46
histotype	ADENOCARCINOMA	2
	BRENNER TUMOR	1
	CARCINOSARCOMA	2
	CLEAR CELLS	6
	ENDOMETRIOID	6
	MUCINOUS	6
	NA	34
	SEROUS	59
tumor/metastasis	METASTASIS	33
	PRIMARY	83
Breast tumor_Age at diagnosis	Median=48 years	
	NA	103
	20-39	2
	40-49	7
	50-59	2
	60-69	2
	70-79	0
	80-89	0
Breast tumor_Local treatment	BCS	5
	M	8
	NA	103
Breast tumor_Stage	I	2
	II	1
	III	10
	NA	103
Breast tumor_Systemic treatment	CT	5
	CT-T	5
	NA	103
	NO	1
	T	2
Breast_ER	plus	10
	minus	3
	NA	103
Breast_Grade	I	2
	II	3
	III	7
	III, THEN II	1
	NA	103
Breast_Histology	IDC	6
	IDC AND ILC (MIXED)	1
	IDC, THEN ILC*	1
	ILC	4
	NA	103
	UNDIFFERENTIATED	1
Breast_PgR	positive	9
	negative	4
	NA	103

Ovarian tumor_Age at diagnosis	Median	52
	NA	103
	20-39	1
	40-49	5
	50-59	5
	60-69	1
	70-79	1
	80-89	0
Ovarian tumor_FIGO stage	IC	1
	IIA	2
	IIC	1
	IIIC	4
	IV	5
	NA	103
Ovarian tumor_Metastatic site	BONE, LIVER	1
	LIVER	2
	NA	111
	PERITONEAL CARCINOMATOSIS, SUPRA	1
	PLEURA	1
Ovarian tumor_Outcome	DOD	9
	NA	103
	NED	3
	SD	1
Ovarian tumor_Time interval from BC (months)	Median=34	
	NA	103
Ovary_ER	positive	4
	negative	2
	NA	103
	ND	7
Ovary_grade	II	2
	III	4
	NA	103
	ND	7
Ovary_Histology	LOBULAR CARCINOMA	5
	NA	103
	POORLY DIFFERENTIATED ADK	3
	SEROUS PAPILLARY	5
Ovary_Pgr	positive	3
	negative	1
	NA	103
	ND	9

Table 2 Ovarian carcinoma vs. Breast cancer metastasis

Normalization of the expression values of each dataset was performed with MBEI algorithm (Li C., *et al.*, 2001). ANOVA method (Kerr M.K. *et al.*, 2001) was utilized to adjust the batch effect among three datasets. All procedures described for TCGA expression data analysis were performed in a similar manner in the GEO expression datasets.

EXAMPLE 1

EVI1 gene expression is a discriminative marker between normal and cancer ovarian tissues

The results are shown in Figure 1 where the expression of EVI1 and MDS1 genes of MECOM locus were shown to be strongly increased in ovarian cancer cells, in comparison with normal epithelium. In the samples taken from ovarian tissues using laser micro-dissection technique, the average expression of EVI1 and MDS1 genes were 3.6 and 1.4 times higher respectively in cancer tissues, in comparison with neighboring normal tissues (Figure 1). This difference in the mean expression value is more significant for EVI1, in comparison with MDS1 (P=0.021). MDS1 and EVI1 expression values correlate across all the samples (Pearson's r=0.84, Kendall's t=0.60).

EXAMPLE 2

EVI1 expression was a sensitive and specific marker of primary EOC tumors among conventional diagnostic biomarkers

To assess the significance of EVI1 in ovarian cancer, its expression was analyzed in three data sets and its significance as a diagnostic marker was tested against a set of genes, which significance in diagnostics of ovarian cancer and patients survival is widely accepted: KRAS, ERBB2, P53, MYC, MUC16 (CA-125) and WFDC2 (HE4). The results against WFDC2 and MUC16 are shown in Figure 2. As can be seen from the results, MECOM expression discriminates between normal and cancerous ovarian epithelia. The combination of expression values of MECOM genes alone Figure 2(A) discriminated EOC with False negatives (FN) =1 and False positives (FP) =1, which was comparable to the combination of MUC16 (CA-125) and WFDC2 (HE-4) markers Figure 2(C). The perfect separation was obtained in the combination of EVI1 and MUC16 genes Figure 2(B). The overall discriminative power of EVI1 gene for normal vs. EOC epithelia was comparable with MUC16 and WFDC2 Figure 2(D).

The expression values of each marker in ovarian tumors emerged from ovarian epithelia (primary tumors) were compared with with the corresponding expression values in the ovarian metastases of breast cancer tumors (secondary tumors). In the given arrays the expression of EVI1 was represented by probesets 226420_at for MDS1 (RefSeq MECOM, transcript variant 4, NM_004991.3) and 208434_at for EVI1 (RefSeq MECOM, transcript variant 2, NM_005241.3) transcripts respectively. The expression cutoff value (given in a log2 scale) was selected in such way that it minimized the number of false negative diagnoses.

The results presented on Figures 3 and 4 demonstrate that both MDS1 and EVI1 genes can discriminate between primary and secondary ovarian tumors with the highest specificity ($P = 6.10^{-14}$ and $P = 6.10^{-12}$ respectively) among the conventional markers (P53, KRAS, ERBB2, MYC, WFDC2 and MUC16).

Cumulative distribution curves of gene expression values presented in Figure 3 to discriminate A) between primary EOC and breast cancer metastases in the ovaries; B) between low malignancy potential tumors and malignant EOC. The larger the difference between the grey and the black cumulative curves observed (by the P-value), the larger the diagnostic value of the given gene. As can be seen in Figure 3, EVI1 and MDS1 genes of MECOM locus were more sensitive and specific in EOC diagnostics than many conventional biomarkers.

Figure 4 shows that both EVI1 and MDS1 genes of MECOM locus are important for diagnostics. Discrimination between Low malignant potential (LMP) and malignant primary EOC as shown in Figure 4 improved after combining the expression values of EVI1 and MDS1 genes together in a two-dimensional discriminant function.

At expression cutoff level 8.5 of EVI1 expression the number of false negative (FN) diagnoses of primary EOC was FN_{EVI1}=7 with the number of false positive (FP) FP_{EVI1}=0. For the next powerful biomarker MUC16 (CA-125) at expression cutoff level 6.7 FN_{MUC16}=8 and FP_{MUC16}=5. Thus, for EVI1 gene the specificity = 33/(33+0)=100% and the sensitivity = (83-7)/83=91%, while for MUC16 the specificity = 33/(33+5)=87% and the sensitivity = (83-5)/83=94%. The robustness of the discrimination between primary and secondary ovarian tumors can be improved by measurement of both EVI1 and MDS1 genes simultaneusly as shown in Figure 4.

As can be seen, the transcripts of genes of MECOM locus are more sensitive and specific in EOC diagnostics than many conventional biomarkers. The Receiver Operating Characteristic (ROC) curves of Figure 5 showed the expression of EVI1 gene of MECOM locus as a biomarker for clinical diagnostics in comparison with two currently most successful diagnostic markers WFDC2 (HE-4) and MUC16 (CA-125). The best predictor was defined by the curve closest to the coordinate axes. The worst predictor was defined by the curve closest to the reference line (false positive rate = true positive rate).

The specificity and the sensitivity of the discrimination between primary and secondary ovarian tumors is further improved by measurement of EVI1 together with transcripts of ERBB family genes and their secreted signalling protein ligands as shown in Figure 6.

The secreted ligands (such as EGF, TGFA, NGR1) demonstrated slightly lower sensitivity, in comparison with ERBB family receptors (such as ERBB1, ERBB2, ERBB3) as shown in Figure 6.

For EVI1-NGR1 combination the sensitivity was (83-4)/83=95% and the specificity was 100%.

For EVI1-EGF and EVI1-TGFA combinations the sensitivity was (83-3)/83=98% and the specificity was 33/(33+1) = 97%.

For EVI1-ERBB1 and EVI1-ERBB2 combinations the sensitivity was (83-2)/83=98% and the specificity was 100%.

For EVI1-ERBB3 combination the sensitivity was (83-1)/83=99% and the specificity was 100%.

EXAMPLE 3

EVI1 expression was a sensitive marker of EOC tumor malignancy potential

In respect to the malignancy potential of the tumors EVI1 gene was ranked 5th (P=0.022), surpassing biomarkers MYC (P=0.03), WFDC2 (P=0.32) and ERBB2 (P=0.74) in the group discrimination significance.

For EVI1 at expression level cutoff 9.5 FN_{ENVI1}=18, FP_{ENVI1}=4, the specificity = 21/(21+4)=84% and the sensitivity = 49/(49+18)=73%, while for MUC16 (CA-125) at expression cutoff level 9.7 FN_{MUC16}=11, FP_{MUC16}=5, the specificity = 21/(21+5)=81% and the sensitivity = 49/(49+11)=82% The advantage of EVI1 as a clinical diagnostic biomarker proposed in the current innovation over the two certified clinical biomarkers MUC16(CA-15) and WFDC2(HE4) in the full range of cutoff values separating the patient groups is presented on the comparative plots of ROC curves on Figure 6. Discrimination between primary EOC tumors and breast cancer metastases in the ovaries was improved after combining the expression values of EVI1 and ERBB1 genes together in a two-dimensional discriminant function as shown in Figure 6. A combination of EVI1 and ERBB1 expression was shown to be the most sensitve marker of primary EOC.

EXAMPLE 4

The prognostic significance of MDS1 gene expression depends on the number of its genomic copies in tumor tissues, while EVI1 expression is survival significant regardless of its gene copy number

The prognostic significance of the expression of EVI1 and, especially, MDS1 was strongly dependent of the copy numbers of their genomic regions in the patients. Survival analysis of the patients with EOC diagnosed at stages 3 and 4 demonstrated that EVI1 is one of the three genes, which expression level separates the patients into two groups with significantly different survival times. As a survival prognostic marker, the expression of EVI1 gene with P=0.010 ranked second after PIK3CA (P=0.0021). The survival significance of MDS1 gene expression with P=0.049 was marginal, ranking behind PIK3CA, EVI1, WFDC2, MUC16, KRAS, and MYC, but better than ERBB2, P53 and BRCA2.

In the studied population EVI1 gene was represented in more than 2 copies per genome in 52% (262/504) of the patients. The values of both average copy number and the fraction of the patients exceed the corresponding values (50 to 70%) reported previously for EOC tumors (Sunde JS, 2006). The site of copy number at the 3' end of MECOM locus was the most highly amplified region on chromosome 3 reaching the strongest gene amplification hot spot in the genomes of EOC tumor cells as shown in Figure 7. The 3' end of the MECOM locus copy number was 2.5 and higher in the tumors of 84% of patients with EOC.

Across the patients, EVI1 region was characterized with mean 3.52 and median 3.19 copies per genome, and MDS1 region with mean 3.39 and 3.18 copies respectively as shown in Figure 8.

The effect of EVI1 and MDS1 expression was analysed in the tumors on the survival of patients as factors depending on the genomic amplification of these genes. The effect of EVI1 and MDS1 copy numbers was studied by separating the population of patients into two groups relative to a given copy number cutoff: the patients with a large number of EVI1 or MDS1 copies (the genes are amplified) in their tumors and the rest of the patients (the genes are not amplified). The results are presented on Figure 8C and D. Survival analysis of EVI1 and MDS1 expression was performed in the high copy number group and the best P-values for the survival groups separated by the expression were calculated. The copy number cutoff value varied from 0 to 4. Thus copy number-dependent expression-defined survival P-value was calculated. Figure 8(A) shows the comparison of the best separation

of the patient survival curves by gene expression markers (grey- high expression, black- low expression); the more separated grey and black curves were (the less the p-value), the more predictive power for patient survival a given marker had. Histograms in Figure 8(B) are of amplification values of EVI1 and MDS1 genes of MECOM locus obtained. The dependence of the p-value of the best survival curves separation (Pmin) on the degree of amplification of the genes from MECOM locus obtained are shown in Figure 8(C). Figure 8(D) are survival curves which show the maximal separation of patients for the subsets of patients with EVI1 and MDS1 obtained. The amplification higher (the upper row) and lower (the middle row) than 3.5 and 3.6 copies/cell were obtained respectively. As a reference of survival curves separation the combined (amplified together with not amplified) are given in the bottom row. In Figures 8(A) and (D) the cutoff expression value (in log2 scale) are denoted with letters above each of the graphs (within the graph frames). For Figure 8A and the 'Combined' graphs in Figure 8(D) the cutoff was obtained with a single one-dimensional optimization by the expression value. In Figure 8(D) the analysis was made with a two-step optimization by 1) copy number, followed by 2) expression. It is shown that amplification of MECOM locus together with the expression of its genes EVI1 and MDS1 were stronger indicators of EOC patients survival prognosis than many conventional biomarkers.

The copy number cutoff giving the best survival P-value for EVI1 and MDS1 was slightly different (3.5 and 3.6 respectively), although these loci are adjacent. In the group of patients, where MDS1 copy number was higher than 3.6, the expression of MDS1 was a strong factor separating the patients by survival (P=0.0028), showing a remarkable improvement of the P-value by a factor of 17, in comparison with the whole population (P=0.049). In the sub-population where MDS1 was not amplified, no survival significance of its expression was observed (P=0.39).

MDS1 expression was a significant factor of EOC patients' survival only if in their tumors MDS1 gene region copy number was higher than 3.6 (227/354, 65% of the patient population). At the same time, EVI1 gene expression was a significant factor of patients' survival for the whole population (P=0.010), regardless of the level of the amplification of its coding region. The results are shown on Figure 8D.

The survival prognostic value of both EVI1 and MDS1 gene depended on their copy number (i.e. the copy number of MECOM locus, which includes both of them). As displayed on Figure 8D, after segregation of the patients ("Combined") into 2 cohorts ("Amplified" and "Not amplified") by the genes copy number, the prognostic P-value for EVI1 was improved. In addition, for EVI1 gene the prognostic value was marginally significant even without prior

patient stratification ("Combined" cohort) by copy number values. Thus, EVI1 gene expression can be used as a prognostic marker regardless its copy number status.

The results demonstrated that MECOM locus, including proto-oncogenes EVI1 and MDS1, was a clinical biomarker for differential diagnostics and prognosis of the human epithelial ovarian cancer.

EXAMPLE 5

EVI1 and MDS1 were strong prognostic factors for shorter survival periods

The P-value of the difference between MECOM copy number distributions for patients with good (survival) and poor (death) prognoses for a given limited time period had a log-linear dependence on the time limit with a strong positive correlation as shown in Figure 14. Detailed correlation structure of surviving (good prognosis) and deceasing (poor prognosis) patient cohorts were defined by follow-up time. Both MDS1 and EVI1 expression P-value (between good and poor prognosis cohorts) correlated with patients follow-up time. For MDS1 two ranges of follow-up times with different correlation coefficients were observed.

For both EVI1 and MDS1 the functional dependence of P-value on the survival time limit consists of two trends with very strong correlation separated by a region of discontinuity around time 1200 days. For time periods shorter than 500 days correlation between P-value and copy number decreases, while MECOM copy number became a stronger marker, in comparison with all conventional cancer biomarkers as can be seen when comparing Figures 9 and 10. As can be seen in Figure 9, expression of EVI1, but not MDS1 gene showed a strong survival significance for the given prognosis time span. Based on Figure 9, EVI1 was shown to be the most powerful marker for discriminating between poor and good prognosis patients at a short prognostic time period (300 days). As can be seen in Figure 10, expression only of either EVI1 or MDS1 did not have a sufficient survival prognostic power for the given prognosis time span. However, the expression of both genes was highly significant for patients (91%) with tumors with amplified MECOM locus (copy number >2.9). EVI1 was thus shown to be a marginally significant marker for discriminating between poor and good prognosis patients at medium prognostic time period (1500 days).

EXAMPLE 6

The genes of MECOM locus (EVI1 and MDS), being analyzed simultaneously, improve the discriminative power of clinical analysis

The combination of expression data of MDS1 with EVI1 could be used together for increasing the discrimination power of ovarian tumor subtypes in. This possibility is illustrated on Figure 3B by a comparison of the cumulative frequency distribution of the genes expressed in low malignancy potential (LMP) tumors versus malignant EOC. On this figure it can be observed that higher expression values of EVI1 gene of MECOM locus correspond to malignant EOC type. Contrary to this, higher expression values of MDS1 gene of this locus correspond to LMP ovarian tumor type. Thus, the maximal instrumental sensitivity (with both genes measured at high expression) of the measurement is obtained. If MDS1 and EVI1 expression was measured together, the power of the resulting two-dimensional discriminant function to separate the LMP and malignant EOC subgroups was increased, as shown on Figure 4.

EXAMPLE 7

EVI1 and MDS1 copy number were predictive markers of primary therapy outcome

The P-value was studied for the difference between EVI1 and MDS1 copy number distributions for patient groups with good prognosis and poor prognosis. For a given follow-up time cutoff the poor prognosis group was limited to the patients who died before this follow-up time. The good prognosis group was limited to the patients who survived longer than the given follow-up time.

These P-values were found to be different for patients grouped by primary therapy outcome success ("complete", "incomplete" and "null" response patient cohorts), as seen, e.g. on Figures 9 and 10. Moreover, the P-values for both EVI1 and MDS1 strongly correlated with the fraction of patients belonging to each cohort in the poor prognosis group as shown in Figure 11 which shows that the primary therapy response outcome depends on EVI1 and MDS1 copy numbers.

The correlation structure for each cohort was complex. For EVI1, with increase of the cohort patient fraction up to a certain limiting value (10% for "complete", 40% for "incomplete" and 50% for "null" response groups for EVI1) the correlation with P-value was strongly negative (-0.58 for "complete", -0.6 for "incomplete" and -0.89 for "null" response groups for EVI1). After this limit, the correlation changed to strongly positive for "complete"

and "null" response groups, but remained negative for "incomplete" response group. For MDS1 the correlation structure was qualitatively similar. Thus, an 'optimal' fraction of patients for a certain therapy response was defined by the expected minimal P-value of EVI1 (or MDS1) copy number distribution different between the group of the patients with good and poor survival prognosis.

This demonstrated a dependence of primary therapy outcome response on the copy number of EVI1 and/or MDS1 in the tumor tissue of a given patient, along with survival prognostic time period. It became evident that every distribution of EVI1 and MDS1 values predicted the 'optimal' fraction of patients with a given time of response. This optimal corresponded to the maximal dependence of therapy response outcome (as survival probability) of patients in a given cohort on the expression of EVI and/or MDS1.

The expression value of EVI1 and/or MDS1, measured at the time of the surgery, simultaneously mapped each individual patient to the probability distribution of the difference between good and poor survival groups, patient survival probability and the probability of having a "complete", "incomplete" and "null" response and the expected survival time. Thus for a given patient primary therapy outcome could be predicted, along with the survival time estimated from this prediction. An example of this analysis is shown on Figure 17 which includes the algorithm for primary therapy outcome prediction and associated patient survival prognosis based on EVI1/MDS1 copy number data.

The specific steps include:

1. Measure EVI1/MDS1 copy number in the tumor extracted during surgery. Map the copy number on the copy number distribution of EVI1/MDS1.

2. For the given copy number find its quantile Q in the copy number distribution. Optional: for a given copy number find its quantile in the copy number distribution in each response cohort ("Complete", "Incomplete" and "Null") separately . These alternative quantiles will lead to 3 alternative prediction scenarios, the results of which at Step 7 can be combined.

3. Map Q to the "response cohort class composition" vs. Q (copy number) nomogram (Figure 11)

4. From the points of intersection of line "$y=Q$" with the approximated cohort fraction - log(Q) dependence the corresponding values on Ox axis are estimated for each cohort. Thus the inference of the plausibilities of the given patient belonging to each of the cohorts (Ox axis) is made. Up to 6 potential expected cohort fraction values (6 plausibility values) can be obtained at this step if all the intersection points are within [0,1] range. Similarly to Step 2, these values can be traced further to be resolved at Step 7.

5. Map the obtained set of expected cohort fraction values (plausibilities) to "Last follow-up time" vs. "cohort fraction" nomogram (Figure 15A).

6. For every given expected cohort fraction (plausibility) find the points of intersection of the corresponding lines "" with the nomogram curves for each corresponding primary therapy outcome prediction cohort. Obtain the corresponding values of expected survival times on Ox axis for each cohort fraction.

7. Obtain the predicted survival time based on the primary therapy outcome predictions by inferring it from the plausible survival times obtained at Step 6. For example, at Step 2, 6 points of intersection (2 points for each of the 3 quantiles were obtained). At Steps 5 and 6 each point was mapped to a single expected survival time. Thus 6 survival times was obtained. If the goal is to obtain the prediction at the time of obtaining copy number data (i.e. at the time of surgery), a range between the minimal and the maximal of the 6 survival times can be retrieved as a prediction.

8. The set of plausible survival times (up to 6) can be reduced to the set of 2 survival times if the survival prognosis is made for the patient after the outcome of the primary therapy is observed. Then, the certainty in the patient primary therapy outcome cohort is obtained and thus the range of survival times is reduced to the smallest and the largest predicted values for the given cohort.

Primary therapy response strongly correlated with patient last follow up time Figure 15(A), which, in its turn, correlates with EVI1 Figure 15(B) and MDS1 Figure 15(C) -log10 of P-values for expression differentiating between patients with good and poor prognosis. Good and poor prognosis cohorts for a given follow-up time are defined by patient survival or death by the given follow-up time.

EXAMPLE 8

EVI1 and MDS1 were predictive markers for secondary therapy outcome

For a short-term survival period (600 days) survival, patients who did not receive secondary chemotherapy, died significantly more often (EVI1 P=0.01, MDS1 P=0.003) if their tumors contained high copy number of MECOM. The 12secondary therapy outcome prediction results is shown in Figure 12. Among the patients who received secondary chemotherapy, MECOM copy number was an insignificant factor of survival. Figure 12(A) show the results of patients who received chemotherapy and Figure 12(B) show the results of patients who did not receive chemotherapy.

EVI1 expression was insignificantly different among the patients, regardless of their receiving chemotherapy. However, MDS1 expression was marginally higher (P=0.04) in the group of patients with poor prognosis who received secondary chemotherapy.

EXAMPLE 9

EVI1 and MDS1 expression and copy number were strong long-term prognostic factors in each group of patients separated by response type to primary therapy

Combinations of EVI1 and MDS1 copy number and expression values separate groups of patients with various types of response to primary therapy. These markers can be also considered as factors increasing the prognostic power of primary therapy outcomes13. A schema for a robust survival prognosis using a combination of MECOM copy number and EVI1 and MDS1 expression with primary therapy response outcome is shown in Figure 13.

For patients with complete response and MECOM copy number between 2.8 and 4.2, EVI1 expression less than 10.3 and MDS1 expression less than 6.76 long survival time was predicted, while for patients with EVI1 and MDS1 expression higher than these values medium survival time was predicted (up to 2000 days).

For patients with partial and no response (stable or progressive disease) to primary chemotherapy, if MECOM copy number was less than 2.7 and MDS1 expression was less than 6.43, medium survival time as predicted. For patients with MECOM copy number > 2.7 and EVI1 expression less than 9.51, medium survival time was predicted.

Short survival time (maximum 1000 days) was predicted for patients without complete response to primary chemotherapy, MECOM copy number <2.7 and MDS1 expression higher than 6.43, or, in the case of MECOM copy number >2.7, if EVI1 expression >9.51.

Overall, lower MECOM copy number and EVI1 and MDS1 expression, as well as complete response in survival time resulted in prognosis of longer survival times.

EXAMPLE 10

The prognostic power of MECOM copy number **was the strongest for patients with shortest surviving times and predicted for them negative outcome of primary therapy**

From Figure 15, the P-value of difference between MECOM copy number distributions for patients with good and poor prognoses correlated not only with the prognostic period, but also with primary therapy response group composition of the patients. As the fraction of patients with partial and complete response increased, both survival time and the copy number P-value increased. Contrary to this, the MECOM copy number (and EVI1 expression value) was the most significant survival-predicting factor for the short periods of time (<300 days, as shown on Figure 9), when most of the patients belonging to the "Null" therapy response group died. -log10 of P-values for expression differentiated between patients with good and poor prognosis. Good and poor prognosis cohorts for a given follow-up time were defined by patient survival or death by the given follow-up time.

Thus, regarding the prognostic value of MECOM copy number (and also EVI1 and MDS1 expression) and survival of patients grouped by primary therapy response, the prognosis time scale could be divided into 3 ranges (Figure 15A). Short term prognostic range, up to 500 days long, was characterized with the strongest prognostic power of MECOM copy number and the expression of its genes, as well as with the prevalence of the "Null"-therapy responding patients. Medium term prognostic range, spanning from 500 to 1500 days - was characterized with poor prognostic power of MECOM copy number and the expression of its genes (where some conventional cancer biomarkers show stronger prognostic power than MECOM), as well as prevalence of the mixed group of "Incomplete" primary therapy response (combining patients with partial response and stable or progressing disease). Long term prognostic range as characterized with MECOM copy number -dependent prognostic power of EVI1 and MDS1 expression (stronger than for conventional cancer biomarkers) and with prevalence of patients with complete response to primary therapy.

EXAMPLE 11

EVI1 and MDS1 were markers for anti-cancer therapy success prediction

EVI1 expression value was a significant marker (P(Mann-Whitney U)=0.0021) for predicting the outcome of primary post-operative therapeutic treatment response at the time of surgery as shown in Figure 16A. MDS1 expression was insignificant with a trend to significance (P=0.074). In comparison with conventional cancer markers (e.g. P(MUC16)=0.7), P(WFDC2)=0.1, P(KRAS)=0.49, P(P53)=0.4, P(ERBB2)=0.37), only EVI1 expression demonstrated a statistically significant P-value and MDS1 expression ranked

second after EVI1. EVI1 and MDS1 transcription was a significant prognostic factor in combination with patients with complete response to the primary therapy ("Complete response" group) are compared with a combined group containing patients demonstrating partial response, stable or progressive disease ("Incomplete response"). Figure 16(A) show the cumulative distribution functions of EVI1 and MDS1 gene expression values for "Complete response" (black dots) and "Incomplete response" (grey dots). Figure 16(B) show that EVI1 and MDS1 copy number (MECOM locus) and expression values as survival prognostic markers for patients with complete response to primary therapy. Figure 16(C) show EVI1 and MDS1 copy number (MECOM locus) and expression values as survival prognostic markers for patients with partial response to primary therapy, stable or progressive disease.

The prognostic value of EVI1 and MDS1 for primary therapy response was further confirmed by studying patients segregated by primary response types. Both genes were significant for survival prognosis in both patient groups (i.e. complete and "incomplete" response) for . MDS1 was a significant prognostic marker for 42% of patients with complete response (P=0.0010) and 64% of patients with "incomplete" response (P=0.0045). EVI1 was a significant prognostic marker for 82% of patients with complete response(P=0.0048) and 50% of patients with "incomplete" response (P=0.018).

An illustration of primary therapy success prediction is given on Figure 13. On the first stage the prediction of primary therapy success is made based on initial data on MECOM copy number and MDS1 and EVI1 expression values obtained from the analysis of tumor sample extracted during the surgery. Based on this data an appropriate course of therapy is selected (more intensive in the case of MECOM amplification and less intensive otherwise). On the second stage, the response to primary therapy is assessed. Based on the type of response further decision is made on the more exact prognosis of survival time of the patient based on the exact range of MECOM copy number and MDS1 and EVI1 expression values.

REFERENCES

1. Anglesio MS, Arnold JM, George J, Tinker AV, Tothill R, et al. (2008) Mutation of erbb2 provides a novel alternative mechanism for the ubiquitous activation of ras-mapk in ovarian serous low malignant potential tumors. Mol Cancer Res 6: 1678-90.

2. Bowen NJ, Walker LD, Matyunina LV, Logani S, Totten KA, et al. (2009) Gene expression profiling supports the hypothesis that human ovarian surface epithelia are multipotent and capable of serving as ovarian cancer initiating cells. BMC Med Genomics 2: 71
3. Edgar R (2002) Gene expression omnibus: Ncbi gene expression and hybridization array data repository. Nucleic Acids Research 30: 207.
4. Jazaeri AA, Ferriss JS, Bryant JL, Dalton MS, Dutta A (2010) Evaluation of evi1 and evi1s (delta324) as potential therapeutic targets in ovarian cancer. Gynecol Oncol 118: 189-95.
5. Kerr MK, Churchill GA (2001) Statistical design and the analysis of gene expression microarray data. Genet Res 77: 123-8.
6. Li C, Hung Wong W (2001) Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. Genome Biology, 2(8):research0032.1–0032.11
7. McLendon RE, Friedman AH, D BD (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455: 1061-8.
8. Meyniel JP, Cottu PH, Decraene C, Stern MH, Couturier J, et al. (2010) A genomic and transcriptomic approach for a differential diagnosis between primary and secondary ovarian carcinomas in patients with a previous history of breast cancer. BMC Cancer 10: 222
9. Sambrook and Russel, Molecular Cloning: A Laboratory Manual, Cold Springs Harbor Laboratory, New York (2001).
10. Sunde JS, Donninger H, Wu K, Johnson ME, Pestell RG, et al. (2006) Expression profiling identifies altered expression of genes that contribute to the inhibition of transforming growth factor-beta signaling in ovarian cancer. Cancer Res 66: 8404-12
11. http://www.ncbi.nlm.nih.gov/UniGene/clust.cgi?ORG=Hs&CID=659873
12. Nanjundan M, Nakayama Y, Cheng KW, Lahad J, Liu J, Lu K, Kuo WL, Smith-McCune K, Fishman D. Gray JW and Mills GB. Amplification of MDS1/EVI1 and EVI1, Located in the 3q26.2 Amplicon, Is Associated with Favorable Patient Prognosis in Ovarian Cancer. *Cancer Res,* 2007 67:3074-3084

CLAIMS

1. An *in vitro* method for diagnosing epithelial ovarian cancer and/or predisposition to epithelial ovarian cancer in a subject, the method comprising determining in a sample of the subject:

    (i) gene expression level of at least one gene in MECOM locus; and/or

    (ii) copy number of at least one gene in the MECOM locus;

    wherein the level against at least one expression cutoff value and/or copy number against at least one copy number cutoff value are indicative of the subject having epithelial ovarian cancer and/or a predisposition to epithelial ovarian cancer.

2. An *in vitro* method for determining survival prognosis of a subject with epithelial ovarian cancer, the method comprising determining in a sample of the subject:

    (i) gene expression level of at least one gene in MECOM locus; and/or

    (ii) copy number of at least one gene in the MECOM locus;

    wherein the level against at least one expression cutoff value and/or copy number against at least one copy number cutoff value are indicative of determining survival prognosis of the subject with epithelial ovarian cancer.

3. An *in vitro* method of determining effectiveness of treatment of epithelial ovarian cancer in a subject, the method comprising determining in a sample of the subject:

    (i) gene expression level of at least one gene in the MECOM locus; and/or

    (ii) copy number of at least one gene in the MECOM locus;

    wherein the level against at least one expression cutoff value and/or copy number against at least one copy number cutoff value are indicative of the effectiveness of the treatment.

4. An *in vitro* method of determining if an epithelial ovarian cancer in a subject is of primary origin or secondary origin, the method comprising determining in a sample of the subject:

    (i) gene expression level of at least one gene in the MECOM locus; and/or

    (ii) copy number of at least one gene in the MECOM locus;

    wherein the level against at least one expression cutoff value and/or copy number against at least one copy number cutoff value are indicative of whether the cancer is of primary or secondary origin.

5. The method according to any one of the preceding claims, wherein the epithelial ovarian cancer is primary epithelial ovarian cancer.

6. The method according to any one of the preceding claims, wherein the gene expression level is determined by measuring the mRNA or protein expression of the gene from the MECOM locus.

7. The method according to any one of the preceding claims, wherein the method comprises determining the gene expression level of EVI1 and/or MDS1 from the MECOM locus in the sample.

8. The method according to any one of the preceding claims, wherein the method comprises determining the gene expression level of at least one further gene selected from the group consisting of MUC16, WFDC2, ERBB1, ERBB2, ERBB3, EGF, NGR1, and TGFA, in the sample.

9. The method according to any one of the preceding claims, wherein the method comprises determining the copy number of EVI1 and/or MDS1 from the MECOM locus in the subject.

10. The method according to any one of the preceding claims, further comprising a training stage prior to diagnosis of the subject, determining survival prognosis and/or determining effectiveness of treatment, the training stage comprising the following steps for the gene in the MECOM locus:

    for each of a plurality of training subjects with known diagnosis relating to ovarian cancer, determining gene expression level of the gene in the training subject; and

determining an expression cutoff value which divides the training subjects into two groups according to whether the gene expression level of the gene in each training subject exceeds the expression cutoff value;

wherein the training subjects comprise two sets of training subjects, each set associated with a different diagnosis and the expression cutoff value minimizes a difference between the division of the training subjects into the two groups and a division of the training subjects into the two sets.

11. The method according to any one of the preceding claims, further comprising a training stage prior to diagnosis of the subject, determining survival prognosis and/or determining effectiveness of treatment, the training stage comprising the following steps for each said gene in the MECOM locus:

(i) for each of a plurality of training subjects with known diagnoses relating to ovarian cancer, determining gene expression level and copy number of the gene in the training subject;

(ii) estimating a sample copy number and dividing the training subjects into two cohorts according to whether the copy number of the gene in each training subject is above or below the sample copy number;

(iii) for each said cohort, determining a sample expression value which divides the training subjects in the cohort into two groups according to whether the gene expression level of the gene in each training subject exceeds the sample expression value, wherein the sample expression value achieves a maximum measure of difference between the two groups;

(iv) repeating steps (ii) – (iii) by varying the sample copy number in a range of copy numbers identified in the training subjects and obtaining a copy number distribution curve for the gene; and

(v) selecting a copy number cutoff value as the copy number associated with the largest maximum measure of difference between the two groups of a cohort and selecting expression cutoff values as the expression values determined for the cohorts obtained with the copy number cutoff value.

12. The method according to claim 11, wherein the measure of difference between the two groups comprises a measure of difference between survival curves of the two groups.

13. The method according to either claim 11 or 12, wherein diagnosis of the subject, determining survival prognosis and/or determining effectiveness of treatment further comprises the following steps for each said gene in the MECOM locus:

comparing the copy number of the gene of the subject against the copy number cutoff value for the gene;

wherein

if the copy number of the gene of the subject exceeds the copy number cutoff value for the gene, further comparing the gene expression level of the gene of the subject against the expression cutoff value determined for the cohort with copy numbers of the gene above the copy number cutoff value;

if the copy number of the gene of the subject is below the copy number cutoff value for the gene, comparing the gene expression level of the gene of the subject against the expression cutoff value determined for the cohort with copy numbers of the gene below the copy number cutoff value; and

obtaining diagnosis the subject, determining survival prognosis and/or determining effectiveness of treatment based on the comparison between the gene expression level of the gene of the subject and the respective expression cutoff value.

14. The method according to claim 13, for determining survival prognosis of the subject, wherein the method further comprises the steps of:

(i) parametrization of a dependence between a patient cohort fraction and the copy number of the gene in the set of training subjects;

(ii) parametrization of a dependence between the patient cohort fraction and the survival time in the set of training subjects; and

(iii) using the copy number of the subject to determine the patient cohort fraction from the dependence of (i) and using the patient cohort fraction of the subject to determine an estimated survival time of the subject from dependence (ii).

15. The method according to claim 14, wherein the survival time of the training subject is based on a last follow-up time for the training subject.

16. The method according to claim 13 for determining effectiveness of treatment further comprises the steps of:

(i) parametrizing a dependence between the copy number of the gene of the training subject and a possible treatment response of the training subject selected from a group consisting of:

- a complete response;

- an incomplete response; and

- a null response; and

(ii) using the copy number of the subject to determine an estimated possible treatment response of the subject.

17. The method according to any one of claims 1 and 5 to 16, wherein the method is capable of determining whether the ovarian cancer in the subject is primary or secondary ovarian cancer.

18. The method according to any one of claims 1 and 5 to 16, wherein the method is capable of determining whether the ovarian cancer has a malignancy potential in the subject.

19. The method according to any one of claims 3 and 5 to 16, wherein the treatment is selected from the group consisting of chemotherapy, surgery and post-surgery chemotherapy.

20. The method according to any one of claims 4 to 16, wherein the cancer of secondary origin is a secondary breast cancer metastatic cancer.

21. A kit for diagnosing epithelial ovarian cancer and/or predisposition to epithelial ovarian cancer in a subject, the kit comprising at least one probe that can identify the level of expression, copy number and/or protein expression of at least one gene in the MECOM locus.

22. The kit according to claim 21, wherein the probe is at least one aptamer that binds to EVI1 and/or MDS1 proteins.

23. The kit according to claim 21, wherein the probe is at least one antibody that binds to EVI1 and/or MDS1 proteins.

24. The kit according to claim 21, wherein the probe is at least one drug that directly interact or binds to EVI1 and/or MDS1 proteins.

25. The kit according to claim 21, wherein the copy number is determined using at least one method selected from the group consisting of quantitative PCR assay, in situ hybridization, Southern blotting, multiplex ligation-dependent probe amplification (MLPA) and Quantitative Multiplex PCR of Short Fluorescent Fragments (QMPSF).

26. A computer system having a processor arranged to perform a method according to any of claims 1 to 19.

27. A computer program product such as a tangible data storage device, readable by a computer and containing instructions operable by a processor of a computer system to cause the processor to perform a method according to any of claims 1 to 19.