KITS AND METHODS FOR SELECTING A TREATMENT FOR OVARIAN CANCER

The present invention relates to methods for selecting a treatment for a subject having ovarian cancer, in particular methods to distinguish between ovarian cancer patients who will respond to the taxol/platinum chemotherapy and survive longer than seven years versus those who will succumb to the disease within three years.

Figure 1

Ovarian Cancer Prognosis (F1) — Discovery Study

- Sensitivity = 0.950
- Specificity = 1.000
- AUC = 0.989

Bars represent standard deviations (SD)
* Significance level: \( p < 0.001 \) (two-tailed)

\( F_1 \) is parametrically distributed with respect to both groups
Designated States (unless otherwise indicated, for every kind of regional protection available): AIPPI (BW, GB, GM, KE, LB, LS, MW, NZ, NA, NW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LV, MC, MT, NL, NO, PL, PT, RO, SK, SI, SE, SL, SK, SM, TE), OAPI (BF, BJ, CP, CG, CI, CM, GA, GN, GQ), GW, ME, MR, NE, SN, TD, TG).

Published:
— with international search report (Art. 21(3))
— with sequence listing part of description (Rule 5.2(a))
KITS AND METHODS FOR SELECTING A TREATMENT FOR OVARIAN CANCER

This application is being filed on 02 May 2012, as a PCT International Patent application in the name of Applied Informatic Solutions, Inc., a U.S. national corporation, applicant for the designation of all countries except the U.S., and, Jason Basil Nikas, a citizen of the U.S., Walter Cheney Low, a citizen of the U.S., Amy Patrice Skubiz, a citizen of the U.S., and Kristin Louise Murgic Boylan, a citizen of the U.S., applicants for the designation of the U.S. only, and claims priority to U.S. Patent Application Serial No. 61/481,556 filed on 02 May 2011, the disclosure of which is incorporated herein by reference in its entirety.

Statement of Government Rights
This invention was made with government support under Grant No. T32 DA007097 awarded by the National Institutes of Health (NIH). The government has certain rights in the invention.

Background of the Invention
Ovarian cancer is the most lethal gynecological malignancy in the U.S., due in part to the subtlety of its symptoms. It accounts for ~3% of all cancers in women in the U.S. Approximately 24,000 new cases of ovarian cancer are diagnosed each year in the U.S., resulting in about 16,000 deaths per year. Current diagnostic tests are neither adequately sensitive nor specific; consequently the majority of ovarian cancer patients are diagnosed with advanced disease. Standard therapy for ovarian cancer involves debulking surgery to reduce tumor burden followed by chemotherapy with a combination of platinum and paclitaxel (e.g., TAXOL®). Initially, up to 80% of ovarian cancer patients respond to chemotherapy, however most patients relapse in less than 2 years.

Taxol® is a mitotic inhibitor, whose mechanism of action is to prevent a) the destabilization of microtubules, necessary to the formation of the mitotic spindle and subsequent chromosomal separation during mitosis and b) the formation of new microtubules, also necessary to the aforementioned mitotic stages. Another cytostructural component to the destabilization of the microtubules and subsequent formation of the mitotic spindle is β actin, a polymeric microfilament that helps hold microtubules together.
Summary of the Invention

One embodiment provides a method to determine if an ovarian cancer patient is a long term survivor or a short term survivor comprising measuring the level of expression of at least one gene in a sample from the patient, wherein the level of expression of the at least one gene in the sample is an indication that the subject is a long term survivor or a short term survivor.

In embodiments, a method of selecting a treatment for a subject having ovarian cancer comprises determining whether a subject having ovarian cancer is likely to have short term or long term survival by a method comprising measuring the level of gene expression of at least a set of genes comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4 in a sample comprising ovarian cancer cells from the subject; inputting the expression levels of the set of genes into a function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer to obtain an output score; determining whether the subject is likely to have long term survival by determining if the output score is less than a cutoff value or whether the subject is likely to have short term survival by determining if the output score is greater than or equal to the cutoff value, wherein the cutoff value is a value determined by identifying a value between the 99% confidence interval of a mean output score of a first set of samples from subjects known to have short term survival and the 99% confidence interval of a mean output score of a second set of samples from subjects known to have long term survival; and optionally, displaying whether the output score is greater than or equal to the cutoff value or less than the cutoff value to a health care worker so that the health care worker can select a treatment for the subject.

In embodiments, a method of selecting a treatment for a subject having ovarian cancer comprises determining whether the subject having ovarian cancer is likely to have short term or long term survival by a method comprising measuring the level of gene expression of at least a set of genes comprising SSR1, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L, and EED in a sample comprising ovarian cancer cells from the subject; inputting the expression levels of the set of genes into a function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer to obtain an output score; determining whether the subject is likely to have long term survival by determining if the output score is less than a cutoff value or whether the subject is likely to have short term survival by determining if the output score is greater than or equal to the cutoff value, wherein the cutoff value is a value.
determined by identifying a value between the 99% confidence interval of a mean output score of a first set of samples from subjects known to have short term survival and the 99% confidence interval of a mean output score of a second set of samples from subjects known to have long term survival; and optionally, displaying whether the output value of the sample is greater than or equal to the cutoff value or less than the cutoff value so that the health care worker can select a treatment for the subject.

In embodiments, a method of selecting a treatment for a subject having ovarian cancer comprises determining whether the subject is likely to have short term or long term survival by a method comprising measuring the level of gene expression of at least a set of genes comprising CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED in a sample comprising ovarian cancer cells from the subject; inputting the expression levels of the set of genes into a function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer to obtain an output score; determining whether the subject is likely to have long term survival by determining if the output score is less than a cutoff value or whether the subject is likely to have short term survival by determining if the output score is greater than or equal to the cutoff value, wherein the cutoff value is a value determined by identifying a value between the 99% confidence interval of a mean output score of a first set of samples from subjects known to have short term survival and the 99% confidence interval of a mean output score of a mean output score of a second set of samples from subjects known to have long term survival; and optionally, displaying whether the output score is less than a cutoff value or greater than or equal to the cutoff so that the health care worker can select a treatment for the subject.

In embodiments, the methods further comprise treating a subject likely to have long term survival with standard chemotherapy. In embodiments, standard chemotherapy comprises taxol and/or platinum. In embodiments, the method further comprises treating a subject likely to have short term survival with therapy in addition to or in place of standard chemotherapy. In embodiments, an alternative therapy comprises a therapy selected from the group consisting of antiangiogenesis compounds, taxane analogues, tubulin binding agents, and ubiquitination inhibitors. In embodiments, a subject likely to have short term survival is treated with an inhibitor of a protein selected from the group consisting of TUBA3C, ACTB, CDC42 and combinations thereof.

In yet other embodiments, the disclosure provides a method for selecting a treatment for a subject that has ovarian cancer comprising, the method comprising: calculating an
output score, using a computing device, by inputting gene expression levels of a first set of
genes comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4, a
second set of genes comprising SSR1, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C,
MED13L, and EED, or a third set of genes comprising CDC42, LYPLA2, TUBA3C, ACTB,
HLCS, MED13L, and EED, into a function that provides a predictive relationship between
gene expression levels of the set of genes and short term or long term survival of subjects
having ovarian cancer; and displaying the output score, using a computing device. In
embodiments, the method further comprises determining whether the output score is greater
than or equal to or less than a cutoff value, using a computing device; and displaying whether
the subject is likely to be a short term or long term survivor.

One embodiment provides a method for diagnosing ovarian cancer in a subject
comprising: measuring the level of expression of at least one gene in a test sample from a
subject and comparing the level of expression with the level of expression of the at least one
gene in a control sample from a healthy subject, wherein a higher or lower level of expression
of the gene in the test sample compared with the level of expression in the control sample is
an indication that the subject has ovarian cancer. In one embodiment, the mRNA levels are
measured. In another embodiment, the protein levels are measured. In one embodiment, the
gene expression levels are measured by microarray analysis.

One embodiment provides that expression of LYPLA2, TUBA3C, ACTB, MED13L,
OSBPL8, EED, PKP4, SSR1, USP5, HLCS, NDUFB1, CDC42 or a combination thereof is
measured. In another embodiment, the expression of LYPLA2, TUBA3C, ACTB, MED13L,
OSBPL8, EED, and PKP4 is measured. In another embodiment, the expression of LYPLA2,
TUBA3C, ACTB and PKP4 is increased and the expression of MED13L, OSBPL8, and EED
is decreased. In one embodiment, the expression of SSR1, USP5, ACTB, HLCS, NDUFB1,
LYPLA2, TUBA3C, MED13L, and EED is measured. In another embodiment, the
expression of SSR1, NDUFB1, MED13L and EED is decreased and the expression USP5,
ACTB, HLCS LYPLA2 and TUBA3C is increased. In one embodiment, the expression of
CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED is measured. In another
embodiment the expression of CDC42, LYPLA2, TUBA3C, ACTB and HLCS is increased
and the expression of MED13L and EED is decreased. In one embodiment, the expression of
LYPLA2, TUBA3C, ACTB, USP5, HLCS, CDC42 or a combination thereof is increased. In
another embodiment, the expression of MED13L, OSBPL8, EED, PKP4, SSR1, NDUFB1 or
a combination thereof is decreased.
In one embodiment, the expression of LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, PKP4, SSR1, USP5, HLCS, NDUFB1, CDC42 or a combination thereof is measured and applied to a mathematical function to yield a diagnosis of ovarian cancer.

In one embodiment, the measurement of gene expression provides a diagnosis which indicates that the subject/patient will survive the cancer longer than about seven years. In another embodiment, the measurement of gene expression provides a diagnosis that the subject/patient will not survive the cancer for longer than about three years.

In one embodiment, the subject/patient is a mammal, such as a human.

In one embodiment, a health care provider or worker is informed. In another embodiment, the subject/patient is treated for ovarian cancer.

In another aspect, the disclosure provides kits for selecting a treatment for an ovarian cancer patient. In embodiments, a kit comprises or consists essentially of primer or a probe or both that specifically hybridizes to each gene of a first set of genes comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4. In embodiments, the kit consists essentially of reagents for detecting expression of the first set of genes and contains other reagents such as primer or probes for housekeeping genes, positive controls and/or negative controls. In other embodiments, a kit comprises or consists essentially of: a primer or a probe or both that specifically hybridizes to each gene of a first set of genes comprising SSR1, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L, and EED. In yet other embodiments, a kit comprises or consists essentially of a primer or a probe or both that specifically hybridizes to each gene of a first set of genes comprising CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED.

In embodiments, the kit contains no more than 200 primers or probes or both, no more than 175 primers, probes or both, no more than 150 primers, probes or both, no more than 125 primers, probes or both, no more than 100 primers, probes or both, no more than 75 primers, probes or both, no more than 50 primers, probes or both, no more than 25 primers, probes or both, or no more than 15 primers, probes or both.

In embodiments, a kit further comprises a computer readable storage medium having computer-executable instructions that, when executed by a computing device, cause the computing device to perform a step comprising: calculating an output score by inputting gene expression levels of a set of genes comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4, a second set of genes comprising SSR1, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L, and EED, or a third set of genes comprising CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED from a sample, into a
function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer.

In embodiments, the disclosure provides a computing device comprising a processing unit; and a system memory connected to the processing unit, the system memory including instructions that, when executed by the processing unit, cause the processing unit to: calculate an output score by inputting gene expression levels of a set of genes comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4, a second set of genes comprising SSR1, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L, and EED, or a third set of genes comprising CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED from a sample, into a function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer; and display the output score. In yet another embodiment, the system memory includes instructions, that when executed by the processing unit, cause the processing unit to determine whether the output score is greater than or equal to or less than a cutoff value; and displaying whether the subject is likely to be a short term or long term survivor.

**Brief Description of the Drawings**

Figure 1. Output scores from F1 ovarian cancer gene expression biomarker for long-term vs. short-term survival (responders vs. non-responders).

Figure 2. Output scores from F2 and F3 ovarian cancer gene expression biomarkers for long-term vs. short-term survival (responders vs. non-responders).

Figure 3. Box plots of the output scores of two survival/(treatment-response) groups (LTS and STS) of the F1 biomarker.

Figure 4. Box plots of the output scores of two survival/(treatment-response) groups (LTS & STS) of the F2 and F3 biomarkers.

Figure 5. 3D plot of output scores from long term and short term survivor subjects from functions F1 vs. F2 vs. F3. It can be seen that, with the exception of one subject, the three biomarkers are able to separate long-term from short-term survivors (responders vs. non-responders) in this 3D space.

Figure 6. Scatter plot & bar graph of output scores of all individual subjects [both LTS (responders) and STS (non-responders)] of the F1 prognostic biomarker. All 10 unknown STS subjects have F1 scores that are higher than the cutoff value (21.4), whereas all 10 unknown LTS subjects have F1 scores that are lower than the cutoff value.
Figure 7. Scatter plot and bar graph of output score of all individual subjects (both LTS (responders) and STS (non-responders)) of the F2 and F3 prognostic biomarkers. All 10 unknown STS subjects have F3 scores that are higher than the cutoff value (14.3 for F2 and 14.7 for F3), whereas all 10 unknown LTS subjects have F2 and F3 scores that are lower than the cutoff value.

Figure 8. Three-dimensional plot of prognostic biomarkers of output scores from each function F1 vs. F2 vs. F3 for the validation (qualification) study of long-term (responders) and short-term (non-responders) ovarian cancer survivors. As can be seen there is a complete segregation of the two survival/(treatment response) groups.

Figure 9 provides mathematical equations.

**Detailed Description of the Invention**

There are currently few reliable prognostic markers available for the diagnosis and/or prognosis of ovarian cancer, in particular, the classification of ovarian patients in relation to short-term (less than about three years from diagnosis, including several weeks, several months, 1 year, 2 year or three years) vs. long-term survivors (at least about 4 years, about 5 years, about 6 years or about 7 years or longer than about 7 years) or in relation to response to the standard aforementioned chemotherapy treatment. The ability to distinguish between these two patient populations would allow the modification of treatment therapies and/or the development of new pharmacological treatments for short-term survivors to potentially prolong their survival time.

Described herein are novel prognostic biomarkers that can distinguish between ovarian cancer patients who will survive longer than seven years versus those who will succumb to the disease within three years using a novel mathematical bioinformatic approach for the analysis of gene expression in each patient's tumor tissue. This novel mathematical bioinformatic approach has resulted in the discovery of novel genes and networks underlying the progression from long-term survival to short-term survival in ovarian cancer patients.

In one embodiment, the gene biomarkers that constitute this novel gene network when combined together into a single complex mathematical function and, thus, treated as a single complex biomarker, have a very high prognostic power (AUC of 0.978). This AUC value indicates that these biomarkers can both independently and collectively be used to identify short-term survivors with a very high accuracy and therefore provide alternative treatments that may extend their survival. In general, this approach demonstrates the potential of
personalized medicine based on the particular gene expression of a patient as it pertains to their specific disease.

One of the discovered genes, namely, TUBA3C, is directly linked to the mechanism of action of taxol, the standard chemotherapy treatment for ovarian cancer. Two of the remaining discovered genes, namely, ACTB and CDC42, are indirectly linked to the mechanism of action of taxol. More specifically, the TUBA3C gene is responsible for the production of microtubules, something which is needed for cell proliferation, and something which taxol is trying to oppose. The gene ACTB is responsible for the production of β-actin, which can be polymerized to form β-actin microfilaments, which are used for the polymerization or depolymerization of microtubules. Taxol and other taxol analogs oppose either the depolymerization or polymerization of microtubules, respectively. The gene CDC42 promotes the polymerization of β-actin into microfilaments, and, furthermore, it can regulate the polarization of both the actin and the microtubule cytoskeleton. All three of those genes were significantly over-expressed in the short-term survivors as compared with those of the long-term survivors. This indicates that in the case of the short-term survivors, taxol cannot overcome the combined effect of the TUBA3C, ACTB, and CDC42 genes, and that those individuals will not respond to the standard treatment of care, i.e. chemotherapy with platinum and taxol. In addition, the findings indicate that chemotherapeutic agents that inhibit the overexpression of these genes are useful to extend the survival of ovarian cancer patients.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, several embodiments with regards to methods and materials are described herein. As used herein, each of the following terms has the meaning associated with it in this section.

The articles "a" and "an" are used herein to refer to one or to more than one (i.e. to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

"Plurality" means at least two.

A "subject" or "patient" is a vertebrate, including a mammal, such as a human. Mammals include, but are not limited to, humans, farm animals, sport animals and pets.

The term "biological sample," as used herein, refers to samples obtained from a subject, including, but not limited to, skin, hair, tissue, blood, plasma, serum, cells, sweat,
saliva, feces, tissue and/or urine.

The term "about," as used herein, means approximately, in the region of, roughly, or around. When the term "about" is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. In general, the term "about" is used herein to modify a numerical value above and below the stated value by a variance of 10%. In one aspect, the term "about" means plus or minus 20% of the numerical value of the number with which it is being used. Therefore, about 50% means in the range of 45%-55%. Numerical ranges recited herein by endpoints include all numbers and fractions subsumed within that range (e.g. 1 to 5 includes 1, 1.5, 2, 2.75, 3, 3.90, 4, and 5). It is also to be understood that all numbers and fractions thereof are presumed to be modified by the term "about."

As used herein, the term "biologically active fragments" or "bioactive fragment" of the polypeptides encompasses natural or synthetic portions of the full length protein that are capable of specific binding to their natural ligand or of performing the function of the protein. For example, a "functional" or "active" biological molecule is a biological molecule in a form in which it exhibits a property by which it is characterized. A functional enzyme, for example, is one which exhibits the characteristic catalytic activity by which the enzyme is characterized.

A "fragment" or "segment" is a portion of an amino acid sequence, comprising at least one amino acid, or a portion of a nucleic acid sequence comprising at least one nucleotide. The terms "fragment" and "segment" are used interchangeably herein. As used herein, the term "fragment," as applied to a protein or peptide, can ordinarily be at least about 3-15 amino acids in length, at least about 15-25 amino acids, at least about 25-50 amino acids in length, at least about 50-75 amino acids in length, at least about 75-100 amino acids in length, and greater than 100 amino acids in length. As used herein, the term "fragment" as applied to a nucleic acid, may ordinarily be at least about 20 nucleotides in length, typically, at least about 50 nucleotides, more typically, from about 50 to about 100 nucleotides, at least about 100 to about 200 nucleotides, at least about 200 nucleotides to about 300 nucleotides, at least about 300 to about 350, at least about 350 nucleotides to about 500 nucleotides, at least about 500 to about 600, at least about 600 nucleotides to about 620 nucleotides, at least about 620 to about 650, and or the nucleic acid fragment will be greater than about 650 nucleotides in length.

The term "binding" refers to the adherence of molecules to one another, such as, but not limited to, enzymes to substrates, ligands to receptors, antibodies to antigens, DNA
binding domains of proteins to DNA, and DNA or RNA strands to complementary strands. "Binding partner," as used herein, refers to a molecule capable of binding to another molecule.

As used herein, "health care provider or worker" includes either an individual or an institution that provides preventive, curative, promotional or rehabilitative health care services to a subject, such as a patient. In one embodiment, the data is provided to a health care provider so that they may use it in their diagnosis/treatment of the patient.

The term "standard," as used herein, refers to something used for comparison, such as control or a healthy subject.

The terms "comprises", "comprising", and the like can have the meaning ascribed to them in U.S. Patent Law and can mean "includes", "including" and the like. As used herein, "including" or "includes" or the like means including, without limitation.

The term "primer" refers to a nucleic acid capable of acting as a point of initiation of synthesis along a complementary strand when conditions are suitable for synthesis of a primer extension product. The synthesizing conditions include the presence of four different bases and at least one polymerization-inducing agent such as reverse transcriptase or DNA polymerase. These are present in a suitable buffer, which may include constituents which are co-factors or which affect conditions such as pH and the like at various suitable temperatures. A primer is preferably a single strand sequence, such that amplification efficiency is optimized, but double stranded sequences can be utilized. Primers are typically at least about 15 nucleotides. In embodiments, primers can have a length of anywhere from 15 to 2000 nucleotides. In embodiments, primers have a melting temp of at least 50°C, 52°C, 55°C, 58°C, 60°C, or 65°C.

The term "probe" refers to a nucleic acid that hybridizes to a target sequence. In some embodiments, a probe includes about eight nucleotides, about 10 nucleotides, about 15 nucleotides, about 20 nucleotides, about 25 nucleotides, about 30 nucleotides, about 40 nucleotides, about 50 nucleotides, about 60 nucleotides, about 70 nucleotides, about 75 nucleotides, about 80 nucleotides, about 90 nucleotides, about 100 nucleotides, about 110 nucleotides, about 115 nucleotides, about 120 nucleotides, about 130 nucleotides, about 140 nucleotides, about 150 nucleotides, about 175 nucleotides, about 187 nucleotides, about 200 nucleotides, about 225 nucleotides, and about 250 nucleotides. A probe can further include a detectable label. Detectable labels include, but are not limited to, a fluorophore (e.g., Texas-Red®, Fluorescein isothiocyanate, etc.) and a hapten, (e.g., biotin). A detectable label can be covalently attached directly to a probe oligonucleotide, e.g., located at the probe's 5' end or
at the probe's 3' end. A probe including a fluorophore may also further include a quencher, e.g., Black Hole Quencher™, Iowa Black™, etc.

Ovarian Cancer

Ovarian cancer is a cancerous growth arising from different parts of the ovary. Most (>90%) ovarian cancers are classified as "epithelial" and were believed to arise from the surface (epithelium) of the ovary. However, recent evidence suggests that the Fallopian tube could also be the source of some ovarian cancers. Other types arise from the egg cells (germ cell tumor) or supporting cells (sex cord/stromal).

In 2004, in the United States, 25,580 new cases were diagnosed and 16,090 women died of ovarian cancer. The risk increases with age and decreases with pregnancy. Lifetime risk is about 1.6%, but women with affected first-degree relatives have a 5% risk. Women with a mutated BRCA1 or BRCA2 gene carry a risk between 25% and 60% depending on the specific mutation. Ovarian cancer is the fifth leading cause of death from cancer in women and the leading cause of death from gynecological cancer.

Ovarian cancer usually has a poor prognosis. It is disproportionately deadly because it lacks any clear early detection or screening test, meaning that most cases are not diagnosed until they have reached advanced stages. More than 60% of patients presenting with this cancer already have stage III or stage IV cancer, when it has already spread beyond the ovaries. Ovarian cancers shed cells into the naturally occurring fluid within the abdominal cavity. These cells can then implant on other abdominal (peritoneal) structures including the uterus, urinary bladder, bowel and the lining of the bowel wall (omentum) forming new tumor growths before cancer is even suspected.

Ovarian cancer causes non-specific symptoms. Most women with ovarian cancer report one or more symptoms such as abdominal pain or discomfort, an abdominal mass, bloating, back pain, urinary urgency, constipation, tiredness and a range of other non-specific symptoms, as well as more specific symptoms such as pelvic pain, abnormal vaginal bleeding or involuntary weight loss. There can be a build-up of fluid (ascites) in the abdominal cavity.

Diagnosis of ovarian cancer starts with a physical examination (including a pelvic examination), a blood test (for CA-125 and sometimes other markers), and transvaginal ultrasound. The diagnosis must be confirmed with surgery to inspect the abdominal cavity, take biopsies (tissue samples for microscopic analysis) and look for cancer cells in the abdominal fluid. Treatment usually involves chemotherapy and surgery, and sometimes radiotherapy.
In most cases, the cause of ovarian cancer remains unknown. Older women, and in those who have a first or second degree relative with the disease, have an increased risk. Hereditary forms of ovarian cancer can be caused by mutations in specific genes (most notably BRCA1 and BRCA2, but also in genes for hereditary nonpolyposis colorectal cancer). Infertile women and those with a condition called endometriosis, those who have never been pregnant and those who use postmenopausal estrogen replacement therapy are at increased risk. Use of combined oral contraceptive pills is a protective factor. The risk is also lower in women who have had their uterine tubes blocked surgically (tubal ligation).

Ovarian cancer is classified according to the histology of the tumor, obtained in a pathology report. Surface epithelial-stromal tumor, also known as ovarian epithelial carcinoma, is the most common type of ovarian cancer. It includes serous tumor, endometrioid tumor and mucinous cystadenocarcinoma. Sex cord-stromal tumor, including estrogen-producing granulosa cell tumor and virilizing Sertoli-Leydig cell tumor or arrhenoblastoma, accounts for 8% of ovarian cancers. Germ cell tumor accounts for approximately 30% of ovarian tumors, but only 5% of ovarian cancers. Germ cell tumor tends to occur in young women and girls. The prognosis depends on the specific histology of germ cell tumor. Mixed tumors, containing elements of more than one of the above classes of tumor histology are also possible.

Ovarian cancer staging is by the FIGO staging system and uses information obtained after surgery, which can include a total abdominal hysterectomy, removal of (usually) both ovaries and fallopian tubes, (usually) the omentum, and pelvic (peritoneal) washings for cytopathology. The AJCC stage is the same as the FIGO stage. The AJCC staging system describes the extent of the primary Tumor (T), the absence or presence of metastasis to nearby lymph Nodes (N), and the absence or presence of distant Metastasis (M).

Stage I - limited to one or both ovaries
1A - involves one ovary; capsule intact; no tumor on ovarian surface; no malignant cells in ascites or peritoneal washings
1B - involves both ovaries; capsule intact; no tumor on ovarian surface; negative washings
1C - tumor limited to ovaries with any of the following: capsule ruptured, tumor on ovarian surface, positive washings

Stage II - pelvic extension or implants
2A - extension or implants onto uterus or fallopian tube; negative washings
2B - extension or implants onto other pelvic structures; negative washings
IIC - pelvic extension or implants with positive peritoneal washings
Stage III - microscopic peritoneal implants outside of the pelvis; or limited to the pelvis with extension to the small bowel or omentum
IIIA - microscopic peritoneal metastases beyond pelvis
IIIB - macroscopic peritoneal metastases beyond pelvis less than 2 cm in size
IIIC - peritoneal metastases beyond pelvis > 2 cm or lymph node metastases
Stage IV - distant metastases to the liver or outside the peritoneal cavity

Para-aortic lymph node metastases are considered regional lymph nodes (Stage IIIC). As there is only one para-aortic lymph node intervening before the thoracic duct on the right side of the body, the ovarian cancer can rapidly spread to distant sites such as the lung.

The AJCC/TNM staging system includes three categories for ovarian cancer, T, N and M. The T category contains three other subcategories, T1, T2 and T3, each of them being classified according to the place where the tumor has developed (in one or both ovaries, inside or outside the ovary). The T1 category of ovarian cancer describes ovarian tumors that are confined to the ovaries, and which may affect one or both of them. The sub-subcategory T1a is used to stage cancer that is found in only one ovary, which has left the capsule intact and which cannot be found in the fluid taken from the pelvis. Cancer that has not affected the capsule, is confined to the inside of the ovaries and cannot be found in the fluid taken from the pelvis but has affected both ovaries is staged as T1b. T1c category describes a type of tumor that can affect one or both ovaries, and which has grown through the capsule of an ovary or it is present in the fluid taken from the pelvis. T2 is a more advanced stage of cancer. In this case, the tumor has grown in one or both ovaries and is spread to the uterus, fallopian tubes or other pelvic tissues. Stage T2a is used to describe a cancerous tumor that has spread to the uterus or the fallopian tubes (or both) but which is not present in the fluid taken from the pelvis. Stages T2b and T2c indicate cancer that metastasized to other pelvic tissues than the uterus and fallopian tubes and which cannot be seen in the fluid taken from the pelvis, respectively tumors that spread to any of the pelvic tissues (including uterus and fallopian tubes) but which can also be found in the fluid taken from the pelvis. T3 is the stage used to describe cancer that has spread to the peritoneum. This stage provides information on the size of the metastatic tumors (tumors that are located in other areas of the body, but are caused by ovarian cancer). These tumors can be very small, visible only under the microscope (T3a), visible but not larger than 2 centimeters (T3b) and bigger than 2 centimeters (T3c).
This staging system also uses N categories to describe cancers that have or not spread to nearby lymph nodes. There are only two N categories, NO which indicates that the cancerous tumors have not affected the lymph nodes, and N1 which indicates the involvement of lymph nodes close to the tumor.

The M categories in the AJCC/TNM staging system provide information on whether the ovarian cancer has metastasized to distant organs such as liver or lungs. MO indicates that the cancer did not spread to distant organs and M1 category is used for cancer that has spread to other organs of the body.

The AJCC/TNM staging system also contains a Tx and a Nx sub-category which indicates that the extent of the tumor cannot be described because of insufficient data, respectively the involvement of the lymph nodes cannot be described because of the same reason.

The ovarian cancer stages are made up by combining the TNM categories in the following manner:

Stage I: T1+N0+M0  
IA: T1a+N0+M0  
IB: T1b+N0+M0  
IC: T1c+N0+M0  
Stage II: T2+N0+M0  
Ila: T2a+N0+M0  
IIB: T2b+N0+M0  
IIC: T2c+N0+M0  
Stage III: T3+N0+M0  
IIIA: T3a+N0+M0  
IIIB: T3b+N0+M0  
IIIC: T3c+N0+M0 or Any T+N1+M0  
Stage IV: Any T+ Any N+M1

Ovarian cancer, as well as any other type of cancer, is also graded, apart from staged. The histologic grade of a tumor measures how abnormal or malignant its cells look under the microscope. There are four grades indicating the likelihood of the cancer to spread and the higher the grade, the more likely for this to occur. Grade 0 is used to describe non-invasive tumors. Grade 0 cancers are also referred to as borderline tumors. Grade 1 tumors have cells that are well differentiated (look very similar to the normal tissue) and are the ones with the best prognosis. Grade 2 tumors are also called moderately well differentiated and they are
made up by cells that resemble the normal tissue. Grade 3 tumors have the worst prognosis and their cells are abnormal, referred to as poorly differentiated.

With regard to treatment, surgical treatment may be sufficient for malignant tumors that are well-differentiated and confined to the ovary. Addition of chemotherapy may be required for more aggressive tumors that are confined to the ovary. For patients with advanced disease a combination of surgical reduction with a combination chemotherapy regimen is standard. Borderline tumors, even following spread outside of the ovary, are managed well with surgery, and chemotherapy is not seen as useful.

Chemotherapy has been a general standard of care for ovarian cancer for decades, although with highly variable protocols. Chemotherapy is used after surgery to treat any residual disease, if appropriate. This depends on the histology of the tumor; some kinds of tumor (particularly teratoma) are not sensitive to chemotherapy. In some cases, there may be reason to perform chemotherapy first, followed by surgery.

For patients with stage IIIC epithelial ovarian adenocarcinomas who have undergone successful optimal debulking, a recent clinical trial demonstrated that median survival time is significantly longer for patient receiving intraperitoneal (IP) chemotherapy. Patients in this clinical trial reported less compliance with IP chemotherapy and fewer than half of the patients received all six cycles of IP chemotherapy. Despite this high "drop-out" rate, the group as a whole (including the patients that didn't complete IP chemotherapy treatment) survived longer on average than patients who received intravenous chemotherapy alone.

**Methods of selecting a treatment**

The disclosure provides methods for selecting a treatment for a subject having ovarian cancer. In embodiments, a method of selecting a treatment for a subject that has ovarian cancer comprises: a)determining whether the subject is likely to have short term or long term survival by a method comprising i)measuring the level of gene expression of at least a set of genes in a sample comprising ovarian cancer cells from the subject; ii)inputting the expression levels of the set of genes into a function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer to obtain an output score; iii)determining whether the subject is likely to have long term survival by determining if the output score is less than a cutoff value or whether the subject is likely to have short term survival by determining if the output score is greater than or equal to the cutoff value, wherein the cutoff value is a value determined by identifying a value between the 99% confidence interval of the mean output score of a first set of samples from subjects known to have short term survival and the 99%
confidence interval of the mean output score of a second set of samples from subjects known to have long term survival; and b) optionally, displaying whether the output score is greater than or equal to the cutoff value or less than the cutoff value to a health care worker so that the health care worker can select a treatment for the subject.

In embodiments, the set of genes comprises at least the genes LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4. In other embodiments the set of genes comprises at least the genes SSR1, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L, and EED. In yet other embodiments a set of genes comprises CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED.

Markers

The expression of certain genes has been demonstrated herein to be prognostic of ovarian cancer. Sequences of the protein, nucleic acid encoding the protein and the Affymetrix sequences are available either at GenBank or the Affymetrix website. These genes include the following:

LYPLA2, such as human lysophospholipase II, is represented, for example, by accession numbers 215566_x_at, NM_007269, NM 007620, or NP_009191(231 aa; gi2032149).

The protein sequence is:

```
MCGNTMSVPVLLDAATVSLHGLGLGDWADALSLPHVKY1CPHAPRIFVTLNMKVMPSWDFLMGLSPDAEDEAGIKKAAENIKALIEHEMKNGIPANRVGGALSPLYTALTCHPSLAGIVALSCWLPRLHRAFPQAAN
```

The mRNA sequence is:

```
AAKDLAILQCHGELDPMVPVFGRALTAEKLRSWTPARVQFKTYPGVMHSSCPQEMA
```

**AVKEFLEKLLP**V (SEQ ID NO:1) and the mRNA sequence is:

```
1 ggaagttccg gcgggggccc ccaggggga aagagtcgtc tcgggagaa agagagaat
61 cggccaaagc gcctcggagc tcgggagcgc cggccggagc cggccggagc cggccggagc
ggctgagttt aacaccatgt ctgtgccctct gctcaccgct gctcaccgct gctcaccgct
121 tggcagggaa agggccgcttg ttatatttatt acatggacgt ggagacacag gcgacacatg
181 ggctgacgcc ccttcacacca cccgggttcc tcaagcgtta tacatcgttc ccactgcggcc
241 tagatcctct gtcggcctcta acatgagatg ggtgtgaccc tccgggtctgcc
301 taggtgatcct gttacgctccta acatgagatg ggtgtgaccc tccgggtctgcc
361 ggtgcagtcct gatgccccag aggacgaggg tggcatacgg aagacagcag aagacatccaa
421 ggcttcgagtt gagcatgaa aagagggcctt gatgccccag aggacgaggg tggcatacgg
481 cttcctcaggg ggccggccct ttgccctctta ccaggggccc aacctggggccc aacctgcggc
```

```
TUBA3C, such as human tubulin, alpha 3c, is represented, for example, by accession numbers 210527_x_at, NM_006001.01 (gi 325053695), or NP_005992 (450aa; gi 17921933). The protein sequence is:

```
MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTIGG
GDDSFTNFFSETGAGKHPRAVFVFDLEPTWDDEVRTGT YRQLFHFPEQLITGKEDAAAN
YARGHYTIKEIVDLVDLRKLADLCTGLQGFHFLFSGTSGFASLLMERLSVD
YGKKSLFFAIYPAPQVSTAWEPYNSILTHHTLHEODCAFMVDNEAIYDICRNLD
IERPTYNLRLIQVISSITALRFDGALNVDELETEQTNLVPRIHFPLATYAPVI
SAEKAYHEQLSVAEITNACFEPANQMKCDPHGKYMACMMLYRGDVNPREDATAIAEAWRLD
HKFDLMYAKRAFVHWYVGEMEGEFESEAREDLAALDKYEEVGVDSVEAEEEGER (SEQ
ID NO: 2)
```

and the mRNA sequence is:

```
1 ggttgaggtc aagtagtgcg gttggcgcnggc gcagctcaaca tcagctcaaca tgcgtctgatg
tatctctatc cacgtggggc aggccaggtc agtgataaaa ccattggtgg
ctgcctggaa catggaattc agcccgatgg tcagatgcca agtgataaaa ccattggtgg
121 tggggacgac tccttcacatc cccttctgggc aggctggcag gccttgagga
tcccacacct ggggggtttg gccttcgcc ccagggttgct cgtcgatgcc
181 tggcacatcgc ccttcgcc ccctgttcct cgcgggtggc tcagctcaaca tgcgtctgatg
241 agcagcgttct gttgccctgc gcagctcaaca tgcgtctgatg
tacccagctg ggcagccgg gcagctcaaca tgcgtctgatg
tcccagctg ggcagccgg gcagctcaaca tgcgtctgatg
361 cagagccagc tcagctcaaca tgcgtctgatg
tggcagctgg gcctggcagc tcagctcaaca tgcgtctgatg
tggcagctgg gcctggcagc tcagctcaaca tgcgtctgatg
421 ctggctgcttc ccttcgcc ccctgttcct cgcgggtggc tcagctcaaca tgcgtctgatg
481 ccttcgcc ccctgttcct cgcgggtggc tcagctcaaca tgcgtctgatg
tgcttctgc gcagctcaaca tgcgtctgatg
tgcttctgc gcagctcaaca tgcgtctgatg
541 ctggctgcttc ccttcgcc ccctgttcct cgcgggtggc tcagctcaaca tgcgtctgatg
601 ctgcggcagc tcagctcaaca tgcgtctgatg
tggcagctgg gcctggcagc tcagctcaaca tgcgtctgatg
tggcagctgg gcctggcagc tcagctcaaca tgcgtctgatg
```
ACTB, such as human beta actin, is represented, for example, by accession numbers 200801\_x\_at, NM\_001 101(gl 168480144), or NP\_001092(375 aa; gl4501885). The protein sequence is:

```
MDDDDIAALWDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVM
VGMGQKDSYVGDEAQSKGRILTLYPIEHGIVTNWDDMEKIWHHTFYNELRAPEEH
VLLTEAPLNPKANREKMTQIMFETFNAPMYVAIQAVLSLYASGRTTGIVMDSGDGV
HTVPITYEGYALPHAILRLDLAGRDLTDYLMIKILTERGYSFTAEREIVRDIKEK
VALDFEQEMATASSSLEKSYELPDQVITIGNERFRCEAEFLQPSFLGMECSIHE
```

and the mRNA sequence is:

```
1 accgccgaga ccgccgagcc ccgccgagca cagagcctcg ccctttgcgga ccgccgagc
61 gtccacaccgc ccgccgagct cccctgctgag cagatctgct gctgctgcac
121 ggcttggggct ttctccgctcg ccgcctgctct gctgctgcac
181 tccaatgccttc gacgccacag gcacccctgc gctgctgcac ccgccgagct
241 tatgtggggg agcagagcgc gctgctgcac ccgccgagct ccgccgagct
301 cagctcctcg ccgccgagct gcgcctccgc ctgcctccgc ccgccgagct
361 gagctctgcag ggagctcctcg ccgccgagct gcgcctccgc ccgccgagct
421 aagggcagac ccgccgagct ccgccgagct gcgcctccgc ccgccgagct
481 tacgcttgac tccctcctcg cctgcgctct cctgcgctct ccgccgagct
541 atgctctctc ggcgccgagct gcgcctccgc ccgccgagct gcgcctccgc
601 catgccttc gcgcctcctcg cctgcgctct cctgcgctct ccgccgagct
661 gcgcctcctcg cctgcgctct cctgcgctct ccgccgagct gcgcctccgc
721 aagggagag ccgccgagct gcgcctccgc ccgccgagct gcgcctccgc
```
MED13L, such as human mediator complex subunit 13-like, is represented, for example, by accession numbers 212209_at, NM_015335.4(gI300360584), or NP_0561150 (2210 aa;gI4477121 i). The protein sequence is:

MTAAANWVANGASLEDCHSNLFSLAELTGIKWRRYNFGGHGDCGPIISAPAQDPILLSFIRCLQANLLCVRRWDVKPDCKELWIFWGWDEPNLVGVIHHELQWEGLWENGLSYECRTLLFKAIHNLLERCLMDKNFVRIGKWFVRPYDEKPVNKS

EHLS CAFTFFLHGE SNVCTS VEJACHQP YILINEEIHMAGQ SPAP FQVLVS PYGLNGTLTGQAYKMDPAPRTKLEIEWQYFPMVLLKKEESKEDELGYDDDFPVAVEVIVGV

RMVPSAFVLISQNDIPVQSASAGGHIAVGQGQLGSVKPSNCMPMIPPTPSEQAILGESGMQAASHLVSQDDGMITMHSKRSKGIPKFLHHNVWKEICLNLRTQSK

RSQMSTPTLEDPEASNPATWDVFDPQVTSCGSCHKLKRCAGVNPFRPTVSQPGFSAGPSSSSLPZPASSHKTARERQEGDKLQKRPLIPFHHRPSVAELCMEQDTPGQKLGLAGIDSSLEVSSRSKYDKQMPSNRTSKOQMNPMDSPHSPISLPLSIQPGRQ

ETESLDPSPVPFYNALGNYLEQSQLSSLDDRTVLMQRLPMAEVSETAICYGIRPSPNESPENWHSYRLPSDAAEFRPPELQGERCAKMEVNSESTALQRLIAQPNKRFKWOQDKQPQPLQLWWLPLPSQPDGSLGEVNDPYPFEDDDIQYFTANKCKQGTEKD

SLKNKSEDQFQTOKDVTTPHSTVPFDQKNMSIFSATKTDVRQDANAGRAGSSSLTVQVTDLAPSLHLDDNIFNDDEDSALGVAPVSSKMMAPVYDPRPLGDGRAAVYPPTVADLQRFPTPSLQEHQAPSFIPNYKDQISSETVTALGMMESPVMVSMVQTELK
MEVEDGLGSPKPEEIKDFSYHKVPSQFVGVSSMFAPKLMLPLKIPDACLF
RPSWAIPPIEQLPMPPAATFIRDGNYGNPSVGLADPYNTPQMTNPVTNLNSAPAD
SNSGAVLPSPATPRSVPFTPRTPFRPRPGRGTTASGQSVKVDSTDQGSPASTPST
TRPLNSVEPATMQPIPEAHSLYVTLLSDSVMNIFKDRNFDSCCICACNMIKGAADV
LYIPDSSNEDQYRCCTGFSAIMNRLGYNLGFLALEDELDFGKNSDIGQAERRLMMC
QSTFLPQVEGTTKQPEEPIISLLLQLNQHQTQFASNLFLDYISSNRQTLPCVWSYD
RVQADNNDYWTFCFNAMLEQGRQYVDNPNTGKVDXQALVSATVHWSPHSNLDSMSLS
QDWRMLSSLQPFQLDAIQKRRTKRTWENIQHVQGFLTTQFWVFHMARTGTYGSEESPE
PLPITLVLVDYKDFLTPSFSLPLFWRLLDLQHGRRDYAVIWCPWNEALEGK
FFRDLSSAYEVMCRLGQHKPICKVLRDGIMRQGKTVQKLTDELVSEWFQNSWGEEN
DHSRLKLYAQVCRHHLAPLQLDLSSLLIPPKYQTTPAAQGGATFNGAPLAPNG
SAAPAGSASFNPSSTNPAPASSASSGSSFPVSSSASSAPGQSITTSSSSFSGS
VGGQPSTGIAADRTQGNICCGDTPQSSQQFSQDOESVTERERIGPETDPSAD
DSHAPPPAIIYMVDFFPTRYAEEDSLTSTSGNMTLSLRCMTNDLPEMRNISFILQI
VPCQMYLQTMKDEQVQYVQILKSMFVSCQCRPLIPQHIKSLTGFPAASIEMTL
KNPERSPILYPSFPFIAPIPDKQTELGETFGAESKYMNLFVGYCLSHDQWLLAS
CTDLHGELLETCA/NIALPNRSRRSVKSAIRKLGKLWECVGIVQMTSLPWWRIG
LGRLLGHGELKDSILLGECQLQTISSKLDVCRMCGISAADSPILSACLWMEQPSGS
FWMPDAVTMGSVFRSTALNMQSSQLNTPQDASCTHILVFPTSTTQVAPANYNED
GFSPPNDDMF-VDLPFPDDMDNDIGILMTGNLHSSPNSSPVPSGSPGIGVGFHQS
RSQGERLSSREAPELKLQOPLALTGFVSTAKAENLPQFWSSCPQAONCPFLKASL
HHHISVAQTDCELLPARNSQRVPHPLDSTKTDAN..RF"TELQYNALSWLTCNPATQDRTS
CLPVHFVLWTQLYNAIMNIL (SEQ ID NO.7) and the mRNA sequence is:

1  ctcggacgcG ctgcctcgac atgcgccgcgt ctgggccccg ggctcgggaa ggctgtaaac
61  tgcggcagcc aactggttgg cgaaccggcc gacgctggag gattgtcat ccaacctctt
121  ttcgctggct gaacatcacc gatcaaatag gctgtagtac aatctttggg ggcataggga
181  ctgtggaccc ataatttcac caccagccca agatgatcca attctttcag cccacccaca
241  tgcgctggaa actaatcact tgtgtctatg gcgtctggtg tccaaccag attgcaaga
301  tattgttgata tctctgatgca gaatgatgca caacccaggt tagtgataac atcatgaaact
361  gcaggtttttg gaagaaggcc tctgagagaa tggccctttcc tatgaatgtg ggaacgctgct
421  ctcgaaagcc atccgaaatc ttgtagaatg gtcaccaaatg cggataaatg cgatcagcct
481  tgggaatgtg tgtctgtgc cctagaaatg ggtggaagaa cgacgtaaa ccagatgcaaa
541  ttttgtcctgt cttctctcag cgctgcagca atctagcag cgtctgtgatg caaggtttggg
601  gattgacccg caccagccca tttatatgtt caatgggac gatacaca cggatcagta
661  ttcgctgctc ctgggtgcag actgtttaag tctcttccttc cggggaaa ttggcaccagt
721  cagaagatc actgatgctc acacgacca ccagagctg tattttcatg aatatttattg
tttatatgtt caatgggac gatacaca cggatcagta
781  ctaccggatg tgtctaaaaa agaaagaaga atcgaagag ggacagagt taggtgatatga
841 tgaatggattc cctgtgccag tgtgaagtaat tgttggtggt gttcggatgg tttacccctc
ttgatctctc agaatgacat cccggttcct acctgatctc tccgattccg tgatgaatat
cttttaaagac agaaactttg acagctgttg catctgtgcc tgcaacatga acatcaaagg
cggccagaggt gttgagaggg tggatggtgg tgggagaagt tgcaaagag tggccacatt
agcctccacc gcagcattgtt ttgatctctc agaatgacat cccggttcct acctgatc
tcctgtcctg catctgtgcc tgcaacatga acatcaaagg
cggccagaggt gttgagaggg tggatggtgg tgggagaagt tgcaaagag tggccacatt
agcctccacc gcagcattgtt ttgatctctc agaatgacat cccggttcct acctgatc
tcctgtcctg catctgtgcc tgcaacatga acatcaaagg
cggccagaggt gttgagaggg tggatggtgg tgggagaagt tgcaaagag tggccacatt
agcctccacc gcagcattgtt ttgatctctc agaatgacat cccggttcct acctgatc
tcctgtcctg catctgtgcc tgcaacatga acatcaaagg
cggccagaggt gttgagaggg tggatggtgg tgggagaagt tgcaaagag tggccacatt
agcctccacc gcagcattgtt ttgatctctc agaatgacat cccggttcct acctgatc
tcctgtcctg catctgtgcc tgcaacatga acatcaaagg
6121 ttagtatg tttgtgaccc ttccatcccc agatgatatg gacaatgata ttggcatatt
6181 aatgactggg aacctccatt ttccattccc agatgatatg gacaatgata ttggcatatt
6241 ttctggaatt ggtgtaaggct tctactttcc gcatagtcgg agccaggtgtg agctctctct
6301 ttctagagaa gcacccagag agctaaagac gcagcccttg gcccttggtt atttttgtac
6361 aacgccaaag gcctgaaatc ttccccagtgt gggtgtaaag tctgtctccc aggcctaaaa
6421 ccaagtgcctt ctctctttaa aagcctgccc gcatcaccac atttcatgag cacagcagac
6481 cgaactctcttg ctctccagga aggtctggag gttcccatcc aatactctta aacaagagtct
6541 gtgcagatttt ttgaaggttg gggttgagct ttttttgggg attttttttt aaaaaaaactt
6601 ctctggaacc aacagccgct ttccccgtct tccccctacc tttgtgctgc ttctcatgct
6661 gtacaatgcct atcatgataa tacattaatt ggaaagcacc ttgctctcct tcggctcattt
6721 cccctctccc gcacctcag ccccaagaaa ctgctactac tggaaataaa cccacactcc
6781 ttctcttcga caacctcota cagctcggca ctgctatttc ttctcagcgc gcacagtaca
6841 atctgtcaggt ctcttacctc agatgtactc cttgacactt ctctccatgg caacggaact
6901 ttagataaag gtgcgctttt ggcagcttga aagctctctc caccgccaac cactgcctag
6961 aggaggtggt gtgtggt tact ggcttttccag aacaagagcg agagtgactc tttttttttt
7021 caaaggactct ctgtgtgtct cttctgtaa catttgaggg cagctcggca cctctggcag
7081 aaccccttta atttcgctgtt tattggctat gggtttcgct aacacccoataa cttcggaccc
7141 ttttaaaaaa agacacagcc caagctatcg ttttctact gggttttctt tctctctagt
7201 gtctcccacc agatctcgat ggtttccttg gtttcacttc aacctcttgct tcaagtgcag
7261 aggtcctggag acagctctct tttggtattt atatatttaatt ttggtttttc tcataactcgt
7321 gtgaaccttatt tttaaattaata caagggacc aggtacagta gcgtgaaccc aattcagactc
7381 caccataagga ccccttcgtct acataccttc gcctcataag ccggaagagg gataaaaaca
7441 catggtgggag atcatgcgcta aatataata aatcagctaa accccagggt aagtttttgtt
7501 aacaacaaac agttctcttc atgtagcttg ggctttggat ggaaagcacc ttcagctaaa
7561 aacacccccca tacagaaggg agtccacgctt aatcatcggac cagtcggcag cttaaatcga
7621 ttgctttttat atgtgctcaca gggaaattta ttttttctat ggacaaggct tgttttttgtt
7681 ttctctctct cattagatca aaagggacct aagttcataac aacaattgaa aagttttgtc
7741 catgaaatcca gttataataa tgtgtaaatt ttctctggcc ataggaaatat tatctcaaaa
7801 aaatttttca accttacacc taaattagta cttgaatggt agcccccttg tgtggaacctt
7861 tttaaaaaat gctttttttt cagatcattg gttaatggga attttattat aacagcattt
7921 cctgccccag aggtttgctct ttgaaaccct ttcttttctt cttggttttct tgtttttgtc
7981 cttggagccag tgggtgtaagt gttgttcctcc gggagccact gttgacaggg ctttggagct
8041 tttgctatcttg gggaagagtt catgcaaatat ttaaattttt ttcaccaaga cttcctatgc
8101 cttggtatata aaaaaaaaag aaaaatgtta aaatttgatt gatttggata aagttttttt
8161 cttgaaataa tccaaagtta aaaaaatttct ataataaaaa aagttgggttt gtgggaaagt
8221 gaagacagagaatctcctctg aataattgta atttttgct ctgtgacacc gttggaaggg
8281 gggttgacttt ctgcgaatgt ataggaaaaa aacatctgata taaaccaaccc ttctttctca
8341 atgtgtcagag ttgaaattgt actttttaaa tatttgctag ctattttttttt ttaatcagaa
8401 aggaatttcc taaggccccttg ttagagacca taagagttga aagttggataa actttttttt
8461 aattctcttg gtggaagagat aatatttgga gtgatatatt gttcttttttt accactttttg
8521 ggtggtctttttt tcctggcaact ctctcgaac ctgtaattttt ttttttttttt ggcggttttgtt
8581 taatccaggg ggccttctttt cctgaaacca aaatttttaa caggaaaga aaaaaaaaaccc
8641 aaaaaataaa aacccccctac aaaaaacttt taaaaaat ggcaagaaag ggtagtttttt
8701 atctgtgtgcct ttttatttaaa gtttttttaag ttaaagaaag cttgtgacat atttttacgt
OSBPL8, such as human oxysterol binding protein-like 8, is represented, for example, by accession numbers 212585_at, NM_001003712.0 and NM_020841.4 (2 alternative transcripts), orNP_001003712 (847 aa;gl5 1243032) and NP_065892 (889aa;gll 8079218).

The protein sequence for variant 1 is:

```
MEGGGLADGEPEPRTSILLGDSDKVLGPTSWANSDEQQLTTPGKMS
```

```
QRGQKEAYPTPTKDLHQPSLPSAPSHQFQFERKEDISQNDESSLSMSKSESKLY
```

```
NGSEKDSSTSSXLTKWESLKVQXKYNREKXKARTKELLSTIDPSVIMADWLKIRGT
```

```
LKSWTKLWLVCVLKPYGELLYTITYKQNGQWGTNVLLNACEIERPSKDKGFCFLFHPLEQ
```

```
SIWAVKPGKGEAVGTSIQLPSYLLIRATSEDSGRCWMDALELACKSSLKRTMIR
```

```
EGKEKDLVSIVSSDSTHTFTYGLLARRNNLHGSDFNLNDSEIERQHFKDQDMYSDKSDE
```

```
NDQEIHSDNVMGKSEESDTDSERQDSYIEPFEPEFVLFKTETTEQSQHELGEAG
```

```
ASQETTVESENKLWTLLKQVRPGDLGSKLWTQPFILEPFRSLKDLSDDYYYHADFLS
```

```
EAALLEENPYPRLKKWLYGSLFYKPKGLKYPNPLGETFRCLWFIHPRTNSKTFY
```

```
AEQVSHHPISAFYVSNRKGDFCCLSGILAKSFYGNLSLAEIGEALRTFLNREGEDY
```

```
VMTMVFAYHCIGLYGTMTELGLGTNVITCQRTQGTSAILEFKLKPFLGSDCVQIQSGK
```

```
LKLGKEVLATLGEHDSSEVFIQTDDKTDNESVFWNNPFDIKQWRLLRHNVTQFEEQDFE
```

```
SEKLNQRVTRAINAKDQTEATQEKVLVEEAEQQRQAARDBRTKKEWESCLFLFDLCPTGE
```

```
WHKYFADTRPWPQLMNDIQLFEKDDGVQLTKVRKHTPMVSVPFRMKHKPQRQVKAVGYS
```

```
SEPPIPQDSGGSAEQSVPKSTRRKKGIELGDIQSSIESIKQTQEEIKRINARLHLV
```

```
ssTPATDYFLQQDYFIIFLLILLLQVYIIFNMFK (SEQ ID NO: 9) and the mRNA sequence for variant 1 is
```

```
1 cactagaatg tgaaggactc tgcgctttct cgggtgccacag aagcggcgcc aagcggcgcc
```

```
61 tgacaaacct aggccgcgac gcggctccttg ccagccgcgac gcgttgactg ttctgggtaa
```

```
121 cgccgcgacta gcgcctgcgg tgcagaacc gcgtggagga aggtggggtt cgcggcgccg
```

```
181 tgtggtcctc ccggcgccgg aatccgccgg cgcctgcgggg gcgtgcctcgc cagtgctcctc
```

```
241 gcgcgcgtgt gcgcgagccc ttgcgtcacc ctcgcgcgcc agtgcgcgcc gcagcagcagc
```

```
301 gcgcgagaga tttggggtct cgtgcagcac tcagccgggc ctcgcgcgcc gcgcgcgagc
```

```
361 gacttcggag gacaccgcct gcgtgctccc gcctggctgg gactgcctcg cgggatgggc
```

```
421 agatattcag aatgctgact ctattagaga aagagacagc taacaccatc ttctttttct
```

```
481 atggagggag gtgtggcctg aggagaaatc tcagcaagct cagctctcctt gcgtgaccgga
```

```
541 gatgtccttg ggccatcaac tgttgtagca aacagtgacg aatctcagct tctgacacca
```

```
(SEQ ID NO: 8)
```
3241 aatattccga aaacaaaaaa acctaggaga 
aaatagttga ggatttgact tttcgatctt
tttagtattt atttgtattt catgttagta cttttttgtt
3301 tgtatctgca ctatttttctt tagatatttt atttgtattt catgttagta cttttttgtt
3361 cttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
3421 cttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
3481 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
3541 gagaatgtata ctatctttctt cataaacag gcaaaaaag tttatcagta aggaattaaca
3601 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
3661 cttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
3721 gagaatgtata ctatctttctt cataaacag gcaaaaaag tttatcagta aggaattaaca
3781 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
3841 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
3901 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
3961 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4021 gagaatgtata ctatctttctt cataaacag gcaaaaaag tttatcagta aggaattaaca
4081 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4141 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4201 gagaatgtata ctatctttctt cataaacag gcaaaaaag tttatcagta aggaattaaca
4261 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4321 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4381 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4441 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4501 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4561 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4621 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4681 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4741 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4801 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4861 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4921 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
4981 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5041 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5101 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5161 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5221 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5281 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5341 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5401 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5461 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5521 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5581 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5641 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5701 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
5761 ttatatcact tataaaggta cagtgtactc tttgtcacag ctcagttggt aaccgcattc
The protein sequence for variant 2 is:

MSQRQGKEAYPTPTKDLHQPSLSPASPHSQGFERGKEDISQNKDESLLLSMSKSKSEKLYNGSEKDSSTSSKLTKESLKVKKNYREEKKRTKELLSTIT

The mRNA sequence for variant 2 is:

1 cactagaatg tgaaggatct tcgcggttct gggtgcccag aaaggcggcg acgcggcgga
2 tcgcgctgccg gtagcaagcc gactgaggga aggtgggggt ccgcccgggc

SEQ ID NO: 10

and the mRNA sequence for variant 2 is:

1 cactagaatg tgaaggatct tcgcggttct gggtgcccag aaaggcggcg acgcggcgga
2 tcgcgctgccg gtagcaagcc gactgaggga aggtgggggt ccgcccgggc

SEQ ID NO: 27
181 tgggtgacct cggggccgaa agtccccgcc cccgctcgggg gctgagccgg cagtgcctcc
241 gcggccgctg ggcagcgccc ttcgctcagg ctcgcgcccc ccgctcgggg gctgagccgg cagtgcctcc
301 ggcggagaga gttggtgctt gcacagctag ggctgacgsc gcgaacgccc ccggccgctg
361 gacctccggc gaaaccgccct ttcgctgctcc gcctcccttg gcgtctccga cggtgttggc
421 agatttacag aaggtgctca caattaatag ggaaagctca ttaaactccac ttctctattt
481 atggaggagg gtttggccag ttaaactccac ttcgaaacct gcggccgctg ggcagcgccc ttcgctcagg
541 gttccggagcac ccgagcgccc ctgctgcggcg cgtgagccgg cagtgcctcc
601 caaaagatct gctcagcggc cagcagctag ggctgacgsc gcgaacgccc ccggccgctg
661 gagggaagga agatactcc caaaaaataag atgaaccttc acttctctctg tcacagctag
721 agtctctgac ttaactcttt atgctctggtc gcagaacctc agcagaacgg trccatctctc
781 caaaaagaagtt ctcctttaag gctacaagaag aaaaattcgc ccaggaaagaa aaaaagacgc
841 caaaagagct gcattgagcc tctcaagact cttctctctct gttttggtta tttttggtta
901 agattcgttgc ttctctctct gcgctacagc gttgaggttc agttgagaaa cggggccgctg
961 taccctgctc taaaacccaa aaaaaaggct ac tgtgtaattt ctggggcttc
1021 ttgaactcatcg tgaacgctca ctaaaaaagtt tcctctctct ctctctctct cttttaaaag
1081 ttgagcaatcc tattttgcga gtaaggtggac caaaaaggta agcgggtggtc tcaccatctc
1141 aacctcttacc ttcaggttgc tttctctctct ctgacgcctgc gcctctctctc ctctctctct
1201 gatgtttgcct tttctctctct cttctctctct ctgacgcctgc gcctctctctc ctctctctct
1261 gggaaagaaa gcagactctgc gcagatcttt tcctctctct cttctctctct cttctctctct
1321 gctctctcttc tggctcaatcg tcctctctct gcgctacagc gttgaggttc agttgagaaa cggggccgctg
1381 ttgagcaatcc tattttgcga gtaaggtggac caaaaaggta agcgggtggtc tcaccatctc
1441 atccaaagagtt tctcctatcc gttgaggttc agttgagaaa cggggccgctg
1501 atcacatcag aagcagactgc ttcagaatgc ttcagaatgc ttcagaatgc ttcagaatgc ttcagaatgc
1561 agactacatct ctcctcattc gcgctacagc gttgaggttc agttgagaaa cggggccgctg
1621 cagaaagacttc acctgagactgc tggagctgctc ctgacgcctgc gcctctctctc ctctctctct
1681 ttcgctgtgc tggagctgctc ctgacgcctgc gcctctctctc ctctctctctc ctctctctctc
1741 tggataaaact ttcagatctc tacatctatag cagatctctc atctgagagc gctctctctctc ctctctctctc
ttctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc
ttctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc
ttctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc
1801 aaaaacctta ttcctggtttg aaggagctag tgaatgtPosition: 0
1861 agcgcacaaa gctctctctct gcgctacagc gttgaggttc agttgagaaa cggggccgctg
1921 gcggcttttg tggagctgctc ctgacgcctgc gcctctctctc ctctctctctc ctctctctctc
1981 caccataata gttgaaactc tggagctgctc ctgacgcctgc gcctctctctc ctctctctctc
2041 tccctcctta ggttcctcttc tattgcctttg cggaggtcttc ttcgctcagg ctcgctcgggg gctgagccgg cagtgcctcc
2101 gtttatccatg gggaggttttt gtttatttttt gttgataaaact ttcagatctc tacatctatag cagatctctc atctgagagc gctctctctctc ctctctctctc
ttctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc
ttctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc
2161 aagggatctt tcttgggac tgacagctcc tcgagggagtt ttcgctcagg ctcgctcgggg gctgagccgg cagtgcctcc
2221 aaaaactggtg caccctgca cattgctgca tggagctgctc ctgacgcctgc gcctctctctc ctctctctctc
2281 actgcttgtaa tcaaaaaactt gggaggttttt gtttatttttt gttgataaaact ttcagatctc tacatctatag cagatctctc atctgagagc gctctctctctc ctctctctctc
ttctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc
2341 aaggtcatttg ggttaggtgaa gttttatttt gttgataaaact ttcagatctc tacatctatag cagatctctc atctgagagc gctctctctctc ctctctctctc
ttctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc
2401 gccggagagt gggaggttttt gtttatttttt gttgataaaact ttcagatctc tacatctatag cagatctctc atctgagagc gctctctctctc ctctctctctc
ttctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc
2461 aagaaagggt gttttgttttg tggagctgctc ctgacgcctgc gcctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc
2521 cccataacagc gcggcttttg tggagctgctc ctgacgcctgc gcctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc ctctctctctc
2581 ctcggagaga ggaagatctc cgaagggcag aaggttggtttg cggaggtcttc ttcgctcagg ctcgctcgggg gctgagccgg cagtgcctcc
2641 caccctgca cagaggttcc ttaaactccac tcgcttggctc cgcgtttgtaa cggctgctcagg ctcgctcgggg gctgagccgg cagtgcctcc
2701 atatgctgata gttgtaaaact ctgagttttg ttcgctcagg ctcgctcgggg gctgagccgg cagtgcctcc
2761 ggtgttagct tcccaaaaatg aaacataagc caaccaggcc acaagaagaa gtagcaaaag
EED, such as human embryonic ectoderm development, is represented, for example, by accession numbers 209572_s_at, NM_003797.2 and NMJ52991.1 (2 alternative transcripts), or NP_003788(441 aa; gI24141020) and NP_694536 (400 aa; gI24041023). The protein sequence for variant 1 is:

```
MSEREVSTAPAGTDMPAAKKQKLSSDENSNPDLSGDENDDAVSI
ESGTNTERPDTPTNTPNAPGRKSWGKGKWKSKKCKYSFKCVNLSKEDHNQPLFGVQFN
WHSGEDPLVFATVGSRNLQVLQYSDADADENFYCAWTYDSNTSH
PLLAVAGSRGIIRIINPITMQCICYGHGNAINELKFHPDNPMLLVSXKDHARLW
NIQTDLVAIFGVEGHRDEVLSDADYLGEKIMSCGMDHSLKLWRINSKRMNM1KE
SYDYNPKNTRPISQIKHFPDFSTRD1HRNYVCVRWGLDLPSKSCENAIYCVWKG
```

SEQ ID NO: 28
The protein sequence for variant 2 is:

```
MSEREVSTAPAGTDMPAAKKQKLSSDENSNPDLSGDENDDAVSI
ESGTNTERPDTPTNTPNAPGRKSWGKGKWKSKKCKYSFKCVNSLKEDHNQPLFGVQFN
PLLAVAGSRGIIRIINPITMQCIKHYVGHGNAINELKFHPRDPNLLLSVSKDHALRLW
NIQTDLVAIFGGVEHRDEVLSADYDLLGEKIMSCGMDHSLKLWRINSKRMMNAIKE
```

SYDNPNKTNRPFISQKIHFPDFSTRDIHRNYVDCVRWLGLILSKSCENAIVCWKPG
KMDI1D1KPSESTWLGRFDYSQCDIWMRFSDQKMLAGNQYKVLYVMDE
VEDPHKAK (SEQ ID NO:29) and the mRNA sequence for variant 2 is:

1 ctagcagcgg gtcggagatc gaagagacgg gccaatgccc gccgtgaacgt ctttggaggg
61 aggaaggggg tgtggagagca ttcttcttggat ttctggcgctgtctcgagggct tgttgggagaa
121 gggagacata ctttaactct ccctcttcaatt ccacagcgacacacc ttcacactctgt tagactgtcggc
181 ggagggggcgg ggagagagagca cctctctcctttt cagagagagagagc ggaagagacacct gagaagagagagac
241 cgcggagagcg ccggagagagcg gcggagagagcg gcggagagagagc cagagagagagagc cgcggagagagcag
301 ggtcgagagcg cggagcgagcg cgccggagagcg cggagcgagcg cggagcgagcg cggagcgagcg cggagcgagcg
361 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
421 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
481 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
541 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
601 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
661 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
721 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
781 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
841 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
901 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
961 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1021 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1081 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1141 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1201 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1261 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1321 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1381 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1441 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1501 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1561 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1621 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1681 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1741 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1801 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1861 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1921 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
1981 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
2041 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
2101 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
2161 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
2221 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
2281 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
2341 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg
2401 cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg cggagcagcg

(SEQ ID NO:30)
PKP4, such as human plakophilin 4, is represented, for example, by accession numbers 201929_s_at, NM_001005476.1 and NM_003628.3 (2 alternative transcripts), or NP_001005476 (149aa;gi53829378) and NP_003619 (1149aa;gi53829374). The protein sequence for variant 1 is:

```plaintext
MPAPEQASLVEEGQPQTRQEAASTGPGMEPETTATTLASVKEQ
ELQFQLRTILEVERQIVASQLERCRLGAESPSTASSTSTEKSFPRWRSTDVPNTGVSK
PRVSDAVQPNNYLRTEQGQGTPQGQPQGSGVQGGSFHSNKSADNRQHSGFHSTNNHV
RVNRSRAEQTQLQPSVANARMSRVSPSAR
AQPSYVISTGVPSRGLRTSLGSGFGPSVTPDPRPLNPSAYSSTTLPAARASPSY
QRPASPTAIRGRSVTSTSQTNPGPTPQYTTARVGSPLTLDAQTRVASPSQGQVG
SSPKRSMTAVPQLHSPSLQRTVHDMEQFGQQYDIYERMVPPRSITLGTLRASYS
QHSQIGQDLSARSVPDHLTIPIEGRTYYSVPYRSPNHGTTVEQLGQSQTALYRTS
GVI
GNLQRSTQRSTLTYQRNNYALNTTAYAEFPRIQYRQVQECYNRLQHAVFADDGTT
RSPSIDTQSDPREAFRDPFEVPEHILQHFPSVQNAAAYLQHLCFDGNKVM
CLEGIHKLDDLHRVLNQRKNQGACNARLNLVFQGKSTDKNIAKMNVGGIPALLRLR
KSIDADEVRELTVGMLNLCSDAVKMTIIIRDALTILTNTVİPHSNGWNNSSFDH
KFPQTSVLNVTNTGCLNLSSAGGAEAKRMRQSCEGILVDSLVLVHTCVNSTDYS
KCVCLTLRNLSYRELVPQARLNLDDLIGKESKDPSECPGKWKKKKKRTTPQ
EDQMDGVGVPILGKLSPKGVEMLHPSWPYPLTIoAESSNPATLEGASGLQNSAG
NWKFAAYI1AAVRKDSGIPSLIVELLAMDNRDWS SVATAILRMLALDVNFLQIGY
RLDLVNLPGSNGPSVLSDEMAICCALHEVTSSKNMENAKALDSSGIEKLNITKR
GDRSLSKWKAAAQVNLNTLWYGDRSLYKQDGWFHVSTLERDFKHFPS\\S
TTNQMSPIIQSVGSTSSSSALLGIRDPRSEYDRTQPMPQYNSQGDATHKGLYPGSS
KPSPIYMSYSSPAEZQRNRLQHQQLYSQDDSONRNDFDAYRLYLVSPHSEYDFDY
RVHFPASTDYSTQYGKLSTTYNDFVSTKRPSRYAEQYPGPSDFS\\V (SEQ ID NO: 13) and the mRNA
sequence of variant 1 is:

```
1 tccggggctg agtccgctgc gacgccggcc gcggagcgcc caccatggcc aaggtgtaga
61 gggccaga ttg ccacccgcc ggcggccgggg ttggtgggag aggcgcctgc cgggccccggg
121 cccgtgctgggg agggcagcgc ggcgcgctgg gcactttttta attttttttagg
181 gtgcgcagcg gggacgccct gcggcgccgat gttctctgtag cttggagcag cgacgcgccc
gg
241 tggctactagt ggaaagagga atggccacgtc ctggagcagc cttcgtagtg gaggagggc
gg
301 aaccaacagac ccgcagcagga gtcgcttcttc ctgcccagcag gatggaacca gaggacacag
361 ccacacattact ttcgagctgcc gtagagccag tggccgctcct ctgtagtaga
421 aactggaggt ggaagggcag attgttgcca gtagctgtaga aatagtagaatt cttgagcag
481 aatcacaacag catgcgcagcc accacgtcaca ctggaagtc atttctgtag agatcaacag
541 acgtagcaca tacgtgtaga agcaacaccgt gtagctgcag cccccacact
601 atctcatcag gacagacgga gaaacagaaa cctctcttta accagacagc acatctccct
661 atgaaagtga gggtcatttg ggtaatctagg aaatggccac aacaatatgct tctttacctgc
721 acagtgcgata ccagggagca gcggagtttc acacacgtgc agaacgcagc aagggcagac
781 acagacagca gcttttcttc ataggatcag ctaacacacc tgtggtaggc aaccagagag
841 ctggaagcaca aaactggttg ccagccatcag tagccaatcgg gcggacagta
```
901 cagttccatc tagacacag tctcctttctt atgttatcag cacaggcgtg tctccttcaa
gggggtctct gagaacttct ctgggtagtg ... gaatacgata ggacccagcc acctatgcag tattacaata
gccaagggga tgccacacat aaaggcctgt accctggctc cagcaaacct tcaccaattt
2101
2041
tctcctttga ttgcttggct cagcagcacg gctgctggct cagcaaacct tcaccaattt
gagatgctct gtgggatacc ttcgaaacct atctggttga acagcattca ataaacaggct
tgatgatcat gcgacagcaa gcgaggggct cttcttttga ttgttacagg ttcacatgct
tcagcaccag tcagcatgca ctacgcggag tgggcaagac tgttcatgac aaccaccgcc
2521
tgatgatcat gcgacagcaa gcgaggggct cttcttttga ttgttacagg ttcacatgct
tcagcaccag tcagcatgca ctacgcggag tgggcaagac tgttcatgac aaccaccgcc
2401
tgatgatcat gcgacagcaa gcgaggggct cttcttttga ttgttacagg ttcacatgct
tcagcaccag tcagcatgca ctacgcggag tgggcaagac tgttcatgac aaccaccgcc
2341
tgatgatcat gcgacagcaa gcgaggggct cttcttttga ttgttacagg ttcacatgct
tcagcaccag tcagcatgca ctacgcggag tgggcaagac tgttcatgac aaccaccgcc
2281
tgatgatcat gcgacagcaa gcgaggggct cttcttttga ttgttacagg ttcacatgct
tcagcaccag tcagcatgca ctacgcggag tgggcaagac tgttcatgac aaccaccgcc
2221
tgatgatcat gcgacagcaa gcgaggggct cttcttttga ttgttacagg ttcacatgct
tcagcaccag tcagcatgca ctacgcggag tgggcaagac tgttcatgac aaccaccgcc
2161
tgatgatcat gcgacagcaa gcgaggggct cttcttttga ttgttacagg ttcacatgct
tcagcaccag tcagcatgca ctacgcggag tgggcaagac tgttcatgac aaccaccgcc
2101
tgatgatcat gcgacagcaa gcgaggggct cttcttttga ttgttacagg ttcacatgct
tcagcaccag tcagcatgca ctacgcggag tgggcaagac tgttcatgac aaccaccgcc
1921
tgatgatcat gcgacagcaa gcgaggggct cttcttttga ttgttacagg ttcacatgct
tcagcaccag tcagcatgca ctacgcggag tgggcaagac tgttcatgac aaccaccgcc
1861
tgatgatcat gcgacagcaa gcgaggggct cttcttttga ttgttacagg ttcacatgct
tcagcaccag tcagcatgca ctacgcggag tgggcaagac tgttcatgac aaccaccgcc
1801
tgatgatcat gcgacagcaa gcgaggggct cttcttttga ttgttacagg ttcacatgct
tcagcaccag tcagcatgca ctacgcggag tgggcaagac tgttcatgac aaccaccgcc
The protein sequence for variant 2 is:
MPAPEQASLVEEGQPQTRQEAASTGPGMEPETTATTILASVKEQELQFQRLTRELEVERQIVASQLERCRLGAESPSIASTSSTEKSFPWRSTDVPNTGVSKPRVSDAVPNYNLIRTEPEQGTLYSPEQTSLHESEGSLGNSRSSTQMNSYSDGYQEAQSFNSVNVSKADNRQHQSFQGSTHHWRSNRAEGQTLVQPSVANRAMRVSVPSTQGSPSYISTVSPSGRSLRTSLGGFGSPSVTDPRPLNPSAYSSTSSTLPAARASPSQRPASTAIIRGSVRSRQTNSPQPPTQYPPTARVGSPLITLADQTRVASPSQQVYGSSPKRSQMTAVPGHLSISQTVHMEQFGQQYDIYERMVPPRPDSLGLRYSYSAQHSQLOQDIRSAPLDDLHITPIEGYRITYSPVYRSNPHPGTVELOQSQTALYRTGVSIGNLQTTSRQSTLTYQRNNYLNTATYAEYRPIYRVQECNELQHAVPADDGITTPRPSDDSSQDPFRADEPPDEPEVIHLHMQFPSVQANAAAYLQLHLCFDKVMVCCRLGIGKLVDLLDDHRVELVQKNACGLRNLFVGBKSTDENKAMKNNVGIPAPAALLLRKSLDAEVRRELTVLWVNNCLSCDSDKMMIIRDAILSTTLTIVPHSGWGNNSFDDDKHICFQATELVRNRTTGGNLASSSQAEEARKMSCLEGLVDSLTYVIHTCVNTSDYSDKTVENCVTLRNLNLYRLEVPQAROLLQNLEDLLGKEPSXDSAESPSWGGKKKKKRTPRQEDCGVGIPPLGSKSRPGVLMCLPVSTSWKPTILAAESSNPATGSAQLSNLAGMKFAAYRAAVERKEGLPLVLLRMYDNDRXSVVATARNMALDVRENKELKYAMRLDNLPGPGNPSLSDTMAACICALHEVTSKMNENAKALADGIEKLVNTGKGRDSSLKWWKAAVLQNLTVQYRDLRSIYKKDGWNQNHFITPSVTLERDFKSHPSLSNTTNQMRPIQGSKSRPSIKYSSYSSPAREQNRRLHQOQLYSSDQSDNSKNFADYRL\(Y_{100}\)PSYEDPFYDFDVFPFPASTDYSTQYGLKSTNTNYDFYSTKRRSYRAEQYQPSD
\(S_{W}\) (SEQ ID NO:34) and the mRNA sequence for variant 2 is:
1 tccggggctg agttcgcgct gcacgcgggcc gcggagccgg gcaccatgggc aaggggtaga
61 gggcgaagtt ggccacccgcc gcgcgcggg ggttgtggag aacgcgttcgc gggcagcggg
121 ccggtgggggg agggaggggcc gcgcgcgggcc gccgcagcgc gccctttttta atcttttccggg
SSR1, such as human signal sequence receptor, alpha, is represented, for example, by accession numbers 200891_s_at, NM_003144.3, or NP_003135 (286aa; gll6904009). The protein sequence is:

MRLLPRLLLLL LVLFPATVLFRRGGPRGLLLAVADLTEDEETVED SIIEDEDDAEVEEDEPDVLEDVEEDEVSGEPEAEPSADTTILFVKGEDFPANNVK FLVGTNKGTEDFIVESLDASFRYPQDYQFYIQNPTAPLNTWPPQRTAFEYSFIP AEPMGRFGVLVINLYKDNLINGVFNQDFAVNQTVIEREDGLDEIFMYMFLAGLGL LLIVVGLHQLLESKRKRPIQVEMGTSSQNDVDMSPQQRTLNQINICASPRLPRKR AQKRSVGSDE (SEQ ID NO: 32) .

The mRNA sequence is:

1 cccacacctc cgcgcgtgagc aaggcagcagc gcggcggtcc cttgcgtgaaaa gagggcgtg
61 gcaaaagccgc tgcgcgtgctg ccctcgcgcgc cctgcgcgcgc gatcgagagc tgaagcaggtc
ttcggcgcgc ctcgagcacg cttgcgtgctg gacagcagcag cggagcacg cgtgcgtgctg
ttcggcgcgc ctcgagcacg cttgcgtgctg gacagcagcag cggagcacg cgtgcgtgctg
ttcggcgcgc ctcgagcacg cttgcgtgctg gacagcagcag cggagcacg cgtgcgtgctg
ttcggcgcgc ctcgagcacg cttgcgtgctg gacagcagcag cggagcacg cgtgcgtgctg
ttcggcgcgc ctcgagcacg cttgcgtgctg gacagcagcag cggagcacg cgtgcgtgctg
ttcggcgcgc ctcgagcacg cttgcgtgctg gacagcagcag cggagcacg cgtgcgtgctg
ttcggcgcgc ctcgagcacg cttgcgtgctg gacagcagcag cggagcacg cgtgcgtgctg
ttcggcgcgc ctcgagcacg cttgcgtgctg gacagcagcag cggagcacg cgtgcgtgctg
ttcggcgcgc ctcgagcacg cttgcgtgctg gacagcagcag cggagcacg cgtgcgtgctg
ttcggcgcgc ctcgagcacg cttgcgtgctg gacagcagcag cggagcacg cgtgcgtgctg
ttcggcgcgc ctcgagcacg cttgcgtgctg gacagcagcag cggagcacg cgtgcgtgctg
301 atgaagaac agtagaagat tccataattg agtagaagag tggagaactc
ggtgtagaag
361 aagastgaacc cacagatttg gtgaagatg aagaggaaga gatgtgctct ggtgaacctg
ggctttagta
421 aagctcacc gagatgcaag acaactaata cttgtttaaa aggagaagat ttctccagcaaa
gtgaagtcc
481 atacatatttgt gtaaggcttta ccaaacaaggg tcacagagat tttatgtgtg
gacatcactg
gtgaagctca
541 aacctcttga gtcctcctcg ctgttataac aggtacactca ttcttataac cagaatccaa
gttggttta
601 cagctcttcc tctgaaacgt gtgtgctcag cccagagaca gccaagctaat gactactctt
661 caactctctgc aagccagctgc ggtagatgcc cttatgggga cttctcctttc
721 aagatctggaa gcggactatg tgtcctcaaa cgtgacttac gcaagaactg acaaggtggt
gtaaatgtg
781 aagaagagga gtaggggttagt gggaagaaacaa aacttttatga ttatgtgtgtg gctgtggctt
841 ggctcctgttg tattgtggcc ctcctaaatgct tagaagagata tgaagagacca
901 taccagaaagt aagaaatgggt acaatcagttc ataatgtgat tgaactatgt gtaggacctc
gtaggttactt
961 aggaaacatt cactaatcaac aattaagtctt cactcaagcag gttggcagc acagggccac
1021 aagaagagatc agtgggctct gtggtagaag tggtagccttgc tgcagcacaa aacagctttac
1081 ttacctctggc ctaaatatttt tcggcctgat ggaagatttg gcgaagacag catgtgcacaaca
gtaggggtgaa
1141 tagaaagcoca ctctcactga gttgtgttcg gacacattg gctgctgtgctg gatataatcg
1201 ctagaaaatgc aacgcaatct caatttttcaaa gtaaacaatt ttggttttttttt ttaaacacctt
1261 aaatactatg tggtagtgttg tgtatatgtg tagagcagtct gctctcagcag aacattctgtt
1321 gaatctacctg aacaagctct ttgaataact ataatataatc tgtggtgtctc tctctcttttc
1381 acatttttctg attttttttcag cccaaactca gtttaattttc attattatgtg
1441 tcctgccttg aaacattttaa aagaaacaagtt tttaattggaat gcctcataatt ctaattgtgat
1501 gcacaacatac gcattttttag gttgttgactg tatataaatg aggttaacag tgggagccag
1561 aagaatagact ctttttggag ctttttgcttg gcacagtaatg cggccttttaa cttttttaatct
1621 tcctcagaaa aaccttccagt gcctctcaag aatggtggga ttctcttggtg gaacataattt
1681 gaacatacag atctctggctga actttatatg atgaaaccta atacctcccct ctttttttttct
1741 aaaaaagac tagtcctttg gtgaanattg tgttggtgtg tccatttctg ctcagtctctt
1801 gcaccgagg tagtcctgga cttgtagttag cactacctcag ccagatgtgtgt ctagttttcg
1861 tttttttagtt tgtagagagata atttttttctc cccataataa gagacatcttt ttaaaagagaa
1921 gttgcctgag gcaacaccat aacgagata ccaaggtgcttg gtttttagga tttttgcttg
1981 ttacccctaa tcaagttaag ctgtgttgat cattttttttttt aacgcatgtta ctattaatttta
2041 ctctaccaat ctctttgggtag ttttggcttg tgggataaca ttttttgtta acactgaagaag
2101 tccaaaaaat tgttctcaag tgcttcagac gggaacaaaa ataaaaagtt ccctctctctaca
2161 atcccccttattgtagct ttggccctagag tggtaaattta cttgtgggcatt ctaattgtgat
2221 cttggtgtaag ggtaagtaatt aataatgttac acacattaaag cccctctctgcc ctcgcttttct
2281 taagccatat ctcctctactg ttatattata aagaaatgatg aaagctaatt gttagctgctc
2341 ttaatttcct gacctctctg ttctctcttt tgcgatacaac tccctcagctc cdcgctccccc
2401 acagacaaaa aacaaaacag aataacagact ggacaagaacag ttttttcaccc aagagggact
2461 ttttaaacatc tgagccaaag cagatatag gtaatttgata atttgggtata actctcatttga agttgtg
2521 tagggcgctt caaatattcct atagttgctta tgactttgata taaactctact atctgtgataa
2581 taagtggttt gatgtttaatt tttggggcttg tggtagctga tttgggttta aggaaagttt
2641 atagcgctta ttagaggcttga aggggttttgg ttttgggttta tattgagagg ctctcctcattt
2701 cccagagatt aatcctcaatc acattttttttt ggttagctgaaaa cttggtgtaatt ctctagggcc
2761 atgatacagg taagttgtttc gccagagata attgtgcttt caccagttta aagttttttaa
2821 atggaatacc tgggtctaca gtagcagctc gcaaaaccat ctactggtgtc actcgtcagcc
2881 cctctcctatt cttaataaacc ttgcccttaac tacagcctttt cccatctctgta aagttcataatc
5581 ggcattaaaa  agaaatatta  gtatgacctg  aagaaagata  atctctgttt  taagtctcca
5641 aattgcagtc  aaagatgttt  ...  aagctgctgg  ccctttgccc  ctccctgtg  ggaatcagat
8161 ctagaggagg  ctgagcctgc  agacacagca  gtggccaaaa  ggtcactcta  agtgttttgt
USP5, such as human ubiquitin specific peptidase 5 (isopeptidase T), is represented, for example, by accession numbers 206031_s_at, NM_001098536.1 and NM_003481.2 (2 alternative transcripts), or NP_001092006(858aa;gI148727331) and NP_003472(835aa;gI148727247). The protein sequence for variant 1 is:

MAELSEEALLSVLPTIRVPKAGDRVHKDECAFSFDTPESEGGLYICMNTFLGFGKQYVERHFNKTGQRVYLHLRRTRRPKEEDPATGTGDPPRKKPTRLAIGVEGGFDLSEEKFELDEDVKIVILPDYLEIAKDGLGGLPDIVRDRVTSAVEALLSADSRKQEVQWADGEVRQVSKHAFSLKQLDNPARI PPCGKCSKCDMRLWLNLTDGS LGCGRRYFDGGNNHAVEHYRETGYPLAVKLGTITPDGADVYSYDEDDMVLDPSIAHLSHFRGIDMLMKQRTDMTEILDMNQIGEWEQKFGVLKPFGPVTGIRNLGNCYLNSGWQFISYPDFOQYVKEFKEFQNAFTPDFTIQVAKHLGGLGSEYKSPVFESGGERVPEQEKEQVQIQWQFMKALIGKHPEFSTNRQODAQEFFHLHLMVERNCRSENPNVEVPLVEKEIKCLTEKVRQVRVYMIQLVPMDAALNKEELYYEKKRQAEEKMALPEVQVFSSCLEAYGAEQVVDIFWSTALQASKVAKTTRFSFP
and the mRNA sequence for variant 1 is:

1  aggggagggg  actcggaacg  gtgggaacg  gtgggaacg  aagagcgtgc  cccggtgctg  cgcctggtgc
61  tggccgagct  gagttgaggc  gcgcgagcct  gcgcgagcct  tcctgctgct  gcgcgagcct
121  ctgcgcctca  ggcacggcag  gcgcgagcct  gcgcgagcct  gcgcgagcct  gcgcgagcct
181  ttcctgctca  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
241  gttttccgct  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
301  tccctgcttc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
361  ttgggagttc  gacatgctgc  gacatgctgc  gacatgctgc  gacatgctgc  gacatgctgc
421  agtagctgat  ttcctgcttc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
481  accttgccat  gacatgctgc  gacatgctgc  gacatgctgc  gacatgctgc  gacatgctgc
541  ccctgcctt  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
601  tccctgcttc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
661  ggtgcgcttc  ccctgcttc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
721  ctgctcttcc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
781  gttgggtcct  ttcctgcttc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
841  atgactagtg  ttcctgcttc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
901  tcgcctgctt  gacatgctgc  gacatgctgc  gacatgctgc  gacatgctgc  gacatgctgc
961  accagcggat  ttgctgcttc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1021  tgcctctctc  accagcggat  ttgctgcttc  cccctgcctc  cccctgcctc  cccctgcctc
1081  acctctgttc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1141  gcgcctgttg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1201  ccctgttgcc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1261  ggcgtgctgg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1321  tccgctgctg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1381  tgcctgtgtc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1441  tggcctgttc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1501  ccctgtgttc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1561  aggctgggtg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1621  acgctgggtg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1681  agcctgggtg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1741  gttttcgctg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1801  actgtggctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1861  agttggaggg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1921  tcgtattaat  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
1981  gttgggcttt  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
2041  gcggcgtgctg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
2101  gatgggcttg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
2161  agctgctgctg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
2221  actgtgctgctg  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc  cccctgcctc
The protein sequence for variant 2 is:

MAELSEEALLSVLPTIRVPKAGDRVHKDECAFSFDTPESEGGLY
ICMNTFLGFKQYVERHFNKTQGRVYLHLLRTRFPKEEPATGDPPRRKPKTRLAIG
VEGGFDLSERKFELEDVKVILPYLIEARGLGGLGLPDIVRVRSTSAEVALSADSA
SRKQEVDQAVDGEVRQVKHFSFLKQLDNIPARPPCGWKSCKCDMRENWNLNLTDSGL
CGRYYFDGSQNGHHVEHYRETGYPGLAVKLGTTIPGDAGVDYSEDMDMVLPLEAL
SHFGIDMLMOKXTDKTMTLEIDMNQRIEGWELIQEGSVPLLFPFGYTGIRNLNGLSN
CYLNSWQVLFSPDFQFRYDKLEKIFQNPNTDPTQDFTSQVAKLHGHLSSGEYKSP
VPESGDGERPEQKERVQDIAPRMPFKALIGKHPFESTNRQDAEFFHLLINVNER
CRSSEPNEVFRLVEEKICLATEKVKYTQRVDYMQLVPVMDAALNKEELLEYEEXK
KRAAEEMALPELVRAPVFSSCLEAYGAPEQVDDFWSTLAQSKSVAATKTRFASFP
DLYVPIJQKFTGFLWVPKLDVSIEMPEELDISQLRGTLQPEEELPDIAAPLVTVP
DEPKAPMLESVDLSQVEMFGMPHMAKCRKAVYTGNSGAEAMNWVMSHDMPDFANPL
ILPSSGGPGSTSSAADPPEDCVTTIVSMSGFSDQALKALRANTNLSERADWIFSHI
DDLDAEAMDIESGRSAADSSEPVPKVRKDGPQYQLFAFISHMTGSTMCGHYVCC
HIIKKEGRWV1YNDQKVCASEKPPKDGL1YFQYRVAS (SEQ ID NO: 33) and the mRNA sequence for variant 2 is:

1  gggaggattgg aacggtggga gcgccgtgtt gttgagaagc gcgtgaggtg gctcatggcg
61  agctgatgta ggaggctggct tcttcagat taccgagcat ccggtctcct aagctggag
121  acgcctctca caaaagcagag tcgccctctct cctcgacac gcccaggtg ggggggcc
181  tctccatctct tatgaacacg ttctcggact ttggagaaca gctatggag cagacattta
241  ataagcctgc ccagcgagtc tacctgccca ccgccgcgac ccggccggcc aggagaggag
301  acgcctctct cctgcctttc cttctgcttt gcgcgctcag ccatggcgct ccatggcgct
c361  ttgaaggcgg atttgacccc aagcagggag aagtttgagg aggagaggag gcatgggctg
421  tcattttgcc agatccagct ggagtttccc ggaggagacct gggggacgc gctgacattc
c481  tcagagctct ggtaccaact gcagtggccc gcgcgtgac gcgcgtgct gcgcgtgct
541  agcagaggtt ccagcgattg gatgagggag tacggcagtt gcgggccag
601 tcaagcagt tggacaaccct gtcgcaatcc ctccctgagg ctggaagtgc tccaagtgtg
tcaagcagtt ggacaaccct gctcgaatcc ctccctgtgg ctggaagtgc tccaagtgtg
661 acatgagaga gaacctgtgg ctcaacctga ctgatggctc ... tctcttcctt
gtccttctct ggaaaatgcc aaaatacacg atgtgaataa aagtacaacg gctaaaaaaa
3061 aaaaaaaaaa (SEQ ID NO: 34) .
ACTB, such as human beta actin, is represented, for example, by accession numbers 213867_x_at, NM_001092 (375aa;gI4501885). The protein sequence is:

```
MDDDIAALWDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMGQKDSVQSKRGLTKYPIEHHGIVTNWDDMEKIHWHFTYNELRVAEEHPVLLTEALPNPKANREKTMQETFNPFTNPAMYAVSLYSAGRTTGVMDSGDGTHTVPIYEYALPHAILRLDLARDLTDYLMKILTERGYSFTTTAEREIVRDIEKLYVALDFEQEMATAAASSSLEKSYELPDGQVITINGERFRCPEALQFPSFLGMSCHGEHTTFNSIMKCDVIRDLYANTVLSGTTMYGIDIARMQKIETALAPSTMKIKIIAPE
```

The mRNA sequence is:

```
1 acgcgccgaga cgcgcctcgc cccgcgcaga cagagcctcg cctttgccga tccgccgccc  
6 gtccacaccc ggcgcagcgt cccatggat gatgatatcg ccgcgctcgt cgtcgacaac  
121 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
181 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
241 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
301 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
361 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
421 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
481 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
541 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
601 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
661 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
721 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
781 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
841 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
901 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
961 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1021 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1081 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1141 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1201 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1261 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1321 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1381 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1441 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1501 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1561 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1621 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1681 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1741 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca  
1801 tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca
```

(SEQ ID NO: 19) and the sequence is:

```
acgccgaga cgcgcctcgc cccgcgcaga cagagcctcg cctttgccga tccgccgccc
gtccacaccc ggcgcagcgt cccatggat gatgatatcg ccgcgctcgt cgtcgacaac
tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca
tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca
tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca
tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca
tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca
tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca
tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca
tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca
tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca
tgcggagcca gacccagacca gacccagacca gacccagacca gacccagacca gacccagacca
```
HLCS, such as human holocarboxylase synthetase (biotin-(propionyl-CoA-carboxylase (ATP-hydrolysing)) ligase), is represented, for example, by accession numbers 209399_at, NM_00041.5 and NM_00 1242785 and NM_00 1242784 (three alt transcripts), or NP_000402 (726aa; gI46255045) and NP_001229713 (726aa; gI338753397) and NP_001229714 (726aa; gI338753400). The protein sequence is:

MEDRLHMDLVPQKIVSVKIDSTLKEVKDQVSNKQAQILEPK
PEPSEIKEQDDMEVGRDPKLGAPEEFQKQRGSASGSEPAGDSRGGPGVEYHLH
LSSCHECLENSTIESVKPASANIPDLPYDDSSLESVADESTPEREGRRVKNLTK
APNILLVGSQEALGRHEVRSYLADCVDSYLYHLLEDALSRLPDWTDNCLL
LIARESIPEDLYKFPMLYSQGGKLGLSSSTPGFQVTSKGLHXTQNLVF
KQVADQSEVKLCSVLSSGCRCYQEGVPRLSPGRQLCHLENEDKRMIVHVFPFTGGEAVL
QVHLELPFSNIVQTPDFNLLKSSFRYEVILRTTLGLSGDMQVPAFLPLYLLSA
AIEIRDPWMLKGVHDSVSEGEISGQLLSRFVSVYSEVEITPSICPWNTNEMAFSSE
FNLEIYRLQNLTKQVLIFAEVPTTMRLGEDLMFQTQEMGLIGLIVAARQTEKGR
GGNWLSPVGCALSTLLISIPLASQLGQRPFPVQLMSVAI贩卖RSIIPYEYQDINL
RVKWPNIIYSSLMKIGVVLVSNLTGETFYILCGGFNTSNPTICINDLITEYNDK
HKAELKPLRADYLAARWTVLEKL KIEFOQDGKPSNVLPPYRYVHSQOVQVHSAG
EAPGGVKSIVGLDSSQFLQVHQEGGTWedTVHPDGSNDFMLNRLILLPKRRR (SEQ ID NO: 21) and the mRNA sequence is:

1 acgagagggc gcagcgcggcg ggcgggcggcg ttccgccggcg accacccggc gcaggagcgg
61 ccgcgctttcg gcctcagact ccattggaag ttggcctccat tcttctgtgtc
121 tccttttata cttgaagctgc aacacatgac atttgtgacg gagacgcatct gggtcaaacg
181 tgagattttta cagagatcat ccctctcagta acacagctgg cctttgttgtc
241 tggagctttct catgaggtta cttggcctctt aagttcaggt atttgagcagc cataagacag
301 ccctctgtgac cccagcacatt tattgtgatg cccctttttt agctcctcagc cactcctcag
361 atattgcaagt gcctcagactg tgtgctttacg ttagctgtca gacctgggacg tccttctcagc
421 ctaattgctgt aagcaagtgt ggcaactcctt agcaagcttg ggtctggcttg cttggaggtg
481 agactcccaac cggctctttgg tcttggagat gctcctcagc ccttcctctc acacctcctcagc
gcccgttgag cagcaagatg cccaggggag ggcgcctcttg cagcttctact caactcctcatcg
781 tgtctcctgt gcagacagtga cggagagggg ggcgcctcttg cagcttctact caactcctcatcg
841 ttggcccggag atccctccag atcccttctt gattaggttt gacagctttg cttggaggttg
901 ggatgacccaa cggctctttgg tcttggagat gctcctcagc ccttcctctc acacctcctcagc
gcccgttgag cagcaagatg cccaggggag ggcgcctcttg cagcttctact caactcctcatcg
961 atcctctcct ctattgtggctcg gcactccgcag gacagccttgc gcggcccttg cagatgctggc
ttggcccggag atccctccag atcccttctt gattaggttt gacagctttg cttggaggttg
1021 ttggcccggag atccctccag atcccttctt gattaggttt gacagctttg cttggaggttg
1081 agtgtcttcat gaggcccttg gcgcgctcagc ggacaggact gcctaccactt ccctcgccag
1141 ttcctcctcg aagccttctcc ccaagctctgc ccaactcctctcc ccaactcctctcc ccaactcctctcc
1201 cttggcctctg ctgctcctctc ccaactcctcc ccaactcctctcc ccaactcctctcc ccaactcctctcc
1261 ccaagacag cttgcatcct gtttttttttt ccaagcctgac gacacagctc ggtggtggggt
gtcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc

tgcttgagca gtgcctgcaag tatacagggaa ggccccgtcc ggctcagccc cgcaggctc cagggccacc tggagaatga ggacaaggac...gccttttcca tgtccctcct atctctgttg aacttatgaa tcatactcat tacttttcag ctttttaaaa ggccaatttt tgtccagttt.tctctcttcc
NDUFB1, such as human NADH dehydrogenase (ubiquinone) I beta subcomplex, 1, 7kDa, is represented, for example, by accession numbers 206790_s_at, NM_004545.3, or NP_38569473 (105aa;gI38569473). The protein sequence is:

```
MICWRHPSAPCGRGEWQVPRSQLPLARVEFPVALGLGVAVGAEA
AAIMWLLQIVI^HWVHVLVPMGFVIGCYLDRKSDERLTAFRNKSMLFKRELQPSEEVTWK
```

and the RNA sequence is:

```
1 agaggagtca accctgagga ggaaaaagta gtatgatttg ctggcgtcac ccctctgctc
```
CDC42, such as human cell division cycle 42 (GTP binding protein, 25kDa), is represented, for example, by accession numbers 208728_s_at, NM_00103489(191aa;89903012) and NM_001782(191aa;gI4757952) and NP_426359(191aa;gI16357472). The protein sequence or variant 1 is:

```
MQTIKCVWGDGAVGKTCLLISYTTNKFPSEYVPTVFDNYAVTV
MIGGEPYTLGLFDTAAQEDYDLRLPILYPTQDVFVCPSWSPSFSFENVKEKWPEIT
HRCPKTPFFLVLGTVQDIIIDPSTIEKLANKKQKIPITPETAEKLARDLKAVKYVEC
EppEpKSSRRCVL, (SEQ ID NO: 25) and the mRNA for variant 1 is:
```
The protein sequence or variant 2 is:

MQTIKCWGDGAVKTCLLISYTTNKFPSEYVPTVFDNYAVTV
MGIGEYTLGLFDTAQGQDYDLRPLPS YQQTVDFVLCFS VSPPSSFE[NXKEMNPEIT
HHCPTTPFILYGTQIDLRRDSPTEIELAKNNKQBPPTETAELRDLRAVKYVECSAL
TQRLKMNVEDAIAAELEPPETQPKRKKCIF (SEQ ID NO: 35) and the mRNA sequence for
variant 2 is:

1 acttccgcgg gcaccccaact gtgcgtctcc tgccgctgta cgtcaggtgc gtgcctctgt
6 cccgcaagcc agagaccccc ggcgcagctg gccacagccc ccgtggagaa gcgtaggctca
12 tcacagatt tgaatatattt aaagtggata ccaaatatt tcaagactg agacatatta
18 gtgtggtgttt gttggtggtgg tggctgttgg caaaatattt taatgatatg tcacacacac
24 ccactctaat ccatcgcggg atgtacccag tcgtcttttg acaactgtgc ctgagctttg
30 gatttgggta gaaccataata cctttgacct ttttacactg gcagggcaag agattatag
36 cagattacga cccgctgagtt atccacaaac agatgtattt ctaaactgc gaaatctttg
42 ctcaccatct cttatgaaag acgtgaagaag aaagtggtggc ccttgagataa cttaaccctg
48 tcctactacat cctttcgagg gcctctgtcag cttaattgtct ccagagctga acccctctac
54 tattgagaaga cttgcccaaga aaccaacagaa gcctatcaact ccagagactg ctgaaaagct
60 ggccccgtac gtcgaagggc tcgaaggtgtt gccacccgctc gcctgtatat ggctgtgtgct
66 gcagaaatgtc tttatttttc gcctttgagtg gtctccctgg cttccccaga ctttaaccac
72 aaggaagtgc tttatttttc gcctttgagtg gtctccctgg cttccccaga ctttaaccac
78 cccgctactgt tagaagacgct gtttacccat ctcgagttctt gtcctgtctg tggctgtctg
84 tctggtgttt tttcttcatct caccctagag cccttctgta tatgtttttg atcagaaaaa
90 cattctacatt aaaaattttt ccctgctgat cttcagtttt ccagagctac ccagagctac
96 gcttgagattc ccttttacta agaggttttt gcgtttgatt acttattttaa aagggacttt
102 gctctgtgctt aaagatgtct cttcgtgttg cttcctgtct ttttttcttt gtcctttttaa
108 gctcttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt
114 ttataattgtc atgggagggc ttaacctttt ttttttttttt ttttttttttt ttttttttttt
120 gatttttttt ccttattttt ttttttattt ttttttattt ttttttattt ttttttattt
126 gttgttgatg atgaggttttt ggtgggagttt gggaggtttt gggaggtttt gggaggtttt
132 aacggctttgatgatttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt
138 gtcgctatct tggccactgta gactcggaca ttttttttttt ttttttttttt ttttttttttt
144 attttattttt gtcagggggact cttcttttccc cttcttttccc cttcttttccc cttcttttccc

MQTIKCWGDGAVKTCLLISYTTNKFPSEYVPTVFDNYAVTV
The protein sequence or variant 3 is:
MQTIKVGDGAVGKTCMLSYYTNKSFSEYPTVFDNYAVTV

The mRNA sequence for variant 3 is:
1 acttccgcgg gcacccacact gcgctgcctcc tgcgcctgtga ctgcaggtgc gtcgccttctgtggaagagacctgcgcagtgctgccaaacgcggtggagaa
gctgagacgg tgttgccagtgcctgcagttcaagtggctgcatctctctggtgatttggctcctgagttttcatttttcggccagagtatagtggagaatcgctccag
tggacttttttgcataaatgcaggagttgattcgtcactctccagttcctagatctagtttagaaacatgtttcgcagataactcagagactgctgacgcatctactattgacacgcggcagagtcctgcgaccagtttgattgctgactaatgttatttgacatcttaccaactcgactcatttcctggtgatactaaacaggtgttttgg
ttttgcaatgaccttcagtttttttttttntgtttttttaaaaagttgtgtgtgttttttgygggtttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Multiple primers and probes can be designed based on these known sequences and are encompassed within the disclosure. Primers can be designed in accord with a number of criteria using Primer design programs such as Premier Primer (biosoft), Oligo Primer Analysis software, and Oligo Perfect (Life Technologies) and other free and commercially available software. Probes can be designed using free and commercially available software including Array Designer (biosoft), and Light Cycler Probe design software (Roche). Primers and/or probes can be detectably labeled in accord with standard methods. Probes can be attached to a solid surface such as a slide, a well in a multiwell plate, and/or a chip. In embodiments, primers and/or probes are designed to specifically bind to each of the nucleic acids encoding CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, EED, SSR1, USP5, NDUFB1, OSBPL8, and PKP4. In embodiments, a custom array can be prepared that contains no more than 200 probes, including at least 12 probes, one for each of the identified genes. In embodiments, the primers or probes are not designed to bind to the polyA tail.

In embodiments, the primers and/or probes specifically bind to the nucleic acid sequences under standard PCR or microarray conditions. In embodiments, those conditions include 7% sodium dodecyl sulfate SDS, 0.5 M NaP04, 1 mM EDTA at 50°C with washing in 2X standard saline citrate (SSC), 0.1% SDS at 50°C; preferably in 7% (SDS), 0.5 M NaP04, 1 mM EDTA at 50°C. with washing in 1X SSC, 0.1% SDS at 50°C; preferably 7% SDS, 0.5 M NaP04, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C; and more preferably in 7% SDS, 0.5 M NaP04, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

Each of the genes identified herein as useful in determining a short term or long term survivor can have one or more variants that are known and primers and probes can be designed to detect all variants and/or each variant. Variants include those nucleic acids or proteins that are "Substantially homologous nucleic acid sequence" or "substantially identical nucleic acid sequence" "substantially homologous amino acid sequences" or "substantially identical amino acid sequences".

"Homologous" as used herein, refers to the subunit sequence similarity between two polymeric molecules, e.g., between two nucleic acid molecules, e.g., two DNA molecules or two RNA molecules, or between two polypeptide molecules. When a subunit position in both of the two molecules is occupied by the same monomeric subunit, e.g., if a position in
each of two DNA molecules is occupied by adenine, then they are homologous at that position. The homology between two sequences is a direct function of the number of matching or homologous positions, e.g., if half (e.g., five positions in a polymer ten subunits in length) of the positions in two compound sequences are homologous then the two sequences are 50% homologous, if 90% of the positions, e.g., 9 of 10, are matched or homologous, the two sequences share 90% homology. By way of example, the DNA sequences 3'ATTGCC5' and 3'TATG GC share 50% homology.

As used herein, "homology" is used synonymously with "identity."

The determination of percent identity between two nucleotide or amino acid sequences can be accomplished using a mathematical algorithm. For example, a mathematical algorithm useful for comparing two sequences is the algorithm of Karlin and Altschul (1990), modified as in Karlin and Altschul (1993). This algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al., and can be accessed, for example at the National Center for Biotechnology Information (NCBI) world wide web site. BLAST nucleotide searches can be performed with the NBLAST program (designated "blastn" at the NCBI web site), using the following parameters: gap penalty = 5; gap extension penalty = 2; mismatch penalty = 3; match reward = 1; expectation value 10.0; and word size = 11 to obtain nucleotide sequences homologous to a nucleic acid described herein. BLAST protein searches can be performed with the XBLAST program (designated "blastp" at the NCBI web site) or the NCBI "blastp" program, using the following parameters: expectation value 10.0, BLOSUM62 scoring matrix to obtain amino acid sequences homologous to a protein molecule described herein. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. Alternatively, PSI-Blast or PHI-Blast can be used to perform an iterated search which detects distant relationships between molecules and relationships between molecules which share a common pattern. When utilizing BLAST, Gapped BLAST, PSI-Blast, and PHI-Blast programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically exact matches are counted.

As used herein, a "substantially homologous amino acid sequences" or "substantially identical amino acid sequences" includes those amino acid sequences which have at least about 92%, or at least about 95% homology or identity, including at least about 96%
homology or identity, including at least about 97% homology or identity, including at least
about 98% homology or identity, and at least about 99% or more homology or identity to an
amino acid sequence of a reference antibody chain. Amino acid sequence similarity or
identity can be computed by using the BLASTP and TBLASTN programs which employ the
BLAST (basic local alignment search tool) 2.0.14 algorithm. The default settings used for
these programs are suitable for determining substantial similarity of nucleic acid sequences
for purposes of the present invention.

As used herein, the term "conservative amino acid substitution" is defined herein as
an amino acid exchange within one of the following five groups:

I. Small aliphatic, nonpolar or slightly polar residues:
   Ala, Ser, Thr, Pro, Gly;
II. Polar, negatively charged residues and their amides:
    Asp, Asn, Glu, Gin;
III. Polar, positively charged residues:
     His, Arg, Lys;
IV. Large, aliphatic, nonpolar residues:
     Met Leu, He, Val, Cys
V. Large, aromatic residues:
    Phe, Tyr, Tip

"Substantially homologous nucleic acid sequence" or "substantially identical nucleic
acid sequence" means a nucleic acid sequence corresponding to a reference nucleic acid
sequence wherein the corresponding sequence encodes a peptide having substantially the
same structure and function as the peptide encoded by the reference nucleic acid sequence;
e.g., where only changes in amino acids not significantly affecting the peptide function occur.
In one embodiment, the substantially identical nucleic acid sequence encodes the peptide
encoded by the reference nucleic acid sequence. The percentage of identity between the
substantially similar nucleic acid sequence and the reference nucleic acid sequence is at least
about 50%, 65%, 75%, 85%, 92%, 95%, 99% or more. Substantial identity of nucleic acid
sequences can be determined by comparing the sequence identity of two sequences, for
example by physical/chemical methods (i.e., hybridization) or by sequence alignment via
computer algorithm.

Suitable computer algorithms to determine substantial similarity between two
nucleic acid sequences include, GCS program package The default settings provided with
these programs are suitable for determining substantial similarity of nucleic acid sequences

54
for purposes of the present invention.

**Determination of Expression Levels**

In one embodiment, the expression of the nucleic acid, such as mRNA of the genes of interest is determined. Levels of mRNA can be quantitatively measured by Northern blotting. A sample of RNA is separated on an agarose gel and hybridized to a radio-labeled RNA probe that is complementary to the target sequence. The radio-labeled RNA is then detected by an autoradiograph.

Another approach for measuring mRNA abundance is an amplification reaction such as polymerase chain reaction (PCR). In embodiments, PCR is RT-PCR. In RT-PCR, a DNA template from the mRNA is generated by reverse transcription, which is called cDNA. This cDNA template is then used for qPCR where the change in fluorescence of a probe changes as the DNA amplification process progresses. With a standard curve qPCR can produce an absolute measurement such as number of copies of mRNA, typically in units of copies per nanolitre of homogenized tissue or copies per cell. qPCR is very sensitive (detection of a single mRNA molecule is possible).

Another approach is to individually tag single mRNA molecules with fluorescent barcodes (nanostrings), which can be detected one-by-one and counted for direct digital quantification (Krassen Dimitrov, NanoString Technologies).

Also, DNA microarrays can be used to determine the transcript levels for many genes at once (expression profiling). Recent advances in microarray technology allow for the quantification, on a single array, of transcript levels for every known gene in several organism's genomes, including humans or smaller custom arrays can be utilized.

Also, "tag based" technologies like Serial analysis of gene expression (SAGE), which can provide a relative measure of the cellular concentration of different mRNAs, can be used.

In other embodiments, the level of expression can be determined using RNA sequencing technology. RNA sequencing technology involves high throughput sequencing of cDNA. mRNA is isolated and reverse transcribed to form a library of cDNA. The cDNA is fragmented to a specific size and optionally may be detectably labeled. The fragments are sequenced and the full sequence is assembled in accord with different platforms such as provided by Illumina, 454 Sequencing or SOLID sequencing. In addition, mRNA can be sequenced directly (without conversion to cDNA) using protocols available from Helicos.

In one embodiment, the expression of the protein from the genes of interest is determined. For genes encoding proteins the expression level can be directly assessed by a number of means with some clear analogies to the techniques for mRNA quantification.
The most commonly used method is to perform a Western blot against the protein of interest - this gives information on the size of the protein in addition to its identity. A sample (often cellular lysate) is separated on a polyacrylamide gel, transferred to a membrane and then probed with an antibody to the protein of interest. Other methods include, for example, Enzyme-linked immunosorbent assay (ELISA), lateral flow test, latex agglutination, other forms of immunochromatography, western blot, and/or magnetic immunoassay.

The use of the word "detect" and its grammatical variants refers to measurement of the species without quantification, whereas use of the word "determine” or "measure” with their grammatical variants are meant to refer to measurement of the species with quantification. The terms "detect" and "identify” are used interchangeably herein.

Reagents to detect the molecules of interest (such as a mRNA, cDNA, a nucleic probe, or antibodies) can be produced by methods available to an art worker or purchased commercially.

Analytical Methods

In embodiments, a method for selecting a treatment of a subject with ovarian cancer comprises inputting the expression levels of the set of genes into a function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer to obtain an output score. In one embodiment, the gene expression analysis of the genes of interest is applied to the equations provided in Figure 9. In embodiments, the gene expression analysis is obtained from microarray analysis using an Affymetrix U133 chip and the data from each gene is produced using the original raw intensity data (CEL files) processed using the MAS5 normalization and background-correction algorithm (510K FDA approved).

If gene expression analysis is conducted using PCR or RNA sequencing the gene expression values can be converted to values of the Affymetrix gene expression analysis algorithm using known methods. For example, the gene expression analysis can be run in parallel using PCR or RNA sequencing and using the Affymetrix U133 chip and software. The gene expression values for each gene from PCR or RNA sequencing can be compared to the values generated using Affymetrix system and a conversion factor identified. Gene expression levels for each gene generated by PCR or RNA sequencing can be generated and converted to the output of the Affymetrix algorithm using the conversion factor before inputting gene expression levels for each gene into the functions.

The variables are defined as the expression of the below genes:

Variables:
The functions are those as presented in Figure 9.

\[
F_1 = f (LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, PKP4)
\]
\[
F_2 = f (SSR1, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L, EED)
\]
\[
F_3 = f (CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, EED).
\]

Each function operates to independently provide a risk assessment of whether the subject is likely to have long term or short term survival. One or more functions can be used together to determine the likelihood that a subject has a risk of short term or long term survival. Once the gene expression data is inputted into a function. A function provides an output score for the subject’s sample.

**Cutoff values**

In embodiments a method for selecting a treatment for a patient having ovarian cancer comprises determining whether the subject is likely to have long term survival by determining if the output score is less than a cutoff value or whether the subject is likely to have short term survival by determining if the output score is greater than or equal to the cutoff value, wherein the cutoff value is a value determined by identifying a value between the 99% confidence interval of the mean output score of a first set of samples from subjects known to have short term survival and the 99% confidence interval of the mean output score of a second set of samples from subjects known to have long term survival. The disclosure also provides methods for determining a cutoff value.

In embodiments, the method for determining the cutoff value comprises determining a mean output score for a first group of patients that are known to have short term survival and a mean output score of a second group of patients known to have long term survival of an original set of patients. The mean output score, the standard deviation, the range of each
group, and 99% confidence interval of each group is determined. A cutoff value is determined that falls between the 99% confidence interval for both groups. For example, the cut-off score of the F1 model was determined to be 21.388. The upper limit of the 99% confidence interval for the long term survivors was 20.663 and the lower limit for the 99% confidence interval for the short term responders was 22.924. The difference between the two groups is 2.261 and in one embodiment, this value is divided in half and then added to the upper value for the long term survivors; that constitutes the middle point between the two groups. The cutoff is set within that difference between the 99% confidence interval of the groups and adjusted up or down from the aforementioned middle point according to the magnitude of the standard deviation of the two groups, i.e. the cutoff is moved away from the middle point from the group that has the larger standard deviation and closer to the other group (the one with the smaller standard deviation).

In another embodiment, the cutoff value is determined by a method comprising calculating an optimal point on the ROC curve based on the 34 scores of the 34 original subjects used in the discovery study [optimal point is defined as the point with the highest sensitivity and the lowest false positive rate (1-specificity)] for first group of short term survivors and a second group of long term survivors. That optimal point (the score of one of the 34 original subjects), which represents, according to ROC curve analysis, the best cutoff point for all of the 34 original subjects' scores, itself may be used as the cutoff point.

In embodiments, the cutoff values for the F1 function is 21.388, for the F2 function is 14.3 and for the F3 function is 14.7.

In embodiments, the method for determining the cutoff values further comprises verifying the validity of the cutoff value by obtaining output scores for a second set of patients (validation set) whose status as a long term survivor or short term survivor is hidden from the tester. The output scores are compared to the cutoff values for each function and if the patient’s sample in the validation set is greater than or equal to the cutoff value then it is predicted that the patient is a short term survivor and if less than the cutoff value a long term survivor. The status of the patient is unblinded and the validity of the cutoff value is determined by determining whether the cutoff value provides a sensitivity of at least 90% and a specificity of at least 90%.

**Selecting a treatment**

In embodiments a method comprises displaying whether the output score is less than a cutoff value indicating that the subject is a long term survivor or greater than or equal to the
cutoff indicating that the subject is a short term survivor so that the health care worker can select a treatment for the subject.

In embodiments, where the output score indicates that the subject is likely to be a long term survivor, the health care worker may select one or more standard therapy options. These standard therapy options include chemotherapy, surgery, and/or radiation. Standard chemotherapeutic options include treatment with one or more of cyclophosphamide, Taxol, Platinum, Carboplatin, Cisplatin, Gemcitabine, Topotecan, Oxaliplatin, Doxorubicin, Paclitaxel, Docetaxel, and combinations thereof.

In embodiments, where the output score indicates that the subject is likely to be a short term survivor, the health care worker may select a more aggressive treatment in addition to or in place of the standard chemotherapy. Such treatment includes treatment with a cancer vaccine, angiogenesis inhibitors, tubulin binding inhibitors, taxane analogs, actin polymerization inhibitors, adoptive cell therapy, and protein ubiquination inhibitors. Examples of compounds that can be utilized include Avastin, Votrient, SIK2 inhibitors, Vinblastine, ixabepilone, epothelin B, imatinib, atorvastatin, sirolimus, bestatin, indomethacin, simvastatin, infliximab, microcystin, Camptothecin, Combretastatin A4 phosphate, ZD 6126, Pomalidimide, Lenalidimide, and Bortezomib. In embodiments, the chemotherapy treatment includes treatment with an inhibitor of ACTB, TUBA3C, CDC42, and combinations thereof.

In some embodiment, the methods of the invention may be employed on a set of patients to identify a responder group or a nonresponder group in a clinical trial, for example. When testing a new therapeutic agent, it is useful to know whether the therapeutic agent has different effects in the responder population versus the nonresponder population. Using the methods of the disclosure, a group of patients having ovarian cancer are identified as responders or nonresponders and are then treated with a potential therapeutic agent. Safety and efficacy of the drug is assessed in responder and nonresponder populations.

**Methods for screening therapeutic agents**

Another aspect of the disclosure includes methods for screening therapeutic agents. Identification of ovarian cancer tissue samples as nonresponders and responders can be used to screen therapeutic effectiveness of the potential therapeutic agent on both types of patient populations. In some embodiments, cell lines may be developed from ovarian cancer tissue using standard methods from nonresponder and responders in order to provide for high throughput analysis.
In an embodiment, a method for screening agents for treating ovarian cancer, comprises contacting an ovarian cancer sample identified as a nonresponder or responder with a potential agent for treating ovarian cancer; and redetermining whether the agent decreases the growth, spread of the ovarian cancer sample, or changes the gene expression profile of the first set of genes, the second set of genes, the third set of genes or all sets of genes.

In embodiments, the method further comprises identifying a ovarian cancer sample as from a responder or nonresponder by determining the expression level of a a first set of genes comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4, a second set of genes comprising SSRI, USP5, ACTB, HLCS, NDUFBI, LYPLA2, TUBA3C, MED13L, and EED, or a third set of genes comprising CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED, in a sample from the patient.

In embodiments, the potential therapeutic agents are those that interact with any one of the genes a first set of genes comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4, a second set of genes comprising SSRI, USP5, ACTB, HLCS, NDUFBI, LYPLA2, TUBA3C, MED13L, and EED, or a third set of genes comprising CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED, or all set of genes in a sample from the patient. Examples of such agents are listed above. Drugs or chemicals similar to those known drugs in mechanism of action may be screened using nonresponder and responder ovarian cancer cells or cell lines as a measure of their efficacy in each of the patient groups. Other drugs or agents may also be those that are selected to act on other genes that are known to interact with any of the genes in the first or second set of genes as. The genes in the first, second, and/or third set of genes are targets to develop new therapeutics which can be tested on ovarian cancer cells identified as responder or nonresponders.

High throughput assays such as multiwell plate assays or arrays with cells attached to nanobeads can be utilized to test a number of therapeutic compounds for any effects on the responder or nonresponder cell types with regard to inhibition of cell growth, cell death, or change is gene expression of one or more of the genes of the first set of genes, the second set of genes, the third set of genes or all sets of genes. Those agents effective on both the responder and nonresponder population may be selected for further development. In other embodiments, an effective agent on either a responder or nonresponder cell types is selected and the patient group is sorted as responders and non responders for further testing of the agent effective in the respective responder or nonresponder cell type.
Another aspect of the disclosure involves a kit. In embodiments, a kit comprises a primer or a probe or both that specifically hybridizes to each gene of a set of genes comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4. In other embodiments, the kit comprises a primer or a probe or both that specifically hybridizes to each gene of a set of genes comprising SSR1, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L, and EED. In yet another embodiment, a kit comprises a primer or a probe or both that specifically hybridizes to each gene of a set of genes comprising CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED.

Multiple primers and probes can be prepared based on these known sequences and are encompassed within the disclosure. Primers can be designed in accord with a number of criteria using Primer design programs such as Premier Primer (biosoft), Oligo Primer Analysis software, and Oligo Perfect (Life Technologies) and other free and commercially available software. Probes can be designed using free and commercially available software including Array Designer (biosoft), and Light Cycler Probe design software (Roche). Primers and/or probes can be detectably labeled in accord with standard methods. Probes can be attached to a solid surface such as a slide, a well in a multiwell plate, and/or a chip. In embodiments, a primer and/or probe is designed to specifically bind to each of the nucleic acids encoding CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, EED, SSR1, USP5, NDUFB1, OSBPL8, and PKP4. In embodiments, a custom array can be prepared that contains no more than 200 probes, including at least 12 probes, one for each of the identified genes.

In embodiments, the primers and/or probes specifically bind to the nucleic acid sequences under standard PCR or microarray conditions. In embodiments, those conditions include 7% sodium dodecyl sulfate SDS, 0.5 M NaP04, 1 mM EDTA at 50°C with washing in 2X standard saline citrate (SSC), 0.1% SDS at 50°C; preferably in 7% (SDS), 0.5 M NaP04, 1 mM EDTA at 50°C with washing in IX SSC, 0.1% SDS at 50°C; preferably 7% SDS, 0.5 M NaP04, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C; and more preferably in 7% SDS, 0.5 M NaP04, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C. In other embodiments, a hybridization buffer includes 25% formamide, 2.5X SSC, 0.5% SDS and 1x Denhardts, and the primers and probes are incubated at 42°C for 1 hour followed by two washes of 0.5 SSC and 0.5% SDS.

In embodiments, the kit contains no more than 200 primers or probes or both, no more than 175 primers, probes or both, no more than 150 primers, probes or both, no more than
125 primers, probes or both, no more than 100 primers, probes or both, no more than 75 primer,
probes or both, no more than 50 primers, probes or both, no more than 25 primers, probes or both,
or no more than 15 primers, probes or both.

In embodiments, the kit can comprise or consist essentially of other reagents for detecting the gene expression
level of the identified genes. In embodiments, the kit may also contain primers or probes for detecting one or
more housekeeping genes as a positive control. In embodiments, the kit does not contain probes for any other
genes that are predictive of short term or long term survivorship of ovarian cancer other than the genes
identified herein.

In embodiments, the kit further comprises instructions for inputting the gene expression values into function
1, function 2, function 3, or combinations thereof to obtain an output score. The instructions further provide
comparing the output score for each function to a cutoff value and determining if the subject is likely to have
long term survival if the output score is less than the cutoff value or if the subject is likely to have short term
survival if the subject has an output score greater than or equal to the cutoff value for each function.

In embodiments, a kit further comprises a computer readable storage medium having computer-executable
instructions that, when executed by a computing device, cause the computing device to perform a step comprising:
calculating an output score by inputting gene expression levels of a set of genes into a function that provides a predictive
relationship between gene expression levels of the set of genes and short term or long term survival of
subjects having ovarian cancer.

In embodiments, the computer readable storage medium having computer-executable instructions that, when
executed by a computing device, cause the computing device to perform a step comprising: comparing the output
score to a cutoff value and displaying whether the subject is likely to have long term survival if the output score
is less than the cutoff value or if the subject is likely to have short term survival if the subject has an
output score greater than or equal to the cutoff value for each function.

In embodiments, the set of genes comprises at least the genes LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED,
and PKP4. In other embodiments the set of genes comprises at least the genes SSR1, USP5, ACTB, HLCS, NDUFB1,
LYPLA2, TUBA3C, MED13L, and EED. In yet other embodiments a set of genes comprises CDC42, LYPLA2,
TUBA3C, ACTB, HLCS, MED13L, and EED. In embodiments, the function is selected from the group consisting of
function 1, function 2, and function 3.
**Computer/Processor**

The detection, prognosis and/or diagnosis method can employ the use of a processor/computer system. For example, a general purpose computer system comprising a processor coupled to program memory storing computer program code to implement the method, to working memory, and to interfaces such as a conventional computer screen, keyboard, mouse, and printer, as well as other interfaces, such as a network interface, and software interfaces including a database interface find use one embodiment described herein.

The computer system accepts user input from a data input device, such as a keyboard, input data file, or network interface, or another system, such as the system interpreting, for example, the microarray or PCR data, and provides an output to an output device such as a printer, display, network interface, or data storage device. Input device, for example a network interface, receives an input comprising detection of the proteins/nucleic acids described herein and/or quantification of those compounds. The output device provides an output such as a display, including one or more numbers and/or a graph depicting the detection and/or quantification of the compounds.

Computer system is coupled to a data store which stores data generated by the methods described herein. This data is stored for each measurement and/or each subject; optionally a plurality of sets of each of these data types is stored corresponding to each subject. One or more computers/processors may be used, for example, as a separate machine, for example, coupled to computer system over a network, or may comprise a separate or integrated program running on computer system. Whichever method is employed these systems receive data and provide data regarding detection/diagnosis in return.

In embodiments, a method for selecting a treatment for a subject that has ovarian cancer comprises calculating an output score, using a computing device, by inputting gene expression levels of a set of genes into a function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer; and displaying the output score, using a computing device. In embodiments, the method further comprises determining whether the output score is greater than or equal to or less than a cutoff value, using a computing device; and displaying whether the subject is likely to be a short term survivor if the output score is greater than or equal to the cutoff value or long term survivor if the output score is less than the cutoff value.

In embodiments, a computing device, comprises a processing unit; and a system memory connected to the processing unit, the system memory including instructions that, when executed by the processing unit, cause the processing unit to: calculate an output
score by inputting gene expression levels of a set of genes into a function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer; and display the output score. In embodiments, the system memory includes instructions that when executed by the processing unit, cause the processing unit to determine whether the output score is greater than or equal to or less than a cutoff value; and displaying whether the subject is likely to be a short term survivor if the output score is greater than or equal to the cutoff value or long term survivor if the output score is less than the cutoff value.

In embodiments the set of genes comprises at least the genes LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4. In other embodiments the set of genes comprises at least the genes SSR1, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L, and EED. In yet other embodiments a set of genes comprises CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED. In embodiments, the function is selected from the group consisting of function 1, function 2, and function 3.

**Examples**

The following examples are provided in order to demonstrate and further illustrate certain embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

**Example 1 - Discovery of Ovarian Cancer Biomarkers**

The raw intensity microarray data (CEL files) by Berchuck et al. posted at the Duke Institute for Genome Sciences & Policy (Clinical Cancer Research 11, 3686-3696 (2005); (data.genome.duke.edu/clinicalcancerresearch.php) were used for this study. Briefly, according to Berchuck et al tumor tissue was harvested from 54 EOC patients with stages III and IV during surgery (48 with stage III and 6 with stage IV) and prior to the commencement of platinum/taxol chemotherapy. Total RNA was extracted from each tumor tissue sample and was analyzed for global gene expression using the GeneChip array U133A by Affymetrix. Following platinum/taxol chemotherapy, all 54 patients were followed for a period greater than seven years. Thirty patients survived for a period less than 3 years [NR/STS (non-responders/short-term survivors)-(median survival = 17.5 months)], and 24 patients survived for a period greater than 7 years [R/LTS (responders/long-term survivors)-(median survival = 107.5 months)]. None of the 30 NR/STS subjects died of causes other than EOC.
The aforementioned original raw intensity data (CEL files) were processed using the MAS5 normalization and background-correction algorithm (510k FDA approved). The Affymetrix U133A chip, which has 22,283 probe sets that can interrogate an equal number of transcripts, was utilized for this study. Microarray data from the aforementioned Affymetrix U133A chip were obtained from 14 long-term ovarian cancer survivors who lived more than seven years and 20 short-term survivors who lived less than three years after initial diagnosis.

We performed ROC curve analysis on the entire data matrix, i.e., on all variables (22,283 transcripts × 54 subjects) in order to assess the discriminating capability of all variables with respect to our two groups, namely, R/LTS and NR/STS. In the final round, we selected only those variables with an AUC ≥ 0.80. Eighty four variables (transcripts) fulfilled this criterion, and they constituted the final pool of the most significant variables. From the aforementioned 84 most significant variables, 13 became the input variables to the three complex mathematical functions (F1, F2, and F3), which we were able to generate, and which we term—and henceforward refer to as—super variables. Those three super variables (complex mathematical functions) are the final prognostic biomarker models. We should point out here that several other super variables were generated employing the remaining of the aforementioned 84 most significant variables, but, following final assessment, they proved to be not as robust as the F1, F2, or F3, and they are consequently not presented here.

The platform technology, as developed by Dr. Jason B. Nikas, and as presented in part in Nikas et al. 2010 (2), in Nikas and Low 2011(a) (3), and in Nikas and Low 2011(b) (4), identified three biomarkers (complex mathematical functions of original mRNA variables, see Figure 9 and discussion above) that allowed one to distinguish between long-term and short-term survivors or between responders and non-responders, respectively. The three biomarkers (panels of markers) are as follows:

F1 = f (LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, PKP4)  
F2 = f (SSR1, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L, EED)  
F3 = f (CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, EED).

The cut-off score of the F1 prognostic biomarker model, as well as those of the other two models, was determined by taking into account the results of the following two analyses: 1) calculation of the optimal point on the ROC curve based on the 34 scores of the 34 original subjects used in the discovery study [optimal point is defined as the point with the highest sensitivity and the lowest false positive rate (1-specificity)] and 2) calculation of the 99.99% confidence intervals for the mean F1 scores of the two groups (R/LTS and NR/STS) and their respective standard deviations. Based on that, the cut-off score of the F1 model was 65
determined to be 21.388. If a subject has an F1 score less than 21.388, then that subject is classified as an R/LTS; otherwise, that subject is classified as an NR/STS. As can be seen from Figure 1, the F1 model correctly identified all (14/14) R/LTS subjects and 19/20 NR/STS subjects. In terms of treatment response, since we would like to identify those subjects that will respond to the platinum/taxol chemotherapy, our target group is the R/LTS (responder /long term survivor) and our reference group is the NR/STS (non responder /short term survivor).

The ROC AUC of the F1 is 0.98929 with a 95% CI = [0.90449, 0.99884]. The mean F1 score of the 14 R/LTS subjects was 17.9358 (top of clear bar) and the standard deviation (whisker above or below the top of the clear bar) was 2.9622; whereas the mean F1 score of the 20 NR/STS subjects was 25.4697 (top of dark bar) and the standard deviation (whisker above or below the top of the dark bar) was 3.3651. The significance level was set at a = 0.001 (two-tailed), and the probability of significance was $P = 1.30 \times 10^{-7}$ (independent t-Test with T-value = -6.7405). The F1 is parametrically distributed with respect to both groups.

Therefore, for the discovery study, insofar as response to treatment is concerned, the F1 model exhibited a sensitivity = 1.000 and a specificity = 19/20 = 0.950. See Table IB. In the case of survival, given that we are interested in identifying the subjects that will be short-term survivors, our target group is the NR/STS and our reference group is the R/LTS. With regard to survival, therefore, for the discovery study, the F1 model exhibited a sensitivity = 0.950 and a specificity = 1.000. Both Figure 1 and Tables 1A and IB show all pertinent statistical results of the F1 prognostic biomarker model in connection with the discovery study in great detail.

The cut-off score of the F2 prognostic biomarker model was determined to be 14.259. If a subject has an F2 score less than 14.259, then that subject is classified as an R/LTS; otherwise, that subject is classified as an NR/STS. As can be seen from Figure 2, the F2 model correctly identified 13/14 R/LTS subjects and all (20/20) NR/STS subjects. In connection with treatment response, therefore, and with regard to the discovery study, the F2 model exhibited a sensitivity = 13/14 = 0.929 and a specificity = 1.000; whereas in connection with survival, its sensitivity and specificity were 1.000 and 0.929, respectively. Figure 2 and Tables 1A and IB show all pertinent statistical results of the F2 prognostic biomarker model in connection with the discovery study in great detail.

Regarding the F3 prognostic biomarker model, the cut-off score was determined to be 14.694, signifying that a score less than 14.694 belongs to an R/LTS subject, whereas a score
greater than 14.694 belongs to an NR/STS subject. As can be seen from Figure 2, the F3 model correctly identified all (14/14) R/LTS subjects and 19/20 NR/STS subjects. For the discovery study, therefore, with regard to treatment response, the sensitivity and specificity of the F3 model were 1.000 and 0.950, respectively; with regard to survival, its sensitivity and specificity were 0.950 and 0.000, respectively. Figure 2 and Tables 1A and 1B show all pertinent statistical results of the F3 prognostic biomarker model in connection with the discovery study in great detail.

The ROC AUC of the F2 is 0.98929 with a 95% CI = [0.90321, 0.99886], whereas the ROC AUC of the F3 is 0.98214 with a 95% CI = [0.86165, 0.99782]. The mean F2 score of

the 14 R/LTS subjects was 13.4223 (top of clear bar) and the standard deviation (whisker above or below the top of the clear bar) was 0.8905; whereas the mean F2 score of the 20 NR/STS subjects was 15.1843 (top of dark bar) and the standard deviation (whisker above or below the top of the dark bar) was 0.6407. The mean F3 score of the 14 R/LTS subjects was 13.8864 and the standard deviation was 0.7017; whereas the mean F3 score of the 20 NR/STS subjects was 15.3433 and the standard deviation was 0.6082. The significance level was set at α = 0.001 (two-tailed), and the probability of significance for the F2 was P = 1.37 x 10⁻⁷ (independent t-Test with T-value = -6.7217), whereas the probability of significance for the F3 was P = 2.93 x 10⁻⁷ (independent t-Test with T-value = -6.4541). Both the F2 and the F3 are parametrically distributed with respect to both groups.

Thirty-four subjects were used (14 long-term survivors (LTS) and 20 short-term survivors (STS)). The prognostic accuracy of the three biomarkers is shown in Table 1A. Both the F1 and F3 prognostic models classified correctly all 14 LTS subjects (Reference Group) and misclassified one STS subject (Target Group) (Sensitivity = 0.950 and Specificity = 1.000). The F2 prognostic model classified correctly all 20 STS subjects and misclassified one LTS subject (Sensitivity = 1.000 and Specificity = 0.929).

Table 1A. Statistical results of all three prognostic models and predicted group mean values for future LTS and STS subjects.
Table IB shows the response to treatment results (here, the LTS subjects are the target group and the STS subjects are the reference group). Table IB. Statistical results of all three prognostic models for future Responders (LTS) and Non-Responders (STS).

### Table IA. SURVIVAL - LTS vs. STS

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>AUC</th>
<th>95% CI of AUC</th>
<th>T-Value (two-tailed)</th>
<th>P</th>
<th>Reference Group (LTS)</th>
<th>Target Group (STS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>0.98929</td>
<td>[0.90449, 0.99884]</td>
<td>-6.7405</td>
<td>&lt; 1 x 10⁻⁷</td>
<td>[15.8665, 20.6631] (2.9622)</td>
<td>[22.9236, 27.6717] (3.3651)</td>
</tr>
<tr>
<td>F2</td>
<td>0.98929</td>
<td>[0.90321, 0.99886]</td>
<td>-6.7217</td>
<td>&lt; 1 x 10⁻⁷</td>
<td>[12.7822, 14.2173] (0.8905)</td>
<td>[14.6982, 15.6030] (0.6407)</td>
</tr>
<tr>
<td>F3</td>
<td>0.98214</td>
<td>[0.86165, 0.99782]</td>
<td>-6.4541</td>
<td>&lt; 1 x 10⁻⁷</td>
<td>[13.3940, 14.5352] (0.7017)</td>
<td>[14.8956, 15.7599] (0.6082)</td>
</tr>
</tbody>
</table>

Differences in the values of the gene expression biomarkers between long-term and short-term survivors are shown in Figure 1 for biomarker F1 and in Figure 2 for the F2 and F3 biomarkers. The LTS and STS box plots of the F1 biomarker are shown in Figure 3, whereas those of F2 and F3 biomarkers are shown in Figure 4.

The 3-dimensional plot of biomarkers F1 vs. F2 vs. F3 is shown in Figure 5. This plot shows distinct and separated clustering for long-term vs. short-term survivors of ovarian cancer patients with the exception of one subject.

### Example 2 - Qualification of Ovarian Cancer Biomarkers

The aforementioned diagnostic biomarkers were validated with 20 new, unknown subjects (10 long-term survivors and 10 short-term survivors). The F1, F2, and F3 prognostic models correctly classified all 20 new, unknown subjects (Sensitivity = 1.000 and Specificity = 1.000). The validation (qualification) results are shown in Table 2A (here, the LTS subjects are the reference group and the STS subjects are the target group).

Table 2A. Statistical results of all three prognostic models with respect to the 20 new, unknown subjects, along with the observed group mean values of those unknown subjects.
The observed group mean values of the 20 new, unknown subjects fall within the respective confidence intervals as predicted by all three models (see Table 1A).

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>AUC</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Reference Group (LTS) Mean Value</th>
<th>Target Group (STS) Mean Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>16.1301</td>
<td>24.3990</td>
</tr>
<tr>
<td>F2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>13.0048</td>
<td>14.9212</td>
</tr>
<tr>
<td>F3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>13.6150</td>
<td>15.0952</td>
</tr>
</tbody>
</table>

Table 2B shows the response to treatment results (here, the LTS subjects are the target group and the STS subjects are the reference group). Table 2B. Statistical results of all three prognostic models for the 20 new, unknown subjects as Responders (LTS) & Non-Responders (STS).

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>AUC</th>
<th>Sensitivity (LTS -&gt; Target Group)</th>
<th>Specificity (STS -&gt; Reference Group)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>F2</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>F3</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Combined scatter plots and bar graphs with all individual subjects (both LTS (responders) and STS (non-responders)) of the three prognostic biomarkers (F1, F2, and F3) are shown in Figure 6 (F1) and Figure 7 (F2 and F3). As can be seen, all three prognostic biomarkers were able to correctly prognose all 20 new, unknown subjects (complete segregation of the two survival/(treatment-response) groups, i.e. the LTS from the STS group).

A 3-dimensional plot of F1 vs. F2 vs. F3 is shown in Figure 8. As can be seen, all three prognostic biomarkers correctly prognosed all 20 new, unknown subjects (10 LTS (responders) and 10 STS (non-responders)) (complete segregation of the long-term and short-term survival groups).

The aforementioned 12 genes can be categorized into three general groups: 1) genes that regulate the expression of cytostructural proteins, 2) genes that regulate cell proliferation, and 3) genes that regulate metabolism.

The genes ACTB, TUBA3C, and CDC42 have functions that pertain to the cytoskeleton, and as such, they compose the first group. Cancer proliferation and metastasis
relies on cytostructural materials, i.e., cytoskeletal proteins, such as microfilaments (actin) and microtubules. The first two genes promote the expression of actin and microtubules, respectively. CDC42 promotes the polymerization of actin into microfilaments; reorganization of the actin cytoskeletonl5; and cell formation, growth, and spreading. There is also evidence that CDC42 can regulate the polarization of both the actin and the microtubule cytoskeleton. In our study all three of the aforementioned cytostructural genes were significantly over-expressed in the NR/STS group relative to the R/LTS group.

The following genes, whose function pertains to cell proliferation in general, compose the second group: MED13L, SSR1, PKP4, EED, and USP5. The MED13L protein (also known as, among other names, THRAP2, TRAP240L, and KIAA1025) is member of the Mediator complex, a group of about 30 transcriptional co-activators that play various regulatory roles in the induction of RNA polymerase II transcription. Compositional differences may account for different functions among the Mediator proteins; for instance some promote transcription, whereas others act as transcriptional repressors.

A number of those Mediator proteins are novel, and, consequently, their exact function is not known, including that of MED13L. Regarding specifically the MED13L gene, it has been observed that over-expression of the TP53 gene (p53) in human colon carcinoma cell lines relative to controls suppresses the expression of MED13L (KIAA1025). That could very well explain our finding that the MED13L gene was significantly under-expressed in the NR/STS group relative to the R/LTS group by affirming the existence of a more aggressive EOC cancer in the case of the former group in comparison with the latter one.

SSR1 is an ER (endoplasmic reticulum) receptor part of the translocon-associated protein (TRAP)complex. In our study, the SSR1 gene was significantly under-expressed in the NR/STS group relative to the R/LTS group.

The PKP4 protein (aka p0071) belongs to the family of arm-repeat proteins, which are involved in cell adhesion. According to the results of our analysis, the PKP4 gene was significantly under-expressed in the NR/STS group relative to the R/LTS group, and that accords with the observation that metastatic cancer cells rely on greater cell mobility and, thus, lower cell adhesion.

The EED protein is part of the Polycomb-group (PcG) proteins involved in repressive transcriptional control mediated via histone deacetylation. We found that the EED gene was significantly under-expressed in the NR/STS group relative to the R/LTS group, indicating that in the case of more aggressive EOC, inhibitory control of gene activity was more diminished.
USP5 belongs to the largest class of deubiquitinating enzymes (USPs) that regulate protein ubiquitination, a post-translational modification of cellular proteins. Compounds that inhibit the regulation of protein ubiquitination, such as bortezomib, have been approved by the FDA and are used for the treatment of certain types of cancer. Moreover, other compounds that specifically suppress the activity of USP5, such as WP1 130, lead to apoptosis of tumor cells. Those findings are in agreement with our results: we found that the USP5 gene was significantly over-expressed in the NR/STS group relative to the R/LTS group. There is also evidence that malignant tumors, via the release of certain factors, may promote the expression of USPs and other deubiquitinating enzymes in order to induce major alterations in the metabolism, such as increased proteolysis and lipolysis.

The third group comprises genes whose function is involved in metabolism in general and lipid metabolism in particular. Those genes are: LYPLA2, OSBPL8, HLCS, and NDUFB1. LYPLA2 is the enzyme that catalyzes the hydrolysis of 2-lysophosphatidylcholine (which, along with arachidonic acid, is derived from the hydrolysis of phosphatidylcholine—a phospholipid that is a major component of cell membranes) to glycerophosphocholine. The protein OSBPL8 is an intracellular lipid receptor that belongs to the family of oxysterols (oxygenated cholesterol derivatives). Oxysterols activate the liver X receptors (LXR) that regulate the expression of a number of genes whose function pertains to cholesterol metabolism. We found that the OSBPL8 gene was significantly under-expressed in the NR/STS group relative to the R/LTS group, suggesting that lipogenesis is more suppressed in the more aggressive cancer cells. That is in agreement with our findings about the overexpression of the LYPLA2 gene in the same group (NR/STS), which points to a greater lipolysis in the case of the more aggressive cancer cells. HLCS is an enzyme that catalyzes the covalent biotinylation of the five crucial mammalian carboxylase enzymes: pyruvate carboxylase (PC), acetyl-CoA carboxylase 1 and 2 (ACC1 and ACC2), 3-methylcrotonyl-CoA carboxylase (MCC), and propionyl-CoA carboxylase (PCC). From an energy production perspective, the most likely targets of HLCS in connection with advanced-stage EOC are PCC and MCC. Along with 45 other subunits, the DUFBI dehydrogenase (ubiquinone) 1 beta subcomplex constitutes the mitochondrial Complex I—a very large multiprotein enzyme which is located in the inner mitochondrial membrane, and which catalyzes the first step of the electron transport chain, the redox machinery of the oxidative phosphorylation. It has been observed by multiple studies that, owing to their surrounding hypoxic environment, tumor cells rely to a much larger extent on anaerobic glycolysis to produce energy rather than
on oxidative phosphorylation. This is in complete agreement with our finding that the
NDUFB I gene was significantly under-expressed in the NR/STS group relative to the R/LTS
group, for it points to a lower utilization of oxidative phosphorylation on the part of the more
aggressive cancer cells.

Another category involves genes related to the mechanism of action of taxol. Taxol is
an anti-tubulin chemotherapeutic agent that acts as a mitotic inhibitor. More specifically, it
increases polymerization of microtubules from α-β-tubulin heterodimers, and it stabilizes
microtubules by preventing their depolymerization. This action prevents the formation of the
mitotic spindle, a necessary step in the process of mitosis, and that results in the arrest of the
mitosis cycle either in the G2 or the M phase. As was mentioned earlier, the gene TUBA3C,
which encodes the production of α-tubulin, and which is common to all three super variables
and one of the top four most significant predictors, was significantly over-expressed in those
patients (NR/STS) who did not respond to taxol (platinum/taxol chemotherapy). Furthermore,
the fact that taxol binds to the β-tubulin subunit would render the over-expression of
TUBA3C on the part of the more aggressive cancer cells a successful strategy for evading the
action of taxol and, thus, for survival. The CDC42 gene not only promotes the polymerization
of actin into microfilaments, the reorganization of the actin cytoskeleton, and cell formation,
growth, and spreading; but also it can regulate the polarization of both the actin and the
microtubule cytoskeleton. Theoretically, therefore, over-expression of the CDC42 gene can
overcome the action of taxol, as well; and that is the finding of our analysis: the CDC42 gene
was significantly over-expressed in the NR/STS group relative to the R/LTS group.

In summary, when over-expressed, the genes TUBA3C, ACTB, and CDC42 can
collectively (or even perhaps in a given combination thereof) overcome the exerted actions of
taxol and diminish its efficacy. It stands to reason, therefore, to expect that a new
pharmacological approach whereby, via a combination of chemotherapeutic agents, all three
of those aforementioned genes are targeted will be more successful than the current standard
treatment in extending the life span of those women with EOC who, because of the specific
pattern of the aforementioned genetic networks, will not respond to the platinum/taxol
chemotherapy and will turn out to be short-term survivors.

All publications, patents and patent applications are incorporated herein by reference.
While in the foregoing specification this invention has been described in relation to certain
preferred embodiments thereof, and many details have been set forth for purposes of
illustration, it will be apparent to those skilled in the art that the invention is susceptible to
additional embodiments and that certain of the details described herein may be varied considerably without departing from the basic principles of the invention.

Bibliography


WHAT IS CLAIMED IS:

1. A method of selecting a treatment for a subject that has ovarian cancer comprising:
   a) determining whether the subject is likely to have short term or long term survival by a method comprising
      i) measuring the level of gene expression of at least a set of genes comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4 in a sample comprising ovarian cancer cells from the subject;
      ii) inputting the expression levels of the set of genes into a function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer to obtain an output score;
   iii) determining whether the subject is likely to have long term survival by determining if the output score is less than a cutoff value or whether the subject is likely to have short term survival by determining if the output score is greater than or equal to the cutoff value, wherein the cutoff value is a value determined by identifying a value between the 99% confidence interval of the mean output scores of a first set of samples from subjects known to have short term survival and a second set of samples from subjects known to have long term survival; and
      b) displaying whether the output score is greater than or equal to the cutoff value or less than the cutoff value to a health care worker so that the health care worker can select a treatment for the subject.

2. The method of claim 1, wherein the cDNA levels are measured.

3. The method of claim 1 or 2, wherein the gene expression levels are measured by microarray analysis.

4. The method of any one of claims 1-2, wherein gene expression levels are measured by polymerase chain reaction.

5. The method of any one of claims 1-2, wherein the gene expression levels are measured by RNA sequencing.
6. The method of any one of claims 1-5, wherein the function is as follows:

\[ F_1 = \left\{ \text{arc sinh} \left\{ \frac{(\ln X_1)^{2.5}(X_2)^{0.4}(X_3)^{0.1}}{(X_4)^{0.2}(X_5)^{0.05}(X_6)^{0.5}(\ln X_7)^{1.2}} \right\} \right\} (10) \]

Wherein \( X_1 \) is LYPLA2, \( X_2 \) is TUBA3C, \( X_3 \) is ACTB, \( X_4 \) is MED13L, \( X_5 \) is OSBPL8, \( X_6 \) is EED, and \( X_7 \) is PKP4.

7. The method of claim 6, wherein the cutoff value is about 2.14.

8. The method of any one of claims 1-7, further comprising treating a subject likely to have long term survival with standard chemotherapy.

9. The method of any one of claim 1-8, further comprising treating a subject likely to have short term survival with an inhibitor of a protein selected from the group consisting of TUBA3C, ACTB, CDC42 and combinations thereof.

10. A method of selecting a treatment for a subject that has ovarian cancer comprising:
   a) determining whether the subject is likely to have short term or long term survival by a method comprising
      i) measuring the level of gene expression of at least a set of genes comprising SSRI, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L, and EED in a sample comprising ovarian cancer cells from the subject;
      ii) inputting the expression levels of the set of genes into a function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer to obtain an output score;
      iii) determining whether the subject is likely to have long term survival by determining if the output score is less than a cutoff value or whether the subject is likely to have short term survival by determining if the output score is greater than or equal to the cutoff value, wherein the cutoff value is a value determined by identifying a value between the 99% confidence interval of a mean output score of a first set of samples from subjects known to have short term survival and a mean output score of a second set of samples from subjects known to have long term survival; and
b) displaying whether the output value of the sample is greater than or equal to the
cutoff value or less than the cutoff value so that the health care worker can select a treatment
for the subject.

11. The method of claim 10, wherein the cDNA levels are measured.

12. The method of claim 10 or 11, wherein the gene expression levels are measured by
microarray analysis.

13. The method of any one of claims 10-11, wherein gene expression levels are measured
by polymerase chain reaction.

14. The method of any one of claims 10-11, wherein the gene expression levels are
measured by RNA sequencing.

15. The method of any one of claims 10-14, wherein the function is as follows:

\[
F_2 = \left\{ \ln \left( \frac{(\ln X_1)^{2.5}(X_2)^{0.4}(X_3)^{0.1}(X_9)^{0.03}(\ln X_{10})^{1.8}(X_{11})^{0.5}}{(X_4)^{0.25}(\ln X_8)(X_6)^{0.3}(\ln X_{12})^{0.5}} \right) \right\}^{1.2}
\]

Wherein X1 is LYPLA2, X2 is TUBA3C, X3 is ACTB, X4 is MED13L, X6 is EED, X8 is
SSR1, X9 is USP5, X10 is ACTB, X11 is HLCS, and X12 is NDUFB1.

16. The method of claim 15, wherein the cutoff value is about 14.3.

17. The method of any one of claims 10-16, further comprising treating a subject likely to
have long term survival with standard chemotherapy.

18. The method of any one of claim 10-16, further comprising treating a subject likely to
have short term survival with an inhibitor of a protein selected from the group consisting of
TUBA3C, ACTB, CDC42 and combinations thereof.
19. A method of selecting a treatment for a subject that has ovarian cancer comprising:
   a) determining whether the subject is likely to have short term or long term survival by a method comprising
      i) measuring the level of gene expression of at least a set of genes comprising CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED in a sample comprising ovarian cancer cells from the subject;
      ii) inputting the expression levels of the set of genes into a function that provides a predictive relationship between gene expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer to obtain an output score;
      iii) determining whether the subject is likely to have long term survival by determining if the output score is less than a cutoff value or whether the subject is likely to have short term survival by determining if the output score is greater than or equal to the cutoff value, wherein the cutoff value is a value determined by identifying a value between the 99% confidence interval of a mean output score of a first set of samples from subjects known to have short term survival and a mean output score of a second set of samples from subjects known to have long term survival; and
   b) displaying whether the output score is less than a cutoff value indicating that the subject is a long term survivor or whether the output score is greater than or equal to the cutoff indicating that the subject is a short term survivor so that the health care worker can select a treatment for the subject.

20. The method of claim 19, wherein the cDNA levels are measured.

21. The method of claim 19 or 20, wherein the gene expression levels are measured by microarray analysis.

22. The method of any one of claims 19-20, wherein gene expression levels are measured by polymerase chain reaction.

23. The method of any one of claims 19-20, wherein the gene expression levels are measured by RNA sequencing.

24. The method of any one of claims 19-23, wherein the function is as follows:
\[ F_3 = \left( \text{arc sinh} \left( \frac{(\ln x_1)^{\frac{3}{4}}}{(X_4)^{0.2}(X_6)^{0.4}} \right)^{0.6} (\ln x_{10})^{0.85} \right) \left( X_{11} \right) \left( X_{13} \right) \right)^{1.44} \]

Wherein \( X_1 \) is LYPLA2, \( X_2 \) is TUBA3C, \( X_3 \) is ACTB, \( X_4 \) is MED13L, \( X_6 \) is EED, \( X_{10} \) is ACTB, \( X_{11} \) is HLCS, and \( X_{13} \) is CDC42.

25. The method of claim 24, wherein the cutoff value is about 14.7.

26. The method of any one of claims 19-25, further comprising treating a subject likely to have long term survival with standard chemotherapy.

27. The method of any one of claim 19-25, further comprising treating a subject likely to have short term survival with an inhibitor of a protein selected from the group consisting of TUBA3C, ACTB, CDC42 and combinations thereof.

28. A kit comprising: a primer or a probe or both that specifically hybridizes to each gene of a first set of genes comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4, and wherein the kit contains no more than 125 primers or probes.

29. A kit comprising: a primer or a probe or both that specifically hybridizes to each gene of a first set of genes comprising SSR1, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L, and EED and wherein the kit contains no more than 125 primers or probes.

30. A kit comprising: a primer or a probe or both that specifically hybridizes to each gene of a first set of genes comprising CDC42, LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED and wherein the kit contains no more than 125 primers or probes.

31. The kit of any one of claims 28-30, further comprising a computer readable storage medium having computer-executable instructions that, when executed by a computing device, cause the computing device to perform a step comprising:
calculating an output score by inputting gene expression levels of a set of genes of
any one of claims 28-30 into a function that provides a predictive relationship between gene
expression levels of the set of genes and short term or long term survival of subjects having
ovarian cancer; and

determining whether the subject is likely to be long term survivor by determining
whether the output score is less than a cutoff value or is likely to be a short term survivor if
the output score is greater than or equal to the cutoff score.

32. A method for selecting a treatment for a subject that has ovarian cancer comprising,
the method comprising:

calculating an output score, using a computing device, by inputting gene expression
levels of a first set of genes comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8,
EED, and PKP4, a second set of genes comprising SSRI, USP5, ACTB, HLCS, NDUFB1,
LYPLA2, TUBA3C, MED13L, and EED, or a third set of genes comprising CDC42,
LYPLA2, TUBA3C, ACTB, HLCS, MED13L, and EED, into a function that provides a
predictive relationship between gene expression levels of the set of genes and short term or
long term survival of subjects having ovarian cancer; and displaying the output score, using a
computing device.

33. A method of claim 32, further comprising

determining whether the output score is greater than or equal to or less than a cutoff
value, using a computing device; and displaying whether the subject is likely to be a short
term survivor if the output score is greater than or equal to the cutoff value or long term
survivor if the output score is less than the cutoff value.

34. A computing device, comprising:

a processing unit; and

a system memory connected to the processing unit, the system memory including
instructions that, when executed by the processing unit, cause the processing unit to:

calculate an output score by inputting gene expression levels of a first set of genes
comprising LYPLA2, TUBA3C, ACTB, MED13L, OSBPL8, EED, and PKP4, a second set
of genes comprising SSRI, USP5, ACTB, HLCS, NDUFB1, LYPLA2, TUBA3C, MED13L,
and EED, or a third set of genes comprising CDC42, LYPLA2, TUBA3C, ACTB, HLCS,
MED13L, and EED, into a function that provides a predictive relationship between gene

expression levels of the set of genes and short term or long term survival of subjects having ovarian cancer; and display the output score.

35. A computing device of claim 34, wherein the system memory includes instructions, that when executed by the processing unit, cause the processing unit to determine whether the output score is greater than or equal to or less than a cutoff value; and displaying whether the subject is likely to be a short term survivor if the output score is greater than or equal to the cutoff value or long term survivor if the output score is less than the cutoff value.
Ovarian Cancer Prognosis (F₁) — Discovery Study

- LTS (N=14)
- STS (N=20)

F₁ Score

P = 1.30 x 10⁻⁷

Sensitivity = 0.950
Specificity = 1.000
AUC = 0.989

Mean LTS ± SD = 17.94 ± 2.96
Mean STS ± SD = 25.47 ± 3.37

Bars represent standard deviations (SD)
(*) Significance level: α = 0.001 (two-tailed)
F₁ is parametrically distributed with respect to both groups

Figure 1
Ovarian Cancer Prognosis ($F_2$ & $F_3$) — Discovery Study

$F_2$
- Sensitivity = 1.000
- Specificity = 0.929
- AUC = 0.989

$F_3$
- Sensitivity = 0.950
- Specificity = 1.000
- AUC = 0.982

Bars represent standard deviations (SD)
(*) Significance level: $\alpha = 0.001$ (two-tailed)
$F_2$ & $F_3$ are parametrically distributed with respect to both groups

Figure 2
Ovarian Cancer Prognosis ($F_1$) — Discovery Study

Figure 3
Ovarian Cancer Prognosis ($F_2 - F_3$) — Discovery Study

Figure 4
Ovarian Cancer Prognosis ($F_1 - F_2 - F_3$) — Discovery Study

- LTS (N=14)
- STS (N=20)

Figure 5
Ovarian Cancer Prognosis (F₁) — Validation Study

- LTS (N=10)
- STS (N=10)

P=3.19x10⁻⁷

Sensitivity = 1.000
Specificity = 1.000
AUC = 1.000

Mean LTS ± SD = 16.13 ± 2.93
Mean STS ± SD = 24.40 ± 1.58

Bars represent standard deviations (SD)
(*) Significance level: α = 0.001 (two-tailed)
F₁ is parametrically distributed with respect to both groups

Figure 6
Ovarian Cancer Prognosis ($F_2$ & $F_3$) — Validation Study

![Bar chart showing the comparison between LTS (N=10) and STS (N=10) groups.]

- LTS (N=10)
- STS (N=10)

Bars represent standard deviations (SD)

(*): Significance level: $\alpha = 0.001$ (two-tailed)

$F_2$ & $F_3$ are parametrically distributed with respect to both groups

Sensitivity = 1.000
Specificity = 1.000
AUC = 1.000

Figure 7
Ovarian Cancer Prognosis ($F_1 - F_2 - F_3$) — Validation Study

- LTS (N=10)
- STS (N=10)

Figure 8
\[ F_1 = \left\{ \text{arc sinh} \left\{ \frac{(\ln X_1)^{2.5}(X_2)^{0.4}(X_3)^{0.1}}{(X_4)^{0.2}(X_5)^{0.05}(X_6)^{0.5}(\ln X_7)^{1.2}} \right\} \right\}^{(10)} \]

\[ F_2 = \left\{ \ln \left\{ \frac{(\ln X_1)^{2.5}(X_2)^{0.4}(X_3)^{0.1}(X_9)^{0.03}(\ln X_{10})^{1.8}(X_{11})^{0.5}}{(X_4)^{0.25}(\ln X_8)(X_6)^{0.3}(\ln X_{12})^{0.5}} \right\} \right\}^{1.2} \]

\[ F_3 = \left\{ \text{arc sinh} \left\{ \frac{(\ln X_1)^{3.6}(X_2)(X_3)^{0.6}(\ln X_{10})^{2.8}(X_{11})(X_{13})^{0.85}}{(X_4)^{0.2}(X_6)^{0.4}} \right\} \right\}^{1.48} \]
INTERNATIONAL SEARCH REPORT

International application No. PCT/US2012/036120

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68 (2012.01)
USPC - 435/7.23

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)
IPC(8) - C07K 16/40; C12Q 1/68; G01N 33/574; G06F 19/00 (2012.01)
USPC - 435/6.12, 6.19, 7.23, 530/388.8

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 practicable, search terms used)
PatBase, Google Patents, Google, PubMed Central

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>
<tbody>
<tr>
<td>Y</td>
<td>US 7,15,370 B2 (RAY et al) 03 October 2006 (03.10.2006) entire document</td>
<td>5, 14, 23</td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C.

* 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 application or patent 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 referring to an oral disclosure, use, exhibition or other means
  "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search: 16 July 2012
Date of mailing of the international search report: 20 AUG 2012

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-3201

Authorized officer: Blaine R. Copenheaver
PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)
This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
   because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 6-9, 15-18, 24-27, 31
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: 6-9, 15-18, 24-27, 31.

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 6-9, 15-18, 24-27, 31.

Remark on Protest ☐ The additional search fees were accompanied by the applicant’s protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant’s protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.