Abstract:
The present invention relates to diagnostic markers comprising variants of the HE4 protein and polynucleotides encoding same, useful in the detection of cancer. The invention further relates to diagnosing cancer in a subject or in a sample obtained from a subject.
HE4 VARIANT NUCLEOTIDE AND AMINO ACID SEQUENCES, AND METHODS OF USE THEREOF

FIELD OF THE INVENTION

The present invention is related to variant nucleotide and protein sequences that are diagnostic markers for cancer, and assays and methods of use thereof.

BACKGROUND OF THE INVENTION

HE4 protein (also known as Whey Acidic Proteins (WAP) four-disulfide core domain protein 2; Major epididymis-specific protein E4; Epididymal secretory protein E4; and Putative protease inhibitor WAP5), is a 124 amino acid protein (SEQ ID NO.3), that was initially identified as a transcript exclusively expressed in the epididymis, and suggested to be a marker specific for this tissue (Kirchhoff et al. Biol. Reprod. 1991; 45, 350-357). This protein contains 2 WAP domains; the WAP domain comprises approximately 50 amino acids and includes 8 cysteines in a conserved arrangement (Ranganathan et al. J. MoL Graph Model. 1999; 17, 106-113, 134-136). WAP-domain containing proteins are typically small secretory proteins that exhibit a variety of functions, among them those affecting growth and differentiation (Ranganathan et al. 1999 ibid; Schalkwijk et al. Biochem. J. 1999; 340, 569-577).

In normal human tissues the expression of the known HE4 protein is highly restricted and is essentially limited to normal epithelial tissues particularly in the reproductive tracts and respiratory airways (Drapkin et al. Cancer Res. 2005; 65, 2162-2169; Galgano et al. Modern Pathology 2006; 19, 847-853). It is highly over-expressed in serous and endometroid ovarian carcinoma (Drapkin et al. 2005, Galgano et al. 2006 ibid). Lung adenocarcinomas, some breast adenocarcinomas, transitional cell carcinomas and pancreatic carcinomas express moderate or high levels of HE4. In addition, some mesotheliomas, and less often, gastrointestinal, renal and transitional cell carcinomas also exhibit over-expression of HE4 (Galgano et al. 2006 ibid). HE4 was also shown to be over-expressed in endometrial cancer and can be useful for detecting endometrial cancer (Li et al. Expert Rev MoI Diagn. 2009; 9, 555-566; Moore et al. Gynecol Oncol. 2008; 110, 196-201). In gastric cancers, especially gastric metaplasia, HE4 was also shown to be over-expressed, while it was absent in normal stomach (Nozaki et al. Gastroenterology. 2008; 34(2), 511-22). Currently, it is believed that HE4 protein is secreted into the bloodstream of patients with ovarian carcinoma (Drapkin et al. Cancer Res. 2005; 65, 2162-2169; Hellstrom et al. Cancer Res. 2003; 63, 3695-3700).
Moreover, due to its small size it is speculated to be detectable in the urine of these patients. There are known polymorphisms for the WAP four-disulfide core domain protein 2 precursor sequence, including a variant comprising a change of serine-leucine to leucine-leucine cysteine (SL to LLC) at positions 71-72 of SEQ ID NO:3 and a variant comprising a change of serine to threonine (S to T) at position 101 of SEQ ID NO:3.

It is recognized in the art that accurate diagnosis and/or prognosis of cancer can not rely on the use of a single diagnostic marker, due to the multidimensional nature of various cancerous diseases. In many cases, individual markers by themselves do not have the required sensitivity and/or specificity to be routinely used for diagnosis. For example combining serum HE4 marker with the CA125 marker results in an increase sensitivity and specificity of detecting ovarian cancer compared with using CA125 alone (Montagnana et al., J. of Clinical Laboratory Analysis 2009; 23, 331-335; Moore et al. Gynecol Oncol. 2009; 112, 40 - 46).

Thus, there is a recognized need for, and it would be highly advantageous to have a panel of HE4 markers for diagnosing various types of cancer, particularly ovarian cancer.

**SUMMARY OF THE INVENTION**

The present invention relates to nucleic acid and amino acid sequences that are variants of known HE4 proteins, which can be used as diagnostic markers, preferably as serum markers and as targets for gene and antibody therapies.

The present invention is based in part on the unexpected discovery that cancerous tissues, particularly ovarian cancerous tissues, over-express a variant of HE4 protein, which is significantly different from the known HE4 protein. The variant polynucleotide having SEQ ID NO:1 encodes a polypeptide comprising 136 amino acids (SEQ ID NO:7) compared to the 124 amino-acid long wild type (SEQ ID NO:3), with a unique sequence of 62 amino acids (SEQ ID NO:10). The present invention now shows that the difference in the variant expression in healthy versus cancerous tissues is highly significant, and moreover, more pronounced compared to the difference found for hitherto known HE4 markers.

Thus, the variant proteins and polynucleotides of the invention provide unique diagnostic markers for ovarian cancer as well as for other HE4-associated cancer types, which may be used alone or in a combination with other markers.

According to certain aspects, the present invention provides HE4 variants that can be used as markers for diagnosis of cancer, particularly a cancer selected from the group consisting of ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma,
endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma, each providing a separate embodiment of the present invention.

According to certain embodiments, the HE4-variant of the invention is an isolated polynucleotide comprising a nucleic acid sequence as set forth in SEQ ID NO:1, or a fragment or a homologue thereof. According to one embodiment, the fragment of SEQ ID NO:1 comprises 15-1,000 nucleotides, typically 50-650 nucleotides comprising the nucleic acid sequence set forth in any one of SEQ ID NOs:37, 36, 14, 8 or a sequence at least about 85%, 90%, 95%, 96%, 97%, 98%, 99% homologous thereto. According to another embodiment the fragment consists of a nucleic acid sequence having SEQ ID NO:37, corresponding to the unique segment of the variant, encoding SEQ ID NO:10.

According to other embodiments, the novel HE4-variant is an isolated protein or polypeptide having an amino acid sequence as set forth in SEQ ID NO:7, or a fragment or a homologue thereof. According to one embodiment, the isolated protein or polypeptide is at least about 85%, 90%, 95%, 96%, 97%, 98%, 99% homologous to SEQ ID NO:7. According to another embodiment, the isolated polypeptide consist of SEQ ID NO:7. According to additional embodiments, the fragment of SEQ ID NO:7 comprises from 5 to 80 amino acids comprising the amino acid sequence set forth in anyone of SEQ ID NOs: 10, 11 or a sequence at least about 85%, 90%, 95%, 96%, 97%, 98%, 99% homologous thereto. According to one embodiment, the fragment of SEQ ID NO:7 consists of SEQ ID NO:10.

According to additional embodiments, the present invention provides an antibody or antibody fragment which specifically binds or recognizes any one of the foregoing HE4-variant polypeptides but which does not significantly interact with or recognize known HE4 proteins. Thus, the antibody or antibody fragment according to at least some embodiments of the present invention distinguishes between the HE4-variant and known HE4 proteins.

Reference to the antibody property of "specific interaction" or "recognition" is to be understood as non-covalent associations with a variance of affinity over several orders of magnitude. These terms are to be understood as relative with respect to an index molecule, for which the antibody is thought to have little to no specific interaction or recognition. In certain embodiments, the antibodies specifically interact or recognize a particular epitope or antigen determinant. According to one embodiment, the antibodies of this invention have a specificity such that the specific interaction with or binding to the antigen is at least about 2, or in another embodiment, at least about 5, or in still further embodiment, at least about 10-fold greater than interaction or binding observed under the same reaction conditions with a molecule that does not include the antigenic determinant.
According to at least some embodiments, the present invention provides an antibody or an antibody fragment capable of specifically binding to at least one epitope of a polypeptide comprising an amino acid sequence as set forth in any one of SEQ ID NOs: 10, 11, 7, or a fragment or a homologoue thereof. According to one embodiment the antibody is capable of specifically binding to at least one epitope of a polypeptides about 85%, 90%, 95%, 96%, 97%, 98%, or 99% homologous to any one of SEQ ID NOs: 10 and 11.

According to at least some embodiments, the antibody and/or antibody fragment are useful in detecting qualitative and/or quantitative changes in the expression of the HE4-variant polypeptides of the present invention.

The present invention now discloses that an increase in the level of HE4-variant(s) in a tissue is associated with cancer, particularly with a cancer selected from the group consisting of ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma. Thus, according to certain embodiments, detection of the HE4-polynucelotide or protein content provides a method for diagnosing these cancer types. It is to be explicitly understood that each of the cancer types is a separate embodiment of the invention.

According to additional aspects of the present invention there is provided a use of any one of the foregoing HE4-variant polypeptides and/or HE4-variant polynucleotides, and/or HE4-variant specific antibodies or antibody fragments, for diagnosis of cancer. According to certain typical embodiments, the cancer is selected from the group consisting of ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma, each is a separate embodiment of the present invention.

As used herein, the term "diagnosis" of cancer and/or "diagnosing" cancer encompasses screening for cancer, diagnosing cancer, detecting the presence of or severity of cancer, prognosis of cancer, early diagnosis of cancer, staging of cancer, monitoring of cancer progression and/or treatment efficacy and/or relapse of cancer, as well as selecting a therapy and/or a treatment for cancer, optimization of a given therapy for cancer, monitoring the treatment of cancer, and/or predicting the suitability of a therapy for specific patients or subpopulations or determining the appropriate dosing of a therapeutic product in patients or subpopulations, wherein each is a separate embodiment of the present invention.

In at least some embodiments the present invention provides a method for diagnosis of cancer, comprising detecting the level of at least one HE4-variant polypeptide and/or HE4-variant polynucleotide in a subject or in a sample obtained from the subject compared to a
control baseline. According to typical embodiments of the present invention a relative increase in the expression or the level of the polynucleotide or polypeptide compared to a control baseline herald the onset, severity, or prognosis of an individual with regard to cancer, particularly a cancer selected from the group consisting of ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma, each being a separate embodiment of the invention. The detection may comprise detection of the expression or level of a specific polypeptide or polynucleotide via any means known in the art, and as described herein.

According to certain embodiments, the control baseline expression is measured in a healthy subject or in a sample obtained therefrom. According to additional embodiments, the control baseline expression is measured in the diagnosed subject or in a sample obtained from said subject at an earlier time point. According to certain embodiments, the level of the HE4 variant is measured in a fluid sample. According to other embodiments, the level of the HE4 variant is measured in a tissue sample. Typically, the comparison to the control baseline is performed in a tissue and/or fluid specific manner.

In at least some embodiments the present invention provides a method for diagnosing of cancer, particularly at least one cancer selected from the group consisting of ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma in a subject, comprising detecting in a subject or in a sample obtained from the subject (a) at least one polynucleotide and/or polypeptide being a member of the group consisting of SEQ ID NOs:37, 36, 14, 8, 1; and/or SEQ ID NOs:10, 11, 7, respectively, or fragments or homologues thereof; and (b) measuring the level of the at least one polynucleotide and/or polypeptide in comparison to its level in a control baseline sample, wherein an increase in the level of said at least one polynucleotide and/or polypeptide compared to the control baseline is indicative of cancer.

According to certain embodiments, detecting any one of the HE4-variant polynucleotides or polypeptides is performed by employing a nucleic acid testing (NAT)-based assay (optionally by employing at least one nucleotide probe or primer), or by employing an immunoassay (optionally by employing an antibody according to any of the embodiments described herein), respectively.

According to at least some embodiments, the present invention provides a method for detecting a HE4-variant nucleic acid sequence in a biological sample, comprising: hybridizing
the isolated variant nucleic acid molecules or oligonucleotide fragments thereof of at least about 12 nucleotides to a nucleic acid material of the biological sample and detecting a hybridization complex; wherein the presence of the hybridization complex correlates with the presence of said variant nucleic acid sequence in the biological sample.

In at least some embodiments of the present invention, the methods are conducted on a subject in vivo.

In at least some embodiments of the present invention, the methods are conducted with a sample isolated from a subject having, being predisposed to, or suspected of having cancer. In at least some embodiments of the present invention, the sample is a cell or tissue or a body fluid sample.

According to typical embodiments, the sample is a body fluid or secretion sample including but not limited to seminal plasma, blood, serum, urine, prostatic fluid, seminal fluid, semen, the external secretions of the skin, respiratory, intestinal, and genitourinary tracts, tears, cerebrospinal fluid, sputum, saliva, milk, peritoneal fluid, pleural fluid, cyst fluid, broncho alveolar lavage, lavage of the reproductive system and/or lavage of any other part of the body or system in the body, and stool; or a tissue sample. The term may also optionally encompass samples of in vitro cell culture constituents. The sample can optionally be diluted or eluted with a suitable eluant before contacting the sample to an antibody and/or performing any other diagnostic assay. Each of the above-described possibilities is a separate embodiment of the present invention.

According to at least some embodiments there is provided any of the forgoing methods, wherein the cancer is selected from the group consisting of ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and urothelial cell carcinoma. Each of the cancer types is a separate embodiment of the present invention.

According to certain embodiments diagnosing at least one HE4 variant polypeptide of the present invention comprises the use of any of the foregoing antibodies of the present invention. According to certain embodiments, the at least one polypeptide is detected by employing a method selected from the group consisting of immunohistochemical assay, radioimaging assays, in-vivo imaging, positron emission tomography (PET), single photon emission computer tomography (SPECT), magnetic resonance imaging (MRI), optical imaging, computer tomography, radioimmunoassay (RIA), ELISA, slot blot, competitive binding assays, fluorimetric imaging assays, Western blot, FACS, and the like. According to at least some embodiments, the present invention includes diagnostic methods and or assays
which use any of the foregoing antibodies or fragments that specifically bind any of the polypeptides having an amino acid sequence as set forth in any one of SEQ ID NOs: 7, 10, 11, or a homologue or a fragment thereof, but not bind the wild type protein.

According to other embodiments, diagnosing at least one HE4 variant polynucleotide comprises the use of a NAT-based assay.

In at least some embodiments of the present invention, the NAT-based assay is selected from the group consisting of a PCR, Real-Time PCR, LCR, Self-Sustained Synthetic Reaction, Q-Beta Replicase, Cycling Probe Reaction, Branched DNA, RFLP analysis, DGGE/TGGE, Single-Strand Conformation Polymorphism, Dideoxy Fingerprinting, Microarrays, Fluorescence In Situ Hybridization or Comparative Genomic Hybridization.

In another embodiment the invention relates to an isolated polynucleotide comprising an amplicon having the nucleic acid sequence of SEQ ID NO: 14, or a segment having the nucleic acid sequence set forth in any one of SEQ ID NO: 8 and SEQ ID NO 37, or fragments or polynucleotides homologous thereto.

According to at least some embodiments of the present invention, detecting any of the foregoing HE4 variant polynucleotides employs a primer pair, comprising a pair of isolated oligonucleotides capable of specifically hybridizing to at least a portion of a polynucleotide having a nucleic acid sequence as set forth in any one of SEQ ID NOs: 37, 14, 8, 1 or a homologue thereof. According to certain embodiments, the primer pair specifically hybridizes to at least one polynucleotide at least 85%, 90%, 95%, 96%, 97%, 98%, 99% homologous to any one of SEQ ID NOs: 37, 14 and 8.

In another embodiment the invention relates to a primer pair comprising a pair of isolated oligonucleotides capable of amplifying the foregoing amplicon or segment.

In another embodiment the invention provides a primer pair comprising a pair of isolated oligonucleotides having the nucleic acid sequence selected from the group consisting of SEQ ID NOs: 12 and 13; 12 and 35; 38 and 13; and 38 and 35.

In at least some embodiments the present invention provides a diagnostic kit for diagnosis of cancer, comprising markers and reagents for detecting qualitative and/or quantitative changes in the expression of a polypeptide or a polynucleotide of the present invention. In at least some embodiments the present invention provides a diagnostic kit for diagnosis of cancer selected from the group consisting of ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and urothelial cell carcinoma, each is a separate embodiment of the invention.
In at least some embodiments of the present invention, the kit comprises markers and reagents for detecting the changes by employing a NAT-based technology.

In at least some embodiments of the present invention, the kit comprises at least one nucleotide probe or primer. In at least some embodiments of the present invention, the kit comprises at least one primer pair capable of selectively hybridizing to at least a portion of a polynucleotide having a nucleic acid sequence as set forth in any one of SEQ ID NOs: 36, 37, 14, 8, 1, or polynucleotides homologous thereto. According to certain embodiments, the kit comprises at least one primer pair capable of selectively hybridizing to at least a portion of a polynucleotide at least 85%, 90%, 95%, 96%, 97%, 98% or 99% homologous to any one of SEQ ID NOs: 36, 37, 14 and 8.

In at least some embodiments of the present invention, the kit comprises an antibody capable of specifically recognizing or interacting with at least a portion of any one of the HE4-variant polypeptides, having an amino acid sequence selected from the group consisting of SEQ ID NOs: 7, 10, 11 or a fragment or a variant or a homologue thereof, but not to a known HE4 protein. In at least some embodiments of the present invention, the kit further comprises at least one reagent for performing an immunohistochemical assay, radioimaging assays, in-vivo imaging, positron emission tomography (PET), single photon emission computer tomography (SPECT), magnetic resonance imaging (MRI), optical imaging, computer tomography, radioimmunoassay (RIA), ELISA, slot blot, competitive binding assays, fluorimetric imaging assays, Western blot, FACS, and the like.

According to certain embodiments, the markers of the present invention can be used alone or in combination with other markers, as described herein.

AU nucleic acid sequences and/or amino acid sequences, according to at least some embodiments of the invention, relate to their isolated form.

**BRIEF DESCRIPTION OF DRAWINGS**

The invention is herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of at least some embodiments of the present invention only, and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of the invention. In this regard, no attempt is made to show structural details of the invention in more detail than is necessary for a fundamental
understanding of the invention, the description taken with the drawings making apparent to those skilled in the art how the several forms of the invention may be embodied in practice.

In the drawings:

FIG. 1 is schematic summary of the oligonucleotide based microarray experimental flow.

FIG. 2 presents a comparison between X63187_1_P7 (SEQ ID NO:7) and the known protein WFDC2_HUMAN (SEQ ID NO:3).

FIG. 3 presents a comparison between X63187_1_P7 (SEQ ID NO:7) and the known protein Q14508-4 (SEQ ID NO:2).

FIGs. 4-5 show scatter plots, demonstrating the expression of X63187 transcripts that encode the X63187 proteins, on a virtual panel of all tissues and conditions using MED discovery engine. Figure 4 demonstrates overexpression of X63187 transcripts in ovarian cancer compared to normal ovarian samples. Description of the tissue samples is provided in Table 3 hereinbelow. Figure 5 demonstrates overexpression of X63187 transcripts in kidney cancer compared to normal kidney samples. Description of the tissue samples is provided in Table 4 hereinbelow.

FIG. 6 is a histogram showing over expression of the WFDC2 transcripts detectable by amplicon as depicted in the sequence X63187_seg7-1 l(100-930x100-933) (SEQ ID NO:14) in normal and cancerous ovary tissues.

FIG. 7 is a histogram showing over expression of the WFDC2 transcripts detectable by amplicon as depicted in the sequence X63187_seg7-1 l(100-930x100-933) (SEQ ID NO:14) in different normal tissues.

FIG. 8 demonstrates ethidium bromide gel analysis of the PCR product corresponding to a nucleic acid sequence encoding WFDC2_T12_P7. Lane 1 represents 1Kb DNA marker (Fermentas, Catalog number SMI 173); lane 4 represents the WFDC2JT12 PCR product.

FIG. 9 demonstrates the DNA specific sequence corresponding to a nucleic acid sequence encoding WFDC2_T12_P7. Sequence corresponding to the open reading frame (ORF) is marked in bold, while the primers sequences used are underlined.

FIG. 10 presents the amino acid sequence of WFDC2_T12_P7 (SEQ ID NO:7).

FIG. 11 demonstrates the expression of the known HE4 wild type (WT, probe 3886961); the variant X63187_1_T9 (SEQ ID NO:1) encoding the protein X63187_1_P7 (SEQ ID NO:7) (SV, probe 3886944); and a sequence corresponding to an unexpressed intron (probe 3886952). Each data point represents an average of three (3) samples (apart from the tissue mix which represents the average of 20 samples); error bars represent standard deviation. Y axis is an arbitrary fluorescent read units.
DETAILED DESCRIPTION OF THE INVENTION

According to at least some embodiments, the present invention provides HE4-variant polynucleotides, polypeptides, and specific antibodies and fragments and uses thereof, particularly as diagnostic markers.

The markers according to at least some embodiments of the present invention, alone or in combination, can be used for diagnosis of cancer, particularly ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma.

Biomolecular sequences (amino acid and/or nucleic acid sequences) uncovered using the methodology according to at least some embodiments of the present invention and described herein can be efficiently utilized as tissue or pathological markers and/or as drug targets for treating or preventing cancer.

The present invention now shows that the level of the HE4 variants detected in samples obtained from subjects having cancer is significantly higher compared to the level of these variants found under normal conditions. According to certain embodiments, the elevated level is found in the bloodstream. According to other embodiments, the elevated level is found in cancerous tissues or cells. According to typical embodiment, the cancer is selected from the group consisting of ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma. The measurement of these markers, alone or in combination, in patient samples provides information that the diagnostician can correlate with a probable diagnosis of cancer, particularly ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma, and/or a condition that it is indicative of a higher risk for cancer, particularly for ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma.

According to at least some embodiments, the present invention therefore also relates to diagnostic assays for cancer, particularly ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma, and methods of use of such markers for detection of cancer, particularly ovarian cancer and/or kidney cancer, optionally in a sample taken from a subject (patient), which is optionally some type of blood sample.
Some embodiments of this invention have been exemplified herein wherein cellular localization was determined according to four different software programs: (i) tmhmm (from Center for Biological Sequence Analysis, Technical University of Denmark DTU, http://wwwxbs.dm.dk/services/TMHMM/TMHMM2.0b.guide.php) or (ii) tmpred (from EMBnet, maintained by the ISREC Bioinformatics group and the LICR Information Technology Office, Ludwig Institute for Cancer Research, Swiss Institute of Bioinformatics, http://www.ch.embnet.org/software/TMPRED_form.html) for transmembrane region prediction; (iii) signalp_hmm or (iv) signalp_nn (both from Center for Biological Sequence Analysis, Technical University of Denmark DTU, http://www.cbs.dtu.dk/services/SignalP/background/prediction.php) for signal peptide prediction. The terms "signalp_hmm" and "signalp_nn" refer to two modes of operation for the program SignalP: hmm refers to Hidden Markov Model, while nn refers to neural networks. Localization was also determined through manual inspection of known protein localization and/or gene structure, and the use of heuristics by the individual inventor. In some cases for the manual inspection of cellular localization prediction inventors used the ProLoc computational platform (Einhart Hazkan-Covo, et al. "Evolution of multicellularity in metazoa: comparative analysis of the subcellular localization of proteins in Saccharomyces, Drosophila and Caenorhabditis." Cell Biology International 2004; 28(3), 171-8.), which predicts protein localization based on various parameters including, protein domains (e.g., prediction of transmembraneous regions and localization thereof within the protein), pi, protein length, amino acid composition, homology to pre-annotated proteins, recognition of sequence patterns which direct the protein to a certain organelle (such as, nuclear localization signal, NLS, mitochondria localization signal), signal peptide and anchor modeling and using unique domains from Pfam that are specific to a single compartment.

Information is given in the text with regard to SNPs (single nucleotide polymorphisms). A description of the abbreviations is as follows. "T -> C", for example, means that the SNP results in a change at the position given in the table from T to C. Similarly, "M -> Q", for example, means that the SNP has caused a change in the corresponding amino acid sequence, from methionine (M) to glutamine (Q). If, in place of a letter at the right hand side for the nucleotide sequence SNP, there is a space, it indicates that a frame shift has occurred. A frame shift may also be indicated with a hyphen (-). A stop codon is indicated with an asterisk at the right hand side (*). As part of the description of an SNP, a comment may be found in parentheses after the above description of the SNP itself. This comment may include an FTId, which is an identifier to a SwissProt entry that was
created with the indicated SNP. An FTId is a unique and stable feature identifier, which allows construction of links directly from position-specific annotation in the feature table to specialized protein-related databases. The FTId is always the last component of a feature in the description field, as follows: FTId=XXX_number, in which XXX is the 3-letter code for the specific feature key, separated by an underscore from a 6-digit number. In the table of the amino acid mutations of the wild type proteins of the selected variants of the invention, the header of the first column is "SNP position(s) on amino acid sequence", representing a position of a known mutation on amino acid sequence. SNPs may optionally be used as diagnostic markers according to at least some embodiments the present invention, alone or in combination with one or more other SNPs and/or any other diagnostic marker. According to at least some embodiments, the present invention comprise such SNPs, including but not limited to novel SNPs on the known (WT or wild type) protein sequences given below, as well as novel nucleic acid and/or amino acid sequences formed through such SNPs, and/or any SNP on a variant amino acid and/or nucleic acid sequence described herein.

Information given in the text with regard to the Homology to the known proteins was determined by Smith-Waterman version 5.1.2 using special (non default) parameters as follows:
- model=sw.model
- GAPEXT=O
- GAPOP=100.0
- MATRIX=blosuml

Some embodiments of this invention have been exemplified herein wherein overexpression of a cluster in cancer was a determination based on microarray use. As a microarray reference, in the specific segment paragraphs, the unabbreviated tissue name was used as the reference to the type of chip for which expression was measured. There are two types of microarray results: those from microarrays prepared according to a design by the present inventors, for which the microarray fabrication procedure is described in detail in Materials and Experimental Procedures section herein; and those results from microarrays using Affymetrix technology. As a microarray reference, in the specific segment paragraphs, the unabbreviated tissue name was used as the reference to the type of chip for which expression was measured. For microarrays prepared according to a design by the present inventors, the probe name begins with the name of the cluster (gene), followed by an identifying number.

Oligonucleotide microarray results taken from Affymetrix data were from chips available from Affymetrix Inc, Santa Clara, CA, USA (see for example data regarding the


It should be noted that the terms "segment", "seg" and "node" (abbreviated as "N" in the names of nodes) are used interchangeably in reference to nucleic acid sequences according to at least some embodiments of the present invention, they refer to portions of nucleic acid sequences that were shown to have one or more properties as described herein. They are also the building blocks that were used to construct complete nucleic acid sequences as described in greater detail elsewhere herein. Optionally, they are examples of oligonucleotides which are at least some embodiments of the present invention, for example as amplicons, hybridization units and/or from which primers and/or complementary oligonucleotides may optionally be derived, and/or for any other use.
In order that the present invention may be more readily understood, certain terms are first defined. Additional definitions are set forth throughout the detailed description.

The term "homology", as used herein, refers to a degree of sequence similarity in terms of shared amino acid or nucleotide sequences. There may be partial homology or complete homology (i.e., identity). For amino acid sequence homology amino acid similarity matrices may be used as are known in different bioinformatics programs (e.g. BLAST, Smith Waterman). Different results may be obtained when performing a particular search with a different matrix. Homologous peptide or polypeptides are characterized by one or more amino acid substitutions, insertions or deletions, such as, but not limited to, conservative substitutions, provided that these changes do not affect the biological activity of the peptide or polypeptide as described herein.

Degrees of homology for nucleotide sequences are based upon identity matches with penalties made for gaps or insertions required to optimize the alignment, as is well known in the art (e.g. Altschul S. F. et al, 1990, J Mol Biol 215(3):403-10; Altschul S.F. et al, 1997, Nucleic Acids Res. 25:3389-3402). The degree of sequence homology is presented in terms of percentage, e.g. "70% homology". As used herein, the term "at least" with regard to a certain degree of homology encompasses any degree of homology from the specified percentage up to 100%.

The terms "correspond" or "corresponding to" or "correspondence with" are used herein to indicate identity between two corresponding amino acid or nucleic acid sequences.

In some embodiments, the proteins or polypeptides of this invention comprise chimeric protein or polypeptides.

As used herein, the terms "chimeric protein or polypeptide", or "chimeric polynucleotide" or "chimera" refers to an assembly or a string of amino acids in a particular sequence, or nucleotides encoding the same, respectively, which does not correspond in their entirety to the sequence of the known (wild type) polypeptide or protein, or the nucleic acid encoding same.

In some embodiments, the variants of this invention are derived from two exons, or an exon and an intron of a known protein, or fragments thereof, or segments having sequences with the indicated homology.

In some embodiments, the phrase "marker" in the context of the present invention refers to a nucleic acid fragment, a peptide, or a polypeptide, which is differentially present in a sample taken from patients (subjects) having one of the herein-described diseases or
conditions, as compared to a comparable sample taken from subjects who do not have one of the above-described diseases or conditions.

In some embodiments, the term "polypeptide" is to be understood to refer to a molecule comprising from at least 2 to several thousand or more amino acids. The term "polypeptide" is to be understood to include, *inter alia*, native peptides (either degradation products, synthetically synthesized peptides or recombinant peptides), peptidomimetics, such as peptoids and semipeptoids or peptide analogs, which may comprise, for example, any desirable modification, *inter alia*, modifications rendering the peptides more stable while in a body or more capable of penetrating into cells, or others as will be appreciated by one skilled in the art. Such modifications include, but are not limited to N terminus modification, C terminus modification, peptide bond modification, backbone modifications, residue modification, or others. Inclusion of such peptides within the polypeptides of this invention may produce a polypeptide sharing identity with the polypeptides described herein, for example, those provided in the sequence listing.

In some embodiments, the phrase "differentially present" refers to differences in the quantity or quality of a marker present in a sample taken from patients having one of the herein-described diseases or conditions as compared to a comparable sample taken from patients who do not have one of the herein-described diseases or conditions. For example, a nucleic acid fragment may optionally be differentially present between the two samples if the amount of the nucleic acid fragment in one sample is significantly different from the amount of the nucleic acid fragment in the other sample, for example as measured by hybridization and/or NAT-based assays. A polypeptide is differentially present between the two samples if the amount of the polypeptide in one sample is significantly different from the amount of the polypeptide in the other sample. It should be noted that if the marker is detectable in one sample and not detectable in the other, then such a marker can be considered to be differentially present. According to certain embodiments, the marker quantity is elevated in samples obtained from a subject having the disease compared to a control baseline, wherein the control baseline can be the marker quantity in a sample obtained from healthy subject or in a sample obtained from the same subject at an earlier stage. One of ordinary skill in the art could easily determine such relative levels of the markers; further guidance is provided in the description of each individual marker below.

In some embodiments, the phrase "diagnostic" means identifying the presence or nature of a pathologic condition. Diagnostic methods differ in their sensitivity and specificity. The "sensitivity" of a diagnostic assay is the percentage of diseased individuals who test
positive (percent of "true positives"). Diseased individuals not detected by the assay are "false negatives." Subjects who are not diseased and who test negative in the assay are termed "true negatives." The "specificity" of a diagnostic assay is 1 minus the false positive rate, where the "false positive" rate is defined as the proportion of those without the disease who test positive. While a particular diagnostic method may not provide a definitive diagnosis of a condition, it suffices if the method provides a positive indication that aids in diagnosis.

In some embodiments, the phrase "qualitative" when in reference to differences in expression levels of a polynucleotide, polypeptide or cluster as described herein, refers to the presence versus absence of expression, or in some embodiments, the temporal regulation of expression, or in some embodiments, the timing of expression, or in some embodiments, the variant expressed, or in some embodiments, any post-translational modifications to the expressed molecule, and others, as will be appreciated by one skilled in the art. In some embodiments, the phrase "quantitative" when in reference to differences in expression levels of a polynucleotide, polypeptide or cluster as described herein, refers to absolute differences in quantity of expression, as determined by any means, known in the art, or in other embodiments, relative differences, which may be statistically significant, or in some embodiments, when viewed as a whole or over a prolonged period of time, etc., indicate a trend in terms of differences in expression.

The term "cancer" as used herein should be understood to encompass any neoplastic disease (whether invasive or metastatic) which is characterized by abnormal and uncontrolled cell division causing malignant growth or tumor. Non-limiting examples of cancer which may be treated with a composition according to at least some embodiments of the present invention are solid tumors, sarcomas, hematological malignancies, including but not limited to breast cancer (e.g. breast carcinoma), cervical cancer, ovary cancer (ovary carcinoma), endometrial cancer, melanoma, bladder cancer (bladder carcinoma), lung cancer (e.g. adenocarcinoma and non-small cell lung cancer), pancreatic cancer (e.g. pancreatic carcinoma such as exocrine pancreatic carcinoma), colon cancer (e.g. colorectal carcinoma, such as colon adenocarcinoma and colon adenoma), gastric cancer, urothelial cell carcinomas, prostate cancer including the advanced disease, hematopoietic tumors of lymphoid lineage (e.g. leukemia, acute lymphocytic leukemia, chronic lymphocytic leukemia, B-cell lymphoma, Burkitt's lymphoma, multiple myeloma, Hodgkin's lymphoma, Non-Hodgkin's lymphoma), myeloid leukemia (for example, acute myelogenous leukemia (AML), chronic myelogenous leukemia), thyroid cancer, thyroid follicular cancer, myelodysplastic syndrome (MDS), tumors of mesenchymal origin (e.g. fibrosarcomas and rhabdomyosarcomas), melanoma,
uveal melanoma, teratocarcinoma, neuroblastoma, glioma, glioblastoma, benign tumor of the skin (e.g. keratoacanthomas), renal cancer, anaplastic large-cell lymphoma, esophageal squamous cells carcinoma, hepatocellular carcinoma, follicular dendritic cell carcinoma, intestinal cancer, muscle-invasive cancer, seminal vesicle tumor, epidermal carcinoma, spleen cancer, head and neck cancer, stomach cancer, liver cancer, bone cancer, brain cancer, cancer of the retina, biliary cancer, small bowel cancer, salivary gland cancer, cancer of uterus, cancer of testicles, cancer of connective tissue, prostatic hypertrophy, myelodysplasia, Waldenstrom's macroglobulaemia, nasopharyngeal, neuroendocrine cancer, myelodysplastic syndrome, mesothelioma, angiosarcoma, Kaposi's sarcoma, carcinoid, oesophagogastric, fallopian tube cancer, peritoneal cancer, papillary serous mullerian cancer, malignant ascites, gastrointestinal stromal tumor (GIST), and a hereditary cancer syndrome such as Li-Fraumeni syndrome and Von Hippel-Lindau syndrome (VHL).

With regard to ovarian cancer, the disease is selected from the group including but not limited to primary and metastatic cancer of the ovary, including epithelial ovarian cancer such as serous, mucinous, endometroid, clear cell, mixed epithelial, undifferentiated carcinomas and Brenner tumor, as well as other non-epithelial neoplasms of the ovary, including germ cell malignancies.

With regard to ovarian cancer, the polypeptides and/or polynucleotide of cluster X63187 according to at least some embodiments of the present invention can be used in the diagnosis, treatment or prognostic assessment of invasive or metastatic ovarian cancer; correlating stage and malignant potential; identification of a metastasis of unknown origin which originated from a primary ovarian cancer; differential diagnosis between benign and malignant ovarian cysts; diagnosing a cause of infertility, for example differential diagnosis of various causes thereof; detecting of one or more non-ovarian cancer conditions that may elevate serum levels of ovary related markers, including but not limited to: cancers of the endometrium, cervix, fallopian tubes, pancreas, breast, lung and colon; nonmalignant conditions such as pregnancy, endometriosis, pelvic inflammatory disease and uterine fibroids; diagnosing conditions which have similar symptoms, signs and complications as ovarian cancer and where the differential diagnosis between them and ovarian cancer is of clinical importance including but not limited to: non-malignant causes of pelvic mass, including, but not limited to: benign (functional) ovarian cyst, uterine fibroids, endometriosis, benign ovarian neoplasms and inflammatory bowel lesions; determining a cause of any condition suggestive of a malignant tumor including but not limited to anorexia, cachexia,
weight loss, fever, hypercalcemia, skeletal or abdominal pain, paraneoplastic syndrome, or ascites.

According to at least some embodiments, the polypeptides and/or polynucleotides of cluster X63187 used in the diagnosis of ovarian cancer can be used alone or in combination with one or more known markers for ovarian cancer, including but not limited to CA-125, CEA, CA125 (Mucin 16), CA72-4TAG, CA-50, CA 54-61, CA-195 and CA 19-9, STN and TAG-72, kallikreins, cathepsin L, urine gonadotropin, inhibins, cytokeratins, such as TPA and TPS, members of the Transforming Growth Factors (TGF) beta superfamily, Epidermal Growth Factor, p53 and HER-2, or any combination thereof, and/or in combination with the known HE4 associated with the indicated polypeptide or polynucleotide, as described herein.

In another embodiment, the X63187 cluster, showing age and/or stage differential diagnostic capability, according to at least some embodiments the present invention provides diagnostic methods, kits and assays for diagnosis, assessment and prognostic indications regarding the indicated disease disorder or condition, according to the age of the subject and/or stage of the condition, as described herein.

Immunohistochemistry may be used to assess the origin of the tumor and staging as part of the methods of this invention, and as protected uses for the polypeptides of this invention.

In some embodiments, this invention provides polypeptides/polynucleotides which serves as markers for ovarian cancer. In some embodiments, the marker is any polypeptide/polynucleotide as described herein. In some embodiments, the marker is X63187, or variants as described herein or markers related thereto. Each variant marker according to at least some embodiments of the present invention described herein may be used alone or in combination with one or more other variant ovarian cancer described herein, and/or in combination with known markers for ovarian cancer, as described herein. Diagnosis of ovarian cancer and/or of other conditions that may be diagnosed by these markers or variants of them, include but are not limited to the presence, risk and/or extent of the following:

1. The identification of a metastasis of unknown origin which originated from a primary ovarian cancer.
2. As a marker to distinguish between different types of ovarian cancer, therefore potentially affect treatment choice (e.g. discrimination between epithelial tumors and germ cell tumors).
3. As a tool in the assessment of abdominal mass and in particular in the differential diagnosis between a benign and malignant ovarian cysts.
4. As a tool for the assessment of infertility.
5. Other conditions that may elevate serum levels of ovary related markers. These include but are not limited to: cancers of the endometrium, cervix, fallopian tubes, pancreas, breast, lung and colon; nonmalignant conditions such as pregnancy, endometriosis, pelvic inflammatory disease and uterine fibroids.
6. Conditions which have similar symptoms, signs and complications as ovarian cancer and where the differential diagnosis between them and ovarian cancer is of clinical importance including but not limited to:
   a. Non-malignant causes of pelvic mass. Including, but not limited to: benign (functional) ovarian cyst, uterine fibroids, endometriosis, benign ovarian neoplasms and inflammatory bowel lesions
   b. Any condition suggestive of a malignant tumor including but not limited to anorexia, cachexia, weight loss, fever, hypercalcemia, skeletal or abdominal pain, paraneoplastic syndrome.
   c. Ascites.
7. Prediction of patient's drug response
8. As surrogate markers for clinical outcome of a treated cancer.
9. Screening for early detection of ovarian cancer.

With regard to renal cancer, the disease is selected from the group consisting of but not limited to primary and metastatic cancer of the kidney, including renal cell carcinoma (i.e. renal adenocarcinoma), as well as other non-epithelial neoplasms of the ovary, including nephroblastoma (i.e. Wilm's tumor), transitional cell neoplasms of the renal pelvis, and various sarcomas of renal origin. According to at least some embodiments, the polypeptides and/or polynucleotides of cluster X63 187 used in the diagnosis of renal cancer can be used alone or in combination with known markers for renal cancer, including but not limited to vascular endothelial growth factor, interleukin-12, the soluble interleukin-2 receptor, intercellular adhesion molecule-1, human chorionic gonadotropin beta, insulin-like growth factor-1 receptor, Carbonic anhydrase 9 (CA 9), endostatin, Thymidine phosphorylase, and/or in combination with the known HE4 associated with the indicated polypeptide or polynucleotide, as described herein.

With regard to breast cancer, the disease is selected from the group including but not limited to primary and metastatic cancer of the breast, including mammary carcinomas such as Infiltrating Ductal carcinoma, Ductal carcinoma in-situ, Infiltrating Lobular carcinoma, Lobular carcinoma in-situ, Inflammatory breast cancer, Paget's disease of the breast, and other
non-epithelial neoplasms of the breast, including fibrosarcomas, leiomyosarcomas, rhabdomyosarcomas, angiosarcomas, cystosarcoma phylloides. According to at least some embodiments, the polypeptides and/or polynucleotides of cluster X63187 used in the diagnosis of breast cancer can be used alone or in combination with known markers for breast cancer, including but not limited to Calcitonin, CA1 5-3 (Mucinl), CA27-29, TPA, a combination of CA 15-3 and CEA, CA 27.29 (monoclonal antibody directed against MUC1), Estrogen 2 (beta), HER-2 (c-erbB2), and/or in combination with the known HE4 associated with the indicated polypeptide or polynucleotide, as described herein.

With regard to lung cancer, the disease is selected from the group consisting of but not limited to squamous cell lung carcinoma, lung adenocarcinoma, carcinoma, small cell lung cancer or non-small cell lung cancer. According to at least some embodiments, the polypeptides and/or polynucleotides of cluster X63187 used in the diagnosis of lung cancer can be used alone or in combination with known markers for lung cancer, including but not limited to CEA, CA1 5-3, Beta-2-microglobulin, CA1 9-9, TPA, and/or in combination with the known HE4 associated with the indicated polypeptide or polynucleotide, as described herein.

With regard to pancreatic cancer, the disease is selected from the group consisting of but not limited to primary and metastatic cancers of the exocrine pancreas, including adenocarcinoma, serous and mucinous cystadenocarcinomas, acinar cell carcinoma, undifferentiated carcinoma, pancreatoblastoma and neuroendocrine tumors such as insulinoma. According to at least some embodiments, the polypeptides and/or polynucleotides of cluster X63187 used in the diagnosis of pancreatic cancer can be used alone or in combination with known markers for pancreatic cancer, including but not limited to CA 19-9, and/or in combination with the known HE4 associated with the indicated polypeptide or polynucleotide, as described herein.

With regard to endometrial cancer, the disease is selected from the group including but not limited to primary and metastatic cancer of the endometrial, including endometrioid adenocarcinoma, adenosquamous carcinoma, including squamous cells, papillary serous carcinoma, clear cell endometrial carcinoma, and uterine sarcoma. According to at least some embodiments, the polypeptides and/or polynucleotides of cluster X63187 used in the diagnosis of endometrial cancer can be used alone or in combination with known markers for endometrial cancer, including but not limited to HER2, CA125, and/or in combination with the known HE4 associated with the indicated polypeptide or polynucleotide, as described herein.
With regard to gastric cancer, the disease is selected from the group including but not limited to primary and metastatic cancer of the stomach; including gastric adenocarcinoma, intestinal type adenocarcinoma, diffuse type adenocarcinoma, gastrointestinal carcinoma, gastric metaplasia, gastrointestinal stromal tumor (GIST, leiomyomas or leiomyosarcomas), carcinoids tumors, distal gastric adenocarcinoma, diffuse carcinoma, cardiac carcinoma MALT lymphoma, small and large cell types. According to at least some embodiments, the polypeptides and/or polynucleotides of cluster X63187 used in the diagnosis of gastric cancer can be used alone or in combination with known markers for gastric cancer, including but not limited to NUDT6 (both transmembrane and circulating), CEA, CA19-9, c-KIT, and/or in combination with the known HE4 associated with the indicated polypeptide or polynucleotide, as described herein.

With regard to lung mesothelial tumor, the disease is selected from the group including but not limited to primary and metastatic cancer of the pleura; including epithelioid mesothelioma, sarcomatoid mesothelioma, desmoplastic mesothelioma, biphasic mesothelioma. According to at least some embodiments, the polypeptides and/or polynucleotides of cluster X63187 used in the diagnosis of lung mesothelial tumor can be used alone or in combination with known markers for lung mesothelial tumor, including but not limited to soluble mesothelin-related peptides (SMRP), and/or in combination with the known HE4 associated with the indicated polypeptide or polynucleotide, as described herein.

With regard to urothelial cell carcinoma, the disease encompasses renal pelvis and ureteral tumor, and urinary bladder tumor. With regard to renal pelvis and ureteral tumor, the disease is selected from the group including but not limited to primary and metastatic cancer of the renal pelvis and ureter, including transitional cell carcinoma and papilloma. With regard to urinary bladder tumor, the disease is selected from the group including but not limited to primary and metastatic cancer of the urinary bladder, including papilloma, adenoma, carcinoma (transitional cell carcinoma including papillary and infiltrating, papillary and noninfiltrating, nonpapillary and infiltrating, carcinoma in situ), squamous cell carcinoma, adenocarcinoma, and undifferentiated carcinoma. According to at least some embodiments, the polypeptides and/or polynucleotides of cluster X63187 used in the diagnosis of urothelial cell carcinoma can be used alone or in combination with known markers for urothelial cell carcinoma, including but not limited to p53, bcl2, p16, p21, and EGFR family mutations, and/or in combination with the known HE4 associated with the indicated polypeptide or polynucleotide, as described herein.

21
As used herein the term "diagnosis" refers to the process of identifying a medical condition or disease by its signs, symptoms, and in particular from the results of various diagnostic procedures, including e.g. detecting the expression of the nucleic acids or polypeptides according to at least some embodiments of the invention in a biological sample (e.g. in cells, tissue or serum, as defined below) obtained from an individual. Furthermore, as used herein the term "diagnosis" encompasses screening for a disease, detecting a presence or a severity of a disease, distinguishing a disease from other diseases including those diseases that may feature one or more similar or identical symptoms, providing prognosis of a disease, monitoring disease progression or relapse, as well as assessment of treatment efficacy and/or relapse of a disease, disorder or condition, as well as selecting a therapy and/or a treatment for a disease, optimization of a given therapy for a disease, monitoring the treatment of a disease, and/or predicting the suitability of a therapy for specific patients or subpopulations or determining the appropriate dosing of a therapeutic product in patients or subpopulations. The diagnostic procedure can be performed in vivo or in vitro. It should be noted that a "biological sample obtained from the subject" may also optionally comprise a sample that has not been physically removed from the subject.

The terms "diagnosing" and "detecting" may also optionally encompass any of the above.

Diagnosis of a disease according to at least some embodiments of the present invention can, in some embodiments, be affected by determining a level of a polynucleotide or a polypeptide according to at least some embodiments of the present invention in a biological sample obtained from the subject, wherein the level determined can be correlated with predisposition to, or presence or absence of the disease.

In some embodiments, the term "level" refers to expression levels of RNA and/or protein or to DNA copy number of a marker according to at least some embodiments of the present invention.

Typically the level of the marker in a biological sample obtained from the subject is different (i.e., increased) from the level of the same variant in a similar sample obtained from a healthy individual (examples of biological samples are described herein). Alternatively, the level of the marker in a biological sample obtained from the subject is different (i.e., increased) from the level of the same variant in a similar sample obtained from the same subject at an earlier time point.
Numerous well known tissue or fluid collection methods can be utilized to collect the biological sample from the subject in order to determine the level of DNA, RNA and/or polypeptide of the variant of interest in the subject.

Examples include, but are not limited to, fine needle biopsy, needle biopsy, core needle biopsy and surgical biopsy (e.g., brain biopsy), and lavage. Regardless of the procedure employed, once a biopsy/sample is obtained the level of the variant can be determined and a diagnosis can thus be made.

Determining the level of the same variant in normal tissues of the same origin is preferably effected along-side to detect an elevated expression and/or amplification of the variant as opposed to the normal tissues.

In some embodiments, the term "test amount" of a marker refers to an amount of a marker in a subject's sample that is consistent with a diagnosis of a particular disease or condition. A test amount can be either in absolute amount (e.g., microgram/ml) or a relative amount (e.g., relative intensity of signals).

In some embodiments, the term "control amount" or "control baseline" or "amount in an index sample" of a marker can be any amount or a range of amounts to be compared against a test amount of a marker. For example, a control amount of a marker can be the amount of a marker in a patient with a particular disease or condition at an earlier stage or a person without such a disease or condition. The control amount may be measured in a subject or in a sample obtained from the subject. According to certain embodiment the sample can be a body fluid or a tissue sample. Typically, measurements of the control amount and test amount are preformed in a fluid/tissue specific manner. A control amount can be either in absolute amount (e.g., microgram/ml) or a relative amount (e.g., relative intensity of signals).

In some embodiments, the term "detect" refers to identifying the presence, absence or amount of the object to be detected.

In some embodiments, the term "label" includes any moiety or item detectable by spectroscopic, photo chemical, biochemical, immunochemical, or chemical means. For example, useful labels include $^{32}\text{P}$, $^{35}\text{S}$, fluorescent dyes, electron-dense reagents, enzymes (e.g., as commonly used in an ELISA), biotin-streptavadin, dioxigenin, haptens and proteins for which antisera or monoclonal antibodies are available, or nucleic acid molecules with a sequence complementary to a target. The label often generates a measurable signal, such as a radioactive, chromogenic, or fluorescent signal, that can be used to quantify the amount of bound label in a sample. The label can be incorporated in or attached to a primer or probe either covalently, or through ionic, van der Waals or hydrogen bonds, e.g., incorporation of
radioactive nucleotides, or biotinylated nucleotides that are recognized by streptavidin. The label may be directly or indirectly detectable. Indirect detection can involve the binding of a second label to the first label, directly or indirectly. For example, the label can be the ligand of a binding partner, such as biotin, which is a binding partner for streptavidin, or a nucleotide sequence, which is the binding partner for a complementary sequence, to which it can specifically hybridize. The binding partner may itself be directly detectable, for example, an antibody may be itself labeled with a fluorescent molecule. The binding partner also may be indirectly detectable, for example, a nucleic acid having a complementary nucleotide sequence can be a part of a branched DNA molecule that is in turn detectable through hybridization with other labeled nucleic acid molecules (see, e.g., P. D. Fahrlander and A. Klauser, Bio/Technology 6:1165 (1988)). Quantitation of the signal is achieved by, e.g., scintillation counting, densitometry, or flow cytometry.

Exemplary detectable labels, optionally for use with immunoassays, include but are not limited to magnetic beads, fluorescent dyes, radiolabels, enzymes (e.g., horse radish peroxide, alkaline phosphatase and others commonly used in an ELISA), and calorimetric labels such as colloidal gold or colored glass or plastic beads. Alternatively, the marker in the sample can be detected using an indirect assay, wherein, for example, a second, labeled antibody is used to detect bound marker-specific antibody, and/or in a competition or inhibition assay wherein, for example, a monoclonal antibody which binds to a distinct epitope of the marker are incubated simultaneously with the mixture.

"Immunoadsay" is an assay that uses an antibody to specifically bind an antigen. The immunoassay is characterized by the use of specific binding properties of a particular antibody to isolate, target, and/or quantify the antigen.

The phrase "specifically (or selectively) binds" to an antibody or "specifically (or selectively) immunoreactive with," or "specifically interacts or binds" when referring to a protein or peptide (or other epitope), refers, in some embodiments, to a binding reaction that is determinative of the presence of the protein in a heterogeneous population of proteins and other biologies. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein at least two times greater than the background (non-specific signal) and do not substantially bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, polyclonal antibodies raised to seminal basic protein from specific species such as rat, mouse, or human can be selected to obtain only those polyclonal antibodies that are specifically immunoreactive with seminal basic protein.
and not with other proteins, except for polymorphic variants and alleles of seminal basic protein. This selection may be achieved by subtracting out antibodies that cross-react with seminal basic protein molecules from other species. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (see, e.g., Harlow & Lane, Antibodies, A Laboratory Manual (1988), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity). Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

According to at least some embodiments, the present invention relates to bridges, tails, heads and/or insertions, and/or analogs, homologs and derivatives of such peptides. Such bridges, tails, heads and/or insertions are described in greater detail below with regard to the Examples.

As used herein a term "tail" refers to a peptide sequence at the end of an amino acid sequence that is unique to a variant according to at least some embodiments of the present invention. Therefore, a variant having such a tail may optionally be considered as a chimera, in that at least a first portion of the variant is typically highly homologous (often 100% identical) to a portion of the corresponding known protein, while at least a second portion of the variant comprises the tail.

As used herein a term "head" refers to a peptide sequence at the beginning of an amino acid sequence that is unique to a variant according to at least some embodiments of the present invention. Therefore, a variant having such a head may optionally be considered as a chimera, in that at least a first portion of the variant comprises the head, while at least a second portion is typically highly homologous (often 100% identical) to a portion of the corresponding known protein.

As used herein a term "an edge portion" refers to a connection between two portions of a variant according to at least some embodiments of the present invention that were not joined in the wild type or known protein. An edge may optionally arise due to a join between the above "known protein" portion of a variant and the tail, for example, and/or may occur if an internal portion of the wild type sequence is no longer present, such that two portions of the sequence are now contiguous in the variant that were not contiguous in the known protein. A "bridge" may optionally be an edge portion as described above, but may also include a join between a head and a "known protein" portion of a variant, or a join between a tail and a "known protein" portion of a variant, or a join between an insertion and a "known protein"
portion of a variant.

In some embodiments, a bridge between a tail or a head or a unique insertion, and a
"known protein" portion of a variant, comprises at least about 10 amino acids, or in some
embodiments at least about 20 amino acids, or in some embodiments at least about 30 amino
acids, or in some embodiments at least about 40 amino acids, in which at least one amino acid
is from the tail/head/insertion and at least one amino acid is from the "known protein" portion
of a variant. In some embodiments, the bridge may comprise any number of amino acids
from about 10 to about 40 amino acids (for example, 10, 11, 12, 13... 37, 38, 39, 40 amino
acids in length, or any number in between).

It should be noted that a bridge cannot be extended beyond the length of the sequence
in either direction, and it should be assumed that every bridge description is to be read in such
manner that the bridge length does not extend beyond the sequence itself.

Furthermore, bridges are described with regard to a sliding window in certain contexts
below. For example, certain descriptions of the bridges feature the following format: a bridge
between two edges (in which a portion of the known protein is not present in the variant) may
optionally be described as follows: a bridge portion of CONTIG-NAME_P1 (representing the
name of the protein), comprising a polypeptide having a length "n", wherein n is at least about
10 amino acids in length, optionally at least about 20 amino acids, at least about 30 amino
acids, at least about 40 amino acids and at least about 50 amino acids in length, wherein at
least two amino acids comprise XX (2 amino acids in the center of the bridge, one from each
end of the edge), having a structure as follows (numbering according to the sequence of
CONTIG-NAME_P1): a sequence starting from any of amino acid numbers 49-x to 49 (for
example); and ending at any of amino acid numbers 50 + ((n-2) - x) (for example), in which x
varies from 0 to n-2. In this example, it should also be read as including bridges in which n is
any number of amino acids between 10-50 amino acids in length. Furthermore, the bridge
polypeptide cannot extend beyond the sequence, so it should be read such that 49-x (for
example) is not less than 1, nor 50 + ((n-2) - x) (for example) greater than the total sequence
length.

In another embodiment, this invention provides a method for detecting a
polynucleotide of this invention in a biological sample, comprising: hybridizing the isolated
nucleic acid molecules or oligonucleotide fragments of at least about a minimum length to a
nucleic acid material of a biological sample and detecting a hybridization complex; wherein
the presence of a hybridization complex correlates with the presence of a the polynucleotide
in the biological sample.
According to some embodiments of the present invention, any marker according to the present invention may optionally be used alone or combination. Such a combination may optionally comprise a plurality of markers described herein, optionally including any subcombination of markers, and/or a combination featuring at least one other marker, for example a known marker. Furthermore, such a combination may optionally be used as described above with regard to determining a ratio between a quantitative or semi-quantitative measurement of any marker described herein to any other marker described herein, and/or any other known marker, and/or any other marker.

The plurality of markers is preferably then correlated with cancer. For example, such correlating may optionally comprise determining the concentration of each of the plurality of markers, and individually comparing each marker concentration to a threshold level. Optionally, if the marker concentration is above or below the threshold level (depending upon the marker and/or the diagnostic test being performed), the marker concentration correlates with cancer. Optionally, a plurality of marker concentrations correlates with cancer.

Alternatively, such correlating may optionally comprise determining the concentration of each of the plurality of markers, calculating a single index value based on the concentration of each of the plurality of markers, and comparing the index value to a threshold level.

Also alternatively, such correlating may optionally comprise determining a temporal change in at least one of the markers, and wherein the temporal change is used in the correlating step.

Also alternatively, such correlating may optionally comprise determining whether at least "X" number of the plurality of markers has a concentration outside of a predetermined range and/or above or below a threshold (as described above). The value of "X" may optionally be one marker, a plurality of markers or all of the markers; alternatively or additionally, rather than including any marker in the count for "X", one or more specific markers of the plurality of markers may optionally be required to correlate with cancer (according to a range and/or threshold).

Also alternatively, such correlating may optionally comprise determining whether a ratio of marker concentrations for two markers is outside a range and/or above or below a threshold. Optionally, if the ratio is above or below the threshold level and/or outside a range, the ratio correlates with cancer.

Optionally, a combination of two or more these correlations may be used with a single panel and/or for correlating between a plurality of panels.
Optionally, the method distinguishes a disease or condition with a sensitivity of at least 70% at a specificity of at least 85% when compared to normal subjects. As used herein, sensitivity relates to the number of positive (diseased) samples detected out of the total number of positive samples present; specificity relates to the number of true negative (non-diseased) samples detected out of the total number of negative samples present. Optionally, the method distinguishes a disease or condition with a sensitivity of at least 80% at a specificity of at least 90% when compared to normal subjects. Optionally, the method distinguishes a disease or condition with a sensitivity of at least 90% at a specificity of at least 90% when compared to normal subjects. Also optionally, the method distinguishes a disease or condition with a sensitivity of at least 70% at a specificity of at least 85% when compared to subjects exhibiting symptoms that mimic disease or condition symptoms.

A marker panel may be analyzed in a number of fashions well known to those of skill in the art. For example, each member of a panel may be compared to a "normal" value, or a value indicating a particular outcome. A particular diagnosis/prognosis may depend upon the comparison of each marker to this value; alternatively, if only a subset of markers is outside of a normal range, this subset may be indicative of a particular diagnosis/prognosis. The skilled artisan will also understand that diagnostic markers, differential diagnostic markers, prognostic markers, time of onset markers, disease or condition differentiating markers, etc., may be combined in a single assay or device. Markers may also be commonly used for multiple purposes by, for example, applying a different threshold or a different weighting factor to the marker for the different purpose(s).

In one embodiment, the panels comprise markers for the following purposes: diagnosis of a disease; diagnosis of disease and indication if the disease is in an acute phase and/or if an acute attack of the disease has occurred; diagnosis of disease and indication if the disease is in a non-acute phase and/or if a non-acute attack of the disease has occurred; indication whether a combination of acute and non-acute phases or attacks has occurred; diagnosis of a disease and prognosis of a subsequent adverse outcome; diagnosis of a disease and prognosis of a subsequent acute or non-acute phase or attack; disease progression (for example for cancer, such progression may include for example occurrence or recurrence of metastasis).

The above diagnoses may also optionally include differential diagnosis of the disease to distinguish it from other diseases, including those diseases that may feature one or more similar or identical symptoms.

In certain embodiments, one or more diagnostic or prognostic indicators are correlated to a condition or disease by merely the presence or absence of the indicator(s). In other
embodiments, threshold level(s) of a diagnostic or prognostic indicator(s) can be established, and the level of the indicator(s) in a patient sample can simply be compared to the threshold level(s). The sensitivity and specificity of a diagnostic and/or prognostic test depends on more than just the analytical "quality" of the test—they also depend on the definition of what constitutes an abnormal result. In practice, Receiver Operating Characteristic curves, or "ROC" curves, are typically calculated by plotting the value of a variable versus its relative frequency in "normal" and "disease" populations, and/or by comparison of results from a subject before, during and/or after treatment. For any particular marker, a distribution of marker levels for subjects with and without a disease will likely overlap. Under such conditions, a test does not absolutely distinguish normal from disease with 100% accuracy, and the area of overlap indicates where the test cannot distinguish normal from disease. A threshold is selected, above which (or below which, depending on how a marker changes with the disease) the test is considered to be abnormal and below which the test is considered to be normal. The area under the ROC curve is a measure of the probability that the perceived measurement will allow correct identification of a condition.

The horizontal axis of the ROC curve represents (1-specificity), which increases with the rate of false positives. The vertical axis of the curve represents sensitivity, which increases with the rate of true positives. Thus, for a particular cutoff selected, the value of (1-specificity) may be determined, and a corresponding sensitivity may be obtained. The area under the ROC curve is a measure of the probability that the measured marker level will allow correct identification of a disease or condition. Thus, the area under the ROC curve can be used to determine the effectiveness of the test.

ROC curves can be used even when test results don't necessarily give an accurate number. As long as one can rank results, one can create an ROC curve. For example, results of a test on "disease" samples might be ranked according to degree (say 1=low, 2=normal, and 3=high). This ranking can be correlated to results in the "normal" population, and a ROC curve created. These methods are well known in the art (see for example Hanley et al, Radiology 143: 29-36 (1982), incorporated by reference as if fully set forth herein).

One marker may lack diagnostic or prognostic value when considered alone, but when used as part of a panel, such marker may be of great value in determining a particular diagnosis/prognosis. In some embodiments, particular thresholds for a marker in a panel are not relied upon to determine if a profile of marker level obtained from a subject are indicative of a particular diagnosis/prognosis. Rather, the present invention may utilize an evaluation of the entire marker profile by plotting Receiver Operating Characteristic (ROC) curves for the
sensitivity of a particular panel of markers versus 1-(specificity) for the panel at various
cutoffs. In these methods, a profile of marker measurements from a subject is considered
together to provide a global probability (expressed either as a numeric score or as a percentage
risk) that an individual has had a disease, is at risk for developing such a disease, optionally
the type of disease which the individual has had or is at risk for, and so forth. In such
embodiments, an increase in a certain subset of markers may be sufficient to indicate a
particular diagnosis/prognosis in one patient, while an increase in a different subset of
markers may be sufficient to indicate the same or a different diagnosis/prognosis in another
patient. Weighting factors may also be applied to one or more markers in a panel, for
example, when a marker is of particularly high utility in identifying a particular
diagnosis/prognosis, it may be weighted so that at a given level it alone is sufficient to signal a
positive result. Likewise, a weighting factor may provide that no given level of a particular
marker is sufficient to signal a positive result, but only signals a result when another marker
also contributes to the analysis.

In some embodiments, markers and/or marker panels are selected to exhibit at least
70%, at least 80%, at least 85%, at least 90%, or at least 95% sensitivity, combined with at
least 70%, 80%, at least 85%, at least 90%, or at least 95% specificity. In some embodiments,
both the sensitivity and specificity are at least 75%, at least 80%, at least 85%, at least 90%, or
at least 95%. Sensitivity and/or specificity may optionally be determined as described above,
with regard to the construction of ROC graphs and so forth, for example.

According to some embodiments of the present invention, individual markers and/or
combinations (panels) of markers may optionally be used for diagnosis of time of onset of a
disease or condition. Such diagnosis may optionally be useful for a wide variety of conditions,
including those conditions with an abrupt onset.

The phrase "determining the prognosis" as used herein refers to methods by which the
skilled artisan can predict the course or outcome of a condition in a patient. The term
"prognosis" does not refer to the ability to predict the course or outcome of a condition with
100% accuracy, or even that a given course or outcome is more likely to occur than not.
Instead, the skilled artisan will understand that the term "prognosis" refers to an increased
probability that a certain course or outcome will occur; that is, that a course or outcome is
more likely to occur in a patient exhibiting a given condition, when compared to those
individuals not exhibiting the condition. For example, in individuals not exhibiting the
condition, the chance of a given outcome may be about 3%. In some embodiments, a
prognosis is about a 5% chance of a given outcome, about a 7% chance, about a 10% chance,
about a 12% chance, about a 15% chance, about a 20% chance, about a 25% chance, about a 30% chance, about a 40% chance, about a 50% chance, about a 60% chance, about a 75% chance, about a 90% chance, and about a 95% chance. The term "about" in this context refers to +/-1%.

The skilled artisan will understand that associating a prognostic indicator with a predisposition to an adverse outcome is a statistical analysis. For example, a marker level of greater than a pre-set value may signal that a patient is more likely to suffer from an adverse outcome than patients with a level less than or equal to the pre-set value, as determined by a level of statistical significance. Additionally, a change in marker concentration from baseline levels may be reflective of patient prognosis, and the degree of change in marker level may be related to the severity of adverse events. Statistical significance is often determined by comparing two or more populations, and determining a confidence interval and/or a p value. See, e.g., Dowdy and Wearden, Statistics for Research, John Wiley & Sons, New York, 1983. In one embodiment the confidence intervals of the invention are 90%, 95%, 97.5%, 98%, 99%, 99.5%, 99.9% and 99.99%, while preferred p values are 0.1, 0.05, 0.025, 0.01, 0.005, 0.001, and 0.0001. Exemplary statistical tests for associating a prognostic indicator with a predisposition to an adverse outcome are described hereinafter.

In other embodiments, a threshold degree of change in the level of a prognostic or diagnostic indicator can be established, and the degree of change in the level of the indicator in a patient sample can simply be compared to the threshold degree of change in the level. A preferred threshold change in the level for markers of the invention is about 5%, about 10%, about 15%, about 20%, about 25%, about 30%, about 50%, about 75%, about 100%, and about 150%. The term "about" in this context refers to +/-10%. In yet other embodiments, a "nomogram" can be established, by which a level of a prognostic or diagnostic indicator can be directly related to an associated disposition towards a given outcome. The skilled artisan is acquainted with the use of such nomograms to relate two numeric values with the understanding that the uncertainty in this measurement is the same as the uncertainty in the marker concentration because individual sample measurements are referenced, not population averages.

The HE4 variant markers of the present invention are advantageous over hitherto known markers relating to HE4 protein, as the level of this HE4 variants in healthy tissues/cell is significantly lower compared to their level in cancerous cells, such that the sensitivity and/or specificity of these markers is significant.
According to at least some embodiments of the present invention, HE4 variant protein or a fragment thereof, or HE4 variant nucleic acid sequence or a fragment thereof, may be featured as a biomarker for detecting cancer, particularly ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma, and/or an indicative condition, such that a biomarker may optionally comprise any of the above.

According to still other embodiments, the present invention optionally encompasses any amino acid sequence or fragment thereof encoded by a nucleic acid sequence corresponding to HE4 variant as described herein. Any oligopeptide or peptide relating to such an amino acid sequence or fragment thereof may optionally also (additionally or alternatively) be used as a biomarker, including but not limited to the unique amino acid sequences of these proteins that are depicted as tails, heads, insertions, edges or bridges.

The present invention also optionally encompasses antibodies capable of recognizing, and/or being elicited by, such oligopeptides or peptides.

In another embodiment, this invention provides antibodies specifically recognizing the HE4 variants and polypeptide fragments thereof of this invention. Typically such antibodies differentially recognize HE4 variants according to at least some embodiments of the present invention but do not recognize a corresponding known protein (such known proteins are discussed with regard to their variants in the Examples below).

In another embodiment, this invention provides a method for detecting HE4 variant in a biological sample, comprising: contacting a biological sample with an antibody specifically recognizing a variant according to at least some embodiments of the present invention under conditions whereby the antibody specifically interacts with the variant in the biological sample but do not recognize known corresponding proteins (wherein the known protein is discussed with regard to its variant(s) in the Examples below), and detecting said interaction; wherein the presence of an interaction correlates with the presence of a variant in the biological sample.

Non-limiting examples of methods or assays are described below.

According to at least some embodiments, the present invention also relates to kits based upon such diagnostic methods or assays.

Nucleic acid sequences and oligonucleotides

Various embodiments of the present invention encompass nucleic acid sequences described hereinabove; fragments thereof, sequences hybridizable therewith, sequences
homologous thereto, sequences encoding similar polypeptides with different codon usage, altered sequences characterized by mutations, such as deletion, insertion or substitution of one or more nucleotides, either naturally occurring or artificially induced, either randomly or in a targeted fashion.

It should be noted that the terms "oligonucleotide", "polynucleotide", and "nucleic acid molecule"; or "peptide", "polypeptide" and "protein", may optionally be used interchangeably.

According to at least some embodiments, the present invention encompasses nucleic acid sequences described herein; fragments thereof, sequences hybridizable therewith, sequences homologous thereto (e.g., at least 85%, at least 95%, at least 96% or at least 97% or at least 98% or at least 99% or more say 100% identical to the nucleic acid sequences set forth below), sequences encoding similar polypeptides with different codon usage, altered sequences characterized by mutations, such as deletion, insertion or substitution of one or more nucleotides, either naturally occurring or man induced, either randomly or in a targeted fashion. According to at least some embodiments, the present invention also encompasses homologous nucleic acid sequences (i.e., which form a part of a polynucleotide sequence according to at least some embodiments of the present invention) which include sequence regions unique to the polynucleotides according to at least some embodiments of the present invention.

In cases where the polynucleotide sequences according to at least some embodiments of the present invention encode previously unidentified polypeptides, the present invention also encompasses novel polypeptides or portions thereof, which are encoded by the isolated polynucleotide and respective nucleic acid fragments thereof described hereinabove.

The terms "nucleic acid fragment" or an "oligonucleotide" or a "polynucleotide" are used herein interchangeably to refer to a polymer of nucleic acids. A polynucleotide sequence according to at least some embodiments of the present invention refers to a single or double stranded nucleic acid sequences which is isolated and provided in the form of an RNA sequence, a complementary polynucleotide sequence (cDNA), a genomic polynucleotide sequence and/or a composite polynucleotide sequences (e.g., a combination of the above).

As used herein the phrase "complementary polynucleotide sequence" refers to a sequence, which results from reverse transcription of messenger RNA using a reverse transcriptase or any other RNA dependent DNA polymerase. Such a sequence can be subsequently amplified in vivo or in vitro using a DNA dependent DNA polymerase.
As used herein the phrase "genomic polynucleotide sequence" refers to a sequence derived (isolated) from a chromosome and thus it represents a contiguous portion of a chromosome.

As used herein the phrase "composite polynucleotide sequence" refers to a sequence, which is composed of genomic and cDNA sequences. A composite sequence can include some exonal sequences required to encode the polypeptide according to at least some embodiments of the present invention, as well as some intronic sequences interposing therebetween. The intronic sequences can be of any source, including of other genes, and typically will include conserved splicing signal sequences. Such intronic sequences may further include cis acting expression regulatory elements.

According to at least some embodiments, the present invention encompass oligonucleotide probes.

An example of an oligonucleotide probe which can be utilized by the present invention is a single stranded polynucleotide which includes a sequence complementary to the unique sequence region of any variant according to at least some embodiments of the present invention, including but not limited to a nucleotide sequence coding for an amino sequence of a bridge, tail, head and/or insertion, and/or the equivalent portions of any nucleotide sequence given herein (including but not limited to a nucleotide sequence of a node, segment or amplicon described herein).

Alternatively, an oligonucleotide probe according to at least some embodiments of the present invention can be designed to hybridize with a nucleic acid sequence encompassed by any of the above nucleic acid sequences, particularly the portions specified above, including but not limited to a nucleotide sequence coding for an amino sequence of a bridge, tail, head and/or insertion, and/or the equivalent portions of any nucleotide sequence given herein (including but not limited to a nucleotide sequence of a node, segment or amplicon described herein).

Oligonucleotides designed according to the teachings of the present invention can be generated according to any oligonucleotide synthesis method known in the art such as enzymatic synthesis or solid phase synthesis. Equipment and reagents for executing solid-phase synthesis are commercially available from, for example, Applied Biosystems. Any other means for such synthesis may also be employed; the actual synthesis of the oligonucleotides is well within the capabilities of one skilled in the art and can be accomplished via established methodologies as detailed in, for example, "Molecular Cloning: A laboratory Manual" Sambrook et al, (1989); "Current Protocols in Molecular Biology"

Oligonucleotides used according to at least some embodiments of the present invention are those having a length selected from a range of about 10 to about 300 bases, or about 10 to about 200 bases, or about 15 to about 150 bases, or about 20 to about 100 bases, or about 20 to about 50 bases. The oligonucleotide according to at least some embodiments of the present invention features at least 17, at least 18, at least 19, at least 20, at least 22, at least 25, at least 30 or at least 40, bases specifically hybridizable with the biomarkers according to at least some embodiments of the present invention.

The oligonucleotides according to at least some embodiments of the present invention may comprise heterocyclic nucleosides consisting of purines and the pyrimidines bases, bonded in a 3′ to 5′ phosphodiester linkage.

The oligonucleotides according to at least some embodiments of the present invention are those modified at one or more of the backbone, internucleoside linkages or bases, as is broadly described hereinunder.

Specific examples of oligonucleotides useful according to this aspect of the present invention include oligonucleotides containing modified backbones or non-natural internucleoside linkages. Oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone, as disclosed in U.S. Pat. Nos: 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050.

Preferred modified oligonucleotide backbones include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkyl phosphotriesters, methyl and other alkyl phosphonates including 3′-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3′-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylyphosphonates, thionoalkylyphosphotriesters, and boranophosphates having normal 3′-5′ linkages, 2′-5′ linked analogs of these, and those having inverted polarity wherein the
adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts and free acid forms can also be used.

Alternatively, modified oligonucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts, as disclosed in U.S. Patent Nos. 5,034,506; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,620,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437; and 5,677,439.

Other oligonucleotides which can be used according to the present invention, are those modified in both sugar and the internucleoside linkage, i.e., the backbone, of the nucleotide units are replaced with novel groups. The base units are maintained for complementation with the appropriate polynucleotide target. An example for such an oligonucleotide mimetic includes peptide nucleic acid (PNA). United States patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent Nos. 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Other backbone modifications, which can be used in the present invention, are disclosed in U.S. Pat. No: 6,303,374.

Oligonucleotides according to at least some embodiments of the present invention may also include base modifications or substitutions. As used herein, "unmodified" or "natural" bases include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified bases include but are not limited to other synthetic and natural bases such as 5-methylecytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-
methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-
deazaguanine and 3-deazaadenine. Further bases particularly useful for increasing the binding
affinity of the oligomeric compounds of the invention include 5-substituted pyrimidines, 6-
azaapyrimidines and N-2, N-6 and 0-6 substituted purines, including 2-aminopropyladenine, 5-
propynyluracil and 5-propynylcytosine. 5-methylcytosine substitutions have been shown to
increase nucleic acid duplex stability by 0.6-1.2 °C and are presently preferred base
substitutions, even more particularly when combined with 2'-O-methoxyethyl sugar
modifications.

Another modification of the oligonucleotides of the invention involves chemically
linking to the oligonucleotide one or more moieties or conjugates, which enhance the activity,
cellular distribution or cellular uptake of the oligonucleotide. Such moieties include but are
not limited to lipid moieties such as a cholesterol moiety, cholic acid, a thioether, e.g., hexyl-
S-tritylthiol, a thiocholesterol, an aliphatic chain, e.g., dodecanediol or undecyl residues, a
phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-
glycero-3-H-phosphonate, a polyamine or a polyethylene glycol chain, or adamantane acetic
acid, a palmitoyl moiety, or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety,
as disclosed in U.S. Patent No: 6,303,374.

It is not necessary for all positions in a given oligonucleotide molecule to be uniformly
modified, and in fact more than one of the aforementioned modifications may be incorporated
in a single compound or even at a single nucleoside within an oligonucleotide.

It will be appreciated that oligonucleotides according to at least some embodiments of
the present invention may include further modifications for more efficient use as diagnostic
agents and/or to increase bioavailability, therapeutic efficacy and reduce cytotoxicity.

To enable cellular expression of the polynucleotides according to at least some
embodiments of the present invention, a nucleic acid construct may be used, which includes
at least a coding region of one of the above nucleic acid sequences, and further includes at
least one cis acting regulatory element. As used herein, the phrase "cis acting regulatory
element" refers to a polynucleotide sequence, preferably a promoter, which binds a trans
acting regulator and regulates the transcription of a coding sequence located downstream
thereto.

Any suitable promoter sequence can be used by the nucleic acid construct according
to at least some embodiments of the present invention.

Preferably, the promoter utilized by the nucleic acid construct according to at least
some embodiments of the present invention is active in the specific cell population
transformed. Examples of cell type-specific and/or tissue-specific promoters include promoters such as albumin that is liver specific, lymphoid specific promoters (Calame et al., Adv. Immunol. 1988; 43, 235-275); in particular promoters of T-cell receptors (Winoto et al., EMBO J. 1989; 8, 729-733) and immunoglobulins; (Banerji et al. (1983) Cell 33729-740), neuron-specific promoters such as the neurofilament promoter (Byrne et alProc. Natl. Acad. Sci. USA 1989 86, 5473-5477), pancreas-specific promoters (Edlunch et al. Science 1985 230, 912-916) or mammary gland-specific promoters such as the milk whey promoter (U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). The nucleic acid construct according to at least some embodiments of the present invention can further include an enhancer, which can be adjacent or distant to the promoter sequence and can function in up regulating the transcription therefrom.

The nucleic acid construct according to at least some embodiments of the present invention further includes an appropriate selectable marker and/or an origin of replication. Preferably, the nucleic acid construct utilized is a shuttle vector, which can propagate both in E. coli (wherein the construct comprises an appropriate selectable marker and origin of replication) and be compatible for propagation in cells, or integration in a gene and a tissue of choice. The construct according to at least some embodiments of the present invention can be, for example, a plasmid, a bacmid, a phagemid, a cosmids, a phage, a virus or an artificial chromosome.

Examples of suitable constructs include, but are not limited to, pcDNA3, pcDNA3.1 (+/-), pGL3, PzeoSV2 (+/-), pDisplay, pEF/myc/cyto, pCMV/myc/cyto each of which is commercially available from Invitrogen Co. (www.invitrogen.com). Examples of retroviral vector and packaging systems are those sold by Clontech, San Diego, Calif, including Retro-X vectors pLNCX and pLXSN, which permit cloning into multiple cloning sites and the transgene is transcribed from CMV promoter. Vectors derived from Mo-MuLV are also included such as pBabe, where the transgene will be transcribed from the 5'LTR promoter.

In vivo nucleic acid transfer techniques include transfection with viral or non-viral constructs, such as adenovirus, lentivirus, Herpes simplex I virus, or adeno-associated virus (AAV) and lipid-based systems. Useful lipids for lipid-mediated transfer of the gene are, for example, DOTMA, DOPE, and DC-Choi (Tonkinson et al., Cancer Investigation, 14(1): 54-65 (1996)). The most preferred constructs for use in gene therapy are viruses, most preferably adenoviruses, AAV, lentiviruses, or retroviruses. A viral construct such as a retroviral construct includes at least one transcriptional promoter/enhancer or locus-defining element(s), or other elements that control gene expression by other means such as alternate splicing,
nuclear RNA export, or post-translational modification of messenger. Such vector constructs also include a packaging signal, long terminal repeats (LTRs) or portions thereof, and positive and negative strand primer binding sites appropriate to the virus used, unless it is already present in the viral construct. In addition, such a construct typically includes a signal sequence for secretion of the peptide from a host cell in which it is placed. Preferably the signal sequence for this purpose is a mammalian signal sequence or the signal sequence of the polypeptide variants according to at least some embodiments of the present invention. Optionally, the construct may also include a signal that directs polyadenylation, as well as one or more restriction sites and a translation termination sequence. By way of example, such constructs will typically include a 5' LTR, a tRNA binding site, a packaging signal, an origin of second-strand DNA synthesis, and a 3' LTR or a portion thereof. Other vectors can be used that are non-viral, such as cationic lipids, polylysine, and dendrimers.

It will be appreciated that the nucleic acid construct can be administered to the individual employing any suitable mode of administration, described hereinbelow (i.e., in-vivo gene therapy). Alternatively, the nucleic acid construct is introduced into a suitable cell via an appropriate gene delivery vehicle/method (transfection, transduction, homologous recombination, etc.) and an expression system as needed and then the modified cells are expanded in culture and returned to the individual (i.e., ex-vivo gene therapy).

Hybridization assays

Detection of a nucleic acid of interest in a biological sample may optionally be effected by hybridization-based assays using an oligonucleotide probe (non-limiting examples of probes according to the present invention were previously described).

Traditional hybridization assays include PCR, RT-PCR, Real-time PCR, RNase protection, in-situ hybridization, primer extension, Southern blots (DNA detection), dot or slot blots (DNA, RNA), and Northern blots (RNA detection) (NAT type assays are described in greater detail below). More recently, PNAs have been described (Nielsen et al. 1999, Current Opin. Biotechnol. 10:71-75). Other detection methods include kits containing probes on a dipstick setup and the like.

Hybridization based assays which allow the detection of a variant of interest (i.e., DNA or RNA) in a biological sample rely on the use of oligonucleotides which can be 10, 15, 20, or 30 to 100 nucleotides long optionally from 10 to 50, or from 40 to 50 nucleotides long.
Thus, the isolated polynucleotides (oligonucleotides) according to at least some embodiments of the present invention are optionally hybridizable with any of the herein described nucleic acid sequences under moderate to stringent hybridization conditions.

Moderate to stringent hybridization conditions are characterized by a hybridization solution such as containing 10% dextrane sulfate, 1 M NaCl, 1% SDS and 5 x 10^6 cpm ^32P labeled probe, at 65°C with a final wash solution of 0.2 x SSC and 0.1% SDS and final wash at 65°C and whereas moderate hybridization is effected using a hybridization solution containing 10% dextrane sulfate, 1 M NaCl, 1% SDS and 5 x 10^6 cpm ^32P labeled probe, at 65°C, with a final wash solution of 1 x SSC and 0.1% SDS and final wash at 50°C.

More generally, hybridization of short nucleic acids (below 200 bp in length, e.g. 17-40 bp in length) can be effected using the following exemplary hybridization protocols which can be modified according to the desired stringency; (i) hybridization solution of 6 x SSC and 1% SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5% SDS, 100 µg/ml denatured salmon sperm DNA and 0.1% nonfat dried milk, hybridization temperature of 1 - 1.5°C below the T_m, final wash solution of 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5% SDS at 1 - 1.5°C below the T_m; (H) hybridization solution of 6 x SSC and 0.1% SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5% SDS, 100 µg/ml denatured salmon sperm DNA and 0.1% nonfat dried milk, hybridization temperature of 2 - 2.5°C below the T_m, final wash solution of 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5% SDS at 1 - 1.5°C below the T_m, final wash solution of 6 x SSC, and final wash at 22°C; (Hi) hybridization solution of 6 x SSC and 1% SDS or 3 M TMACI, 0.01 M sodium phosphate (pH 6.8), 1 mM EDTA (pH 7.6), 0.5% SDS, 100 µg/ml denatured salmon sperm DNA and 0.1% nonfat dried milk, hybridization temperature.

The detection of hybrid duplexes can be carried out by a number of methods. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Such labels refer to radioactive, fluorescent, biological or enzymatic tags or labels of standard use in the art. A label can be conjugated to either the oligonucleotide probes or the nucleic acids derived from the biological sample.

Probes can be labeled according to numerous well known methods. Non-limiting examples of radioactive labels include 3H, 14C, 32P, and 35S. Non-limiting examples of detectable markers include ligands, fluorophores, chemiluminescent agents, enzymes, and antibodies. Other detectable markers for use with probes, which can enable an increase in
sensitivity of the method of the invention, include biotin and radio-nucleotides. It will become
evident to the person of ordinary skill that the choice of a particular label dictates the manner
in which it is bound to the probe.

For example, oligonucleotides according to at least some embodiments of the present
invention can be labeled subsequent to synthesis, by incorporating biotinylated dNTPs or
rNTP, or some similar means (e.g., photo-cross-linking a psoralen derivative of biotin to
RNAs), followed by addition of labeled streptavidin (e.g., phycoerythrin-conjugated
streptavidin) or the equivalent. Alternatively, when fluorescently-labeled oligonucleotide
probes are used, fluorescein, lissamine, phycoerythrin, rhodamine (Perkin Elmer Cetus), Cy2,
Cy3, Cy3.5, Cy5, Cy5.5, Cy7, FluorX (Amersham) and others [e.g., Kricka et al. (1992),
Academic Press San Diego, Calif] can be attached to the oligonucleotides.

Those skilled in the art will appreciate that wash steps may be employed to wash away
excess target DNA or probe as well as unbound conjugate. Further, standard heterogeneous
assay formats are suitable for detecting the hybrids using the labels present on the
oligonucleotide primers and probes.

It will be appreciated that a variety of controls may be usefully employed to improve
accuracy of hybridization assays. For instance, samples may be hybridized to an irrelevant
probe and treated with RNase A prior to hybridization, to assess false hybridization.

Although the present invention is not specifically dependent on the use of a label for
the detection of a particular nucleic acid sequence, such a label might be beneficial, by
increasing the sensitivity of the detection. Furthermore, it enables automation. Probes can be
labeled according to numerous well known methods.

As commonly known, radioactive nucleotides can be incorporated into probes of the
invention by several methods. Non-limiting examples of radioactive labels include $^3$H, $^{14}$C,
$^{32}$P, and $^{35}$S.

Those skilled in the art will appreciate that wash steps may be employed to wash away
excess target DNA or probe as well as unbound conjugate. Further, standard heterogeneous
assay formats are suitable for detecting the hybrids using the labels present on the
oligonucleotide primers and probes.

It will be appreciated that a variety of controls may be usefully employed to improve
accuracy of hybridization assays.

Probes of the invention can be utilized with naturally occurring sugar-phosphate
backbones as well as modified backbones including phosphorothioates, dithionates, alkyl
phosphonates and a-nucleotides and the like. Probes of the invention can be constructed of either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA).

**NAT assays**

detection of a nucleic acid of interest in a biological sample may also optionally be effected by NAT-based assays, which involve nucleic acid amplification technology, such as PCR for example (or variations thereof such as real-time PCR for example).

As used herein, a "primer" defines an oligonucleotide which is capable of annealing to (hybridizing with) a target sequence, thereby creating a double stranded region which can serve as an initiation point for DNA synthesis under suitable conditions.

Amplification of a selected, or target, nucleic acid sequence may be carried out by a number of suitable methods. See generally Kwoh et al., 1990, Am. Biotechnol. Lab. 8:14 Numerous amplification techniques have been described and can be readily adapted to suit particular needs of a person of ordinary skill. Non-limiting examples of amplification techniques include polymerase chain reaction (PCR), ligase chain reaction (LCR), strand displacement amplification (SDA), transcription-based amplification, the q3 replicate system and NASBA (Kwoh et al., 1989, Proc. Natl. Acad. Sci. USA 86, 1173-1177; Lizardi et al., 1988, BioTechnology 6:1 197-1202; Malek et al., 1994, Methods Mol. Biol., 28:253-260; and Sambrook et al., 1989, supra).

The terminology "amplification pair" (or "primer pair") refers herein to a pair of oligonucleotides (oligos) according to at least some embodiments of the present invention, which are selected to be used together in amplifying a selected nucleic acid sequence by one of a number of types of amplification processes, preferably a polymerase chain reaction. Other types of amplification processes include ligase chain reaction, strand displacement amplification, or nucleic acid sequence-based amplification, as explained in greater detail below. As commonly known in the art, the oligos are designed to bind to a complementary sequence under selected conditions.

In one particular embodiment, amplification of a nucleic acid sample from a patient is amplified under conditions which favor the amplification of the most abundant differentially expressed nucleic acid. In one embodiment, RT-PCR is carried out on an mRNA sample from a patient under conditions which favor the amplification of the most abundant mRNA. In another embodiment, the amplification of the differentially expressed nucleic acids is carried out simultaneously. It will be realized by a person skilled in the art that such methods could be
adapted for the detection of differentially expressed proteins instead of differentially expressed nucleic acid sequences.

The nucleic acid (i.e. DNA or RNA) for practicing the present invention may be obtained according to well known methods.

Oligonucleotide primers according to at least some embodiments of the present invention may be of any suitable length, depending on the particular assay format and the particular needs and targeted genomes employed. Optionally, the oligonucleotide primers are at least 12 nucleotides in length, preferably between 15 and 24 molecules, and they may be adapted to be especially suited to a chosen nucleic acid amplification system. As commonly known in the art, the oligonucleotide primers can be designed by taking into consideration the melting point of hybridization thereof with its targeted sequence (Sambrook et al., 1989, Molecular Cloning - A Laboratory Manual, 2nd Edition, CSH Laboratories; Ausubel et al., 1989, in Current Protocols in Molecular Biology, John Wiley & Sons Inc., N.Y.).

It will be appreciated that antisense oligonucleotides may be employed to quantify expression of an isoform of interest. Such detection is effected at the pre-mRNA level. Essentially the ability to quantitate transcription from a splice site of interest can be effected based on splice site accessibility. Oligonucleotides may compete with splicing factors for the splice site sequences. Thus, low activity of the antisense oligonucleotide is indicative of splicing activity.

The polymerase chain reaction and other nucleic acid amplification reactions are well known in the art (various non-limiting examples of these reactions are described in greater detail below). The pair of oligonucleotides according to this aspect of the present invention are preferably selected to have compatible melting temperatures (Tm), e.g., melting temperatures which differ by less than that 7 °C, preferably less than 5 °C, more preferably less than 4 °C, most preferably less than 3 °C, ideally between 3 °C and 0 °C.

Polymerase Chain Reaction (PCR): The polymerase chain reaction (PCR), as described in U.S. Pat. Nos. 4,683,195 and 4,683,202 to Mullis and Mullis et al., is a method of increasing the concentration of a segment of target sequence in a mixture of genomic DNA without cloning or purification. This technology provides one approach to the problems of low target sequence concentration. PCR can be used to directly increase the concentration of the target to an easily detectable level. This process for amplifying the target sequence involves the introduction of a molar excess of two oligonucleotide primers which are complementary to their respective strands of the double-stranded target sequence to the DNA mixture containing the desired target sequence. The mixture is denatured and then allowed to hybridize.
Following hybridization, the primers are extended with polymerase so as to form complementary strands. The steps of denaturation, hybridization (annealing), and polymerase extension (elongation) can be repeated as often as needed, in order to obtain relatively high concentrations of a segment of the desired target sequence.

The length of the segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and, therefore, this length is a controllable parameter. Because the desired segments of the target sequence become the dominant sequences (in terms of concentration) in the mixture, they are said to be "PCR-amplified."

Ligase Chain Reaction (LCR or LAR): The ligase chain reaction [LCR; sometimes referred to as "Ligase Amplification Reaction" (LAR)] has developed into a well-recognized alternative method of amplifying nucleic acids. In LCR, four oligonucleotides, two adjacent oligonucleotides which uniquely hybridize to one strand of target DNA, and a complementary set of adjacent oligonucleotides, which hybridize to the opposite strand are mixed and DNA ligase is added to the mixture. Provided that there is complete complementarity at the junction, ligase will covalently link each set of hybridized molecules. Importantly, in LCR, two probes are ligated together only when they base-pair with sequences in the target sample, without gaps or mismatches. Repeated cycles of denaturation and ligation amplify a short segment of DNA. LCR has also been used in combination with PCR to achieve enhanced detection of single-base changes: see for example Segev, PCT Publication No. WO90001069 A1 (1990). However, because the four oligonucleotides used in this assay can pair to form two short ligatable fragments, there is the potential for the generation of target-independent background signal. The use of LCR for mutant screening is limited to the examination of specific nucleic acid positions.

Self-Sustained Synthetic Reaction (3SR/NASBA): The self-sustained sequence replication reaction (3SR) is a transcription-based in vitro amplification system that can exponentially amplify RNA sequences at a uniform temperature. The amplified RNA can then be utilized for mutation detection. In this method, an oligonucleotide primer is used to add a phage RNA polymerase promoter to the 5' end of the sequence of interest. In a cocktail of enzymes and substrates that includes a second primer, reverse transcriptase, RNase H, RNA polymerase and ribo-and deoxyribonucleoside triphosphates, the target sequence undergoes repeated rounds of transcription, cDNA synthesis and second-strand synthesis to amplify the area of interest. The use of 3SR to detect mutations is kinetically limited to screening small segments of DNA (e.g., 200-300 base pairs).
**Q-Beta (Qβ) Replicase:** In this method, a probe which recognizes the sequence of interest is attached to the replicatable RNA template for Qβ replicase. A previously identified major problem with false positives resulting from the replication of unhybridized probes has been addressed through use of a sequence-specific ligation step. However, available thermostable DNA ligases are not effective on this RNA substrate, so the ligation must be performed by T4 DNA ligase at low temperatures (37 degrees C). This prevents the use of high temperature as a means of achieving specificity as in the LCR, the ligation event can be used to detect a mutation at the junction site, but not elsewhere.

A successful diagnostic method must be very specific. A straight-forward method of controlling the specificity of nucleic acid hybridization is by controlling the temperature of the reaction. While the 3SR/NASBA, and Qβ systems are all able to generate a large quantity of signal, one or more of the enzymes involved in each cannot be used at high temperature (i.e., >55 degrees C). Therefore the reaction temperatures cannot be raised to prevent non-specific hybridization of the probes. If probes are shortened in order to make them melt more easily at low temperatures, the likelihood of having more than one perfect match in a complex genome increases. For these reasons, PCR and LCR currently dominate the research field in detection technologies.

The basis of the amplification procedure in the PCR and LCR is the fact that the products of one cycle become usable templates in all subsequent cycles, consequently doubling the population with each cycle. The final yield of any such doubling system can be expressed as: \((1+X)^n = y\), where "X" is the mean efficiency (percent copied in each cycle), "n" is the number of cycles, and "y" is the overall efficiency, or yield of the reaction. If every copy of a target DNA is utilized as a template in every cycle of a polymerase chain reaction, then the mean efficiency is 100%. If 20 cycles of PCR are performed, then the yield will be \(2^{20}\), or 1,048,576 copies of the starting material. If the reaction conditions reduce the mean efficiency to 85%, then the yield in those 20 cycles will be only \(1.85^{20}\), or 220,513 copies of the starting material. In other words, a PCR running at 85% efficiency will yield only 21% as much final product, compared to a reaction running at 100% efficiency. A reaction that is reduced to 50% mean efficiency will yield less than 1% of the possible product.

In practice, routine polymerase chain reactions rarely achieve the theoretical maximum yield, and PCRs are usually run for more than 20 cycles to compensate for the lower yield. At 50% mean efficiency, it would take 34 cycles to achieve the million-fold amplification theoretically possible in 20, and at lower efficiencies, the number of cycles required becomes
prohibitive. In addition, any background products that amplify with a better mean efficiency than the intended target will become the dominant products.

Also, many variables can influence the mean efficiency of PCR, including target DNA length and secondary structure, primer length and design, primer and dNTP concentrations, and buffer composition, to name but a few. Contamination of the reaction with exogenous DNA (e.g., DNA spilled onto lab surfaces) or cross-contamination is also a major consideration. Reaction conditions must be carefully optimized for each different primer pair and target sequence, and the process can take days, even for an experienced investigator. The laboriousness of this process, including numerous technical considerations and other factors, presents a significant drawback to using PCR in the clinical setting. Indeed, PCR has yet to penetrate the clinical market in a significant way. The same concerns arise with LCR, as LCR must also be optimized to use different oligonucleotide sequences for each target sequence. In addition, both methods require expensive equipment, capable of precise temperature cycling.

Additional NAT tests are Fluorescence In Situ Hybridization (FISH) and Comparative Genomic Hybridization (CGH). Fluorescence In Situ Hybridization (FISH) - The test uses fluorescent single-stranded DNA probes which are complementary to the DNA sequences that are under examination (genes or chromosomes). These probes hybridize with the complementary DNA and allow the identification of the chromosomal location of genomic sequences of DNA.

Comparative Genomic Hybridization (CGH) - allows a comprehensive analysis of multiple DNA gains and losses in entire genomes. Genomic DNA from the tissue to be investigated and a reference DNA are differentially labeled and simultaneously hybridized in situ to normal metaphase chromosomes. Variations in signal intensities are indicative of differences in the genomic content of the tissue under investigation.

Many applications of nucleic acid detection technologies, such as in studies of allelic variation, involve not only detection of a specific sequence in a complex background, but also the discrimination between sequences with few, or single, nucleotide differences. One method of the detection of allele-specific variants by PCR is based upon the fact that it is difficult for Taq polymerase to synthesize a DNA strand when there is a mismatch between the template strand and the 3’ end of the primer. An allele-specific variant may be detected by the use of a primer that is perfectly matched with only one of the possible alleles; the mismatch to the other allele acts to prevent the extension of the primer, thereby preventing the amplification of that sequence. This method has a substantial limitation in that the base composition of the
mismatch influences the ability to prevent extension across the mismatch, and certain mismatches do not prevent extension or have only a minimal effect.

A similar 3'-mismatch strategy is used with greater effect to prevent ligation in the LCR. Any mismatch effectively blocks the action of the thermostable ligase, but LCR still has the drawback of target-independent background ligation products initiating the amplification. Moreover, the combination of PCR with subsequent LCR to identify the nucleotides at individual positions is also a clearly cumbersome proposition for the clinical laboratory.

The direct detection method according to at least some embodiments of the present invention may be, for example a cycling probe reaction (CPR) or a branched DNA analysis.

When a sufficient amount of a nucleic acid to be detected is available, there are advantages to detecting that sequence directly, instead of making more copies of that target, (e.g., as in PCR and LCR). Most notably, a method that does not amplify the signal exponentially is more amenable to quantitative analysis. Even if the signal is enhanced by attaching multiple dyes to a single oligonucleotide, the correlation between the final signal intensity and amount of target is direct. Such a system has an additional advantage that the products of the reaction will not themselves promote further reaction, so contamination of lab surfaces by the products is not as much of a concern. Recently devised techniques have sought to eliminate the use of radioactivity and/or improve the sensitivity in automatable formats. Two examples are the "Cycling Probe Reaction" (CPR), and "Branched DNA" (bDNA).

Cycling probe reaction (CPR): The cycling probe reaction (CPR) uses a long chimeric oligonucleotide in which a central portion is made of RNA while the two termini are made of DNA. Hybridization of the probe to a target DNA and exposure to a thermostable RNase H causes the RNA portion to be digested. This destabilizes the remaining DNA portions of the duplex, releasing the remainder of the probe from the target DNA and allowing another probe molecule to repeat the process. The signal, in the form of cleaved probe molecules, accumulates at a linear rate. While the repeating process increases the signal, the RNA portion of the oligonucleotide is vulnerable to RNases that may carried through sample preparation.

Branched DNA: Branched DNA (bDNA), involves oligonucleotides with branched structures that allow each individual oligonucleotide to carry 35 to 40 labels (e.g., alkaline phosphatase enzymes). While this enhances the signal from a hybridization event, signal from non-specific binding is similarly increased.

The detection of at least one sequence change according to at least some embodiments of the present invention may be accomplished by, for example restriction fragment length
polymorphism (RFLP analysis), allele specific oligonucleotide (ASO) analysis, Denaturing/Temperature Gradient Gel Electrophoresis (DGGE/TGGE), Single-Strand Conformation Polymorphism (SSCP) analysis or Dideoxy fingerprinting (ddF).

The demand for tests which allow the detection of specific nucleic acid sequences and sequence changes is growing rapidly in clinical diagnostics. As nucleic acid sequence data for genes from humans and pathogenic organisms accumulates, the demand for fast, cost-effective, and easy-to-use tests for as yet mutations within specific sequences is rapidly increasing.

A handful of methods have been devised to scan nucleic acid segments for mutations. One option is to determine the entire gene sequence of each test sample (e.g., a bacterial isolate). For sequences under approximately 600 nucleotides, this may be accomplished using amplified material (e.g., PCR reaction products). This avoids the time and expense associated with cloning the segment of interest. However, specialized equipment and highly trained personnel are required, and the method is too labor-intense and expensive to be practical and effective in the clinical setting.

In view of the difficulties associated with sequencing, a given segment of nucleic acid may be characterized on several other levels. At the lowest resolution, the size of the molecule can be determined by electrophoresis by comparison to a known standard run on the same gel. A more detailed picture of the molecule may be achieved by cleavage with combinations of restriction enzymes prior to electrophoresis, to allow construction of an ordered map. The presence of specific sequences within the fragment can be detected by hybridization of a labeled probe, or the precise nucleotide sequence can be determined by partial chemical degradation or by primer extension in the presence of chain-terminating nucleotide analogs.

**Restriction fragment length polymorphism (RFLP):** For detection of single-base differences between like sequences, the requirements of the analysis are often at the highest level of resolution. For cases in which the position of the nucleotide in question is known in advance, several methods have been developed for examining single base changes without direct sequencing. For example, if a mutation of interest happens to fall within a restriction recognition sequence, a change in the pattern of digestion can be used as a diagnostic tool (e.g., restriction fragment length polymorphism [RFLP] analysis).

Single point mutations have been also detected by the creation or destruction of RFLPs. Mutations are detected and localized by the presence and size of the RNA fragments generated by cleavage at the mismatches. Single nucleotide mismatches in DNA heteroduplexes are also recognized and cleaved by some chemicals, providing an alternative strategy to detect single
base substitutions, generically named the "Mismatch Chemical Cleavage" (MCC). However, this method requires the use of osmium tetroxide and piperidine, two highly noxious chemicals which are not suited for use in a clinical laboratory.

RFLP analysis suffers from low sensitivity and requires a large amount of sample. When RFLP analysis is used for the detection of point mutations, it is, by its nature, limited to the detection of only those single base changes which fall within a restriction sequence of a known restriction endonuclease. Moreover, the majority of the available enzymes has 4 to 6 base-pair recognition sequences, and cleaves too frequently for many large-scale DNA manipulations. Thus, it is applicable only in a small fraction of cases, as most mutations do not fall within such sites.

A handful of rare-cutting restriction enzymes with 8 base-pair specificities have been isolated and these are widely used in genetic mapping, but these enzymes are few in number, are limited to the recognition of G+C-rich sequences, and cleave at sites that tend to be highly clustered. Recently, endonucleases encoded by group I introns have been discovered that might have greater than 12 base-pair specificity, but again, these are few in number.

**Allele specific oligonucleotide (ASO):** If the change is not in a recognition sequence, then allele-specific oligonucleotides (ASOs), can be designed to hybridize in proximity to the mutated nucleotide, such that a primer extension or ligation event can be used as the indicator of a match or a mis-match. Hybridization with radioactively labeled allelic specific oligonucleotides (ASO) also has been applied to the detection of specific point mutations. The method is based on the differences in the melting temperature of short DNA fragments differing by a single nucleotide. Stringent hybridization and washing conditions can differentiate between mutant and wild-type alleles. The ASO approach applied to PCR products also has been extensively utilized by various researchers to detect and characterize point mutations in ras genes and gsp/gip oncogenes. Because of the presence of various nucleotide changes in multiple positions, the ASO method requires the use of many oligonucleotides to cover all possible oncogenic mutations.

With either of the techniques described above (i.e., RFLP and ASO), the precise location of the suspected mutation must be known in advance of the test. That is to say, they are inapplicable when one needs to detect the presence of a mutation within a gene or sequence of interest.

**Denaturing/Temperature Gradient Gel Electrophoresis (DGGE/TGGE):** Two other methods rely on detecting changes in electrophoretic mobility in response to minor sequence changes. One of these methods, termed "Denaturing Gradient Gel Electrophoresis"
(DGGE) is based on the observation that slightly different sequences will display different patterns of local melting when electrophoretically resolved on a gradient gel. In this manner, variants can be distinguished, as differences in melting properties of homoduplexes versus heteroduplexes differing in a single nucleotide can detect the presence of mutations in the target sequences because of the corresponding changes in their electrophoretic mobilities. The fragments to be analyzed, usually PCR products, are "clamped" at one end by a long stretch of G-C base pairs (30-80) to allow complete denaturation of the sequence of interest without complete dissociation of the strands. The attachment of a GC "clamp" to the DNA fragments increases the fraction of mutations that can be recognized by DGGE. Attaching a GC clamp to one primer is critical to ensure that the amplified sequence has a low dissociation temperature. Modifications of the technique have been developed, using temperature gradients, and the method can be also applied to RNA:RNA duplexes.

Limitations on the utility of DGGE include the requirement that the denaturing conditions must be optimized for each type of DNA to be tested. Furthermore, the method requires specialized equipment to prepare the gels and maintain the needed high temperatures during electrophoresis. The expense associated with the synthesis of the clamping tail on one oligonucleotide for each sequence to be tested is also a major consideration. In addition, long running times are required for DGGE. The long running time of DGGE was shortened in a modification of DGGE called constant denaturant gel electrophoresis (CDGE). CDGE requires that gels be performed under different denaturant conditions in order to reach high efficiency for the detection of mutations.

A technique analogous to DGGE, termed temperature gradient gel electrophoresis (TGGE), uses a thermal gradient rather than a chemical denaturant gradient. TGGE requires the use of specialized equipment which can generate a temperature gradient perpendicularly oriented relative to the electrical field. TGGE can detect mutations in relatively small fragments of DNA therefore scanning of large gene segments requires the use of multiple PCR products prior to running the gel.

**Single-Strand Conformation Polymorphism (SSCP):** Another common method, called "Single-Strand Conformation Polymorphism" (SSCP) was developed by Hayashi, Sekya and colleagues and is based on the observation that single strands of nucleic acid can take on characteristic conformations in non-denaturing conditions, and these conformations influence electrophoretic mobility. The complementary strands assume sufficiently different structures that one strand may be resolved from the other. Changes in sequences within the fragment will
also change the conformation, consequently altering the mobility and allowing this to be used as an assay for sequence variations.

The SSCP process involves denaturing a DNA segment (e.g., a PCR product) that is labeled on both strands, followed by slow electrophoretic separation on a non-denaturing polyacrylamide gel, so that intra-molecular interactions can form and not be disturbed during the run. This technique is extremely sensitive to variations in gel composition and temperature. A serious limitation of this method is the relative difficulty encountered in comparing data generated in different laboratories, under apparently similar conditions.

Dideoxy fingerprinting (ddF): The dideoxy fingerprinting (ddF) is another technique developed to scan genes for the presence of mutations. The ddF technique combines components of Sanger dideoxy sequencing with SSCP. A dideoxy sequencing reaction is performed using one dideoxy terminator and then the reaction products are electrophoresed on non-denaturing polyacrylamide gels to detect alterations in mobility of the termination segments as in SSCP analysis. While ddF is an improvement over SSCP in terms of increased sensitivity, ddF requires the use of expensive dideoxynucleotides and this technique is still limited to the analysis of fragments of the size suitable for SSCP (i.e., fragments of 200-300 bases for optimal detection of mutations).

In addition to the above limitations, all of these methods are limited as to the size of the nucleic acid fragment that can be analyzed. For the direct sequencing approach, sequences of greater than 600 base pairs require cloning, with the consequent delays and expense of either deletion sub-cloning or primer walking, in order to cover the entire fragment. SSCP and DGGE have even more severe size limitations. Because of reduced sensitivity to sequence changes, these methods are not considered suitable for larger fragments. Although SSCP is reportedly able to detect 90% of single-base substitutions within a 200 base-pair fragment, the detection drops to less than 50% for 400 base pair fragments. Similarly, the sensitivity of DGGE decreases as the length of the fragment reaches 500 base-pairs. The ddF technique, as a combination of direct sequencing and SSCP, is also limited by the relatively small size of the DNA that can be screened.

According to at least some embodiments of the present invention the step of searching for any of the nucleic acid sequences described here, in tumor cells or in cells derived from a cancer patient is effected by any suitable technique, including, but not limited to, nucleic acid sequencing, polymerase chain reaction, ligase chain reaction, self-sustained synthetic reaction, Qβ-Replicase, cycling probe reaction, branched DNA, restriction fragment length polymorphism analysis, mismatch chemical cleavage, heteroduplex analysis, allele-specific
oligonucleotides, denaturing gradient gel electrophoresis, constant denaturant gel electrophoresis, temperature gradient gel electrophoresis and dideoxy fingerprinting.

Detection may also optionally be performed with a chip or other such device. The nucleic acid sample which includes the candidate region to be analyzed is optionally isolated, amplified and labeled with a reporter group. This reporter group can be a fluorescent group such as phycoerythrin. The labeled nucleic acid is then incubated with the probes immobilized on the chip using a fluidics station.

Once the reaction is completed, the chip is inserted into a scanner and patterns of hybridization are detected. The hybridization data is collected, as a signal emitted from the reporter groups already incorporated into the nucleic acid, which is now bound to the probes attached to the chip. Since the sequence and position of each probe immobilized on the chip is known, the identity of the nucleic acid hybridized to a given probe can be determined.

It will be appreciated that when utilized along with automated equipment, the above described detection methods can be used to screen multiple samples for a disease and/or pathological condition both rapidly and easily.

Small interfering nucleic acids and antisense molecules

According to other embodiments, the present invention encompasses inhibitory RNA oligonucleotides, in particular small interfering nucleic acids (siNA) comprising complementary sequences capable of specifically hybridizing with the polynucleotides according to at least some embodiments of the invention and specifically silencing these genes. According to at least some embodiments the present invention also relates to sequences and constructs encoding such nucleic acids and to the uses of such nucleic acids or constructs to modify HE4 variant expression, particularly to reduce or inhibit its expression.

Certain single stranded nucleic acid molecules are able to form a self-complementary double stranded region where part of the nucleotide sequence is able to interact with another part of the sequence by Watson-Crick base pairing between inverted repeats of the sequence. Where the repeated regions are adjacent or in close proximity to each other, the double stranded regions may form structures known as hairpin structures. The hairpin structure forms with an unpaired "loop" of nucleotides at one end of the hairpin structure, with the inverted repeat sequence annealed. The loop may also facilitate the folding of the nucleic acid chain.

Hairpin RNA sequences have been used in interfering RNA and gene silencing technologies. Such techniques are described for example in US patent no. 6,573,099 and in Grimm D. (Adv. Drug Deliv. Rev. 2009 61 (9): 672-703). According to at least some
embodiments the present invention further contemplates antisense RNA molecules complementary to the polynucleotides according to at least some embodiments of the invention, or to any fragment thereof. Antisense RNA may be introduced into a cell to inhibit translation of the complementary mRNA by hybridizing with the polynucleotides of the according to at least some embodiments of the invention and obstructing the translation machinery.

siNA or antisense molecules according to at least some embodiments of the invention may be used as a therapeutic tool to inhibit HE4 variant expression in vivo.

Amino acid sequences and peptides

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an analog or mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. Polypeptides can be modified, e.g., by the addition of carbohydrate residues to form glycoproteins. The terms "polypeptide," "peptide" and "protein" include glycoproteins, as well as non-glycoproteins.

Polypeptide products can be biochemically synthesized such as by employing standard solid phase techniques. Such methods include but are not limited to exclusive solid phase synthesis, partial solid phase synthesis methods, fragment condensation, classical solution synthesis. These methods are optionally used when the peptide is relatively short (i.e., 10 kDa) and/or when it cannot be produced by recombinant techniques (i.e., not encoded by a nucleic acid sequence) and therefore involves different chemistry.

Solid phase polypeptide synthesis procedures are well known in the art and further described by John Morrow Stewart and Janis Dillaha Young, Solid Phase Peptide Syntheses (2nd Ed., Pierce Chemical Company, 1984).

Synthetic polypeptides can optionally be purified by preparative high performance liquid chromatography [Creighton T. (1983) Proteins, structures and molecular principles. WH Freeman and Co. N.Y.], after which their composition can be confirmed via amino acid sequencing.


According to at least some embodiments, the present invention also encompasses polypeptides encoded by the polynucleotide sequences of the present invention, as well as polypeptides according to the amino acid sequences described herein. According to at least some embodiments the present invention also encompasses homologues of these polypeptides, such homologues can be at least 85%, at least 95% or more say 100% homologous to the amino acid sequences set forth below, as can be determined using BlastP software of the National Center of Biotechnology Information (NCBI) using default parameters, optionally including the following: filtering on (this option filters repetitive or low-complexity sequences from the query using the Seg (protein) program), scoring matrix is BLOSUM62 for proteins, word size is 3, E value is 10, gap costs are 11, 1 (initialization and extension), and number of alignments shown is 50. Optionally, nucleic acid sequence homology/identity is determined by using BlastN software of the National Center of Biotechnology Information (NCBI) using default parameters, which optionally include using the DUST filter program, and also optionally include having an E value of 10, filtering low complexity sequences and a word size of 11. According to at least some embodiments the present invention also encompasses fragments of the above described polypeptides and polypeptides having mutations, such as deletions, insertions or substitutions of one or more amino acids, either naturally occurring or artificially induced, either randomly or in a targeted fashion.

It will be appreciated that peptides according to at least some embodiments of the present invention may be degradation products, synthetic peptides or recombinant peptides as well as peptidomimetics, typically, synthetic peptides and peptoids and semipeptoids which are peptide analogs, which may have, for example, modifications rendering the peptides more stable while in a body or more capable of penetrating into cells. Such modifications include, but are not limited to N terminus modification, C terminus modification, peptide bond modification, including, but not limited to, CH2-NH, CH2-S, CH2-S=O, O=C-NH, CH2-O, CH2=CH2, S=C-NH, CH=CH or CF=CH, backbone modifications, and residue modification. Methods for preparing peptidomimetic compounds are well known in the art and are specified. Further details in this respect are provided hereunder.

Peptide bonds (-CO-NH-) within the peptide may be substituted, for example, by N-methylated bonds (-N(CH3)-CO-), ester bonds (-C(R)H-C-O-O-C(R)-N-), ketomethylen bonds (-CO-CH2-), α-aza bonds (-NH-N(R)-CO-), wherein R is any alkyl, e.g., methyl, carba bonds (-CH2-NH-), hydroxyethylene bonds (-CH(OH)-CH2-), thioamide bonds (-CS-NH-), olefinic
double bonds (-CH=CH-), retro amide bonds (-NH-CO-), peptide derivatives (-N(R)-CH2-CO-)
), wherein R is the "normal" side chain, naturally presented on the carbon atom.

These modifications can occur at any of the bonds along the peptide chain and even at
several (2-3) at the same time.

Natural aromatic amino acids, Trp, Tyr and Phe, may be substituted for synthetic non-
natural acid such as Phenylglycine, TIC, naphthylelanine (NoI), ring-methylated derivatives of
Phe, halogenated derivatives of Phe or o-methyl~Tyr.

In addition to the above, the peptides according to at least some embodiments of the
present invention may also include one or more modified amino acids or one or more non-
amino acid monomers (e.g. fatty acids, complex carbohydrates etc).

As used herein in the specification and in the claims section below the term "amino
acid" or "amino acids" is understood to include the 20 naturally occurring amino acids; those
amino acids often modified post-translationally in vivo, including, for example,
hydroxyproline, phosphoserine and phosphothreonine; and other unusual amino acids
including, but not limited to, 2-aminoacidic acid, hydroxylysine, isodesmosine, nor-valine,
nor-leucine and ornithine. Furthermore, the term "amino acid" includes both D- and L-amino
acids. Non-conventional or modified amino acids can be incorporated in the polypeptides
of this invention as well, as will be known to one skilled in the art.

Since the peptides according to at least some embodiments of the present invention are
optionally utilized in diagnostics which require the peptides to be in soluble form, the peptides
according to at least some embodiments of the present invention optionally include one or
more non-natural or natural polar amino acids, including but not limited to serine and
threonine which are capable of increasing peptide solubility due to their hydroxyl-containing
side chain.

The peptides according to at least some embodiments of the present invention are
optionally utilized in a linear form, although it will be appreciated that in cases where
cyclization does not severely interfere with peptide characteristics, cyclic forms of the peptide
can also be utilized.

The peptides according to at least some embodiments of present invention can be
biochemically synthesized such as by using standard solid phase techniques. These methods
include exclusive solid phase synthesis well known in the art, partial solid phase synthesis
methods, fragment condensation, classical solution synthesis. These methods are optionally
used when the peptide is relatively short (i.e., 10 kDa) and/or when it cannot be produced by
recombinant techniques (i.e., not encoded by a nucleic acid sequence) and therefore involves different chemistry.

Synthetic peptides can be purified by preparative high performance liquid chromatography and the composition of which can be confirmed via amino acid sequencing.


Antibodies:

"Antibody" refers to a polypeptide ligand that is optionally substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof, which specifically binds and recognizes an epitope (e.g., an antigen). The recognized immunoglobulin genes include the kappa and lambda light chain constant region genes, the alpha, gamma, delta, epsilon and mu heavy chain constant region genes, and the myriad-immunoglobulin variable region genes. Antibodies exist, e.g., as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. This includes, e.g., Fab' and F(ab')2 fragments. The term "antibody," as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized de novo using recombinant DNA methodologies. It also includes polyclonal antibodies, monoclonal antibodies, chimeric antibodies, humanized antibodies, or single chain antibodies.

"Fc" portion of an antibody refers to that portion of an immunoglobulin heavy chain that comprises one or more heavy chain constant region domains, CH1, CH2 and CH3, but does not include the heavy chain variable region.

The functional fragments of antibodies, such as Fab, F(ab')2, and Fv that are capable of binding to macrophages, are described as follows: (1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule, can be produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; (2) Fab', the fragment of an antibody molecule that can be obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion
of the heavy chain; two Fab' fragments are obtained per antibody molecule; (3) (Fab')2, the fragment of the antibody that can be obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; F(ab')2 is a dimer of two Fab' fragments held together by two disulfide bonds; (4) Fv, defined as a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and (5) Single chain antibody ("SCA"), a genetically engineered molecule containing the variable region of the light chain and the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule.

Methods of producing polyclonal and monoclonal antibodies as well as fragments thereof are well known in the art (See for example, Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York, 1988, incorporated herein by reference).

Monoclonal antibody development may optionally be performed according to any method that is known in the art. The method described below is provided for the purposes of description only and is not meant to be limiting in any way.

**Antibody engineering in phage display libraries**

Antibodies of this invention may be prepared through the use of phage display libraries, as is known in the art, for example, as described in PCT Application No. WO 94/18219, U.S. Patent No. 6096551, both of which are hereby fully incorporated by reference.

The method involves inducing mutagenesis in a complementarity determining region (CDR) of an immunoglobulin light chain gene for the purpose of producing light chain gene libraries for use in combination with heavy chain genes and gene libraries to produce antibody libraries of diverse and novel immuno-specificities. The method comprises amplifying a CDR portion of an immunoglobulin light chain gene by polymerase chain reaction (PCR) using a PCR primer oligonucleotide. The resultant gene portions are inserted into phagemids for production of a phage display library, wherein the engineered light chains are displayed by the phages, for example for testing their binding specificity.

Antibody fragments according to at least some embodiments of the present invention can be prepared by proteolytic hydrolysis of the antibody or by expression in E. coli or mammalian cells (e.g. Chinese hamster ovary cell culture or other protein expression systems) of DNA encoding the fragment. Antibody fragments can be obtained by pepsin or papain digestion of whole antibodies by conventional methods. For example, antibody fragments can be produced by enzymatic cleavage of antibodies with pepsin to provide a 5S fragment.
denoted F(ab')2. This fragment can be further cleaved using a thiol reducing agent, and optionally a blocking group for the sulphydryl groups resulting from cleavage of disulfide linkages, to produce 3.5S Fab1 monovalent fragments. Alternatively, an enzymatic cleavage using Papain produces two monovalent Fab' fragments and an Fc fragment directly. These methods are described, for example, by Goldenberg, U.S. Pat. Nos. 4,036,945 and 4,331,647, and references contained therein, which patents are hereby incorporated by reference in their entirety. See also Porter, R. R. (Biochem. J. 1959 73, 119-126). Other methods of cleaving antibodies, such as separation of heavy chains to form monovalent light-heavy chain fragments, further cleavage of fragments, or other enzymatic, chemical, or genetic techniques may also be used, so long as the fragments bind to the antigen that is recognized by the intact antibody.

Fv fragments comprise an association of VH and VL chains. This association may be noncovalent, as described in Inbar et al. (Proc. Nat'l Acad. Sci. USA 1972 69, 2659-62). Alternatively, the variable chains can be linked by an intermolecular disulfide bond or cross-linked by chemicals such as glutaraldehyde. Optionally, the Fv fragments comprise VH and VL chains connected by a peptide linker. These single-chain antigen binding proteins (sFv) are prepared by constructing a structural gene comprising DNA sequences encoding the VH and VL domains connected by an oligonucleotide. The structural gene is inserted into an expression vector, which is subsequently introduced into a host cell such as E. coli. The recombinant host cells synthesize a single polypeptide chain with a linker peptide bridging the two V domains. A scFv antibody fragment is an engineered antibody derivative that includes heavy- and light chain variable regions joined by a peptide linker. The minimal size of antibody molecules are those that still comprise the complete antigen binding site. ScFv antibody fragments are potentially more effective than unmodified IgG antibodies. The reduced size of 27-30 kDa permits them to penetrate tissues and solid tumors more readily. Methods for producing sFvs are described, for example, by (Whitlow and Filpula, Methods 1991 2, 97-105); Bird et al., Science 1988 242, 423-426); Pack et al., Bio/Technology 1993 11, 1271-77); and U.S. Patent No. 4,946,778, which is hereby incorporated by reference in its entirety.

Another form of an antibody fragment is a peptide coding for a single complementarity-determining region (CDR). CDR peptides ("minimal recognition units") can be obtained by constructing genes encoding the CDR of an antibody of interest. Such genes are prepared, for example, by using the polymerase chain reaction to synthesize the variable region from RNA of antibody-producing cells. See, for example, Larrick and Fry (Methods,
1991 2, 106-10). Optionally, there may be 1, 2 or 3 CDRs of different chains, but preferably there are 3 CDRs of 1 chain. The chain could be the heavy or the light chain.


The antibody according to at least some embodiments of the present invention specifically binds at least one epitope of the polypeptide variants according to at least some embodiments of the present invention. As used herein, the term "epitope" refers to any antigenic determinant on an antigen to which the paratope of an antibody binds.

Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or carbohydrate side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics.

Optionally, a unique epitope may be created in a variant due to a change in one or more post-translational modifications, including but not limited to glycosylation and/or phosphorylation, as described below. Such a change may also cause a new epitope to be created, for example through removal of glycosylation at a particular site.

An epitope according to at least some embodiments the present invention may also optionally comprise part or all of a unique sequence portion of a variant according to the present invention in combination with at least one other portion of the variant which is not contiguous to the unique sequence portion in the linear polypeptide itself, yet which are able to form an epitope in combination. One or more unique sequence portions may optionally
combine with one or more other non-contiguous portions of the variant (including a portion which may have high homology to a portion of the known protein) to form an epitope.

**Imunoassays**

According to at least some embodiments of the present invention, an immunoassay can be used to qualitatively or quantitatively detect and analyze markers in a sample. This method comprises: providing an antibody that specifically binds to a marker; contacting a sample with the antibody; and detecting the presence of a complex of the antibody bound to the marker in the sample.

To prepare an antibody that specifically binds to a marker, purified protein markers can be used. Antibodies that specifically bind to a protein marker can be prepared using any suitable methods known in the art.

After the antibody is provided, a marker can be detected and/or quantified using any of a number of well recognized immunological binding assays. Useful assays include, for example, an enzyme immune assay (EIA) such as enzyme-linked immunosorbent assay (ELISA), a radioimmune assay (RIA), a Western blot assay, or a slot blot assay see, e.g., U.S. Patent Nos. 4,366,241; 4,376,110; 4,517,288; and 4,837,168). Generally, a sample obtained from a subject can be contacted with the antibody that specifically binds the marker.

Optionally, the antibody can be fixed to a solid support to facilitate washing and subsequent isolation of the complex, prior to contacting the antibody with a sample. Examples of solid supports include but are not limited to glass or plastic in the form of, e.g., a microtiter plate, a stick, a bead, or a microbead. Antibodies can also be attached to a solid support.

After incubating the sample with antibodies, the mixture is washed and the antibody-marker complex formed can be detected. This can be accomplished by incubating the washed mixture with a detection reagent. Alternatively, the marker in the sample can be detected using an indirect assay, wherein, for example, a second, labeled antibody is used to detect bound marker-specific antibody, and/or in a competition or inhibition assay wherein, for example, a monoclonal antibody which binds to a distinct epitope of the marker are incubated simultaneously with the mixture.

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. However, the incubation time will depend upon the assay format, marker, volume of solution, concentrations and the like. Usually the
assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

The immunoassay can be used to determine a test amount of a marker in a sample from a subject. First, a test amount of a marker in a sample can be detected using the immunoassay methods described above. If a marker is present in the sample, it will form an antibody-marker complex with an antibody that specifically binds the marker under suitable incubation conditions described above. The amount of an antibody-marker complex can optionally be determined by comparing to a standard. As noted above, the test amount of marker need not be measured in absolute units, as long as the unit of measurement can be compared to a control amount and/or signal.

Preferably used are antibodies which specifically interact with the polypeptides according to at least some embodiments of the present invention and not with wild type proteins or other isoforms thereof, for example. Such antibodies are directed, for example, to the unique sequence portions of the polypeptide variants according to at least some embodiments of the present invention, including but not limited to bridges, heads, tails and insertions described in greater detail below. Preferred embodiments of antibodies according to at least some embodiments the present invention are described in greater detail with regard to the section entitled "Antibodies".

Radioimmunoassay (RIA): In one version, this method involves precipitation of the desired substrate and in the methods detailed hereinbelow, with a specific antibody and radiolabeled antibody binding protein (e.g., protein A labeled with I^{125}) immobilized on a precipitable carrier such as agarose beads. The number of counts in the precipitated pellet is proportional to the amount of substrate.

In an alternate version of the RIA, a labeled substrate and an unlabelled antibody binding protein are employed. A sample containing an unknown amount of substrate is added in varying amounts. The decrease in precipitated counts from the labeled substrate is proportional to the amount of substrate in the added sample.

Enzyme linked immunosorbent assay (ELISAV). This method involves fixation of a sample (e.g., fixed cells or a proteinaceous solution) containing a protein substrate to a surface such as a well of a microtiter plate. A substrate specific antibody coupled to an enzyme is applied and allowed to bind to the substrate. Presence of the antibody is then detected and quantitated by a colorimetric reaction employing the enzyme coupled to the antibody. Enzymes commonly employed in this method include horseradish peroxidase and alkaline
phosphatase. If well calibrated and within the linear range of response, the amount of substrate present in the sample is proportional to the amount of color produced. A substrate standard is generally employed to improve quantitative accuracy.

**Western blot:** This method involves separation of a substrate from other protein by means of an acrylamide gel followed by transfer of the substrate to a membrane (e.g., nylon or PVDF). Presence of the substrate is then detected by antibodies specific to the substrate, which are in turn detected by antibody binding reagents. Antibody binding reagents may be, for example, protein A, or other antibodies. Antibody binding reagents may be radiolabeled or enzyme linked as described hereinabove. Detection may be by autoradiography, colorimetric reaction or chemiluminescence. This method allows both quantitation of an amount of substrate and determination of its identity by a relative position on the membrane which is indicative of a migration distance in the acrylamide gel during electrophoresis.

**Immunohistochemical analysis:** This method involves detection of a substrate in situ in fixed cells by substrate specific antibodies. The substrate specific antibodies may be enzyme linked or linked to fluorophores. Detection is by microscopy and subjective evaluation. If enzyme linked antibodies are employed, a colorimetric reaction may be required.

**Fluorescence activated cell sorting (FACS):** This method involves detection of a substrate in situ in cells by substrate specific antibodies. The substrate specific antibodies are linked to fluorophores. Detection is by means of a cell sorting machine which reads the wavelength of light emitted from each cell as it passes through a light beam. This method may employ two or more antibodies simultaneously.

**Radio-imaging methods**

These methods include but are not limited to, positron emission tomography (PET) single photon emission computed tomography (SPECT). Both of these techniques are non-invasive, and can be used to detect and/or measure a wide variety of tissue events and/or functions, such as detecting cancerous cells for example. Unlike PET, SPECT can optionally be used with two labels simultaneously. SPECT has some other advantages as well, for example with regard to cost and the types of labels that can be used. For example, US Patent No. 6,696,686 describes the use of SPECT for detection of breast cancer, and is hereby incorporated by reference as if fully set forth herein.

**Display libraries**

According to still another aspect of the present invention there is provided a display library comprising a plurality of display vehicles (such as phages, viruses or bacteria) each
displaying at least 6, at least 7, at least 8, at least 9, at least 10, 10-15, 12-17, 15-20, 15-30 or 20-50 consecutive amino acids derived from the polypeptide sequences according to at least some embodiments of the present invention.


Theranostics

The term theranostics describes the use of diagnostic testing to diagnose the disease, choose the correct treatment regime according to the results of diagnostic testing and/or monitor the patient response to therapy according to the results of diagnostic testing. Theranostic tests can be used to select patients for treatments that are particularly likely to benefit them and unlikely to produce side-effects. They can also provide an early and objective indication of treatment efficacy in individual patients, so that (if necessary) the treatment can be altered with a minimum of delay. For example: DAKO and Genentech together created HercepTest and Herceptin (trastuzumab) for the treatment of breast cancer, the first theranostic test approved simultaneously with a new therapeutic drug. In addition to HercepTest (which is an immunohistochemical test), other theranostic tests are in development which use traditional clinical chemistry, immunoassay, cell-based technologies and nucleic acid tests. PPGx's recently launched TPMT (thiopurine S-methyltransferase) test, which is enabling doctors to identify patients at risk for potentially fatal adverse reactions to 6-mercaptopurine, an agent used in the treatment of leukemia. Also, Nova Molecular pioneered SNP genotyping of the apolipoprotein E gene to predict Alzheimer's disease
patients' responses to cholinomimetic therapies and it is now widely used in clinical trials of
ew drugs for this indication. Thus, the field of theranostics represents the intersection of
diagnostic testing information that predicts the response of a patient to a treatment with the
selection of the appropriate treatment for that particular patient.

5 **Surrogate markers**

A surrogate marker is a marker, that is detectable in a laboratory and/or according to a
physical sign or symptom on the patient, and that is used in therapeutic trials as a substitute
for a clinically meaningful endpoint. The surrogate marker is a direct measure of how a
patient feels, functions, or survives which is expected to predict the effect of the therapy. The
need for surrogate markers mainly arises when such markers can be measured earlier, more
conveniently, or more frequently than the endpoints of interest in terms of the effect of a
treatment on a patient, which are referred to as the clinical endpoints. Ideally, a surrogate
marker should be biologically plausible, predictive of disease progression and measurable by
standardized assays (including but not limited to traditional clinical chemistry, immunoassay,
cell-based technologies, nucleic acid tests and imaging modalities).

Surrogate endpoints were used first mainly in the cardiovascular area. For example,
antihypertensive drugs have been approved based on their effectiveness in lowering blood
pressure. Similarly, in the past, cholesterol-lowering agents have been approved based on their
ability to decrease serum cholesterol, not on the direct evidence that they decrease mortality
from atherosclerotic heart disease. The measurement of cholesterol levels is now an accepted
surrogate marker of atherosclerosis. In addition, currently two commonly used surrogate
markers in HIV studies are CD4+ T cell counts and quantitative plasma HIV RNA (viral
load). In some embodiments of this invention, the polypeptide/polynucleotide expression
pattern may serve as a surrogate marker for a particular disease, as will be appreciated by one
skilled in the art.

**Monoclonal antibody therapy**

In some embodiments, monoclonal antibodies are useful for the identification of
cancer cells. In some embodiments, monoclonal antibody therapy is a form of passive
immunotherapy useful in cancer treatment. Such antibodies may comprise naked monoclonal
antibodies or conjugated monoclonal antibodies - joined to a chemotherapy drug, radioactive
particle, or a toxin (a substance that poisons cells). In some embodiments, the former is
directly cytotoxic to the target (cancer) cell, or in another embodiment, stimulates or
otherwise participates in an immune response ultimately resulting in the lysis of the target cell.

In some embodiments, the conjugated monoclonal antibodies are joined to drugs, toxins, or radioactive atoms. They are used as delivery vehicles to take those substances directly to the cancer cells. The MAb acts as a homing device, circulating in the body until it finds a cancer cell with a matching antigen. It delivers the toxic substance to where it is needed most, minimizing damage to normal cells in other parts of the body. Conjugated MAbs are also sometimes referred to as "tagged," "labeled," or "loaded" antibodies. MAbs with chemotherapy drugs attached are generally referred to as chemolabeled. MAbs with radioactive particles attached are referred to as radiolabeled, and this type of therapy is known as radioimmunotherapy (RIT). MAbs attached to toxins are called immunotoxins.

An illustrative, non-limiting example is provided herein of a method of treatment of a patient with an antibody to a variant as described herein, such that the variant is a target of the antibody. A patient with breast cancer is treated with a radiolabeled humanized antibody against an appropriate breast cancer target as described herein. The patient is optionally treated with a dosage of labeled antibody ranging from 10 to 30 mCi. Of course any type of therapeutic label may optionally be used.

The following examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way.

All patent and literature references cited in the present specification are hereby incorporated by reference in their entirety.

All technical and scientific terms used herein should be understood to have the meaning commonly understood by a person skilled in the art to which this invention belongs, as well as any other specified description. The following references provide one of skill in the art with a general definition of many of the terms used in this invention: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). All of these are hereby incorporated by reference as if fully set forth herein.
EXAMPLES

EXAMPLE 1: DESCRIPTION OF THE METHODOLOGY UNDERTAKEN TO UNCOVER THE HE4 VARIANTS

Human ESTs and cDNAs were obtained from GenBank versions 136 (June 15, 2003 ftp.ncbi.nih.gov/genbank/release.notes/gbl36.release.notes); NCBI genome assembly of April 2003; RefSeq sequences from June 2003;  Genbank version 139 (December 2003); Human Genome from NCBI (Build 34) (from Oct 2003); and RefSeq sequences from December 2003. With regard to GenBank sequences, the human EST sequences from the EST (GBEST) section and the human mRNA sequences from the primate (GBPRI) section were used; also the human nucleotide RefSeq mRNA sequences were used (see for example www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html and for a reference to the EST section, see www.ncbi.nlm.nih.gov/dbEST/; a general reference to dbEST, the EST database in GenBank, may be found in Boguski et al, Nat Genet. 1993; 4(4), 332-3; all of which are hereby incorporated by reference as if fully set forth herein).

Novel variants including splice variants were predicted using the LEADS clustering and assembly system as described in Sorek, R., Ast, G. & Graur, D. Alu-containing exons are alternatively spliced. Genome Res 12, 1060-7 (2002); US patent No: 6,625,545; and U.S. Pat. Appl. No. 10/426,002, published as US20040101876 on May 27 2004; all of which are hereby incorporated by reference as if fully set forth herein. Briefly, the software cleans the expressed sequences from repeats, vectors and immunoglobulins. It then aligns the expressed sequences to the genome taking alternatively splicing into account and clusters overlapping expressed sequences into "clusters" that represent genes or partial genes.

These were annotated using the GeneCarta (Compugen, Tel-Aviv, Israel) platform. The GeneCarta platform includes a rich pool of annotations, sequence information (particularly of spliced sequences), chromosomal information, alignments, and additional information such as SNPs, gene ontology terms, expression profiles, functional analyses, detailed domain structures, known and predicted proteins and detailed homology reports.

A brief explanation is provided with regard to the method of selecting the candidates. However, it should be noted that this explanation is provided for descriptive purposes only, and is not intended to be limiting in any way. The potential markers were identified by a computational process that was designed to find genes and/or their variants (including splice variants) that are differentially expressed in cancer tissues as opposed to non-cancerous tissues. Various parameters related to the information in the EST libraries, determined according to classification by library annotation, and/or manual classification process, were
used to assist in locating genes and/or variants thereof that are specifically and/or
differentially expressed in cancerous tissues.

**Identification of differentially expressed gene products - Algorithm**

In order to distinguish between differentially expressed gene products and constitutively
expressed genes (i.e., house keeping genes) an algorithm based on an analysis of frequencies was
configured. A specific algorithm for identification of transcripts over expressed in cancer is
described hereinafter.

**Data analysis**

**Library annotation** - EST libraries are manually classified according to:

(i) Tissue origin

(ii) Biological source - Examples of frequently used biological sources for
construction of EST libraries include cancer cell-lines; normal tissues; cancer tissues;
fetal tissues; and others such as normal cell lines and pools of normal cell-lines, cancer
cell-lines and combinations thereof. A specific description of abbreviations used below
with regard to these tissues/cell lines etc is given above.

(iii) Protocol of library construction - various methods are known in the art for
library construction including normalized library construction; non-normalized library
construction; subtracted libraries; ORESTES and others. It will be appreciated that at
times the protocol of library construction is not indicated.

The following rules are followed:

EST libraries originating from identical biological samples are considered as a single
library.

EST libraries which include above-average levels of DNA contamination are eliminated.

**Computational Method** - development of engines which are capable of identifying
genomes and variants (including splice variants) that are temporally and specifically expressed.

Clusters (genes) having at least five sequences including at least two sequences from the
tissue of interest are analyzed.

**Identification of genes over expressed in cancer**

Two different scoring algorithms were developed.

**Libraries score** - candidate sequences which are supported by a number of cancer
libraries are more likely to serve as specific and effective diagnostic markers.

**The basic algorithm** - for each cluster, the number of cancer and normal libraries
contributing sequences to the cluster was counted. Fisher exact test was used to check if cancer
libraries are significantly over-represented in the cluster as compared to the total number of cancer and normal libraries.

Library **counting** - small libraries (e.g., less than 1000 sequences) were excluded from consideration unless they participate in the cluster. For this reason, the total number of libraries is actually adjusted for each cluster.

**Clones no. score** - generally, when the number of ESTs is much higher in the cancer libraries relative to the normal libraries it might indicate actual over-expression.

**The algorithm** -

**Clone counting:** For counting EST clones each library protocol class was given a weight based on our belief of how much the protocol reflects actual expression levels:

(i) non-normalized: 1
(ii) normalized: 0.2
(iii) all other classes: 0.1

**Clones number score** - The total weighted number of EST clones from cancer libraries was compared to the EST clones from normal libraries. To avoid cases where one library contributes to the majority of the score, the contribution of the library that gives most clones for a given cluster was limited to 2 clones.

The score was computed as

$$\frac{c+1}{C} \frac{n+1}{N}$$

where:

- c - weighted number of "cancer" clones in the cluster.
- C - weighted number of clones in all "cancer" libraries.
- n - weighted number of "normal" clones in the cluster.
- N - weighted number of clones in all "normal" libraries.

**Clones number score significance** - Fisher exact test was used to check if EST clones from cancer libraries are significantly over-represented in the cluster as compared to the total number of EST clones from cancer and normal libraries.

Two search approaches were used to find either general cancer-specific candidates or tumor specific candidates.
• Libraries/sequences originating from tumor tissues are counted as well as libraries originating from cancer cell-lines ("normal" cell-lines were ignored).
• Only libraries/sequences originating from tumor tissues are counted.

**Identification of tissue specific genes**

For detection of tissue specific clusters, tissue libraries/sequences were compared to the total number of libraries/sequences in cluster. Similar statistical tools to those described in above were employed to identify tissue specific genes. Tissue abbreviations are the same as for cancerous tissues, but are indicated with the header "normal tissue".

The algorithm - for each tested tissue T and for each tested cluster the following were examined:

1. Each cluster includes at least 2 libraries from the tissue T. At least 3 clones (weighed - as described above) from tissue T in the cluster; and
2. Clones from the tissue T are at least 40% from all the clones participating in the tested cluster

Fisher exact test P-values were computed both for library and weighted clone counts to check that the counts are statistically significant.

**EXAMPLE 2: OLIGONUCLEOTIDE-BASED MICRO-ARRAY EXPERIMENT PROTOCOL**

Microarray fabrication

Microarrays (chips) were printed by pin deposition using the MicroGrid II MGII 600 robot from BioRobotics Limited (Cambridge, UK). 50-mer oligonucleotides target sequences were designed by Compugen Ltd (Tel-Aviv, IL) as described by A. Shoshan et al, "Optical technologies and informatics", Proceedings of SPIE. Vol 4266, pp. 86-95 (2001). The designed oligonucleotides were synthesized and purified by desalting with the Sigma-Genosys system (The Woodlands, TX, US) and all of the oligonucleotides were joined to a C6 amino-modified linker at the 5' end, or being attached directly to CodeLink slides (Cat #25-6700-01. Amersham Bioscience, Piscataway, NJ, US). The 50-mer oligonucleotides, forming the target sequences, were first suspended in Ultra-pure DDW (Cat # 01-866-IA Kibbutz Beit-Haemek, Israel) to a concentration of 50µM. Before printing the slides, the oligonucleotides were resuspended in 300mM sodium phosphate (pH 8.5) to final concentration of 150mM and printed at 35-40% relative humidity at 21°C.
Each slide contained a total of 9792 features in 32 subarrays. Of these features, 4224 features were sequences of interest according to at least some embodiments of the present invention and negative controls that were printed in duplicate. An additional 288 features (96 target sequences printed in triplicate) contained housekeeping genes from Human Evaluation Library2, Compugen Ltd, Israel. Another 384 features are E. coli spikes 1-6, which are oligos to E. coli genes which are commercially available in the Array Control product (Array control- sense oligo spots, Ambion Inc. Austin, TX. Cat #1781, Lot #112K06).

Post-coupling processing of printed slides

After the spotting of the oligonucleotides to the glass (CodeLink) slides, the slides were incubated for 24 hours in a sealed saturated NaCl humidification chamber (relative humidity 70-75%).

Slides were treated for blocking of the residual reactive groups by incubating them in blocking solution at 50°C for 15 minutes (10ml/slide of buffer containing 0.1M Tris, 50mM ethanolamine, 0.1% SDS). The slides were then rinsed twice with Ultra-pure DDW (double distilled water). The slides were then washed with wash solution (10ml/slide. 4X SSC, 0.1% SDS) at 50°C for 30 minutes on the shaker. The slides were then rinsed twice with Ultra-pure DDW, followed by drying by centrifugation for 3 minutes at 800 rpm.

Next, in order to assist in automatic operation of the hybridization protocol, the slides were treated with Ventana Discovery hybridization station barcode adhesives. The printed slides were loaded on a Bio-Optica (Milan, Italy) hematology staining device and were incubated for 10 minutes in 50ml of 3-Aminopropyl Triethoxysilane (Sigma A3648 lot #122K589). Excess fluid was dried and slides were then incubated for three hours in 20 mm/Hg in a dark vacuum desiccator (Pelco 2251, Ted Pella, Inc. Redding CA).

The following protocol was then followed with the Genisphere 900-RP (random primer), with mini elute columns on the Ventana Discovery HybStation™, to perform the microarray experiments. Briefly, the protocol was performed as described with regard to the instructions and information provided with the device itself. The protocol included cDNA synthesis and labeling. cDNA concentration was measured with the TBS-380 (Turner Biosystems. Sunnyvale, CA.) PicoFlour, which is used with the OliGreen ssDNA Quantitation reagent and kit.

Hybridization was performed with the Ventana Hybridization device, according to the provided protocols (Discovery Hybridization Station Tuscon AZ).
The slides were then scanned with GenePix 4000B dual laser scanner from Axon Instruments Inc, and analyzed by GenePix Pro 5.0 software.

Briefly, the oligonucleotide based microarray fabrication and the experimental flow is summarized below: DNA oligonucleotides at 25uM were deposited (printed) onto Amersham 'CodeLink' glass slides generating a well defined 'spot'. These slides are covered with a long-chain, hydrophilic polymer chemistry that creates an active 3-D surface that covalently binds the DNA oligonucleotides 5'-end via the C6-amine modification. This binding ensures that the full length of the DNA oligonucleotides is available for hybridization to the cDNA and also allows lower background, high sensitivity and reproducibility.

Figure 1 shows a schematic method for performing the microarray experiments. It should be noted that stages on the left-hand or right-hand side may optionally be performed in any order, including in parallel, until stage 4 (hybridization). Briefly, on the left-hand side, the target oligonucleotides are being spotted on a glass microscope slide (although optionally other materials could be used) to form a spotted slide (stage 1)- On the right hand side, control sample RNA and cancer sample RNA are Cy3 and Cy5 labeled, respectively (stage 2), to form labeled probes. It should be noted that the control and cancer samples come from corresponding tissues (for example, normal prostate tissue and cancerous prostate tissue). Furthermore, the tissue from which the RNA was taken is indicated below in the specific examples of data for particular clusters, with regard to overexpression of an oligonucleotide from a "chip" (microarray), as for example "prostate" for chips in which prostate cancerous tissue and normal tissue were tested as described above. In stage 3, the probes are mixed. In stage 4, hybridization is performed to form a processed slide. In stage 5, the slide is washed and scanned to form an image file, followed by data analysis in stage 6.

**EXAMPLE 3: METHODS USED TO ANALYZE THE EXPRESSION OF THE RNA ENCODING THE HE4-VARIANT PROTEINS**

The markers according to at least some embodiments of the present invention were tested with regard to their expression in ovarian cancerous and non-cancerous tissue samples. Unless otherwise noted, all experimental data relates to variants according to at least some embodiments of the present invention, named according to the segment being tested (as expression was tested through RT-PCR as described). A description of the samples used in the ovarian cancer testing panel is provided in Table 1_1 below. A description of the samples used in normal tissue panel is provided in Table 1_2 below. Tests were then performed as described in the "Materials and Experimental Procedures" section below.
Table 1-1: Tissue samples in ovary panel

<table>
<thead>
<tr>
<th>Source</th>
<th>sample name</th>
<th>Ovary tissue type</th>
<th>Stage</th>
<th>Sample ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asterand</td>
<td>1-As-Ser SI</td>
<td>Serous carcinoma</td>
<td>I</td>
<td>71900A2</td>
</tr>
<tr>
<td>Asterand</td>
<td>2-As-Ser SI</td>
<td>Serous carcinoma</td>
<td>I</td>
<td>70270A1</td>
</tr>
<tr>
<td>Asterand</td>
<td>3-As-Ser SIB</td>
<td>Serous carcinoma</td>
<td>IB</td>
<td>40771B1</td>
</tr>
<tr>
<td>Asterand</td>
<td>4-As-Ser SIB</td>
<td>Serous carcinoma</td>
<td>IB</td>
<td>32667B1</td>
</tr>
<tr>
<td>Asterand</td>
<td>5-As-Ser SIC</td>
<td>Serous carcinoma</td>
<td>IC</td>
<td>22996A1</td>
</tr>
<tr>
<td>Asterand</td>
<td>6-As-Ser SIIA</td>
<td>Serous carcinoma</td>
<td>IIA</td>
<td>40773C1</td>
</tr>
<tr>
<td>GCI</td>
<td>7-GC-Ser SIIIB</td>
<td>Serous carcinoma</td>
<td>IIB</td>
<td>20370</td>
</tr>
<tr>
<td>GCI</td>
<td>8-GC-Ser SIIIB</td>
<td>Serous carcinoma</td>
<td>IIB</td>
<td>7B3DP</td>
</tr>
<tr>
<td>Asterand</td>
<td>9-As-Ser SIIIC</td>
<td>Serous carcinoma</td>
<td>IIIC</td>
<td>19832A1</td>
</tr>
<tr>
<td>GCI</td>
<td>10-GC-Ser SIIIC</td>
<td>Serous carcinoma</td>
<td>IIIC</td>
<td>3NTIS</td>
</tr>
<tr>
<td>GCI</td>
<td>11-GC-Ser SIIIC</td>
<td>Serous carcinoma</td>
<td>IIIC</td>
<td>CEJUS</td>
</tr>
<tr>
<td>GCI</td>
<td>12-GC-Ser SIIIC</td>
<td>Serous carcinoma</td>
<td>IIIC</td>
<td>5NCLK</td>
</tr>
<tr>
<td>GCI</td>
<td>13-GC-Ser SIIIC</td>
<td>Serous carcinoma</td>
<td>IIIC</td>
<td>1HISH</td>
</tr>
<tr>
<td>GCI</td>
<td>14-GC-Ser SIIIC</td>
<td>Serous carcinoma</td>
<td>IIIC</td>
<td>7RMHZ</td>
</tr>
<tr>
<td>GCI</td>
<td>15-GC-Ser SIIIC</td>
<td>Serous carcinoma</td>
<td>IIIC</td>
<td>4WAAB</td>
</tr>
<tr>
<td>GCI</td>
<td>16-GC-Ser SIIIC</td>
<td>Serous carcinoma</td>
<td>IIIC</td>
<td>79Z67</td>
</tr>
<tr>
<td>GCI</td>
<td>17-GC-Ser SIIIC</td>
<td>Serous carcinoma</td>
<td>IIIC</td>
<td>DDSNL</td>
</tr>
<tr>
<td>GCI</td>
<td>18-GC-Ser SIV</td>
<td>Serous carcinoma</td>
<td>IV</td>
<td>DH8SP</td>
</tr>
<tr>
<td>GCI</td>
<td>19-GC-Endo SIA</td>
<td>Endometrioid Carcinoma</td>
<td>IA</td>
<td>E2WKF</td>
</tr>
<tr>
<td>GCI</td>
<td>20-GC-Endo SIA</td>
<td>Endometrioid Carcinoma</td>
<td>IA</td>
<td>5895C</td>
</tr>
<tr>
<td>GCI</td>
<td>21-GC-Endo SIA</td>
<td>Endometrioid Carcinoma</td>
<td>IA</td>
<td>533DX</td>
</tr>
<tr>
<td>GCI</td>
<td>22-GC-Endo SIA</td>
<td>Endometrioid Carcinoma</td>
<td>IA</td>
<td>HZ2EY</td>
</tr>
<tr>
<td>GCI</td>
<td>23-GC-Endo SIA</td>
<td>Endometrioid Carcinoma</td>
<td>IA</td>
<td>RWOIV</td>
</tr>
<tr>
<td>GCI</td>
<td>24-GC-Endo SIIA</td>
<td>Endometrioid Carcinoma</td>
<td>IIA</td>
<td>1U52X</td>
</tr>
<tr>
<td>GCI</td>
<td>25-GC-Endo SIIIB</td>
<td>Endometrioid Carcinoma</td>
<td>IIB</td>
<td>A17WS</td>
</tr>
<tr>
<td>GCI</td>
<td>26-GC-Endo SIIIC</td>
<td>Endometrioid Carcinoma</td>
<td>IIIC</td>
<td>1VT3I</td>
</tr>
<tr>
<td>GCI</td>
<td>27-GC-Endo SIIIC</td>
<td>Endometrioid Carcinoma</td>
<td>IIIC</td>
<td>PZQXH</td>
</tr>
<tr>
<td>GCI</td>
<td>28-GC-Endo SIV</td>
<td>Endometrioid Carcinoma</td>
<td>IV</td>
<td>18VHZ</td>
</tr>
<tr>
<td>GOG</td>
<td>29-(21)-GO- Muc SIA</td>
<td>Mucinous Carcinoma</td>
<td>IA</td>
<td>95-10-G020</td>
</tr>
<tr>
<td>GCI</td>
<td>30-GC-Muc SIC</td>
<td>Mucinous Carcinoma</td>
<td>IC</td>
<td>IMDA1</td>
</tr>
<tr>
<td>Asterand</td>
<td>31-As-Muc SIC</td>
<td>Mucinous Carcinoma</td>
<td>IC</td>
<td>18920A1</td>
</tr>
<tr>
<td>ABS</td>
<td>32-(22)-AB-Muc SIC</td>
<td>Mucinous Carcinoma</td>
<td>IC</td>
<td>A0139</td>
</tr>
<tr>
<td>GCI</td>
<td>33-GC-Muc SIIA</td>
<td>Mucinous Carcinoma</td>
<td>IIA</td>
<td>NJM4U</td>
</tr>
<tr>
<td>ABS</td>
<td>34-(20)- AB-Muc SIIA</td>
<td>Mucinous Carcinoma</td>
<td>IIA</td>
<td>USA-00273</td>
</tr>
<tr>
<td>GCI</td>
<td>35-GC-Muc SIIIA</td>
<td>Mucinous Carcinoma</td>
<td>IIA</td>
<td>RAFCW</td>
</tr>
<tr>
<td>Asterand</td>
<td>36-As-Muc SIIIC</td>
<td>Mucinous Carcinoma</td>
<td>IIC</td>
<td>72888A1</td>
</tr>
<tr>
<td>Asterand</td>
<td>37-As-Muc SIIIC</td>
<td>Mucinous Carcinoma</td>
<td>IIIC</td>
<td>29374B1</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>-------------------------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>GCI</td>
<td>38-GC-Muc Border SIA</td>
<td>Mucinous borderline tumor</td>
<td>IA</td>
<td>SC656</td>
</tr>
<tr>
<td>GCI</td>
<td>39-GC-Muc Border SIA</td>
<td>Mucinous borderline tumor</td>
<td>IA</td>
<td>3D5FO</td>
</tr>
<tr>
<td>GCI</td>
<td>40-GC-Muc Border SIA</td>
<td>Mucinous borderline tumor</td>
<td>IA</td>
<td>7JP3F</td>
</tr>
<tr>
<td>GOG</td>
<td>41-(62)-Go-Ben Muc</td>
<td>Benign mucinous</td>
<td></td>
<td>99-10-G442</td>
</tr>
<tr>
<td>GCI</td>
<td>43-GC-Ben Muc</td>
<td>Benign mucinous</td>
<td></td>
<td>QL9KY</td>
</tr>
<tr>
<td>Asterand</td>
<td>44-As-Ben Muc</td>
<td>Benign mucinous</td>
<td></td>
<td>30534A1</td>
</tr>
<tr>
<td>GOG</td>
<td>45-(56)-GO-Ben Muc</td>
<td>Benign mucinous</td>
<td></td>
<td>99-01-G407</td>
</tr>
<tr>
<td>GCI</td>
<td>46-GC-Ben Muc</td>
<td>Benign mucinous</td>
<td></td>
<td>943EC</td>
</tr>
<tr>
<td>GCI</td>
<td>47-GC-Ben Muc</td>
<td>Benign mucinous</td>
<td></td>
<td>JO8W7</td>
</tr>
<tr>
<td>Asterand</td>
<td>48-As-Ben Ser</td>
<td>Benign serous</td>
<td></td>
<td>30645B1</td>
</tr>
<tr>
<td>GOG</td>
<td>49-(64)-GO-Ben Ser</td>
<td>Benign serous</td>
<td></td>
<td>99-06-G039</td>
</tr>
<tr>
<td>GCI</td>
<td>50-GC-Ben Ser</td>
<td>Benign serous</td>
<td></td>
<td>DQQ2F</td>
</tr>
<tr>
<td>Asterand</td>
<td>51-As-Ben Ser</td>
<td>Benign serous</td>
<td></td>
<td>8275A1</td>
</tr>
<tr>
<td>Asterand</td>
<td>52-As-N BM</td>
<td>Normal</td>
<td></td>
<td>23054A1</td>
</tr>
<tr>
<td>Asterand</td>
<td>53-As-N BM</td>
<td>Normal</td>
<td></td>
<td>30488A1</td>
</tr>
<tr>
<td>Asterand</td>
<td>54-As-N BM</td>
<td>Normal</td>
<td></td>
<td>30496B1</td>
</tr>
<tr>
<td>Asterand</td>
<td>55-As-N BM</td>
<td>Normal</td>
<td></td>
<td>30499C1</td>
</tr>
<tr>
<td>GCI</td>
<td>56-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>WPU1U</td>
</tr>
<tr>
<td>GCI</td>
<td>58-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>76VM9</td>
</tr>
<tr>
<td>GCI</td>
<td>59-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>DWHTZ</td>
</tr>
<tr>
<td>GCI</td>
<td>60-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>SJ2R2</td>
</tr>
<tr>
<td>GCI</td>
<td>61-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>9RQMN</td>
</tr>
<tr>
<td>GCI</td>
<td>62-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>TOAE5</td>
</tr>
<tr>
<td>GCI</td>
<td>63-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>TW9PM</td>
</tr>
<tr>
<td>GCI</td>
<td>64-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>2VND2</td>
</tr>
<tr>
<td>GCI</td>
<td>65-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>L629F</td>
</tr>
<tr>
<td>GCI</td>
<td>66-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>XLB23</td>
</tr>
<tr>
<td>GCI</td>
<td>67-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>IDUVY</td>
</tr>
<tr>
<td>GCI</td>
<td>68-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>2CXAD</td>
</tr>
<tr>
<td>GCI</td>
<td>69-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>PEQ6C</td>
</tr>
<tr>
<td>GCI</td>
<td>70-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>DD73B</td>
</tr>
<tr>
<td>GCI</td>
<td>71-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>E2UF7</td>
</tr>
<tr>
<td>GCI</td>
<td>74-GC-N PS</td>
<td>Normal</td>
<td></td>
<td>FDPL9</td>
</tr>
<tr>
<td>BioChain</td>
<td>76-(46)-Bc-N PM</td>
<td>Normal</td>
<td></td>
<td>A504086</td>
</tr>
<tr>
<td>Ichilov</td>
<td>77-(71)-Ic-N PM</td>
<td>Normal</td>
<td></td>
<td>CG-188-7</td>
</tr>
<tr>
<td>BioChain</td>
<td>78-(48)-Bc-N PM</td>
<td>Normal</td>
<td>Stage</td>
<td>Reference</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>--------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>GOG</td>
<td>79-(32)-GO-SIB-SER</td>
<td>Serous papillary cystadenocarcinoma</td>
<td>Stage 1B</td>
<td>93-09-4901</td>
</tr>
<tr>
<td>GOG</td>
<td>80-(30)-GO-SIIIA-SER</td>
<td>Papillary serous carcinoma</td>
<td>Stage 3A</td>
<td>2001-08-G011</td>
</tr>
<tr>
<td>GOG</td>
<td>81-(70)-GO-SIIIB-SER</td>
<td>Papillary serous adenocarcinoma</td>
<td>Stage 3B</td>
<td>95-08-G069</td>
</tr>
<tr>
<td>GOG</td>
<td>82-(3)-GO-SIIC-SER</td>
<td>Adenocarcinoma</td>
<td>Stage 3C</td>
<td>99-12-G432</td>
</tr>
<tr>
<td>GOG</td>
<td>83-(29)-GO-SIIC-SER</td>
<td>Serous adenocarcinoma</td>
<td>Stage 3C</td>
<td>2001-12-G035</td>
</tr>
<tr>
<td>ABS</td>
<td>84-(25)-AB-SIIC-SER</td>
<td>Papillary serous adenocarcinoma</td>
<td>Stage 3C</td>
<td>N0021</td>
</tr>
<tr>
<td>GOG</td>
<td>85-(13)-GO-SIIC-SER</td>
<td>Adenocarcinoma from primary peritoneal</td>
<td>stage3C</td>
<td>94-05-7603</td>
</tr>
<tr>
<td>Biochain</td>
<td>86-(33)-BC-SER</td>
<td>Serous papillary cystadenocarcinoma</td>
<td></td>
<td>A503175</td>
</tr>
<tr>
<td>Biochain</td>
<td>87-(14)-Bc-SER</td>
<td>Adenocarcinoma</td>
<td></td>
<td>A501111</td>
</tr>
<tr>
<td>Biochain</td>
<td>88-(12)-Bc-SER</td>
<td>Adenocarcinoma</td>
<td></td>
<td>A406023</td>
</tr>
<tr>
<td>Biochain</td>
<td>89-(11)-Bc-SER</td>
<td>Adenocarcinoma</td>
<td></td>
<td>A407068</td>
</tr>
<tr>
<td>ABS</td>
<td>90-(4)-AB-SER</td>
<td>Papillary cystadenocarcinoma</td>
<td>UN</td>
<td>ILS-7286</td>
</tr>
<tr>
<td>ABS</td>
<td>91-(6)-AB-SER</td>
<td>adenocarcinoma</td>
<td>UN</td>
<td>A0106</td>
</tr>
<tr>
<td>ABS</td>
<td>92-(3)-AB-SER</td>
<td>Papillary adenocarcinoma</td>
<td>UN</td>
<td>ILS-1431</td>
</tr>
<tr>
<td>Biochain</td>
<td>93-(31)-Bc-SER</td>
<td>Serous papillary cystadenocarcinoma</td>
<td></td>
<td>A503176</td>
</tr>
<tr>
<td>ABS</td>
<td>94-(2)-AB-SER</td>
<td>Papillary adenocarcinoma</td>
<td>UN</td>
<td>ILS-1408</td>
</tr>
<tr>
<td>ABS</td>
<td>95-(7)-AB-SER</td>
<td>adenocarcinoma</td>
<td>UN</td>
<td>IND-00375</td>
</tr>
<tr>
<td>Biochain</td>
<td>96-(8)-Bc-SER</td>
<td>adenocarcinoma</td>
<td></td>
<td>A501113</td>
</tr>
<tr>
<td>Biochain</td>
<td>97-(10)-Bc-SER</td>
<td>Adenocarcinoma</td>
<td></td>
<td>A407069</td>
</tr>
<tr>
<td>ABS</td>
<td>98-(1)-AB-SER</td>
<td>Papillary adenocarcinoma</td>
<td>UN</td>
<td>ILS-1406</td>
</tr>
<tr>
<td>GOG</td>
<td>99-(41)-GO-Ser Mix SIIC-OTR</td>
<td>Mixed epithelial cystadenocarcinoma with mucinous, endometrioid, squamous and papillary serous</td>
<td>Stage2</td>
<td>98-03-G803</td>
</tr>
<tr>
<td>GOG</td>
<td>100-(40)-GO-Ser Mix SIIC-OTR</td>
<td>Papillary serous and endometrioid cystadenocarcinoma</td>
<td>Stage3C</td>
<td>95-11-G006</td>
</tr>
<tr>
<td>GOG</td>
<td>101-(37)-GO-Ser Mix SIIC-OTR</td>
<td>Mixed serous and endometrioid adenocarcinoma</td>
<td>Stage3C</td>
<td>2002-05-G513</td>
</tr>
<tr>
<td>GOG</td>
<td>102-(38)-GO-Ser Mix SIIC-OTR</td>
<td>Mixed serous and endometrioid adenocarcinoma of mullerian</td>
<td>Stage3C</td>
<td>2002-05-G509</td>
</tr>
<tr>
<td>GOG</td>
<td>103-(34)-GO-Ser Mix SIIC-OTR</td>
<td>Papillary endometrioid adenocarcinoma</td>
<td>Stage3C</td>
<td>95-04-2002</td>
</tr>
<tr>
<td>Biochain</td>
<td>104-(19)-Bc-MUC</td>
<td>Mucinous adenocarcinoma</td>
<td></td>
<td>A504085</td>
</tr>
<tr>
<td>Biochain</td>
<td>105-(18)-Bc-MUC</td>
<td>Mucinous adenocarcinoma</td>
<td></td>
<td>A504083</td>
</tr>
<tr>
<td>Biochain</td>
<td>106-(17)-Bc-MUC</td>
<td>Mucinous adenocarcinoma</td>
<td></td>
<td>A504084</td>
</tr>
<tr>
<td>Biochain</td>
<td>107-(15)-Bc-Car-OTR</td>
<td>Carcinoma</td>
<td></td>
<td>A407065</td>
</tr>
<tr>
<td>Clontech</td>
<td>Tissue Samples in normal panel:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Carcinoma NOS</td>
<td>Source</td>
<td>Sample id</td>
<td></td>
</tr>
<tr>
<td>GOG</td>
<td>Clear cell adenocarcinoma</td>
<td>GCI</td>
<td>1090387</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stage 1A</td>
<td></td>
<td>2001-07-G084</td>
<td></td>
</tr>
<tr>
<td>GOG</td>
<td>Clear cell adenocarcinoma</td>
<td>GCI</td>
<td>2001-10-G002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stage 3A</td>
<td></td>
<td>2001-10-G002</td>
<td></td>
</tr>
<tr>
<td>ABS</td>
<td>Mucinous cystadenocarcinoma with low malignant</td>
<td>GCI</td>
<td>VNM-00187</td>
<td></td>
</tr>
<tr>
<td>GOG</td>
<td>Endometroid adenocarcinoma of borderline malignancy</td>
<td>Biochain</td>
<td>98-08-G001</td>
<td></td>
</tr>
</tbody>
</table>

### Table 1-2: Tissue samples in normal panel:

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Tissue</th>
<th>Source</th>
<th>Sample ID</th>
<th>Tissue</th>
<th>Source</th>
<th>Sample ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-(7)-Bc-Rectum</td>
<td>Rectum</td>
<td>Biochain</td>
<td>A610297</td>
<td>34-(78)-GC-Ovary</td>
<td>GCI</td>
<td>GWXUZN5M</td>
</tr>
<tr>
<td>2-(8)-Bc-Rectum</td>
<td>Rectum</td>
<td>Biochain</td>
<td>A610298</td>
<td>36-GC-cervix</td>
<td>GCI</td>
<td>E2P2N</td>
</tr>
<tr>
<td>3-GC-Colon</td>
<td>Colon</td>
<td>GCI</td>
<td>CDSUV</td>
<td>38-(26)-Bc-Uterus</td>
<td>Biochain</td>
<td>A504090</td>
</tr>
<tr>
<td>4-As-Colon</td>
<td>Colon</td>
<td>Asterand</td>
<td>31802</td>
<td>39-(30)-Am-Placenta</td>
<td>Ambion</td>
<td>021P33A</td>
</tr>
<tr>
<td>5-As-Colon</td>
<td>Colon</td>
<td>Asterand</td>
<td>74446</td>
<td>40-(32)-Bc-Placenta</td>
<td>Ambion</td>
<td>A411073</td>
</tr>
<tr>
<td>6-GC-Small bowel</td>
<td>Small bowel</td>
<td>GCI</td>
<td>V9L7D</td>
<td>41-GC-Breast</td>
<td>GCI</td>
<td>DHLR1</td>
</tr>
<tr>
<td>7-GC-Small bowel</td>
<td>Small bowel</td>
<td>GCI</td>
<td>M3GVT</td>
<td>42-GC-Breast</td>
<td>GCI</td>
<td>TG6J6</td>
</tr>
<tr>
<td>8-GC-Small bowel</td>
<td>Small bowel</td>
<td>GCI</td>
<td>196S2</td>
<td>43-GC-Breast</td>
<td>GCI</td>
<td>E6UDD</td>
</tr>
<tr>
<td>9-(9)-Am-Stomach</td>
<td>Stomach</td>
<td>Ambion</td>
<td>110P04A</td>
<td>44-(38)-Am-Prostate</td>
<td>Ambion</td>
<td>25955</td>
</tr>
<tr>
<td>11-(11)-Bc-Esoph</td>
<td>Esophagus</td>
<td>Biochain</td>
<td>A603814</td>
<td>45-Bc-Prostate</td>
<td>Biochain</td>
<td>A609258</td>
</tr>
<tr>
<td>12-(12)-Bc-Esoph</td>
<td>Esophagus</td>
<td>Biochain</td>
<td>A603813</td>
<td>46-As-Testis</td>
<td>Testis</td>
<td>Asterand</td>
</tr>
<tr>
<td>13-As-Panc</td>
<td>Panc</td>
<td>Asterand</td>
<td>9442</td>
<td>47-As-Testis</td>
<td>Testis</td>
<td>Asterand</td>
</tr>
<tr>
<td>14-As-Panc</td>
<td>Panc</td>
<td>Asterand</td>
<td>11134</td>
<td>49-GC-Artery</td>
<td>Artery</td>
<td>GCI</td>
</tr>
</tbody>
</table>

75
Real Time PCR analysis: materials and experimental procedures

RNA preparation - RNA was obtained from ABS (Wilmington, DE 19801, USA, absbioreagents.com), BioChain Inst. Inc. (Hayward, CA 94545 USA biochain.com), GOG for...
ovary samples- Pediatric Cooperative Human Tissue Network, Gynecologic Oncology Group Tissue Bank, Children Hospital of Columbus (Columbus OH 43205 USA), Clontech (Franklin Lakes, NJ USA 07417, clontech.com), Ambion (Austin, TX 78744 USA, ambion.com), Asternad (Detroit, MI 48202-3420, USA, asterand.com), and from Genomics Collaborative Inc., a Division of Seracare (Cambridge, MA 02139, USA, genomicsinc.com). Alternatively, RNA was generated from tissue samples using TRI-Reagent (Molecular Research Center), according to Manufacturer’s instructions. Tissue and RNA samples were obtained from patients or from postmortem. Total RNA samples were treated with DNasel (Ambion).

RT PCR - Purified RNA (10 µg) was mixed with 150 ng Random Hexamer primers (Invitrogen) and 500 µM dNTP in a total volume of 156µl. The mixture was incubated for 5 min at 65°C and then quickly chilled on ice. Thereafter, 50µl of 5X SuperscriptII first strand buffer (Invitrogen), 24µl 0.1M DTT and 400 units RNasin (Promega) were added, and the mixture was incubated for 10 min at 25°C, followed by further incubation at 42°C for 2 min. Then, 10µl (2000units) of SuperscriptII (Invitrogen) was added and the reaction (final volume of 250µl) was incubated for 50 min at 42°C and then inactivated at 70°C for 15min. The resulting cDNA was diluted 1:20 in TE buffer (10 mM Tris pH=8, 1 mM EDTA pH=8).

Real-Time RT-PCR analysis- cDNA (5µl) was prepared as described above, was used as a template in Real-Time PCR reactions using the SYBR Green I assay (PE Applied Biosystem) with specific primers and UNG Enzyme (Eurogentech or ABI or Roche). The amplification was effected as follows: 50°C for 2 min, 95°C for 10 min, and then 40 cycles of 95°C for 15sec, followed by 60°C for 1 min, following by dissociation step. Detection was performed by using the PE Applied Biosystem SDS 7000. The cycle in which the reactions achieved a threshold level (Ct) of fluorescence was registered and was used to calculate the relative transcript quantity in the RT reactions. Non-detected samples and samples that are out of the detectable range were assigned Ct value of 41 and were calculated accordingly. The relative quantity was calculated using the equation Q=efficiency×Λ-Ct. The efficiency of the PCR reaction was calculated from a standard curve, created by using serial dilutions of several reverse transcription (RT) reactions. To minimize inherent differences in the RT reaction, the resulting relative quantities were normalized to normalization factor calculated as follows: the expression of several housekeeping (HSKP) genes was checked on every panel. The relative quantity (Q) of each housekeeping gene in each sample, calculated as described above, was divided by the median quantity of this gene in all panel samples to obtain the "relative Q rel to MED". Then, for each sample the median of the "relative Q rel to MED" of the selected
housekeeping genes was calculated and served as normalization factor of this sample for further calculations. It should be noted that this type of analysis provides relative quantification.

The sequences of the housekeeping genes measured in all the examples on ovarian cancer panel were as follows: SDHA (GenBank Accession No. NM_004168 (SEQ ID NO: 15), SDHA Forward primer (SEQ ID NO: 16), SDHA Reverse primer (SEQ ID NO: 17), and SDHA-amplicon (SEQ ID NO.18); HPRT1 (GenBank Accession No. NM_000194 (SEQ ID NO: 19), HPRT1 Forward primer (SEQ ID NO:20), HPRT1 Reverse primer (SEQ ID NO:21); and HPRT1-amplicon (SEQ ID NO:22); G6PD (GenBank Accession No. NM_000402 (SEQ ID NO:23), G6PD Forward primer (SEQ ID NO:24), G6PD Reverse primer (SEQ ID NO:25), and G6PD-amplicon (SEQ ID NO:26).

The sequences of the housekeeping genes measured in all the examples on normal tissue samples panel were as follows: TATA box (GenBank Accession No. NM_003194 (SEQ ID NO:27)), TATA box Forward primer (SEQ ID NO:28), TATA box Reverse primer (SEQ ID NO:29), and TATA box -amplicon (SEQ ID NO:30); Ubiquitin (GenBank Accession No. BC000449 (SEQ ID NO:31)), Ubiquitin Forward primer (SEQ ID NO:32), Ubiquitin Reverse primer (SEQ ID NO:33), and Ubiquitin-amplicon (SEQ ID NO:34); SDHA (GenBank Accession No. NM_004168 (SEQ ID NO:15), SDHA Forward primer (SEQ ID NO:16), SDHA Reverse primer (SEQ ID NO:17), and SDHA-amplicon (SEQ ID NO:18).

Differential expression of the X63187_1_T9 (SEQ ID NO:1) transcript in ovarian cancer samples as compared to normal samples was measured using:

Forward Primer >X63187_seg7-ll(100-930)F (SEQ ID NO: 12):
CCACACCATGCTGCTGTC
Reverse Primer >X63187_seg7-ll(100-933)R (SEQ ID NO:13):
TCGGCCGCGCTGCTGTC
Amplicon >X63187_seg7-ll(100-930xl00-933) (SEQ ID NO:14):
CCACACCATGCTGCTGTCCTAGGCACAGCAGAGAAGACTGGCGTGCTGCCCCGAGCTCCAGGCTGACCAGAACTGCACGCAAGAGTGCGTCTCGGACAGCGAATGCGCCGACAACACTCAAGTGCTGCAGCAGCGGGCTGTGCCACCTTCTGCTCTCTGCCCAATGGTAACCCCACGGCGGCCGA
MED discovery engine

Another methodology used to predict the expression pattern of the proteins of the invention was MED discovery engine.

MED is a platform for collection of public gene-expression data, normalization, annotation and performance of various queries. Expression data from the most widely used Affymetrix microarrays is downloaded from the Gene Expression Omnibus (GEO - www.ncbi.nlm.nih.gov/GEO). Data is multiplicatively normalized by setting the 95 percentile to a constant value (normalized expression=1200), and noise is filtered by setting the lower 30% to 0. Experiments are annotated, first automatically, and then manually, to identify tissue and condition, and chips are grouped according to this annotation, and cross verification of this grouping by comparing the overall expression pattern of the genes of each chip to the overall average expression pattern of the genes in this group. Each probeset in each group is assigned an expression value which is the median of the expressions of that probeset in all chips included in the group. The vector of expression of all probesets within a certain group is the virtual chip of that group, and the collection of all such virtual chips is a virtual panel. The panel (or sub-panels) can be queried to identify probesets with a required behavior (e.g. specific expression in a sub-set of tissues, or differential expression between disease and healthy tissues). These probesets are linked to LEADS contigs and to RefSeqs (http://www.ncbi.nlm.nih.gov/RefSeq/) by probe-level mapping, for further analysis.

The Affymetrix platforms that are downloaded are HG-U95A and the HG-U1 33 family (A,B, A2.0 and PLUS 2.0). Than three virtual panels were created: U95 and U133 Plus 2.0, based on the corresponding platforms, and U133 which uses the set of common probe sets for HG-U133A, HG-U133A2.0 and HG-U133 PLUS 2.0+.

The results of the MED discovery engine are presented in scatter plots. The scatter plot is a compact representation of a given panel (collection of groups). The y-axis is the (normalized) expression and the x-axis describes the groups in the panel.

For each group, the median expression is represented by a marker, and the expression values of the different chips in the group are represented by small dashes ("-"). The groups are ordered and marked as follows - "Other" groups (e.g. benign, non-cancer diseases, etc.) with an "x", Treated cells with a black square, Normal with a circle, Matched with a "+", and Cancer with a diamond

Note - when the "+" sign falls on marking of expression of the chips, it will actually appear as a "-" sign, as the vertical part blends in with the marking of the expression. The number of chips in each group is written adjacent to it's name.
EXAMPLE 4: DESCRIPTION FOR CLUSTER X63187

According to at least some embodiments of the present invention there are novel variants of the known WAP four-disulfide core domain protein 2 precursor, shown in Cluster X63 187 herein.

Cluster X63187 (internal ID 97635836) features 1 transcript of interest, X63187_1_T9 (SEQ ID NO:1), encoding protein variant X63 187_1_P7 (SEQ ID NO:7).

Various oligonucleotides were tested for being differentially expressed in various disease conditions, particularly cancer, using a Microarray (chip) data analysis, as described in Example 3 herein above. The oligonucleotide HSHE4MR_0_0_10633 (SEQ ID NO:8), specific for X63187_1_T9 (SEQ ID NO:1) transcript but not for other segments/transcripts, was found to be overexpressed in ovarian carcinomas. The nucleic acid sequence of the oligonucleotide of HSHE4MR_0_0_10633 (SEQ ID NO:8) is given below:

TAGGAAAATGCCCCGCTCATCTAAGATGTGTAAGGGAGCATCGGTGAG

An alignment of the X63187_1_P7 (SEQ ID NO:7) HE4-variant protein to the known HE4 protein is given in Figures 2-3. A brief description of the relationship of the HE4-variant protein according to at least some embodiments of the present invention to such aligned proteins is as follows:

1. Comparison report between X63187_1_P7 (SEQ ID NO:7) and a known protein WFDC2_HUMAN (SEQ ID NO: 3) (Figure 2):
   A. An isolated chimeric polypeptide, comprising a first amino acid sequence being at least 90% homologous to the amino acids sequence:
   MPACRLGPLAAAALLLSLLFGFTLVSGTGAEKTCMELQADQNCTQECVSDSECAN
   LKCCSAAGCATFCSLPN corresponding to amino acids 1 - 74 of the known protein WFDC2_HUMAN (SEQ ID NO: 3), which also corresponds to amino acids 1 - 74 of X63187_1_P7 (SEQ ID NO:7), and a second amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence:
   GNPTAAERERGGAALGWEVGGPFGREQGRPRTRPGPGKSRLKPDPSVLSLARPRG
   RQRRR (SEQ ID NO:10) corresponding to amino acids 75 - 136 of X63187_1_P7 (SEQ ID NO:7), wherein said first amino acid sequence and second amino acid sequence are contiguous and in a sequential order.
2. Comparison report between X63187_1_P7 (SEQ ID NO:7) and known protein Q14508-4 (SEQ ID NO:2) (Figure 3):

A. An isolated chimeric polypeptide, comprising a first amino acid sequence being at least 90% homologous to the amino acid sequence:

```
MPACRLGPLAAALLLSLLFGFTLVSGTGAEKTCPELQADQNCTQECVSDSECAN
```

LKCCSAGCATFC corresponding to amino acids 1 - 70 of the known protein Q14508-4 (SEQ ID NO:2), which also corresponds to amino acids 1 - 70 of X63187_1_P7 (SEQ ID NO:7), abridging amino acids SL corresponding to amino acids 71-72 of X63187_1_P7 (SEQ ID NO:7), a second amino acid sequence PNG corresponding to amino acids 74 - 76 of known protein Q14508-4 (SEQ ID NO:2), which also corresponds to amino acids 73 - 75 of X63187_1_P7 (SEQ ID NO:7), and a third amino acid sequence being at least 70%, optionally at least 80%, preferably at least 85%, more preferably at least 90% and most preferably at least 95% homologous to a polypeptide having the sequence:

```
NPTAAERERGGAALGWEEVGGPGFREQGRPRTRGPPGKSRRLKPDPSVLSLARPRGQ
```

RRR (SEQ ID NO:11) corresponding to amino acids 76 - 136 of X63187_1_P7 (SEQ ID NO:7), wherein said first amino acid sequence, bridging amino acid, second amino acid sequence, and third amino acid sequence are contiguous and in a sequential order.

B. An isolated polypeptide encoding for an edge portion of X63 187_1_P7 (SEQ ID NO:7), comprising an amino acid sequence being at least 70%, optionally at least about 80%, preferably at least about 85%, more preferably at least about 90% and most preferably at least about 95% homologous to the sequence:

```
NPTAAERERGGAALGWEEVGGPGFREQGRPRTRGPPGKSRRLKPDPSVLSLARPRGQ
```

RRR (SEQ ID NO:11) of X63187_1_P7 (SEQ ID NO:7).

The localization of the HE4-variant protein was determined according to results from a number of different software programs and analyses, including analyses from SignalP and other specialized programs. The variant protein is believed to be secreted.

Variant protein X63187_1_P7 (SEQ ID NO:7) also has the following non-silent SNPs (Single Nucleotide Polymorphisms): A -> T at position 12 of SEQ ID NO:7; T -> P or T -> S at position 23 of SEQ ID NO:7.

The glycosylation site at position 44 of the known protein WAP four-disulfide core domain protein 2 is also present in a variant protein X63187JJP7 (SEQ ID NO:7). The coding portion of transcript X63187_1_T9 (SEQ ID NO:1) starts at position 281 and ends at position 688. The transcript has the following SNPs as listed in Table 2 (given according to their position on the nucleotide sequence, with the alternative nucleic acid listed).
Table 2 - Nucleic acid SNPs

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>SNP position(s) on nucleotide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A -&gt; G</td>
<td>11, 26, 1216</td>
</tr>
<tr>
<td>A -&gt; T</td>
<td>26, 347</td>
</tr>
<tr>
<td>C -&gt; G</td>
<td>152, 835, 1468, 1492, 1542, 1654</td>
</tr>
<tr>
<td>G -&gt; T</td>
<td>186, 1557</td>
</tr>
<tr>
<td>G -&gt; C</td>
<td>186, 1491</td>
</tr>
<tr>
<td>C -&gt; A</td>
<td>237, 667</td>
</tr>
<tr>
<td>C -&gt; T</td>
<td>255, 835, 960, 1401</td>
</tr>
<tr>
<td>G -&gt; A</td>
<td>314, 1508</td>
</tr>
<tr>
<td>A -&gt; C</td>
<td>347</td>
</tr>
<tr>
<td>T -&gt; C</td>
<td>735, 737</td>
</tr>
<tr>
<td>G -&gt;</td>
<td>1491</td>
</tr>
<tr>
<td>C -&gt;</td>
<td>1492</td>
</tr>
<tr>
<td>T -&gt;</td>
<td>1628</td>
</tr>
</tbody>
</table>

EXAMPLE 5: ANALYSIS OF THE EXPRESSION OF X63187 TRANSCRIPTS

MED discovery engine described in Example 3 hereinabove was used to assess the expression of X63187 transcripts. X63187 transcripts were found to be over-expressed in ovarian cancer, and kidney cancer, as is demonstrated in Figures 4 and 5, respectively. Figures 4 and 5 show expression graphs of Affymetrix probe set 203892_at. Figure 4 shows the expression of X63187 transcripts in microarray chips from ovarian cancer and ovary normal tissue samples. Table 3 below shows the tissue samples examined and the corresponding number in the X axis of Figure 4. As can be seen X63187 transcripts are overexpressed in ovarian cancer tissues (diamond markers) relative to its expression in normal ovary (open square markers). Figure 5 shows the expression of X63187 transcripts in microarray chips from kidney cancereous and kidney normal tissues. Table 4 below shows the tissue samples examined and the corresponding number in the X axis of Figure 5 As can be seen, X63187 transcripts are overexpressed in kidney cancer tissues (diamond markers) relative to its expression in normal kidney (open square markers).

For each group, the median expression is represented by a marker, and the expression values of the different chips in the group are represented by small dashes ("-".). The groups are ordered and marked as follows - "Other" groups (e.g. benign, non-cancer diseases, etc.) - with
an "x"; Treated cells - celles treated as specified in the figure legend - with a black square; Normal with an open square; Matched (cancer and normal tissue from the same patient) with a "+"; and Cancer with a diamond.

Table 3: Tissue samples presented in Figure 4

<table>
<thead>
<tr>
<th>Tissue Number</th>
<th>Tissue Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ovary_benign.1 HU133 3</td>
</tr>
<tr>
<td>2</td>
<td>ovary_theca_cells_primary_culture_pcos HU133 5</td>
</tr>
<tr>
<td>3</td>
<td>ovary_benign HU133 2.0+ 3</td>
</tr>
<tr>
<td>4</td>
<td>ovary_theca_cells_primary_culture_treated_valproic_acid HU133 4</td>
</tr>
<tr>
<td>5</td>
<td>ovary_normal.1 HU133 3</td>
</tr>
<tr>
<td>6</td>
<td>ovary_normal.2 HU133 14</td>
</tr>
<tr>
<td>7</td>
<td>ovary_pcos_normal HU133 12</td>
</tr>
<tr>
<td>8</td>
<td>ovary_theca_cells_primary_culture_normal HU133 4</td>
</tr>
<tr>
<td>9</td>
<td>ovary_normal HU133 2.0+ 9</td>
</tr>
<tr>
<td>10</td>
<td>ovary_pcos_normal HU133 2.0+ 12</td>
</tr>
<tr>
<td>11</td>
<td>ovary_adenocarcinoma_cancer HU133 19</td>
</tr>
<tr>
<td>12</td>
<td>ovary_adenocarcinoma_metastatic_cancer HU133 26</td>
</tr>
<tr>
<td>13</td>
<td>ovary_cancer.1 HU133 149</td>
</tr>
<tr>
<td>14</td>
<td>ovary_cancer.2 HU133 4</td>
</tr>
<tr>
<td>15</td>
<td>ovary_carcinoma_cancer HU133 6</td>
</tr>
<tr>
<td>16</td>
<td>ovary_clear_cell_adenocarcinoma_cancer HU133 9</td>
</tr>
<tr>
<td>17</td>
<td>ovary_clear_cell_adenocarcinoma_cancer HU133 3</td>
</tr>
<tr>
<td>18</td>
<td>ovary_clear_cell_cancer HU133 6</td>
</tr>
<tr>
<td>19</td>
<td>ovary_cystadenocarcinoma_cancer HU133 3</td>
</tr>
<tr>
<td>20</td>
<td>ovary_endometrioid_adenocarcinoma_cancer HU133 40</td>
</tr>
<tr>
<td>21</td>
<td>ovary_endometrioid_cancer HU133 13</td>
</tr>
<tr>
<td>22</td>
<td>ovary_endometrioid_carcinoma_cancer HU133 8</td>
</tr>
<tr>
<td>23</td>
<td>ovary_metastatic_carcinoma_cancer HU133 6</td>
</tr>
<tr>
<td>24</td>
<td>ovary_metastatic_papillary_serous_adenocarcinoma_metastatic_cancer HU133 3</td>
</tr>
<tr>
<td>25</td>
<td>ovary_mucinous_adenocarcinoma_cancer HU133 15</td>
</tr>
<tr>
<td>26</td>
<td>ovary_mucinous_cancer HU133 3</td>
</tr>
<tr>
<td>27</td>
<td>ovary_papillary_serous_adenocarcinoma_cancer HU133 7</td>
</tr>
</tbody>
</table>
Table 4: Tissue samples presented in Figure 5

<table>
<thead>
<tr>
<th>Tissue Number</th>
<th>Tissue Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td>ovary_papillary_serous_carcinoma_cancer HU133 5</td>
</tr>
<tr>
<td>29</td>
<td>ovary_papillary_serous_carcinoma_metastatic_cancer HU133 6</td>
</tr>
<tr>
<td>30</td>
<td>ovary_papillary_serous_tumor_cancer HU133 3</td>
</tr>
<tr>
<td>31</td>
<td>ovary_serous_cancer.1 HU133 81</td>
</tr>
<tr>
<td>32</td>
<td>ovary_serous_metastatic_cancer HU133 23</td>
</tr>
<tr>
<td>33</td>
<td>ovary_adenocarcinoma_cancer HU133_2.0+ 18</td>
</tr>
<tr>
<td>34</td>
<td>ovary_adenocarcinoma_metastatic_cancer HU133_2.0+ 25</td>
</tr>
<tr>
<td>35</td>
<td>ovary_cancer HU133_2.0+ 6</td>
</tr>
<tr>
<td>36</td>
<td>ovary_carcinoma_cancer HU133_2.0+ 6</td>
</tr>
<tr>
<td>37</td>
<td>ovary_clear_cell_adenocarcinoma_cancer HU133_2.0+ 3</td>
</tr>
<tr>
<td>38</td>
<td>ovary_clear_cell_cancer HU133_2.0+ 6</td>
</tr>
<tr>
<td>39</td>
<td>ovary_cystadenocarcinoma_cancer HU133_2.0+ 3</td>
</tr>
<tr>
<td>40</td>
<td>ovary_endometrioid_adenocarcinoma_cancer HU133_2.0+ 3</td>
</tr>
<tr>
<td>41</td>
<td>ovary_endometrioid_cancer HU133_2.0+ 13</td>
</tr>
<tr>
<td>42</td>
<td>ovary_endometrioid_carcinoma_cancer HU133_2.0+ 8</td>
</tr>
<tr>
<td>43</td>
<td>ovary_metastatic_cancer HU133_2.0+ 6</td>
</tr>
<tr>
<td>44</td>
<td>ovary_metastatic_papillary_serous_adenocarcinoma_metastatic_cancer HU133_2.0+ 3</td>
</tr>
<tr>
<td>45</td>
<td>ovary_papillary_serous_carcinoma_cancer HU133_2.0+ 7</td>
</tr>
<tr>
<td>46</td>
<td>ovary_papillary_serous_carcinoma_cancer HU133_2.0+ 5</td>
</tr>
<tr>
<td>47</td>
<td>ovary_papillary_serous_carcinoma_metastatic_cancer HU133_2.0+ 6</td>
</tr>
<tr>
<td>48</td>
<td>ovary_papillary_serous_tumor_cancer HU133_2.0+ 3</td>
</tr>
<tr>
<td>49</td>
<td>ovary_serous_cancer.1 HU133_2.0+ 39</td>
</tr>
<tr>
<td>50</td>
<td>ovary_serous_metastatic_cancer HU133_2.0+ 23</td>
</tr>
</tbody>
</table>
Expression of HE4 variant transcripts which are detectable by amplicon X63187_seg7-11 (100-930x100-933) (SEQ ID NO:14) in normal and cancerous ovary tissues

Expression of WFDC2 transcripts detectable by or according to X63187_seg7-ll (100-930x100-933) amplicon (SEQ ID NO:14) and primers X63187_seg7-ll (100-930)F (SEQ ID NO:12) and X63187_seg7-l (100-933)R (SEQ ID NO:13) was measured by real time PCR. Non-detected samples (samples no. 81, 83, 107, 110, 41, 44, 45, 53, 54, 57, 62, 64, 65, 67, 69,
71, 72, 74, 75 and 76, Table 1_1) were assigned Ct value of 41 and were calculated accordingly. In parallel the expression of several housekeeping genes - SDHA (GenBank Accession No. NM_004168 (SEQ ID NO:15); amplicon - SDHA-amplicon (SEQ ID NO:18), HPRT1 (GenBank Accession No. NM_000194 (SEQ ID NO:19); amplicon - HPRT1-amplicon (SEQ ID NO:22)), G6PD (GenBank Accession No. NM_000402 (SEQ ID NO:23); amplicon - G6PD-amplicon (SEQ ID NO:26) were measured similarly. For each RT sample, the expression of the above amplicon was normalized to the normalization factor calculated from the expression of these house keeping genes as described in Example 6. The normalized quantity of each RT sample was then divided by the median of the quantities of the normal samples (sample numbers 53, 54, 57, 62, 64, 65, 67, 69, 71, 72, 74, 75 and 76, Table 1_1), to obtain a value of fold up-regulation for each sample relative to median of the normal samples.

Figure 6 is a histogram showing over expression of the above-indicated WFDC2 transcripts in cancerous ovary samples relative to the normal samples.

As is evident from Figure 6, the expression of WFDC2 transcripts detectable by the above amplicon in serous carcinoma, mucinous carcinoma and endometroid carcinoma samples was significantly higher than in the non-cancerous samples (sample numbers 53, 54, 57, 62, 64, 65, 67, 69, 71, 72, 74, 75 and 76, Table 1_1). Notably an over-expression of at least 5 fold was found in 36 out of 38 serous carcinoma samples, in 12 out of 12 mucinous carcinoma samples and in 10 out of 10 endometroid carcinoma samples.

Statistical analysis was applied to verify the significance of these results, as described below.

The P value for the difference in the expression levels of WFDC2 transcripts detectable by the above amplicon in ovary serous carcinoma samples versus the normal tissue samples was determined by T test as 3.02e-005. The P value for the difference in the expression levels of WFDC2 transcripts detectable by the above amplicon in ovary mucinous carcinoma samples versus the normal tissue samples was determined by T test as 9.6e-003. The P value for the difference in the expression levels of WFDC2 transcripts detectable by the above amplicon in ovary endometroid carcinoma samples versus the normal tissue samples was determined by T test as 7.04e-003. The P value for the difference in the expression levels of WFDC2 transcripts detectable by the above amplicon in ovary adenocarcinoma samples versus the normal tissue samples was determined by T test as 1.30e-008.

Threshold of 5 fold over expression was found to differentiate between serous carcinoma and normal samples with P value of 7.08e-009 as checked by exact Fisher test. Threshold of 5 fold over expression was found to differentiate between mucinous carcinoma
and normal samples with P value of 2.50e-006 as checked by exact Fisher test. Threshold of 5 fold over expression was found to differentiate between endometrioid carcinoma and normal samples with P value of 9.61e-006 as checked by exact Fisher test. Threshold of 5 fold over expression was found to differentiate between adenocarcinoma and normal samples with P value of 3.16e-010 as checked by exact Fisher test.

The above values demonstrate statistical significance of the results.

Primer pairs are also encompassed within the present invention; for example, for the above experiment, the following primer pair was used as a non-limiting illustrative example only of a suitable primer pair: X63187_seg7-ll(100-930)F (SEQ ID NO:12) forward primer; and X63187_seg7-ll(100-933)R (SEQ ID NO:13) reverse primer.

According to at least some embodiments, the present invention also encompasses any amplicon obtained through the use of any suitable primer pair; for example, for the above experiment, the following amplicon was obtained as a non-limiting illustrative example only of a suitable amplicon: X63187_seg7-ll (100-930x100-933) (SEQ ID NO:14).

Expression of HE4-variant transcripts which are detectable by amplicon as depicted in sequence name x63187_seg7-ll(100-930x100-933) (seq id no:14) in different normal tissues

Expression of WFDC2 transcripts detectable by or according to X63187_seg7-ll(100-930x100-933) amplicon (SEQ ID NO:14) and primers X63187_seg7-ll (100-930)F (SEQ ID NO:12) and X63187_seg7-ll (100-933)R (SEQ ID NO:13) was measured by real time PCR. Non-detected samples (samples no. 14, 32, 42, 50, 56, 62 and 66, Table 1_2) were assigned Ct value of 41 and were calculated accordingly. In parallel, the expression of several housekeeping genes - SDHA (GenBank Accession No. NM_004168 (SEQ ID NO:15); amplicon - SDHA-amplicon (SEQ ID NO:18), Ubiquitin (GenBank Accession No. BC000449 (SEQ ID NO:31); amplicon - Ubiquitin-amplicon (SEQ ID NO:34)) and TATA box (GenBank Accession No. NM_003194 (SEQ ID NO:27); TATA amplicon (SEQ ID NO:30) was measured similarly. For each RT sample, the expression of the above amplicon was normalized to the normalization factor calculated from the expression of these housekeeping genes as described in Example 6. The normalized quantity of each RT sample was then divided by the median of the quantities of the ovary samples (sample number 32, Table 1_2), to obtain a value of relative expression of each sample relative to median of the ovary samples.
Figure 7 is a histogram showing over-expression of the WFDC2 transcripts detectable by amplicon as depicted in sequence name X63187_seg7-1 (100-930x100-933) (SEQ ID NO: 14) in different normal tissues.

It should be noted that for X63187_seg7-II (100-930x100-933) (SEQ ID NO:14), no differential expression was observed in one Q-PCR experiment carried out with cancerous and normal breast and lung panels.

**Evaluation of the expression of the known HE4 wild type (WT) and the HE4 variant X63187_1_T9 (SEQ ID NO:1) in a panel of normal human tissues**

Using exon level expression data ((Affymetrix Exon Array ST 1 chip, downloaded from Affymetrix web site: http://www.affymetrix.com/support/help/faqs/exon_array_analysis/faq_8.jsp) for a panel of normal tissues, it was demonstrated that the expression of the known HE4 WT is much higher (fold range between 4 to >20) than the expression of the variant X63187J._T9 (SEQ ID NO:1) encoding the protein X63187_1_P7 (SEQ ID NO:7). The results are presented in Figure 11.

More over, as shown in Figure 11, the expression of the variant X63187_1_T9 (SEQ ID NO:1) encoding the protein X63187_1_P7 (SEQ ID NO:7) has the same magnitude as a probe derived from an unexpressed intron of the genes, suggesting that the variant X63187_1_T9 (SEQ ID NO:1) encoding the protein X63187_1_P7 (SEQ ID NO:7) is very poorly expressed in normal tissues.

The expression of the known HE4 wild type (WT) was demonstrated using the Affymetrix probe 3886961; the expression of the variant X63187_1_T9 (SEQ ID NO:1) encoding the protein X63187_1_P7 (SEQ ID NO:7) (SV), was demonstrated using the Affymetrix probe 3886944; and the expression of the sequence corresponding to an unexpressed intron was demonstrated using the Affymetrix probe 3886952. Each data point represents an average of three (3) samples (apart from the tissue mix which represents the average of 20 samples), error bars represent standard deviation. Y axis is an arbitrary fluorescent read units.

Thus, using the variant X63187_1_T9 (SEQ ID NO:1) and the corresponding protein X63187_1JP7 (SEQ ID NO:7) as diagnostic markers has an advantage as compared to the known HE4, as it has the potential to reduce significantly the rate of false positives detection.
EXAMPLE 6: PCR ANALYSIS OF A NUCLEIC ACID SEQUENCE ENCODING
HE4-VARIANT POLYPEPTIDE WFDC2_T12_P7

To verify the foil length ORF sequence of the WFDC2_T12_P7 the following analysis
was carried out:

A reverse transcription reaction was carried out as follows: \(10\mu\text{g}\) of purified ovary
cancer RNA was mixed with \(150\text{ng}\) Random Hexamer primers (Invitrogen, Carlsbad, CA,
USA, catalog number: 48190-01 i) and \(500\mu\text{M}\) dNTPs in a total volume of \(156\mu\text{l}\). The
mixture was incubated for 5 min at 65°C and then quickly chilled on ice. Thereafter, \(50\mu\text{l}\) of
5X SuperscriptII first strand buffer (Invitrogen, catalog number: 18064-014, part number:
Y00146), \(24\mu\text{l}\) 0.1M DTT and 400 units RNasin (Promega, Milwaukee, WS, U.S.A., catalog
number: N251 i) were added, and the mixture was incubated for 10 min at 25°C, followed by
further incubation at 42°C for 2 min. Then, \(10\mu\text{l}\) (2000 units) of SuperscriptII (Invitrogen,
catalog number: 18064-014) was added and the reaction (final volume of \(250\mu\text{l}\)) was
incubated for 50 min at 42°C and then inactivated at 70°C for 15min. The resulting cDNA
was diluted 1:20 in TE buffer (10mM Tris, 1 mM EDTA pH 8). PCR was done using GoTaq
ReadyMix (Promega, catalog number M122) under the following conditions: \(10.5\mu\text{l}\) cDNA
from the above; \(1\mu\text{l}\) of each primer (\(100\mu\text{M}\)) and \(12.5 \mu\text{l}\) ReadyMix in a total reaction volume of
\(25\mu\text{l}\); with a reaction program of 5 minutes in 95°C; 35 cycles of: 30 seconds at 94°C, 30
seconds at 55°C, 60 seconds at 72°C; then 10 minutes at 72°C. Primers which were used
include gene specific sequences delimiting the ORF coordinates of the protein. Forward
primer 100-930 WAP_n7_For2 (SEQ ID NO: 12) and a reverse primer 100-936
WAP_nl2_Rev1 (SEQ ID NO: 35).

\(25\mu\text{l}\) of the PCR product described above were loaded onto a 1% agarose gel stained
with ethidium bromide, electrophoresed in \(1\times\)TBE solution at 100V, and visualized with UV
light. The results are shown in Figure 8, demonstrating an expected band size of 462bp,
corresponding to a nucleic acid sequence encoded by WFDC2_T12_P7. The PCR product was
then excised, extracted from the gel using QiaQuick™ Gel Extraction kit (Qiagen, catalog
number: 28707), and sequenced. The DNA sequence of the resulting PCR product is shown in
Figure 9 and in SEQ ID NO: 36. Sequence corresponding to the open reading frame (ORF) is
marked in bold, while the primers sequences used are underlined. The amino acid sequence of
WFDC2_T12_P7 encoded by the resulting nucleotide sequence above is shown in Figure 10
(SEQ ID NO:7).
EXAMPLE 7

1. GENERATION OF ANTIBODY LIBRARY

The gene coding for WFDC2_T12_P7 was cloned into a mammalian expression vector and the construct was transfected into cells. Following isolation of a stable recombinant cell line, the WFDC2 _P7 protein was expressed and purified. An aliquot of this protein was biotinylated for library selection and screening of HE4-variant specific antibodies.

Next, the purified protein was used to immunize mice. After the mice develop titers, they are sacrificed and the RNA is purified from spleens. Phage display antibody libraries is made and selected with a biotinylated construct of the same protein, as described in U.S. Patent No. 6,057,098. This phage library is then subjected to sequential selection with biotinylated WFDC2 _P7 protein. Finally, the enriched phage library is cloned into a plasmid expression vector and electroporated into E. coli to generate an Omiclonal® antibody library.

π. ANTIBODY SCREENING TO CONSTRUCT SANDWICH ASSAYS FOR WFDC2_P7 AND WFDC2 _ WT (HE4 WT)

The cells representing the antibody library are streaked on agar plates to generate colonies coding for monoclonal antibodies. Serial cultures are inoculated to express and purify the monoclonal antibodies in microgram quantities. These antibodies are tested by indirect ELISA using biotinylated WFDC2 JP7 protein coated in assay plates to identify the best binders and then are tested on the wild type WFDC2 to exclude antibodies which bind to the WT WFDC2 as well. Similar methods are used to screen antibodies from another antibody library that was previously raised with wild type WFDC2 protein.

The foregoing HE4-variant specific antibodies are scaled up to generate milligram quantities. Aliquots of these antibodies are conjugated to either biotin or fluorscein. In sandwich ELISAs, the biotinylated antibodies are coated in Neutravidin plates as capture antibodies while the fluorsceininated antibodies are used as detection antibodies, using an anti-fluorscein antibody conjugated to Alkaline Phosphatase as a secondary label to generate sandwich assay signals. Serum sample pools are generated using aliquots of multiple clinical serum samples representing normal donors and ovarian cancer patients. The levels of the WFDC2 _P7 and WFDC2 _ WT proteins are measured in the serum samples using the ELISA, as described above.

It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single
embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable subcombination.

Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims. AU publications, patents and patent applications mentioned in this specification are herein incorporated in their entirety by reference into the specification, to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated herein by reference. In addition, citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the present invention.
WHAT IS CLAIMED IS:

1. An isolated polypeptide comprising an amino acid sequence as set forth in SEQ ID NO: 10, a homologue or a fragment thereof.

2. The isolated polypeptide of claim 1, comprising amino acids 2-62 of SEQ ID NO:10.

3. The isolated polypeptide of claim 1, comprising an amino acid sequence at least about 85%, 90% or 95% homologous to SEQ ID NO: 10.

4. The isolated polypeptide of claim 1 comprising an amino acid sequence as set forth in SEQ ID NO:7, a homologue or a fragment thereof.

5. The isolated polypeptide of claim 4, comprising an amino acid sequence at least about 85%, 90% or 95% homologous to SEQ ID NO:7.

6. The isolated polypeptide of claim 4, consisting of the amino acid sequence set forth in SEQ ID NO:7.

7. An isolated polynucleotide comprising a nucleic acid sequence as set forth in SEQ ID NO:37, a homologue or a fragment thereof.

8. The isolated polynucleotide of claim 7, comprising a nucleic acid sequence at least about 85%, 90% or 95% homologous to SEQ ID NO:37.

9. The isolated polynucleotide of claim 7, comprising a nucleic acid sequence as set forth in SEQ ID NO:1, or a fragment thereof.

10. An isolated polynucleotide of from 50 to 650 nucleotides, comprising an amplicon as set forth in SEQ ID NO:14, or a sequence at least about 85%, 90% or 95% homologous thereto, or a fragment thereof.

11. An isolated oligonucleotide of at least 10 nucleotides but not more than 200 nucleotides that hybridizes at high stringency to at least one polynucleotide having a nucleic acids sequence selected from the group consisting of SEQ ID NOs:37, 14, and 1, a fragment or a complement thereof.

12. An isolated primer pair comprising a pair of nucleic acid sequences of from at least 10 nucleotides and no more than 200 nucleotide that specifically amplify at least one polynucleotide having a nucleic acids sequence selected from the group consisting of SEQ ID NOs:37, 14, and 1.
13. The primer pair of claim 12, comprising a pair of isolated polynucleotides having a
sequences selected from the group consisting of SEQ ID NOs: 12 and 13; SEQ ID
NOs: 12 and 35; SEQ ID NOs: 38 and 13; and SEQ ID NOs: 38 and 35.

14. The oligonucleotide of claim 11, having the nucleic acid sequence set forth in SEQ ID
NO: 8.

15. An antibody or a fragment thereof capable of specifically binding to at least one
epitope of a polypeptide having the amino acid sequence selected from the group
consisting of SEQ ID NOs: 10, H, 7, or a fragments or a homolog thereof, wherein
said antibody does not substantially bind known HE4 proteins having the amino acid
sequence set forth in any one of SEQ ID NOs: 2-6.

16. A kit for detecting cancer, comprising at least one isolated oligonucleotide of at least
10 nucleotides but not more than 200 nucleotides, wherein said oligonucleotide is
capable of specifically detecting at least one HE4-variant polynucleotide having a
nucleic acid sequence selected from the group consisting of SEQ ID NOs: 37, 14, and
1, a fragment or a complement thereof.

17. The kit of claim 16, wherein said oligonucleotide has a nucleic acid sequence as set
forth in SEQ ID NO: 8.

18. The kit of claim 16, comprising a pair of isolated oligonucleotides.

19. The kit of claim 18, wherein the pair of isolated oligonucleotide comprises the nucleic
acid sequence selected from the group consisting of SEQ ID NOs: 12 and 13; SEQ ID
NOs: 12 and 35; SEQ ID NOs: 38 and 13; and SEQ ID NOs: 38 and 35.

20. The kit according to claim 16, wherein the cancer is selected from the group consisting
of ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma,
endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and
urothelial cell carcinoma, and wherein the cancer is selected from invasive, non
metastatic or metastatic cancer.

21. The kit of claim 16, wherein said kit further comprises at least one reagent for
performing a NAT (nucleic acid testing)-based assay.

22. The kit of claim 21, wherein the NAT-based assay is selected from the group
consisting of a PCR, Real-Time PCR, LCR, Self-Sustained Synthetic Reaction, Q-
Beta Replicase, Cycling Probe Reaction, Branched DNA, RFLP analysis,
DGGE/TGGE, Single-Strand Conformation Polymorphism, Dideoxy fingerprinting, Microarrays, Fluorescence In Situ Hybridization and Comparative Genomic Hybridization.

23. A kit for diagnosing cancer, comprising an antibody or an antibody fragment according to claim 15.

24. The kit of claim 23, wherein said kit further comprises at least one reagent for performing an immunoassay.

25. The kit of claim 24, wherein said immunoassay is selected from the group consisting of immunohistochemical assay, radioimaging assays, in-vivo imaging, positron emission tomography (PET), single photon emission computer tomography (SPECT), magnetic resonance imaging (MRI), optical imaging, computer tomography, radioimmunoassay (RIA), ELISA, slot blot, competitive binding assays, fluorimetric imaging assays, western blot, FACS.

26. A method for diagnosing cancer in a subject, comprising detecting in the subject or in a sample obtained from said subject the level of at least one polypeptide having an amino acid sequence set forth in any one of SEQ ID Nos: 10, 11, 7, a homologue or a fragment thereof compared to a control baseline, wherein an increase in the level of the at least one polypeptide is indicative of cancer.

27. The method of claim 26, wherein the detection is conducted by immunoassay.

28. The method of claim 27, wherein the immunoassay utilizes an antibody according to claim 15.

29. A method for diagnosing cancer in a subject, comprising detecting in the subject or in a sample obtained from said subject a polynucleotide having a nucleic acid sequence as set forth in any one of SEQ ID Nos: 37, 14, 1, a homologue or a fragment thereof compared to a control baseline, wherein an increase in the level of the at least one polynucleotide is indicative of cancer.

30. The method according to any one of claims 26 and 29, wherein the control baseline is selected from the group consisting of the level of the polypeptide or polynucleotide in a healthy subject and the level of the polypeptide or polynucleotide in same subject at an earlier time point.
31. The method according to any one of claims 26 and 29, wherein diagnosing cancer comprises any one of screening for cancer, diagnosing cancer, detecting the presence or a severity of cancer, prognosis of cancer, early diagnosis of cancer, staging of cancer, monitoring of cancer progression, monitoring of cancer treatment efficacy, monitoring of cancer relapse, selecting a therapy or a treatment for cancer, optimization of a given therapy for cancer, predicting the suitability of a therapy for specific patients or subpopulations and determining the appropriate dosing of a therapeutic product in patients or subpopulations.

32. The method according to any one of claims 26 and 29 wherein the cancer is selected from the group consisting of ovarian cancer, renal cancer, lung adenocarcinoma, breast adenocarcinoma, endometrial cancer, gastric cancer, lung mesothelial tumor, pancreatic carcinoma, and/or urothelial cell carcinoma, and wherein the cancer is invasive, non metastatic or metastatic.

33. The method of claim 29, wherein the detection is performed using at least one oligonucleotide of at least 10 nucleotides but not more than 200 nucleotides that hybridizes at high stringency to at least one polynucleotide having the nucleic acid sequence selected from the group consisting of SEQ ID NOs: 37, 14, and 1, a fragment or a complement thereof.

34. The method of claim 33, wherein the oligonucleotide has the sequence set forth in SEQ ID NO.8.

35. The method of claim 29, wherein the detection is performed using at least one primer pair comprising a pair of nucleic acid sequences of at least 10 nucleotides but not more than 200 nucleotides that specifically detect at least one polynucleotide having a nucleic acids sequence selected from the group consisting of SEQ ID NOs: 37, 14, and 1.

36. The method of claim 35, wherein the pair comprises a pair of isolated polynucleotides having the sequence set forth in any one of SEQ ID NOs: 12 and 13; SEQ ID NOs: 12 and 35; SEQ ID NOs: 38 and 13; and SEQ ID NOs: 38 and 35.

37. An expression vector containing a nucleic acid sequence according to claim 7.

38. A host cell comprising the expression vector of claim 37.
OVER EXPRESSION RELATIVE TO MEDIAN OF OVARY SAMPLES
FIG. 8

CCACACCATGCGCTGCTCTGCGCTAGGCCGCCCTAGCCGCCGCCCTCCTCTCAGCCCTGCTGCT
GTTCGGCTTCAACCTAGTCTCAGCCACAGGAGCAGGAAGAAGACTGGCGTGTGCCCCGAGCT
CCAGGCTGACCAAGAACTGCAACGCAAGAGTCTGCTCAGCAAGCGGAAATGCGCCGACAAACCT
CAAGTGCTGCAGGGCGGCCTGTGCAGACCTTCTGCTCTCTGTCCAAATGTAACCCCACCGCC
GGCCGAGCGGAACGGGCGGCGGCCGCGCTGGGCTGGGAGAGTGCGAGGAGGAGGAGGCGGCGGT
CCGGGAAAGCGGGCGCCCGCGGAGGACCGGGGAAAGGTCAGCGCGGTGTTGAAACCC
AGATGCTGCTAGTCCTCTCCTGCGACGCGCCAGGGGCTAGACAAAGGCCGCTGTTGAAACGCG
AGCCAAGGGGGGTGCCACCGGGCGCTAGCTGGGATTCGACT

FIG. 9

MPACRLGPLAAALLSLLFLGFTLSGVTGAETGVCPELQADQNCTQECVSDECAADNLK
CCSAGCATFCSLPENGPTAAERERGGAALGWEVEVGGPFGFREQGRPRTRGPPGKSSRLKPD
PSVSLARPRGRQRRR

FIG. 10
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

INV. C07K14/47 C12N15/11 G01N33/48 G01N33/50 G01N33/53
C12Q1/68 C12N5/10 C12N1/15 C12N1/21 C07K16/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K C12Q G01N C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, BIOSIS, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
</table>

[X] Further documents are listed in the continuation of Box C  
[X] See patent family annex

Special categories of cited documents

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referred to in an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

Date of the actual completion of the international search
12 March 2010

Date of mailing of the international search report
30/03/2010

Name and mailing address of the ISA/Authorized officer
European Patent Office, P B 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel (+31-70) 340-2040, Fax (+31-70) 340-3016
Kools, Patrick

Form PCT/ISA/210 (second sheet) (April 2005)
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>DATABASE EMBL [Online], 7 May 2008 (2008-05-07), &quot;Human DNA sequence from clone RP3-461P17 on chromosome 20q12-13.2 Contains two novel genes, the WFDC2 gene for WAP four-disulfide core domain 2, the RPL5P2 gene for ribosomal protein L5 pseudogene 2, a COX6C (Cytochrome C Oxidase subunit VIC) pseudogene, a pseudogene similar to part of heat shock 60&quot; XP002572896 retrieved from EBI accession no. EMBL:AL031663 Database accession no. AL031663 only relevant sequence part is provided</td>
<td>7-9, 37-38</td>
</tr>
<tr>
<td>X</td>
<td>DATABASE EMBL [Online], 24 March 1999 (1999-03-24), &quot;th60e02.xl NCI_CGAP_0v23 Homo sapiens cDNA clone IMAGE: 2122682 3', mRNA sequence.&quot; XP002572897 retrieved from EBI accession no. EMBL:AI521428 Database accession no. AI521428 the whole document</td>
<td>11</td>
</tr>
<tr>
<td>Category</td>
<td>Citation of document, with indication, where appropriate, of the relevant passages</td>
<td>Relevant to claim No</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>WO 2008112514 A1 18-09-2008</td>
<td>AU 2008226587 A1 18-09-2008</td>
<td></td>
</tr>
<tr>
<td>CA 2679696 A1 18-09-2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US 2008254048 A1 16-10-2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WO 2008121391 A2 09-10-2008</td>
<td>CA 2682132 A1 09-10-2008</td>
<td></td>
</tr>
<tr>
<td>EP 2126577 A1 02-12-2009</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>