(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 December 2005 (29.12.2005)

PCT

(10) International Publication Number
WO 2005/123993 A2

(51) International Patent Classification⁷: **C40B 40/02**

(21) International Application Number: PCT/US2005/020107

(22) International Filing Date: 8 June 2005 (08.06.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/578,406	9 June 2004 (09.06.2004)	US
Not furnished	6 June 2005 (06.06.2005)	US

(71) Applicant *(for all designated States except US)*: **THE REGENTS OF THE UNIVERSITY OF MICHIGAN** [US/US]; 3003 South State Street, Ann Arbor, MI 48109-1280 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **CHINNAIYAN, Arul** [US/US]; 51300 Plymouth Valley Drive, Plymouth, MI 48170 (US). **WANG, Xiaoju** [CN/US]; 6145 Gabrielle Avenue, Ann Arbor, MI 48103 (US).

(74) Agent: ARENSON, Tanya, A.; Medlen & Carroll, LLP, 101 Howard Street, Suite 350, San Francisco, CA 94105 (US).

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

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

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

(54) Title: PHAGE MICROARRAY PROFILING OF THE HUMORAL RESPONSE TO DISEASE

(57) Abstract: The present invention relates to compositions and methods for cancer diagnostics, including but not limited to, cancer markers. In particular, the present invention provides methods and compositions for phage microarray profiling of cancer (e.g., prostate, lung, or breast cancer). The present invention further provides novel markers useful for the diagnosis, characterization, and treatment of cancers.

WO 2005/123993 A2

PHAGE MICROARRAY PROFILING OF THE HUMORAL RESPONSE TO DISEASE

This application claims priority to provisional application serial number 60/578,406, filed June 9, 2004, which is herein incorporated by reference in its entirety.

The present invention was funded in part by NIH grant PSOCA69568. The government has certain rights in the invention.

FIELD OF THE INVENTION

The present invention relates to compositions and methods for disease diagnostics. In particular, the present invention provides methods and compositions for phage microarray profiling of cancer (e.g., prostate, lung or breast cancer). The present invention further provides novel markers useful for the diagnosis, characterization, and treatment of disease (e.g., cancers).

BACKGROUND OF THE INVENTION

Afflicting one out of nine men over age 65, prostate cancer (PCA) is a leading cause of male cancer-related death, second only to lung cancer (Abate-Shen and Shen, Genes Dev 14:2410 [2000]; Ruijter *et al.*, Endocr Rev, 20:22 [1999]). The American Cancer Society estimates that about 184,500 American men will be diagnosed with prostate cancer and 39,200 will die in 2001.

Prostate cancer is typically diagnosed with a digital rectal exam and/or prostate specific antigen (PSA) screening. An elevated serum PSA level can indicate the presence of PCA. PSA is used as a marker for prostate cancer because it is secreted only by prostate cells. A healthy prostate will produce a stable amount -- typically below 4 nanograms per milliliter, or a PSA reading of "4" or less -- whereas cancer cells produce escalating amounts that correspond with the severity of the cancer. A level between 4 and 10 may raise a doctor's suspicion that a patient has prostate cancer, while amounts above 50 may show that the tumor has spread elsewhere in the body.

When PSA or digital tests indicate a strong likelihood that cancer is present, a transrectal ultrasound (TRUS) is used to map the prostate and show any suspicious areas. Biopsies of various sectors of the prostate are used to determine if prostate cancer is present. Treatment options depend on the stage of the cancer. Men with a 10-year life expectancy or less who have a low Gleason number and whose tumor has not spread beyond the prostate are often treated with watchful waiting (no treatment). Treatment options for more aggressive cancers include surgical treatments such as radical prostatectomy (RP), in which the prostate is completely removed (with or without nerve sparing techniques) and radiation, applied through an external beam that directs the dose to the prostate from outside the body or via low-dose radioactive seeds that are implanted within the prostate to kill cancer cells locally. Anti-androgen hormone therapy is also used, alone or in conjunction with surgery or radiation. Hormone therapy uses luteinizing hormone-releasing hormones (LH-RH) analogs, which block the pituitary from producing hormones that stimulate testosterone production. Patients must have injections of LH-RH analogs for the rest of their lives.

While surgical and hormonal treatments are often effective for localized PCA, advanced disease remains essentially incurable. Androgen ablation is the most common therapy for advanced PCA, leading to massive apoptosis of androgen-dependent malignant cells and temporary tumor regression. In most cases, however, the tumor reemerges with a vengeance and can proliferate independent of androgen signals.

The advent of prostate specific antigen (PSA) screening has led to earlier detection of PCA and significantly reduced PCA-associated fatalities. However, the impact of PSA screening on cancer-specific mortality is still unknown pending the results of prospective randomized screening studies (Etzioni *et al.*, J. Natl. Cancer Inst., 91:1033 [1999]; Maattanen *et al.*, Br. J. Cancer 79:1210 [1999]; Schroder *et al.*, J. Natl. Cancer Inst., 90:1817 [1998]). A major limitation of the serum PSA test is a lack of prostate cancer sensitivity and specificity especially in the intermediate range of PSA detection (4-10 ng/ml). Elevated serum PSA levels are often detected in patients with non-malignant conditions such as benign prostatic hyperplasia (BPH) and prostatitis, and provide little information about the aggressiveness of the cancer detected. Coincident with increased serum PSA testing, there has been a dramatic increase in the number of prostate needle biopsies performed (Jacobsen *et al.*, JAMA 274:1445 [1995]). This has resulted in a surge of equivocal prostate needle biopsies (Epstein and Potter J. Urol., 166:402 [2001]). Thus, development of additional serum and tissue biomarkers to supplement or replace PSA screening is needed.

SUMMARY OF THE INVENTION

The present invention relates to compositions and methods for disease diagnostics. In particular, the present invention provides methods and compositions for phage microarray profiling of cancer (e.g., prostate, breast, or lung cancer). The present invention further provides novel markers useful for the diagnosis, characterization, and treatment of disease (e.g., cancers).

Accordingly, in some embodiments, the present invention provides a method, comprising: providing a phage library, wherein the phage library comprises a plurality of phage clones, each of the phage clones comprising a cDNA obtained from a disease (e.g., cancer, autoimmune disease, inflammatory disease, cardiovascular disease and diabetes) mRNA sample; enriching the phage library for phage clones comprising cDNAs specific to the disease, where the enriching comprises binding the phage library to a control IgG to remove non-disease specific phage clones followed by binding the phage library to a disease specific IgG to enrich the phage library for disease specific phage clones, thereby generating an enriched phage library; exposing the enriched phage library to serum from disease patients and optionally serum from non-diseased control subjects to generate a immunoglobulin bound phage library; and identifying phage clones that react with the serum from the disease patients. In some embodiments, the method further comprises the step of identifying phage clones that react with serum from the disease subjects, but not with the serum from non-diseased control subjects. In some embodiments, the identifying comprises contacting the immunoglobulin bound phage library with a first immunoglobulin that binds to immunoglobulins from the serum from patients having the disease and a second immunoglobulin that binds to a phage capsid protein. In some embodiments, the identifying further comprises the step of exposing the first and second immunoglobulins to third and fourth immunoglobulins wherein the third immunoglobulin binds to the first immunoglobulin and wherein the third immunoglobulin comprises a first label, and wherein the fourth immunoglobulin binds to the second immunoglobulin and wherein the fourth immunoglobulin comprises a second label. In some embodiments, the first and second labels are fluorescent dyes and the first label emits fluorescence at a different wavelength than the second label. In some embodiments, the method further comprises the step of exposing the labeled phage library to an image scanner to identify phage clones that react with the serum from the disease patients but not with the serum from non-diseased control subjects. In some embodiments, the method further comprises the step of determining the identity of genes contained in the phage clones that react with the serum from the disease patients but not with the serum from non-diseased control subjects. In some embodiments, the disease is prostate, lung, or breast cancer. In certain embodiments, the enriched phage library is arrayed on a solid surface. In some embodiments, the disease specific IgG is purified from the serum of a patient with the disease. In some preferred embodiments, the enriching step is repeated 2 or more, and preferably 5 or more times. In preferred embodiments, the disease is cancer and the phage clones that react with the serum from the cancer patients but not with the serum from non-cancer control subjects comprise cDNAs encoding tumor antigens. In certain embodiments, the present invention provides a tumor antigen identified by the above-described method.

In further embodiments, the present invention provides a method for detecting cancer (e.g., prostate, breast or lung cancer), comprising: providing a sample (e.g., including, but not limited to, a blood sample or a tumor sample) from a subject (e.g., a human) suspected of having cancer; and detecting the presence or absence of a humoral response to a tumor antigen (e.g., BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 or heat shock 70kDa protein 8 (HSPA70)), thereby detecting cancer. In some embodiments, the detecting comprises exposing the sample to an antibody and detecting the antibody binding to the tumor antigen. In other embodiments, the detecting comprises detecting the presence of an autoantibody to the tumor antigen (e.g., by exposing the sample to an autoantibody specific antibody and detecting the autoantibody specific antibody binding to the antibody). In some further embodiments, the method further comprises the step of providing a prognosis to the subject. In some embodiments, the detecting cancer further comprises detecting a stage of the cancer or a sub-type of the cancer.

In yet other embodiments, the present invention provides a kit for detecting the presence of cancer (e.g., prostate, lung or breast cancer) in a subject, comprising: a reagent capable of (e.g., sufficient to) specifically detecting the presence of a tumor antigen (e.g., BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 or heat shock 70kDa protein 8 (HSPA70)); and instructions for using the reagent for detecting the presence of cancer in the subject. In some embodiments, the reagent is a tumor antigen specific antibody. In other embodiments, the reagent is an antibody specific for an autoantibody to the tumor antigen. In certain embodiments, the instructions comprise instructions required by the food and drug administration for labeling of in vitro diagnostics.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides a schematic overview of the phage-microarray profiling method of some embodiments of the present invention.

Figure 2 shows supervised analyses and validation of humoral immune response candidates of prostate cancer. Figure AB shows a Receiver Operator Characteristic (ROC) curve based on multiplex analysis of the 22 epitomic biomarkers. AUC, area under the curve. Figure 2B shows immunoreactivity of three representative clones validated by ELISA. Figure 2C shows titration curves of the humoral immune response to a representative phage-epitope clone (5'-UTR_BMI1).

Figure 3 shows a gene expression meta-analysis of humoral immune response candidates. Figure 3A shows a heatmap representation of the humoral immune response for four in frame phage-epitope clones assessed across 129 serum samples. Figure 3B shows the relative gene expression levels of in frame phage-epitope clones assessed using publicly available DNA microarray data housed in ONCOMINE. Figure 3C shows immunoblot validation of the overexpression of humoral response candidates at the protein level in prostate cancer.

Figure 4 shows a Table of clinical and pathology information of prostate cancer patients used for biopanning and epitope profiling in the training cohort of sera.

Figure 5 shows a Table of clinical and pathology information of prostate cancer patients used for epitope profiling in the validation cohort of sera.

Figure 6 shows a Table of Clinical and pathology information of hormone-refractory prostate cancer patients.

Figure 7 shows a Table of prediction accuracy of KNN models.

Figure 8 shows a Table that summarizes class predictions for the training sample set.

Figure 9 shows a Table of class predictions for the independent testing sample set.

Figure 10 shows a Table of class predictions of prostate cancer sera in which PSA levels are less than 4 ng/ml.

Figure 11 shows a Table of protein sequences of in-frame phage epitope clones.

Figure 12 shows a Table of significant protein list for epitope protein sequence alignment.

Figure 13 shows a schematic of the approach used to identify epitomic biomarkers of lung cancer in some embodiments of the present invention.

Figure 14 shows performance of the immune response profile in the test set.

Figure 15 shows humoral immune response profiles and patient survival.

Figure 16 shows characterization of UBQLN1.

DEFINITIONS

To facilitate an understanding of the present invention, a number of terms and phrases are defined below:

The term "epitope" as used herein refers to that portion of an antigen that makes contact with a particular antibody.

When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as "antigenic determinants". An antigenic determinant may compete with the intact antigen (*i.e.*, the "immunogen" used to elicit the immune response) for binding to an antibody.

The terms "specific binding" or "specifically binding" when used in reference to the interaction of an antibody and a protein or peptide means that the interaction is dependent upon the presence of a particular structure (*i.e.*, the antigenic determinant or epitope) on the protein; in other words the antibody is recognizing and binding to a specific protein structure rather than to proteins in general. For example, if an antibody is specific for epitope "A," the presence of a protein containing epitope A (or free, unlabelled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

As used herein, the terms "non-specific binding" and "background binding" when used in reference to the interaction of an antibody and a protein or peptide refer to an interaction that is not dependent on the presence of a particular structure (*i.e.*, the antibody is binding to proteins in general rather that a particular structure such as an epitope).

As used herein, the term "subject" refers to any animal (*e.g.*, a mammal), including, but not limited to, humans, non-human primates, rodents, and the like, which is to be the recipient of a particular treatment. Typically, the terms "subject" and "patient" are used interchangeably herein in reference to a human subject.

As used herein, the term "subject suspected of having cancer" refers to a subject that presents one or more symptoms indicative of a cancer (*e.g.*, a noticeable lump or mass) or is being screened for a cancer (*e.g.*, during a routine physical). A subject suspected of having cancer may also have one or more risk factors. A subject suspected of having cancer has generally not been tested for cancer. However, a "subject suspected of having cancer" encompasses an individual who has received an initial diagnosis (*e.g.*, a CT scan showing a mass or increased PSA level) but for whom the stage of cancer is not known. The term further includes people who once had cancer (*e.g.*, an individual in remission).

As used herein, the term "subject at risk for cancer" refers to a subject with one or more risk factors for developing a specific cancer. Risk factors include, but are not limited to, gender, age, genetic predisposition, environmental expose, previous incidents of cancer, preexisting non-cancer diseases, and lifestyle.

As used herein, the term "characterizing cancer in subject" refers to the identification of one or more properties of a cancer sample in a subject, including but not limited to, the presence of benign, pre-cancerous or cancerous tissue, the stage of the cancer, and the subject's prognosis. Cancers may be characterized by the identification of the expression of one or more cancer marker or tumor antigen genes, including but not limited to, the cancer markers disclosed herein.

As used herein, the term "characterizing prostate tissue in a subject" refers to the identification of one or more properties of a tissue sample (*e.g.*, including but not limited to, the presence of cancerous tissue, the presence of pre-cancerous tissue that is likely to become cancerous, and the presence of cancerous tissue that is likely to metastasize). In some embodiments, tissues are characterized by the identification of the expression of one or more cancer marker or tumor antigen genes, including but not limited to, the cancer markers disclosed herein.

As used herein, the term "cancer marker genes" refers to a gene whose expression level, alone or in combination with other genes, is correlated with cancer or prognosis of cancer. The correlation may relate to either an increased or decreased expression of the gene. For example, the expression of the gene may be indicative of cancer, or lack of expression of the gene may be correlated with poor prognosis in a cancer patient. Cancer marker expression may be characterized using any suitable method, including but not limited to, those described in illustrative Examples below.

As used herein, the term "a reagent that specifically detects expression levels" refers to reagents used to detect the expression of one or more genes (*e.g.*, including but not limited to, the cancer markers of the present invention). Examples of suitable reagents include but are not limited to, nucleic acid probes capable of specifically hybridizing to the gene of interest, PCR primers capable of specifically amplifying the gene of interest, and antibodies capable of specifically binding to proteins expressed by the gene of interest. Other non-limiting examples can be found in the description and examples below.

As used herein, the term "detecting a decreased or increased expression relative to non-cancerous control" refers to measuring the level of expression of a gene (*e.g.*, the level of mRNA or protein) relative to the level in a non-cancerous prostate control sample. Gene expression can be measured using any suitable method, including but not limited to, those described herein.

As used herein, the term "detecting a change in gene expression in said cell sample in the presence of said test compound relative to the absence of said test compound" refers to measuring an altered level of expression (*e.g.*, increased or decreased) in the presence of a test compound relative to the absence of the test compound. Gene expression can be measured using any suitable method, including but not limited to, those described herein.

As used herein, the term "tumor antigen" refers to an immunogenic epitope (*e.g.*, protein) expressed by a tumor cell. The protein may be expressed by non tumor cells but be immunogenic only when expressed by a tumor cell. Alternatively, the protein may be expressed by tumor cells, but not normal cells. Exemplary tumor antigens include, but are not limited to, BRD2, eIF4G1, RPL22, RPL13A, HES1, and hypothetical protein XP_373908.

As used herein, the term "autoantibody" refers to an antibody produced by a host (with or without immunization) and directed to a host antigen (*e.g.*, a tumor antigen).

As used herein, the term "cancer vaccine" refers to a composition (*e.g.*, a tumor antigen and a cytokine) that elicits a tumor-specific immune response. The response is elicited from the subject's own immune system by administering the cancer vaccine composition at a site (*e.g.*, a site distant from the tumor). In preferred embodiments, the immune response results in the eradication of tumor cells everywhere in the body (*e.g.*, both primary and metastatic tumor cells).

As used herein, the term "instructions for using said kit for detecting cancer in said subject" includes instructions for using the reagents contained in the kit for the detection and characterization of cancer in a sample from a subject. In some embodiments, the instructions further comprise the statement of intended use required by the U.S. Food and Drug Administration (FDA) in labeling *in vitro* diagnostic products. As used herein, the term "cancer expression profile map" refers to a presentation of expression levels of genes in a particular type of tissue (*e.g.*, primary, metastatic, and pre-cancerous tissues). The map may be presented as a graphical representation (*e.g.*, on paper or on a computer screen), a physical representation (*e.g.*, a gel or array) or a digital representation stored in computer memory. Each map corresponds to a particular type of tissue (*e.g.*, primary, metastatic, and pre-cancerous) and thus provides a template for comparison to a patient sample. In preferred embodiments, maps are generated from pooled samples comprising tissue samples from a plurality of patients with the same type of tissue.

As used herein, the terms "computer memory" and "computer memory device" refer to any storage media readable by a computer processor. Examples of computer memory include, but are not limited to, RAM, ROM, computer chips, digital video disc (DVDs), compact discs (CDs), hard disk drives (HDD), and magnetic tape.

As used herein, the term "computer readable medium" refers to any device or system for storing and providing information (*e.g.*, data and instructions) to a computer processor. Examples of computer readable media include, but are not limited to, DVDs, CDs, hard disk drives, magnetic tape and servers for streaming media over networks.

As used herein, the terms "processor" and "central processing unit" or "CPU" are used interchangeably and refer to a device that is able to read a program from a computer memory (*e.g.*, ROM or other computer memory) and perform a set of steps according to the program.

As used herein, the term "stage of cancer" refers to a qualitative or quantitative assessment of the level of advancement of a cancer. Criteria used to determine the stage of a cancer include, but are not limited to, the size of the tumor, whether the tumor has spread to other parts of the body and where the cancer has spread (*e.g.*, within the same organ or region of the body or to another organ).

As used herein, the term "providing a prognosis" refers to providing information regarding the impact of the presence of cancer (*e.g.*, as determined by the diagnostic methods of the present invention) on a subject's future health (*e.g.*, expected morbidity or mortality, the likelihood of getting cancer, and the risk of metastasis).

As used herein, the term "prostate specific antigen failure" refers to the development of high prostate specific antigen levels in a patient following prostate cancer therapy (*e.g.*, surgery). As used herein, the term "risk of developing prostate specific antigen failure" refers to a subject's relative risk (*e.g.*, the percent chance or a relative score) of developing prostate specific antigen failure following prostate cancer therapy.

As used herein, the term "post surgical tumor tissue" refers to cancerous tissue (*e.g.*, prostate tissue) that has been removed from a subject (*e.g.*, during surgery).

As used herein, the term "subject diagnosed with a cancer" refers to a subject who has been tested and found to have cancerous cells. The cancer may be diagnosed using any suitable method, including but not limited to, biopsy, x-ray, blood test, and the diagnostic methods of the present invention.

As used herein, the term "initial diagnosis" refers to results of initial cancer diagnosis (*e.g.* the presence or absence of cancerous cells). An initial diagnosis does not include information about the stage of the cancer of the risk of prostate specific antigen failure.

As used herein, the term "biopsy tissue" refers to a sample of tissue (*e.g.*, prostate tissue) that is removed from a subject for the purpose of determining if the sample contains cancerous tissue. In some embodiment, biopsy tissue is obtained because a subject is suspected of having cancer. The biopsy tissue is then examined (*e.g.*, by microscopy) for the presence or absence of cancer.

As used herein, the term "inconclusive biopsy tissue" refers to biopsy tissue for which histological examination has not determined the presence or absence of cancer.

As used herein, the term "non-human animals" refers to all non-human animals including, but are not limited to, vertebrates such as rodents, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc.

As used herein, the term "disease" refers to any deviation from a normal state in a subject. In preferred embodiments, the methods and compositions of the present invention are useful in the diagnosis and treatment of diseases where the immunological reaction (e.g., generation of immunoglobulins to native proteins) differs in subjects with disease and subjects not having disease. The present invention finds use with any number of diseases including, but not limited to, cancer, autoimmune disease, inflammatory disease, cardiovascular disease and diabetes.

The term "label" as used herein refers to any atom or molecule that can be used to provide a detectable (preferably quantifiable) effect, and that can be attached to a nucleic acid or protein. Labels include but are not limited to dyes; radiolabels such as ^{32}P; binding moieties such as biotin; haptens such as digoxgenin; luminogenic, phosphorescent or fluorogenic moieties; mass tags; and fluorescent dyes alone or in combination with moieties that can suppress or shift emission spectra by fluorescence resonance energy transfer (FRET). Labels may provide signals detectable by fluorescence, radioactivity, colorimetry, gravimetry, X-ray diffraction or absorption, magnetism, enzymatic activity, characteristics of mass or behavior affected by mass (*e.g.*, MALDI time-of-flight mass spectrometry), and the like. A label may be a charged moiety (positive or negative charge) or alternatively, may be charge neutral. Labels can include or consist of nucleic acid or protein sequence, so long as the sequence comprising the label is detectable.

The term “siRNAs” refers to short interfering RNAs. In some embodiments, siRNAs comprise a duplex, or double-stranded region, of about 18-25 nucleotides long; often siRNAs contain from about two to four unpaired nucleotides at the 3' end of each strand. At least one strand of the duplex or double-stranded region of a siRNA is substantially homologous to or substantially complementary to a target RNA molecule. The strand complementary to a target RNA molecule is the "antisense strand;" the strand homologous to the target RNA molecule is the "sense strand," and is also complementary to the siRNA antisense strand. siRNAs may also contain additional sequences; non-limiting examples of such sequences include linking sequences, or loops, as well as stem and other folded structures. siRNAs appear to function as key intermediaries in triggering RNA interference in invertebrates and in vertebrates, and in triggering sequence-specific RNA degradation during posttranscriptional gene silencing in plants.

The term "RNA interference" or "RNAi" refers to the silencing or decreasing of gene expression by siRNAs. It is the process of sequence-specific, post-transcriptional gene silencing in animals and plants, initiated by siRNA that is homologous in its duplex region to the sequence of the silenced gene. The gene may be endogenous or exogenous to the organism, present integrated into a chromosome or present in a transfection vector that is not integrated into the genome. The expression of the gene is either completely or partially inhibited. RNAi may also be considered to inhibit the function of a target RNA; the function of the target RNA may be complete or partial.

As used herein, the term "gene transfer system" refers to any means of delivering a composition comprising a nucleic acid sequence to a cell or tissue. For example, gene transfer systems include, but are not limited to, vectors (*e.g.*, retroviral, adenoviral, adeno-associated viral, and other nucleic acid-based delivery systems), microinjection of naked nucleic acid, polymer-based delivery systems (*e.g.*, liposome-based and metallic particle-based systems), biolistic injection, and the like. As used herein, the term "viral gene transfer system" refers to gene transfer systems comprising viral elements (*e.g.*, intact viruses, modified viruses and viral components such as nucleic acids or proteins) to facilitate delivery of the sample to a desired cell or tissue. As used herein, the term "adenovirus gene transfer system" refers to gene transfer systems comprising intact or altered viruses belonging to the family Adenoviridae.

As used herein, the term "site-specific recombination target sequences" refers to nucleic acid sequences that provide recognition sequences for recombination factors and the location where recombination takes place.

As used herein, the term "nucleic acid molecule" refers to any nucleic acid containing molecule, including but not limited to, DNA or RNA. The term encompasses sequences that include any of the known base analogs of DNA and RNA including, but not limited to, 4-acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinylcytosine, pseudoisocytosine, 5-(carboxyhydroxylmethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethylaminomethyluracil, dihydrouracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxy-aminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarbonylmethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, and 2,6-diaminopurine.

The term "gene" refers to a nucleic acid (*e.g.*, DNA) sequence that comprises coding sequences necessary for the production of a polypeptide, precursor, or RNA (*e.g.*, rRNA, tRNA). The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or functional properties (*e.g.*, enzymatic activity, ligand binding, signal transduction, immunogenicity, etc.) of the full-length or fragment are retained. The term also encompasses the coding region of a structural gene and the sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb or more on either end such that the gene corresponds to the length of the full-length mRNA. Sequences located 5' of the coding region and present on the mRNA are referred to as 5' non-translated sequences. Sequences located 3' or downstream of the coding region and present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

As used herein, the term "heterologous gene" refers to a gene that is not in its natural environment. For example, a heterologous gene includes a gene from one species introduced into another species. A heterologous gene also includes a gene native to an organism that has been altered in some way (*e.g.*, mutated, added in multiple copies, linked to non-native regulatory sequences, etc). Heterologous genes are distinguished from endogenous genes in that the heterologous gene sequences are typically joined to DNA sequences that are not found naturally associated with the gene sequences in the chromosome or are associated with portions of the chromosome not found in nature (*e.g.*, genes expressed in loci where the gene is not normally expressed).

As used herein, the term "gene expression" refers to the process of converting genetic information encoded in a gene into RNA (*e.g.*, mRNA, rRNA, tRNA, or snRNA) through "transcription" of the gene (*i.e.*, via the enzymatic action of an RNA polymerase), and for protein encoding genes, into protein through "translation" of mRNA. Gene expression can be regulated at many stages in the process. "Up-regulation" or "activation" refers to regulation that increases the production of gene expression products (*i.e.*, RNA or protein), while "down-regulation" or "repression" refers to regulation that decrease production. Molecules (*e.g.*, transcription factors) that are involved in up-regulation or down-regulation are often called "activators" and "repressors," respectively.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences that are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers that control or influence the transcription of the gene. The 3' flanking region may contain sequences that direct the termination of transcription, post-transcriptional cleavage and polyadenylation.

The term "wild-type" refers to a gene or gene product isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designed the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" refers to a gene or gene product that displays modifications in sequence and or functional properties (*i.e.*, altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally occurring mutants can be isolated; these are identified by the fact that they have altered characteristics (including altered nucleic acid sequences) when compared to the wild-type gene or gene product.

As used herein, the terms "nucleic acid molecule encoding," "DNA sequence encoding," and "DNA encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the polypeptide (protein) chain. The DNA sequence thus codes for the amino acid sequence.

As used herein, the terms "an oligonucleotide having a nucleotide sequence encoding a gene" and "polynucleotide having a nucleotide sequence encoding a gene," means a nucleic acid sequence comprising the coding region of a gene or in other words the nucleic acid sequence that encodes a gene product. The coding region may be present in a cDNA, genomic DNA or RNA form. When present in a DNA form, the oligonucleotide or polynucleotide may be single-stranded (*i.e.*, the sense strand) or double-stranded. Suitable control elements such as enhancers/promoters, splice junctions, polyadenylation signals, etc. may be placed in close proximity to the coding region of the gene if needed to permit proper initiation of transcription and/or correct processing of the primary RNA transcript. Alternatively, the coding region utilized in the expression vectors of the present invention may contain endogenous enhancers/promoters, splice junctions, intervening sequences, polyadenylation signals, etc. or a combination of both endogenous and exogenous control elements.

As used herein, the term "oligonucleotide," refers to a short length of single-stranded polynucleotide chain. Oligonucleotides are typically less than 200 residues long (*e.g.*, between 15 and 100), however, as used herein, the term is also intended to encompass longer polynucleotide chains. Oligonucleotides are often referred to by their length. For example a 24 residue oligonucleotide is referred to as a "24-mer". Oligonucleotides can form secondary and tertiary structures by self-hybridizing or by hybridizing to other polynucleotides. Such structures can include, but are not limited to, duplexes, hairpins, cruciforms, bends, and triplexes.

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (*i.e.*, a sequence of nucleotides) related by the base-pairing rules. For example, for the sequence "A-G-T," is complementary to the sequence "T-C-A." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids.

The term "homology" refers to a degree of complementarity. There may be partial homology or complete homology (*i.e.*, identity). A partially complementary sequence is a nucleic acid molecule that at least partially inhibits a completely complementary nucleic acid molecule from hybridizing to a target nucleic acid is "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (*i.e.*, the hybridization) of a completely homologous nucleic acid molecule to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target that is substantially non-complementary (*e.g.*, less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe that can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above.

A gene may produce multiple RNA species that are generated by differential splicing of the primary RNA transcript. cDNAs that are splice variants of the same gene will contain regions of sequence identity or complete homology (representing the presence of the same exon or portion of the same exon on both cDNAs) and regions of complete non-identity (for example, representing the presence of exon "A" on cDNA 1 wherein cDNA 2 contains exon "B" instead). Because the two cDNAs contain regions of sequence identity they will both hybridize to a probe derived from the entire gene or portions of the gene containing sequences found on both cDNAs; the two splice variants are therefore substantially homologous to such a probe and to each other.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe that can hybridize (*i.e.*, it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described above.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

As used herein, the term "T_m" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (See *e.g.*, Anderson and Young, Quantitative Filter Hybridization, in Nucleic Acid Hybridization [1985]). Other references include more sophisticated computations that take structural as well as sequence characteristics into account for the calculation of T_m.

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. Under "low stringency conditions" a nucleic acid sequence of interest will hybridize to its exact complement, sequences with single base mismatches, closely related sequences (*e.g.*, sequences with 90% or greater homology), and sequences having only partial homology (*e.g.*, sequences with 50-90% homology). Under "medium stringency conditions," a nucleic acid sequence of interest will hybridize only to its exact complement, sequences with single base mismatches, and closely relation sequences (*e.g.*, 90% or greater homology). Under "high stringency conditions," a nucleic acid sequence of interest will hybridize only to its exact complement, and (depending on conditions such a temperature) sequences with single base mismatches. In other words, under conditions of high stringency the temperature can be raised so as to exclude hybridization to sequences with single base mismatches.

"High stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH_2PO_4 H_2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 0.1X SSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

"Medium stringency conditions" when used in reference to nucleic acid hybridization comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH_2PO_4 H_2O and 1.85 g/l EDTA, pH adjusted to 7.4 with NaOH), 0.5% SDS, 5X Denhardt's reagent and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 1.0X SSPE, 1.0% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

"Low stringency conditions" comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH_2PO_4 H_2O and 1.85 g/l

EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5X Denhardt's reagent [50X Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharamcia), 5 g BSA (Fraction V; Sigma)] and 100 μg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5X SSPE, 0.1% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

The art knows well that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions that promote hybridization under conditions of high stringency (*e.g.*, increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, etc.) (see definition above for "stringency").

"Amplification" is a special case of nucleic acid replication involving template specificity. It is to be contrasted with non-specific template replication (*i.e.*, replication that is template-dependent but not dependent on a specific template). Template specificity is here distinguished from fidelity of replication (*i.e.*, synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out.

Template specificity is achieved in most amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under conditions they are used, will process only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. For example, in the case of Qβ replicase, MDV-1 RNA is the specific template for the replicase (Kacian *et al.*, Proc. Natl. Acad. Sci. USA 69:3038 [1972]). Other nucleic acids will not be replicated by this amplification enzyme. Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (Chamberlin *et al.*, Nature 228:227 [1970]). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (Wu and Wallace, Genomics 4:560 [1989]). Finally, Taq and Pfu polymerases, by virtue of their ability to function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the target sequences and not hybridization with non-target sequences (H.A. Erlich (ed.), PCR Technology, Stockton Press [1989]).

As used herein, the term "amplifiable nucleic acid" is used in reference to nucleic acids that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template."

As used herein, the term "sample template" refers to nucleic acid originating from a sample that is analyzed for the presence of "target." In contrast, "background template" is used in reference to nucleic acid other than sample template that may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

As used herein, the term "probe" refers to an oligonucleotide (*i.e.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, that is capable of hybridizing to at least a portion of another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be labeled with any "reporter molecule," so that is detectable in any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

As used herein the term "portion" when in reference to a nucleotide sequence (as in "a portion of a given nucleotide sequence") refers to fragments of that sequence. The fragments may range in size from four nucleotides to the entire nucleotide sequence minus one nucleotide (10 nucleotides, 20, 30, 40, 50, 100, 200, etc.).

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

The terms "in operable combination," "in operable order," and "operably linked" as used herein refer to the linkage of nucleic acid sequences in such a manner that a nucleic acid molecule capable of directing the transcription of a given gene and/or the synthesis of a desired protein molecule is produced. The term also refers to the linkage of amino acid sequences in such a manner so that a functional protein is produced.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated oligonucleotide" or "isolated polynucleotide" refers to a nucleic acid sequence that is identified and separated from at least one component or contaminant with which it is ordinarily associated in its natural source. Isolated nucleic acid is such present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids as nucleic acids such as DNA and RNA found in the state they exist in nature. For example, a given DNA sequence (*e.g.*, a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs that encode a multitude of proteins. However, isolated nucleic acid encoding a given protein includes, by way of example, such nucleic acid in cells ordinarily expressing the given protein where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid, oligonucleotide, or polynucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid, oligonucleotide or polynucleotide is to be utilized to express a protein, the oligonucleotide or polynucleotide will contain at a minimum the sense or coding strand (*i.e.*, the oligonucleotide or polynucleotide may be single-stranded), but may contain both the sense and anti-sense strands (*i.e.*, the oligonucleotide or polynucleotide may be double-stranded).

As used herein, the term "purified" or "to purify" refers to the removal of components (*e.g.*, contaminants) from a sample. For example, antibodies are purified by removal of contaminating non-immunoglobulin proteins; they are also purified by the removal of immunoglobulin that does not bind to the target molecule. The removal of non-immunoglobulin proteins and/or the removal of immunoglobulins that do not bind to the target molecule results in an increase in the percent of target-reactive immunoglobulins in the sample. In another example, recombinant polypeptides are expressed in bacterial host cells and the polypeptides are purified by the removal of host cell proteins; the percent of recombinant polypeptides is thereby increased in the sample.

"Amino acid sequence" and terms such as "polypeptide" or "protein" are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

The term "native protein" as used herein to indicate that a protein does not contain amino acid residues encoded by vector sequences; that is, the native protein contains only those amino acids found in the protein as it occurs in nature. A native protein may be produced by recombinant means or may be isolated from a naturally occurring source.

As used herein the term "portion" when in reference to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino acid sequence minus one amino acid.

The term "Southern blot," refers to the analysis of DNA on agarose or acrylamide gels to fractionate the DNA according to size followed by transfer of the DNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized DNA is then probed with a labeled probe to detect DNA species complementary to the probe used. The DNA may be cleaved with restriction enzymes prior to electrophoresis. Following electrophoresis, the DNA may be partially depurinated and denatured prior to or during transfer to the solid support.

Southern blots are a standard tool of molecular biologists (J. Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, NY, pp 9.31-9.58 [1989]).

The term "Northern blot," as used herein refers to the analysis of RNA by electrophoresis of RNA on agarose gels to fractionate the RNA according to size followed by transfer of the RNA from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized RNA is then probed with a labeled probe to detect RNA species complementary to the probe used. Northern blots are a standard tool of molecular biologists (J. Sambrook, *et al.*, *supra,* pp 7.39-7.52 [1989]).

The term "Western blot" refers to the analysis of protein(s) (or polypeptides) immobilized onto a support such as nitrocellulose or a membrane. The proteins are run on acrylamide gels to separate the proteins, followed by transfer of the protein from the gel to a solid support, such as nitrocellulose or a nylon membrane. The immobilized proteins are then exposed to antibodies with reactivity against an antigen of interest. The binding of the antibodies may be detected by various methods, including the use of radiolabeled antibodies.

The term "transgene" as used herein refers to a foreign gene that is placed into an organism by, for example, introducing the foreign gene into newly fertilized eggs or early embryos. The term "foreign gene" refers to any nucleic acid (*e.g.*, gene sequence) that is introduced into the genome of an animal by experimental manipulations and may include gene sequences found in that animal so long as the introduced gene does not reside in the same location as does the naturally occurring gene.

As used herein, the term "vector" is used in reference to nucleic acid molecules that transfer DNA segment(s) from one cell to another. The term "vehicle" is sometimes used interchangeably with "vector." Vectors are often derived from plasmids, bacteriophages, or plant or animal viruses.

The term "expression vector" as used herein refers to a recombinant DNA molecule containing a desired coding sequence and appropriate nucleic acid sequences necessary for the expression of the operably linked coding sequence in a particular host organism. Nucleic acid sequences necessary for expression in prokaryotes usually include a promoter, an operator (optional), and a ribosome binding site, often along with other sequences. Eukaryotic cells are known to utilize promoters, enhancers, and termination and polyadenylation signals.

The terms "overexpression" and "overexpressing" and grammatical equivalents, are used in reference to levels of mRNA to indicate a level of expression approximately 3-fold higher (or greater) than that observed in a given tissue in a control or non-transgenic animal. Levels of mRNA are measured using any of a number of techniques known to those skilled in the art including, but not limited to Northern blot analysis. Appropriate controls are included on the Northern blot to control for differences in the amount of RNA loaded from each tissue analyzed (*e.g.*, the amount of 28S rRNA, an abundant RNA transcript present at essentially the same amount in all tissues, present in each sample can be used as a means of normalizing or standardizing the mRNA-specific signal observed on Northern blots). The amount of mRNA present in the band corresponding in size to the correctly spliced transgene RNA is quantified; other minor species of RNA which hybridize to the transgene probe are not considered in the quantification of the expression of the transgenic mRNA.

The term "transfection" as used herein refers to the introduction of foreign DNA into eukaryotic cells. Transfection may be accomplished by a variety of means known to the art including calcium phosphate-DNA co-precipitation, DEAE-dextran-mediated transfection, polybrene-mediated transfection, electroporation, microinjection, liposome fusion, lipofection, protoplast fusion, retroviral infection, and biolistics.

The term "calcium phosphate co-precipitation" refers to a technique for the introduction of nucleic acids into a cell. The uptake of nucleic acids by cells is enhanced when the nucleic acid is presented as a calcium phosphate-nucleic acid co-precipitate. The original technique of Graham and van der Eb (Graham and van der Eb, Virol., 52:456 [1973]), has been modified by several groups to optimize conditions for particular types of cells. The art is well aware of these numerous modifications.

The term "stable transfection" or "stably transfected" refers to the introduction and integration of foreign DNA into the genome of the transfected cell. The term "stable transfectant" refers to a cell that has stably integrated foreign DNA into the genomic DNA.

The term "transient transfection" or "transiently transfected" refers to the introduction of foreign DNA into a cell where the foreign DNA fails to integrate into the genome of the transfected cell. The foreign DNA persists in the nucleus of the transfected cell for several days. During this time the foreign DNA is subject to the regulatory controls that govern the expression of endogenous genes in the chromosomes. The term "transient transfectant" refers to cells that have taken up foreign DNA but have failed to integrate this DNA.

As used herein, the term "selectable marker" refers to the use of a gene that encodes an enzymatic activity that confers the ability to grow in medium lacking what would otherwise be an essential nutrient (*e.g.* the HIS3 gene in yeast cells); in addition, a selectable marker may confer resistance to an antibiotic or drug upon the cell in which the selectable marker is expressed. Selectable markers may be "dominant"; a dominant selectable marker encodes an enzymatic activity that can be detected in any eukaryotic cell line. Examples of dominant selectable markers include the bacterial aminoglycoside 3' phosphotransferase gene (also referred to as the neo gene) that confers resistance to the drug G418 in mammalian cells, the bacterial hygromycin G phosphotransferase (hyg) gene that confers resistance to the antibiotic hygromycin and the bacterial xanthine-guanine phosphoribosyl transferase gene (also referred to as the gpt gene) that confers the ability to grow in the presence of mycophenolic acid. Other selectable markers are not dominant in that their use must be in conjunction with a cell line that lacks the relevant enzyme activity. Examples of non-dominant selectable markers include the thymidine kinase (tk) gene that is used in conjunction with tk $^-$ cell lines, the CAD gene that is used in conjunction with CAD-deficient cells and the mammalian hypoxanthine-guanine phosphoribosyl transferase (hprt) gene that is used in conjunction with hprt $^-$ cell lines. A review of the use of selectable markers in mammalian cell lines is provided in Sambrook, J. *et al.*, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, New York (1989) pp.16.9-16.15.

As used herein, the term "cell culture" refers to any in vitro culture of cells. Included within this term are continuous cell lines (*e.g.*, with an immortal phenotype), primary cell cultures, transformed cell lines, finite cell lines (*e.g.*, non-transformed cells), and any other cell population maintained in vitro.

As used, the term "eukaryote" refers to organisms distinguishable from "prokaryotes." It is intended that the term encompass all organisms with cells that exhibit the usual characteristics of eukaryotes, such as the presence of a true nucleus bounded by a nuclear membrane, within which lie the chromosomes, the presence of membrane-bound organelles, and other characteristics commonly observed in eukaryotic organisms. Thus, the term includes, but is not limited to such organisms as fungi, protozoa, and animals (*e.g.*, humans).

As used herein, the term "*in vitro*" refers to an artificial environment and to processes or reactions that occur within an artificial environment. In vitro environments can consist of, but are not limited to, test tubes and cell culture. The term "*in vivo*" refers to the natural environment (*e.g.*, an animal or a cell) and to processes or reaction that occur within a natural environment.

The terms "test compound" and "candidate compound" refer to any chemical entity, pharmaceutical, drug, and the like that is a candidate for use to treat or prevent a disease, illness, sickness, or disorder of bodily function (*e.g.*, cancer). Test compounds comprise both known and potential therapeutic compounds. A test compound can be determined to be therapeutic by screening using the screening methods of the present invention. In some embodiments of the present invention, test compounds include antisense compounds.

As used herein, the term "sample" is used in its broadest sense. In one sense, it is meant to include a specimen or culture obtained from any source, as well as biological and environmental samples. Biological samples may be obtained from animals (including humans) and encompass fluids, solids, tissues, and gases. Biological samples include blood products, such as plasma, serum and the like. Environmental samples include environmental material such as surface matter, soil, water, crystals and industrial samples. Such examples are not however to be construed as limiting the sample types applicable to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to compositions and methods for disease diagnostics. In particular, the present invention provides methods and compositions for phage microarray profiling of cancer (e.g., prostate, lung or breast cancer). The present invention further provides novel markers useful for the diagnosis, characterization, and treatment of disease (e.g., cancers). The below description illustrates the present invention in the context of cancer diagnosis and treatment. However, the present invention is not limited to use in the diagnosis and treatment of cancer. The methods and compositions of the present invention find use in the diagnosis and treatment of a variety of diseases including, but not limited to, inflammatory disease, autoimmune disease, cancer, cardiovascular disease, and diabetes.

When cancer is identified at the earliest stages, the probability of cure is very high and therefore diagnostic screening tests that can detect these early stages are crucial. Tumor-associated antigens recognized by humoral effectors of the immune system are an attractive target for diagnostic and therapeutic approaches to human cancer. Efforts toward the development of early detection assays for cancers have traditionally depended on single biomarker molecule. Current technologies have been disappointing and have not resulted in diagnostic tests suitable for clinical practice.

Serologic identification of antigens by recombinant expression cloning (SEREX) has been used for identification of few types of antigen over recent years through screening expression cDNA libraries from human solid tumors with sera of the autologous patients. This type of screening of a cDNA expression library by conventional methods, however, requires the preparation of a large number of membrane filters blotted with bacteriophage plaques that are then searched with a specific probe. In the case of the SEREX experiments, the screening is performed using large amounts of sera from cancer patients, which are usually available in very limited quantity. The second limitation is that such immunoscreening procedure does not allow selection of antigens that are recognized by sera from different patients. In addition, due to the filter screening procedure, SEREX does not allow for high throughput screening and thus makes it difficult to perform replicated experiments for the selection of antigens that can be recognized by sera from a subset of cancer patients. Furthermore, SEREX relies upon a one-step screening technique without affinity selection steps (biopanning).

The methods and compositions of the present invention overcome many of these limitations. In some embodiments, the present invention provides an effective screening test to overcome these limitations and simplify the screening procedure by performing affinity selection of cDNA libraries in very small volumes using, for example, T7 phage display cDNA libraries. The platform of phage-epitope microarrays is capable of detecting over 2300 phage clones in one microarray using only microliters of sera. Highly parallel assays using different patient samples are easily compared using protein microarray technology that allows for the molecular classification of cancer based on epitomic profiles (akin to molecular profiles based on gene expression). In some embodiments, the methods of the present invention employ the recognition of a pattern of immunologic response as a diagnostic strategy. The present invention is not limited by the nature of the peptide display system used.

Phage-display technology is typically based on the insertion of foreign nucleotide sequences into genes encoding for various capsid proteins of T7 phage, resulting in a heterogeneous mixture of phages, each displaying the different peptide sequence encoded by a corresponding insert. A physical link between a displayed fusion protein and DNA encoded for it make this phage target selectable. In some embodiments, the methods of the present invention detect antibodies that are produced by patients in reaction to proteins expressed in their tumors. These markers find use as diagnostic biomarkers and therapeutic targets. In some embodiments, the methods of the present invention employ pattern recognition of multiple markers as a diagnostic rather than any single marker. Features of the approach include acknowledging the heterogeneous nature of any specific kind of cancer, and using specialized bioinformatics techniques to interpret the results.

Experiments conducted during the course of development of the present invention resulted in the detection of a serum reaction with large numbers of epitopes using a highly parallel phage display assay on protein microarrays. Once the chosen epitope markers are spotted on the final version of the array, serum from both cancer patients and controls are tested. In some embodiments, the results of the reaction of the sera with the various subjects are used to train a machine learning device to build a predictor and further to test unknown samples.

The methods and compositions of the present invention provide several advantages over existing methods. For example, in some embodiments, the methods of the present invention utilize fluorescent probes and laser scanner, resulting in high sensitivity and the detection of very small signal differences. In addition, the methods of the present invention allow for detection at the protein expression level rather than cDNA level as compared to cDNA or oligo arrays. In preferred embodiments, the methods of the present invention utilize an analytical approach rather that a visual assessment, which results in greater consistency and reproducibility. Further, due to the high sensitivity of this technique, low amounts (e.g., only 1-2 μl) of serum samples may be used. The methods of the present invention are rapid and allow for the analysis of protein-protein interactions.

I. Markers for Cancer

In some embodiments, the present invention provides markers whose expression is specifically altered in cancerous prostate tissues. Such markers find use in the diagnosis and characterization of cancer (e.g., prostate, lung or breast cancer).

A. Identification of Markers

In some embodiments, the phage expression profiling methods of the present invention (See e.g., the experimental section for a detailed description) are used to identify cancer markers or tumor antigens. Exemplary prostate tumor antigens include, but are not limited to, BRD2, eIF4G1, RPL22, RPL13A, HES1, and hypothetical protein XP_373908. Exemplary breast cancer tumor antigens include, but are not limited to, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70).

B. Detection of Cancer Markers

In some embodiments, the present invention provides methods for detection of expression of cancer markers (*e.g.*, BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70)). In preferred embodiments, expression is measured directly (*e.g.*, at the RNA or protein level). In some embodiments, expression is detected in tissue samples (*e.g.*, biopsy tissue). In other embodiments, expression is detected in bodily fluids (*e.g.*, including but not limited to, plasma, serum, whole blood, mucus, and urine). The present invention further provides panels and kits for the detection of markers. In preferred embodiments, the presence of a cancer marker is used to provide a prognosis to a subject. The information provided is also used to direct the course of treatment. For example, if a subject is found to have a marker indicative of a highly metastasizing tumor, additional therapies (*e.g.*, hormonal or radiation therapies) can be started at a earlier point when they are more likely to be effective (*e.g.*, before metastasis). In addition, if a subject is found to have a tumor that is not responsive to hormonal therapy, the expense and inconvenience of such therapies can be avoided.

In some embodiments, the present invention provides a panel for the analysis of a plurality of markers. The panel allows for the simultaneous analysis of multiple markers correlating with carcinogenesis and/or metastasis. For example, a panel may include markers identified as correlating with cancerous tissue, metastatic cancer, localized cancer that is likely to metastasize, pre-cancerous tissue that is likely to become cancerous, and pre-cancerous tissue that is not likely to become cancerous. Depending on the subject, panels may be analyzed alone or in combination in order to provide the best possible diagnosis and prognosis. Markers for inclusion on a panel are selected by screening for their predictive value using any suitable method, including but not limited to, those described in the illustrative examples below.

In other embodiments, the present invention provides a phage array profile map comprising protein array profiles of cancers of various stages or prognoses (*e.g.*, likelihood of future metastasis). Such maps can be used for comparison with patient samples. Any suitable method may be utilized, including but not limited to, by computer comparison of digitized data. The comparison data is used to provide diagnoses and/or prognoses to patients.

i) Detection of RNA

In some preferred embodiments, detection of prostate cancer markers (*e.g.*, including but not limited to, BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70)) is detected by measuring the expression of corresponding mRNA in a tissue sample (*e.g.*, prostate, breast, or lung tissue). mRNA expression may be measured by any suitable method.

ii) Detection of Protein

In other embodiments, gene expression of cancer markers is detected by measuring the expression of the corresponding protein or polypeptide. Protein expression may be detected by any suitable method. In other embodiments, proteins are detected by their binding to an antibody raised against the protein. The generation of antibodies is described below.

Antibody binding is detected by techniques known in the art (*e.g.*, radioimmunoassay, ELISA (enzyme-linked immunosorbant assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitation reactions, immunodiffusion assays, *in situ* immunoassays (*e.g.*, using colloidal gold, enzyme or radioisotope labels, for example), Western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays, etc.), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc.

In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many methods are known in the art for detecting binding in an immunoassay and are within the scope of the present invention.

In some embodiments, an automated detection assay is utilized. Methods for the automation of immunoassays include those described in U.S. Patents 5,885,530, 4,981,785, 6,159,750, and 5,358,691, each of which is herein incorporated by reference. In some embodiments, the analysis and presentation of results is also automated. For example, in some embodiments, software that generates a prognosis based on the presence or absence of a series of proteins corresponding to cancer markers is utilized.

In other embodiments, the immunoassay described in U.S. Patents 5,599,677 and 5,672,480; each of which is herein incorporated by reference.

iii) Data Analysis

In some embodiments, a computer-based analysis program is used to translate the raw data generated by the detection assay (*e.g.*, the presence, absence, or amount of a given marker or markers) into data of predictive value for a clinician. The clinician can access the predictive data using any suitable means. Thus, in some preferred embodiments, the present invention provides the further benefit that the clinician, who is not likely to be trained in genetics or molecular biology, need not understand the raw data. The data is presented directly to the clinician in its most useful form. The clinician is then able to immediately utilize the information in order to optimize the care of the subject.

The present invention contemplates any method capable of receiving, processing, and transmitting the information to and from laboratories conducting the assays, information provides, medical personal, and subjects. For example, in some embodiments of the present invention, a sample (*e.g.*, a biopsy or a serum or urine sample) is obtained from a subject and submitted to a profiling service (*e.g.*, clinical lab at a medical facility, genomic profiling business, etc.), located in any part of the world (*e.g.*, in a country different than the country where the subject resides or where the information is ultimately used) to generate raw data. Where the sample comprises a tissue or other biological sample, the subject may visit a medical center to have the sample obtained and sent to the profiling center, or subjects may collect the sample themselves (*e.g.*, a urine sample) and directly send it to a profiling center. Where the sample comprises previously determined biological information, the information may be directly sent to the profiling service by the subject (*e.g.*, an information card containing the information may be scanned by a computer and the data transmitted to a computer of the profiling center using an electronic communication systems). Once received by the profiling service, the sample is processed and a profile is produced (*i.e.*, expression data), specific for the diagnostic or prognostic information desired for the subject.

The profile data is then prepared in a format suitable for interpretation by a treating clinician. For example, rather than providing raw expression data, the prepared format may represent a diagnosis or risk assessment (e.g., likelihood of metastasis or PSA failure) for the subject, along with recommendations for particular treatment options. The data may be displayed to the clinician by any suitable method. For example, in some embodiments, the profiling service generates a report that can be printed for the clinician (*e.g.*, at the point of care) or displayed to the clinician on a computer monitor.

In some embodiments, the information is first analyzed at the point of care or at a regional facility. The raw data is then sent to a central processing facility for further analysis and/or to convert the raw data to information useful for a clinician or patient. The central processing facility provides the advantage of privacy (all data is stored in a central facility with uniform security protocols), speed, and uniformity of data analysis. The central processing facility can then control the fate of the data following treatment of the subject. For example, using an electronic communication system, the central facility can provide data to the clinician, the subject, or researchers.

In some embodiments, the subject is able to directly access the data using the electronic communication system. The subject may choose further intervention or counseling based on the results. In some embodiments, the data is used for research use. For example, the data may be used to further optimize the inclusion or elimination of markers as useful indicators of a particular condition or stage of disease.

C. Detection of Tumor Antigens

As described above, the presence of an immune response to specific proteins expressed in cancerous cells is indicative of the presence of cancer. Accordingly, in some embodiments, the present invention provides methods (*e.g.*, diagnostic methods) for detecting the presence of tumor antigens identified using the methods of the present invention (e.g., BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70)). In some embodiments (*e.g.*, where tumor antigens are expressed in cancerous cells but not non-cancerous cells), tumor antigen proteins are detected directly. In other embodiments (*e.g.*, where the presence of an autoantibody in cancerous but not cancerous cells is indicative of the presence of cancer), autoantibodies to the tumor antigens are detected. In preferred embodiments, tumor antigens are detected directly in tumors or cells suspected of being cancerous.

The diagnostic methods of the present invention find utility in the diagnosis and characterization of cancers. For example, the presence of an autoantibody to a specific protein may be indicative of a cancer. In addition, certain autoantibodies may be indicative of a specific stage or sub-type of the same cancer.

The information obtained is used to determine prognosis and appropriate course of treatment. For example, it is contemplated that individuals with a specific autoantibody or stage of cancer may respond differently to a given treatment than individuals lacking the antibody. The information obtained from the diagnostic methods of the present invention thus provides for the personalization of diagnosis and treatment.

i) Detection of Antigens

In some embodiments, antibodies are used to detect tumor antigens in a biological sample from an individual. The biological sample can be a biological fluid, such as, but not limited to, blood, serum, plasma, interstitial fluid, urine, cerebrospinal fluid, and the like, containing cells. In preferred embodiments, the biological sample comprises cells suspected of being cancerous (*e.g.*, cells obtained from a biopsy).

The biological samples can then be tested directly for the presence of tumor antigens using an appropriate strategy (*e.g.*, ELISA or radioimmunoassay) and format (*e.g.*, microwells, dipstick (*e.g.*, as described in International Patent Publication WO 93/03367), etc). Alternatively, proteins in the sample can be size separated (*e.g.*, by polyacrylamide gel electrophoresis (PAGE), in the presence or not of sodium dodecyl sulfate (SDS), and the presence of tumor antigens detected by immunoblotting (*e.g.*, Western blotting). Immunoblotting techniques are generally more effective with antibodies generated against a peptide corresponding to an epitope of a protein, and hence, are particularly suited to the present invention.

Antibody binding is detected by techniques known in the art (*e.g.*, radioimmunoassay, ELISA (enzyme-linked immunosorbant assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitation reactions, immunodiffusion assays, *in situ* immunoassays (*e.g.*, using colloidal gold, enzyme or radioisotope labels, for example), Western blots, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays, etc.), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc.

In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention. As is well known in the art, the immunogenic peptide should be provided free of the carrier molecule used in any immunization protocol. For example, if the peptide was conjugated to KLH, it may be conjugated to BSA, or used directly, in a screening assay.)

In some embodiments, an automated detection assay is utilized. Methods for the automation of immunoassays are well known in the art (*See e.g.*, U.S. Patents 5,885,530, 4,981,785, 6,159,750, and 5,358,691, each of which is herein incorporated by reference). In some embodiments, the analysis and presentation of results is also automated. For example, in some embodiments, software that generates a prognosis based on the presence or absence of a series of antigens is utilized.

ii) Detection of Autoantibodies

In some embodiments, the presence of autoantibodies to a tumor antigen is detected. This approach to diagnosing and typing tumors is particularly suited to tumor antigens that are present, but not immunogenic, in normal cells and immunogenic in tumor cells. For example, in some embodiments, antibodies (*e.g.*, monoclonal or polyclonal) are generated to the autoantibodies identified during the development of the present invention. Such antibodies are then used to detect the presence of autoantibodies using any suitable technique, including but not limited to, those described above.

In other embodiments, tumor proteins are attached to a solid surface. The presence of autoantibodies is identified by contacting the solid surface (*e.g.*, microarray) with serum from the subject and detecting binding to a tumor marker. One exemplary method for performing such an assay is described in the experimental section below.

iii) Other Detection Methods

The present invention is not limited to the detection methods described above. Any suitable detection method that allows for the specific detection of cancerous cells may be utilized. For example, in some embodiments, the expression of RNA corresponding to a tumor antigen gene is detected by hybridization to an antisense oligonucleotide (*e.g.*, those described below). In other embodiments, RNA expression is detected by hybridization assays such as Northern blots, RNase assays, reverse transcriptase PCR amplification, and the like.

In further embodiments of the present invention, the presence of particular sequences in the genome of a subject are detected. Such sequences include tumor antigen sequences associated with abnormal expression of tumor antigens (*e.g.*, overexpression or expression at a physiological inappropriate time). These sequences include polymorphisms, including polymorphisms in the transcribed sequence (*e.g.*, that effect tumor antigen processing and/or translation) and regulatory sequences such as promoters, enhances, repressors, and the like. These sequences may also include polymorphisms in genes or control sequences associated with factors that affect expression such as transcription factors, and the like. Any suitable method for detecting and/or identifying these sequences is within the scope of the present invention including, but not limited to, nucleic acid sequencing, hybridization assays (*e.g.*, Southern blotting), single nucleotide polymorphism assays (*See e.g.*, U.S. Pat. No. 5,994,069, herein incorporated by reference in its entirety), and the like.

Direct and/or indirect measures of tumor antigen expression may be used as a marker within the scope of the present invention. Because the present invention provides a link between tumor antigen expression and cancer, any indication of tumor expression may be used. For example, the expression, activation, or repression of factors involved in tumor antigen signaling or regulation may be used as surrogate measures of expression, so long as they are reliably correlated with tumor antigen expression and/or cancer.

D. Molecular Fingerprint

In some embodiments, the present invention provides "molecular fingerprints" or "expression profile maps" of cancer markers or tumor antigens. Such molecular fingerprints and expression profiles provide a profile of the presence of autoantibodies or cancer markers in particular cancers or cancer sub-types. The profiles find use in providing cancer diagnoses and prognoses. Such prognoses can be used to determine treatment course of action. For example, in some embodiments, the profile of a particular cancer subtype is indicative of a cancer that is responsive to a particular choice of therapy. In other embodiments, profiles are indicative of the aggressiveness of a particular cancer sub-type and are used to determine the aggressiveness of treatment to be pursued.

E. Prognostic Applications

In some embodiments, cancer markers identified using the methods and compositions of the present invention find use in providing cancer prognoses (e.g., probability of cancer metastasis, recurrence or death from cancer). In experiments conducted during the course of development of the present invention (See e.g., Examples 3 and 4) a correlation between expression profiles and cancer prognosis was observed. For example, a correlation between expression of tripartite motif-containing 7 isoform 4, cytochrome c oxidase subunit I, nucleolar protein 3 (apoptosis repressor with CARD domain), hypothetical protein AM638, putative p150, MUP1, similar to CG9996-PA, hypothetical protein Magn028940, COG0568: DNA-directed RNA polymerase, sigma subunit, IgG kappa light chain variable region and lung cancer prognosis was observed (See Example 3).

F. Kits

In yet other embodiments, the present invention provides kits for the detection and characterization of cancer (e.g., prostate, breast, or lung cancer). In some embodiments, the kits contain antibodies specific for a cancer marker or tumor antigen, in addition to detection reagents and buffers. In other embodiments, the kits contain reagents specific for the detection of mRNA or cDNA (*e.g.*, oligonucleotide probes or primers). In preferred embodiments, the kits contain all of the components necessary to perform a detection assay, including all controls, directions for performing assays, and any necessary software for analysis and presentation of results.

G. In vivo Imaging

In some embodiments, in vivo imaging techniques are used to visualize the expression of cancer markers or tumor antigens in an animal (*e.g.*, a human or non-human mammal). For example, in some embodiments, cancer marker mRNA or protein is labeled using a labeled antibody specific for the cancer marker. A specifically bound and labeled antibody can be detected in an individual using an *in vivo* imaging method, including, but not limited to, radionuclide imaging, positron emission tomography, computerized axial tomography, X-ray or magnetic resonance imaging method, fluorescence detection, and chemiluminescent detection. Methods for generating antibodies to the cancer markers of the present invention are described below.

The in vivo imaging methods of the present invention are useful in the diagnosis of cancers that express the cancer markers or tumor antigens of the present invention (*e.g.*, prostate cancer). In vivo imaging is used to visualize the presence of a marker indicative of the cancer. Such techniques allow for diagnosis without the use of an unpleasant biopsy. The in vivo imaging methods of the present invention are also useful for providing prognoses to cancer patients. For example, the presence of a marker indicative of cancers likely to metastasize can be detected. The in vivo imaging methods of the present invention can further be used to detect metastatic cancers in other parts of the body.

In some embodiments, reagents (*e.g.*, antibodies) specific for the cancer markers or tumor antigens of the present invention are fluorescently labeled. The labeled antibodies are introduced into a subject (*e.g.*, orally or parenterally). Fluorescently labeled antibodies are detected using any suitable method (*e.g.*, using the apparatus described in U.S. Patent 6,198,107, herein incorporated by reference).

In other embodiments, antibodies are radioactively labeled. The use of antibodies for in vivo diagnosis is well known in the art. Sumerdon *et al.*, (Nucl. Med. Biol 17:247-254 [1990] have described an optimized antibody-chelator for the radioimmunoscintographic imaging of tumors using Indium-111 as the label. Griffin *et al.*, (J Clin Onc 9:631-640 [1991]) have described the use of this agent in detecting tumors in patients suspected of having recurrent colorectal cancer. The use of similar agents with paramagnetic ions as labels for magnetic resonance imaging is known in the art (Lauffer, Magnetic Resonance in Medicine 22:339-342 [1991]). The label used will depend on the imaging modality chosen. Radioactive labels such as Indium-111, Technetium-99m, or Iodine-131 can be used for planar scans or single photon emission computed tomography (SPECT). Positron emitting labels such as Fluorine-19 can also be used for positron emission tomography (PET). For MRI, paramagnetic ions such as Gadolinium (III) or Manganese (II) can be used.

Radioactive metals with half-lives ranging from 1 hour to 3.5 days are available for conjugation to antibodies, such as scandium-47 (3.5 days) gallium-67 (2.8 days), gallium-68 (68 minutes), technetiium-99m (6 hours), and indium-111 (3.2 days), of which gallium-67, technetium-99m, and indium-111 are preferable for gamma camera imaging, gallium-68 is preferable for positron emission tomography.

A useful method of labeling antibodies with such radiometals is by means of a bifunctional chelating agent, such as diethylenetriaminepentaacetic acid (DTPA), as described, for example, by Khaw *et al.* (Science 209:295 [1980]) for In-111 and Tc-99m, and by Scheinberg *et al.* (Science 215:1511 [1982]). Other chelating agents may also be used, but the 1-(p-carboxymethoxybenzyl)EDTA and the carboxycarbonic anhydride of DTPA are advantageous because their use permits conjugation without affecting the antibody's immunoreactivity substantially.

Another method for coupling DPTA to proteins is by use of the cyclic anhydride of DTPA, as described by Hnatowich *et al.* (Int. J. Appl. Radiat. Isot. 33:327 [1982]) for labeling of albumin with In-111, but which can be adapted for labeling of antibodies. A suitable method of labeling antibodies with Tc-99m which does not use chelation with DPTA is the pretinning method of Crockford *et al.*, (U.S. Pat. No. 4,323,546, herein incorporated by reference).

A preferred method of labeling immunoglobulins with Tc-99m is that described by Wong *et al.* (Int. J. Appl. Radiat. Isot., 29:251 [1978]) for plasma protein, and recently applied successfully by Wong *et al.* (J. Nucl. Med., 23:229 [1981]) for labeling antibodies.

In the case of the radiometals conjugated to the specific antibody, it is likewise desirable to introduce as high a proportion of the radiolabel as possible into the antibody molecule without destroying its immunospecificity. A further improvement may be achieved by effecting radiolabeling in the presence of the specific cancer marker of the present invention, to insure that the antigen binding site on the antibody will be protected. The antigen is separated after labeling.

In still further embodiments, *in vivo* biophotonic imaging (Xenogen, Almeda, CA) is utilized for in vivo imaging. This real-time *in vivo* imaging utilizes luciferase. The luciferase gene is incorporated into cells, microorganisms, and animals (*e.g.*, as a fusion protein with a cancer marker of the present invention). When active, it leads to a reaction that emits light. A CCD camera and software is used to capture the image and analyze it.

II. Antibodies

The present invention provides isolated antibodies. In preferred embodiments, the present invention provides monoclonal antibodies that specifically bind to an isolated polypeptide comprised of at least five amino acid residues of the cancer markers or tumor antigens described herein (e.g., BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70)). These antibodies find use in the diagnostic and therapeutic methods described herein.

An antibody against a protein of the present invention may be any monoclonal or polyclonal antibody, as long as it can recognize the protein. Antibodies can be produced by using a protein of the present invention as the antigen according to a conventional antibody or antiserum preparation process.

The present invention contemplates the use of both monoclonal and polyclonal antibodies. Any suitable method may be used to generate the antibodies used in the methods and compositions of the present invention, including but not limited to, those disclosed herein. For example, for preparation of a monoclonal antibody, protein, as such, or together with a suitable carrier or diluent is administered to an animal (*e.g.*, a mammal) under conditions that permit the production of antibodies. For enhancing the antibody production capability, complete or incomplete Freund's adjuvant may be administered. Normally, the protein is administered once every 2 weeks to 6 weeks, in total, about 2 times to about 10 times. Animals suitable for use in such methods include, but are not limited to, primates, rabbits, dogs, guinea pigs, mice, rats, sheep, goats, etc.

For preparing monoclonal antibody-producing cells, an individual animal whose antibody titer has been confirmed (*e.g.*, a mouse) is selected, and 2 days to 5 days after the final immunization, its spleen or lymph node is harvested and antibody-producing cells contained therein are fused with myeloma cells to prepare the desired monoclonal antibody producer hybridoma. Measurement of the antibody titer in antiserum can be carried out, for example, by reacting the labeled protein, as described hereinafter and antiserum and then measuring the activity of the labeling agent bound to the antibody. The cell fusion can be carried out according to known methods, for example, the method described by Koehler and Milstein (Nature 256:495 [1975]). As a fusion promoter, for example, polyethylene glycol (PEG) or Sendai virus (HVJ), preferably PEG is used.

Examples of myeloma cells include NS-1, P3U1, SP2/0, AP-1 and the like. The proportion of the number of antibody producer cells (spleen cells) and the number of myeloma cells to be used is preferably about 1:1 to about 20:1. PEG (preferably PEG 1000-PEG 6000) is preferably added in concentration of about 10% to about 80%. Cell fusion can be carried out efficiently by incubating a mixture of both cells at about 20°C to about 40°C, preferably about 30°C to about 37°C for about 1 minute to 10 minutes.

Various methods may be used for screening for a hybridoma producing the antibody (*e.g.*, against a tumor antigen or autoantibody of the present invention). For example, where a supernatant of the hybridoma is added to a solid phase (*e.g.*, microplate) to which antibody is adsorbed directly or together with a carrier and then an anti-immunoglobulin antibody (if mouse cells are used in cell fusion, anti-mouse immunoglobulin antibody is used) or Protein A labeled with a radioactive substance or an enzyme is added to detect the monoclonal antibody against the protein bound to the solid phase. Alternately, a supernatant of the hybridoma is added to a solid phase to which an anti-immunoglobulin antibody or Protein A is adsorbed and then the protein labeled with a radioactive substance or an enzyme is added to detect the monoclonal antibody against the protein bound to the solid phase.

Selection of the monoclonal antibody can be carried out according to any known method or its modification. Normally, a medium for animal cells to which HAT (hypoxanthine, aminopterin, thymidine) are added is employed. Any selection and growth medium can be employed as long as the hybridoma can grow. For example, RPMI 1640 medium containing 1% to 20%, preferably 10% to 20% fetal bovine serum, GIT medium containing 1% to 10% fetal bovine serum, a serum free medium for cultivation of a hybridoma (SFM-101, Nissui Seiyaku) and the like can be used. Normally, the cultivation is carried out at 20°C to 40°C, preferably 37°C for about 5 days to 3 weeks, preferably 1 week to 2 weeks under about 5% CO_2 gas. The antibody titer of the supernatant of a hybridoma culture can be measured according to the same manner as described above with respect to the antibody titer of the anti-protein in the antiserum.

Separation and purification of a monoclonal antibody (*e.g.*, against a cancer marker of the present invention) can be carried out according to the same manner as those of conventional polyclonal antibodies such as separation and purification of immunoglobulins, for example, salting-out, alcoholic precipitation, isoelectric point precipitation, electrophoresis, adsorption and desorption with ion exchangers (*e.g.*, DEAE), ultracentrifugation, gel filtration, or a specific purification method wherein only an antibody is collected with an active adsorbent such as an antigen-binding solid phase, Protein A or Protein G and dissociating the binding to obtain the antibody.

Polyclonal antibodies may be prepared by any known method or modifications of these methods including obtaining antibodies from patients. For example, a complex of an immunogen (an antigen against the protein) and a carrier protein is prepared and an animal is immunized by the complex according to the same manner as that described with respect to the above monoclonal antibody preparation. A material containing the antibody against is recovered from the immunized animal and the antibody is separated and purified.

As to the complex of the immunogen and the carrier protein to be used for immunization of an animal, any carrier protein and any mixing proportion of the carrier and a hapten can be employed as long as an antibody against the hapten, which is cross-linked on the carrier and used for immunization, is produced efficiently. For example, bovine serum albumin, bovine cycloglobulin, keyhole limpet hemocyanin, etc. may be coupled to a hapten in a weight ratio of about 0.1 part to about 20 parts, preferably, about 1 part to about 5 parts per 1 part of the hapten.

In addition, various condensing agents can be used for coupling of a hapten and a carrier. For example, glutaraldehyde, carbodiimide, maleimide activated ester, activated ester reagents containing thiol group or dithiopyridyl group, and the like find use with the present invention. The condensation product as such or together with a suitable carrier or diluent is administered to a site of an animal that permits the antibody production. For enhancing the antibody production capability, complete or incomplete Freund's adjuvant may be administered. Normally, the protein is administered once every 2 weeks to 6 weeks, in total, about 3 times to about 10 times.

The polyclonal antibody is recovered from blood, ascites and the like, of an animal immunized by the above method. The antibody titer in the antiserum can be measured according to the same manner as that described above with respect to the supernatant of the hybridoma culture. Separation and purification of the antibody can be carried out according to the same separation and purification method of immunoglobulin as that described with respect to the above monoclonal antibody.

The protein used herein as the immunogen is not limited to any particular type of immunogen. For example, a cancer marker of the present invention (further including a gene having a nucleotide sequence partly altered) can be used as the immunogen. Further, fragments of the protein may be used. Fragments may be obtained by any methods including, but not limited to expressing a fragment of the gene, enzymatic processing of the protein, chemical synthesis, and the like.

III. Drug Screening

In some embodiments, the present invention provides drug screening assays (*e.g.*, to screen for anticancer drugs). The screening methods of the present invention utilize cancer markers and tumor antigens identified using the methods of the present invention. For example, in some embodiments, the present invention provides methods of screening for compound that alter (*e.g.*, increase or decrease) the expression of cancer marker or tumor antigen genes. In some embodiments, candidate compounds are antisense agents (*e.g.*, oligonucleotides) directed against cancer markers. See below for a discussion of antisense therapy. In other embodiments, candidate compounds are antibodies that specifically bind to a cancer marker or tumor antigen of the present invention.

In one screening method, candidate compounds are evaluated for their ability to alter cancer marker expression by contacting a compound with a cell expressing a cancer marker and then assaying for the effect of the candidate compounds on expression. In some embodiments, the effect of candidate compounds on expression of a cancer marker gene is assayed for by detecting the level of cancer marker or tumor antigen mRNA expressed by the cell. mRNA expression can be detected by any suitable method. In other embodiments, the effect of candidate compounds on expression of cancer marker or tumor antigen genes is assayed by measuring the level of polypeptide encoded by the cancer markers. The level of polypeptide expressed can be measured using any suitable method, including but not limited to, those disclosed herein.

Specifically, the present invention provides screening methods for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to cancer markers or tumor antigens of the present invention, have an inhibitory (or stimulatory) effect on, for example, cancer marker or tumor antigen expression or activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a cancer marker or tumor antigen substrate. Compounds thus identified can be used to modulate the activity of target gene products (*e.g.*, cancer marker or tumor antigen genes) either directly or indirectly in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions. Compounds that inhibit the activity or expression of cancer markers or tumor antigens are useful in the treatment of proliferative disorders, *e.g.*, cancer, particularly metastatic (*e.g.*, androgen independent) prostate cancer.

In one embodiment, the invention provides assays for screening candidate or test compounds that are substrates of a cancer marker or tumor antigen protein or polypeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds that bind to or modulate the activity of a cancer marker or tumor antigen protein or polypeptide or a biologically active portion thereof.

The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone, which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, *e.g.*, Zuckennann *et al.*, J. Med. Chem. 37: 2678-85 [1994]); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are preferred for use with peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) Anticancer Drug Des. 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.,* Proc. Natl. Acad. Sci. U.S.A. 90:6909 [1993]; Erb *et al.,* Proc. Nalt. Acad. Sci. USA 91:11422 [1994]; Zuckermann *et al.,* J. Med. Chem. 37:2678 [1994]; Cho *et al.,* Science 261:1303 [1993]; Carrell *et al.,* Angew. Chem. Int. Ed. Engl. 33.2059 [1994]; Carell *et al.,* Angew. Chem. Int. Ed. Engl. 33:2061 [1994]; and Gallop *et al.,* J. Med. Chem. 37:1233 [1994].

Libraries of compounds may be presented in solution (*e.g.*, Houghten, Biotechniques 13:412-421 [1992]), or on beads (Lam, Nature 354:82-84 [1991]), chips (Fodor, Nature 364:555-556 [1993]), bacteria or spores (U.S. Patent No. 5,223,409; herein incorporated by reference), plasmids (Cull *et al.,* Proc. Natl. Acad. Sci. USA 89:18651869 [1992]) or on phage (Scott and Smith, Science 249:386-390 [1990]; Devlin Science 249:404-406 [1990]; Cwirla *et al.,* Proc. Natl. Acad. Sci. 87:6378-6382 [1990]; Felici, J. Mol. Biol. 222:301 [1991]).

In one embodiment, an assay is a cell-based assay in which a cell that expresses a cancer marker or tumor antigen protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to the modulate cancer marker's activity is determined. Determining the ability of the test compound to modulate cancer marker activity can be accomplished by monitoring, for example, changes in enzymatic activity. The cell, for example, can be of mammalian origin.

The ability of the test compound to modulate cancer marker or tumor antigen binding to a compound, *e.g.*, a cancer marker substrate, can also be evaluated. This can be accomplished, for example, by coupling the compound, *e.g.*, the substrate, with a radioisotope or enzymatic label such that binding of the compound, *e.g.*, the substrate, to a cancer marker can be determined by detecting the labeled compound, *e.g.*, substrate, in a complex.

Alternatively, the cancer marker or tumor antigen is coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate cancer marker binding to a cancer marker or tumor antigen substrate in a complex. For example, compounds (*e.g.*, substrates) can be labeled with ^{125}I, ^{35}S ^{14}C or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

The ability of a compound (*e.g.*, a cancer marker substrate) to interact with a cancer marker with or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with a cancer marker without the labeling of either the compound or the cancer marker (McConnell *et al.* Science 257:1906-1912 [1992]). As used herein, a "microphysiometer" (*e.g.*, Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and cancer markers.

In yet another embodiment, a cell-free assay is provided in which a cancer marker or tumor antigen protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the cancer marker or tumor antigen protein or biologically active portion thereof is evaluated. Preferred biologically active portions of the cancer marker or tumor antigen proteins to be used in assays of the present invention include fragments that participate in interactions with substrates or other proteins, *e.g.*, fragments with high surface probability scores.

Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

The interaction between two molecules can also be detected, *e.g.*, using fluorescence energy transfer (FRET) (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos *et al.*, U.S. Patent No. 4,968,103; each of which is herein incorporated by reference). A fluorophore label is selected such that a first donor molecule's emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy.

Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in 1 5 the assay should be maximal. An FRET binding event can be conveniently measured through standard fluorometric detection means well known in the art (*e.g.*, using a fluorimeter).

In another embodiment, determining the ability of the cancer marker or tumor antigen protein to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (*see, e.g.*, Sjolander and Urbaniczky, Anal. Chem. 63:2338-2345 [1991] and Szabo *et al.* Curr. Opin. Struct. Biol. 5:699-705 [1995]). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal that can be used as an indication of real-time reactions between biological molecules.

In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

It may be desirable to immobilize cancer markers, an anti-cancer marker antibody or its target molecule to facilitate separation of complexed from non-complexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a cancer marker protein, or interaction of a cancer marker protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase-cancer marker fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione Sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione-derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or cancer marker protein, and the mixture incubated under conditions conducive for complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above.

Alternatively, the complexes can be dissociated from the matrix, and the level of cancer markers binding or activity determined using standard techniques. Other techniques for immobilizing either cancer markers protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated cancer marker protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, EL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (*e.g.*, by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; *e.g.*, using a labeled antibody specific for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, *e.g.*, a labeled anti-IgG antibody).

This assay is performed utilizing antibodies reactive with cancer marker or tumor antigen protein or target molecules but which do not interfere with binding of the cancer markers protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or cancer markers protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the cancer marker protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the cancer marker or tumor antigen protein or target molecule.

Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including, but not limited to: differential centrifugation (see, for example, Rivas and Minton, Trends Biochem Sci 18:284-7 [1993]); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (*see, e.g.*, Ausubel *et al.*, eds. Current Protocols in Molecular Biology 1999, J. Wiley: New York.); and immunoprecipitation (*see*, for example, Ausubel *et al.*, eds. Current Protocols in Molecular Biology 1999, J. Wiley: New York). Such resins and chromatographic techniques are known to one skilled in the art (*See e.g.*, Heegaard J. Mol. Recognit 11:141-8 [1998]; Hageand Tweed J. Chromatogr. Biomed. Sci. Appl 699:499-525 [1997]). Further, fluorescence energy transfer may also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

The assay can include contacting the cancer marker or tumor antigen protein or biologically active portion thereof with a known compound that binds the cancer marker or tumor antigen to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a cancer marker or tumor antigen protein, wherein determining the ability of the test compound to interact with a cancer marker or tumor antigen protein includes determining the ability of the test compound to preferentially bind to cancer markers or tumor antigens or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

To the extent that cancer markers can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins, inhibitors of such an interaction are useful. A homogeneous assay can be used can be used to identify inhibitors.

For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared such that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, *e.g.*, U.S. Patent No. 4,109,496, herein incorporated by reference, that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified. Alternatively, cancer markers protein can be used as a "bait protein" in a two-hybrid assay or three-hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.,* Cell 72:223-232 [1993]; Madura *et al.*, J. Biol. Chem. 268.12046-12054 [1993]; Bartel *et al.*, Biotechniques 14:920-924 [1993]; Iwabuchi *et al.*, Oncogene 8:1693-1696 [1993]; and Brent W0 94/10300; each of which is herein incorporated by reference), to identify other proteins, that bind to or interact with cancer markers or tumor antigens ("cancer marker-binding proteins" or "cancer marker-bp") and are involved in cancer marker or tumor antigen activity. Such cancer marker-bps can be activators or inhibitors of signals by the cancer marker proteins or targets as, for example, downstream elements of a cancer markers-mediated signaling pathway.

Modulators of cancer marker or tumor antigen expression can also be identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of cancer marker or tumor antigen mRNA or protein evaluated relative to the level of expression of cancer marker or tumor antigen mRNA or protein in the absence of the candidate compound. When expression of cancer marker or tumor antigen mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of cancer marker or tumor antigen mRNA or protein expression. Alternatively, when expression of cancer marker or tumor antigen mRNA or protein is less (i.e., statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of cancer marker or tumor antigen mRNA or protein expression. The level of cancer marker or tumor antigen mRNA or protein expression can be determined by methods described herein for detecting cancer marker or tumor antigen mRNA or protein.

A modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a cancer marker or tumor antigen protein can be confirmed *in vivo*, *e.g.*, in an animal such as an animal model for a disease (*e.g.*, an animal with prostate, breast or lung cancer or metastatic prostate, breast, or lung cancer; or an animal harboring a xenograft of a prostate, lung, or breast cancer from an animal (*e.g.*, human) or cells from a cancer resulting from metastasis of a prostate, breast, or lung cancer (*e.g.*, to a lymph node, bone, or liver), or cells from a prostate, breast, or lung cancer cell line.

This invention further pertains to novel agents identified by the above-described screening assays (*See e.g.*, below description of cancer therapies). Accordingly, it is within the scope of this invention to further use an agent identified as described herein (*e.g.*, a cancer marker modulating agent, an antisense cancer marker nucleic acid molecule, a siRNA molecule, a cancer marker specific antibody, or a cancer marker-binding partner) in an appropriate animal model (such as those described herein) to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be, *e.g.*, used for treatments as described herein.

IV. Cancer Therapies

In some embodiments, the present invention provides therapies for cancer (*e.g.*, prostate cancer). In some embodiments, therapies target cancer markers or tumor antigens identified using the phage array profiling methods of the present invention (e.g., BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70)).

A. Immunotherapy

The tumor antigens identified during the development of the present invention find use in cancer immunotherapy. Such methods are improvements over the non-specific chemotherapeutic cancer therapies currently available. For example, in some embodiments, tumor antigens are used to generate therapeutic antibodies. In other embodiments, the tumor antigens of the present invention find use in the generation of cancer vaccines.

i) Antibody Immunotherapy

In some embodiments, the present invention provides therapy for cancer comprising the administration of therapeutic antibodies (*See e.g.*, U.S. Patents 6,180,357; and 6,051,230; both of which are herein incorporated by reference).

In some embodiments, the therapeutic antibodies comprise an antibody generated against a tumor antigen of the present invention (e.g., BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70)) conjugated to a cytotoxic agent. Such antibodies are particularly suited for targeting tumor antigens expressed on tumor cells but not normal cells. In such embodiments, a tumor specific therapeutic agent is generated that does not target normal cells, thus reducing many of the detrimental side effects of traditional chemotherapy. For certain applications, it is envisioned that the therapeutic agents will be pharmacologic agents will serve as useful agents for attachment to antibodies or growth factors, particularly cytotoxic or otherwise anticellular agents having the ability to kill or suppress the growth or cell division of endothelial cells. The present invention contemplates the use of any pharmacologic agent that can be conjugated to an antibody, and delivered in active form. Exemplary anticellular agents include chemotherapeutic agents, radioisotopes, and cytotoxins. The therapeutic antibodies of the present invention may include a variety of cytotoxic moieties, including but not limited to, radioactive isotopes (*e.g.*, iodine-131, iodine-123, technicium-99m, indium-111, rhenium-188, rhenium-186, gallium-67, copper-67, yttrium-90, iodine-125 or astatine-211), hormones such as a steroid, antimetabolites such as cytosines (*e.g.*, arabinoside, fluorouracil, methotrexate or aminopterin; an anthracycline; mitomycin C), vinca alkaloids (*e.g.*, demecolcine; etoposide; mithramycin), and antitumor alkylating agent such as chlorambucil or melphalan. Other embodiments may include agents such as a coagulant, a cytokine, growth factor, bacterial endotoxin or the lipid A moiety of bacterial endotoxin. For example, in some embodiments, therapeutic agents will include plant-, fungus- or bacteria-derived toxin, such as an A chain toxins, a ribosome inactivating protein, α-sarcin, aspergillin, restrictocin, a ribonuclease, diphtheria toxin or pseudomonas exotoxin, to mention just a few examples. In some preferred embodiments, deglycosylated ricin A chain is utilized.

In any event, it is proposed that agents such as these may, if desired, be successfully conjugated to an antibody, in a manner that will allow their targeting, internalization, release or presentation to blood components at the site of the targeted tumor cells as required using known conjugation technology (*See, e.g.*, Ghose *et al.*, Methods Enzymol., 93:280 [1983]).

For example, in some embodiments the present invention provides immunotoxins targeted to tumor antigens of the present invention. Immunotoxins are conjugates of a specific targeting agent typically a tumor-directed antibody or fragment, with a cytotoxic agent, such as a toxin moiety. The targeting agent directs the toxin to, and thereby selectively kills, cells carrying the targeted antigen. In some embodiments, therapeutic antibodies employ crosslinkers that provide high *in vivo* stability (Thorpe *et al.*, Cancer Res., 48:6396 [1988]).

In other embodiments, particularly those involving treatment of solid tumors, antibodies are designed to have a cytotoxic or otherwise anticellular effect against the tumor vasculature, by suppressing the growth or cell division of the vascular endothelial cells. This attack is intended to lead to a tumor-localized vascular collapse, depriving the tumor cells, particularly those tumor cells distal of the vasculature, of oxygen and nutrients, ultimately leading to cell death and tumor necrosis.

In preferred embodiments, antibody based therapeutics are formulated as pharmaceutical compositions and described above. In preferred embodiments, administration of an antibody composition of the present invention results in a measurable decrease in cancer (*e.g.*, decrease or elimination of tumor).

ii) Cancer Vaccines

In some embodiments, the present invention provides cancer vaccines directed against a specific cancer. Cancer vaccines induce a systemic tumor-specific immune response. Such a response is capable of eradicating tumor cells anywhere in the body (*e.g.*, metastatic tumor cells). Methods for generating tumor vaccines are well known in the art (*See e.g.*, U.S. Patents 5,994,523; 5,972,334; 5,904,920; 5,674,486; and 6,207,147; each of which is herein incorporated by reference).

In some embodiments, tumor vaccines are administered when cancer is first detected (*e.g.*, concurrently with other therapeutics such as chemotherapy). In other embodiments, cancer vaccines are administered following treatment (*e.g.*, surgical resection or chemotherapy) to prevent relapse or metastases. In yet other embodiments, cancer vaccines are administered prophylactically (*e.g.*, to those at risk of a certain cancer).

In some embodiments, the cancer vaccines of the present invention comprise one or more tumor antigens in a pharmaceutical composition (*e.g.*, those described above). In some embodiments, the tumor antigen is inactivated prior to administration. In other embodiments, the vaccine further comprises one or more additional therapeutic agents (*e.g.*, cytokines or cytokine expressing cells).

In some embodiments (*e.g.*, the method described in U.S. Patent 5,674,486, herein incorporated by reference), selected cells from a patient, such as fibroblasts, obtained, for example, from a routine skin biopsy, are genetically modified to express one or more cytokines. Alternatively, patient cells that may normally serve as antigen presenting cells in the immune system such as macrophages, monocytes, and lymphocytes may also be genetically modified to express one or more cytokines. The cytokine expressing cells are then mixed with the patient's tumor antigens (*e.g.*, a tumor antigen of the present invention), for example in the form of irradiated tumor cells, or alternatively in the form of purified natural or recombinant tumor antigen, and employed in immunizations, for example subcutaneously, to induce systemic anti-tumor immunity.

The vaccines of the present invention may be administered using any suitable method, including but not limited to, those described above. In preferred embodiments, administration of a cancer vaccine of the present invention results in elimination (*e.g.*, decrease or elimination of tumors) or prevention of detectable cancer cells.

B. Antisense Therapies

In some embodiments, the present invention targets the expression of cancer markers. For example, in some embodiments, the present invention employs compositions comprising oligomeric antisense compounds, particularly oligonucleotides (*e.g.*, those identified in the drug screening methods described above), for use in modulating the function of nucleic acid molecules encoding cancer markers of the present invention (e.g., BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70)), ultimately modulating the amount of cancer marker expressed. This is accomplished by providing antisense compounds that specifically hybridize with one or more nucleic acids encoding cancer markers of the present invention. The specific hybridization of an oligomeric compound with its target nucleic acid interferes with the normal function of the nucleic acid. This modulation of function of a target nucleic acid by compounds that specifically hybridize to it is generally referred to as "antisense." The functions of DNA to be interfered with include replication and transcription. The functions of RNA to be interfered with include all vital functions such as, for example, translocation of the RNA to the site of protein translation, translation of protein from the RNA, splicing of the RNA to yield one or more mRNA species, and catalytic activity that may be engaged in or facilitated by the RNA. The overall effect of such interference with target nucleic acid function is modulation of the expression of cancer markers of the present invention. In the context of the present invention, "modulation" means either an increase (stimulation) or a decrease (inhibition) in the expression of a gene. For example, expression may be inhibited to potentially prevent tumor proliferation.

It is preferred to target specific nucleic acids for antisense. "Targeting" an antisense compound to a particular nucleic acid, in the context of the present invention, is a multistep process. The process usually begins with the identification of a nucleic acid sequence whose function is to be modulated. This may be, for example, a cellular gene (or mRNA transcribed from the gene) whose expression is associated with a particular disorder or disease state, or a nucleic acid molecule from an infectious agent. In the present invention, the target is a nucleic acid molecule encoding a cancer marker of the present invention. The targeting process also includes determination of a site or sites within this gene for the antisense interaction to occur such that the desired effect, *e.g.*, detection or modulation of expression of the protein, will result. Within the context of the present invention, a preferred intragenic site is the region encompassing the translation initiation or termination codon of the open reading frame (ORF) of the gene. Since the translation initiation codon is typically 5'-AUG (in transcribed mRNA molecules; 5'-ATG in the corresponding DNA molecule), the translation initiation codon is also referred to as the "AUG codon," the "start codon" or the "AUG start codon". A minority of genes have a translation initiation codon having the RNA sequence 5'-GUG, 5'-UUG or 5'-CUG, and 5'-AUA, 5'-ACG and 5'-CUG have been shown to function in vivo. Thus, the terms "translation initiation codon" and "start codon" can encompass many codon sequences, even though the initiator amino acid in each instance is typically methionine (in eukaryotes) or formylmethionine (in prokaryotes). Eukaryotic and prokaryotic genes may have two or more alternative start codons, any one of which may be preferentially utilized for translation initiation in a particular cell type or tissue, or under a particular set of conditions. In the context of the present invention, "start codon" and "translation initiation codon" refer to the codon or codons that are used *in vivo* to initiate translation of an mRNA molecule transcribed from a gene encoding a tumor antigen of the present invention, regardless of the sequence(s) of such codons.

Translation termination codon (or "stop codon") of a gene may have one of three sequences (*i.e.*, 5'-UAA, 5'-UAG and 5'-UGA; the corresponding DNA sequences are 5'-TAA, 5'-TAG and 5'-TGA, respectively). The terms "start codon region" and "translation initiation codon region" refer to a portion of such an mRNA or gene that encompasses from about 25 to about 50 contiguous nucleotides in either direction (*i.e.*, 5' or 3') from a translation initiation codon. Similarly, the terms "stop codon region" and "translation termination codon region" refer to a portion of such an mRNA or gene that encompasses from about 25 to about 50 contiguous nucleotides in either direction (*i.e.*, 5' or 3') from a translation termination codon.

The open reading frame (ORF) or "coding region," which refers to the region between the translation initiation codon and the translation termination codon, is also a region that may be targeted effectively. Other target regions include the 5' untranslated region (5' UTR), referring to the portion of an mRNA in the 5' direction from the translation initiation codon, and thus including nucleotides between the 5' cap site and the translation initiation codon of an mRNA or corresponding nucleotides on the gene, and the 3' untranslated region (3' UTR), referring to the portion of an mRNA in the 3' direction from the translation termination codon, and thus including nucleotides between the translation termination codon and 3' end of an mRNA or corresponding nucleotides on the gene. The 5' cap of an mRNA comprises an N7-methylated guanosine residue joined to the 5'-most residue of the mRNA via a 5'-5' triphosphate linkage. The 5' cap region of an mRNA is considered to include the 5' cap structure itself as well as the first 50 nucleotides adjacent to the cap. The cap region may also be a preferred target region.

Although some eukaryotic mRNA transcripts are directly translated, many contain one or more regions, known as "introns," that are excised from a transcript before it is translated. The remaining (and therefore translated) regions are known as "exons" and are spliced together to form a continuous mRNA sequence. mRNA splice sites (*i.e.*, intron-exon junctions) may also be preferred target regions, and are particularly useful in situations where aberrant splicing is implicated in disease, or where an overproduction of a particular mRNA splice product is implicated in disease. Aberrant fusion junctions due to rearrangements or deletions are also preferred targets. It has also been found that introns can also be effective, and therefore preferred, target regions for antisense compounds targeted, for example, to DNA or pre-mRNA.

In some embodiments, target sites for antisense inhibition are identified using commercially available software programs (e.g., Biognostik, Gottingen, Germany; SysArris Software, Bangalore, India; Antisense Research Group, University of Liverpool, Liverpool, England; GeneTrove, Carlsbad, CA). In other embodiments, target sites for antisense inhibition are identified using the accessible site method described in U.S. Patent WO0198537A2, herein incorporated by reference.

Once one or more target sites have been identified, oligonucleotides are chosen that are sufficiently complementary to the target (*i.e.*, hybridize sufficiently well and with sufficient specificity) to give the desired effect. For example, in preferred embodiments of the present invention, antisense oligonucleotides are targeted to or near the start codon.

In the context of this invention, "hybridization," with respect to antisense compositions and methods, means hydrogen bonding, which may be Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary nucleoside or nucleotide bases. For example, adenine and thymine are complementary nucleobases that pair through the formation of hydrogen bonds. It is understood that the sequence of an antisense compound need not be 100% complementary to that of its target nucleic acid to be specifically hybridizable. An antisense compound is specifically hybridizable when binding of the compound to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA to cause a loss of utility, and there is a sufficient degree of complementarity to avoid non-specific binding of the antisense compound to non-target sequences under conditions in which specific binding is desired (*i.e.*, under physiological conditions in the case of *in vivo* assays or therapeutic treatment, and in the case of *in vitro* assays, under conditions in which the assays are performed).

Antisense compounds are commonly used as research reagents and diagnostics. For example, antisense oligonucleotides, which are able to inhibit gene expression with specificity, can be used to elucidate the function of particular genes. Antisense compounds are also used, for example, to distinguish between functions of various members of a biological pathway.

The specificity and sensitivity of antisense is also applied for therapeutic uses. For example, antisense oligonucleotides have been employed as therapeutic moieties in the treatment of disease states in animals and man. Antisense oligonucleotides have been safely and effectively administered to humans and numerous clinical trials are presently underway. It is thus established that oligonucleotides are useful therapeutic modalities that can be configured to be useful in treatment regimes for treatment of cells, tissues, and animals, especially humans.

While antisense oligonucleotides are a preferred form of antisense compound, the present invention comprehends other oligomeric antisense compounds, including but not limited to oligonucleotide mimetics such as are described below. The antisense compounds in accordance with this invention preferably comprise from about 8 to about 30 nucleobases (*i.e.*, from about 8 to about 30 linked bases), although both longer and shorter sequences may find use with the present invention. Particularly preferred antisense compounds are antisense oligonucleotides, even more preferably those comprising from about 12 to about 25 nucleobases.

Specific examples of preferred antisense compounds useful with the present invention include oligonucleotides containing modified backbones or non-natural internucleoside linkages. As defined in this specification, oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone. For the purposes of this specification, modified oligonucleotides that do not have a phosphorus atom in their internucleoside backbone can also be considered to be oligonucleosides.

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

Preferred modified oligonucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage (*i.e.*, the backbone) of the nucleotide units are replaced with novel groups. The base units are maintained for hybridization with an appropriate nucleic acid target compound. One such oligomeric compound, an oligonucleotide mimetic that has been shown to have excellent hybridization properties, is referred to as a peptide nucleic acid (PNA). In PNA compounds, the sugar-backbone of an oligonucleotide is replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The nucleobases are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone. Representative United States patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Pat. Nos.: 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Further teaching of PNA compounds can be found in Nielsen *et al.*, Science 254:1497 (1991).

Most preferred embodiments of the invention are oligonucleotides with phosphorothioate backbones and oligonucleosides with heteroatom backbones, and in particular --CH_2, --NH--O--CH_2--, --CH_2--N(CH_3)--O--CH_2-- [known as a methylene (methylimino) or MMI backbone], --CH_2--O--N(CH_3)--CH_2--, --CH_2--N(CH_3)--N(CH_3)--CH_2--, and --O--N(CH_3)--CH_2--CH_2-- [wherein the native phosphodiester backbone is represented as --O--P--O--CH_2--] of the above referenced U.S. Pat. No. 5,489,677, and the amide backbones of the above referenced U.S. Pat. No. 5,602,240. Also preferred are oligonucleotides having morpholino backbone structures of the above-referenced U.S. Pat. No. 5,034,506.

Modified oligonucleotides may also contain one or more substituted sugar moieties. Preferred oligonucleotides comprise one of the following at the 2' position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C_1 to C_{10} alkyl or C_2 to C_{10} alkenyl and alkynyl. Particularly preferred are $O[(CH_2)_nO]_mCH_3$, $O(CH_2)_nOCH_3$, $O(CH_2)_nNH_2$, $O(CH_2)_nCH_3$, $O(CH_2)_nONH_2$, and $O(CH_2)_nON[(CH_2)_nCH_3)]_2$, where n and m are from 1 to about 10. Other preferred oligonucleotides comprise one of the following at the 2' position: C_1 to C_{10} lower alkyl, substituted lower alkyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, SO_2CH_3, ONO_2, NO_2, N_3, NH_2, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. A preferred modification includes 2'-methoxyethoxy (2'-O--$CH_2CH_2OCH_3$, also known as 2'-O-(2-methoxyethyl) or 2'-MOE) (Martin *et al.*, Helv. Chim. Acta 78:486 [1995]) *i.e.*, an alkoxyalkoxy group. A further preferred modification includes 2'-dimethylaminooxyethoxy (*i.e.*, a $O(CH_2)_2ON(CH_3)_2$ group), also known as 2'-DMAOE, and 2'-dimethylaminoethoxyethoxy (also known in the art as 2'-O-dimethylaminoethoxyethyl or 2'-DMAEOE), *i.e.*, 2'-O--CH_2--O--CH_2--$N(CH_2)_2$.

Other preferred modifications include 2'-methoxy(2'-O--CH_3), 2'-aminopropoxy(2'-$OCH_2CH_2CH_2NH_2$) and 2'-fluoro (2'-F). Similar modifications may also be made at other positions on the oligonucleotide, particularly the 3' position of the sugar on the 3' terminal nucleotide or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Oligonucleotides may also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar.

Oligonucleotides may also include nucleobase (often referred to in the art simply as "base") modifications or substitutions. As used herein, "unmodified" or "natural" nucleobases include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified nucleobases include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Further nucleobases include those disclosed in U.S. Pat. No. 3,687,808. Certain of these nucleobases are particularly useful for increasing the binding affinity of the oligomeric compounds of the invention. These include 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2. degree ^{o}C and are presently preferred base substitutions, even more particularly when combined with 2'-O-methoxyethyl sugar modifications.

Another modification of the oligonucleotides of the present invention involves chemically linking to the oligonucleotide one or more moieties or conjugates that enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. Such moieties include but are not limited to lipid moieties such as a cholesterol moiety, cholic acid, a thioether, (*e.g.*, hexyl-S-tritylthiol), a thiocholesterol, an aliphatic chain, (*e.g.*, dodecandiol or undecyl residues), a phospholipid, (*e.g.*, di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate), a polyamine or a polyethylene glycol chain or adamantane acetic acid, a palmityl moiety, or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety.

One skilled in the relevant art knows well how to generate oligonucleotides containing the above-described modifications. The present invention is not limited to the antisense oligonucleotides described above. Any suitable modification or substitution may be utilized.

It is not necessary for all positions in a given compound to be uniformly modified, and in fact more than one of the aforementioned modifications may be incorporated in a single compound or even at a single nucleoside within an oligonucleotide. The present invention also includes antisense compounds that are chimeric compounds. "Chimeric" antisense compounds or "chimeras," in the context of the present invention, are antisense compounds, particularly oligonucleotides, which contain two or more chemically distinct regions, each made up of at least one monomer unit, *i.e.*, a nucleotide in the case of an oligonucleotide compound. These oligonucleotides typically contain at least one region wherein the oligonucleotide is modified so as to confer upon the oligonucleotide increased resistance to nuclease degradation, increased cellular uptake, and/or increased binding affinity for the target nucleic acid. An additional region of the oligonucleotide may serve as a substrate for enzymes capable of cleaving RNA:DNA or RNA:RNA hybrids. By way of example, RNaseH is a cellular endonuclease that cleaves the RNA strand of an RNA:DNA duplex. Activation of RNase H, therefore, results in cleavage of the RNA target, thereby greatly enhancing the efficiency of oligonucleotide inhibition of gene expression. Consequently, comparable results can often be obtained with shorter oligonucleotides when chimeric oligonucleotides are used, compared to phosphorothioate deoxyoligonucleotides hybridizing to the same target region. Cleavage of the RNA target can be routinely detected by gel electrophoresis and, if necessary, associated nucleic acid hybridization techniques known in the art.

Chimeric antisense compounds of the present invention may be formed as composite structures of two or more oligonucleotides, modified oligonucleotides, oligonucleosides and/or oligonucleotide mimetics as described above.

The present invention also includes pharmaceutical compositions and formulations that include the antisense compounds of the present invention as described below.

C. RNAi Therapies

In other embodiments, RNAi is used to regulate expression of tumor antigens or cancer markers of the present invention. RNAi represents an evolutionary conserved cellular defense for controlling the expression of foreign genes in most eukaryotes, including humans. RNAi is triggered by double-stranded RNA (dsRNA) and causes sequence-specific mRNA degradation of single-stranded target RNAs homologous in response to dsRNA. The mediators of mRNA degradation are small interfering RNA duplexes (siRNAs), which are normally produced from long dsRNA by enzymatic cleavage in the cell. siRNAs are generally approximately twenty-one nucleotides in length (e.g. 21-23 nucleotides in length), and have a base-paired structure characterized by two nucleotide 3'-overhangs. Following the introduction of a small RNA, or RNAi, into the cell, it is believed the sequence is delivered to an enzyme complex called RISC (RNA-induced silencing complex). RISC recognizes the target and cleaves it with an endonuclease. It is noted that if larger RNA sequences are delivered to a cell, RNase III enzyme (Dicer) converts longer dsRNA into 21-23 nt ds siRNA fragments.

Chemically synthesized siRNAs have become powerful reagents for genome-wide analysis of mammalian gene function in cultured somatic cells. Beyond their value for validation of gene function, siRNAs also hold great potential as gene-specific therapeutic agents (Tuschl and Borkhardt, Molecular Intervent. 2002; 2(3):158-67, herein incorporated by reference).

The transfection of siRNAs into animal cells results in the potent, long-lasting post-transcriptional silencing of specific genes (Caplen et al, Proc Natl Acad Sci U.S.A. 2001; 98: 9742–7; Elbashir et al., Nature. 2001; 411:494–8; Elbashir et al., Genes Dev. 2001;15: 188–200; and Elbashir et al., EMBO J. 2001; 20: 6877–88, all of which are herein incorporated by reference). Methods and compositions for performing RNAi with siRNAs are described, for example, in U.S. Patent 6,506,559, herein incorporated by reference.

siRNAs are extraordinarily effective at lowering the amounts of targeted RNA, and by extension proteins, frequently to undetectable levels. The silencing effect can last several months, and is extraordinarily specific, because one nucleotide mismatch between the target RNA and the central region of the siRNA is frequently sufficient to prevent silencing

Brummelkamp et al, Science 2002; 296:550–3; and Holen et al, Nucleic Acids Res. 2002; 30:1757–66, both of which are herein incorporated by reference.

C. Genetic Therapies

The present invention contemplates the use of any genetic manipulation for use in modulating the expression of cancer markers (e.g., BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70)) of the present invention. Examples of genetic manipulation include, but are not limited to, gene knockout (*e.g.*, removing the cancer marker gene from the chromosome using, for example, recombination), expression of antisense constructs with or without inducible promoters, and the like. Delivery of nucleic acid construct to cells *in vitro* or *in vivo* may be conducted using any suitable method. A suitable method is one that introduces the nucleic acid construct into the cell such that the desired event occurs (*e.g.*, expression of an antisense construct).

Introduction of molecules carrying genetic information into cells is achieved by any of various methods including, but not limited to, directed injection of naked DNA constructs, bombardment with gold particles loaded with said constructs, and macromolecule mediated gene transfer using, for example, liposomes, biopolymers, and the like. Preferred methods use gene delivery vehicles derived from viruses, including, but not limited to, adenoviruses, retroviruses, vaccinia viruses, and adeno-associated viruses. Because of the higher efficiency as compared to retroviruses, vectors derived from adenoviruses are the preferred gene delivery vehicles for transferring nucleic acid molecules into host cells *in vivo*. Adenoviral vectors have been shown to provide very efficient *in vivo* gene transfer into a variety of solid tumors in animal models and into human solid tumor xenografts in immune-deficient mice. Examples of adenoviral vectors and methods for gene transfer are described in PCT publications WO 00/12738 and WO 00/09675 and U.S. Pat. Appl. Nos. 6,033,908, 6,019,978, 6,001,557, 5,994,132, 5,994,128, 5,994,106, 5,981,225, 5,885,808, 5,872,154, 5,830,730, and 5,824,544, each of which is herein incorporated by reference in its entirety.

Vectors may be administered to subject in a variety of ways. For example, in some embodiments of the present invention, vectors are administered into tumors or tissue associated with tumors using direct injection. In other embodiments, administration is via the blood or lymphatic circulation (*See e.g.*, PCT publication 99/02685 herein incorporated by reference in its entirety). Exemplary dose levels of adenoviral vector are preferably 10^8 to 10^{11} vector particles added to the perfusate.

V. Pharmaceutical Compositions

In some embodiments, the present invention provides pharmaceutical compositions that may comprise all or portions of tumor antigen or cancer marker polynucleotide sequences, tumor antigen polypeptides, inhibitors or antagonists of tumor antigen bioactivity, including antibodies, alone or in combination with at least one other agent, such as a stabilizing compound, and may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The pharmaceutical compositions find use as therapeutic agents and vaccines for the treatment of cancer.

The methods of the present invention find use in treating cancers as described in greater detail above. Antibodies can be administered to the patient intravenously in a pharmaceutically acceptable carrier such as physiological saline. Standard methods for intracellular delivery of antibodies can be used (*e.g.*, delivery via liposome). Such methods are well known to those of ordinary skill in the art. The formulations of this invention are useful for parenteral administration, such as intravenous, subcutaneous, intramuscular, and intraperitoneal.

As is well known in the medical arts, dosages for any one patient depends upon many factors, including the patient's size, body surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and interaction with other drugs being concurrently administered.

Accordingly, in some embodiments of the present invention, compositions (*e.g.*, antibodies and vaccines) can be administered to a patient alone, or in combination with other nucleotide sequences, drugs or hormones or in pharmaceutical compositions where it is mixed with excipient(s) or other pharmaceutically acceptable carriers. In one embodiment of the present invention, the pharmaceutically acceptable carrier is pharmaceutically inert. In another embodiment of the present invention, compositions may be administered alone to individuals suffering from cancer.

Depending on the type of cancer being treated, these pharmaceutical compositions may be formulated and administered systemically or locally. Techniques for formulation and administration may be found in the latest edition of "Remington's Pharmaceutical Sciences" (Mack Publishing Co, Easton Pa.). Suitable routes may, for example, include oral or transmucosal administration; as well as parenteral delivery, including intramuscular, subcutaneous, intramedullary, intrathecal, intraventricular, intravenous, intraperitoneal, or intranasal administration.

For injection, the pharmaceutical compositions of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. For tissue or cellular administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

In other embodiments, the pharmaceutical compositions of the present invention can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral or nasal ingestion by a patient to be treated.

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. For example, an effective amount of antibody or vaccine may be that amount that decreases the presence of cancerous cells (*e.g.*, shrinks or eliminates a tumor or reduces the number of circulating cancer cells). Determination of effective amounts is well within the capability of those skilled in the art, especially in light of the disclosure provided herein.

In addition to the active ingredients these pharmaceutical compositions may contain suitable pharmaceutically acceptable carriers comprising excipients and auxiliaries that facilitate processing of the active compounds into preparations that can be used pharmaceutically. The preparations formulated for oral administration may be in the form of tablets, dragees, capsules, or solutions.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is itself known (*e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes).

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical preparations for oral use can be obtained by combining the active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, etc; cellulose such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; and gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid or a salt thereof such as sodium alginate.

Dragee cores are provided with suitable coatings such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, (*i.e.*, dosage).

Pharmaceutical preparations that can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients mixed with filler or binders such as lactose or starches, lubricants such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycol with or without stabilizers.

Compositions comprising a compound of the invention formulated in a pharmaceutical acceptable carrier may be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition. For antibodies to a tumor antigen of the present invention, conditions indicated on the label may include treatment of conditions related to cancer.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, *etc.* Salts tend to be more soluble in aqueous or other protonic solvents that are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder in 1 mM-50 mM histidine, 0.1%-2% sucrose, 2%-7% mannitol at a pH range of 4.5 to 5.5 that is combined with buffer prior to use.

For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. Then, preferably, dosage can be formulated in animal models (particularly murine models) to achieve a desirable circulating concentration range that adjusts antibody levels.

A therapeutically effective dose refers to that amount of antibody that ameliorates symptoms of the disease state. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index, and it can be expressed as the ratio LD_{50}/ED_{50}. Compounds that exhibit large therapeutic indices are preferred. The data obtained from these cell culture assays and additional animal studies can be used in formulating a range of dosage for human use. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage is chosen by the individual physician in view of the patient to be treated. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Additional factors which may be taken into account include the severity of the disease state; age, weight, and gender of the patient; diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long acting pharmaceutical compositions might be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature (*See*, U.S. Pat. Nos. 4,657,760; 5,206,344; or 5,225,212, all of which are herein incorporated by reference).

In some embodiments, the pharmaceutical compositions of the present invention further include one or more agents useful in the treatment of cancer. For example, in some embodiments, one or more antibodies or vaccines are combined with a chemotherapeutic agent. Chemotherapeutic agents are well known to those of skill in the art. Examples of such chemotherapeutics include alkylating agents, antibiotics, antimetabolitic agents, plant-derived agents, and hormones. Among the suitable alkylating agents are nitrogen mustards, such as cyclophosphamide, aziridines, alkyl alkone sulfonates, nitrosoureas, nonclassic alkylating agents, such as dacarbazine, and platinum compounds, such as carboplatin and cisplatin. Among the suitable antibiotic agents are dactinomycin, bleomycin, mitomycin C, plicamycin, and the anthracyclines, such as doxorubicin (also known as adriamycin) and mitoxantrone. Among the suitable antimetabolic agents are antifols, such as methotrexate, purine analogues, pyrimidine analogues, such as 5-fluorouracil (5-FU) and cytarabine, enzymes, such as the asparaginases, and synthetic agents, such as hydroxyurea. Among the suitable plant-derived agents are vinca alkaloids, such as vincristine and vinblastine, taxanes, epipodophyllotoxins, such as etoposide, and camptothecan. Among suitable hormones are steroids. Currently, the preferred drug is adriamycin. However, other suitable chemotherapeutic agents, including additional agents within the groups of agents identified above, may be readily determined by one of skill in the art depending upon the type of cancer being treated, the condition of the human or veterinary patient, and the like.

Suitable dosages for the selected chemotherapeutic agent are known to those of skill in the art. One of skill in the art can readily adjust the route of administration, the number of doses received, the timing of the doses, and the dosage amount, as needed. Such a dose, which may be readily adjusted depending upon the particular drug or agent selected, may be administered by any suitable route, including but not limited to, those described above. Doses may be repeated as needed.

VI. Transgenic Animals Expressing Cancer Marker Genes or Knockouts

The present invention contemplates the generation of transgenic animals comprising an exogenous cancer marker or tumor antigen (BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70)) gene of the present invention or mutants and variants thereof (*e.g.*, truncations or single nucleotide polymorphisms). In other embodiments, the transgenic animals comprise a knock-out of a cancer marker or tumor antigen gene. In preferred embodiments, the transgenic animal displays an altered phenotype (*e.g.*, increased or decreased presence of markers) as compared to wild-type animals. Methods for analyzing the presence or absence of such phenotypes include but are not limited to, those disclosed herein. In some preferred embodiments, the transgenic animals further display an increased or decreased growth of tumors or evidence of cancer.

The transgenic animals of the present invention find use in drug (*e.g.*, cancer therapy) screens. In some embodiments, test compounds (*e.g.*, a drug that is suspected of being useful to treat cancer) and control compounds (*e.g.*, a placebo) are administered to the transgenic animals and the control animals and the effects evaluated.

The transgenic animals can be generated via a variety of methods. In some embodiments, embryonal cells at various developmental stages are used to introduce transgenes for the production of transgenic animals. Different methods are used depending on the stage of development of the embryonal cell. The zygote is the best target for micro-injection. In the mouse, the male pronucleus reaches the size of approximately 20 micrometers in diameter that allows reproducible injection of 1-2 picoliters (pl) of DNA solution. The use of zygotes as a target for gene transfer has a major advantage in that in most cases the injected DNA will be incorporated into the host genome before the first cleavage (Brinster *et al.*, Proc. Natl. Acad. Sci. USA 82:4438-4442 [1985]). As a consequence, all cells of the transgenic non-human animal will carry the incorporated transgene. This will in general also be reflected in the efficient transmission of the transgene to offspring of the founder since 50% of the germ cells will harbor the transgene. U.S. Patent No. 4,873,191 describes a method for the micro-injection of zygotes; the disclosure of this patent is incorporated herein in its entirety.

In other embodiments, retroviral infection is used to introduce transgenes into a non-human animal. In some embodiments, the retroviral vector is utilized to transfect oocytes by injecting the retroviral vector into the perivitelline space of the oocyte (U.S. Pat. No. 6,080,912, incorporated herein by reference). In other embodiments, the developing non-human embryo can be cultured *in vitro* to the blastocyst stage. During this time, the blastomeres can be targets for retroviral infection (Janenich, Proc. Natl. Acad. Sci. USA 73:1260 [1976]). Efficient infection of the blastomeres is obtained by enzymatic treatment to remove the zona pellucida (Hogan *et al.*, in *Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [1986]). The viral vector system used to introduce the transgene is typically a replication-defective retrovirus carrying the transgene (Jahner *et al.*, Proc. Natl. Acad Sci. USA 82:6927 [1985]). Transfection is easily and efficiently obtained by culturing the blastomeres on a monolayer of virus-producing cells (Stewart, *et al.*, EMBO J., 6:383 [1987]). Alternatively, infection can be performed at a later stage. Virus or virus-producing cells can be injected into the blastocoele (Jahner *et al.*, Nature 298:623 [1982]). Most of the founders will be mosaic for the transgene since incorporation occurs only in a subset of cells that form the transgenic animal. Further, the founder may contain various retroviral insertions of the transgene at different positions in the genome that generally will segregate in the offspring. In addition, it is also possible to introduce transgenes into the germline, albeit with low efficiency, by intrauterine retroviral infection of the midgestation embryo (Jahner *et al.*, *supra* [1982]). Additional means of using retroviruses or retroviral vectors to create transgenic animals known to the art involve the micro-injection of retroviral particles or mitomycin C-treated cells producing retrovirus into the perivitelline space of fertilized eggs or early embryos (PCT International Application WO 90/08832 [1990], and Haskell and Bowen, Mol. Reprod. Dev., 40:386 [1995]).

In other embodiments, the transgene is introduced into embryonic stem cells and the transfected stem cells are utilized to form an embryo. ES cells are obtained by culturing pre-implantation embryos *in vitro* under appropriate conditions (Evans *et al.*, Nature 292:154 [1981]; Bradley *et al.*, Nature 309:255 [1984]; Gossler *et al.*, Proc. Acad. Sci. USA 83:9065 [1986]; and Robertson *et al.*, Nature 322:445 [1986]). Transgenes can be efficiently introduced into the ES cells by DNA transfection by a variety of methods known to the art including calcium phosphate co-precipitation, protoplast or spheroplast fusion, lipofection and DEAE-dextran-mediated transfection. Transgenes may also be introduced into ES cells by retrovirus-mediated transduction or by micro-injection. Such transfected ES cells can thereafter colonize an embryo following their introduction into the blastocoel of a blastocyst-stage embryo and contribute to the germ line of the resulting chimeric animal (for review, *See*, Jaenisch, Science 240:1468 [1988]).

Prior to the introduction of transfected ES cells into the blastocoel, the transfected ES cells may be subjected to various selection protocols to enrich for ES cells which have integrated the transgene assuming that the transgene provides a means for such selection. Alternatively, the polymerase chain reaction may be used to screen for ES cells that have integrated the transgene. This technique obviates the need for growth of the transfected ES cells under appropriate selective conditions prior to transfer into the blastocoel.

In still other embodiments, homologous recombination is utilized to knock-out gene function or create deletion mutants (*e.g.*, truncation mutants). Methods for homologous recombination are described in U.S. Pat. No. 5,614,396, incorporated herein by reference.

EXPERIMENTAL

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

In the experimental disclosure which follows, the following abbreviations apply: N (normal); M (molar); mM (millimolar); μM (micromolar); mol (moles); mmol (millimoles); μmol (micromoles); nmol (nanomoles); pmol (picomoles); g (grams); mg (milligrams); μg (micrograms); ng (nanograms); l or L (liters); ml (milliliters); μl (microliters); cm (centimeters); mm (millimeters); μm (micrometers); nm (nanometers); and°C (degrees Centigrade).

Example 1

Phage Array Profiling of Prostate Cancer

This Example describes a phage array profiling method of the present invention as applied to prostate cancer.

A. Methods

Patient Population and Samples. At the time of diagnosis and prior to radical prostatectomy, sera from biopsy-proven clinically localized prostate cancer participants were collected by the University of Michigan Specialized Research Program in Prostate Cancer (SPORE) tissue/serum bank between January 1995 to January of 2003. The average age of all prostate cancer patients was 59.6 (range 41-74). For post-prostatectomy prostate cancer patients, the average age and PSA value were 58.1 and 0.169 ng/ml respectively. Sera from lung adenocarcinoma patients (average age 53.9) without any known history of prostate cancer were used. As controls, serum samples from 85 age-matched males (average age 62.5, range 50-80) with no known history of cancer were used for the study. All sera were stored in aliquots at -20°C until use.

Construction of T7 phage display prostate cancer cDNA libraries. Total RNA was isolated separately from six prostate cancer tissue samples according to the standard Trizol protocol (Dhanasekaran et al., Nature 412, 822-826. (2001)). The integrity of each RNA preparation was assessed by confirming that the A_{260}/A_{280} ratio was greater than 1.8 and gel electrophoresis. Equal amounts of total RNA from six tissues were combined to make a pool. Poly(A) RNA was purified from the total RNA pool following Straight A's mRNA Isolation System protocol (Novagen). A total of 8.7 μg of mRNA was eluted and its integrity was judged by gel electrophoresis.

OrientExpression cDNA Synthesis and Cloning System (Novagen) was used for the construction of the T7 phage prostate cancer cDNA libraries. In order to ensure the representation of both N-terminal and C-terminal amino acid sequences and eliminate the 3' bias inherent from oligo(dT)-primed strands, equal amounts of mRNA from each was used to construct two cDNA libraries using directional oligo(dT) primers and random primers in parallel.

After vector ligation and T7 packaging, two cDNA phage display libraries were constructed and the library titers were determined by plaque assay with $4.2 \times 10_6$ pfu for the oligo(dT) primer library and $2.2 \times 10_6$ pfu for the random primer library, respectively. Phage particles from two libraries were combined to make phage library pool. After amplification, glycerol was added and the libraries were stored at -80°C.

Amplification of Libraries. Five milliliters of LB with carbenicillin was inoculated at 37°C overnight with a single colony of BLT5615 from a freshly streaked plate. Overnight culture was added to 100 ml of LB with carbenicillin and grew to an OD_{600} of 0.5. One mM of IPTG was added and the cells were allowed to grow for further 30 min. An appropriate volume of culture was infected with phage library at multiplicity of infection (MOI) of 0.001-0.01 (i.e. 100-1000 cells for each pfu). The infected bacteria were incubated with shaking at 37°C for 1-2 hr until lysis was observed. The phage lysate was then clarified by spinning at 8000 x g for 10 min. The supernatant is collected and stored at -80°C.

Biopanning for Phage-Epitope Clones Specific for Prostate Cancer. To enrich for phageepitopes that bind to IgGs specifically associated with prostate cancer, a positive and negative selection strategy was performed. First, a pre-clearing step was used to remove non-specific epitope-clones by pre-adsorbing the phage libraries onto purified IgG pool from 10 normal sera. Next, the pre-cleared phage libraries were selected onto the pool of IgGs purified from the sera of 19 localized prostate cancer patients. Protein A/G agarose beads (Pierce) were then used to purify IgGs from the sera of prostate cancer patients. Briefly, 10 µl protein-A/G agarose beads were placed into 1.5 ml eppendorf tubes and washed two times with 1X PBS. Washed beads were blocked with 1% BSA at 4°C for 1 hr. The beads were then incubated at 4°C with 15 µl of individual serum from control or prostate cancer patients at 1:50 dilution in 1% BSA. After incubation overnight, the beads were washed with 1X PBS by centrifuging at 1000 g for 2 min. After three washes, 10 µl of 1X PBS was added to each tube, and 10 tubes of protein A/G-IgG complex from 10 control sera and 19 tubes of prostate cancer sera were combined to make IgG pools of control and prostate cancer respectively. These control and prostate cancer IgG pools associated with protein A/G beads were stored at 4°C as stocks for subsequent biopanning.

Twenty microliters of control IgG pool was incubated with 30 µl amplified phage library pool diluted at 1:40 with 10% BSA at 4°C. After 2 hrs, the mixture was centrifuged at 1000 g for 2 min. The beads with non-specifically bound phage particles were discarded, and the supernatant was collected. Next, the supernatant was incubated with 30 µl of the prostate cancer IgG pool at 4°C overnight. The mixture was centrifuged at 1000 g for 2 min and the supernatant was discarded. To elute the bound phage, 100 µl of 1% SDS was added and incubated at room temperature for 10 min to break up the antibody-antigen reaction without disrupting T7 phage particles. The bound phages were removed from the beads by centrifuging at 5500 g for 8 min. Eluted phages were transferred to 10 ml culture of BLT5615 cells for amplification. Five cycles of affinity selections and biopanning were carried out for enrichment of prostate cancer-specific epitope phages.

Construction of the Phage-Epitope Microarrays. The phage library (~10^{10} pfu) from the fifth cycle of biopanning was diluted at 1:10^8 and allowed to grow on LB agar plates with carbenicillin. A total number of 2300 random phage colonies were picked and amplified in 96-well plates. The phage lysates were spotted onto on FAST slides (Schleicher & Schuell) to make high density phage epitope microarrays using a GMS 417 printer (Affymetrix). T7 phage without any cDNA insert and anti-human IgG at 1:1000 dilution were spotted in triplicate as negative and positive controls, respectively. The arrays were dried overnight at room temperature. Before processing, the arrays were rinsed briefly in a 4% nonfat milk/PBS with 0.1% tween-20 to remove unbound phage, and then transferred immediately to 4% nonfat milk/PBS as a blocking solution for 1 hr at room temperature. Without allowing to dry, 2 ml of PBS containing human serum and T7-tag antibody (Novagen) at a dilution of 1:500 and 1:5000 respectively was applied to the surface of the slides in a screw-top slide hybridization tube. To test the specificity of the immune response, reactive serum was first quenched of non-specific activity by pre-adsorbing with 50 fold higher amount (v/v) of bacterial lysate (OD_{600} of 0.5) and then used for incubation as described below. The arrays were incubated with sera from prostate cancer or control individuals for 1 hour at room temperature and then washed 5 times in PBS/0.1% Tween-20 solution for 5 min each. All washes were performed at room temperature.

After washing, the arrays were incubated with 2 ml of PBS containing Cy3-labeled goat antimouse antibody and Cy5-labeled goat anti-human antibody (Jackson ImmunoResearch) at a dilution of 1:5,000 for both for 1 hr in the dark. Five washes were performed using PBS/0.1% Tween-20 solution with 5 mins each. The arrays were dried by centrifuging at 500 g for 5 min and scanned.

Scanning and Primary Analysis of Phage-Epitope Microarrays. All slides were scanned using 532 nm and 635 nm lasers (Axon Laboratories). After scanning, the array images were quantified using GenePix software (Axon Laboratories). According to the experimental design, the median of Cy5/Cy3 was utilized so as to control the small variations in the amount of phage epitope being spotted. Ratio of Cy5/Cy3 for each spot was subtracted by median of Cy5/Cy3 of the negative T7 empty spots with the observation that the signal for the T7 empty phage on each chip highly correlated with the signal intensity for whole array. A Z-transformation was applied to clones so that the mean of each clone was zero across arrays and the standard deviation was 1.

Normalized data was subjected to two-way clustering analysis with use of Cluster and TreeView (Eisen et al., Proc Natl Acad Sci U S A 95, 14863-14868 (1998)). To filter the data, the criteria of at least 1 observation with absolute values greater than 1.2 was applied and 186 clones were selected. An unsupervised hierarchical clustering analysis was performed with correlation (uncentered) similarity matrix and average linkage clustering.

Supervised Analysis of Humoral Immune Response Profiles. In order to efficiently screen hundreds of sera on phage epitope clones, a focused protein microarray comprised of 180 phage clones selected from the primary analysis of high-density epitope microarrays described above was utilized. This focused microarray included four T7 empty phages as negative controls. By employing this small microarray platform, 129 sera included 59 sera from prostate cancer patients obtained prior to prostatectomy and 70 control sera from age-matched males were screened as mentioned above.

The entire dataset from 129 samples was used to build a class prediction model by a leave one out cross-validation (LOOCV) strategy in genetic algorithm/K-nearest neighbors (GA/KNN) ($k=3$ in this study) method (Li et al., 4, 727-739 (2001)). The raw phage-epitope microarray data was normalized as described for the high-density epitope microarrays. The normalized array data was then applied to GA for selection of feature epitopes and assessment of the relative predictive importance of the epitope by ranking them based on their frequency of occurrence in GA solutions. Different numbers of the top-most epitopes were used to build a different KNN prediction model.

Prediction accuracy and error were calculated using LOOCV to evaluate the performance of different KNN model. Finally, a top-ranked 22 clones were selected based on their best performance on specificity and sensitivity. Prediction sensitivity and specificity were computed based on the number of misclassified samples in the cancer and control groups.

Class Prediction on Independent Data. A weighted voting scheme was adopted to predict "test samples", as described previously (Golub et al., Science 286, 531-537 (1999)). Briefly, each epitope in the feature set casts a weighted vote for a class 0 or 1: $V_x = T_x (e_x - b_x)$ where e_x is expression value of epitope x, T_x is the t-statistic for comparing the two class means of epitope x in the training set, and b_x is $(\mu_{class0} + \mu_{class1})/2$. The final vote for class 0 or 1 is sign $(\Sigma_x V_x)$ and the prediction strength (*PS*) or *confidence* in the prediction of the winning class *is* $(V_{win}-V_{lose})/(V_{win}+V_{lose})$, where V_i is the votes for class *i*.

Statistical Analysis. Principal Components Analysis (PCA) (Crescenzi and Giuliani, FEBS Lett 507, 114-118 (2001)) was applied on the epitomic profiles of the 22 phage clones. The first five components contained 90% of the variation in the data set and were subsequently used as covariates in the logistic regression fitting cancer versus normal as binary diagnostic outcome. Fitted probabilities were obtained and used to generate the ROC curve to assess the prediction accuracy of the epitomic profile. All statistical analysis was performed with SPSS 11.1 (SPSS Inc). The mean values for phage epitope humoral response were presented as mean plots with the error bars signifying a 95 % confidence interval of the mean. P values less than 0.05 were considered statistically significant.

Sequence Analysis of Humoral Response Candidates. The top 22 phage epitope clones were amplified by PCR using T7 capsid forward and reverse primers (Novagen). Briefly, 2 µl of fresh phage lysate with titer of ~ 10^{10} pfu was incubated with 100 µl of 10 mM EDTA, pH 8.0 at 60°C for 10 min. After centrifuging at 14,000 X g for 3 min, 2 µl of denatured phage was used for PCR in 100 µl volume of reaction under standard condition. PCR products were confirmed on 1% agarose gel containing ethidium bromide. After purifying with MultiScreen-FB filter plate (Millipore) following manufacturer's protocol, PCR products were sequenced using T7 capsid forward primer to determine the cDNA inserts. DNA sequence and potential protein sequence were aligned using NCBI BLAST.

Development of an ELISA to Validate Humoral Response Candidates. ELISAs were developed for the phage epitopes to confirm their immunoreactivity with different patient serum. Ninety-six well MAX-SORB microtiter plates (NUNC) were coated with 100 µl of diluted T7-tag antibody (Novagen) using 1X PBS at 1:1000 overnight at 4°C on an orbital shaker. All the additions were in 100 µl volumes unless otherwise mentioned. Dilutions of serum and secondary detection reagents were carried out in 1:5 HPE buffer (R&D systems). After washing 5 times with PBS/Tween-20 using EL404 microplate autowasher (Bio-Tek), the plates were blocked first with 200 µl of 2% BSA/PBS for 2 hrs followed by 200 µl of superblock (Pierce) for 2 mins, both at room temperature. Phages and the T7 empty phage as negative control were separately diluted at 1:25 to a final titration of ~10^9 pfu. After washing as above, the plate was incubated with 100 µl of diluted phages for 2 hrs at RT. Serially diluted (1:500, 1:1000 and 1:2000) serum samples were added to each well, and incubated for 1 hr at RT. After washing, the plates were then incubated with 1:10000 diluted HRP-conjugated anti-human IgG for 1 hr at RT. The plates were then developed using 100 µl TMB substrate system (Sigma) for 30 min after final washing. The reaction was stopped using 50 µl of 1.5 M H_2SO_4 and read at 450 nm using ELx 800 universal microplate reader (Bio-Tek).

Meta-Analysis of Gene Expression of Humoral Response Candidates. The gene expression level of four genes, namely BRD2, eIF4G1, RPL13A and RPL22, were studied using ONCOMINE. Briefly, each gene was searched on the database, and the results were filtered by selecting prostate cancer. The data from study class of benign prostate, prostate cancer and / or metastatic prostate cancer with $p<0.05$ were used to plot the box plots with SPSS11.1. P values for each group were calculated using student t-test.

Immunoblot Analysis. Tissues were homogenized in NP-40 lysis buffer containing 50 mmol/L Tris-HCl, pH 7.4, 1% Nonidet P-40 (Sigma) and complete protease inhibitor cocktail (Roche). Fifteen µg of protein extracts were mixed with SDS sample buffer and electrophoresed onto a 4-15% linear gradient SDS-polyacrylamide gel under reducing conditions. The separated proteins were transferred onto polyvinyl difluoride membranes (Amersham). The membranes were then incubated for 1 hour in blocking buffer (Tris-buffered saline with 0.1% Tween (TBS-T) and 5% nonfat dry milk). Membranes were incubated with purified eIF4G1 rabbit polyclonal at 1:4000 dilution (Bethyl), RPL22 mouse monoclonal (BD biosciences) at 1:400 dilution, BRD2 rabbit polyclonal (Abgent) diluted at 1:400 and RPL13a rabbit polyclonal (kind gift of Dr. Paul Fox) used at 1:4000 dilution and incubated overnight at 4°C. After washing three times with TBS-T buffer, the membrane was incubated with horseradish peroxidase-linked donkey anti-rabbit IgG or rabbit anti-mouse IgG HRP conjugate (Amersham) at 1:5000 for 1 hour at room temperature.

After washing the blots with TBS-T and TBS, the signals were visualized with the ECL detection system (Amersham) and autoradiography. To monitor equal loading, the membranes were incubated with anti-human GAPDH antibody (Abcam) at 1:25,000 dilution for two hours and the signals were visualized.

Tissue microarray (TMA) and Immunohistochemistry. In order to determine the expression of eIF4G1 protein *in situ* across a wide range of prostate tissues, a prostate cancer progression TMA composed of benign prostate tissue, localized prostate cancers and metastatic prostate cancer was employed. Antigen retrieval was carried out by heating the slides in citrate buffer pH 6.0 in a microwave oven for 15 minutes. Rabbit anti-eIF4G1 (Bethyl) antibodies were applied (1:100 dilution) and incubated for 1 hour at room temperature. Secondary anti-mouse antibodies avidin-conjugated were applied before washing. Enzymatic reaction was completed using a streptavidin biotin detection kit (Dako).

Immunofluorescence and confocal microscopy. The prostate cancer tissue section slides were soaked in xylene to remove paraffin. Antigen was retrieved by heating the slides in citrate buffer pH6.0 for 15 minutes in a pressure cooker. The slides were then blocked in PBS-T with 5% normal donkey serum for 1 hour. A mixture of rabbit anti-eIF4G1 (Bethyl) antibody and mouse anti-Ecadherin (BD biosciences) antibody was added to the slides at 1:40 and 1:250 dilutions respectively and incubated for 1 hour at room temperature. Slides were then incubated with secondary antibodies (anti-mouse Alexa 488 and anti-rabbit Alexa 555 at 1:1000 dilution) were incubated for 1 hour. After washing the slides with PBS-T and PBS, the slides were mounted using vectashield mounting medium containing DAPI. Confocal images were taken with Ziess LSM510

META (Carl Zeiss) imaging system using ultraviolet, Argon and Helium Neon 1 light source. The triple color images were exported as TIFF images and color balanced.

B. Results

An overview of the method used in the present invention to identify epitomic biomarkers of prostate cancer is described in Figure 1. To develop a T7 phage display library for prostate cancer, RNA was isolated from prostate cancer tissues derived from six patients with clinically localized disease (three patients with Gleason grade 6 and three patients with Gleason grade 7 prostate cancer). To generate a wide range of epitopes (both representing C-terminal and N terminal epitopes), parallel libraries were constructed using oligo(dT) and random primers.

Once packaged into the T7 phage system, epitopes from the library were expressed as a fusion protein with the capsid 10B protein on the surface of the phage. This serves as "bait" to capture potential autoantibodies found in serum. To enrich for epitopes that specifically generate a humoral response in prostate cancer patients, the phage-epitope libraries were subjected to five rounds of biopanning (Fig. 1). In order to remove non-specific immunoreactivity, the phage epitope particles were pre-adsorbed to a pool of immunoglobulins (IgG) isolated from ten control individuals. The "flow-thru" or nonbonding supernatant was then enriched for prostate cancer-specific epitopes by incubating with IgGs from a pool of 19 patients with clinically localized prostate cancer (see Figures 4, 5, and 6 for clinical and pathological information for patients). Protein A/G beads were used to isolate phage-epitope particles that specifically bound antibodies from prostate cancer patients. The bound phages were eluted and amplified in bacteria, thus completing one round of biopanning (Fig. 1). After five rounds of biopanning, it is expected that the pool will be enriched for epitopes that specifically elicit a humoral immune response in prostate cancer patients. Approximately 2300 (2.3K) phage-epitope clones were selected randomly from the biopanned material in order to generate protein microarrays. Once in a microarray format, these enriched phage epitope clones are used to interrogate serum samples for humoral immune response markers.

Using this 2.3K phage-epitope microarray, sera from prostate cancer patients and controls was evaluated. A two-color system was used in which a green fluorescent dye (Cy3) was used to measure levels of the capsid 10B fusion protein as a control for protein spotting, and a red fluorescent (Cy5) was used to measure levels of bound IgG (Fig. 1). Therefore, increased Cy5/Cy3 ratios represented varying levels of immune reactivity. As an initial discovery approach, 31 serum samples consisting of 20 sera from prostate cancer patients and 11 controls were evaluated. Most of the sera from prostate cancer patients exhibited antibody repertoires that reacted with phage-epitope clones on the microarrays while most of the controls did not. After normalization, the data was filtered for elements that have a Cy5/Cy3 ratio with an absolute value greater than 1.2 in at least one of the serum samples. This resulted in 186 phage-epitope clones, which were used for subsequent analyses. Using an unsupervised learning method, Cy5/Cy3 values from these immunoreactive clones were hierarchically clustered. The sera from prostate cancer patients and those from controls segregated into two predominant clusters. Samples in the cluster containing primarily sera from prostate cancer patients, exhibited a robust humoral response to specific phage epitope clones (represented by intensities of yellow color). In this set of 31 sera there was one mis-classified sample from both the prostate cancer cohort as well as the control group. This resulted in a sensitivity and specificity of 95% and 91%, respectively.

To expand the population of sera tested, a focused phage-epitope microarray consisting of the 180 of clones used in the unsupervised analysis (above) as well as additional control elements (i.e., T7 empty phage) was developed. Using these focused protein microarrays, 129 serum samples including 59 patients with biopsy-confirmed prostate cancer and 70 controls were evaluated. Unsupervised analysis using the total 176 epitope clones (excluding four negative clones) revealed 80% specificity and 83% sensitivity for 129 serum samples (see Figure 7). To increase the classification accuracy, a class prediction model was developed by employing a non-parametric pattern recognition approach, Genetic Algorithm (GA) combined with k-Nearest Neighbor (KNN), to discriminate different serum samples. The predictive importance of each epitope for sample classification was evaluated and the epitopes were then ranked with the top-most epitope assigned a rank of 1. Eleven different KNN class prediction models were constructed using different numbers of the top-most epitopes (10, 20-26, 30, 50, and 100 features) to evaluate their predictive performances by leave-one-out-cross-validation. The prediction accuracy improved as more epitopes were involved in the models, whereas too many epitopes introduced excess error in the model thus decreasing the prediction accuracy. The 22 phage epitope clones yielded the best performance in classifying the serum samples with 97% specificity (2 out of 70 controls misclassified) and 88% sensitivity (7 out of 59 prostate cancer patients mis-classified) . Thus, in a substantially larger cohort of sera, it was possible to predict prostate cancer status based on the humoral response to 22 phage epitopes.

The receiver operator characteristics (ROC) of a multiplex panel of humoral response markers was next evaluated to assess prediction accuracy. In order to develop an ROC curve, the 22 predictive phage epitope biomarkers were considered as covariates and the dimension of the dataset from humoral immune response was reduced by principal components analysis (PCA). The first five components accounting for 90% of the variation were applied to logistic regression to predict prostate cancer versus control. The fitted probabilities from the logistic model

($p<0.001$ for the overall model) were used as threshold points to calculate sensitivities and specificities (Fig. 2A). The area under the curve equaled 0.95.

The 22 top discriminating clones identified by supervised analysis were sequenced. Six out of the 22 clones were found to be in-frame and in known expressed sequences. These Six included Bromodomain Containing Protein 2 (BRD2), Eukaryotic Translation Initiation Factor 4 Gamma 1 (eIF4G1), Ribosomal Protein L22 (RPL22), Ribosomal Protein L13A (RPL13A), HES1 (hairy and enhancer of split 1, homolog of Drosophila), and hypothetical protein XP_373908. None of these proteins have been associated with prostate cancer previously as either an over-expressed protein or as a humoral response target. Except hypothetical protein XP_373908, four of the in-frame phage-epitope clones were intracellular proteins involved in regulating transcription or translation in rapidly growing cells. BRD2, also known as RING3, is a nuclear transcription factor kinase known to be up-regulated in human leukemias (Denis and Green, Genes Dev 10, 261-271 (1996); Denis et al., Cell Growth Differ 11, 417-424 (2000)). BRD2 has been shown to specifically interact with acetylated lysine 12 on histone H4 (Kanno et al., Mol Cell 13, 33-43 (2004)). Initiation factors of the eIF4 group are important in the recognition of the 5' cap region of messenger RNAs (mRNA) as well as unwinding of mRNA structure (Gingras et al., Genes Dev 15, 807-826 (2001)). Among them, eIF4G1 plays a central role in the assembly of the preinitiation complex (Morino et al., Mol Cell Biol 20, 468-477 (2000)). eIF4G1 has been shown to be overexpressed in head and neck squamous cell carcinoma (Cromer et al., Oncogene (2003)) and squamous lung carcinoma patients (Bauer, C. et al. Int J Cancer 98, 181-185 (2002); Bauer et al., Cancer 92, 822-829 (2001)) and produces a humoral immune response (Brass et al., Hum Mol Genet 6, 33-39 (1997)). Overexpression of eIF4G1 has been shown to transform NIH3T3 cells (Fukuchi-Shimogori et al., Cancer Res 57, 5041-5044 (1997)). RPL22 and RPL13A are cytoplasmic ribosomal proteins that are the components of the 60S subunit (Mazumder et al., Cell 115, 187-198 (2003)). RPL22 has been shown to be overexpressed in lung cancer (Miura et al., Cancer Res 62, 3244-3250 (2002); Racz et al., Eur J Cancer 35, 641-646 (1999)). RPL13a was identified as a candidate interferon-Gamma Activated Inhibitor of Translation (GAIT) and thus mediates transcript-specific translational control (Mazumder et al., supra). HES1 is basic helix-loop-helix transcription factor of the achaete-scute family. Human achaete-scut homolog 1 (hASH1) is highly expressed in neuroendocrine cancers such as medullary thyroid cancer and small cell lung cancer. HES1 genes encode helix-loop-helix transcription repressors with structural homology to the Drosophila hairy and Enhancer-to-split. HES1 protein is detected at abundant levels in most non-neuroendocrine human lung cancer cell lines.

The remaining 17 prostate cancer specific phage epitope clones were either in un-translated regions of expressed genes or out of frame in the coding sequence of known genes (see Figures 11 and 12)). These clones likely represent "mimotopes" or epitopes that are structurally similar to expressed proteins but unrelated or weakly related at the protein sequence level. Three of the remaining 17 discriminating clones represented an epitope encoded by overlapping sequence from the 5' un-translated region (UTR) of the BMI1 gene (5'-UTR_BMI1), which is a Polycomb Group (PcG) protein implicated in various cellular processes including self-renewal (Park et al., Nature 423, 302-305 (2003); Molofsky et al., Nature 425, 962-967 (2003)). PcG proteins function as multi-component complexes. Protein BLAST analysis of the peptide sequence shared by the three phage clones representing the 5'-UTR_BMI1 identified significant homology (E value = 5 x 10^{-4}) to a glycine-rich stretch of the androgen receptor (Figure 12). Androgens are known to play an important role in prostate cancer progression (Singh and Figg, Cancer Biol Ther 3 (2004); Taplin et al., J Cell Biochem 91, 483-190 (2004)). This was the only phage epitope clone picked up by the methods of the present invention that was represented by multiple independent clones suggesting consistency and robustness of this humoral response in prostate cancer patients (Fig. 2B, C). In 1985, Liao and Witte reported that that 37% of males and only 3% of females had significant autoantibodies to androgen receptor (Liao and Witte, Proc Natl Acad Sci U S A 82, 8345-8348 (1985)). Males older than 66 more often had higher-titer autoantibodies to androgen receptor than younger males or females.

To validate the observations we made using phage-epitope protein microarrays, an ELISA was generated using three of the phage epitope clones including the 5'-UTR_BMI1, eIF4G1 and RPL22. Phage particles were purified and coated onto 96-well plates for subsequent incubation with representative sera from prostate cancer patients and controls. As shown in Figure 2B, prostate cancer patients produce a humoral response to these epitopes relative to controls. Titration of the humoral immune response to the 5'-UTR_BMI1 clone is shown as a representative example in Figure 2C.

In order to validate the 22-clone epitomic profile, an independent cohort of sera from 48 clinically localized prostate cancer patients (pre-prostatectomy), 14 prostate cancer patients (post-prostatectomy), 11 hormone refractory prostate cancer patients, 15 age-matched controls and 10 lung cancer patients was employed. A prediction model was built by a weighted voting algorithm using the 22 phage epitope profile derived from the "training" cohort of 129 samples (Figure 8). As an independent test cohort, a class prediction was made for 63 samples (48 localized prostate cancer and 15 controls) using this model (Figure 9). In total, only 2 out of 15 controls and 8 out of 42 cancers were misclassified, which resulted in 87% specificity and 81% sensitivity. An additional 6 cancer samples were considered as unclassified due to a low prediction strength (confidence) of 0.1 (See Figures 8, 9 and 10). After prostatectomy, the humoral response was generally decreased especially in patients that did not exhibit a recurrence suggesting that the immune response is attenuated upon removal of the "immunogen". 4/4 patients that exhibited PSA recurrence post-prostatectomy, also maintained the 22-epitope humoral response. Only 3 out 11 patients with hormone-refractory disease exhibited a humoral response to the 22 selected epitopes. This suggests that the humoral immune response is attenuated in advanced prostate cancer or those patients treated with anti-androgens and/or chemotherapeutics. To determine if this 22-epitope profile is specific to prostate cancer, sera from 10 lung cancer patients was also examined. Only 2/10 sera from lung cancer patients exhibited reactivity to the prostate cancer epitopes. This is in contrast to the over 80% sensitivity achieved for prostate cancer patients using this platform, suggesting that the epitomic profile is prostate cancer-specific (proportion test, $P<0.001$).

To determine whether the four in-frame phage epitope clones (Fig. 3A) are dysregulated in prostate cancer, a meta-analysis of publicly available prostate cancer gene expression data was performed (LaTulippe et al., Cancer Res 62, 4499-4506 (2002); Luo et al., Mol Carcinog 33, 25-35 (2002); Luo et al., Cancer Res 61, 4683-4688. (2001); Singh et al., Cancer Cell 1, 203-209. (2002); Welsh et al., Cancer Res 61, 5974-5978. (2001); Dhanasekaran et al., supra). This *in silico* analysis suggested there was ample evidence in multiple profiling studies for over-expression of the four in-frame phage epitope clones (Fig. 3B). Immunoblot analyses of benign prostate and prostate cancer tissue extracts demonstrated overexpression of these humoral response candidates at the protein level confirming the *in silico* analyses (Fig. 3C).

To assess the expression of the humoral response candidates *in situ,* immunohistochemistry and immunofluorescence analysis was performed. One out of the four antibodies used for immunoblot analysis (Fig. 3C) were compatible for tissue staining purposes. The antibody that was successful for these applications was directed against the eIF4G1 protein. Weak cytoplasmic staining of eIF4G1 was observed in benign prostate epithelia, and strong staining was observed in clinically localized prostate cancer. These immunohistochemical analyses were further confirmed by immunofluorescence staining for eIF4G1. A strong cytoplasmic staining of eIF4G1 was observed in prostate cancer epithelia as compared to negative staining in benign epithelia.

In summary, the present example describes a robust approach of combining phage display with protein microarrays to detect cancer based on the endogenous humoral immune response. As this approach relies on a multiplex set of markers, it is less likely to suffer from the drawbacks of monitoring single biomarkers such as PSA.

Example 2

Breast Cancer Detection by Epitomic Profiling of the Humoral Immune Response

This Example describes an investigation of the humoral immune signature in breast cancer. The phage display breast cancer cDNA library was purchased commercially from Novagen. The library was enriched for breast cancer specific phage epitopes using a pool of IgG from 10 breast cancer sera and 10 normal controls. A total of 2,304 phage clones were picked and printed on slides to make a high-density phage epitope microarray. By applying this platform, 77 sera samples were screened, including 42 breast cancers and 35 normal controls. The images and data were analyzed and normalized as for prostate cancer (See Example 1). In order to build a predictor, a total of 28 cancers and 24 controls were randomly selected and assigned as training set, and the remaining 14 cancers and 11 controls served as test set. The best performing clones were selected from the training set by t-test with 1000X permutation. A total of 21 clones were selected with 81% specificity (5/24) and 79% sensitivity (6/28). When applying these 21 phage epitopes on independent test set, the same level of accuracy was achieved with 91% specificity (10/11) and 50% sensitivity (7/14).

Example 3

Humoral immune response profiles associated with diagnosis and prognosis in lung adenocarcinomas

A. Construction of phage-epitope protein microarray

The approach described above for profiling of prostate cancer (See Example 1) was used to identify epitomic biomarkers of lung cancer (Figure 13). To develop a phage display library for lung cancer, total RNA was isolated from 7 lung cancer tissues (3 lung adenocarcinomas and 4 squamous). The phage library was then enriched by affinity purification (biopanning) using individual serum samples from 6 adenocarcinomas, 4 squamous and 3 non-cancer controls. Thus, a total of 13 enriched phage libraries were created. After four rounds of biopanning, epitopes that specifically elicit a humoral immune response in lung cancer patients or controls were enriched for. Totally, 2304 phage-epitope clones were selected randomly from the 13 biopanned libraries in order to generate epitope microarrays . Once in a microarray format, these enriched phage epitope clones were used to interrogate serum samples for humoral immune response markers.

Using this high-density phage-epitope microarray platform, sera from 150 lung adenocarcinomas and 101 non-cancer controls were evaluated. As described above (See Example 1), a two-color system was employed in which a green fluorescent dye (Cy3) was used to measure levels of the capsid 10B fusion protein as a control for protein spotting, and a red fluorescent (Cy5) was used to measure levels of bound IgG. Therefore, increased Cy5/Cy3 ratios represented varying levels of immune reactivity. After normalization, data were used for subsequent diagnosis and survival analyses. Results are shown in Tables 1 and 2.

Table 1. Clinical information for Training/Test set samples

	Training set	Test set
Adenocarcinomas (n)	75	75
Age average (year)	63.6	66.3
Age range	44-90	34-88
Male	37	37
Female	38	38
stage I-II	57	59
stage III-IV	18	16
Dead	35	33
Alive	40	42
survival time (ms)	31.5	32.4

No-cancer control (n)	50	51
Age average (year)	60.8	60.8
Age range	36 - 77	40 - 77
Male	30	31
Female	20	20

Table 2. Prediction accuracy of training and test sets

	Training set	Test set
Sensitivity	82.7% (62/75)	82.7% (62/75)
Specificity	94.0% (47/50)	84.3% (43/51)
Accuracy	87.2% (109/125)	83.3% (105/126)

For diagnosis analysis, 251 samples were first randomly assigned to training set (75 tumors and 50 controls) and test set (75 tumors and 51 controls) with matched age, sex, stage and survival (Figure 13 and Table 1). In the training set, t-test combined with leave-one-out-cross-validation (LOOCV) was performed to build a class prediction model, and the top-ranked 59 epitope clones were selected based on their best performance on 82.7 % (62/75) sensitivity and 94.0 % (47/50) specificity (Table 2). Prediction sensitivity and specificity were computed based on the number of misclassified samples in the cancer and control groups. This prediction model consisting of 59 phage-epitopes was then applied to the independent test set. The test samples were correctly classified into cancer and normal groups with 82.7 % (62/75) sensitivity and 84.3 % (43/51) specificity, respectively (Table 2).

In order to investigate the predictive performance of the immune response profile, receiver operator characteristics (ROC) analysis was performed using the 59 phage-epitopes derived from the training set to assess the prediction accuracy in the test set. The discriminative ability of the panel of 59 phage-epitopes between cancers and controls was statistically significant ($p<0.0001$) with an area under the curve (AUC) equal to 0.88 (95% CI = 0.82 to 0.94) (Figure 14).

A leave-one-out cross-validation approach was performed on entire 251 samples (150 tumors and 101 controls) to select the best diagnosis related phage epitopes. The top-ranked 113 clones were found to give the best predict values with 83 % (125/150) sensitivity and 87.1 % (88/101) specificity.

B. Humoral immune response profiles predict survival

The association between phage epitopes and patient survival was next investigated. First, the 150 cancer samples were randomly assigned to a training (n=100) set and test set (n=50) with matched stage and dead/alive. LOOCV with Cox proportional-hazard regression model was used to select the survival related epitopes in the training set. An epitope risk index was created from 7 top-ranked survival related clones based on median cutoff point of the index, which give the best overall survival prediction in the training set (*P* = 0.004, Fig. 15a). The risk index and cutoff point were then applied to the test set. This risk index of the top 7 clones correctly identified low- and high-risk individuals within the independent test set (*P* = 0.02, Fig. 15b).

In order to select the most robust set of survival related clones, the LOOCV approach was used to identify epitopes associated with survival from all 150 tumor samples. A risk index of the top 8 clones can significantly separate 150 patients to high- and low-risk groups (median cutoff point, *P* = 0.0008, Fig. 15c). This risk index can also predict patients with stage I, Ia or Ib cancer (Fig. 15d, e and f). Further analysis with univariate Cox model showed that patient stage, T or N status were also related to survival, but age and sex were not (Table 3). To analyze whether this epitope risk index is an independent factor from other clinical variables, multivate Cox model was performed on age, sex, stage and risk index. The result showed that this epitope risk index is an independent survival predictor (*P* = 0.003, Table 4).

Table 3. Univariate Cox's proportional hazards model

Variable		*P* value
Age		0.96
Sex		0.48
Stage	II	0.02
	III-IV	<0.0001
T status		0.02
N status		<0.0001
Epitope Risk index		0.0008

Table 4 Multivariate Cox's proportional hazards model

Variable	HR	95% CI	*P* value
Age	1.02	0.999 - 1.05	0.06

Sex	1.13	0.693 - 1.85	0.6
Stage II	2.61	1.233 - 5.54	0.01
Stage III-IV	5.89	3.352 - 10.35	<0.00001
Epitomic risk index	2.23	1.328 - 3.76	0.003

C. Identification of phage epitopes

The phage display peptide microarray strategy allows for the easy identification of humoral response targets by sequencing and BLAST searching. The top 400 clones identified by previously LOOCV analysis based on all samples were sequenced (Table 5). Some sequences were found to be in-frame of known protein sequence, such as ubiquilin 1, nucleoar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70). Most of the humoral immuno response peptide targets were mimotopes.

Among the in-frame known proteins, heat shock 70kDa protein was previously reported to be a humoral immune response target in lung cancer by another group. Two different sizes (113-197 and 113-219 CDS region) of HSP70 were found with the same humoral immune response pattern. Three clones of nucleoar protein 3 and alpha-2-glycoprotein 1 were uncovered respectively although the serum antibody to NOL3 was decreased in tumors as compared to no-cancer controls and this humoral immune response was related to an unfavorable survival in lung adenocarcinomas ($P < 0.006$).

A total of 9 clones with 2 different sizes (112 aa and 125aa) of UBQLN1 were found in this study. The mRNA was increased in lung adenocarcinomas (Fig. 16a). Two forms of protein were found by 2D Western blot, of which the native form was increased in tumors as compared to normal lung tissue and the phosphorylated form was decreased in tumors (Fig. 16b and c). A second phosphorylated form of UBQLN1 was found in normal tissue only.

Table 5. Sequence Identity for phage clones associated with diagnosis and prognosis.

Clone ID	Associated with		Translated Protein Sequences	SEQ ID NO	No. of Clones	Protein Identity
	Diagnosis	Prognosis				
12G5	X		PGLIPGFTPGLGALGSTGGSSGTNGSNATPSENTSPTAGTTEPGHQQFIQQMLQALAGVNPQLQNPEVRFQQQLEQLSAMGFLNREANLQALIAT	1	7	Ubiquilin 1

		GGDINAAIERLLGSQPS			
12G9	X	QIQQGLQTLATEAPGLIPGFTPGLGALGSTGGSSGTNGSNATPSENTSPTAGTTEPGHQQFIQQMLQALAGVNPQLQNPEVRFQQQLEQLSAMGFLNREANLQALIATGGDINAAIERLLGSQPS	2	2	Ubiquilin 1
7A2	X	NSLESYAFNMKATVEDEKLQGKINDEDKQKILDKCNEIINWLDKNQTAEKEEFEHQQKELEKVCNPIITKLYQSAGGMPGGMPGGFPGGGAPPSGGASSGPTIEEVD	3	2	Heat shock 70kDa protein 8 (HSPA8)
18D11	X	GAYPSTYDLDIEVHGGLQPCLELEYGAEPIVGIKGSLDSLASEEATMKVESWGSRKHEALYCIQNTEI	4	2	hypothetical protein OB1516
4C10	X	QAFPQQTGRRATSEPTAM	5	2	PREDICTED: similar to Coagulation factor II receptor precursor
2D5	X	VTRPPSGRRPPTS	6	2	PREDICTED: similar to B-cell receptor-associated protein 29
17H12	X	AVAQMRMRMKMRMRMGQEGTQQEPQQQNILEDDTRDQGAHTGGPPGKPDADE	7	2	TPA: HDC18596
19G8	X	QERQTRAQKKGTSSSGHSTTKVIP	8	2	putative protein
4C4	X	GTEIDGRSISLYYTGEKGQNQDYRGGKNSTWSGESKTLVLSNLSYSATEETLQEVFEKATFIKVPQNQNGKSKGYAFIEFASFEDAKEALNSCNKREIEGRAIRLELQGPRGSPNARSQPSKTLFVKGLSEDTTEETLKESFDGSVRARIVTDRETGSSKGFGFVDFNSEEDAKAAKEAMEDGEIDGNKVTLDWAKPKGEGGFGGRGGGQACGRTRVTS	9	1	Nucleolin (NCL protein)

11G4	X	LGTAIGPVGPVTPIGPIGPIVPFTPIGPIGPIGPTGPAAPPGSTGSGGPTGPTVSSAAPSETTSPTSESGPNQQFIQQMVQALAGANAPQLPNPEVRFQQQLEQLNAMGFLNREANLQALIATGGDINAAIERLLGSQPS	10	1	Ubiquilin 2
5B4	X	AERVSETWYMKGTVQHCDFN	11	1	apolipoprotein B
22A10	X	AKHSSAYTFFHPHSNPVSHYHPRFI	12	1	hypothetical protein UM00661.1
7D8	X	ARWGLRMG	13	1	acetyl-CoA acetyltransferase
7G8	X	CCLPRFTESTSV	14	1	similar to ENSANGP00000005259
8D5	X	GELKGKEK	15	1	adenine phosphoribosyltransferase 1, APRT
13D2	X	GKVGGGFLI	16	1	COG0730: Predicted permeases
22F5	X	GPQTDRPPQDRRPRHAPCPQEGCVPLESNAGRPHNLLSDYSCDKSPGRSMTRG	17	1	hypothetical protein MCA0617
17D3	X	GSRGQEFKTSLANMVKLHLY	18	1	PRO0478
1H8	X	HLHNPGDPCRVMSQRPL	19	1	PREDICTED: similar to VPS10 domain receptor protein SORCS 3
18A7	X	HPWAPKGWARWGAAPWAAGWPGTPALSAGTPKLAAALE	20	1	PREDICTED: similar to Zinc finger protein 43
22C1	X	IISRRGTNTAPLTSSSATTRTPARLWCCRS	21	1	hypothetical protein FG05539.1
1E8	X	IKTKENMLREARQKGLVTNGSPSD	22	1	hypothetical protein
6B5	X	IRIAPLEVKFLDRRKTDQSESICQECFH	23	1	solute carrier family 9, member 4
4D1	X	KKKDNL	24	1	COG0628: Predicted permease
4E8	X	KKTSGPDGFTGERYQXI	25	1	ORF2 contains a reverse transcriptase domain
2B6	X	KYWRSIEDRKI	26	1	cytochrome D ubiquinol oxidase subunit II
2G4	X	LELQRQSSL	27	1	spalt4

13F6	X	LEPSFSANYHKDKKTPHVLTHRWELNNENTWTQEEEQHTLGPVL	28	1	PREDICTED: similar to glycogenin 2
13F9	X	LIFRGNGQGMREGNKK	29	1	hypothetical protein AN5619.2
1B8	X	LLLKLEPISQQ	30	1	glycosyl transferase, group 2 family protein
1F4	X	LRQEDCLNPGGRGCSEPRSCHCTPAWATE	31	1	KIAA1556 protein
7E6	X	LRSHAWWWT	32	1	trbI
10G2	X	LSISCL	33	1	hypothetical protein FG08221.1
2C6	X	MVLVNLKP	34	1	heparan Sulfate-glucuronic acid-5-Epimerase (hse-5)
7F9	X	NKTPSVPHNHFSLIK	35	1	PREDICTED: similar to zinc finger protein 300
8B6	X	NSCILKEDKDILKKPLNSRFSSNSKVKNMRLLEHSTFSAPLNRVM	36	1	asparagine-rich protein, putative
7E10	X	NSDFYDFFHK	37	1	Hypothetical protein CBG01255
2D10	X	NSEGRLLS	38	1	Hypothetical protein ZC443.6
2F9	X	NSFDLVGTGGLEESRLSIPWPLGSLLYAKSPRK	39	1	TPA: olfactory receptor OR11-50
3C5	X	NSKESI	40	1	ATP-dependent helicase
3D1	X	NSKNTVLQLDSVRSMSESRAITT	41	1	immunoglobulin heavy chain variable region
2B9	X	NSLPGLPSLYFVSMAKHKNNTSTTIS	42	1	GH05757p
7B7	X	NSPNTLFRSASTKPK	43	1	genral secretion protein E
2C5	X	NSQECLSQILLIPSSCLKKNICV	44	1	ENSANGP00000011065
7E11	X	NSRLRGIL	45	1	COG0330: Membrane protease subunits, stomatin/prohibitin homologs
4B6	X	NSVFLPFINMFIRKWYHSEHISYILFFFCVWIFTLR	46	1	sensor-histidine kinase VanSc
11D1	X	NVTRVFK	47	1	hypothetical protein
7A10	X	PASTLKGQDARNRLTQK	48	1	similar to AF15q14 protein isoform 2
1B12	X	PIHMCYTGAKKEGCFVGKSS-EEVPRTWLLSLKGDG	49	1	CIR protein, putative

		VNSPCWGSY			
13D1	X	PQIASHSLFLLPRVLSTSIIS	50	1	hypothetical protein GZ28G717
5A4	X	PQMTKTKRTHKNI	51	1	FP588
17A5	X	QAYVNV	52	1	COG1538: Outer membrane protein
8B3	X	QEASVSGLKMKSMSTKQVWNQIAFDEKGSGFWRLYFRCCYNASSNQD	53	1	S2 gene product
6A6	X	QTCKQLQFLPFAS	54	1	PREDICTED: hypothetical protein XP537924
7B10	X	RMTYLWGLNHKPTDNVNCHSQFLP	55	1	putative permease (MFS superfamily)
5D5	X	RSQFQQGNVPVQSRLR	56	1	hypothetical protein having cryptosporidium-specific paralog
2B3	X	RVTPTAEQSPIPGCRK	57	1	TonB-dependent receptor
1A8	X	VCSSSIHRSPQVERVSPPHHFPEEQT	58	1	PMF31
3D5	X	VESASLHLDCF	59	1	hypothetical protein BH11560
3B7	X	VGGGRASGRIANGCWA	60	1	AMPA GLutamate Receptor subunit (glr-2)
1A10	X	VPIQMPPEATCVT	61	1	hypothetical protein Bcep02003282
6D2	X	VSNSMKI	62	1	ORFveg109
1F6	X	VVSGSGHLERSQDCGEKGNIFQ	63	1	likely glycerol-3-phosphate dehydrogenase
20A12	X	AHSPTKGCQICQDQEK	64	1	putative retroelement pol polyprotein
20D12	X	AHSRRKTAGN	65	1	recombination activating gene 2
6G7	X	EHIPAPASPRFSIQGS	66	1	PREDICTED: similar to Hypothetical protein 4832420M10, partial
10D10	X	GNRDPVAC	67	1	TPA: 52K
17H8	X	GPWHQMPSPTKGWLGRISQ	68	1	flagellum-specific ATP synthase FliI
15B6	X	IAHSGSSVF	69	1	Niemann-Pick disease, type C1

15A12	X		IQCVYKPNSHFV	70	1	Similar to RIKEN cDNA 4930429O20
19B12	X		IYISLNVVTLKACTLKFGCINATFNLN	71	1	ENSANGP00000025688
23E12	X		LFYGGMGGWKNGSRASEAD	72	1	NIb protein
15E9	X		LLQRNTVPQKQRNKAGWRMTLTS	73	1	PREDICTED: similar to ankyrin repeat-containing SOCS box protein 5
16H8	X		LPSVARRSPGLGPQLRQQGGCGPVCHHHQDIPPPQGLPFPLAPSPFL	74	1	parathymosin-like
8B12	X		NSALGNHGEGKPIVECLLRC	75	1	two-component system, sensor protein
6H3	X		NSASSKCPSY	76	1	hypothetical protein PMM1351
21G10	X		NSFKAIRK	77	1	CDC27 D-618 protein
17H10	X		NSFLEGEEQIL	78	1	hypothetical protein LIC11950
14E12	X		NSSVTLMRQRVTMMGRHTT	79	1	DNA topoisomerase II
21C12	X		PDWDAVVQSWLTAASNS	80	1	ADAM 32 precursor (A disintegrin and metalloprotease domain 32)variant
16C7	X		PRRTGEGAPPARLARRAGEVEHERTC	81	1	PREDICTED: similar to testin
17G10	X		SKLSKGYEKLVF	82	1	putative transcriptional regulator
16E8	X		TMPKGNVKLGN	83	1	mitogen-activated protein kinase kinase kinase 3 isoform 2
8F11	X		VITLIYR	84	1	hypothetical protein OB0069
16H11	X	X	GPEGSEAVQSGTPEEPEPELEAEASKEAEPEPEPEPELEPEAEAEPEPELEPEPDPEPEPDFEERDESEGIPEGQSSDRRCPAHAG	85	3	nucleolar protein 3 (apoptosis repressor with CARD domain)
16E9	X	X	PQCREKTKFN	86	1	tripartite motif-containing 7 isoform 4
16B11		X	SGMPRRYSDYPDAYTT	87	1	cytochrome c oxidase subunit I
16E11		X	DVRVSIHKHILG	88	1	nucleolar protein 3 (apoptosis repressor with

					CARD domain)
8E11	X	GKRRDSFFSF	89	1	hypothetical protein AM638
14E11	X	LETIILSKLAQEQKTKHRMFSLISGS	90	1	putative p150
16G11	X	NSPSVGLFTH	91	1	MUP1
10G9	X	NSRLYQKYKN	92	1	similar to CG9996-PA
5E3	X	PARLARRAGEVEHERTC	93	1	hypothetical protein Magn028940
16F11	X	SLTSTASDGDYSARTVM	94	1	COG0568: DNA-directed RNA polymerase, sigma subunit
10G11	X	TQSPTTLNVAGTPQQ	95	1	IgG kappa light chain variable region
21C5		PSQLKCSPSANVKMGGGKGLKIRENCMHLRT	96	14	glycine decarboxylase
13E11		GERGKRTFQKESDTAL**ILRECPICL**	97	11	BRCA1 protein
11A12		NSLEWTKVYLGKKIWTPEKGNSSYK	98	7	FAM53B protein
13B3		RPQTDRPPQDRRPRHAPCPQEGCVPLESNAGRPHNLLSDYSCDKSPGRSMTRG	99	5	PSIP1 protein
19H9		GQQRKPCLGGKKKT	100	3	CGI-143 protein
22A9		NSTATTSSSSLKDPGSRRPSWTSLAKERSQEQAKRNLEFQSPTLSPPMKATLSKPS	101	3	Oncogene EMS1
16B9		PCSKH	102	3	Siah2 protein
15D6		QERPSETIDRERKRLVETLQADSG...EPDFEERDESEDS	103	3	nucleolar protein 3
13C8		RICPTHTKPQNTVPLHLLRPTIDQL	104	3	FAT tumor suppressor 2 precursor
12E5		WVSEPHCVVVNM	105	3	Kinesin-like protein KIF13B
15E3		GAGTGARARARAGAALTWS	106	2	ALEX2 protein
17A7		ILLMRRRMTRMSGGAEQTQTMQMGVKTK	107	2	CREB-binding protein
17B10		LHHIGQQHPQRFWHQRPIS	108	2	telomerase catalytic subunit

18H6	LMRVLKTEVTGYQEVCTPKRNWNSRQE	109	2	EF hand domain family, member A1
13E12	NSLIQHQHLGQI	110	2	ZFP-95
19G6	NSQGLDFSKATLRSRQRL	111	2	TIP30
18F11	NSSDSLRIVWLLSDVYESFLHLPFQISHCSWYKYLS	112	2	CCAAT/enhancer binding protein alpha
14H12	NSSPADLPCRIC	113	2	UbcH 7-binding protein
21E12	RTPSSPCWPPGPVLAE...EPEPDFEERDESEDS	114	2	nucleolar protein 3 (apoptosis repressor with CARD domain)
13A6	RVPKQRYRSMEQNRALRNNAVYLQLSDL	115	2	tumor-related protein DRC2
10G3	STKKMGTQALSKAAPHC	116	2	kringle-containing protein
15B12	TRSGSSSWAVLTGARPKRLCAATFPNMEKS	117	2	HSPC017
8G11	AEEYRLQRHYCSY	118	1	Pleckstrin and Sec7 domain containing 2
23D12	AESTPVQDPSIFCEYSTPTSMGGGK	119	1	Chain B, Binary Complex Structure Of Human Tau Protein Kinase I
11A2	AEVPILFIPP	120	1	solute carrier family 4 sodium bicarbonate cotransporter-like member 10
17A1	AGGSFSPWPVLLPPPPPGGKSGHNRGQRPH	121	1	frizzled 8
10D4	AHIRTKDSINCI	122	1	TRIM14 isoform alpha
6F3	AICSIL	123	1	
10E5	AIGKIAKNNP	124	1	SFRS protein kinase 2
16E4	ANNLLNGGLYTGKPYCGN	125	1	RAD51D
10C11	ANQLNELNPK	126	1	
9G11	AQGPRCAGCTGKGRTTAG	127	1	
6D4	AQVLCHIEDQVPDQILPGVPLELLGEFCQESGRRK	128	1	
12B5	ARGPSWRSNELWLHHLSSSSRHLMSS	129	1	

1A11	ASCYLTSNCTTRVQ	130	1
1F11	ASRKIWYELNSGYAEWRTEEAIRRSGRHQVQ	131	1
1E7	ATLSV	132	1
4E5	AVYFFKAK	133	1
13H8	AWYKICKICL	134	1
14B9	AYNKFLHL	135	1
21A12	CWPGWSQTPDLR	136	1
7H8	DEWKNTFQGELKGLKC	137	1
14A5	DKKFLIETSI	138	1
7G7	DVFNTVGPLGWSVFHPQTNADQNGVF	139	1
1G7	ECQGQC	140	1
6G6	EEEHSDKYVLSLLMNSLSLRS	141	1
6G2	EFFLMTIGKN	142	1
17G8	EKEKNLNCFFGRTTTKKR	143	1
7A5	EKLATSMYLQNPNWRLSSESEVSME	144	1
9F11	ELESCCVTQAGVPCYDLCSLQPPSPGFK	145	1
12H4	ELLFL	146	1
21B8	EMLNGGRVLWM	147	1
12B3	EQLQT	148	1
4F10	ERKVF	149	1
8B1	ETSIKYT	150	1
17G12	GAGKFLREKEKEISLGLMLGK	151	1
8E5	GCLG	152	1
1F7	GCLGFWGRG	153	1
15C3	GEACLSTATSW	154	1

6C12	GFLTMERKKITPPTTKTYISTLPTDSIKQLRNGDYKATS	155	1
7C9	GGCDHCRDTTHGGCGHCGLRGNPSRPPDLQDCLC	156	1
3A6	GIFFVSKI	157	1
3A1	GIGNVKDGRHGESF	158	1
14A11	GISPTKEDVIHSDVQDELVHSACYVCI	159	1
23F5	GKHEGEG	160	1
3C10	GKIDERGRQGGRERDRNRDRERQRERE	161	1
17B6	GKPKRHWDERAAGGL	162	1
1A1	GKPTPLIQ	163	1
9F9	GKVKELNKEVREKKGKIKQYNTXQKGKKSRRQCKNS	164	1
7E7	GLPLWRRERVKVMR	165	1
5G11	GLWWKRKYLHLNTREKHSQKLLCDDCIHLTELNIPIDRAVWKHSCCGMCKWRFGAL	166	1
24D5	GMST	167	1
21D2	GNYAK	168	1
21D12	GNYARQ	169	1
11H11	GPAFVLMKPGASPYPILALTLITNQMLQNKSNNDPN	170	1
1F9	GPFCHQRSGNPRIHHQHSQAHPWSGLQEACTSGTQRDSEICHEGDGNSRCAH	171	1
8G5	GPTSN	172	1
21H11	GQHYPNTKARQKITTRKL	173	1
10F2	GQRLIIING	174	1
1G3	GRCVVATEINSRNRDSACQEFEFRV	175	1
13G1	GRGRTRWGMGMLLKKIQ	176	1

3E1	GRPGIGATHSSRFIPLK	177	1
19F5	GRVPFTFFNLSL	178	1
2B10	GTSSSHDPLSRLPKLNLSRGGVWASWVK	179	1
3H10	GVERVAYSIHPASPTSVSHSLVERMAMAPPVMESMRSPPQSTRPRVPLS	180	1
17B12	GWGRRIA	181	1
6D6	HCHCLPDLP	182	1
3G10	HILSSTCCFLTF	183	1
7D6	HLWAQHHSVSSLKGRTTLEYF	184	1
17B4	HTFKNTWELKNENTWTQGGEYHTPGPAGGFGGKGRESIRTKI	185	1
8F4	IASYM	186	1
16G2	IDLKSNL	187	1
12G12	IFRN	188	1
4F5	IGTRDQGKRLRMK	189	1
7G1	ILLQGYPGSSSTSLRPHSSN	190	1
16E3	INQKYTWLDKSHYALTTNASS	191	1
4F11	IQNSKKS	192	1
17C8	IQSATELVGRLGMHPRIQSATELVVS	193	1
14B10	IRASNQYRSSVKYISVH	194	1
6A3	ITPRAVFWY	195	1
20D10	IYFKKKKT	196	1
7H2	KDHAQSNKYLTSL	197	1
4E9	KGMNKTSKNCGTM	198	1
15G5	KGTTRSGSLGCK	199	1
2G11	KIYNI	200	1

4D5	KKAERSTK	201	1
1C8	KKEESSSRMWPL	202	1
22C12	KKHFICTSFLDLGYTVPVY	203	1
12D2	KSFCRIFLCW	204	1
20B6	KSTAHSLCKGLM	205	1
11H2	KTTIF	206	1
21E6	LAYVSNSHQGKFGWLSGLSR	207	1
7G11	LDGMLAAQTEEDPET	208	1
15F4	LETEAGESLEPRRWRLQ	209	1
22D3	LEVRISRPSWLTR	210	1
13A12	LHKPQSQWTR	211	1
4A9	LHQNPKGLGSESFWITLPGR	212	1
20C1	LKDVTVSVRLAPLYISM	213	1
14F2	LKHENCLNPGGRGCSESRWCRCTPTRTTE	214	1
10A9	LKQILSSVLNSEIELLL	215	1
9H8	LLHMAAARRSAEQRGKSPS	216	1
7C2	LLPQPPE	217	1
16G1	LLSHLQDWQHH	218	1
12G1	LLSKSLRNEDTAVV	219	1
7B8	LQTGKEKASHPPPTLFSPIIYNNTDLRAVKVILKYYIKWVRRE	220	1
14G11	LQVTLPRRGRDTCGSHREATER	221	1
16G12	LRIT	222	1
23B7	LRLSTPWPTLKPHLKGKVMSL	223	1
16C10	LSESIWFAFHFDDCK	224	1
15F5	LSHGTG	225	1

1C11	LTRNDI	226	1	
11B9	MKEYA	227	1	
11D12	NELWLHHLSSSSRHLMSS	228	1	
10C12	NGCVYLSKFKL	229	1	TBC1 domain family, member 2
17A3	NKEREVFSTNGTGYPHGKKRTTQ	230	1	
15D12	NNQK	231	1	
1E4	NRGKHRG	232	1	
4A5	NSACL	233	1	
1C12	NSAQN	234	1	
8D1	NSASTEPSTNRLQLPWVGGLMQTGRLPGSLTA	235	1	
18D4	NSASTRPISHIRRRTLLSSA	236	1	
11B10	NSDLVRHQFKGKTTLKVH	237	1	
5D4	NSDQIQNTGAESREKVRMSITADEFVG	238	1	
3E4	NSDVI	239	1	
3B8	NSECTCIIVKGNTFSPCKFIV	240	1	
4D2	NSEG	241	1	
13H12	NSEGAT	242	1	
2A7	NSEQQRLKELKSEHTNNKKVKQPCC	243	1	
15D1	NSESNSFASKNKFN	244	1	
21B1	NSFCVCVFNSQS	245	1	
8C2	NSFGFST	246	1	
18C9	NSFLLEIQEPSLGVWIRTPFL	247	1	
10E11	NSFLSF	248	1	
11F3	NSFPSSICFNS	249	1	
1E10	NSFQGLQDYLIKSSMNLVL	250	1	TRHDE

15F11	NSFRKQRHWKG	251	1
6C6	NSFRL	252	1
20E10	NSFRPHRFKSNA	253	1
7C12	NSFRYFA	254	1
11E7	NSGVSW	255	1
9E3	NSHCDI	256	1
4C9	NSHNPKLEK	257	1
7A3	NSIHHVLLSLHPPLYK	258	1
3A2	NSIHM	259	1
22C3	NSIIPRAIWLSVERMWQLRW	260	1
2A6	NSIKCKKM	261	1
12H7	NSIKRFSASCVARICPG	262	1
18D6	NSIL	263	1
17E4	NSILIKYGDTWN	264	1
1G10	NSILQSAGESFLLHNLNLCS	265	1
2G3	NSITHLEKHTILYTNSSTK	266	1
3A4	NSKETSSNGTEWNPH	267	1
17B5	NSKGRRV	268	1
9E4	NSKHR	269	1
21H6	NSKIMFSKMFLSQITE	270	1
19H5	NSKQRFFLKKK	271	1
17C5	NSLCGICI	272	1
7C11	NSLKKL	273	1
19H7	NSLLCLICLT	274	1
10B2	NSLNKIQNTFESSTID	275	1

21B4	NSLPLT	276	1	
10B10	NSLPWKQKV	277	1	Chain A, Structurally Distinct Recognition Motifs In Lymphotoxin-B Receptor And Cd40 For Traf-Mediated Signaling
12D7	NSLS	278	1	
11H12	NSLSFADWFWKRS	279	1	
5H5	NSLSSFHCSSHCF	280	1	
8B2	NSMMDHVTNNATGMNIMEK	281	1	
4G1	NSMSMPRLCGRMKECVPATNAPTSTS	282	1	
13C9	NSMVVTATSYSTPIPEDRLSTRGKEQMPHEMS	283	1	
7E5	NSNEE	284	1	
22E11	NSNPYPGGRSTSGDPKFKPRNCSVPQWLGYNPFWP	285	1	
4F2	NSPAGISRELVDKLAAALE	286	1	
1C6	NSPASAS	287	1	
10B1	NSPKMGSPSLLKYYT	288	1	
9D1	NSPKMGSPSLLKYYT-RS	289	1	
6A2	NSPPAN	290	1	
3D4	NSPSQPACLGAQR	291	1	
5F1	NSPVPSVTTDYQNISLLT	292	1	
10H10	NSQAVCIFF	293	1	
21H10	NSQNVFNSSSFHFMALERYRRK	294	1	
1H5	NSQRLIWLSN	295	1	
14H6	NSQVGLSSSYPQ	296	1	
3D3	NSRCHCPA	297	1	

8A1	NSRFDF	298	1	
11D4	NSSDITLIEKKELIKANI	299	1	TAK1-like protein
2D11	NSSFLMT	300	1	
4E11	NSSFLQGALVPLSGE	301	1	
17D6	NSSGLLKVSLLKYHPS FMNSRGFSLQVL	302	1	
16G8	NSSRQPHPLLTSLNILY I	303	1	
3B10	NSSRTAFSFHSLLLL	304	1	
10G5	NSSSSQHREHEKEEKY	305	1	HGDF-related pro 2
4D7	NSSSSSNPILSHGTTKN KVCSAPEALYAGDGQ LNENLKGKPSGLRCVP LRDFT	306	1	
17A9	NSSSYRPQRVWCGSIC SRASTGIPIPQGLPPKY LAFKELSYLNSAGTSC	307	1	
7F8	NSSV		1	
18C5	NSSVTLMRQRVMMM GRHTT	308	1	dipeptide ABC transporter, dipeptide-binding protein
11H8	NSSWHIRSQGEDNRET ALVYRKQIFSETLHYY KKKK	309	1	
20E7	NSTDK	310	1	
16B6	NSTGNMKGIHLTFQLK RMGKPTPLLF	311	1	
1D4	NSTR	312	1	
19A2	NSTSKSVEHS	313	1	
9A3	NSTVLKYVTLPHLRE	314	1	
5F2	NSVCV	315	1	
10C6	NSVIIESLVVNV	316	1	
1C7	NSVNFILIPLDLEG	317	1	
12C8	NSVQGRAVLLCHGLT GRAWFYLYGLFCV	318	1	
6C2	NSVVH	319	1	

4E4	NSVYMI	320	1	
3F3	NSYCVNQAGLELLASSDPLALASGMLGL	321	1	
1H4	NSYLFSR	322	1	
1D12	PAWATKSKTPS	323	1	
13H6	PGLGEWCRVCV	324	1	
6B10	PGRHLAEAQHGHPRPCLHSEVFS	325	1	
3E6	PHATSHLRVKHEISQIQHPPLLS	326	1	
14F11	PISLRGATAGRAERIREEEVRGAVHHKRH	327	1	
7B1	PQRTTLNFLLGQPARLPLGLSVGDRPTSQGR	328	1	
1B9	PRFPSSAQQRMK	329	1	
5E11	PSRPPRRGGGARAHVLGPERW	330	1	
1A9	QGHTGVSHK	331	1	
1B7	QKTKHRIFSLIGGN	332	1	
2A2	QMLLLPAI	333	1	
3E12	QRSRVAEGWRGPLNPELTPKCIDPSMHGWR	334	1	
20F1	QSLPPARNCNKPDSML	335	1	
1E9	QVPRVLPQHRLGLAGEEAGAPSIPATDHRRLRSGQL	336	1	GADD45 gamma
2E2	QVSGPPSKI	337	1	
2H3	QWLTPVIPTLWEAKAGG	338	1	breast cancer suppressor element Ishmael Upper RP2
2B4	RALQQLRHPDLHLQRRSQAQQHQGGQDS	339	1	
14E10	RAVRREASHRPSPPLASRRPLDALS	340	1	
4D8	RDDSDYSVE	341	1	
18F10	RECTRCRRKTESTAQRVKKPATLLASVKPPANAVSTM	342	1	
1B5	RGPKRLL	343	1	

20G3	RISILKR	344	1	
18E11	RIVRVTPRRSWNHYETIESKE	345	1	
8G6	RLGPQARHG	346	1	
18F2	RLHR	347	1	
1E3	RMKQIVRKVEPIMT	348	1	
19D3	RMMSSSIQSLRKAGSEP	349	1	
2E7	RNWNKPSKRNCP	350	1	
8C11	RPQP	351	1	
14C10	RPQTDLPRTDVPGTLLVLRRAASPWSPTRGDPITCCLITVVISPREGA	352	1	PSIP1 protein
13B11	RPTDRQTSPGQTSPARSLSSGGLRPPGVQRGATP	353	1	
2F6	RQDCF	354	1	
19C12	RRLLGLYMVL	355	1	
6F7	RRRLW	356	1	
14B8	RRSRPSWPTG	357	1	
1D7	RRWTKAHCK	358	1	
10H11	RTLKAEVEKGSM	359	1	
20E12	RVPFTFFNLSL	360	1	
22B12	SFSRG	361	1	
12A11	SLSSTHFDICAGSGGRRSTKCKGLSTSVQCVYEEAH	362	1	
23H10	SNEGLKEVKISTCRLSKQSVSKLLNEKKS	363	1	
15F12	SNSHSPSTQGSLDCVFQETHLIWSDFVSPPKSHLEL	364	1	
6D9	SRRMA	365	1	
12E11	SRSASFMVGTTTVSDRLRTSDFRS	366	1	
2H5	SXARXPIQRESRMGD	367	1	

13D4	TIPGLRTPVSTRPTGTVPIPPIL	368	1	
1G11	TPTRDTSVMQIEETGRGKESSTMVVATTIHHGEATGTISMSSTGTRTTIMGTGDIWMPTVPEAIDPTTCPERGLMTSTSLRPHSSN	369	1	
15H6	TRLAWDLNWKLNVV	370	1	
2A10	TRPPSGRRPPTS	371	1	
7H12	TVLFGV	372	1	
21H4	VAQRPAGPVGWAAGGEALIG	373	1	
1E11	VFEDLKKYLKF	374	1	putative prolyl oligopeptidase
20F12	VFTVVISTSGARCQRQY	375	1	
8C10	VGSWERAGGPPRGEPPPVPAPCLSAPPRCS	376	1	
24H12	VGTIY	377	1	
4E6	VGVGIILS	378	1	
2D6	VHYHNINNLVK	379	1	
21D5	VIGSLMGMALNL	380	1	
16A12	VKKLVVGSWERAGGPPRGEPPPVPAPCLSAPPRCS	381	1	
17D12	VKNYF	382	1	
9G3	VLLYLKR	383	1	
8C3	VPGHARWLTPIIPALRDAEAGGS	384	1	
9D8	VVCSISLLSF	385	1	
2E8	VVFLR	386	1	
14A9	VVQTESLKSPSTYRCAQQDQVTSSSDCHHK	387	1	
3E11	VVVVVETGAI	388	1	
1G1	VYGRNYDGI	389	1	
13A3	WELNSEKTWTQGGEHHTPGPLWGRGARGGIALG	390	1	

16D10	WKKNSRCY	391	1
22H4	WKSGRS	392	1
24F10	WMQSKYSKKSCCYVYG	393	1
11F5	WPPELRLLTDQWQHSILMGM	394	1
20H3	WPPSSGPDCRFTHAIKL	395	1
16B7	WRSSFPSTIYGKD	396	1
19A1	WSGWPT	397	1
11F11	YWTNPPTLTIPRHHLSTVLA	398	1

Example 4

Humoral immune response profiles associated with prognosis in prostate cancer

This example describes the investigation of association of phage epitope clones with prognosis of prostate cancer. The prostate cancer cDNA phage display library described in Example 1 was biopanned using a pool of IgG from 16 prostate cancer sera (7 samples with Gleason = 6 and 9 samples with Gleason = 8 and 9). After construction of phage epitope microarray platform, 32 sera samples were screened. Raw data scanned were normalized as described in Example 1 for prostate caner diagnosis. In order to identify the phage clones for prognosis, the samples were randomly assigned to a training set (31 samples) or a test set (11 samples) with an equal proportion of samples having the same Gleason score. T-test combined with leave-one-out cross validation was applied on the training set. Low risk patients with a Gleason score ≤6 and high risk patients with a Gleason score ≥8 were considered as two groups. A total of 21 clones were selected based on their best performance on the training set with 100% specificity (13/13) and 62.5% sensitivity (5/8). When applying these 21 phage epitopes on an independent test set, its performance was shown to be 100% specificity (4/4) and 75% sensitivity (5/6).

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the relevant fields are intended to be within the scope of the following claims.

CLAIMS

We claim:

1. A method, comprising:

a) providing a phage library, wherein said phage library comprises a plurality of phage clones, each of said phage clones comprising a cDNA obtained from a mRNA sample of a subject with a disease;

b) enriching said phage library for phage clones comprising cDNAs acid specific to said disease, where said enriching comprises binding said phage library to a control IgG to remove non-disease specific phage clones followed by binding said phage library to a disease specific IgG to enrich said phage library for disease specific phage clones, thereby generating an enriched phage library;

c) exposing said enriched phage library to serum from patients having said disease to generate a immunoglobulin bound phage library; and

d) identifying phage clones that react with said serum from said patients having said disease.

2. The method of claim 1, wherein said exposing step further comprises the step of exposing said enriched phage library to serum from control subjects not having said disease.

3. The method of claim 2, further comprising the step of identifying phage clones that react with said serum from said patients having said disease but not with said serum from said control subjects not having said disease.

4. The method of Claim 1, wherein said identifying comprises contacting said immunoglobulin bound phage library with a first immunoglobulin that binds to immunoglobulins from said serum from patients having said disease and a second immunoglobulin that binds to a phage capsid protein.

5. The method of claim 4, wherein said identifying further comprises the step of exposing said first and second immunoglobulins to third and fourth immunoglobulins wherein

said third immunoglobulin binds to said first immunoglobulin and wherein said third immunoglobulin comprises a first label, and wherein said fourth immunoglobulin binds to said second immunoglobulin and wherein said fourth immunoglobulin comprises a second label.

6. The method of claim 5, wherein said first and second labels are fluorescent dyes and wherein said first label emits fluorescence at a different wavelength than said second label.

7. The method of claim 4, further comprising the step of exposing said labeled phage library to an image scanner to identify phage clones that react with said serum from said patients with said disease but not with said serum from said control subjects not having said disease.

8. The method of claim 7, further comprising the step of determining the identify of genes contained in said phage clones that react with said serum from said patients with said disease but not with said serum from said control subjects not having said disease.

9. The method of claim 1, wherein said disease is selected from the group consisting of cancer, autoimmune disease, inflammatory disease, cardiovascular disease, and diabetes.

10. The method of claim 7, wherein said cancer is selected from the group consisting of prostate cancer, breast cancer, and lung cancer.

11. The method of claim 1, wherein said enriched phage library is arrayed on a solid surface.

12. The method of claim 1, wherein said disease specific IgG is purified from the serum of a patient having said disease.

13. The method of claim 1, wherein step b) is repeated 2 or more times.

14. The method of claim 1, wherein step b) is repeated 5 or more times.

15. The method of claim 3, wherein said disease is cancer and said phage clones that react with said serum from said patients with said disease but not with said serum from said control subjects not having said disease are cDNAs encoding tumor antigens.

16. A method for detecting cancer, comprising:

a) providing a sample from a subject suspected of having cancer; and

b) detecting the presence or absence of a humoral response to a tumor antigen selected from the group consisting of BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70), thereby detecting cancer.

17. The method of Claim 16, wherein said cancer is selected from the group consisting of prostate cancer, lung cancer, and breast cancer.

18. The method of Claim 16, wherein said detecting comprises detecting the presence of an autoantibody to said tumor antigen.

19. The method of Claim 16, wherein said method further comprises step c) providing a prognosis to said subject.

20. A kit for detecting the presence of cancer in a subject, comprising:

a) a reagent capable of specifically detecting the presence of a tumor antigen selected from the group consisting of BRD2, eIF4G1, RPL22, RPL13A, HES1, hypothetical protein XP_373908, ubiquilin 1, nucleolar protein 3 (NOL3), alpha-2-glycoprotein 1 and heat shock 70kDa protein 8 (HSPA70); and

b) instructions for using said reagent for detecting the presence of cancer in said subject.

21. The kit of Claim 20, wherein said reagent is an antibody specific for a tumor antigen or an autoantibody to said tumor antigen.

Figure 1

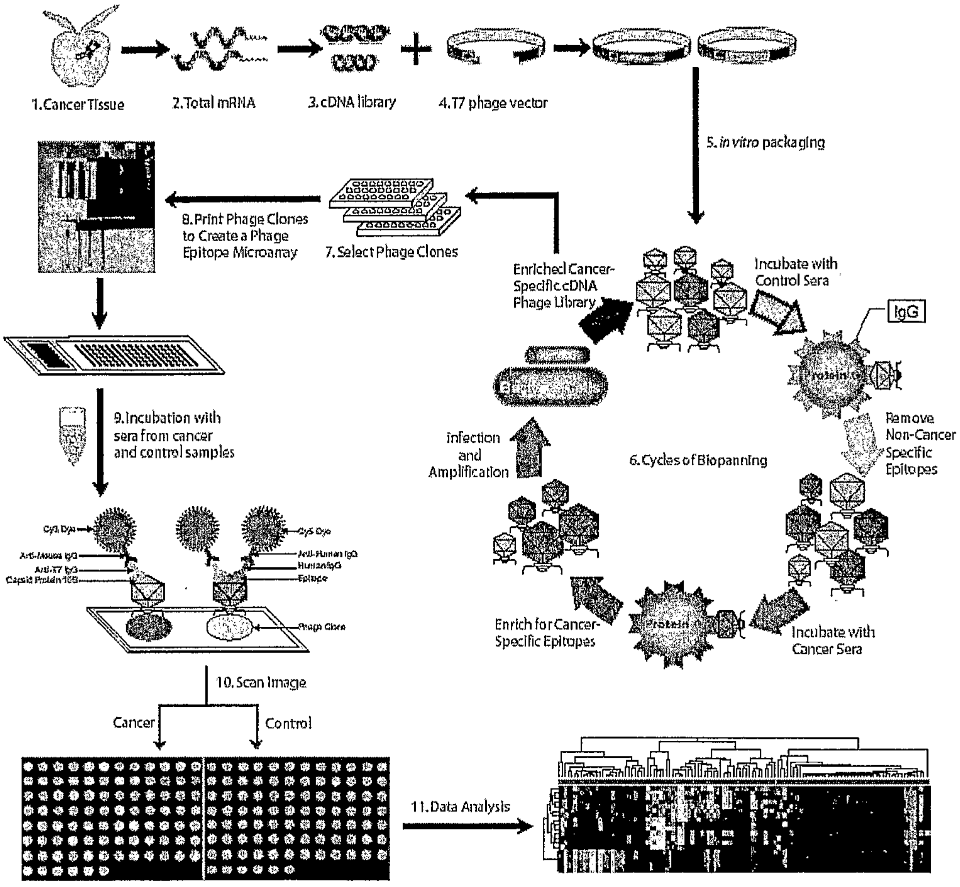
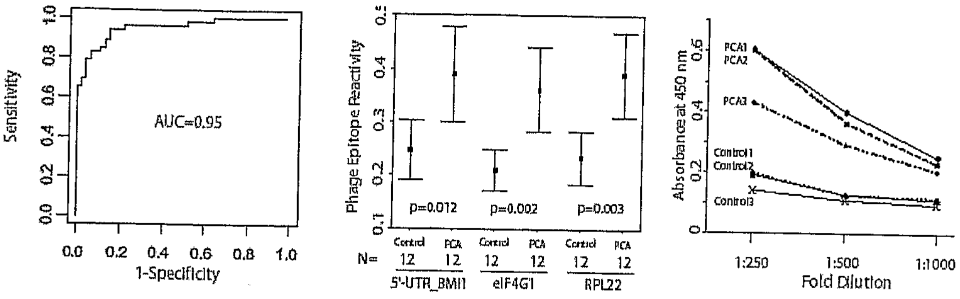

1.Cancer Tissue
2.Total mRNA
3.cDNA library
4.T7 phage vector
5. in vitro packaging
6.Cycles of Biopanning
Incubate with Control Sera
IgG
Protein G
Remove Non-Cancer Specific Epitopes
Incubate with Cancer Sera
Protein G
Enrich for Cancer-Specific Epitopes
Infection and Amplification
Enriched Cancer-Specific cDNA Phage Library
7.Select Phage Clones
8.Print Phage Clones to Create a Phage Epitope Microarray
9.Incubation with sera from cancer and control samples
Cy3 Dye
Anti-Mouse IgG
Anti-T7 IgG
Capsid Protein 10B
Cy5 Dye
Anti-Human IgG
Human IgG
Epitope
Phage Clone
10. Scan Image
Cancer
Control
11.Data Analysis

Figure 2


Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
AUC=0.95
1-Specificity
Phage Epitope Reactivity
0.1
0.2
0.3
0.4
0.5
p=0.012
p=0.002
p=0.003
Control
PCA
N=
12
5'-UTR_BMI1
eIF4G1
RPL22
Absorbance at 450 nm
0
PCA1
PCA2
PCA3
Control1
Control2
Control3
1:250
1:500
1:1000
Fold Dilution

A B C

Figure 3

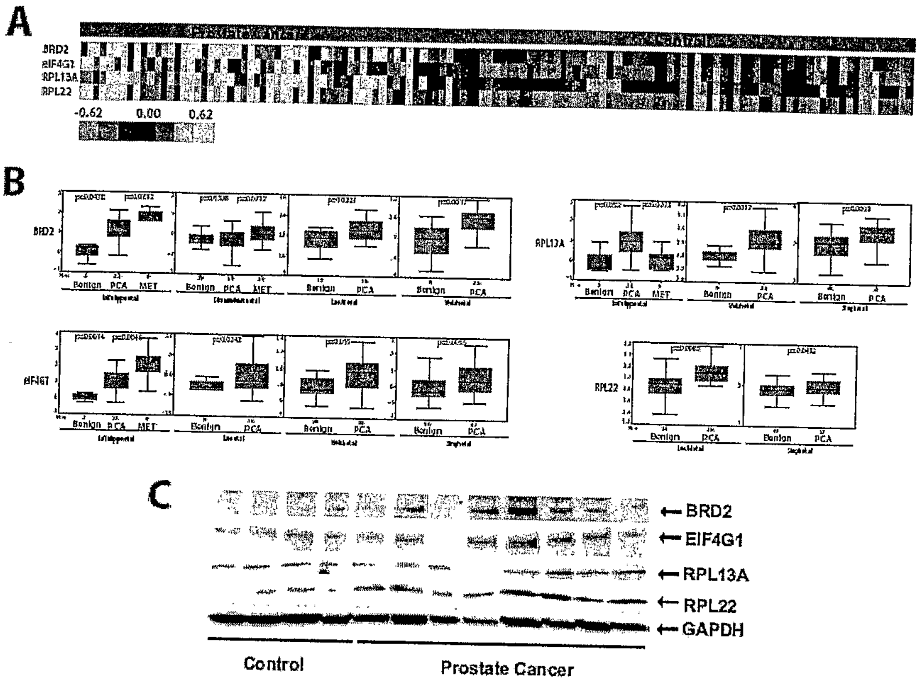

A
BRD2
EIF4G1
RPL13A
RPL22
-0.62
0.00
0.62
B
BRD2
Benign
PCA
MET
Benign
PCA
MET
Benign
PCA
Benign
PCA
RPL13A
Benign
PCA
MET
Benign
PCA
Benign
PCA
EIF4G1
Benign
PCA
MET
Benign
PCA
Benign
PCA
Benign
PCA
RPL22
Benign
PCA
Benign
PCA
C
BRD2
EIF4G1
RPL13A
RPL22
GAPDH
Control
Prostate Cancer

Figure 4

Characteristic	**Value**
Mean age (yr) ± SEM [*]	59.508 ± 1.05
Mean gland weight (g) ± SEM [#]	57.717 ± 8.78
Mean gland size (cm) ± SEM [!]	1.416 ± 0.096
PSA [@]	
Mean (ng/ml) ± SEM	8.276 ± 1.073
0-2.5 ng/ml (%)	8.5
2.6 –10 ng/ml (%)	69.5
4- 10 ng/ml (%)	57.6
> 10 ng/ml (%)	22
Biochemical Recurrence (%)	13.6
Gleason grade [•]	
=< 6 (%)	36.2
>=7 (%)	63.8
Primary tumor identification [†]	
T2a (%)	23.7
T2b (%)	64.4
T3a (%)	3.4
T3b (%)	6.8
T4 (%)	1.7
Ethnicity [‡]	
White (Not Hispanic origin) (%)	83
Hispanic (%)	1.9
Black (Not Hispanic origin) (%)	3.8
Unknown (%)	11.3

[*] Data were available for 59 patients.
[#] Data were available for 59 patients.
[!] Data were available for 58 patients.
[@] Data were available for 59 patients.
[•] Data were available for 58 patients.
[†] Data were available for 59 patients.
[‡] Data were available for 53 patients.

Figure 5

Characteristic	Value
Mean age (yr) ± SEM [#]	59.72 ± 1.06
Mean gland weight (g) ± SEM [#]	57.750 ± 2.49
Mean gland size (cm) ± SEM [1]	1.628 ± 0.145
PSA [#]	
Mean (ng/ml) ± SEM	7.633 ± 1.053
0-2.5 ng/ml (%)	24.6
2.6 –10 ng/ml (%)	62.5
4- 10 ng/ml (%)	43.8
> 10 ng/ml (%)	23
Biochemical Recurrence (%)	35.4
Gleason grade [#]	
=< 6 (%)	52.1
>=7 (%)	47.9
Primary tumor identification [#]	
T2a (%)	22.9
T2b (%)	58.3
T3a (%)	8.3
T3b (%)	10.4
Ethnicity [#]	
White (Not Hispanic origin) (%)	80.9
Hispanic (%)	2.1
Black (Not Hispanic origin) (%)	2.1
Unknown (%)	14.9

[#] Data were available for 62 patients.
[1] Data were available for 61 patients.

Figure 6

Sample Number	Age	RRP	Gleason	Disease Stage at Disease	PSA (ng/ml)	Current Bone Disease	Current Soft Tissue Disease	Hormone Treatment	Chemotherapy
P1	58	Y	5+4=9	T3b	4	N	N	Y	N
P2	76	N	N/A	Metastatic	85	Y	Y	Y	N
P3	76	Y	N/A	T2	16	Y	Y	Y	N
P4	68	N	3+4=7	T1a	15	N	N	N	N
P5	69	N	4+3=7	T2c	239	Y	N	Y	Y
P6	76	N	4+4=8	T2	6	Y	N	Y	Y
P7	44	N	5+4=9	Metastatic	674	Y	Y	Y	N
P8	72	N	1+2=3	Metastatic	7	Y	Y	Y	Y
P9	72	N	4+4=8	Metastatic	412	Y	N	Y	Y
P10	45	N	4+3=7	Metastatic	0.1	Y	N	Y	Y
P11	52	N	4+3=7	Metastatic	183	Y	N	Y	Y

Figure 7

Number of features	Prediction Accuracy (%)	Error
10	93	9/129
20	93	9/129
21	93	9/129
22	93	9/129
23	91	12/129
24	90	13/129
25	90	13/129
26	91	12/129
30	91	12/129
50	88	15/129
100	86	18/129

Figure 8

Sample Name	Call	Confidence	Pathology	Error
TrainSet_175	Normal	0.828	Normal	
TrainSet_116	Normal	0.827	Normal	
TrainSet_178	Normal	0.769	Normal	
TrainSet_168	Normal	0.768	Normal	
TrainSet_111	Normal	0.726	Normal	
TrainSet_182	Normal	0.695	Normal	
TrainSet_190	Normal	0.684	Normal	
TrainSet_187	Normal	0.678	Normal	
TrainSet_115	Normal	0.648	Normal	
TrainSet_208	Normal	0.645	Normal	
TrainSet_219	Normal	0.638	Normal	
TrainSet_114	Normal	0.606	Normal	
TrainSet_203	Normal	0.598	Normal	
TrainSet_179	Normal	0.581	Normal	
TrainSet_189	Normal	0.562	Normal	
TrainSet_117	Normal	0.558	Normal	
TrainSet_184	Normal	0.538	Normal	
TrainSet_214	Normal	0.527	Normal	
TrainSet_164	Normal	0.527	Normal	
TrainSet_174	Normal	0.507	Normal	
TrainSet_108	Normal	0.487	Normal	
TrainSet_109	Normal	0.463	Normal	
TrainSet_112	Normal	0.453	Normal	
TrainSet_113	Normal	0.439	Normal	
TrainSet_107	Normal	0.437	Normal	
TrainSet_166	Normal	0.433	Normal	
TrainSet_198	Normal	0.430	Cancer	*
TrainSet_167	Normal	0.410	Normal	
TrainSet_207	Normal	0.403	Normal	
TrainSet_165	Normal	0.394	Normal	
TrainSet_181	Normal	0.370	Normal	
TrainSet_103	Normal	0.367	Normal	
TrainSet_170	Normal	0.358	Normal	
TrainSet_225	Normal	0.355	Normal	
TrainSet_169	Normal	0.354	Normal	
TrainSet_105	Normal	0.345	Normal	
TrainSet_102	Normal	0.331	Normal	
TrainSet_135	Cancer	0.329	Cancer	
TrainSet_157	Normal	0.326	Cancer	*
TrainSet_177	Normal	0.324	Normal	
TrainSet_173	Normal	0.323	Normal	
TrainSet_212	Cancer	0.236	Normal	*
TrainSet_136	Cancer	0.233	Cancer	
TrainSet_217	Normal	0.233	Normal	
TrainSet_221	Normal	0.231	Normal	
TrainSet_110	Cancer	0.227	Normal	*

TrainSet_106	Cancer	0.226	Cancer	
TrainSet_210	Normal	0.226	Normal	
TrainSet_101	Normal	0.222	Normal	
TrainSet_229	Normal	0.216	Normal	
TrainSet_133	Cancer	0.211	Cancer	
TrainSet_143	Cancer	0.202	Cancer	
TrainSet_171	Normal	0.197	Normal	
TrainSet_140	Cancer	0.195	Cancer	
TrainSet_159	Cancer	0.193	Cancer	
TrainSet_162	Cancer	0.191	Cancer	
TrainSet_131	Cancer	0.180	Cancer	
TrainSet_196	Cancer	0.176	Cancer	
TrainSet_156	Cancer	0.176	Cancer	
TrainSet_160	Cancer	0.168	Cancer	
TrainSet_222	Normal	0.166	Normal	
TrainSet_193	Cancer	0.165	Cancer	
TrainSet_161	Cancer	0.163	Cancer	
TrainSet_122	Cancer	0.156	Cancer	
TrainSet_130	Cancer	0.153	Cancer	
TrainSet_213	Normal	0.151	Normal	
TrainSet_119	Cancer	0.148	Cancer	
TrainSet_146	Cancer	0.146	Cancer	
TrainSet_163	Cancer	0.146	Cancer	
TrainSet_185	Normal	0.142	Normal	
TrainSet_183	Normal	0.140	Normal	
TrainSet_211	Normal	0.138	Normal	
TrainSet_180	Normal	0.136	Normal	
TrainSet_191	Cancer	0.132	Cancer	
TrainSet_188	Normal	0.130	Normal	
TrainSet_199	Cancer	0.121	Cancer	
TrainSet_125	Cancer	0.116	Cancer	
TrainSet_150	Cancer	0.109	Cancer	
TrainSet_223	Cancer	0.108	Cancer	
TrainSet_144	Cancer	0.105	Cancer	
TrainSet_201	Normal	0.102	Normal	
TrainSet_176	Normal	0.097	Normal	
TrainSet_202	Normal	0.097	Normal	
TrainSet_128	Cancer	0.093	Cancer	
TrainSet_141	Cancer	0.089	Cancer	
TrainSet_194	Cancer	0.088	Cancer	
TrainSet_209	Normal	0.087	Normal	
TrainSet_124	Cancer	0.085	Cancer	
TrainSet_126	Cancer	0.075	Cancer	
TrainSet_172	Normal	0.075	Normal	
TrainSet_151	Cancer	0.074	Cancer	
TrainSet_227	Normal	0.074	Normal	
TrainSet_186	Normal	0.069	Normal	
TrainSet_137	Cancer	0.069	Cancer	
TrainSet_138	Cancer	0.069	Cancer	
TrainSet_120	Cancer	0.067	Cancer	
TrainSet_129	Cancer	0.066	Cancer	
TrainSet_200	Cancer	0.063	Cancer	
TrainSet_226	Normal	0.058	Normal	
TrainSet_104	Cancer	0.056	Cancer	
TrainSet_206	Cancer	0.056	Normal	*
TrainSet_152	Cancer	0.052	Cancer	

TrainSet_195	Cancer	0.050	Cancer	
TrainSet_228	Normal	0.045	Normal	
TrainSet_192	Cancer	0.042	Cancer	
TrainSet_149	Cancer	0.041	Cancer	
TrainSet_142	Cancer	0.040	Cancer	
TrainSet_204	Normal	0.038	Normal	
TrainSet_145	Cancer	0.033	Cancer	
TrainSet_158	Cancer	0.027	Cancer	
TrainSet_197	Cancer	0.024	Cancer	
TrainSet_123	Cancer	0.024	Cancer	
TrainSet_153	Normal	0.021	Cancer	*
TrainSet_121	Cancer	0.019	Cancer	
TrainSet_155	Cancer	0.018	Cancer	
TrainSet_205	Cancer	0.016	Normal	*
TrainSet_218	Normal	0.016	Normal	
TrainSet_216	Normal	0.016	Normal	
TrainSet_148	Cancer	0.013	Cancer	
TrainSet_134	Normal	0.010	Cancer	*
TrainSet_220	Normal	0.007	Normal	
TrainSet_118	Cancer	0.003	Cancer	
TrainSet_224	Cancer	0.002	Normal	*
TrainSet_132	Cancer	0.001	Cancer	
TrainSet_127	Cancer	0.000	Cancer	
TrainSet_139	Normal	0.000	Cancer	*
TrainSet_147	Normal	0.000	Cancer	*
TrainSet_154	Cancer	0.000	Cancer	
TrainSet_215	Normal	0.000	Normal	

Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R. and Korsmeyer, S. J., MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet 2002. 30: 41-47.

Figure 9

Sample Name	Call	Confidence	Pathology	Error
TestSet_101	Normal	0.147	Normal	
TestSet_102	Normal	0.239	Normal	
TestSet_103	Normal	0.270	Normal	
TestSet_104	Cancer	0.317	Normal	*
TestSet_105	Normal	0.386	Normal	
TestSet_106	Cancer	0.460	Normal	*
TestSet_107	Normal	0.471	Normal	
TestSet_108	Normal	0.570	Normal	
TestSet_109	Normal	0.577	Normal	
TestSet_110	Normal	0.583	Normal	
TestSet_111	Normal	0.601	Normal	
TestSet_112	Normal	0.635	Normal	
TestSet_113	Normal	0.660	Normal	
TestSet_114	Normal	0.711	Normal	
TestSet_115	Normal	0.748	Normal	
TestSet_116	Normal	0.002	Cancer	*
TestSet_117	Cancer	0.002	Cancer	
TestSet_118	Normal	0.012	Cancer	*
TestSet_119	Cancer	0.047	Cancer	
TestSet_120	Cancer	0.072	Cancer	
TestSet_121	Cancer	0.084	Cancer	
TestSet_122	Cancer	0.137	Cancer	
TestSet_123	Cancer	0.186	Cancer	
TestSet_124	Normal	0.226	Cancer	*
TestSet_125	Normal	0.256	Cancer	*
TestSet_126	Cancer	0.263	Cancer	
TestSet_127	Normal	0.279	Cancer	*
TestSet_128	Cancer	0.297	Cancer	
TestSet_129	Normal	0.343	Cancer	*
TestSet_130	Normal	0.351	Cancer	*
TestSet_131	Cancer	0.382	Cancer	
TestSet_132	Normal	0.386	Cancer	*
TestSet_133	Cancer	0.396	Cancer	
TestSet_134	Cancer	0.474	Cancer	
TestSet_135	Cancer	0.504	Cancer	
TestSet_136	Cancer	0.506	Cancer	
TestSet_137	Cancer	0.543	Cancer	
TestSet_138	Cancer	0.604	Cancer	

TestSet_139	Cancer	0.610	Cancer
TestSet_140	Cancer	0.631	Cancer
TestSet_141	Normal	0.657	Cancer
TestSet_142	Cancer	0.676	Cancer
TestSet_143	Cancer	0.700	Cancer
TestSet_144	Cancer	0.704	Cancer
TestSet_145	Cancer	0.708	Cancer
TestSet_146	Cancer	0.733	Cancer
TestSet_147	Cancer	0.761	Cancer
TestSet_148	Cancer	0.825	Cancer
TestSet_149	Cancer	0.836	Cancer
TestSet_150	Cancer	0.857	Cancer
TestSet_151	Cancer	0.872	Cancer
TestSet_152	Cancer	0.907	Cancer
TestSet_153	Normal	0.917	Cancer
TestSet_154	Cancer	0.921	Cancer
TestSet_155	Cancer	0.930	Cancer
TestSet_156	Cancer	0.963	Cancer
TestSet_157	Cancer	0.984	Cancer
TestSet_158	Cancer	0.988	Cancer
TestSet_159	Cancer	1.000	Cancer
TestSet_160	Cancer	1.000	Cancer
TestSet_161	Cancer	1.000	Cancer
TestSet_162	Cancer	1.000	Cancer
TestSet_163	Cancer	1.000	Cancer

* Golub, T.R. et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-537 (1999).

Figure 10

Sample Name	Call	Confidence	Pathology	Error
Sample_1	Cancer	1.000	Cancer	
Sample_2	Cancer	1.000	Cancer	
Sample_3	Cancer	1.000	Cancer	
Sample_4	Cancer	0.996	Cancer	
Sample_5	Cancer	0.970	Cancer	
Sample_6	Cancer	0.966	Cancer	
Sample_7	Cancer	0.820	Cancer	
Sample_8	Normal	0.615	Cancer	*
Sample_9	Cancer	0.578	Cancer	
Sample_10	Cancer	0.495	Cancer	
Sample_11	Cancer	0.477	Cancer	
Sample_12	Normal	0.319	Cancer	*
Sample_13	Normal	0.314	Cancer	*
Sample_14	Cancer	0.133	Cancer	
Sample_15	Normal	0.065	Cancer	*

Figure 11

Protein Symbol	Protein Sequence
RPL13A	RCEGINISGNFYRNKLKYLAFLRKRMNTNPSRGPYHFRAPSRIFWRTVRGMLPHKTKRGQAALDRLKVFDGIPPPYDKKKADGGSCCPQGRASEAYKKVCLSGAPGSRGWLEVPGSDSHPGGEEEACGRTRVTS
RPL22	SSITVTSEVPFSKRYLKYLTKKYLKKNNLRDWLRVVANSKESYELRYFQINQDEEEEESLRPHSSN
Hypothetical Protein XP_373908	PASASILAGVPMYRNEFTAWYRRMSVVYGIGTWSVLGSLLYYSRTMAKSSVDQKDGSASEVPSELSERPSLRPHSSN
EIF4G1	QTKEERISQXEIMSGARTASTPTPPQTGGGLEPQANGETPQVAVIVRPDDRSQGAIIADRPGLPGPEHSPSESQPSSPSPTPSPSPVLEPGSEPNLAVLSIPGDTMTTIQMSVEEACGRTRVTS
BRD2	SSESRPMSYDEKRQLSLDINKLPGEKLGRVVHIIQAREPSLRDSNPEEIEIDFETLKPSTLRELERYVLSCLRKKPRKPYSTYEMRFISWF

Figure 12

Lab #	Mimotope Candidate	E value	Sequence Alignment(Query, epitope sequence; Sbjct, protein hit)	Protein Sequence
Epitope 1	unnamed protein product	1.00E-18	Query: ILYPETLLKLLISLRRFWAEMMEFSRYTIMSSENRDNLTSS ILYP+ LLKLLISLR F E MEFSRY IMSS N+DN+TSS Sbjct: ILYPQALLKLLIELRSFCIETMEFERYRIMSSANKDNNTSS	ILYPETLLKLLISLRRFWAEMMEFSRYTIMSSENRDNLTSSFPN
Epitope 2	hypothetical protein XP_373740	5.00E-05	Query: MVPRMRRTSRASIHHIKP MVP R R+S+ SIHHIKP Sbjct: MVPCMHRSSQTSIHHIKP	REMVPRMRRTSRASIHHIKPTE
5'-UTR_BMI1	Androgen Receptor *	5.00E-04	Query: GVGGRGGGGGGGRGAGGGRGAGAGGGRPEA G GG GGGGGGG G G+G G G GGG EA Sbjct: GGGGGGGGGGGGGGGGGGGGGGGGGG--EA	GVGGRGGGGGGGRGAGGGRGAGAGGGRPEAA
Epitope 3	Serine/threonine-protein kinase DCAMKL1	0.78	Query: IVKKQKSLCSGFKEHL--NEASIL I+KK K C G KEH+ NE SIL Sbjct: IIKKEK--CRG-KEHNIQNEVSIL	KAECFKNLIVKKQKSLCSGFKEHLNEASILAQVSVSSSKRWKSWENLISSFMVWNPAHLIISIPNLEKTSDLSMMSKLIFLLGSRRFFRSSPRGIF
Epitope 4	hypothetical protein MGC20470	1.8	Query: PLKIPVA----TSRT---QASLGK--QK PLKIP A TSRT ASL K QK Sbjct: PLKIPPAEVTLTSRTTAGAASLTEKIQK	RMPKEPLKIPVATSRTQASLGKQKCRRRIMMSLRQRWQMGISWMGRLKPTQW
Epitope 5	hypothetical protein DKFZp564I1171.1	2.6	Query: CCEKGKKQ CCEKGK+Q Sbjct: CCEKGKRQ	EGSVYQCCEKGKKQVCSQR
Epitope 6	hypothetical protein MGC13275	2.7	Query: QSSVALTNPESYHILKPKLEAD QSSVAL+ P S KP+L+ D Sbjct: QSSVALSTPIS----KPQLDTD	QSSVALTNPESYHILKPKLEADLRWLKLRKRKQVSKLLVLSCCLLKNLGFWKGRMGKTQQRYARLTLWRLWTLQVQPSTLT
Epitope 7	PTPL1-associated RhoGAP 1	3.4	Query: QKLCQAKEKGN +KLCQA E GN Sbjct: EKLCQALEKGN	QKLCQAKEKGMCMKKLRMLWECQKLYSLGF
Epitope 8	Unknown Protein	3.6	Query: PRTRTLRARRSPRMEI-AQK P TRTLR PRME AQK Sbjct: PTTRTLR----PRMEAPAQK	APRTRTLRARRSPRMEIAQKWMMKTVKEEEWNVWMKCPILKNSLPISKINFIKND
Epitope 9	organic cation transporter	3.8	Query: PQWLLPQKGERNT P+WLL QK RNT Sbjct: PRWLLSQK--RNT	PNTFSISSEGNSDVQTNFNKRIKRFIWVHGPQWLLPQKGERNTRVDDF
Epitope 10	BRF1 protein	7.6	Query: PRCLSDATQWP PRCL A+QWP Sbjct: PRCL-HASQWP	PFCKFRILSPRCLSDATQWPFKVLFKWDCSSNSFLGPN
Epitope 11	MOV10-like 1	28	Query: VSALLGWQK V+ LLGW+K Sbjct: VTSLLGWEK	NNVSALLGWQK
Epitope 12	hypothetical protein FLJ40243	28	Query: PENFKI PENFKI Sbjct: PENFKI	QSQHGGPENFKI

Epitope 13	zinc finger protein 292	38	Query: LLTKTIY LLTKT+Y Sbjct: LLTKTVY	LVSILLTKTIY
Epitope 14	HVEC cell-cell adhesion molecule	51	Query: FATEPT FATEPT Sbjct: FATEPT	EFFATEPTPKQHGA

* Besides androgen receptor, this sequence also matches the following proteins with significant E value: autoantigen P542, BAF250b subunit, Brn-1, CAPNS1 protein, FBRL_HUMAN, fibrillarin, FUSE binding protein 3, GAR1 protein, HMG-box transcription factor TCF-3, Homeobox protein SIX3, homeotic protein EVX2, keratin 1, MHC class II regulatory factor, RFX, RDC-1, SHOX2 protein, signal recognition particle 68, SRP68 protein, SRY, SWI/SNF chromatin remodeling complex subunit OSA2.

Figure 13

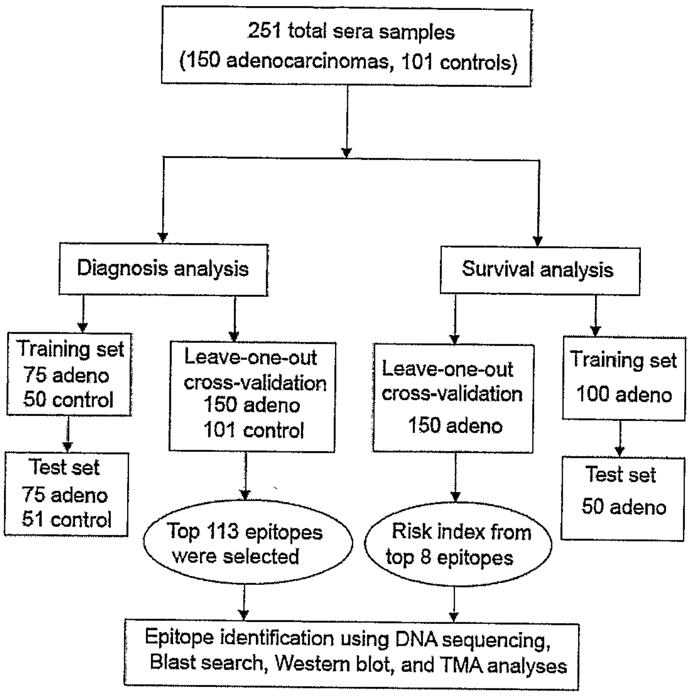

251 total sera samples
(150 adenocarcinomas, 101 controls)
Diagnosis analysis
Survival analysis
Training set
75 adeno
50 control
Test set
75 adeno
51 control
Leave-one-out
cross-validation
150 adeno
101 control
Top 113 epitopes
were selected
Leave-one-out
cross-validation
150 adeno
Risk index from
top 8 epitopes
Training set
100 adeno
Test set
50 adeno
Epitope identification using DNA sequencing,
Blast search, Western blot, and TMA analyses

Figure 14

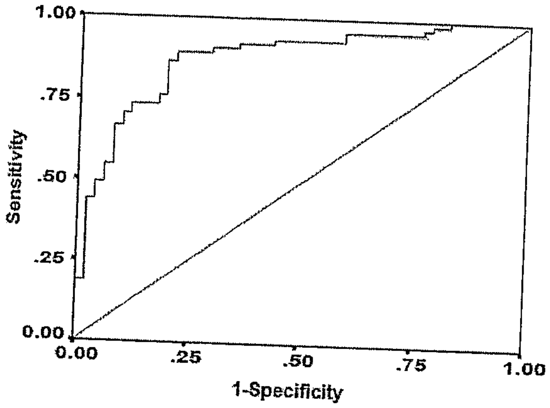


Figure 15

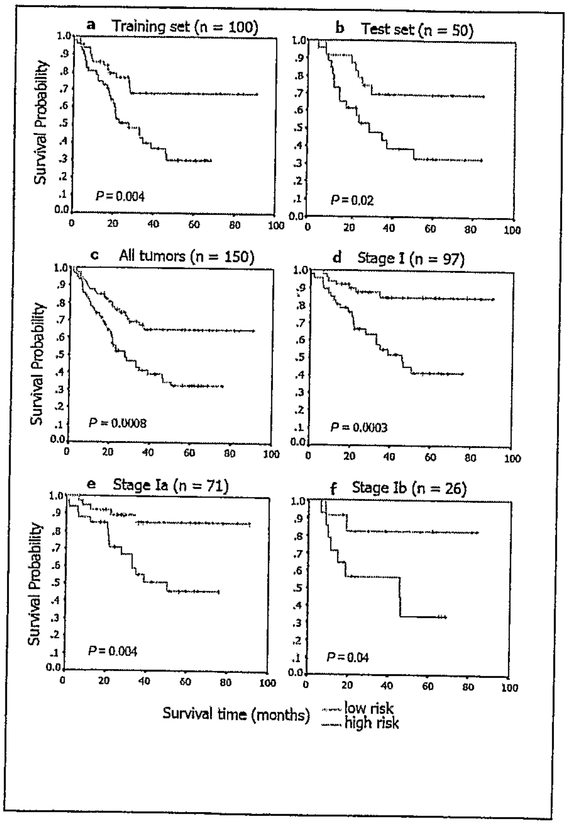
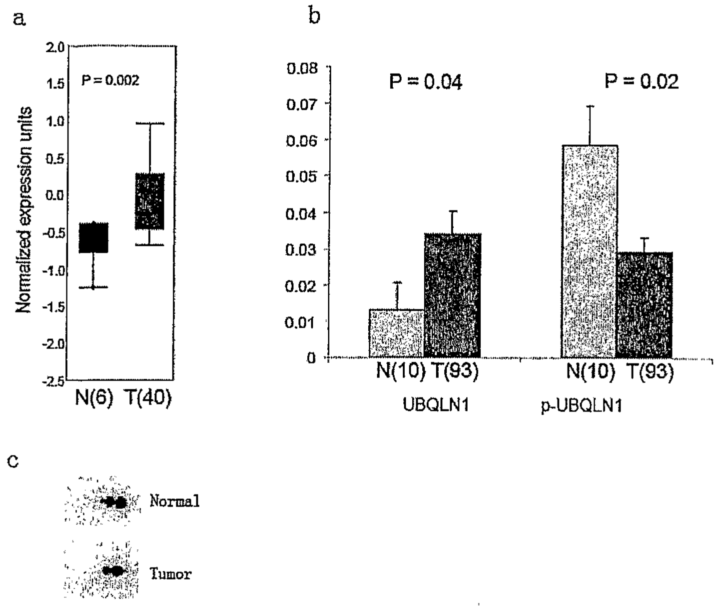

a Training set (n = 100)
P = 0.004
b Test set (n = 50)
P = 0.02
c All tumors (n = 150)
P = 0.0008
d Stage I (n = 97)
P = 0.0003
e Stage Ia (n = 71)
P = 0.004
f Stage Ib (n = 26)
P = 0.04
Survival Probability
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0.0
0
20
40
60
80
100
Survival time (months)
low risk
high risk

Figure 16


a
2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
Normalized expression units
P = 0.002
N(6) T(40)
b
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
P = 0.04
P = 0.02
N(10) T(93)
N(10) T(93)
UBQLN1
p-UBQLN1
c
Normal
Tumor