

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 July 2006 (27.07.2006)

PCT

(10) International Publication Number
WO 2006/078766 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2006/001819

(22) International Filing Date: 19 January 2006 (19.01.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/645,060 19 January 2005 (19.01.2005) US

(71) Applicant (for all designated States except US): **UNIVERSITY OF CONNECTICUT** [US/US]; 263 Farmington Avenue, Farmington, Connecticut 06030-6207 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **CLAFFEY, Kevin** [US/US]; 263 FARMINGTON AVENUE, MC-3501, UCHC, Farmington, Connecticut 06034-3501 (US).

(74) Agent: **HEUSCH, Marina**; EDWARDS ANGELL PALMER & DODGE LLP, P.O. Box 55874, Boston, Massachusetts 02205 (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, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, 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, LV, 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

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR METASTATIC TUMOR SCREENING

(57) Abstract: Methods are described for identifying immunologically reactive lymph nodes in human patients, isolating or preparing variable region recombinant proteins, identifying antigens in tumor and tissue extracts, and identifying activated germinal centers in tumor-draining and pathologically reactive-lymph nodes. The antigens and proteins may be employed as diagnostics and therapeutics for cancers and other pathological conditions. Also described are methods for screening biological samples for early tumor metastases. Stromal cell derived receptor-1/Neuroplastin gp65/55 is described as a marker of invasive potential in human tumors.



WO 2006/078766 A2

5 **TITLE OF THE INVENTION**

METHODS AND COMPOSITIONS FORMETASTATIC TUMOR SCREENING

RELATED APPLICATIONS & INCORPORATIONS BY REFERENCE

This application claims priority to U.S. Provisional Application Serial No. 60/645,060, filed January 19, 2005, the contents of which are incorporated herein by reference.

10

Each of the applications and patents cited in this text, as well as each document or reference cited in each of the applications and patents (including during the prosecution of each issued patent; "application cited documents"), and each of the PCT and foreign applications or patents corresponding to and/or claiming priority from any of these applications and patents, and each of the documents cited or referenced in each of the application cited documents, are hereby expressly incorporated herein by reference. More generally, documents or references are cited in this text, either in a Reference List before the claims, or in the text itself; and, each of these documents or references ("herein-cited references"), as well as each document or reference cited in each of the herein-cited references (including any manufacturer's specifications, instructions, etc.), is hereby expressly incorporated herein by reference.

15
20

BACKGROUND OF THE INVENTION

It is generally understood that human immune defense against cancer is functionally suppressed at many levels, resulting in little active repression of aggressive tumors.

25 However, many of the initial activation and selective immune reactions do occur in and around tumors, particularly within tumor-draining lymph nodes. One of these events is immune cell recognition of uncommon or aberrant proteins derived from the tumor. In these instances, the immune cells respond by producing antibodies that recognize these proteins. Importantly, the immune system has identified these proteins as non-self or foreign, and, thus, they constitute biologically determined tumor antigens.

30

Common surgical procedures currently require the identification of local tumor-draining lymph nodes using a method called sentinel lymph node mapping. These lymph nodes are removed with the tumor at the time of surgery and used to determine if tumor cells have spread or metastasized from the primary lesion. The presence of tumor cells in the tumor-draining lymph nodes indicates a higher tumor grade, and aggressive systemic chemotherapy is, consequently, instituted to block metastatic tumor growth.

35

5 It is difficult to derive antibodies from human B-cells. Current methods require either large-scale gene insertion into bacterial phage libraries and selection based upon non-specific phage binding to tumor preparations, or the isolation of B-cells that recognize "known" antigen targets such as the DNA in lupus auto-immune patients.

There are no current methods that detail the selection of immune cells and synthesis of
10 antibodies in a manner sufficient to isolate and identify novel tumor antigens.

There is, furthermore, a particular need for diagnostic and therapeutic reagents applicable in a number of clinical venues. The current methods for screening patients for breast cancer are restricted mainly to diagnostic mammograms. Breast tumor samples are characterized for several major identifiers related to therapeutic options; 1) tumor size and grade, 2) local
15 metastasis, 3) estrogen receptor status, and 4) gene amplification of the HER-2/neu kinase receptor.

Large tumors can be detected with advanced CAT or MRI scanning techniques that are typically employed when patients already present with symptoms of disseminating disease. There are, however, no current methods for detecting small microscopic tumors earlier on
20 that have spread to other tissues, thus highlighting the need for systemic tumor antigen biomarker identification. There are, for example, no available biomarkers employed for breast cancer similar to the Prostate Specific Antigen for male prostate cancers.

In addition, there are no methods to determine when women may have micro-metastases that have spread from the breast to other parts of the body, which ultimately results in
25 terminal disease. In addition, when patients present with symptomatic disseminating breast cancer, there are few effective treatment options, and, certainly, there are none that are specific to individual patients.

In summary, with the exception of mammograms, there are no effective clinical diagnostic assays for breast cancer that can be used for common samples, either serum or urine. There
30 are no diagnostic markers to determine whether therapies are effective in reducing tumor burden in metastatic disease. There is only one example of a selective antibody-based therapeutic based upon initial diagnostics, Herceptin (a targeted antibody to HER-2/neu receptor kinase). There are no current methods to screen post-surgical patients for development of microscopic cancers that have spread and/or are expanding in disseminated
35 tissues prior to symptom development.

5 SUMMARY OF THE INVENTION

In one embodiment, the invention provides a method for identifying a reactive B-cell germinal center in a lymph node sample, comprising detecting three histological markers, CD20, CD23, and Ki67 at a location in said lymph node sample and comparing the detected signals wherein a positive signal for all three markers (i.e., CD20+, CD23+, Ki67+) at said
10 location is indicative of a reactive B-cell germinal center, thereby identifying a reactive B-cell germinal center in a lymph node sample. The method may further comprise excising a sample of lymph node and freezing said sample. In other embodiments of the method of the invention, the lymph node is tumor-draining or pathologically inflamed.

15 In another embodiment, the invention provides a method for identifying immune cytokine profiles in a germinal center microenvironment, comprising:

- i) obtaining a tissue sample of tumor-draining sentinel or distal lymph node;
- ii) isolating mRNA from the tissue sample from a reactive germinal
20 center and a non-reactive germinal center;
- iii) determining the levels of TH1 and TH2 cytokines from mRNA of ii);
- iv) comparing the levels of the cytokine mRNAs from the reactive germinal center and the non-reactive germinal center as determined in step (iii) where
25 higher TH2 cytokines in the germinal center indicates an active immune reaction,

thus identifying immune cytokine profiles in a germinal center microenvironment. In a further embodiment, the tumor is associated with breast cancer. A non-reactive germinal center is defined as CD20+ (CD23-, Ki67-). Other lymphatic organs including the spleen are likewise contemplated as sources for the tissue samples or extracts employed in
30 various aspects of the present invention.

In another embodiment, the invention provides a method for identifying antibody variable regions with somatic mutations in a lymph node, comprising:

- i) obtaining a tissue sample from a lymph node;
- ii) preparing VH, V λ and V κ germinal center antibody variable region
35 cDNA libraries;
- iii) compiling CDR/FR ratios from sequenced variable region cDNAs from the libraries; and
- iv) selecting variable region sequences with CDR/FR ratios greater than about 0.8 in the libraries,

5 thus identifying antibody variable regions with somatic mutations in a lymph node. A presence of antibody variable regions with somatic mutations may be indicative of immune reaction in the lymph node. In a further embodiment of the method of the invention, the lymph node is a tumor-draining lymph node.

10 In another embodiment, the invention provides an isolated tumor antigen identified by the steps comprising:

- i) expressing an antibody variable region as a purified, variable region recombinant protein fused to at least one epitope tag;
- ii) contacting the variable region recombinant protein with a tumor extract;
- 15 iii) determining binding of a tumor antigen with the variable region recombinant protein; and
- iv) identifying said tumor antigen. The antibody variable region may, for example, be identified as described above. The term "extract" includes the total or a subfraction of the proteins derived from the tumor (or other tissue). An epitope tag
- 20 includes, for example, without limitation, c-myc, V5, or 6xHis tag. In yet another embodiment, the antibody variable region is expressed in a bacterial expression system.

25 In another embodiment, the invention provides a variable region recombinant protein fused to at least one epitope tag prepared by the steps comprising expressing the antibody variable region identified in claim 6 as a purified, variable region recombinant protein fused to at least one epitope tag and isolating the protein.

 In another embodiment, the invention provides an isolated nucleic acid molecule selected from the group consisting of:

- i) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% homologous to the nucleotide sequence of SEQ ID NO:1, or a
- 30 complement thereof;
- ii) a nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 50% homologous to the amino acid sequence of SEQ ID NO:2;
- iii) a nucleic acid molecule which encodes a fragment of a polypeptide
- 35 comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2; and
- iv) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2,

5 wherein the nucleic acid molecule hybridizes to a complement of a nucleic acid molecule comprising SEQ ID NO:1 under stringent conditions. In a preferred embodiment, the isolated nucleic acid molecule is selected from the group consisting of:

i) a nucleic acid molecule comprising the nucleotide sequence of SEQ
10 ID NO:1 or a complement thereof; and

ii) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2. In a further embodiment, the invention provides an expression vector comprising the nucleic acid molecule. Yet another embodiment of the invention provides a cell transformed with said expression vector. The term "transformed"
15 includes transfected, transduced, infected, microinjected, and the like.

In another embodiment, the invention provides an isolated polypeptide selected from the group consisting of:

i) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acids
20 of SEQ ID NO:2;

ii) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a complement of a nucleic acid molecule comprising SEQ ID NO:1 under stringent conditions;

25 iii) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1; and

iv) a polypeptide comprising an amino acid sequence which is at least 30% homologous to the amino acid sequence of SEQ ID NO:2. In a preferred embodiment,
30 the isolated polypeptide comprises the amino acid sequence of SEQ ID NO:2. In a further embodiment, the invention provides an antibody or conjugate thereof which selectively binds to the isolated polypeptide. In a preferred embodiment, the antibody is monoclonal. The term "conjugate" means that a molecule or molecular complex is linked to the antibody, including, without limitation, an imaging agent, a radioactive molecule, a toxin, or a
35 molecule that modifies the antibody's binding efficacy or activity.

In another embodiment, the invention provides a composition comprising the variable region recombinant protein described above.

5 In yet another embodiment, the invention provides a pharmaceutical composition comprising the antibody described above and a pharmaceutically acceptable excipient.

 In still another embodiment, the invention provides a fusion protein comprising the antibody described above.

 In another embodiment, the invention provides a use of the variable region
10 recombinant protein, antibody, or fusion protein described above for detecting and/or quantifying tumor-related antigens present in biological samples.

 In yet another embodiment, the invention provides a use of the variable region recombinant protein, antibody, or fusion protein described above for treating cancer, wherein the recombinant protein, antibody, or fusion protein targets tumor-specific antigens.

15 In an additional embodiment, the cancer is primary or disseminated.

 In another embodiment, the invention provides a method of identifying a tumor cell with a predominance for lobular carcinomas or ductal carcinomas in a subject, comprising:

- i) obtaining a tumor extract or histological section of a tumor;
- ii) contacting the extract or section with the monoclonal antibody
20 described above or the variable region recombinant protein described above or the fusion protein described above; and
- iii) measuring binding of said antibody with an antigen in said extract or section;

 wherein the binding identifies a tumor cell with a predominance for lobular
25 carcinomas or ductal carcinomas in the subject. In a further embodiment, the carcinoma is invasive or highly invasive. In still a further embodiment, the subject has breast cancer or some other invasive or highly invasive cancer.

 In another embodiment, the invention provides a method for identifying an antigen associated with a pathological condition in a subject, comprising:

- 30 i) obtaining, from the subject, a tissue sample from a B-cell germinal center positive for histological markers CD20, CD23, and Ki67;
- ii) preparing VH, V λ and V κ germinal center antibody variable region cDNA libraries;
- iii) compiling CDR/FR ratios from sequenced variable region cDNAs
35 from the libraries;
- iv) selecting variable region sequences with CDR/FR ratios greater than about 0.8 in the libraries to identify antibody variable regions with somatic mutations;
- v) expressing the antibody variable region as a purified, variable region recombinant protein fused to at least one epitope tag;

- 5 vi) contacting the variable region recombinant protein with a tissue sample from the subject; and
- vii) determining binding of an antigen with the variable region recombinant protein; thereby identifying an antigen associated with a pathological condition. The pathological condition may be selected, without limitation, from the group
- 10 consisting of: chronic inflammatory disease, autoimmunity, toxin exposure, and bacterial, virus, or parasitic infection.

In another embodiment, the invention provides a kit comprising the composition comprising the variable region recombinant protein described above and instructions for administering the composition to identify a specific pathological event in a subject.

- 15 In yet another embodiment, the invention provides a kit comprising the pharmaceutical composition described above and instructions for administering the composition to treat cancer in a subject.

In one embodiment, the present invention is directed to a method for isolating antigen-recognizing antibodies from immunologically reactive lymph nodes in pathological

20 conditions. In particular embodiments of the invention, the antibodies can be used for diagnostic, biological sample screening, and therapeutic treatment of such condition. Lymph nodes become reactive to pathologic antigens through antigen-presenting cells, T-cell activation, and activation of B-cells to form reactive germinal centers in lymph nodes and initiate antigen-dependent antibody production.

- 25 According to one embodiment of the invention, surgical or post-mortem-obtained lymph nodes are used to evaluate germinal center B-cell activation markers and proliferation status. In a particular embodiment of the invention, molecular cloning of antibody variable domains from these reactive germinal centers allow identification of stimulatory antigens by using recombinant antibody variable domain proteins to trap and identify the antigens
- 30 present in the primary pathologic condition, *e.g.*, cancer, chronic inflammatory diseases, autoimmunity, toxin exposure, bacterial, virus and parasitic infections. In additional embodiments, the invention provides methods of lymph node analysis, genetic recovery and bioinformatic analysis of antibody sequences, recombinant protein synthesis and isolation, details as to antigen trapping with recombinant antibodies, and identification and
- 35 verification by mass spectrometry analysis.

In one particular embodiment, the invention is directed to a novel tumor antigen known as stromal cell derived factor receptor-1 (SDR1) and neuropilin gp65/55.

- 5 Other aspects of the invention are described in or are obvious from the following disclosure and are within the ambit of the invention.

BRIEF DESCRIPTION OF THE FIGURES

The following Detailed Description, given by way of example, but not intended to limit the invention to specific embodiments described, may be understood in conjunction with the
10 accompanying drawings, incorporated herein by reference. Various preferred features and embodiments of the present invention will now be described by way of non-limiting example and with reference to the accompanying drawings in which:

FIG. 1A shows the histological phenotype, germinal center hyperplastic phenotype, of
sentinel lymph nodes from metastatic node negative patients. Box indicates an isolated
15 germinal center structure readily apparent in H&E stain.

FIG. 1B. shows the histological phenotype, sinus histocytic phenotype with central sinusoidal expansion and peripheral small GC clusters in a sentinel lymph node from a metastatic node negative patient.

FIG. 2A shows differential immunohistological staining of germinal centers; utilizing CD20
20 staining as a B-cell marker.

FIG. 2B shows differential immunohistological staining of germinal centers, with proliferative GC zones indicated via Ki-67 marker. Thin arrows denote proliferative GCs, and thick arrows non-proliferative GCs.

FIG. 3 depicts the overlap of B-cell germinal center markers in serial sections from breast
25 cancer sentinel lymph nodes. The histological profile of CD20, CD23 and Ki-67 in each section is registered as a merged image.

FIG. 4 depicts the results of a qualitative analysis of sentinel lymph node (SLN) mRNA expression for TH1- and TH2-specific cytokines. The libraries analyzed constituted "Total": non-specific core section composed of both germinal center and T-cell zone, Germinal
30 Center A (GC-A): CD20⁺/CD23⁺/Ki67⁺, and Germinal Center (GC-B):CD20⁺/CD23⁺/Ki67⁻.

FIG. 5 shows a histological profile of overlapping areas defined by histological staining and CD20 marker and the specific germinal center area cored from the sample.

- 5 FIG. 6A depicts the amplification of VH, V λ and V κ chains from germinal center cores, wherein primary amplification was carried out with degenerate human PCR primers using VENT polymerase.

FIG. 6B depicts total VH amplification from a single germinal center core, prior to TOPO cloning.

- 10 FIG. 6C provides a comparison of PCR-amplified Ki-67⁺ vs. Ki-67⁻ VH fragments as a quality control screen prior to di-deoxy sequencing.

FIG. 7 shows, in graph form, that VH CDR/FR ratios indicate a higher degree of antigen-driven somatic hypermutation in Ki67⁺ sentinel lymph node. Completely sequenced VH gene fragments from each of the three samples (Library A (sentinel lymph node – Ki67⁺), Library B (sentinel lymph node – Ki67⁻), and Library C (distal lymph node – Ki67⁺) were split into two groups: those with CDR/FR ratios below 0.8 and those with ratios of 0.8 and above.

- FIG. 8 shows V λ and V κ germline usage for sentinel and distal lymph node germinal-center Libraries A-C. The frequency of each light-chain germline gene is depicted in respective charts. The total numbers of sequenced light-chain genes and quantitation for the four most common germline sequences are listed for each library.

FIG. 9 provides a graphic display of phylogenetic associations of VH clones obtained from Ki-67-positive versus -negative germinal centers.

- FIG. 10 provides a comparison of closely related VH sequences that distinguish high somatic hypermutational (SHM) events, as well as CDR-3 alterations. The SHM observed in the C08 clone is statistically significant (using the method of Lossos (56), wherein $p < 0.05$ are significant). CDR/FR ratios exceeding 0.3 also constitute SHM. Asterisks indicate amino acid differences. R:S corresponds to replacement vs. substitution ratio. FIG. 10 additionally shows SEQ ID NO: 28 ("C08") and SEQ ID NO: 29 ("E06").

- 30 FIG. 11A schematically depicts a single VH domain (VH AB-1) fused to C-terminal c-myc and 6x-his tags. Somatic hypermutation of this clone is indicated by a 1.2 CDR/FR ratio, as statistically confirmed was carried out by the method of Lossos (56) ($p < 0.05$).

FIG. 11B shows the purification path of bacterially expressed recombinant VH Ab-1 as detected by anti-c-myc antibody immunoblots. The lanes correspond to: UI-uninduced

5 extract, IP-induced insoluble pellet, IS-induced soluble supernatant, FT-Talon flow through, and E1-E3-successive elutions off Talon column with 50, 100 and 200 mM imidazole, respectively.

FIG. 12A schematically depicts the microscale "antigen-trap" assay. The microplate well-based method binds VH to Ni covalently attached to the plate. Non-specific binding is
10 blocked by BSA. Antigens bound from tumor extracts were washed and eluted through direct trypsinization.

FIG. 12 B shows trypsinized LC and MS profiles (LC-MS/MS elution profiles) of VH-bound peptides.

FIG. 12C shows the identification of SDR-1/NP, specifically, one of five repeat matches
15 (red font) of peptide for SDR-1/NP. FIG. 12C additionally shows SEQ ID NO: 30.

FIG. 13 shows the results of direct ELISA assay using recombinant VH proteins, or no antibody controls to detect surface-bound tumor extracts. Four VH proteins, as well as no ab, and the fluorescent reagent only were used in a direct ELISA against 50µg of total tumor extracts as a Triton X-100 soluble fraction from four patients A-D. The signal indicated
20 constitutes an average of duplicates, all with error less than 8%. Asterisks denote detection above background determined and confirmed with repeat experiments.

FIG. 14A constitutes a representation of SDR-1/NP protein isoforms in human breast tumors and normal and tumor colon. Specifically, FIG. 14A shows a direct immunoblot of 5 human breast cancer extracts using anti-SDR-1/NP polyclonal antibodies.

25 FIG. 14B shows the blot (from FIG. 14A, above) stripped and detected with VH Ab-1.

FIG. 14C shows an immunological evaluation of SDR-1/NP in normal colon tissue using the polyclonal antibody.

FIG. 14D shows an immunological evaluation of colon tumor samples using the anti-SDR-1/NP polyclonal antibody.

30 FIG. 15 shows the distribution of high levels of SDR-1/NP in invasive sub-types of human breast cancer. Archived breast cancer samples were transferred into a tissue array as a panel of 40 cases with multiple cores from each. Immunohistochemical detection of SDR-1/NP was evaluated. Invasive ductal carcinoma and Lobular carcinoma positive cores are shown as representative examples (A-C), with sample A showing a mixed phenotype. Invasive

- 5 tumors that displayed uniform staining are also shown (D,E). Staining of a stromal tissue is negative for SDR-1/NP (F.)

FIG. 16A shows the results of a molecular analysis of SDR-1/NP expression in human breast cancer cells, employing standard RT-PCR of RNA from MDA-MB-231 cells. GAPDH control is shown in lane 1, and a 5' coding region of alpha SDR-1 is shown in lane 2.

- 10 FIG. 16B shows the 5' and 3' RACE evaluation of SDR-1 cDNA structure in MDA-MB-231 cells. Lane 1 expected 5' RACE clone at 1353bp, Lane 2 transferin receptor control for PCR, Lane 3 lack of expected 3' amplification using a C-terminal primer to SDR-1/NP.

FIG. 16C provides a comparison of mRNA levels in matched normal versus tumor tissues using a Cancer cDNA Profiling Array. Hybridization of radiolabeled 5' coding fragment
15 (panel A, lane 2) was performed against the Clontech Cancer Profiling II cDNA Array. A subset of normal (N) matched to breast tumor (T) from the indicated tissues is presented.

DETAILED DESCRIPTION OF THE INVENTION

In order for the full scope of the invention to be clearly understood, the following definitions are provided.

20

I. Definitions

- The term "homologous" is intended to include a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent amino acid residues or nucleotides, *e.g.*, an amino acid residue which has a similar side chain, to a
25 second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains and/or a common functional activity.

- The term an "isolated polypeptide" is substantially free of cellular material or other
30 contaminating polypeptides from the microorganism from which the polypeptide is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized.

- The term "nucleic acid" is intended to include nucleic acid molecules, *e.g.*, polynucleotides
35 which include an open reading frame encoding a polypeptide, and can further include non-coding regulatory sequences, and introns. In addition, the terms are intended to include one or more genes that map to a functional locus. In addition, the terms are intended to include

- 5 a specific gene for a selected purpose. The gene can be endogenous to the host cell or can be recombinantly introduced into the host cell, *e.g.*, as a plasmid maintained episomally or a plasmid (or fragment thereof) that is stably integrated into the genome.

The term "recombinant nucleic acid molecule" includes a nucleic acid molecule (*e.g.*, a
10 DNA molecule) that has been altered, modified or engineered such that it differs in nucleotide sequence from the native or natural nucleic acid molecule from which the recombinant nucleic acid molecule was derived (*e.g.*, by addition, deletion or substitution of one or more nucleotides). Advantageously, a recombinant nucleic acid molecule (*e.g.*, a recombinant DNA molecule) includes an isolated nucleic acid molecule or gene of the
15 present invention (*e.g.*, an isolated *pdc* gene) operably linked to regulatory sequences.

II. Isolated Nucleic Acid Molecules and Genes

In one embodiment, the present invention features an isolated nucleic acid molecule. The nucleic acid molecule includes DNA molecules (*e.g.*, linear, circular, cDNA or
20 chromosomal DNA) and RNA molecules (*e.g.*, tRNA, rRNA, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but advantageously is double-stranded DNA. The isolated nucleic acid molecule of the invention includes a nucleic acid molecule which is free of sequences which naturally flank the nucleic acid molecule (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid molecule) in the chromosomal DNA of the organism from
25 which the nucleic acid is derived. In various embodiments, an isolated nucleic acid molecule can contain less than about 10 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, 0.1 kb, 50 bp, 25 bp or 10 bp of nucleotide sequences which naturally flank the nucleic acid molecule in chromosomal DNA of the microorganism from which the nucleic acid molecule is
30 derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular materials when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

In another embodiment, an isolated nucleic acid molecule of the present invention
35 comprises a nucleotide sequence which is at least about 60-65%, advantageously at least about 70-75%, more preferable at least about 80-85%, and even more advantageously at least about 90-95% or more identical to a nucleotide sequence set forth as SEQ ID NO:1.

In another embodiment, an isolated nucleic acid molecule hybridizes under stringent
40 conditions to a nucleic acid molecule having a nucleotide sequence set forth as SEQ ID NO:1. Such stringent conditions are known to those skilled in the art and can be found in

5 *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A particular, non-limiting example of stringent (*e.g.* high stringency) hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Advantageously, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the
10 sequence of SEQ ID NO:1 corresponds to a naturally-occurring nucleic acid molecule. Typically, a naturally-occurring nucleic acid molecule includes an RNA or DNA molecule having a nucleotide sequence that occurs in nature.

A nucleic acid molecule of the present invention (*e.g.*, a nucleic acid molecule having the
15 nucleotide sequence of SEQ ID NO:1) can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, nucleic acid molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press,
20 Cold Spring Harbor, NY, 1989) or can be isolated by the polymerase chain reaction using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1. A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. In another embodiment, an isolated nucleic acid molecule of
25 the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1.

Additional contemplated nucleic acid sequences are those that comprise the nucleotide
sequence of SEQ ID NO:1, that encode a homologue of the polypeptide having the amino
30 acid sequence set forth in SEQ ID NO:2 (*e.g.*, encode a polypeptide having at least 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or more identity to the polypeptide having the amino acid sequence as set forth in SEQ ID NO:2, and having a substantially identical activity as the polypeptide), hybridize under stringent conditions to all or a portion of a
nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1 or to all or a portion
35 of a nucleic acid molecule that encodes a polypeptide having the amino acid sequence of SEQ ID NO:2.

In yet another embodiment, the isolated nucleic acid molecule of the invention encodes a
homologue of the polypeptide having the amino acid sequence of SEQ ID NO:2. Typically,
40 the term "homologue" includes a polypeptide or polypeptide sharing at least about 30-35%, advantageously at least about 35-40%, more advantageously at least about 40-50%, and

- 5 even more advantageously at least about 60%, 70%, 80%, 90% or more identity with the amino acid sequence of a wild-type polypeptide or polypeptide described herein and having a substantially equivalent functional or biological activity as the wild-type polypeptide or polypeptide. For example, a homologue contemplated herein shares at least about 30-35%, advantageously at least about 35-40%, more advantageously at least about 40-50%, and
- 10 even more advantageously at least about 60%, 70%, 80%, 90% or more identity with the polypeptide having the amino acid sequence set forth as SEQ ID NO:2, and has a substantially equivalent functional or biological activity (*i.e.*, is a functional equivalent) of the polypeptide having the amino acid sequence set forth as SEQ ID NO:2.
- 15 In an embodiment, the isolated nucleic acid molecule of the invention comprises a nucleotide sequence that encodes a polypeptide as set forth in SEQ ID NO:2.

- In another embodiment, the isolated nucleic acid molecule of the invention hybridizes to all or a portion of a nucleic acid molecule having the nucleotide sequence set forth in SEQ ID
- 20 NO:1 or hybridizes to all or a portion of a nucleic acid molecule having a nucleotide sequence that encodes a polypeptide having the amino acid sequence of SEQ ID NO:2.

- Such hybridization conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., John Wiley & Sons, Inc.
- 25 (1995), sections 2, 4 and 6. Additional stringent conditions can be found in *Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), chapters 7, 9 and 11.

III. Recombinant Nucleic Acid Molecules and Vectors

- 30 The present invention further features recombinant nucleic acid molecules (*e.g.*, recombinant DNA molecules) that include nucleic acid molecules described herein (*e.g.*, isolated nucleic acid molecules).

- The present invention further features expression vectors (*e.g.*, recombinant vectors) that
- 35 include nucleic acid molecules (*e.g.*, isolated or recombinant nucleic acid molecules and/or genes) described herein. The recombinant vector (*e.g.*, plasmid, phage, phasmid, virus, cosmid or other purified nucleic acid vector) can be altered, modified or engineered such that it contains greater, fewer or different nucleic acid sequences than those included in the native or natural nucleic acid molecule from which the recombinant vector was derived.

40

- 5 Typically, the expression vector further comprises regulatory sequence(s) which allows for expression (*e.g.*, enhanced, increased, constitutive, basal, attenuated, decreased or repressed expression) of the nucleotide sequence, advantageously expression of a gene product encoded by the nucleotide sequence.
- 10 The regulatory sequence includes nucleic acid sequences which affect (*e.g.*, modulate or regulate) expression of other nucleic acid sequences. In one embodiment, a regulatory sequence is included in a recombinant nucleic acid molecule or recombinant vector in a similar or identical position and/or orientation relative to a particular gene of interest as is observed for the regulatory sequence and gene of interest as it appears in nature, *e.g.*, in a
- 15 native position and/or orientation. For example, a gene of interest can be included in a recombinant nucleic acid molecule or recombinant vector operably linked to a regulatory sequence which accompanies or is adjacent to the gene of interest in the natural organism (*e.g.*, operably linked to "native" regulatory sequences, for example, to the "native" promoter). Alternatively, a gene of interest can be included in a recombinant nucleic acid
- 20 molecule or recombinant vector operably linked to a regulatory sequence. In one embodiment, a regulatory sequence is a non-native or non-naturally-occurring sequence (*e.g.*, a sequence which has been modified, mutated, substituted, derivatized, deleted including sequences which are chemically synthesized). Advantageous regulatory sequences include promoters, enhancers, termination signals, anti-termination signals and
- 25 other expression control elements (*e.g.*, sequences to which repressors or inducers bind and/or binding sites for transcriptional and/or translational regulatory polypeptides, for example, in the transcribed mRNA). Such regulatory sequences are described, for example, in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold
- 30 Spring Harbor, NY, 1989.

IV. Isolated Polypeptides

- Another aspect of the present invention features isolated polypeptides. An isolated or purified polypeptide is substantially free of cellular material or other contaminating
- 35 polypeptides from the organism from which the polypeptide is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, an isolated or purified polypeptide has less than about 30% (by dry weight) of contaminating polypeptide or chemicals, more advantageously less than about 20% of contaminating polypeptide or chemicals, still more advantageously less than about 10% of
- 40 contaminating polypeptide or chemicals, and most advantageously less than about 5% contaminating polypeptide or chemicals.

5 It is well understood that one of skill in the art can mutate (*e.g.*, substitute) nucleic acids which, due to the degeneracy of the genetic code, encode for an identical amino acid as that encoded by the naturally-occurring gene. This may be desirable in order to improve the codon usage of a nucleic acid to be expressed in a particular organism. Moreover, it is well understood that one of skill in the art can mutate (*e.g.*, substitute) nucleic acids which
10 encode for conservative amino acid substitutions. It is further well understood that one of skill in the art can substitute, add or delete amino acids to a certain degree without substantially affecting the function of a gene product as compared with a naturally-occurring gene product, each instance of which is intended to be included within the scope of the present invention.

15 In an embodiment, an isolated polypeptide of the present invention has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, an isolated polypeptide of the present invention is a homologue of the polypeptide set forth as SEQ ID NO:2 (*e.g.*, comprises an amino acid sequence at least about 30-40% identical, advantageously about
20 40-50% identical, more advantageously about 50-60% identical, and even more advantageously about 60-70%, 70-80%, 80-90%, 90-95% or more identical to the amino acid sequence of SEQ ID NO:2, and has an activity that is substantially similar to that of the polypeptide encoded by the amino acid sequence of SEQ ID NO:2.

25 In one embodiment, the invention is directed to a method for isolating and identifying lymph node immune reactive cells responsive to tumor-derived antigens. The immune B-cells that produce antibodies that identify these tumor antigens may be selected *in situ*. Molecular techniques may be employed to recover the gene sequences encoding the antibody proteins. These sequences may then be analyzed for structural characteristics that
30 indicate whether they have been produced in response to tumor antigen exposure. The antibodies identified as antigen-stimulated may then be synthesized in the laboratory and used to purify its specific target, thus identifying novel tumor antigens. In a further embodiment of the invention, tumor antigen identification can be verified for development of the antigens as diagnostic and therapeutic targets for a particular tumor.

35 In one embodiment of the invention, the molecular antibodies developed have utility for novel diagnostic panels for screening tumor antigen profiles and detecting antigens present in blood or urine samples as a biomarker of metastatic cancer. In another embodiment of the invention, the antibodies can be used to treat disseminated cancer by means of targeting tumor-specific antigen(s). Such a therapeutic approach may be selective to antigen

- 5 expression by the cancer of each individual patient, thus providing patient-selective therapy for late metastatic disease, where few effective methods exist.

Further embodiments of the invention include methods for identifying lymph node immune-reactive areas, for isolating anti-tumor antibody sequences, for producing recombinant antibody antigen-recognizing domain proteins, for identifying novel breast cancer antigens,
10 and for establishing methods to assay for tumor antigens in biological samples. The methodology of the invention is additionally applicable to solid and/or metastatic tumors, such as melanoma, ovarian, cervical, pancreatic, esophageal, stomach, colon and prostate.

In one embodiment, the method of the invention integrates molecular and biochemical technologies in an efficient throughput method. Each of the steps described is performed in
15 a linked and quality assured process. In another embodiment of the invention, the time from sample acquisition to a complete sequencing profile is minimal (less than two weeks), and several samples can be processed at the same time.

In one embodiment of the invention, the key elements of the method of the invention comprises at least one of the following: 1) identification of reactive B-cell germinal centers
20 as opposed to those that are not; 2) germinal center excision and RNA recovery; 3) VH, V λ , and V κ library synthesis and medium throughput sequencing; 4) bioinformatics of variable region sequences to define antigen-dependent somatic hypermutation and clonal expansion; 5) recombinant protein synthesis and purification methods; 6) effective microscale antigen-trap assays and antigen identification using recombinant VH proteins; and 7) methods to
25 verify antigen relevance to breast cancer via immunochemical, biochemical, and molecular means.

In one embodiment, the invention provides antibody reagents that define the common antigen presentation in a specific cancer type found in an individual undergoing surgical tumor resection, along with sentinel and proximal lymph node mapping. In a further
30 embodiment of the invention, the antibody reagents are employed as clinical diagnostic screens of two sample formats: i) primary tumor extracts, defining the antigen presentation profile for each patient; and ii) serum or urine samples, detecting antigen peptides released from tumors in asymptomatic people, being applied as cancer screens for annual physicals, insurance screens, or for post-surgical patients to indicate distal metastatic relapse.

35 In one embodiment of the invention, recombinant antibody molecules produced in mammalian expression systems are incorporated into analytical assays to effectively screen

5 primary cancers such as urinary cancer, blood cancer, biopsy cancer, and surgical cancer, histological material, and biological fluids in an effort to evaluate their potential as diagnostic biomarkers. In a particular embodiment of the invention, the antibody molecules are incorporated as fusion proteins, such as antibody heavy chain-fusion proteins with C-terminal tags or heavy chain-linker-light chain fusion proteins with C-terminal tags or the
10 like.

Thus, one embodiment of the invention comprises the use of recombinant antibody fusion proteins to identify tumor-related antigens expressed in tumor samples. In a further embodiment of the invention, the recombinant antibody fusion proteins have significant utility for diagnostic identification and classification of primary tumors, as well as a
15 diagnostic screen for metastatic tumor burden in the form of a laboratory test. Such a diagnostic screen may have as many as 10-50 such antigens to profile prospective tumor burden in patient blood, urine, biopsy or surgical samples.

In one embodiment, the invention is directed to a method for identifying and selecting reactive lymph node germinal centers from tumor-draining lymph nodes or lymph nodes in
20 pathogenic inflammation. In a further embodiment of the invention, the human antibody variable domains are isolated and cloned. The antigen-stimulated clonally-derived subgroups may subsequently be analyzed.

In one embodiment of the invention, stromal cell derived receptor-1 is identified as a potential cancer antigen. SDR1 has been identified as a cDNA encoding an Ig-domain
25 containing transmembrane protein in a genetic screen of bone marrow stromal cells. A unique isoform of SDRI in breast cancer (45kDa) has been characterized and also found in other epithelial cancers. This gene produces, by alternative splicing, 16 types of transcripts, predicted to encode 15 distinct proteins, many of which are likely to be potential cancer antigens.

30 In additional embodiments of the invention, the extracellular domains of SDR1 are employed as: 1) unique tumor biomarkers in urine or plasma indicating tumor progression and protein shedding *in vivo*; or 2) a potential target for directed therapeutics, tumor imaging and tumor-selective delivery systems. Accordingly, SDR1 reagents have been developed and include specific immune detection molecules and soluble recombinant
35 protein isoforms.

- 5 In one embodiment of the invention, the selection of self-activated antibodies in cancer-associated lymph nodes utilizes each individual patient's biological threshold to identify select and targeted tumor antigens. As a result, the antibody sequences obtained are all "unique".

10 In additional embodiments, the potential diagnostic and therapeutic utility of the methods, antibodies, and antigens of the invention is also applicable to inflammation and pathological infections.

As used here in, the term "recombinant" is used to describe nucleic acid sequences (or the amino acid sequences encoded thereby) that have been joined to other nucleic acid sequences by genetic modification (e.g., enzymatic or chemical processes).

- 15 A "fusion polypeptide" as used herein is a polypeptide joined to another polypeptide or amino acid sequence (e.g., biomolecule partition motif) by peptide-bond formation or affinity-based techniques.

20 Isolated proteins of the present invention have an amino acid sequence sufficiently homologous to the amino acid sequence of SDR1 or are encoded by a nucleotide sequence sufficiently homologous to that of SDR1. As used herein, the term "sufficiently homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue-which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common
25 structural domains or motifs or a common functional activity. All degenerate variants of the nucleotide sequence of SDR1 are considered to be "sufficiently homologous" to the same.

Accordingly, another embodiment of the invention features isolated SDR1 subunit proteins and polypeptides, fragments, and variants thereof having SDR1 subunit activity.

30 In still another embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 60% or more homologous to the entire length of the nucleotide sequence shown herein for SDR1 (SEQ ID NO: 1), or a portion of this nucleotide sequence.

5 SEQ ID NO:1

LOCUS: NM_012428 2388 bp mRNA linear PRI 23-AUG-2004
 DEFINITION: Homo sapiens stromal cell derived factor receptor 1
 (SDFR1), transcript variant beta, mRNA.
 ACCESSION: NM_012428

```

10      1 gaaggacgga gccgagccgc ggctgcctcc ctgctcact ccctcgcgca
      ctgcccgc
      61 cctccctcc ctcccctccc ttcccgggc ccgggtctgg ccccgccca
      ttgctgttg
15     121 ggtctctgc tagggaggat gtgggttcg tctgtccca gcgccctggc
      cctctcgtg
      181 ttgctggtct ctggctccct cctcccagg ccaggcgccg ctcaaacgc
      tgggtttgtc
      241 aagtgcacca tgtcagaaac taagctcac ggggacgcct ttgagctgta
20     ctgtgacgtg
      301 gtcgggagcc ccacgccaga gatccagtgg tggtagcag aagtcaaccg
      ggcagagtct
      361 ttcagacagc tgtgggacgg tgctcggaag cgccgtgtca ccgtaaacac
      cgctacggg
25     421 tcaaacggcg tgagtgtgct gagaataacc cggctcacct tggaggactc
      tgggacttac
      481 gagtgcaggg ccagcaacga cccaagagg aatgacttga ggcaaaacc
      ctocataaca
      541 tggattcgag ccaggccac cataagcgtc cttcagaagc caaggattgt
30     caccagtga
      601 gaggtcatta ttcgagacag ccctgttctc cctgtcacc tgcatgtaa
      cctcacctcc
      661 agctctcaca cccttacata cagctactgg acaagaatg gggtggaact
      gagtgcact
35     721 cgtaagaatg ccagcaacat ggagtacagg atcaataagc cgagagctga
      ggattcaggc
      781 gaataccact gcgtatatca cttgtcagc gtcctaaag caaacgccac
      cattgaagtg
      841 aaagccgctc ctgacatcac tggccataaa cggagtgaga acaagaatga
40     agggcaggat
      901 gccactatgt attgcaagtc agttggctac cccacccag actggatatg
      gcgcaagaag
      961 gagaacggga tgcccatgga cattgtcaat acctctggcc gcttcttcat
      catcaacaag
45     1021 gaaaattaca ctgagttgaa cattgtgaac ctgcagatca cggaagacc
      tggcgagtat
      1081 gaatgtaatg ccaccaacgc cattggctcc gcctctgttg toactgtcct
      cagggtgcgg
      1141 agccacctgg cccactctg gcctttcttg ggaattctgg ctgaaattat
50     catccttctg
      1201 gtgatcattg ttgtgtatga gaagaggaag aggcagatg aggttctcta
      cgatgatgaa
      1261 ccagctggac caatgaaaac caactctacc aacaatcaca aagataaaaa
      cttgcgccag
55     1321 agaaacacaa attaatgact gcttacaata tctttagggt cctgaaactg
      gtggcaacat
      1381 gacctgttaa aattttctgc ttggacctct ttggttctct cccctttcaa
      gtgagcaaca
      1441 ccacaatgac tgtctaaagc atgccttatt tagcctctcc tgtaaggggtg
60     atctagccag
  
```

5 1501 gtacatttta aacaatgctt cagtgtagaa ggtgtaaact attttgggct
tgatgtgctg
1561 tgaatgttgc tttttttttt cctttgttaa aatatttaaa tagaagtga
aaggtcctct
10 1621 gaggatcaga tcatgcatgc gccatttttt acttaatgca gctgttaa
tggaagct
1681 ctaaaatgca ctgctgccat ctagtgtatc acttttgtaa agtacagca
aacctacagg
1741 tatatacagc atataaatat atatataat atatttatat ttttgggggt
gggagaaatc
15 1801 caaaataaag taaatgcttg tttcattttt aagctgctga tattcattcc
ttattgtatg
1861 ttgtcagatg aggaaattgt gcagttctgg tacataaaga tgagtaatat
aaactgaaat
1921 ctataatttt aagggtctaa cctgtgactt taataagctg gaacagtcca
20 ctgaatgggt
1981 ataataaatt gcagtataat cgtatgattg ctttttaagt gattatcttt
tcttctgtta
2041 agtcatgtaa attcataaat ccttttgcac tgatgtgttg aaccttattc
ttgtacattc
25 2101 attcaatcaa ggcaaacctt tataattttt cttttgtttc caatgacctt
gaaatgttat
2161 agcatggtaa tattctatgc aactatagtt ataacttttg gtttgacact
gtattttttc
2221 acattgattt actggttgat gatagatttt ataacctaac ggttctcatg
30 cgggtgcgtaa
2281 ttgtagatgc atgtacttgt gtgttttgtg taactattga agtgcaatga
tgtataaaaa
2341 agtggattca cctgttttta aaaataaaac attgataaaa aaaaaaaa
35

- 5 **SEQ ID NO:2** corresponds to the amino acid sequence as follows:

CDS: 139..1335

10 MSGSSLPALALSILLVSGSLLPGPGAQNAQNFVKSPMSETKLTGDAFELYCDVVGSPPTPEIQWYAE
VNRAESFRQLWDGARKRRVTVNTAYGSNGVSVLRITRLTLEDSTGYECRASNDPKRNDLRQNPSITWI
RAQATISVLQKPRIVTSEEVIIRDSPVLPVTLQCNTLSSSHTLTYSYWTKNGVELSATRKNASNMEYR
INKPRAEDSGEYHCVYHFVSAPKANATIEVKAAPDITGHKRSENKNEGQDATMYCKSVGYPHPDWIWR
KKENGMPMDIVNTSGRFFIINKENYTELNIVNLQITEDPGEYECNATNAIGSASVVTVLVRSHLAPL
WPFLGILAEIIILVVIIVVYEKRKRKPDEVPDDDEPAGPMKTNSTNNHKDKNLRQNTN

- 15 An isolated nucleic acid molecule of the invention that hybridizes under moderate or highly stringent conditions to the sequence of SDR1 can correspond to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

- 20 In addition to naturally-occurring allelic variants of SDR1 sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of the protein, thereby leading to changes in the amino acid sequence of the encoded protein, without altering its functional ability. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential"
- 25 amino acid residues can be made in the sequence of SDR1. A "non-essential" amino acid residue is a residue that can be altered from the disclosed sequence of SDR1 without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding SDR1 protein that contain changes in amino acid residues that are not
- 30 essential for activity.

- The invention involves, in one aspect, defined nucleic acid molecules and, in another, defined polypeptides, while also embracing isolated functionally equivalent variants, degenerate variants, useful analogs and fragments of the foregoing nucleic acids and polypeptides; complements of the foregoing nucleic acids; and molecules which selectively
- 35 bind the foregoing nucleic acids and polypeptides.

- Nucleic acid molecules that code for the defined polypeptides have the nucleotide sequence of SEQ ID NO:1 and nucleotide sequences that differ from SEQ ID NO:1 in codon sequence due to the degeneracy of the genetic code (i.e. degenerate variants). Contemplated
- 40 nucleic acid molecules also include alleles of the foregoing nucleic acid molecules, as well as fragments thereof. Such fragments can be used, for example, as probes in hybridization

- 5 assays and as primers in a polymerase chain reaction (PCR). Complements of the foregoing nucleic acid molecules also are embraced by the invention.

The nucleic acid molecules and polypeptides of the invention may be isolated. As used herein with respect to nucleic acid molecules, the term "isolated" means, without limitation:

10 (i) amplified *in vitro* by, for example, polymerase chain reaction (PCR); (ii) recombinantly produced by cloning; (iii) purified, as by cleavage and gel separation; or (iv) synthesized by, for example, chemical synthesis. An isolated nucleic acid molecule is one which is readily manipulated by recombinant DNA techniques well known in the art. Thus, a nucleotide sequence contained in a vector in which 5' and 3' restriction sites are known or

15 for which polymerase chain reaction (PCR) primer sequences have been disclosed is considered isolated but a nucleic acid sequence existing in its native state in its natural host is not. An isolated nucleic acid molecule may be substantially purified, but need not be. For example, a nucleic acid molecule that is isolated within a cloning or expression vector is not pure in that it may comprise only a tiny percentage of the material in the cell in which it

20 resides. Such a nucleic acid molecule is isolated, however, as the term is used herein because it is readily manipulated by standard techniques known to those of ordinary skill in the art. An isolated nucleic acid molecule as used herein is not a naturally occurring chromosome.

25 As used herein with respect to polypeptides, "isolated" means separated from its native environment and present in sufficient quantity to permit its identification or use. Isolated, when referring to a protein or polypeptide, means, for example: (i) selectively produced by expression cloning or (ii) purified as by chromatography or electrophoresis. Isolated proteins or polypeptides may be, but need not be, substantially pure. The term

30 "substantially pure" means that the proteins or polypeptides are essentially free of other substances with which they may be found in nature or *in vivo* systems to an extent practical and appropriate for their intended use. Substantially pure polypeptides may be produced by techniques well known in the art. Because an isolated protein may be admixed with a pharmaceutically acceptable carrier in a pharmaceutical preparation, the protein may

35 comprise only a small percentage by weight of the preparation. The protein is nonetheless isolated in that it has been separated from the substances with which it may be associated in living systems, i.e. isolated from other proteins.

5 The invention also encompasses allelic nucleic acids. Alleles of the defined nucleic acid molecules of the invention can be identified by conventional techniques. For example, they can be isolated by hybridizing a probe which includes at least a fragment of SEQ ID NO:1 under stringent conditions with a nucleic acid library and selecting positive clones. Thus, the invention embraces nucleic acid molecules that code for the defined polypeptides and
10 that hybridize to a nucleic acid molecule consisting of SEQ ID NO:1 under stringent conditions. The term "stringent conditions" as used herein refers to parameters with which the art is familiar. Nucleic acid hybridization parameters may be found in references which compile such methods, e.g. *Molecular Cloning: A Laboratory Manual*, J. Sambrook, et al., eds., Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York,
15 1989, or *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York.

In general homologs and alleles typically will share at least 85% nucleotide identity and/or at least 90% amino acid identity to the sequences of the defined nucleic acid molecules and
20 polypeptides of the invention, respectively, in some instances will share at least 90% nucleotide identity and/or at least 95% amino acid identity and in still other instances will share at least 95% nucleotide identity and/or at least 99% amino acid identity and in still yet other instances will share at least 99% nucleotide identity and/or at least 99.5% amino acid identity. The identity can be calculated using various, publicly available software tools
25 developed by NCBI (Bethesda, Maryland) that can be obtained through the internet. Exemplary tools include the BLAST system also available through the NCBI website, Pairwise and ClustalW alignments (BLOSUM30 matrix setting) as well as Kyte-Doolittle hydropathic analysis can be obtained using the MacVector sequence analysis software (Oxford Molecular Group, Hunt Valley, MD). Watson-Crick complements of the foregoing
30 nucleic acids also are embraced by the invention.

The nucleic acid molecules of the invention also include degenerate nucleic acid molecules which include alternative codons to those present in the native materials. For example, serine residues are encoded by the codons TCA, AGT, TCC, TCG, TCT and AGC. Each of
35 the six codons is equivalent for the purposes of encoding a serine residue. Thus, it will be apparent to one of ordinary skill in the art that any of the serine-encoding nucleotide triplets may be employed to direct the protein synthesis apparatus, *in vitro* or *in vivo*, to incorporate a serine residue into an elongating polypeptide of the invention. Similarly, nucleotide sequence triplets which encode other amino acid residues include, but are not limited to:

5 CCA, CCC, CCG and CCT (proline codons); CGA, CGC, CGG, CGT, AGA and AGG (arginine codons); ACA, ACC, ACG and ACT (threonine codons); AAC and AAT (asparagine codons); and ATA, ATC and ATT (isoleucine codons). Other amino acid residues may be encoded similarly by multiple nucleotide sequences. Thus, the invention embraces degenerate nucleic acid molecules that differ from the biologically isolated
10 nucleic acids in codon sequence due to the degeneracy of the genetic code.

The invention also provides modified nucleic acid molecules which include additions, substitutions and deletions of one or more nucleotides. In preferred embodiments, these modified nucleic acid molecules and/or the polypeptides they encode retain at least one
15 activity or function of the unmodified nucleic acid molecule and/or the polypeptides, such as T-cell chemoattractant activity, etc. In certain embodiments, the modified nucleic acid molecules encode modified polypeptides, preferably polypeptides having conservative amino acid substitutions as are described elsewhere herein. The modified nucleic acid molecules are structurally related to the unmodified nucleic acid molecules and in preferred
20 embodiments are sufficiently structurally related to the unmodified nucleic acid molecules so that the modified and unmodified nucleic acid molecules hybridize under stringent conditions known to one of skill in the art.

For example, modified nucleic acid molecules which encode polypeptides having single
25 amino acid changes can be prepared. Each of these nucleic acid molecules can have one, two or three nucleotide substitutions exclusive of nucleotide changes corresponding to the degeneracy of the genetic code as described herein. Likewise, modified nucleic acid molecules which encode polypeptides having two amino acid changes can be prepared which have, e.g., 2-6 nucleotide changes. Numerous modified nucleic acid molecules like
30 these will be readily envisioned by one of skill in the art, including for example, substitutions of nucleotides in codons encoding amino acids 2 and 3, 2 and 4, 2 and 5, 2 and 6, and so on. In the foregoing example, each combination of two amino acids is included in the set of modified nucleic acid molecules, as well as all nucleotide substitutions which code for the amino acid substitutions. Additional nucleic acid molecules that encode
35 polypeptides having additional substitutions (i.e., 3 or more), additions or deletions (e.g., by introduction of a stop codon or a splice site(s)) also can be prepared and are embraced by the invention as readily envisioned by one of ordinary skill in the art. Any of the foregoing nucleic acid molecules or polypeptides can be tested by routine experimentation for

- 5 retention of structural relation or activity to the nucleic acids and/or polypeptides disclosed herein.

The invention also provides isolated fragments of SEQ ID NO:2. The fragments can be used as probes in Southern blot assays to identify such nucleic acids, or can be used in
10 amplification assays such as those employing PCR. Fragments also can be used to produce fusion proteins for generating antibodies or determining binding of the polypeptide fragments. Likewise, fragments can be employed to produce non-fused fragments of the defined polypeptides, useful, for example, in the preparation of antibodies, in immunoassays, and the like.

15 The invention also includes functionally equivalent variants, which include variant nucleic acids and polypeptides which retain one or more of the activities of the SDR1. For example, variants include truncations, deletions, point mutations, or additions of amino acids to the sequence of SEQ ID NO:2 which retain one or more activities of SDR1. Functionally
20 equivalent variants also include a SDF1X polypeptide which has had a portion removed or replaced by a similar domain from another SDR1 (e.g. a "domain-swapping" variant). Other functionally equivalent variants will be known to one of ordinary skill in the art, as will methods for preparing such variants. The activity of a functionally equivalent variant can be determined using the methods provided herein, and in references that have described
25 assays using SDR1 polypeptides. Such variants are useful, *inter alia*, for determining the portions of the SDR1 which are required for activity. Variants which are non-functional also can be prepared as described above. Such variants are useful, for example, as negative controls in experiments testing SDR1 activity.

30 The defined nucleic acid molecules and polypeptides of the invention can be delivered to the target cell alone or in association with a vector. In its broadest sense, a "vector" is any vehicle capable of facilitating: (1) delivery of the nucleic acid molecule or polypeptide to a target cell, (2) uptake of the nucleic acid or polypeptide by a target cell, or (3) expression of the nucleic acid molecule or polypeptide in a target cell. Preferably, the vectors transport
35 the nucleic acid or polypeptide into the target cell with reduced degradation relative to the extent of degradation that would result in the absence of the vector. Optionally, a "targeting ligand" can be attached to the vector to selectively deliver the vector to a cell which expresses on its surface the cognate receptor for the targeting ligand (e.g. a receptor, an antigen recognized by an antibody). In this manner, the vector (containing the nucleic acid

5 molecule or polypeptide) can be selectively delivered to a specific cell. In general, the vectors useful in the invention are divided into two classes: biological vectors and chemical/physical vectors. Biological vectors are more useful for delivery and/or uptake of the nucleic acid. Chemical/physical vectors are more useful for delivery and/or uptake of nucleic acids or proteins.

10

Biological vectors include, but are not limited to, plasmids, phagemids, viruses, other vehicles derived from viral or bacterial sources that have been manipulated by the insertion or incorporation of the nucleic acid sequences of the invention, and free nucleic acid fragments which can be linked to the nucleic acid sequences of the invention.

15

Expression vectors containing all the necessary elements for expression are commercially available and known to those skilled in the art. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989. Cells are genetically engineered by the introduction into the cells of heterologous DNA encoding the defined polypeptide or fragment or variant thereof. The heterologous DNA is placed under operable control of transcriptional elements to permit the expression of the heterologous DNA in the host cell.

20

In addition to the biological vectors, chemical/physical vectors may be used to deliver the nucleic acid molecule or polypeptide to a target cell and facilitate uptake thereby. As used herein, a "chemical/physical vector" refers to a natural or synthetic molecule, other than those derived from bacteriological or viral sources, capable of delivering the isolated nucleic acid molecule or polypeptide to a cell.

25

30 Fragments of a polypeptide preferably are those fragments which retain a distinct functional capability of the defined (SDR1) polypeptide. Functional capabilities which can be retained in a fragment of the defined polypeptide include interaction with antibodies and interaction with other polypeptides (such as would be found in a protein complex). Those skilled in the art are well versed in methods for selecting fragments which retain a functional capability of the defined polypeptide. Confirmation of the functional capability of the fragment can be carried out by synthesis of the fragment and testing of the capability according to standard methods.

35

5 The invention embraces variants of the defined polypeptides described above. As used herein, a “variant” of a polypeptide is a polypeptide which contains one or more modifications to the primary amino acid sequence of the defined polypeptide. Modifications which create such a variant can be made to the polypeptide for a variety of reasons, including to enhance a property of the polypeptide, such as protein stability in an expression system or the stability of protein-protein binding; or to provide a novel activity or property to the polypeptide, such as addition of an antigenic epitope or addition of a detectable moiety. Modifications to the defined polypeptide are typically made to the nucleic acid molecule which encodes the polypeptide, and can include deletions, point mutations, truncations, amino acid substitutions and additions of amino acids or non-amino acid moieties. Alternatively, modifications can be made directly to the polypeptide, such as by cleavage, addition of a linker molecule, addition of a detectable moiety, such as biotin, addition of a fatty acid, and the like. Modifications also embrace fusion proteins comprising all or part of the SDR1 amino acid sequence. One of skill in the art will be familiar with methods for predicting the effect on protein conformation of a change in protein sequence, and can thus “design” a variant polypeptide according to known methods.

Mutations of a nucleic acid molecule which encodes the SDR1 polypeptide or a fragment thereof preferably preserve the amino acid reading frame of the coding sequence, and preferably do not create regions in the nucleic acid which are likely to hybridize to form secondary structures, such as hairpins or loops, which can be deleterious to expression of the variant polypeptide.

Mutations can be made by selecting an amino acid substitution, or by random mutagenesis of a selected site in a nucleic acid molecule which encodes the polypeptide. Variant polypeptides are then expressed and tested for one or more activities to determine which mutation provides a variant polypeptide with a desired property. Further mutations can be made to variants (or to non-variant polypeptides) which are silent as to the amino acid sequence of the polypeptide, but which provide preferred codons for translation in a particular host. The preferred codons for translation of a nucleic acid in, e.g., *E. coli*, are well known to those of ordinary skill in the art. Still other mutations can be made to the noncoding sequences of the SDR1 gene or cDNA clone to enhance expression of the polypeptide.

5 The skilled artisan will also realize that conservative amino acid substitutions may be made in the defined polypeptides to provide functionally equivalent variants of the foregoing polypeptides, i.e., variants which retain the functional capabilities of the polypeptides. As used herein, a "conservative amino acid substitution" refers to an amino acid substitution which does not alter the relative charge or size characteristics of the polypeptide in which
10 the amino acid substitution is made.

As used herein, the compounds of this invention, e.g., antigens and antibodies identified by the methods described herein, are defined to include pharmaceutically acceptable derivatives or prodrugs thereof. A "pharmaceutically acceptable derivative or prodrug"
15 means any pharmaceutically acceptable salt, ester, salt of an ester, or other derivative of a compound of this invention which, upon administration to a recipient, is capable of providing (directly or indirectly) a compound of this invention. Particularly favored derivatives and prodrugs are those that increase the bioavailability of the compounds of this invention when such compounds are administered to a mammal (e.g., by allowing an orally
20 administered compound to be more readily absorbed into the blood) or which enhance delivery of the parent compound to a biological compartment (e.g., the brain or lymphatic system) relative to the parent species. Prodrugs include derivatives where a group which enhances aqueous solubility or active transport through the gut membrane is appended to the structure of formulae described herein. Pharmaceutically acceptable salts of the compounds
25 of this invention include those derived from pharmaceutically acceptable inorganic and organic acids and bases.

The compounds of the formulae described herein can, for example, be administered by injection, intravenously, intraarterially, subdermally, intraperitoneally, intramuscularly, or subcutaneously; or orally, buccally, nasally, transmucosally, topically, in an ophthalmic
30 preparation, or by inhalation. Administration of an effective amount of compound or compound composition is contemplated to achieve the desired or stated effect. The amount of active ingredient that may be combined with the carrier materials to produce a single dosage form will vary depending upon the host treated and the particular mode of administration. A typical preparation will contain from about 5% to about 95% active
35 compound (w/w). Alternatively, such preparations contain from about 20% to about 80% active compound.

Lower or higher doses than those recited above may be required. Specific dosage and treatment regimens for any particular patient will depend upon a variety of factors, including

5 the activity of the specific compound employed, the age, body weight, general health status, sex, diet, time of administration, rate of excretion, drug combination, the severity and course of the disease, condition or symptoms, the patient's disposition to the disease, condition or symptoms, and the judgment of the treating physician.

The compositions delineated herein include the compounds delineated herein, as well as
10 additional therapeutic agents if present, in amounts effective for achieving a modulation of disease or disease symptoms. When the compositions of this invention comprise a combination of a compound of the formulae described herein and one or more additional therapeutic or prophylactic agents, both the compound and the additional agent should be present at dosage levels of between about 1 to 100%, and more preferably between about 5
15 to 95% of the dosage normally administered in a monotherapy regimen. The additional agents may be administered separately, as part of a multiple dose regimen, from the compounds of this invention. Alternatively, those agents may be part of a single dosage form, mixed together with the compounds of this invention in a single composition.

A composition according to the invention may include a "pharmaceutically acceptable
20 carrier or adjuvant", a carrier or adjuvant that may be administered to a patient, together with a compound of this invention, and which does not destroy the pharmacological activity thereof and is nontoxic when administered in doses sufficient to deliver a therapeutic amount of the compound.

The pharmaceutical compositions of this invention may be administered orally, parenterally,
25 by inhalation spray, topically, rectally, nasally, buccally, vaginally or via an implanted reservoir, preferably by oral administration or administration by injection. The pharmaceutical compositions of this invention may contain any conventional non-toxic pharmaceutically-acceptable carriers, adjuvants or vehicles. In some cases, the pH of the formulation may be adjusted with pharmaceutically acceptable acids, bases or buffers to
30 enhance the stability of the formulated compound or its delivery form. The term parenteral as used herein includes subcutaneous, intracutaneous, intravenous, intramuscular, intra-articular, intra-arterial, intrasynovial, intrasternal, intrathecal, intralesional and intracranial injection or infusion techniques.

The pharmaceutical compositions may be in the form of a sterile injectable preparation, for
35 example, as a sterile injectable aqueous or oleaginous suspension. This suspension may be formulated according to techniques known in the art using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation may also be a sterile

5 injectable solution or suspension in a non-toxic parenterally acceptable diluent or solvent. Fatty acids, such as oleic acid and its glyceride derivatives are useful in the preparation of injectables, as are natural pharmaceutically-acceptable oils. Other commonly used surfactants and/or emulsifying agents or bioavailability enhancers which are commonly used in the manufacture of pharmaceutically acceptable solid, liquid, or other dosage forms may
10 also be used for the purposes of formulation.

The pharmaceutical compositions of this invention may be orally administered in any orally acceptable dosage form including, but not limited to, capsules, tablets, emulsions and aqueous suspensions, dispersions and solutions. In the case of tablets for oral use, carriers are commonly used. Lubricating agents are also typically added. When aqueous
15 suspensions and/or emulsions are administered orally, the active ingredient may be suspended or dissolved in an oily phase is combined with emulsifying and/or suspending agents. If desired, certain sweetening and/or flavoring and/or coloring agents may be added.

The pharmaceutical compositions of this invention may also be administered in the form of suppositories for rectal administration. These compositions can be prepared by mixing a
20 compound of this invention with a suitable non-irritating excipient which is solid at room temperature but liquid at the rectal temperature and therefore will melt in the rectum to release the active components.

Topical administration of the pharmaceutical compositions of this invention is useful when the desired treatment involves areas or organs readily accessible by topical application. For
25 application topically to the skin, the pharmaceutical composition should be formulated with a suitable ointment containing the active components suspended or dissolved in a carrier. Alternatively, the pharmaceutical composition can be formulated with a suitable lotion or cream containing the active compound suspended or dissolved in a carrier with suitable emulsifying agents.

30 The pharmaceutical compositions of this invention may be administered by nasal aerosol or inhalation. Such compositions are prepared according to techniques well-known in the art of pharmaceutical formulation and may be prepared as solutions in saline, employing benzyl alcohol or other suitable preservatives, absorption promoters to enhance bioavailability, fluorocarbons, and/or other solubilizing or dispersing agents known in the art.

35 In this disclosure, "comprises," "comprising," "containing" and "having" and the like can have the meaning ascribed to them in U.S. Patent law and can mean "includes,"

5 "including," and the like; "consisting essentially of" or "consists essentially" likewise has the meaning ascribed in U.S. Patent law and the term is open-ended, allowing for the presence of more than that which is recited so long as basic or novel characteristics of that which is recited is not changed by the presence of more than that which is recited, but excludes prior art embodiments.

10 Additional definitions are provided in context throughout the disclosure.

In a particular embodiment of the invention, an expanded breast cancer sample cohort is compiled. In a particular embodiments of the invention, optimal tumor-lymph node matched cases are selected for complete VH, V λ and V κ library synthesis and analyzed from reactive B-cell germinal centers.

15 Tumor and sentinel node banking. Preliminary studies have established effective mechanisms for obtaining surgical samples in a manner that does not compromise patient diagnosis of tumor grade and metastasis. This close collaboration with the surgical, nursing and pathological staff has led to the maintenance of clinical specimens with respect to molecular integrity in the form of mRNA and protein without significant degradation. Of
20 note, an integrated system has been developed to obtain surgical samples that have been approved through patient consent and maintain patient confidentiality as defined by HIPAA regulations.

The amount of tumor samples obtained is variable and can range from a 1.5mm medial slice of a tumor less than 1cm in diameter to those exceeding 4 cm. Lymph nodes are also
25 provided as a medial section provided by trained pathologists. The medial 1-2mm section is sufficient to provide both histological and gross sample quantity to obtain library quality RNA from core-excised pieces as described in the Examples below. Clinical data is retained in a tumor acquisition database that is compliant with HIPAA privacy regulations, but matched in a de-identified manner to each sample, so that initial diagnosis, pre-surgical
30 treatments, and follow-up events are retained in the research database (institutional supported cancer tumor acquisition core).

IRB-approved protocols provide for tumor acquisition and mapped sentinel lymph nodes. The samples are obtained within 10 minutes of surgical resection to assure diagnostic pathology samples are retained to prioritize patient diagnosis. The non-essential materials
35 are provided for research purposes only for those patients that have consented to their use. Samples of tumor provided are immediately flash frozen in liquid nitrogen. Medial gross

- 5 sections of lymph nodes are embedded into cryosectioning media for immunohistological screening for B-cell and proliferation markers as indicated in the Examples below.

Banked materials are collected until a defined group of ten or more cases exist that have both a minimal tumor sample of at least 150-200mg which correlates to a medial slice of tumor tissue of 0.5x0.3cmx0.15cm, or approximately equal to a 1.5mm thick cross section
10 through the middle of an almond. Two criteria apply to sample selection for complete antibody variable region cDNA analysis: in the first, there is, ideally, sufficient quantity and quality of primary tumor available for the analysis and antigen-trap assays. Of note, tumor features such as excess tumor necrosis or inflammatory cell infiltrate will be declined for further analysis. Secondly, those cases with mapped sentinel lymph nodes that do not have
15 infiltrating tumor cells (indicative of increased immune response to tumor-elicited antigens as defined by our preliminary studies) will be positively selected, although metastatic nodes will not necessarily be excluded.

VH, V λ and V κ cDNA library synthesis and bioinformatic evaluation of sequences for clonally expanded groups and antibody-specific somatic hypermutation. Those cases where
20 lymph node activation is occurring in an antigen-dependent manner are primarily selected for complete analysis. Histological features evaluated include the B-cell identifiers, CD20 and CD23, proliferation via Ki-67, and the number of total activated germinal centers within each particular lymph node and in all nodes from each case.

For selected cases, the reactive germinal centers are analytically excised with a 0.6mm
25 tissue array core. Total RNA is isolated using a modification of the Qiagen RNeasy kit that uses tRNA as a carrier to facilitate recovery of minimal amounts of total RNA from the small tissue core obtained from a frozen piece of the lymph node. Random primed reverse transcription is performed with SuperScript II reverse transcriptase (Invitrogen), according to the manufacturer's instructions.

- 30 i) Variable region cDNA cloning, sequence and bioinformatics analysis. Human antibody variable domain cDNAs are amplified from total random primed libraries using degenerate oligonucleotide primers specific to human VH, V1 and V κ genes. The amplified fragments are then cloned into pCR-CT-TOPO (Invitrogen, Inc.) according to the manufacturer's instructions. Individual clones are then directly amplified using pCR-CT-TOPO-specific
35 primers and the V region cDNA clone fragments isolated as individuals. The fragments are then directly sequenced using the 5'-primer of the amplicon using modified terminal di-deoxy sequencing using fluorescent dye detection.

- 5 To interpret a large amount of cDNA sequence data for variable regions and identify those antibodies that have undergone clonal expansion and somatic hypermutation in response to antigen-induced B-cell responses, the first pass single orientation sequences of at least 192 clones are obtained from each VH, V λ or V κ PCR amplification and cloning, typically resulting in ~130 acceptable sequences for further analysis. Genetic alignments and
- 10 phylogenetic clustering are performed using MacVector7 Software (Accelrys, San Diego, CA). The number of individual V(D)J clonal subgroups and R:S ratios for FR and CDR regions are determined for each sequence using the VBASE antibody analysis site in DNAPLOT (<http://www.mrc-cpe.cam.ac.uk/vbase-ok.php?menu=901>). CDR/FR mutation ratios that exceed 0.3 are defined as antigen-driven processes. A more refined mathematical
- 15 test for R:S ratios exceeding gene-specific FR and CDR mutation rates is analyzed using the method and analysis program of Lossos (56).

Antibody variable domain sequences that have clones with significant R:S ratios and CDR/FR ratios of greater than 2X baseline at 0.6 are selected as antigen-matured variable domains. An identical analysis is performed for the V λ and V κ amplified libraries. The

20 primary utilization of either V λ or V κ isotypes is determined (similar to the selective use of V κ described in the Examples below) and the V-J germline clones identified. These sequences are selected for matrix synthesis of single chain variable fragment (scFv) proteins with the VH sequences from the same germinal center (SEQ ID NO: 2-9)

25

SEQ ID NO:3 Human VH1.1a

```

      10      20      30      40      50      60      70      80      90     100
110      120
ATGGCGCAGGTGCAGCTGGTGGAGTCTGGAGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGGTCTCCTGCACGGCTTCTGGTTACACCTTTATCAA
TTACGGTATCATCTGGGTGCGA
M A Q V Q L V E S G A E V K K P G A S V K V S C T A S G Y T F I N
Y G I I W V R>

      130     140     150     160     170     180     190     200     210     220
230     240
CAGGCCCTGGACAAGGACTTGAGTGGATGGGATGGATCAACGGTTACAGTGGTCACACAAGCTATGCACAGAAATTCAGGACAAAGTCACCATGAC
CACAGACACATCCACGAGCACA
Q A P G Q G L E W M G W I N G Y S G H T S Y A Q K F Q D K V T M T
T D T S T S T>

      250     260     270     280     290     300     310     320     330     340
350     360
GTCTACATGGAGCTGAGGAACCTGGGATCTGACGACACGGCCGTCTATTACTGTGCGAGAGGATCATCGGGGCGGGAACCTACTCCCGTTTGACCACTG
GGGCCAGGGAACCTGGTCACC
V Y M E L R N L G S D D T A V Y Y C A R G S S G R E L L P F D H W
G Q G T L V T>

      370
GTCTCCTCAAGATCT
V S S R S
```

SEQ ID NO:4 Human VH1.1a+Myc+His

```

      10      20      30      40      50      60      70      80      90     100
110      120
ATGGCGCAGGTGCAGCTGGTGGAGTCTGGAGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGGTCTCCTGCACGGCTTCTGGTTACACCTTTATCAATTACGGTATCATCTG
GGTGCGA
M A Q V Q L V E S G A E V K K P G A S V K V S C T A S G Y T F I N Y
G I I W V R>

      130     140     150     160     170     180     190     200     210     220
230     240
CAGGCCCTGGACAAGGACTTGAGTGGATGGGATGGATCAACGGTTACAGTGGTCACACAAGCTATGCACAGAAATTCAGGACAAAGTCACCATGACCACAGACATCCAC
GAGCACA
Q A P G Q G L E W M G W I N G Y S G H T S Y A Q K F Q D K V T M T T
D T S T S T>

      250     260     270     280     290     300     310     320     330     340
350     360
GTCTACATGGAGCTGAGGAACCTGGGATCTGACGACACGGCCGTCTATTACTGTGCGAGAGGATCATCGGGGCGGGAACCTACTCCCGTTTGACCACTGGGGCCAGGGAACCT
GGTCACC
V Y M E L R N L G S D D T A V Y Y C A R G S S G R E L L P F D H W G
Q G T L V T>

      370     380     390     400     410     420     430     440     450
460     470
GTCTCCTCAAGATCTGCAGCTGGTACCATATGGGAATTCGAAGCTTACGTAGAACAAAACCTCATCTCAGAAGAGGATCTGAATAGCGCCGTCGACCATCATCATCATCA
TTGA
V S S R S A A G T I W E F E A Y V E Q K L I S E E D L N S A V D H H
H H H H *
```

SEQ ID NO:5 Human VH1.1b

```

110      120      10      20      30      40      50      60      70      80      90      100
ATGCAGGTGCAGCTGGTGGAGTCTGGAGCTGAGGTGAAGAAGCCTGGGSCCTCAGTGAAGGTCTCCTGCACGGCTTCTGGTTACACCTTTATCAATTACGGTATCATCTGGGT
GCGACAG
      M Q V Q L V E S G A E V K K P G A S V K V S C T A S G Y T F I N Y G
I I W V R Q>

230      240      130      140      150      160      170      180      190      200      210      220
GCCCCTGGACAAGGACTTGAGTGGATGGGATGGATCAACGGTTACAGTGGTCACACAAGCTATGCACAGAAATTCAGGACAAAGTCACCATGACCACAGACACATCCACGAG
CACAGTC
      A P G Q G L E W M G W I N G Y S G H T S Y A Q K F Q D K V T M T T D
T S T S T V>

350      360      250      260      270      280      290      300      310      320      330      340
TACATGGAGCTGAGGAACCTGGGATCTGACGACACGGCCGCTCTATTACTGTGCGAGAGGATCATCGGGCGGGAACTACTCCCCTTTGACCACTGGGGCCAGGGAACCATGGT
CACCGTC
      Y M E L R N L G S D D T A V Y Y C A R G S S G R E L L P F D H W G Q
G T M V T V>

      TCCTCA
      S S>

```

SEQ ID NO:6 Human VH1.1b+V5+His

```

110      120      10      20      30      40      50      60      70      80      90      100
ATGCAGGTGCAGCTGGTGGAGTCTGGAGCTGAGGTGAAGAAGCCTGGGSCCTCAGTGAAGGTCTCCTGCACGGCTTCTGGTTACACCTTTATCAATTACGGTATCATCTGGGT
GCGACAG
      M Q V Q L V E S G A E V K K P G A S V K V S C T A S G Y T F I N Y G
I I W V R Q>

230      240      130      140      150      160      170      180      190      200      210      220
GCCCCTGGACAAGGACTTGAGTGGATGGGATGGATCAACGGTTACAGTGGTCACACAAGCTATGCACAGAAATTCAGGACAAAGTCACCATGACCACAGACACATCCACGAG
CACAGTC
      A P G Q G L E W M G W I N G Y S G H T S Y A Q K F Q D K V T M T T D
T S T S T V>

350      360      250      260      270      280      290      300      310      320      330      340
TACATGGAGCTGAGGAACCTGGGATCTGACGACACGGCCGCTCTATTACTGTGCGAGAGGATCATCGGGCGGGAACTACTCCCCTTTGACCACTGGGGCCAGGGAACCATGGT
CACCGTC
      Y M E L R N L G S D D T A V Y Y C A R G S S G R E L L P F D H W G Q
G T M V T V>

      370      380      390      400      410      420      430      440      450
      TCCTCAAAGGGCAATTCGAAGCTTGAAGGTAAAGCCATCCCTAACCCCTCTCCTCGGTCTCGATTCTACGGGTACCGGTATCATCACCATCACCATTGA
      S S K G N S K L E G K P I F N P L L G L D S T R T G H H H H H H *

```

SEQ ID NO:7 Human VH1.1c

```

110      120      10      20      30      40      50      60      70      80      90      100
ATGCAGGTGCAGCTGGTGGAGTCTGGAGCTGAGGTGAAGAAGCCTGGGSCCTCAGTGAAGGTCTCCTGCACGGCTTCTGGTTACACCTTTATCAATTACGGTATCATCTGGGT
GCGACAG
      M Q V Q L V E S G A E V K K P G A S V K V S C T A S G Y T F I N Y G
I I W V R Q>

230      240      130      140      150      160      170      180      190      200      210      220
GCCCCTGGACAAGGACTTGAGTGGATGGGATGGATCAACGGTTACAGTGGTCACACAAGCTATGCACAGAAATTCAGGACAAAGTCACCATGACCACAGACACATCCACGAG
CACAGTC
      A P G Q G L E W M G W I N G Y S G H T S Y A Q K F Q D K V T M T T D
T S T S T V>

350      360      250      260      270      280      290      300      310      320      330      340
TACATGGAGCTGAGGAACCTGGGATCTGACGACACGGCCGCTCTATTACTGTGCGAGAGGATCATCGGGCGGGAACTACTCCCCTTTGACCACTGGGGCCAGGGAACCCCTGGT
CACCGTC
      Y M E L R N L G S D D T A V Y Y C A R G S S G R E L L P F D H W G Q
G T L V T V>

      TCCTCA

```

S S>

SEQ ID NO:8 Human VH1.1c+V5+His

```

      10      20      30      40      50      60      70      80      90     100
110      120
ATGCAGGTGCAGCTGGTGGAGTCTGGAGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGGTCCTCTGCACGGCTTCTGGTTACACCTTTATCAATTACGGTATCATCTGGGT
GCGACAG
      M Q V Q L V E S G A E V K K P G A S V K V S C T A S G Y T F I N Y G
I I W V R Q>
      130      140      150      160      170      180      190      200      210      220
230      240
GCCCCTGGACAAGGACTTGGAGTGGGATGGATCAACGGTTACAGTGGTCACACAAGCTATGCACAGAAATTCAGGACAAAGTCACCATGACCACAGACACATCCACGAG
CACAGTC
      A P G Q G L E W M G W I N G Y S G H T S Y A Q K F Q D K V T M T T D
T S T S T V>
      250      260      270      280      290      300      310      320      330      340
350      360
TACATGGAGCTGAGGAACCTGGGATCTGACGACACGGCGTCTATTACTGTGCGAGAGGATCATCGGGGCGGGAACTACTCCCGTTTGACCACTGGGGCCAGGGAAACCTGGT
CACCGTC
      Y M E L R N L G S D D T A V Y Y C A R G S S G R E L L P F D H W G Q
G T L V T V>
      370      380      390      400      410      420      430      440      450
TCCTCAAAGGGCAATTGGAAGCTTGAAGGTAAGCCTATCCCTAACCCCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATTGA
S S K G N S K L E G K P I P N P L L G L D S T R T G H H H H H H *
```

SEQ ID NO:9 Human VH2.1

```

      10      20      30      40      50      60      70      80      90     100
110      120
ATGCAGGTGCAGCTGGTGGAGTCTGGAGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGGTCCTCTGCAAGGCTTCTGGTTACACCTTTACCAGCTACGGTATCACTTGGGT
GCGACAG
      M Q V Q L V E S G A E V K K P G A S V R V S C K A S G Y T F T S Y G
I T W V R Q
      130      140      150      160      170      180      190      200      210      220
230      240
GCCCCTGGACAAAACTTGNNTGGCTGGGATGGATCAGCGCTTACAATGGTAACACACATTATGCACAGAAGCTCCAGGGCAGAGTCACCATGACCACAGACACATCCACGAC
CACAGCC
      A P G Q K L X W L G W I S A Y N G N T H Y A Q K L Q G R V T M T T D
T S T T T A
      250      260      270      280      290      300      310      320      330      340
350      360
TACATGGAGCTTAGGAGCCTGAGATCTGACGACACGGCCGTGTATTACTGTGCGAGGGTCGAGGCTCGCTTCGACCTCTTTGACTACTGGGGCCAGGGAACCATGGTCACCGT
CTCCTCA
      Y M E L R S L R S D D T A V Y Y C A R V E A A F D L F D Y W G Q G T
M V T V S S
```

SEQ ID NO:10 Human VH2.1+V5+His

```

      10      20      30      40      50      60      70      80      90     100
110      120
ATGCAGGTGCAGCTGGTGGAGTCTGGAGCTGAGGTGAAGAAGCCTGGGGCCTCAGTGAAGGTCCTCTGCAAGGCTTCTGGTTACACCTTTACCAGCTACGGTATCACTTGGGT
GCGACAG
      M Q V Q L V E S G A E V K K P G A S V R V S C K A S G Y T F T S Y G
I T W V R Q
      130      140      150      160      170      180      190      200      210      220
230      240
GCCCCTGGACAAAACTTGNNTGGCTGGGATGGATCAGCGCTTACAATGGTAACACACATTATGCACAGAAGCTCCAGGGCAGAGTCACCATGACCACAGACACATCCACGAC
CACAGCC
      A P G Q K L X W L G W I S A Y N G N T H Y A Q K L Q G R V T M T T D
T S T T T A
      250      260      270      280      290      300      310      320      330      340
350      360
TACATGGAGCTTAGGAGCCTGAGATCTGACGACACGGCCGTGTATTACTGTGCGAGGGTCGAGGCTCGCTTCGACCTCTTTGACTACTGGGGCCAGGGAACCATGGTCACCGT
CTCCTCA
      Y M E L R S L R S D D T A V Y Y C A R V E A A F D L F D Y W G Q G T
M V T V S S
      370      380      390      400      410      420      430      440      450
```

AAGGGCAATTCGAAGCTTGAAGGTAGCCTATCCCTAACCCCTCTCCTCGGTCTCGATTCTACGCGTACCGGTCATCATCACCATCACCATTGA
K G N S K L E G K P I P N P L L G L D S T R T G H H H H H H *

5 These selection processes define a group of approximately 4 VH sequences and combinations with possibly 5 V λ or V κ light chains. Thus, for each selected lymph node, 20-40 VH-V λ / κ scFv combinations are expected.

10 ii) Production and purification of human variable region chimeric proteins in prokaryotic cells. To provide VH antibody chimeric proteins for use in diagnostic and antigen identification assays rapid and abundant protein production methods are required. One embodiment of this invention utilizes the production of human variable region chimeric proteins in *E. coli* bacteria. Individual VH cDNA clones selected from the germinal center libraries are amplified by expanding the pCR-CT-TOPO plasmid containing the individual VH clone in *E. coli* (XL-1 blue, Stratagene, Inc.) under ampicillin resistance and
15 purification of the plasmid using a NucleoSpin plasmid purification kit (Machery-Nagel, Inc.)

The plasmids contain a T7 RNA polymerase promoter, a bacterial translation ribosome binding site as well as C-terminal V5 epitope and 6X-histidine protein tags. The individual plasmids containing variable region clones are transfected into BL21 (DE3) pLysS *E. coli*
20 strain. The bacteria are selected in resistance media (ampicillin) and induced to express T7 RNA polymerase with IPTG. The bacteria are incubated to allow induction of RNA polymerase activity on the pCR-T7-CT-TOPO vector to produce high levels of recombinant variable region chimeric protein. The protein is then extracted from the bacteria by chemical lysis and purification using metal affinity chromatography selective for the 6X-
25 histidine tag on the C-terminal domain of the recombinant protein. The protein production levels are analyzed by direct SDS-PAGE and coomassie staining for polypeptide of an expected size. Confirmation of V5 and 6X-histidine tags on the c-terminus are defined by direct immunoblotting using anti-V5 and anti-6X-histidine specific antibodies.

30 iii) Tumor antigen identification using antigen-trap assays. The purified antigen-binding VH or VH-V λ /V κ scFv combination proteins are used to antigen trap on high capacity Ni-affinity plates as described in the Examples below. Controls include no recombinant protein, VH, the same VH as a scFv with V λ /V κ combinations as well as a non-somatic hypermutated V(D)J germline control, all from the same germinal center library(ies). The high sensitivity Finnigan LTQ mass spectrometer allows screening for tumor antigens via
35 VH or scFv trapping using only 50-200ug of primary tumor extracts. Alternatively, immunoblots are performed on accumulated breast cancer extracts to test VH domain or scFv protein antigen specificities. Bioinformatic analysis of peptide identification is

5 somewhat easier with such assays, since multiple peptides of the target antigen are expected to be detected in the LC-MS/MS analysis. If multiple antigens are indicated by proteomic analysis, the more direct assays such as immunoblot and IP-Westerns with known antibodies should determine antigen specificity, if needed.

iv) Characterization of identified antigen distribution, expression level, and possible
10 mutations in breast cancer. Confirmation of any identified antigen expression and association with breast cancer progression is screened in three ways: 1) RT-PCR(q) analysis in primary breast cancer and human breast cancer cell line samples to evaluate mRNA expression level and possible mutations in primary tumor (sequencing PCR products) as described in the Examples below (64, 65); 2) analysis of cDNA expression in normal versus
15 matched tumor samples for breast, lung, liver, colon and ovarian cancers using the cDNA Cancer Array II (Clontech); 3) definition of distribution *in situ* using breast cancer tissue arrays (66, 67). The antigen expression level is evaluated for associations with tumor size, grade, estrogen receptor/progesterone receptor status, aromatase expression, HER-2/neu, regional metastatic status and long-term outcome, if available. Appropriate and valid
20 statistical tests are performed for correlations between each variable.

In one embodiment of the invention, recombinant antibody molecules are produced in mammalian expression systems. In particular embodiments of the invention, antigen binding specificity and affinity are tested. To create a mammalian system of protein synthesis that can be integrated into the other pre-existing modules, *Pichia pastoris* yeast expression
25 systems are developed that will facilitate recombinant protein expression at the level and structural integrity needed for antigen screening applications. The *Pichia* expression system is tested by expression of the SDR-1/NP antigen protein defined by a VH antigen-trap method. In addition, the VH Ab-1 recombinant protein and a V(D)J-germline matched control without somatic hypermutation are synthesized alone and in combination with
30 variable light chains to evaluate their influence on VH binding specificity and affinity.

Pichia expression system for antigen production. SDR-1/NP cDNAs are incorporated into *Pichia* methanol-induced protein secretion expression vectors. The pPICZ α -E expression vector (Invitrogen) is used to synthesize SDR-1/NP, VH, and scFv fusions. The pPICZ α -E plasmid fuses an open reading frame protein sequence C-terminal to the yeast alpha-factor secretion sequence. Upon secretion of the fusion gene product, the α factor secretion
35 sequence is cleaved from the rest of the expressed protein, yielding only the open reading

- 5 frame inserted. This system has been demonstrated to be effective for antibody domain and full antibody production (68, 69).

The cloned VH and scFv open reading frames are generated with proof reading polymerase chain reaction from their respective pCR T7 clones (the vector of primary library synthesis), such that both the 6X -His and V5 epitope tags are retained. For generating an scFv clone,
 10 the addition of a V λ or V κ chain is subsequently performed using a double-fragment insertion C-terminal to the VH sequence and in frame with the V5 and 6X-His fusion tags. The scFv requires a (gly₄-pro)₃ flexible linker to juxtapose the heavy and light chains for functional binding. All expression vectors are linearized and electroporated into several yeast *Pichia* strains according to the manufacturer's protocols to determine those strains that
 15 most effectively express secreted protein. Optimal transfection, selection, and induction procedures necessary for optimal protein production are developed. The secreted form of alpha SDR-1/NP (NM_017455)(C-terminal deletion which removes the TM domain), as well as subsequent C-terminal deletions that aid in mapping epitope binding domains for the VH and scFv antibody fragments, are expressed.

- 20 SDR-1/NP serves as a model for the *Pichia* expression system and as a positive control for VH and scFv binding specificity assays. While VH and scFv clones contain both 6X His and V5 epitope tags, the recombinant SDR-1/NP proteins contain only the 6X-His required for purification. VH candidates are chosen from ~200 sequenced VH clones for any given germinal center as described above. Clonal V(D)J subclones are estimated between 5-10
 25 per germinal center from the 200 sequenced. From these, the VH clones that demonstrate antigenic-driven somatic hypermutation with respect R:S and CDR/FR ($> 0.6 = 2 \times 0.3$ threshold for SHM) ratios (~4) are selected. A panel of light chains that are similarly clonally selected from the same germinal center to use for scFv construction is also chosen.

i) Methanol induction of protein synthesis. Open reading frame sequences cloned into
 30 pPICZ α -E are under the transcriptional control of the methanol-inducible promoter AOX1. Without induction, there is no expression from the plasmid. Upon induction with 0.5% methanol, high levels of gene expression (up to 30% of total protein) are driven off the AOX-1 promoter in the pPICZ α -E vector. Initially, timepoints from 12h to 5 days post-induction are screened for the optimal induction period and accumulation of secreted
 35 protein.

ii) Maximizing protein production. At the optimal induction time-point, 6-10 *Pichia* transformants are screened on a small scale for the optimal expressing strain. Since the

5 cloned ORFs are fused to α -factor secretion sequence, proteins are harvested from cell supernatants (which have little contaminating secreted proteins and do not have proteinaceous media). Protein expression levels and detection are accomplished via anti-V5 epitope immunoblot (Vh and VH-V λ / κ proteins) and/or Coomassie blue staining of SDS-PAGE gels where necessary (where SDR-1/NP does not contain V5 epitope). Once optimal
10 expression conditions and clones are identified, those strains and conditions are utilized in large-scale production up to 1 liter expected to yield 1-2 g of protein.

iii) Recombinant protein purification. Cellular supernatants are prepared from Pichia cells via centrifugation. Clarified supernatants are applied to Talon Metal Affinity Resin (Clontech), a specific affinity resin for proteins containing 6X-His tags while the rest of the
15 growth medium is separated as flow-thru. The bound protein is eluted from the Talon column as per the manufacturer's protocol, using increasing imidazole concentrations and dialyzation (as described in the Examples below). FPLC application of these procedures maintains a uniform and improve final product quality and purification (Pharmacia/AKTA system).

20 iv) Evaluation of recombinant proteins and quality control. Purified proteins are verified for yield by running samples on SDS-PAGE gels and staining with Coomassie blue dye. SDR-1/NP is subjected to glycosidases and resolved on gels to compare with untreated SDR-1/NP. Samples are also run on gels with and without α -ME to determine if reducing agent has an effect on its protein structure, indicating the absence of inter-protein disulphide
25 linkages, as opposed to intra-protein bridges. In addition, VH and scFv fusion proteins are confirmed for appropriate size by running each sample on an SDS-PAGE gel, transferring the DNA to nitrocellulose membrane, immunoblotting with the anti-V5 primary antibody conjugate, and detection using standard HRP-enzyme activity visualized through chemiluminescent reagent (KPL, ECL kit).

30 Expression of VH recombinant proteins and testing binding capacity. To express the VH Ab-1 and a V(D)J matched non-hypermutated control in Pichia and compare binding specificity and affinity using the SDR-1/NP protein and tumor extracts as test substrates, the anti-SDR-1/NP VH domain Ab-1 and a germline matched control using the Pichia expression system (PIC-EZ, Invitrogen) are first synthesized. For binding purposes,
35 recombinant SDR-1/NP is subjected to glycosidase and compared with fully glycosylated SDR-1/NP to determine if this affects VH or scFv binding (i.e. define a protein-specific epitope as opposed to sugar antigen). VH binding is epitope-mapped to SDR-1/NP by

5 generating a series of SDR-1/NP recombinant proteins with serial C-terminal deletions in Pichia. Binding assays are performed with the SDR-1/NP isoforms as substrates in immunoblots, direct and sandwich ELISAs, and IP assays using the VH and scFv antibody fragments.

10 i) Direct ELISA. The recombinant SDR1/NP protein and its C-terminal deletions are bound to protein-binding ELISA plates overnight at 4°C. The unbound material is removed via several washes of PBS-Tween binding buffer. Purified VH is then titrated into the SDR1-containing wells and allowed to bind to antigen for 1-2h at room temperature. Wells are again washed of unbound material. A secondary anti-V5-HRP conjugate is added to associate with any bound VH within each well (this is selective for the VH or scFv proteins, since the SDR-1 antigens do not have V5 epitope). Finally, TSA-linked fluorescence of an HRP substrate converts it to a nucleophilic reactive fluorescent molecule that will attack free amino sites on protein in the well and result in correlative fluorescence (70, 71). The cells are washed again, and fluorescence is quantified using the BioRad cytofluor as described in the Examples below.

20 Upon determination of an optimal concentration for VH or scFv detection of SDR-1/NP, a serial titration of SDR-1/NP is performed in a 96-well format to determine the detection limit for each of the recombinant proteins synthesized, a direct molar comparison of binding affinity. SDR-1/NP antigens are in their native, glycosylated state (from Pichia expression) or following glycosidase treatment (Calbiochem Glycoprotein Deglycosylation Kit) to determine the potential requirement, or not, of sugar residues in epitope specificity. Non-specific binding for the C-terminal 6X-His tag is reduced using a competitive concentration of imidazole at 5mM.

30 ii) Immunoprecipitation and antigen-trap. Immunoprecipitation assays are performed to evaluate the efficiency and specificity of purified VH or scFv to bind antigen. Purified SDR-1/NP are added in increasing amounts into MDA-MB-231 human breast cancer cell extract, or tumor extracts are tested directly. Secondary anti-V5 epitope antibodies and Protein G-agarose are serially bound to precipitate VH or scFv bound protein. Washed immunoprecipitates are resolved on SDS-PAGE gels and either 1) directly silver-stained or 2) transferred to nitrocellulose membranes and immunoblotted with anti-SDR-1/NP polyclonal antibody. Alternatively, purified VH or scFv are bound to high capacity Ni-affinity plates (Sigma His-Tag plates). Cell extracts are spiked with increasing amounts of SDR-1/NP as a standard addition control and introduced to the well-bound antibody

- 5 fragments. Following binding and extensive washing, protein within the wells are trypsin-cleaved and prepared for LC-MS/MS and analyzed as described above.
- iii) Sandwich ELISA. ELISA plates are employed to bind anti-SDR-1/NP polyclonal as a capture antibody. Purified antigen are added alone or in a mixture with breast cancer cell extracts and allowed to bind. Unbound material is washed out of the wells. VH or scFv
- 10 proteins are then added, and a secondary anti-V5 detector antibody is used with TSA-fluorescent amplification. The resulting fluorescence is quantified using a multi-well fluorimeter. Comparison of binding curves for titrated antigen (PBS vs. cell extract) indicates the effectiveness of the antibody fragments to detect the same antigen in a test diagnostic scenario.
- 15 Combine VH with V light chains and compare to VH alone. The VH Ab-1 clone identified is combined with selected light chains (either lambda or kappa, as defined by library clone sequence analysis) in a matrix fashion for expression in Pichia. The specificity of these scFv proteins is evaluated in comparison to the VH alone in antigen specificity and affinity assays.
- 20 It is determined whether or not maximal antigen specificity has been attained with a VH region alone compared to light chain scFv combinations. By adding light chains expanded in the same germinal center, the potential for increasing binding efficiency or specificity for antigen is determined. Any improvement in antigen binding is useful for applications in the primary tumor extract antigen trap assay. If the combined VH-V λ / κ scFv recombinants
- 25 prove to have higher specificity and affinity, these proteins are synthesized prior to subsequent antigen trap assays. To optimize potential diagnostic antigen profiling panels such as multiplex direct ELISA assays, V λ / κ light chains may be added to the VH domains to reduce overall background and increase signal-to-noise ratios.
- In one embodiment of the invention, analytical assays incorporating the recombinant
- 30 antibody molecules are refined to effectively screen primary cancer, histological material, and biological fluids to evaluate their potential as diagnostic biomarkers. In an effort to evaluate the distribution of any potential biomarker antigen in primary cancers or common testing samples such as urine or serum, methods that increase the sensitivity of detection of any particular antigen are evaluated, and unique formats applicable to high-throughput
- 35 testing for breast cancer antigen biomarkers are defined.

5 In a particular embodiment of the invention, high affinity/specificity VH or scFv recombinant proteins are utilized for rapid histological screening using high sensitivity quantum dot fluorescent detection. Clinical pathological analysis of primary tumor resection requires concurrent histopathological evaluation of frozen margin sections to assure tumor tissue is within the surgical borders. These in-surgery frozen blocks can be used for antigen
10 profiling by obtaining additional sections from patients at the time of tumor and lymph node harvest and banking. These slides can be preserved and rapidly analyzed for antigen presence by using a simplified immunofluorescence assay using the VH or scFv recombinant proteins determined to be selective for their respective antigens. The presence of a specific antigen can be determined quite rapidly in a qualitative manner. Application of
15 this type of screen is employed primarily for diagnostic antigens that distinguish specific tumor types or have linkage to long-term prognosis.

VH and scFv detection in standard immunofluorescence assay is amplified using the anti-V5 epitope antibody (V5 epitope is present in the C-terminal domains of the expressed VH and scFv recombinant proteins), which will be covalently linked to non-quenching fluorescent
20 quantum dots (qDot) using available conjugation kits (Quantumdot, Inc.). The binding efficiency for each batch of the conjugated anti-V5 antibody is evaluated using a direct ELISA and cytofluorometric quantification, since the conjugation occurs through cysteine covalent bonds that may adversely affect binding capacity. Frozen tissue sections (cryostat cut at 10 μ m) are re-hydrated and blocked in BSA blocking solution.

25 Primary VH or scFv are allowed to bind for 1h at RT. The slides are washed 3X and subsequently incubated with the qDot-labeled anti-V5 antibody, as well as a control epithelial membrane antigen antibody labeled with alternative emission qDot (i.e. different color) for localization of epithelial cells for 15 mins, or incubation times determined to suffice for specific binding. Appropriate washes are performed, and the slide is imaged
30 under a Zeiss Axiophot fluorescent microscope using a double-emission filter to determine expression of the antigen, or not, and the epithelial marker. Images are captured using a digital camera (AxioCam, Zeiss) and analyzed for co-expression using image analysis software (ImagePro Plus, Media Cybernetics). Antigen expression is correlated to pathological diagnosis, metastasis to local lymph nodes, ER/PR, and HER-2/neu status and
35 tumor size. Statistical analyses are performed as appropriate.

In another particular embodiment of the invention, the surface plasmon resonance is evaluated in microarray format as a method to screen biological samples for antigen binding

- 5 to multiple VH or scFv recombinant proteins. Surface plasmon resonance (SPR) is a unique application of physical characteristics of gold film and light absorption properties (72). Matrices attached to one side of a gold film within a microfluid chamber can bind proteins and establish a surface-dependent light reflectance on the opposite side. This reflective property is detected and measured in real time.
- 10 Alteration of the matrix interaction with the gold film in the sample chamber can be achieved by protein-protein interactions and, thus, antibody-antigen binding can be detected in real time and with calculated affinities (73). Biosensor hardware incorporating SPR technology has been developed, so that VH or scFv proteins can be multiplexed onto an array slide, and multiple concurrent interactions of proteins in the sample and their binding
- 15 to each can be quantified (ArrayIt). Employing the Pichia expression system supplies a sufficient quantity and purity of recombinant proteins for placement into such multiplexed arrays.

- VH and/or scFv proteins characterized as specific and high affinity to their respective antigens are purified from Pichia expression using scale-up procedures. After quality control
- 20 for purity, disulphide linkages, and post-translational modifications, the purified proteins are aliquoted into bar-coded, non-protein-binding storage plates and stored at -80°C . These plates are used in a microarray robotic spotter (Microgrid II microarray spotter, Apogent-BioRobotics, Inc) to be placed onto SPR array slides for use in the ArrayIt SPR microarrayer. Sample evaluation is tested with breast cancer cell extracts provided with
- 25 standard addition of a test antigen (SDR-1/NP) to determine the relative sensitivity for each spot to detect a select antigen and at what protein concentration.

- Development of these analyses using the matched antigen for each VH or scFv protein confirms select detection limits and/or cross-reactivity and background levels. Subsequently, total tumor lysates will be evaluated for SPR via antigen-trap assays. Diluted
- 30 patient serum and processed urine are screened, where total proteins are precipitated with acetone and resuspended in volumes applicable to the microfluidics over extended flow times to accrue sufficient sample screening to pick up trace peptides that might be elicited from the tumors.

- In another particular embodiment of the invention, the possible utility of peptide
- 35 identification from large volume urine samples for VH or scFv-specific antigens through peptide trap, as well as of detection/identification using automated throughput LC-MS/MS, is determined. The application of matrix-assisted laser desorption/ionization time-of-flight

- 5 mass spectrometry to identify peptides and match to their protein through multiple matches and mass probability to databases is a powerful tool for biological discovery. To screen for low abundance breast cancer antigens as systemic biomarkers, application of a modified antigen-trap assay and identification of peptide levels using LC-MS/MS will be evaluated for large volume urine samples, employing an autosample quadrapole TOF mass
10 spectrometer (Finnigan LTQ Mass Spectrometer) capable of high throughput proteomic screening.

- Low abundance antigen peptides may be elicited from tumor microenvironments as protease-digested peptides contained in secreted urine. To improve the ability to detect trace levels of tumor-elicited antigens as clinical biomarkers, urine is relied upon as a cumulative
15 source of minor peptides potentially released from primary or metastatic tumors. In the first, patients that have been positively diagnosed with breast cancer through aspirate biopsy and are awaiting surgical resection are tested. Enrollment proceeds through informed consent to obtain multiple 24h urine output samples. These large volume urine samples are reduced for peptide/protein using a large volume acetone precipitation and desalting procedure. The
20 concentrated peptide/protein pool is subjected to antigen-trap array plates that have multiple VH or scFv proteins bound to Ni-affinity plates. Unbound proteins and peptides are washed with high stringent salt conditions.

- Preparation for LC-MS/MS analysis is performed with complete digestion of bound proteins with trypsin. Microplate *in situ* drying and resuspension in column solvent is performed and
25 transferred to a sample feed plate for automated sampling for LC-MS/MS to detect antigen-specific peptides. The specific detection of tumor-elicited antigen profiles dictated by the VH or scFv proteins obtained previously is extended to post-surgical patients undergoing chemotherapy and thereafter to determine if secondary metastasis can be detected as an early diagnostic prior to symptomatic presentation.

- 30 In one embodiment of the invention, tumors such as melanoma and ovarian cancers are evaluated for specific antigens.

EXAMPLES

All examples are provided as illustrations of the invention and are in no way to be considered limiting.

5 They are presented in a format to describe a complete process that has been developed to
evaluate the anti-tumor immune reactions present in regional lymph nodes of breast cancer
patients. Included are histological analysis of frozen lymph nodes, selection of activated B-
cell germinal centers, molecular synthesis of antibody variable region cDNA libraries, VH
clone sequence analysis, recombinant protein synthesis, antigen-trap identification from
10 primary tumor extracts, tumor extract screening and evaluation, and verification of antigen
profiles in breast cancer using immunological, biochemical and molecular methods.

Example 1. Analysis of tumor-draining lymph node B-cell germinal center activation status.

A. Breast tumor-draining lymph nodes have diverse histological structures representing
ranges from germinal center hyperplasia (GCH) to nodes with sinus histiocytotic appearance
15 (FIG. 1A). Since primary B-cell activation and germinal center formation requires T-cell
reaction to tumor-elicited antigens presented by antigen-presenting cells, recently activated
B-cell germinal centers might selectively represent anti-tumor antigen antibody maturation
sites. Since tumors and lymph node environments and cytokines can significantly repress T-
cell activation events, the relative number of humorally stimulated B-cell germinal centers
20 in tumor draining lymph nodes may be highly variable.

B. To address whether germinal centers are activated upon surgical resection, a multi-target
immunohistological analysis was applied to discern B-cell types (CD20+) combined with an
activation marker (CD23+), the FcεRII immunoglobulin-E (IgE) receptor (36, 54, 55), and
cellular proliferation (Ki-67). FIG. 2 shows how CD20 defines GC and the selective subset
25 of those that are proliferative.

C. Breast cancer sentinel lymph nodes were serially sectioned and stained with all three B-
cell germinal center markers (CD20, CD23, and Ki67). Samples of primary breast tumor
(>1.5cm diameter) and medial sections of local draining dye/radiolabeled-mapped lymph
nodes were obtained within 5-10 min of resection by a staff pathologist. The samples
30 were placed into optical cutting temperature (OCT) media (Shandon Lipshaw) and
immediately frozen in liquid nitrogen. All samples were devoid of any personal
identification information, tracking marker numbers or other information relating to the
patient.

Primary samples of tumor and lymph node were cryosectioned and stained with
35 hematoxylin and eosin to characterize general architectural features, the type and grade of the
tumor lesion and the distribution of germinal center or follicular hyperplasia features

- 5 within the tumor-draining lymph nodes. Immunohistochemistry was performed on post fixed cryosections (7-10 μ m in thickness) using antibodies to CD20 (Dako, Carpinteria, CA), CD23 (BD Pharmingen, San Jose, CA) and Ki67 (Dako) according to manufacturer's protocols. Controls were performed for all immunostaining without addition of primary antibody to assure signal specificity.
- 10 Immunostained sections were imaged with a color digital camera (Zeiss AxioCam HRc) and individual serial sections stained with the different antibodies and the sections registered for overlapping signals using Adobe Photoshop (Adobe Systems, Inc., San Jose, CA) and Image-Pro Plus (Media Cybernetics, Inc., Silver Spring, MD) software. Selected areas of interest were aligned with the H&E stained section and marked with a
- 15 marker pen. The slide was then aligned with the coordinate OCT tissue block and the area of interest cored out of the block using a cold (-20°C) 0.6mm copper core tube sterilized with detergent (0.5% SDS) and 70% ethanol.
- Large volume cores were obtained using a 5 mm dermal biopsy punch. The core depth ranged from 1-1.5mm and the whole core was transferred into RNA Lysis Buffer
- 20 containing 1% b-mercaptoethanol (Qiagen RNAeasy Kit). Overlap of staining (or lack thereof) was noted in a histological profile (FIG. 3). Histological profiling of 32 archived breast cancer regional lymph nodes from both metastatic and non-metastatic samples was performed with the same markers.

5

IH Marker	Total Average + GC/node (n=32)	Non-Metastatic Internodal Average + GC/node (n=23)	Metastatic (cytokeratin +) Internodal Average + GC/node (n=9)
CD20	42.5+/-19.6	44.2+/-22.4	36.7+/-16.2
CD20/CD23	11.4+/-12.8	7.9+/-3.3	6.4+/-2.1
CD20/CD23/Ki67	1.13+/-1.4	1.78+/-1.24 max=6	0.48+/-1.2 max=3 * p=0.015

* = statistically significant compared to non-metastatic nodes.

Table 1. Histological results of regional lymph node analysis for CD20, CD23 and Ki-67 as total and intranodal matched analyses.

This regional lymph node analysis revealed that the level of CD20/CD23 and Ki67 positive GCs was dramatically reduced in tumors that were identified as metastatic using a pan-cytokeratin positive staining within the lymph node, indicating infiltration of epithelial tumor cells. Thus, those lymph nodes with increased tumor burden showed dramatic reductions in B-cell responses that should repress antibody production and humoral activity.

Example 2. Immune cytokine profiles in total tumor-draining lymph nodes and germinal center microenvironments.

An analysis was made to determine if a non-metastatic breast cancer sentinel lymph node possessed cytokine expression profiles characteristic of either a cellular (TH1) or humoral (TH2) immune response at the time of resection. Messenger RNA expression of respective cytokines, characteristic of either a TH1 or TH2 response, was evaluated from core RNA samples with qualitative RT-PCR analysis. Random-hexamer primed RT reactions were amplified via polymerase chain reaction (PCR) using the following oligonucleotide primers: GAPDH forward primer: 5'-CACCCATGGCAAATTCCATG-3' (SEQ ID NO:11); GAPDH reverse primer: 5'-GCTTGACAAAGTGGTCGTTG-3' (SEQ ID NO:12); IL-4 forward primer: 5'-GGACACAAGTGCGATATCACC-3' (SEQ ID NO:13); IL-4 reverse primer: 5'-ATTTCTCTCTCATGATCGTC-3' (SEQ ID NO:14); IL-10 forward primer: 5'-GCCTAACATGCTTCGAGATC-3' (SEQ ID NO:15); IL-10 reverse primer: 5'-CTCATGGCTTTGTAGATGCC-3' (SEQ ID NO:16); IL-2 forward primer: 5'-TGAGGAGACGGTGACCAKGGT-3' (SEQ ID NO:17); IL-2 reverse primer: 5'-

- 5 CCCTGGGTCTTAAGTGAAAG-3' (SEQ ID NO:18); IFN-g forward primer: 5'-
 CAAGTTATATCTTGGCTTTTCAGC-3' (SEQ ID NO:19); IFN-g reverse primer: 5'-
 CTGGGATGCTCTTCGACCTCG-3' (SEQ ID NO:20).

Amplifications were initiated with a 5 min hot start at 95°C followed by 35 cycles at
 95°C for 30 sec, 48°C for 30 sec, and 72°C for 1 min each, and a terminal extension at
 10 72°C for 10 min. Equal volumes of GAPDH PCR products were resolved on 1.3%
 agarose gels stained with ethidium bromide. PCR products for individual cytokines
 were normalized to GAPDH signals, resolved on agarose gels, and images captured on a
 Gel Documentation System (BioRad).

First, a total, large volume core sample (i.e. tissue derived from both germinal center
 15 and T-cell zone) was evaluated for expression of cytokines specific for TH1 (IL-2, IFN-
 y) and TH2 (IL-4, IL10) profiles. RNA was isolated from cores and subjected to
 reverse transcription (RT). RTs were then amplified, via polymerase chain reaction
 (PCR), for TH1- (IL-2, IFN-y) and TH2-specific (IL-4, IL-10) cytokines. Figure 4A shows
 that the large volume total sentinel lymph node sample contained mRNA encoding the
 20 TH2 cytokines IL-4 and a moderate amount of IL-10, as well as the TH1 cytokine, IL-2.

To evaluate whether individual germinal center microenvironments might have
 different cytokine profiles, the cores of individual germinal centers were evaluated for
 cytokine mRNA similarly. Of note, amplification of mRNAs from germinal center A,
 Figure 4B, a CD20⁺/CD23⁺/Ki67⁺ histological pattern, showed no detectable
 25 amplification of any of the cytokine mRNAs. In contrast, germinal center B, Figure 4C,
 defined by CD20⁺/CD23⁺/Ki67⁻ histology marker profile, showed the presence of IL-10,
 INF-y and IL-2, suggesting a stronger TH1 environment. These observations demonstrate a
 mixed TH1/TH2 cytokine profile in the total node. However, germinal center
 microenvironments might show more selectivity for cytokine expression that could
 30 influence local humoral responses depending on the predominant cytokine levels.

Example 3. Production of V domain cDNA libraries.

To evaluate how these germinal centers, characterized by histological activation markers,
 may relate to progressive antibody production, the antibody variable domain
 repertoire was evaluated from tumor-draining sentinel lymph node germinal centers with
 35 CD20/CD23 positive staining and positive or negative for the proliferative marker, Ki67.
 This was carried out by synthesizing antibody variable domain libraries from select

5 germinal center core samples obtained from the same lymph node in a non-metastatic tumor-draining lymph node. The germinal centers were cored by alignment of the histological cryosection with the frozen block as shown in Figure 5, and total RNA was isolated and reverse-transcribed using random primed synthesis. In brief, RNA was extracted from cored tissue utilizing the Qiagen RNeasy system. RNA yields from
 10 isolated cores were typically 2-6 µg. Total RNA isolated (2µg) was subjected to reverse transcription (RT) using Superscript II (Invitrogen, Carlsbad, CA) with random hexamer primers in a total reaction volume of 20ul as described in the manufacturer's protocol.

For cDNA amplification and library generation of antibody VH,λ,κ fragments, Superscript II
 15 RT reactions were amplified via PCR using a modification of degenerate oligonucleotides described previously (33). VH forward primer containing a methionine start codon: 5'-ATGCAGGTGCAGCTGGTGSAGTCTGG-3' (SEQ ID NO:21); VH reverse primer: 5'-TGAGGAGACGGTGACCAKGGT-3' (SEQ ID NO:22); Vλ forward primer: 5'-CAGTCTGTGCTGACKCAGCCRCC-3' (SEQ ID NO:23); Vλ reverse primer: 5'-
 20 KAGGACGGTCAGCTKGGTSC-3' (SEQ ID NO:24); Vκ forward primer: 5'-GAMATTGTGMTGACSCAGTCTCC-3' (SEQ ID NO:25); Vκ reverse primer: 5'-TTTGATYTCCACCTTGGTCC-3' (SEQ ID NO:26).

One µl of random-heximer primed RT reactions described above was utilized in PCR with Platinum *Pfx* DNA Polymerase (Invitrogen). Amplifications were initiated with a
 25 5 min hot start at 94°C followed by 30 cycles at 94°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min each, and a terminal extension at 72°C for 10 min. PCR products were agarose gel purified and terminal 3'-dATP was added to purified fragments with Taq polymerase (Invitrogen, Carlsbad, CA) and 200 µM dATP for 10 min at 72°C. Fragments were ligated into pCR-T7TOPO expression vector (Invitrogen) and 1/20th ligation
 30 reactions used to transform *E. coli*.

The resultant transformants were screened by PCR using Vent DNA Polymerase (New England Biolabs, Beverly, MA), forward primer: 5'-CGCGAAATTAATACGACTCACTATAGGG-3' (SEQ ID NO:27), and 3' primer: 5'-CCTAAATTGTAAGCGTTAATCCGG-3' (SEQ ID NO:28). A positive screen resulted in
 35 a ~730 bp PCR fragment (~410 bp V_H, λ, κ sequence + ~320 bp vector sequence). PCR fragments that screened positive were separated from their primer precursors on a Qiagen (Valencia, CA) MinElute 96 UF PCR purification plate.

5 To assess how lymph-node microenvironment and activation status might affect B-cell antibody production and antigen-dependent somatic mutation, the germinal-center RNA samples were used to synthesize antibody variable domain libraries. Antibody variable regions were amplified with proof-reading DNA polymerase using degenerate primers that select VH, V λ , and V κ gene usage in humans (74). Figure 6A shows one
10 representative VH, V λ , and V κ amplification derived from a CD20⁺/CD23⁺/Ki67⁺ germinal center.

The VH, V λ , and V κ fragment libraries were then independently subcloned into a topoisomerase cloning vector (pCR-T7-TOPO), and individual clones were screened for insert by re-amplification and analyzed by agarose gel electrophoresis. Examples of
15 individual fragment amplification and analysis are shown in Figure 6B. Of note, VH fragments obtained from the sentinel lymph node, Ki67⁺ germinal center (Library A), demonstrated heterogeneity in gel mobility, possibly indicating antigen-driven somatic hypermutation/deletion-insertion events as compared to isolates derived from the Ki67⁻ germinal center of the same lymph node. The amplified fragments were subsequently
20 purified in 96-well format using PCR purification plates and the fragments analyzed by direct di-deoxy sequencing using a uni-directional primer.

Purified VH, λ , κ PCR fragments were sequenced, using a standard primer to the T7 promoter, a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA), and an ABI automated cycle sequencer (Applied Biosystems). VH, V λ ,
25 and V κ candidates were each directly compared to the VBASE directory of human Ig genes (DNAPLOT) (75). Variation of mutation and ratio of replacement to silent mutations were calculated for framework region (FR)1, complementarity-determining region (CDR)1, FR2, CDR2, FR3, and CDR3 (light chain only) regions for each sequence.

30 Example 4. VH germline usage.

B-cell clonal expansion within lymph nodes could be verified by identical V(D)J rearrangements within multiple clones that were not exactly identical and differ by several or more nucleotides within the VH region. The ability to detect multiple clones from a single B-cell lineage and, perhaps, several different B-cell lineages was evaluated
35 as a potential indicator of antigen-driven hypermutation within the proliferative Ki67⁺ compared to the Ki67⁻ VH libraries derived from the non-metastatic sentinel lymph node.

- 5 Individual VH sequences were obtained from sentinel lymph node germinal center library that were Ki67⁺ (Library A) and Ki67⁻ (Library B), respectively, as well as from a Ki67⁺ germinal center library from a more distal lymph node (Library C) from the same patient. Sequences for all three libraries were analyzed for their specific V(D)J rearrangements and clustering as B-cell subgroups.

Node	VH Library	Ki67 Phenotype ^a	# of Sequences ^b	% Seq Clonal ^c	% Mutation/R:S ^d			
					FR1	CDR1	FR2	CDR2
Sentinel	A	+	40	25	4.42/2.13	5.45/8.00	1.83/1.54	21.2/3.45
Sentinel	B	-	40	5.0	6.12/1.68	6.06/3.21	1.78/0.883	20.7/3.46
Distal	C	+	54	22	6.04/2.52	6.57/5.16	2.30/1.15	23.6/3.06

- 10 ^aPhenotype as measured via immunohistochemistry.
^bTotal number of complete V_H sequences in group.
^cPercentage of sequences in each group that belonged to a clonal subgroup.
^dPercentage of mutation from total of each V_H region and replacement to silent mutation ratio compared to respective germline sequence.
^ePercentage of sequences that contained nonrandom replacement and silent mutations as described by Lossos et al (40).
- 15 Table 2. Summary of molecular characteristics for isolated, sequenced VH genes.

- VH clones derived from the sentinel and proximal lymph-node Ki67⁺ germinal centers (Library A and C) generated more B-cell clonal subgroups than those derived from the Ki67⁻ germinal center (Library B, Table 2). The numbers of replacement and silent mutations were determined for each clone, and the ratio between the two (R:S ratio) was calculated. Generally, R:S ratios of 2.9 or less in the CDR region have been defined as random mutations while ratios greater than 2.9 indicate antigen-driven selection (76, 77). Conversely, antigen-driven selection should not affect the structural integrity of the variable domain; therefore, the R/S ratio in the FR regions should be low.
- 20
25

- Additionally, selective antigen-driven mutational events can be measured from the total number of mutations for each CDR and FR region. Based upon size of the respective regions alone, a random pattern of R and S mutations would localize three times more frequently in the FR region than the CDR region (78). As a result, a CDR/FR mutation ratio of greater than 0.3 would suggest antigen selection over random mutational events. Finally, the algorithm derived from Lossos et al. (79) was utilized to determine the significance of replacement and silent mutations for all regions within a VH sequence selective to actual mutation rates for each specific V(D)J gene usage combination.
- 30

5

Germ-Line Gene

V _H Library/Clone Phenotype ^a	Ki67	VH	DH	JH	FR Mutations ^b	CDR Mutations ^b	FR R:S ^c	CDR R:S ^c	CDR/ FR ^d	Nonrando R,S'
A/C4	+	V _H 3-23	D _H 6-25	J _H 3b	8 (1)	6 (0)	7.0	6:0	0.67	
A/C8	+				4 (2)	5 (0)	1.0	5:0	0.83	
A/E8	+				12 (6)	9 (1)	1.0	9:0	0.75	+
A/H8	+				1 (0)	5 (1)	1:0	5:0	5.0	+
A/A11	+	V _H 1-18	D _H 5-12	J _H 4b	10 (6)	11 (4)	0.67	1.75	1.1	+
A/H5	+				5 (3)	6 (0)	0.67	6:0	1.2	+
A/B10 ^f	+	V _H 3-21	D _H 5-12	J _H 3b	1 (0)	5 (0)	1:0	5:0	5.0	+
A/D11 ^f	+				1 (0)	5 (0)	1:0	5:0	5.0	+
A/E6	+	V _H 3-23	D _H 6-13	J _H 4b	3 (1)	0	3:0	-	0	
A/F5	+				13 (6)	14 (4)	0.86	2.5	1.1	+
B/B2	-	V _H 1-e	D _H 1-26	J _H 4b	15 (5)	8 (4)	2.0	1.0	0.53	
B/E6	-				22 (7)	4 (0)	0.47	4:0	0.18	
C/A4	+	V _H 3-73	D _H 3-22	J _H 4b	16 (8)	10 (3)	1.0	2.3	0.63	
C/C6	+				18 (10)	14 (5)	0.80	1.8	0.78	
C/E1	+				14 (6)	10 (3)	1.3	2.3	0.71	
C/B2	+	V _H 3-23	D _H 6-13	J _H 3b	4 (2)	4 (0)	1.0	4:0	0.67	
C/B12	+				4 (2)	5 (0)	1.0	5:0	0.83	
C/E6	+				5 (2)	5 (0)	1.5	5:0	1.0	
C/B10 ^g	+	V _H 3-74	D _H 1-26	J _H 4b	12 (2)	7 (2)	5.0	2.5	0.58	
C/D12 ^g	+				12 (2)	7 (2)	5.0	2.5	0.58	
C/C1	+				14 (2)	7 (1)	6.0	6.0	0.50	
C/F5	+				12 (3)	7 (2)	3.3	2.5	0.58	
C/C10 ^h	+	V _H 1-46	D _H 5-05	J _H 4b	14 (5)	14 (3)	1.8	3.7	1.0	+
C/F9 ^h	+				14 (5)	14 (3)	1.8	3.7	1.0	+

^aPhenotype as measured via immunohistochemistry.^bTotal mutations measured in FR and CDR regions respectively; silent mutations are in parenthesis.^cReplacement to silent mutation ratios for FR and CDR

regions respectively.

^dTotal mutation ratio for CDR and FR regions respectively.^e "+" indicates nonrandom replacement and silent mutations as calculated by algorithm from Lossos et al. (40).^{f,g,h} Each pair had identical sequences.

10

15

Table 3. Molecular characterization of selected clonal subgroups derived from 3 V_H libraries.

20

Table 3 summarizes the mutation analysis results for the V_H sequences from each germinal center library analyzed. Twenty-five percent of the total sequences (10/40) from the Ki67⁺ sentinel node (Library A) demonstrated clonality. Of this subset, nine out of ten had CDR/FR ratios that indicated antigenic selection. Of these nine, seven were deemed

- 5 significant relative to their respective germline V(D)J sequences based on the calculated algorithm described by Lassos et al. (40).

Sequences from the Ki67⁺ distal node (Library C) also demonstrated B-cell clonality (12/54 or 22% of total sequences). While these sequences had CDR/FR ratios to indicate antigen selection, only two out of twelve were considered significant. In contrast,
10 sequences from the Ki67⁻ sentinel node (Library B) identified little clonality (2/40 or 5%). In addition, neither of the two sequences showed significance in their mutations compared to random germline and only one of the two clones had a CDR/FR ratio >0.3.

Example 5. Somatic hypermutation in libraries.

In order to determine whether individual clones within each library may have undergone
15 significant somatic hypermutational events, CDR/FR ratios from all sequenced genes from each of the three germinal center VH libraries were compiled. Selective antigen-driven hypermutation among individual sequences within each library was determined irrespective of B-cell clonality. To purposely select a significant VH somatic hypermutation quantification profile, CDR/FR ratios were grouped with a higher
20 stringency by selecting those that fell above or below a CDR/FR ratio of 0.8.

The results indicated that the Ki67⁺ sentinel lymph node (Library A) contained 26 sequences representing 65% of the total library with CDR/FR ratios >0.8 as compared to only 13 clones representing 33% the Ki67⁻ library from the same lymph node (Library B) (Figure 7). In contrast, the distal Ki67⁺ node (Library C) showed only 15 or 28%
25 of the total sequences with a high CDR/FR ratio >0.8. Individual B-cell mutational events in the distal node, therefore, were significantly lower than those observed in the sentinel lymph node despite having an increased number of B-cell subgroups.

Example 6. V_κ and V_λ germline usage.

Light chain usage for each of the germinal-center cDNA libraries was determined in a
30 similar manner to that of VH. Lambda and kappa light-chain genes were amplified from RT reactions, and individual clones were analyzed for germline usage and mutational events. With some exceptions, there was little evidence of significant mutational events when light chain sequences were matched to germline in DNAPLOT.

5

Library A (Ki67 ⁺)							
V _κ region	J region	# isolates	CDR/FR > 1.5 ^a	V _λ region	J region	# isolates	CDR/FR > 1
DPK24/VKIVKlob	JK4	36	8 (22%)	1e.10.2/DPL8	JL1	15	0
DPK9/012	JK2	5	0	3r9C5/DPL23	JL2a/JL3a	11	2 (18%)
Vg/38K	JK4	4	0	1c.10.2/DPL2	JL3b	5	1 (20%)
DPK9/012	JK1	3	0	1e.10.2/DPL8	JL2a/JL3a	5	1 (20%)
DPK9/012	JK4	3	0	3j118D9/V2-6	JL2a/JL3a	5	0
DPK22/A27	JK1	2	0	1e.10.2/DPL8	JL3b	3	0
DPK3/L11	JK1	2	0	1g.400B5/DPL3	JL3b	3	0
DPK5/Vb	JK4	2	0	2a.2.272A12/DP11	JL2a/JL3a	3	0
DPK6/V''	JK4	2	0	2a.2.272A12/DP11	JL3b	2	0
DPK15/A19	JK1	2	0	2e.2.2/V1-3	JL2a/JL3a	2	0
DPK15/A19	JK4	2	0	1v318	JL2a/JL3a	2	0
DPK16/A23	JK4	2	1 (50%)				
LFVK431	JK4	2	0	Non-clonal group		15	3 (20%)
Non-clonal group		12	1 (8%)				
79 sequences analyzed – 67 clonal (85%) (n>2) Total # sequences CDR/FR > 0.8 = 43 (54%) Clonal sequences (n~4)and CDR/FR > 0.8 – 35/45 (78%)				71 sequences analyzed – 56 clonal (79%) (n>2) Total # sequences CDR/FR > 0.8 = 22 (31%) 13/41 (32%) Clonal sequences (n~4)and CDR/FR > 0.8 –			

Library B (Ki67 ⁺)							
V _κ region	J region	# isolates	CDR/FR > 1.1 ^a	V _λ region	J region	# isolates	CDR/FR >
L12a/PCR dil 6-5	JK1	9	0	1c.10.2/DPL2	JL3b	6	3 (50%)
DPK24/VKIVKlob	JK1	8	3 (8%)	1g.400B5/DPL3	JL3b	5	1 (20%)
DPK24/VKIVKlob	JK4	7	2 (29%)	1g.400B5/DPL3	JL2a/JL3a	4	1 (25%)
DPK9/012	JK4	6	1 (17%)	2e.2.2/V1-3	JL3b	4	0
DPK3/L11	JK1	3	0	1c.10.2/DPL2	JL1	3	0
DPK9/012	JK2	3	0	IGLV3S2	JL3b	3	0
DPK5/Vb	JK1	2	0	1e.10.2/DPL8	JL1	2	0
DPK9/012	JK1	2	1 (50%)	2e.2.2/V1-3	JL1	2	0
DPK22/A27	JK1	2	2	2e.2.2/V1-3	JL2a/JL3a	2	1 (50%)
DPK15/A19	JK4	2	0	6a.366F5/V1-22	JL3b	2	0
DPK24/Vd	JK4	2	0	2c.118D9/V1-2	JL3b	2	0
Vg/38K	JK4	2	0	Non-clonal group		16	4 (25%)
Non-clonal group		17	1 (6%)				

10	65 sequences analyzed – 48 clonal (74%) (n>2) clonal (69%) (n>2) Total # sequences CDR/FR > 0.8 = 19 (29%) 0.8 = 27 (53%) Clonal sequences (n~4)and CDR/FR > 0.8 – 11/30 (37%) CDR/FR > 0.8 – 12/19 (63%)	51 sequences analyzed – 35 Total # sequences CDR/FR > Clonal sequences (n~4)and
----	--	---

^aHigher CDR/FR stringency set for each region to delineate 7-10 clones with highest ratios.

15 Table 4. Clonally related sets for V light chains.

There was a higher level of V-J clonality in the sentinel Ki67⁺ germinal center library (Table 4), as compared to the number of B-cell clusters obtained for the VH sequences.

5 For example, the V_{κ} - J_{κ} combination of DPK24/VKIVKlob-JK4 was present in 45% of the clones sequenced from Library A, the $Ki67^{+}$ germinal center library of the sentinel lymph node. The V-J combinations for the V_{λ} libraries were less clonal than observed for V_{κ} . In contrast, the non-proliferative $Ki67^{-}$ sentinel-node germinal center and $Ki67^{+}$ distal lymph node libraries (Libraries B and C, respectively) demonstrated a higher V-J
10 gene usage diversity for both V_{κ} and V_{λ} (Figure 8). Consequently, selective V-J gene usage was observed in the sentinel lymph node $Ki67^{+}$ germinal center at levels far exceeding what could be determined as associated with B-cell clonal expansion of the V(D)J recombined VH regions. It is possible that the light chain usage is very selective to the resident B-cells within the germinal center and might not necessarily represent
15 antigen-driven subgroups.

Example 7. Method for evaluating VH open reading frames and selection for recombinant protein synthesis

20 The analysis of VH libraries for clonal expansion and somatic hypermutation indicated that the selection of proliferative germinal centers was appropriate to define B-cell-activated mechanisms compared to non-proliferative GCs. However, it was still necessary to directly evaluate the potential of these clones to be functional VH protein domains. To carry out such an evaluation, all of the sequences in the clonal groups were investigated for complete open reading frame sequences.

25 Translated sequences showed that approximately 60~70% of the sequences contained complete open reading frames, with 20% of the remaining having single mutations within the fragment ends, indicating PCR/TOPO cloning artifacts, and the remainder having insertional stop codons within the VH sequences. Figure 10 shows a translated VH sequence of one of the clonal selected clones compared to one with a baseline mutation rate relevant
30 that germline locus. Most of the sequence differences (as demarcated by asterisks) were concentrated within the CDR3 domain, as would be expected for antigen-dependent somatic hypermutational events. By analyzing the clonal expanded VH clones, it is possible to define those that contain complete open reading frames and have significant antigen-driven mutational events.

35 Example 8. Utility of recombinant VH domain proteins in primary screens of tumor extracts.

5 VH sequences encoding complete open reading frames will be useful in defining antigen through specific interactions. To develop a rapid method to incorporate the VH domains in tumor screening and antigen identification assays, VH sequences were used in a bacterial protein expression system that employs C-terminal epitope fusions of c-myc and 6X-His tag domains for affinity purification and detection. Clones of somatic hypermutated VH and
10 controls with low R:S ratios and insufficient CDR/FR ratios were transferred into pTrc-His bacterial expression vectors (Invitrogen). In-frame clones were confirmed by sequencing.

These clones were used to express and purify the protein employing IPTG-inducible expression and purification from French press lysates using a single step metal affinity chromatography with cobalt columns (BD Biosciences- Talon matrices). Figures 11A and
15 11B show the recombinant VH fusion protein domain structure and an example of Talon purification of the proteins to homogeneity, respectively. This protein synthesis and purification platform thus provided yielded mg levels of fusion proteins -- sufficient to apply in the assays to detect and identify tumor antigens.

Example 9. Identification of breast cancer antigens.

20 Recombinant VH proteins synthesized using the inducible bacterial system provide several advantages useful in the identification of their selective target antigens. Multiple clones can be used to compare and contrast for their selectivity and specificity, and their fusion structure facilitates an NPHS. terminal orientation, when the C-terminal 6X-His domain is bound to an affinity matrix. These features were used to implement a microscale affinity
25 selection process to pull VH-specific antigens from the original patient tumor extracts.

To avoid non-specific interactions in the form of extraneous non-specific binding of proteins from tumor or cellular upon use of the standard metal affinity matrices, such as Talon and nickel-agarose, a micro-well based assay was employed for binding the recombinant VH proteins to nickel coated plates (Sigma High capacity His-Select plates). After washing
30 away unbound material, non-specific binding surfaces were blocked with albumin. Triton X-100 (1%) soluble tumor extracts from the same patient (500 µg) were applied to each well that contained either no antibody, somatic hypermutated VH Ab-1, or a non-specific, non-hypermutated VH Ab-4. The extracts were allowed to bind for 2 hours and were then washed eight times for 15 minutes each and with significant salt at 0.3M, a competitive
35 level of imidazole to reduce histidine interactions and 0.05% non-ionic detergent.

- 5 The wells were subsequently washed with PBS to eliminate unwanted detergent and salt. They were prepared for direct liquid chromatography linked to quadrupole mass spectrometry/mass spectrometry (LC-MS/MS) via surface trypsinization to release bound proteins and peptides. The digested samples were prepared for the LC-MS/MS by drying and resuspension into column solvent (57-60).
- 10 The VH antigen-trap process (FIG. 12A) yielded an indication that only the VH with somatic hypermutated sequences identified a protein with at least five high probability matched peptides and 3 additional peptides. All other proteins bound had single peptide matches and, thus, constitute unconfirmed sequences. The controls of no antibody did not detect selective proteins. The VH selected peptide matched to a cell surface glycosylated
- 15 protein with a triple extracellular Ig-like repeat matching to stromal cell derived receptor-1 originally cloned from a genomic screen (61). Of note, alternative forms of the protein exist; in neuronal tissues, it has been investigated as a synaptic membrane protein thought to provide anti-adhesiveness to the neuronal synapses, termed the neuropilins(NP)(62, 63). The neuropilins have two apparent molecular weight isoforms, one selective to neurons
- 20 (gp65) and the other observed in peripheral tissues, including liver and lung (gp55).

To determine if these recombinant Vh proteins show similar selectivity, an inverse assay was carried out to evaluate their ability to directly detect potential target antigens in tumor extracts. Total tumor extracts were bound onto high affinity immune assay plates overnight at 4°C, including the original tumor extracts, as well as those from three additional non-

25 autologous tumors. The recombinant VH antibodies were then allowed to bind to blocked plates in a standard direct ELISA format.

To increase the sensitivity to bound VH proteins, secondary anti-epitope antibodies (anti-c-myc or anti-His, of which anti-His was more effective with the least background) were incorporated, followed by tertiary antibodies covalently linked to horseradish peroxidase

30 (HRP). The reaction of HRP was performed with the sensitive tyramide fluorescent reagent (Molecular Probes), and the plates were quantified by cytofluorimetric analysis in a multiwell plate fluorimeter (BioRad CytoFluor) (FIG. 13). Of note, the VH Ab-1 had a significant signal when compared with its own extract and one of the three others. No other antibodies reacted to the primary extract from the same patient, indicating that the VH Ab-1

35 was selective and somatic mutation-dependent. One of the other antibodies did have a positive threshold signal in another sample (Ab-3, patient C).

5 The results indicate that it is possible for recombinant VH domains to be used to screen tumor extracts and, possibly, other biological samples in a sensitive manner. Furthermore, it is possible to use recombinant variable region proteins to perform antigen screening and identification assays, especially by combining functional affinity binding with LC-MS/MS peptide identification.

10 Example 10. Verification and evaluation of the breast cancer antigens identified by the antigen trap method

The expression of SDR-1/NP protein isoforms in human breast cancer tumor extracts was evaluated using immunoblot analysis with anti-SDR-1/NP polyclonal antibodies provided
15 by Dr. Eckhart Gundelfinger (FIG. 14A-D). Of note, some of the tumor extracts demonstrated a predominant 45kDa isoform similar but not exactly corresponding with the staining of the 45kDa forms observed upon stripping the same blot and immunoblotting with the recombinant VH Ab-1 protein. For example the signal for sample 2 was amplified over sample 1 with the VH Ab-1 detection but was shown to be similar to sample 1 when
20 using the polyclonal antibody.

The recognition of SDR-1/NP by the VH Ab-1 antibody was confirmed via immunoprecipitation with the polyclonal antibody and the detection via immunoblot of the precipitated proteins with the Ab-1 antibody. In light of the indication of approximately 50% distribution of the 45kDa isoform of the SDR-1/NP protein in some breast cancers, a
25 breast cancer tissue array was examined for SDR-1/NP expression. The SDR-1/NP polyclonal antibody staining showed that some of the tumors exhibited rather high expression, compared to positive tumors with a moderate uniform expression (FIG. 15).

Of note, the staining was selective to tumor cells with predominance for invasive lobular carcinomas or highly invasive ductal carcinomas, with a dispersed presentation through
30 surrounding stroma. Several cases even indicated low expression in the solid tumor areas and increased expression in areas where the tumors are dispersed as small foci through the stroma. The percentage positive in all samples was 14/32. Of those positive for SDR-1/NP, 9/14 were metastatic to local lymph nodes upon primary resection. The distribution with respect to estrogen receptor status was nearly equally split at 8/14 estrogen receptor positive.

35 The SDR-1/NP antibody's detection of both ubiquitous 55kDa and the 45kDa alternative isoforms in breast cancer prompted the examination of molecular mRNA expression by cloning the cDNA for SDR-1. RT-PCR of MDA-MB-231 human invasive carcinoma cells was carried out (FIG. 16A). The 5' coding region SDR-1 amplicon (lane 2) indicates that

5 the alternative spliced alpha isoform was predominant in the cells, consistent with the expression of the 55kDa protein upon post-translational glycosylation.

To determine if the 45kDa protein isoform was the result of possible alternative splicing elsewhere in the cDNA, 5' and 3' RACE cloning was performed using N-terminal protein primers and C-terminal primers to extend 5' and 3', respectively, after selective reverse
10 transcription with terminal linkers. Amplification of the 3' coding and UTR was ineffective (FIG. 16B, lane 3), indicating that the 3' coding primers used may not correspond to the coding sequence. To further investigate whether this indicates a 3' alternative splicing event, additional primer pairs residing further upstream were used to correspond to those that were effective for the 5' RACE.

15 The 5' coding region amplicon was confirmed as human SDR-1/NP by direct sequencing and, subsequently, used as a hybridization probe for a cDNA array comparing normal and tumor patient matched samples from 19 different tumors (BD Clontech, Cancer profiling Array II). Of note, breast tumor samples showed increases in SDR-1 expression compared to their matched normal tissues (FIG. 16C). Many other tissues express SDR-1/NP, but
20 virtually all have reduced levels in the tumor samples, with a number of case-specific exceptions.

The previously described evaluation process is a relevant approach to screen independent antigens defined by VH or scFv antibody proteins recovered from reactive tumor-draining lymph nodes. Confirmation of antigen identification may be effected via: 1) direct
25 immunoanalysis methods such as direct SDS-PAGE blots and precipitation experiments; 2) breast tissue histological array analysis; and 3) cDNA amplification and molecular expression profiling using cDNA arrays, qualitative RT-PCR, and development of qRT-PCR for genes with specific isotype regulation, *i.e.*, a specific expression of an alternative spliced form.

30 Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be apparent to those skilled in the art that certain changes and modifications can be practiced. Therefore, the description and examples should not be construed as limiting the scope of the invention, which is delineated by the appended numbered claims.

35

5 **REFERENCES**

1. Barry, M.M. and J.S. Lee, *Cloning and expression of an autoimmune DNA-binding single chain Fv. Only the heavy chain is required for binding*. Mol Immunol, 1993. **30**(9): p. 833-40.
2. Blank, M., et al., *Characteristics and pathogenic role of anti-beta2-glycoprotein I single-chain Fv domains: induction of experimental antiphospholipid syndrome*. Int Immunol, 1999. **11**(12): p. 1917-26.
3. Finnern, R., et al., *Human autoimmune anti-proteinase 3 scFv from a phage display library*. Clin Exp Immunol, 1997. **107**(2): p. 269-81.
4. van den Brink, E.N., et al., *Two classes of germline genes both derived from the V(H)1 family direct the formation of human antibodies that recognize distinct antigenic sites in the C2 domain of factor VIII*. Blood, 2002. **99**(8): p. 2828-34.
5. Cai, X. and A. Garen, *A melanoma-specific VH antibody cloned from a fusion phage library of a vaccinated melanoma patient*. Proc Natl Acad Sci U S A, 1996. **93**(13): p. 6280-5.
6. Cai, X. and A. Garen, *Comparison of fusion phage libraries displaying VH or single-chain Fv antibody fragments derived from the antibody repertoire of a vaccinated melanoma patient as a source of melanoma-specific targeting molecules*. Proc Natl Acad Sci U S A, 1997. **94**(17): p. 9261-6.
7. Pruneri, G., et al., *Monoclonal proliferation of germinal center cells (incipient follicular lymphoma) in an axillary lymph node of a melanoma patient*. Hum Pathol, 2001. **32**(12): p. 1410-3.
8. Liang, M., et al., *Baculovirus expression cassette vectors for rapid production of complete human IgG from phage display selected antibody fragments*. J Immunol Methods, 2001. **247**(1-2): p. 119-30.
9. Mansfield, C.M., et al., *Ten-year results in 1070 patients with stages I and II breast cancer treated by conservative surgery and radiation therapy*. Cancer, 1995. **75**(9): p. 2328-36.
10. Neff, P.T., et al., *Long-term results of breast conservation therapy for breast cancer*. Ann Surg, 1996. **223**(6): p. 709-16; discussion 716-7.
11. Fisher, B., et al., *Findings from recent National Surgical Adjuvant Breast and Bowel Project adjuvant studies in stage I breast cancer*. J Natl Cancer Inst Monogr, 2001(30): p. 62-6.

- 5 12. Hunter, R.L., D.J. Ferguson, and L.W. Coppleson, *Survival with mammary cancer related to the interaction of germinal center hyperplasia and sinus histiocytosis in axillary and internal mammary lymph nodes*. *Cancer*, 1975. **36**(2): p. 528-39.
13. Whitford, P., et al., *Flow cytometric analysis of tumour-draining lymph nodes in breast cancer patients*. *Eur J Cancer*, 1992. **28**(2-3): p. 350-6.
- 10 14. Bright, J.J. and S. Sriram, *TGF-beta inhibits IL-12-induced activation of Jak-STAT pathway in T lymphocytes*. *J Immunol*, 1998. **161**(4): p. 1772-7.
15. Chen, W. and S.M. Wahl, *TGF-beta: the missing link in CD4+CD25+ regulatory T cell-mediated immunosuppression*. *Cytokine Growth Factor Rev*, 2003. **14**(2): p. 85-9.
- 15 16. Kuerer, H.M., et al., *Axillary lymph node cellular immune response to HER-2/neu peptides in patients with carcinoma of the breast*. *J Interferon Cytokine Res*, 2002. **22**(5): p. 583-92.
17. Shevach, E.M., *Certified professionals: CD4(+)CD25(+) suppressor T cells*. *J Exp Med*, 2001. **193**(11): p. F41-6.
- 20 18. Baecher-Allan, C., V. Viglietta, and D.A. Hafler, *Human CD4+CD25+ regulatory T cells*. *Semin Immunol*, 2004. **16**(2): p. 89-98.
19. Shimizu, J., S. Yamazaki, and S. Sakaguchi, *Induction of tumor immunity by removing CD25+CD4+ T cells: a common basis between tumor immunity and autoimmunity*. *J Immunol*, 1999. **163**(10): p. 5211-8.
- 25 20. Suttmuller, R.P., et al., *Synergism of cytotoxic T lymphocyte-associated antigen 4 blockade and depletion of CD25(+) regulatory T cells in antitumor therapy reveals alternative pathways for suppression of autoreactive cytotoxic T lymphocyte responses*. *J Exp Med*, 2001. **194**(6): p. 823-32.
21. Onizuka, S., et al., *Tumor rejection by in vivo administration of anti-CD25 (interleukin-2 receptor alpha) monoclonal antibody*. *Cancer Res*, 1999. **59**(13): p. 3128-33.
- 30 22. Woo, E.Y., et al., *Regulatory CD4(+)CD25(+) T cells in tumors from patients with early-stage non-small cell lung cancer and late-stage ovarian cancer*. *Cancer Res*, 2001. **61**(12): p. 4766-72.
- 35 23. Liyanage, U.K., et al., *Prevalence of regulatory T cells is increased in peripheral blood and tumor microenvironment of patients with pancreas or breast adenocarcinoma*. *J Immunol*, 2002. **169**(5): p. 2756-61.

- 5 24. Rossen, R.D., et al., *The C1q binding test for soluble immune complexes: clinical correlations obtained in patients with cancer*. J Natl Cancer Inst, 1977. **58**(5): p. 1205-15.
25. Slamon, D.J., et al., *Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene*. Science, 1987. **235**(4785): p. 177-82.
- 10 26. McGuckin, M.A., et al., *Prognostic significance of MUC1 epithelial mucin expression in breast cancer*. Hum Pathol, 1995. **26**(4): p. 432-9.
27. Chu, J.S. and K.J. Chang, *Mucin expression in mucinous carcinoma and other invasive carcinomas of the breast*. Cancer Lett, 1999. **142**(1): p. 121-7.
28. Coronella-Wood, J.A. and E.M. Hersh, *Naturally occurring B-cell responses to breast cancer*. Cancer Immunol Immunother, 2003. **52**(12): p. 715-38.
- 15 29. Leyland-Jones, B., *Trastuzumab: hopes and realities*. Lancet Oncol, 2002. **3**(3): p. 137-44.
30. Hadden, J.W., *The immunology and immunotherapy of breast cancer: an update*. Int J Immunopharmacol, 1999. **21**(2): p. 79-101.
- 20 31. MacLean, G.D., et al., *Enhancing the effect of THERATOPE STn-KLH cancer vaccine in patients with metastatic breast cancer by pretreatment with low-dose intravenous cyclophosphamide*. J Immunother Emphasis Tumor Immunol, 1996. **19**(4): p. 309-16.
32. Cote, R.J., et al., *Generation of human monoclonal antibodies reactive with cellular antigens*. Proc Natl Acad Sci U S A, 1983. **80**(7): p. 2026-30.
- 25 33. Goldstein, N.I., et al., *Isolation and characterization of a human monoclonal antibody which reacts with breast and colorectal carcinoma*. Anticancer Res, 1990. **10**(6): p. 1491-500.
34. Posner, M.R., et al., *An IgG human monoclonal antibody reactive with a surface membrane antigen expressed on malignant breast cancer cells*. Hum Antibodies Hybridomas, 1991. **2**(2): p. 74-83.
- 30 35. Petrarca, C., et al., *Isolation of MUC1-primed B lymphocytes from tumour-draining lymph nodes by immunomagnetic beads*. Cancer Immunol Immunother, 1999. **47**(5): p. 272-7.
- 35 36. Clark, E.A., et al., *Activation of human B cells. Comparison of the signal transduced by IL-4 to four different competence signals*. J Immunol, 1989. **143**(12): p. 3873-80.
37. Winthrop, M.D., et al., *Selection and characterization of anti-MUC-1 scFvs intended for targeted therapy*. Clin Cancer Res, 2003. **9**(10 Pt 2): p. 3845S-53S.

- 5 38. Wong, C., et al., *Human scFv antibody fragments specific for the epithelial tumour marker MUC-1, selected by phage display on living cells*. Cancer Immunol Immunother, 2001. **50**(2): p. 93-101.
39. Mao, S., et al., *Phage-display library selection of high-affinity human single-chain antibodies to tumor-associated carbohydrate antigens sialyl Lewisx and Lewisx*. Proc Natl Acad Sci U S A, 1999. **96**(12): p. 6953-8.
- 10 40. Kupsch, J.M., et al., *Isolation of human tumor-specific antibodies by selection of an antibody phage library on melanoma cells*. Clin Cancer Res, 1999. **5**(4): p. 925-31.
41. Sahin, U., et al., *Human neoplasms elicit multiple specific immune responses in the autologous host*. Proc Natl Acad Sci U S A, 1995. **92**(25): p. 11810-3.
- 15 42. Minenkova, O., et al., *Identification of tumor-associated antigens by screening phage-displayed human cDNA libraries with sera from tumor patients*. Int J Cancer, 2003. **106**(4): p. 534-44.
43. Mintz, P.J., et al., *Fingerprinting the circulating repertoire of antibodies from cancer patients*. Nat Biotechnol, 2003. **21**(1): p. 57-63.
- 20 44. Hansen, M.H., B. Ostenstad, and M. Sioud, *Identification of immunogenic antigens using a phage-displayed cDNA library from an invasive ductal breast carcinoma tumour*. Int J Oncol, 2001. **19**(6): p. 1303-9.
45. Coronella, J.A., et al., *Antigen-driven oligoclonal expansion of tumor-infiltrating B cells in infiltrating ductal carcinoma of the breast*. J Immunol, 2002. **169**(4): p. 1829-36.
- 25 46. Nzula, S., J.J. Goings, and D.I. Stott, *Antigen-driven clonal proliferation, somatic hypermutation, and selection of B lymphocytes infiltrating human ductal breast carcinomas*. Cancer Res, 2003. **63**(12): p. 3275-80.
47. Chin, Y., et al., *Phenotypic analysis of tumor-infiltrating lymphocytes from human breast cancer*. Anticancer Res, 1992. **12**(5): p. 1463-6.
- 30 48. Marsigliante, S., et al., *Computerised counting of tumour infiltrating lymphocytes in 90 breast cancer specimens*. Cancer Lett, 1999. **139**(1): p. 33-41.
49. Bisca, A., et al., *Proteomic evaluation of core biopsy specimens from breast lesions*. Cancer Lett, 2004. **204**(1): p. 79-86.
- 35 50. Hamler, R.L., et al., *A two-dimensional liquid-phase separation method coupled with mass spectrometry for proteomic studies of breast cancer and biomarker identification*. Proteomics, 2004. **4**(3): p. 562-77.
51. Somiari, R.I., et al., *High-throughput proteomic analysis of human infiltrating ductal carcinoma of the breast*. Proteomics, 2003. **3**(10): p. 1863-73.

- 5 52. Rosenblatt, K.P., et al., *Serum proteomics in cancer diagnosis and management*. Annu Rev Med, 2004. **55**: p. 97-112.
53. Kennedy, S., *Proteomic profiling from human samples: the body fluid alternative*. Toxicol Lett, 2001. **120**(1-3): p. 379-84.
54. Rousset, F., et al., *Regulation of Fc receptor for IgE (CD23) and class II MHC*
10 *antigen expression on Burkitt's lymphoma cell lines by human IL-4 and IFN-*
gamma. J Immunol, 1988. **140**(8): p. 2625-32.
55. Defrance, T., et al., *Human recombinant interleukin 4 induces Fc epsilon receptors*
(CD23) on normal human B lymphocytes. J Exp Med, 1987. **165**(6): p. 1459-67.
56. Lossos, I.S., et al., *The inference of antigen selection on Ig genes*. J Immunol, 2000.
15 **165**(9): p. 5122-6.
57. Griffin, T.J., et al., *Toward a high-throughput approach to quantitative proteomic*
analysis: expression-dependent protein identification by mass spectrometry. J Am
Soc Mass Spectrom, 2001. **12**(12): p. 1238-46.
58. Han, D.K., et al., *Quantitative profiling of differentiation-induced microsomal*
20 *proteins using isotope-coded affinity tags and mass spectrometry*. Nat Biotechnol,
2001. **19**(10): p. 946-51.
59. Lundgren, D.H., et al., *PROTEOME-3D: An Interactive Bioinformatics Tool for*
Large-Scale Data Exploration and Knowledge Discovery. Mol Cell Proteomics,
2003. **2**(11): p. 1164-76.
- 25 60. Taylor, C.M., et al., *Proteomic mapping provides powerful insights into functional*
myelin biology. Proc Natl Acad Sci U S A, 2004. **101**(13): p. 4643-8.
61. Shirozu, M., et al., *Characterization of novel secreted and membrane proteins*
isolated by the signal sequence trap method. Genomics, 1996. **37**(3): p. 273-80.
62. Langnaese, K., P.W. Beesley, and E.D. Gundelfinger, *Synaptic membrane*
30 *glycoproteins gp65 and gp55 are new members of the immunoglobulin superfamily*.
J Biol Chem, 1997. **272**(2): p. 821-7.
63. Zhai, R.G., et al., *Assembling the presynaptic active zone: a characterization of an*
active one precursor vesicle. Neuron, 2001. **29**(1): p. 131-43.
64. Agarwal, A., et al., *N-acetyl-cysteine promotes angiostatin production and vascular*
35 *collapse in an orthotopic model of breast cancer*. Am J Pathol, 2004. **164**(5): p.
1683-96.
65. Skobe, M., et al., *Induction of tumor lymphangiogenesis by VEGF-C promotes*
breast cancer metastasis. Nat Med, 2001. **7**(2): p. 192-8.

- 5 66. Rimm, D.L., et al., *Tissue microarray: a new technology for amplification of tissue resources*. Cancer J, 2001. 7(1): p. 24-31.
67. Camp, R.L., L.A. Charette, and D.L. Rimm, *Validation of tissue microarray technology in breast carcinoma*. Lab Invest, 2000. 80(12): p. 1943-9.
68. Freyre, F.M., et al., *Very high expression of an anti-carcinoembryonic antigen single chain Fv antibody fragment in the yeast Pichia pastoris*. J Biotechnol, 2000. 10 76(2-3): p. 157-63.
69. Holliger, P., *Expression of antibody fragments in Pichia pastoris*. Methods Mol Biol, 2002. 178: p. 349-57.
70. Bobrow, M.N., et al., *Catalyzed reporter deposition, a novel method of signal amplification. Application to immunoassays*. J Immunol Methods, 1989. 15 125(1-2): p. 279-85.
71. von Wasielewski, R., et al., *Tyramine amplification technique in routine immunohistochemistry*. J Histochem Cytochem, 1997. 45(11): p. 1455-9.
72. Homola, J., *Present and future of surface plasmon resonance biosensors*. Anal 20 Bioanal Chem, 2003. 377(3): p. 528-39.
73. Wadhwa, M., et al., *Strategies for detection, measurement and characterization of unwanted antibodies induced by therapeutic biologicals*. J Immunol Methods, 2003. 278(1-2): p. 1-17.
74. Sblattro, D., and A. Bradbury. 1998. A definitive set of oligonucleotide primers for 25 amplifying human V regions. *Immunotechnology* 3:271.
75. Giudicelli, V., D. Chaume, G. Mennessier, H.H. Althaus, W. Muller, J. Bodmer, A. Malik, and M.P. Lefranc. 1998. IMGT, the international ImMuno GeneTics database: a new design for immunogenetics data access. Medinfo 9 Pt 1:351.
76. Jukes, T.H., and J.L. King. 1979. Evolutionary nucleotide replacements in DNA. 30 *Nature* 281:605.
77. Schlomchik, M.J., A. Marshak-Rothstein, C.B. Wolfowicz, T.L. Rothstein, and M.G. Weigert, 1987. The role of clonal selection and somatic mutation in autoimmunity. *Nature* 328:805.
78. Davi, F., K. Maloum, O. Pritsch, C. Magnac, E. Macintyre, F. Salomon-Nguyen, 35 J.L. Binet, G. Dighiero, and H. Merle-Beral. 1996. High frequency of somatic mutations in the VH genes expressed in prolymphocytic leukemia. *Blood* 88:3953.
79. Lossos, I.S., R. Tibshirani, B. Narasimhan, and R. Levy. 2000. The Inference of Antigen Selection on Ig Genes. *J Immunol* 165:512

WHAT IS CLAIMED IS:

1. A method for identifying a reactive B-cell germinal center in a lymph node sample, comprising detecting three histological markers, CD20, CD23, and Ki67 at a location in the sample and comparing the detected signals, wherein a positive signal for all three markers at said location is indicative of a reactive B-cell germinal center, thereby identifying a reactive B-cell germinal center in a lymph node sample.
2. The method of claim 1, further comprising excising a sample of lymph nodes and freezing said sample.
3. The method of claim 1 or 2, wherein the lymph nodes are tumor-draining or pathologically inflamed.
4. A method for identifying immune cytokine profiles in a germinal center microenvironment, comprising:
 - i) obtaining a tissue sample of tumor-draining sentinel or distal lymph node;
 - ii) isolating mRNA from the tissue sample from a reactive germinal center and a non-reactive germinal center;
 - iii) determining the levels of TH1 and TH2 cytokines from mRNA of ii);
 - iv) comparing the levels of the cytokine mRNAs from the reactive germinal center and the non-reactive germinal center as determined in step (iii) where higher TH2 cytokines in the germinal center indicates an active immune reaction, thus identifying immune cytokine profiles in a germinal center microenvironment.
5. The method of claim 4, wherein the tumor is associated with breast cancer.
6. A method for identifying antibody variable regions with somatic mutations in a lymph node, comprising:
 - i) obtaining a tissue sample from a lymph node;
 - ii) preparing VH, V λ and V κ germinal center antibody variable region cDNA libraries;
 - iii) compiling CDR/FR ratios from sequenced variable region cDNAs from the libraries; and

- iv) selecting variable region sequences with CDR/FR ratios greater than about 0.8 in the libraries,
thus identifying antibody variable regions with somatic mutations in a lymph node.
- 7. The method of claim 6, wherein a presence of antibody variable regions with somatic mutations is indicative of immune reaction in the lymph node.
- 8. The method of claim 6 or 7, wherein the lymph node is a tumor-draining lymph node.
- 9. An isolated tumor antigen identified by the steps comprising:
 - i) expressing an antibody variable region as a purified, variable region recombinant protein fused to at least one epitope tag;
 - ii) contacting the variable region recombinant protein with a tumor extract;
 - iii) determining binding of a tumor antigen with the variable region recombinant protein; and
 - iv) identifying said tumor antigen.
- 10. The antigen of claim 9, wherein the antibody variable region is expressed in a bacterial expression system.
- 11. A variable region recombinant protein fused to at least one epitope tag prepared by the steps comprising:
 - expressing an antibody variable region as a purified, variable region recombinant protein fused to at least one epitope tag; and
 - isolating the protein.
- 12. An isolated nucleic acid molecule selected from the group consisting of:
 - i) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% homologous to the nucleotide sequence of SEQ ID NO:1, or a complement thereof;
 - ii) a nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 50% homologous to the amino acid sequence of SEQ ID NO:2;

- iii) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2; and
 - iv) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the nucleic acid molecule hybridizes to a complement of a nucleic acid molecule comprising SEQ ID NO:1 under stringent conditions.
13. The isolated nucleic acid molecule of claim 12 which is selected from the group consisting of:
- i) a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or a complement thereof; and
 - ii) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2.
14. An expression vector comprising the nucleic acid molecule of claim 12 or 13.
15. A cell transformed with the expression vector of claim 14.
16. An isolated polypeptide selected from the group consisting of:
- i) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2;
 - ii) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a complement of a nucleic acid molecule comprising SEQ ID NO:1 under stringent conditions;
 - iii) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 50% homologous to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1; and
 - iv) a polypeptide comprising an amino acid sequence which is at least 30% homologous to the amino acid sequence of SEQ ID NO:2.

17. The isolated polypeptide of claim 16 comprising the amino acid sequence of SEQ ID NO:2.
18. An antibody or conjugate thereof which selectively binds to the isolated polypeptide of claim 16 or 17.
19. The antibody of claim 18 which is monoclonal.
20. A composition comprising the variable region recombinant protein of claim 11.
21. A pharmaceutical composition comprising the antibody of claim 18 or 19 and a pharmaceutically acceptable excipient.
22. A fusion protein comprising the antibody of claim 18 or 19.
23. The use of the variable region recombinant protein of claim 11 for detecting and/or quantifying tumor-related antigens present in biological samples.
24. The use of the monoclonal antibody of claim 18 or 19 for detecting and/or quantifying tumor-related antigens present in biological samples.
25. The use of the fusion protein of claim 22 for detecting and/or quantifying tumor-related antigens present in biological samples.
26. The use of the variable region recombinant protein of claim 11 for treating cancer, wherein the recombinant protein targets tumor-specific antigens.
27. The use of the monoclonal antibody of claim 18 or 19 for treating cancer, wherein the antibody targets tumor-specific antigens.
28. The use of the fusion protein of claim 22 for treating cancer, wherein the fusion protein targets tumor-specific antigens.
29. The use of any one of claims 26, 27, or 28, wherein the cancer is primary or disseminated.

30. A method of identifying a tumor cell with a predominance for lobular carcinomas or ductal carcinomas in a subject, comprising:
- i) obtaining a tumor extract or histological section of a tumor;
 - ii) contacting the extract or section with the monoclonal antibody of claim 18 or 19; and
 - iii) measuring binding of said antibody with an antigen in said extract or section;
- wherein the binding identifies a tumor cell with a predominance for lobular carcinomas or ductal carcinomas in the subject.
31. A method of identifying a tumor cell with a predominance for lobular carcinomas or ductal carcinomas in a subject, comprising:
- i) obtaining a tumor extract or histological section of a tumor;
 - ii) contacting the extract or section with the variable region recombinant protein of claim 11; and
 - iii) measuring binding of said protein with an antigen in said extract or section;
- wherein the binding identifies a tumor cell with a predominance for lobular carcinomas or ductal carcinomas in the subject.
32. A method of identifying a tumor cell with a predominance for lobular carcinomas or ductal carcinomas in a subject, comprising:
- i) obtaining a tumor extract or histological section of a tumor;
 - ii) contacting the extract or section with the fusion protein of claim 22;
- and
- iii) measuring binding of said fusion protein with an antigen in said extract or section;
- wherein the binding identifies a tumor cell with a predominance for lobular carcinomas or ductal carcinomas in the subject.
33. The method of any one of claims 30, 31, or 32, wherein the carcinoma is invasive or highly invasive.

34. The method of any one of claims 30, 31, 32, or 33, wherein the subject has breast cancer or other invasive or highly invasive cancer.

35. A method for identifying an antigen associated with a pathological condition in a subject, comprising:

- i) obtaining, from the subject, a tissue sample from a B-cell germinal center positive for histological markers CD20, CD23, and Ki67;
- ii) preparing VH, V λ and V κ germinal center antibody variable region cDNA libraries;
- iii) compiling CDR/FR ratios from sequenced variable region cDNAs from the libraries;
- iv) selecting variable region sequences with CDR/FR ratios greater than about 0.8 in the libraries to identify antibody variable regions with somatic mutations;
- v) expressing the antibody variable region as a purified, variable region recombinant protein fused to at least one epitope tag;
- vi) contacting the variable region recombinant protein with a tissue sample from the subject; and
- vii) determining binding of an antigen with the variable region recombinant protein, thereby identifying an antigen associated with a pathological condition in a subject.

36. The method of claim 35, wherein the pathological condition is selected from the group consisting of: chronic inflammatory disease, autoimmunity, toxin exposure, and bacterial, virus, or parasitic infection.

37. A kit comprising the composition of claim 20 and instructions for administering the composition to identify a specific pathological event in a subject.

38. A tumor antigen comprising the amino acid sequence of SEQ ID NO:2.

38. A kit comprising the pharmaceutical composition of claim 21 and instructions for administering the composition to treat cancer in a subject.

1/15

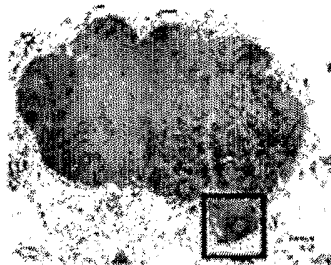


FIG. 1A

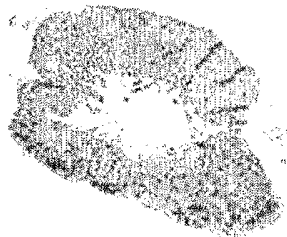


FIG. 1B

2/15

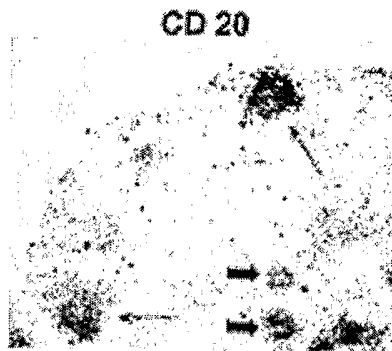


FIG. 2A

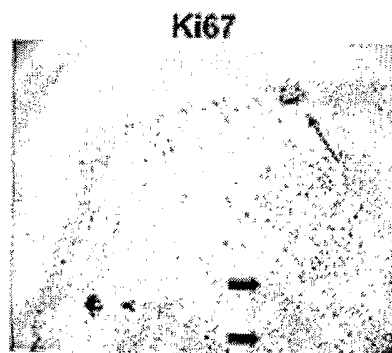


FIG. 2B

3/15

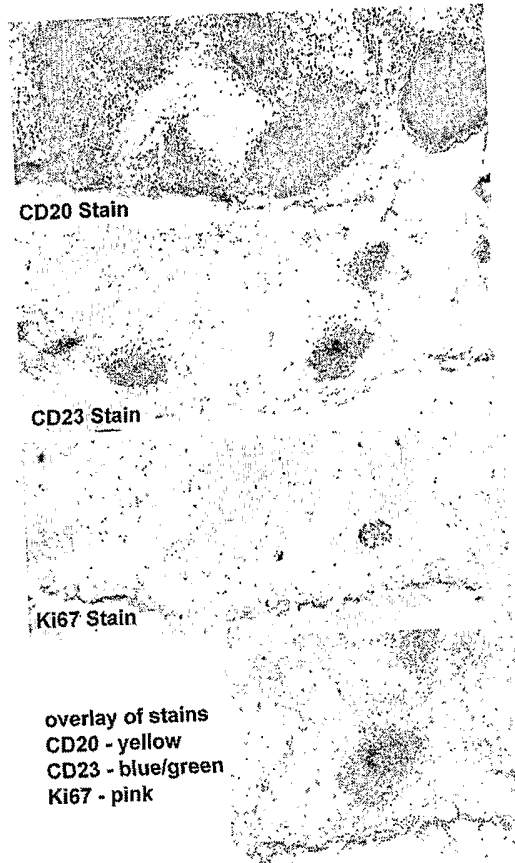


FIG. 3

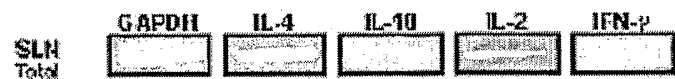


FIG. 4A



FIG. 4B



FIG. 4C

5/15

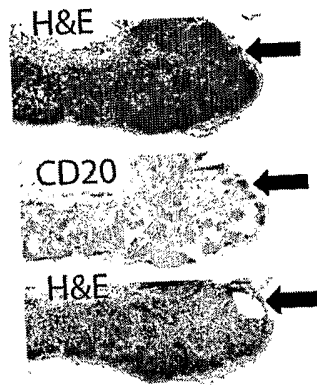


FIG. 5

6/15

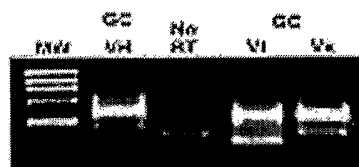


FIG. 6A

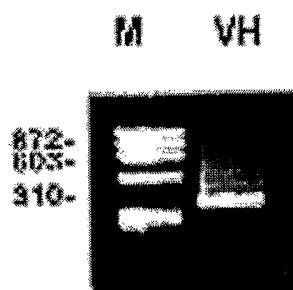


FIG. 6B

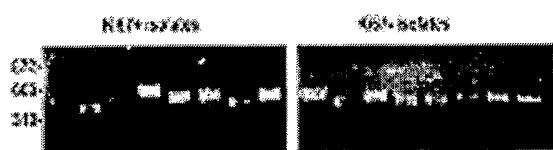


FIG. 6C

7/15

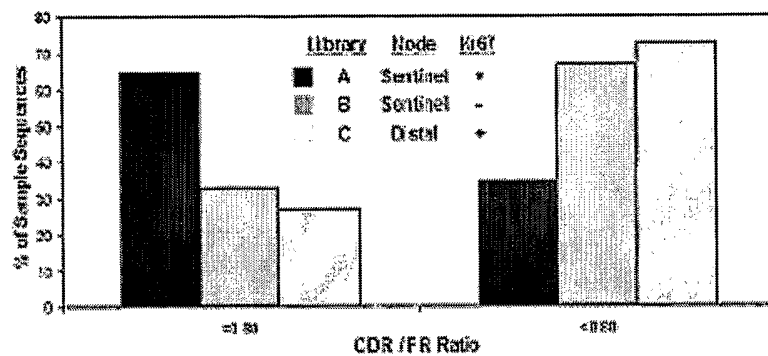


FIG. 7

8/15

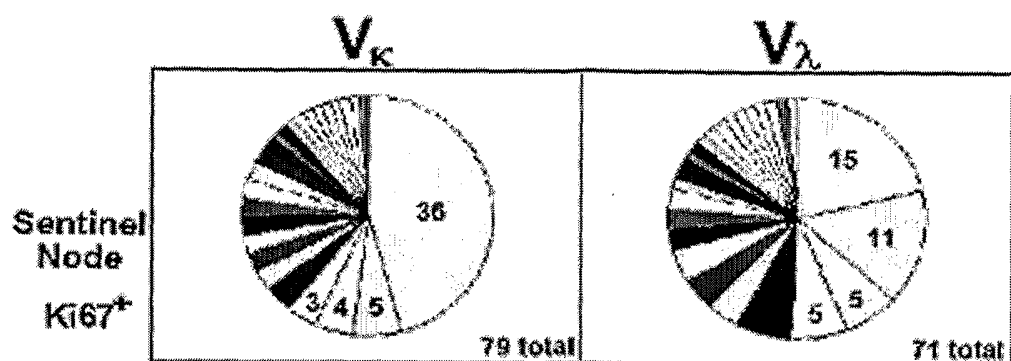


FIG. 8A

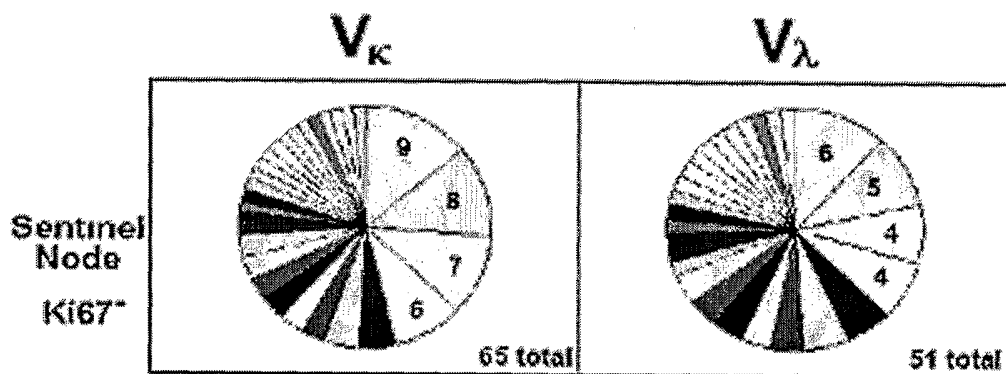


FIG. 8B

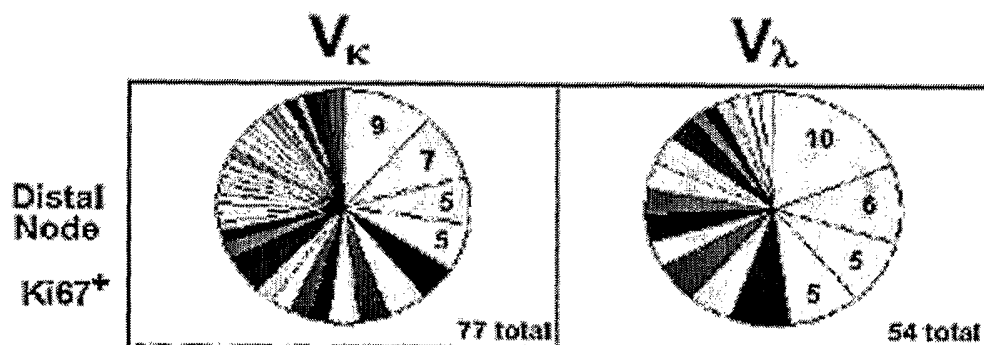


FIG. 8C

9/15

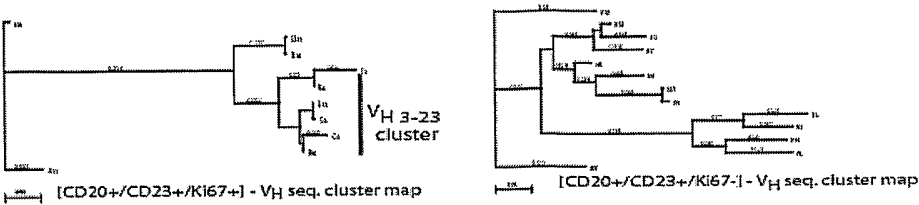


FIG. 9

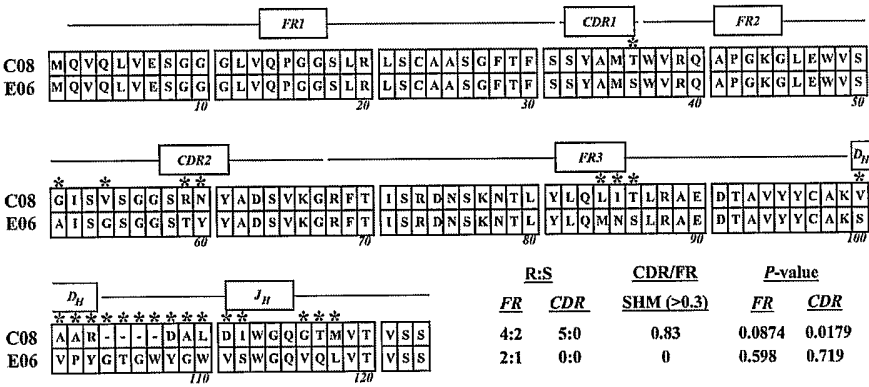


FIG. 10

10/15

VH domain		C-myc	6x-His	
R:S				
FR	CDR	CDR/FR	FR-p	CDR-p
8:2	9:3	1.2	0.024	0.006

FIG. 11A

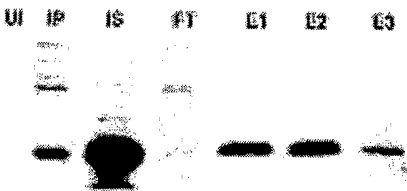


FIG. 11B

11/15

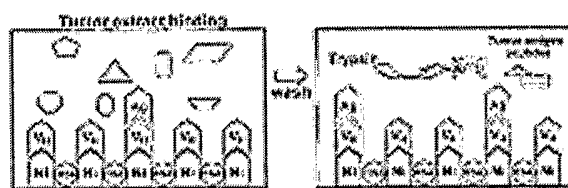


FIG. 12A

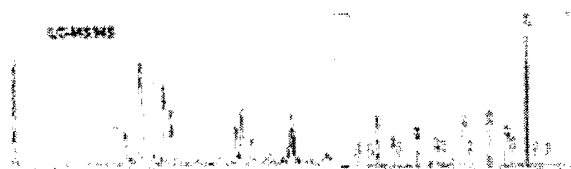


FIG. 12B

Keywords: ERK1/2, angiotensin (Ang), cell cycle, proliferation, growth[illegible]

FIG. 12C

12/15

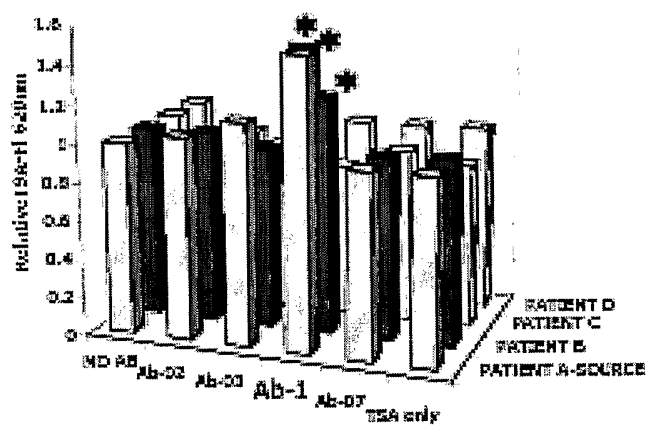


FIG. 13

13/15

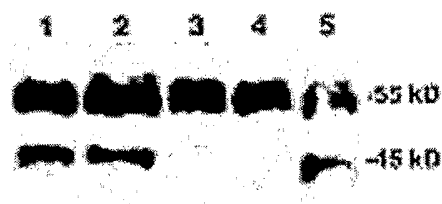


FIG. 14A



FIG. 14B

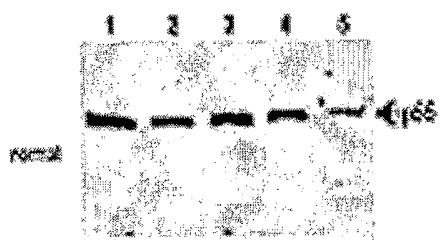


FIG. 14C

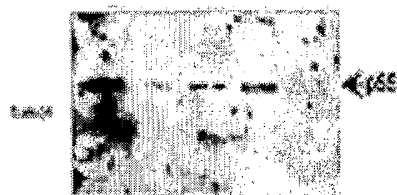


FIG. 14D

14/15

FIG. 15A

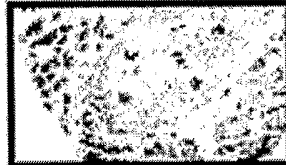


FIG. 15B

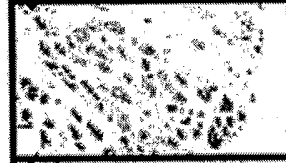


FIG. 15C

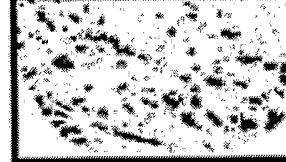


FIG. 15D

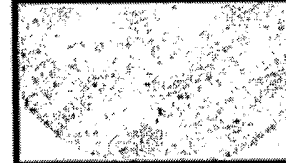


FIG. 15E

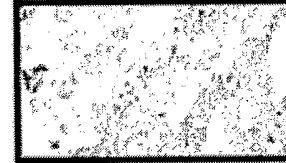
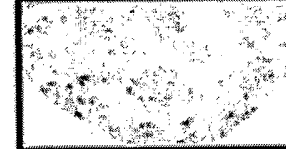


FIG. 15F



15/15

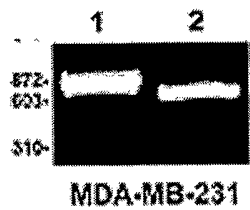


FIG. 16A

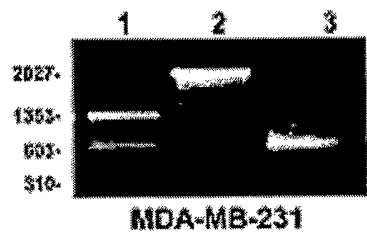


FIG. 16B

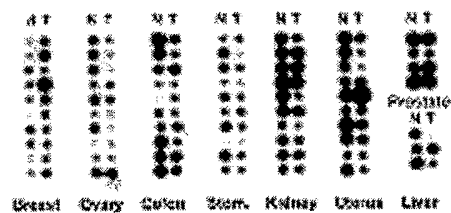


FIG. 16C

1/22

SEQUENCE LISTING

<110> UNIVERSITY OF CONNECTICUT

<120> METHODS AND COMPOSITIONS FOR METASTATIC TUMOR SCREENING

<130> 30471.62424WO

<140>

<141>

<150> 60/645,060

<151> 2005-01-19

<160> 40

<170> PatentIn Ver. 3.3

<210> 1

<211> 398

<212> PRT

<213> Homo sapiens

<400> 1

Met Ser Gly Ser Ser Leu Pro Ser Ala Leu Ala Leu Ser Leu Leu Leu
1 5 10 15

Val Ser Gly Ser Leu Leu Pro Gly Pro Gly Ala Ala Gln Asn Ala Gly
20 25 30

Phe Val Lys Ser Pro Met Ser Glu Thr Lys Leu Thr Gly Asp Ala Phe
35 40 45

Glu Leu Tyr Cys Asp Val Val Gly Ser Pro Thr Pro Glu Ile Gln Trp
50 55 60

Trp Tyr Ala Glu Val Asn Arg Ala Glu Ser Phe Arg Gln Leu Trp Asp
65 70 75 80

Gly Ala Arg Lys Arg Arg Val Thr Val Asn Thr Ala Tyr Gly Ser Asn
85 90 95

Gly Val Ser Val Leu Arg Ile Thr Arg Leu Thr Leu Glu Asp Ser Gly
100 105 110

Thr Tyr Glu Cys Arg Ala Ser Asn Asp Pro Lys Arg Asn Asp Leu Arg
115 120 125

Gln Asn Pro Ser Ile Thr Trp Ile Arg Ala Gln Ala Thr Ile Ser Val
130 135 140

Leu Gln Lys Pro Arg Ile Val Thr Ser Glu Glu Val Ile Ile Arg Asp
145 150 155 160

Ser Pro Val Leu Pro Val Thr Leu Gln Cys Asn Leu Thr Ser Ser Ser
165 170 175

2/22

His Thr Leu Thr Tyr Ser Tyr Trp Thr Lys Asn Gly Val Glu Leu Ser
180 185 190

Ala Thr Arg Lys Asn Ala Ser Asn Met Glu Tyr Arg Ile Asn Lys Pro
195 200 205

Arg Ala Glu Asp Ser Gly Glu Tyr His Cys Val Tyr His Phe Val Ser
210 215 220

Ala Pro Lys Ala Asn Ala Thr Ile Glu Val Lys Ala Ala Pro Asp Ile
225 230 235 240

Thr Gly His Lys Arg Ser Glu Asn Lys Asn Glu Gly Gln Asp Ala Thr
245 250 255

Met Tyr Cys Lys Ser Val Gly Tyr Pro His Pro Asp Trp Ile Trp Arg
260 265 270

Lys Lys Glu Asn Gly Met Pro Met Asp Ile Val Asn Thr Ser Gly Arg
275 280 285

Phe Phe Ile Ile Asn Lys Glu Asn Tyr Thr Glu Leu Asn Ile Val Asn
290 295 300

Leu Gln Ile Thr Glu Asp Pro Gly Glu Tyr Glu Cys Asn Ala Thr Asn
305 310 315 320

Ala Ile Gly Ser Ala Ser Val Val Thr Val Leu Arg Val Arg Ser His
325 330 335

Leu Ala Pro Leu Trp Pro Phe Leu Gly Ile Leu Ala Glu Ile Ile Ile
340 345 350

Leu Val Val Ile Ile Val Val Tyr Glu Lys Arg Lys Arg Pro Asp Glu
355 360 365

Val Pro Asp Asp Asp Glu Pro Ala Gly Pro Met Lys Thr Asn Ser Thr
370 375 380

Asn Asn His Lys Asp Lys Asn Leu Arg Gln Arg Asn Thr Asn
385 390 395

<210> 2

<211> 125

<212> PRT

<213> Homo sapiens

<400> 2

Met Ala Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro
1 5 10 15

Gly Ala Ser Val Lys Val Ser Cys Thr Ala Ser Gly Tyr Thr Phe Ile
20 25 30

Asn Tyr Gly Ile Ile Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu
35 40 45

3/22

Trp Met Gly Trp Ile Asn Gly Tyr Ser Gly His Thr Ser Tyr Ala Gln
 50 55 60
 Lys Phe Gln Asp Lys Val Thr Met Thr Thr Asp Thr Ser Thr Ser Thr
 65 70 75 80
 Val Tyr Met Glu Leu Arg Asn Leu Gly Ser Asp Asp Thr Ala Val Tyr
 85 90 95
 Tyr Cys Ala Arg Gly Ser Ser Gly Arg Glu Leu Leu Pro Phe Asp His
 100 105 110
 Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ser Arg Ser
 115 120 125

<210> 3
 <211> 158
 <212> PRT
 <213> Homo sapiens

<400> 3
 Met Ala Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro
 1 5 10 15
 Gly Ala Ser Val Lys Val Ser Cys Thr Ala Ser Gly Tyr Thr Phe Ile
 20 25 30
 Asn Tyr Gly Ile Ile Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu
 35 40 45
 Trp Met Gly Trp Ile Asn Gly Tyr Ser Gly His Thr Ser Tyr Ala Gln
 50 55 60
 Lys Phe Gln Asp Lys Val Thr Met Thr Thr Asp Thr Ser Thr Ser Thr
 65 70 75 80
 Val Tyr Met Glu Leu Arg Asn Leu Gly Ser Asp Asp Thr Ala Val Tyr
 85 90 95
 Tyr Cys Ala Arg Gly Ser Ser Gly Arg Glu Leu Leu Pro Phe Asp His
 100 105 110
 Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ser Arg Ser Ala Ala Gly
 115 120 125
 Thr Ile Trp Glu Phe Glu Ala Tyr Val Glu Gln Lys Leu Ile Ser Glu
 130 135 140
 Glu Asp Leu Asn Ser Ala Val Asp His His His His His His
 145 150 155

<210> 4
 <211> 122
 <212> PRT
 <213> Homo sapiens

4/22

<400> 4

```

Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly
 1             5             10             15

Ala Ser Val Lys Val Ser Cys Thr Ala Ser Gly Tyr Thr Phe Ile Asn
      20             25             30

Tyr Gly Ile Ile Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu Trp
      35             40             45

Met Gly Trp Ile Asn Gly Tyr Ser Gly His Thr Ser Tyr Ala Gln Lys
      50             55             60

Phe Gln Asp Lys Val Thr Met Thr Thr Asp Thr Ser Thr Ser Thr Val
      65             70             75             80

Tyr Met Glu Leu Arg Asn Leu Gly Ser Asp Asp Thr Ala Val Tyr Tyr
      85             90             95

Cys Ala Arg Gly Ser Ser Gly Arg Glu Leu Leu Pro Phe Asp His Trp
      100            105            110

Gly Gln Gly Thr Met Val Thr Val Ser Ser
      115            120

```

<210> 5

<211> 152

<212> PRT

<213> Homo sapiens

<400> 5

```

Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly
 1             5             10             15

Ala Ser Val Lys Val Ser Cys Thr Ala Ser Gly Tyr Thr Phe Ile Asn
      20             25             30

Tyr Gly Ile Ile Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu Trp
      35             40             45

Met Gly Trp Ile Asn Gly Tyr Ser Gly His Thr Ser Tyr Ala Gln Lys
      50             55             60

Phe Gln Asp Lys Val Thr Met Thr Thr Asp Thr Ser Thr Ser Thr Val
      65             70             75             80

Tyr Met Glu Leu Arg Asn Leu Gly Ser Asp Asp Thr Ala Val Tyr Tyr
      85             90             95

Cys Ala Arg Gly Ser Ser Gly Arg Glu Leu Leu Pro Phe Asp His Trp
      100            105            110

Gly Gln Gly Thr Met Val Thr Val Ser Ser Lys Gly Asn Ser Lys Leu
      115            120            125

Glu Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr Arg
      130            135            140

```

5/22

Thr Gly His His His His His His
145 150

<210> 6
<211> 122
<212> PRT
<213> Homo sapiens

<400> 6
Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly
1 5 10 15
Ala Ser Val Lys Val Ser Cys Thr Ala Ser Gly Tyr Thr Phe Ile Asn
20 25 30
Tyr Gly Ile Ile Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu Trp
35 40 45
Met Gly Trp Ile Asn Gly Tyr Ser Gly His Thr Ser Tyr Ala Gln Lys
50 55 60
Phe Gln Asp Lys Val Thr Met Thr Thr Asp Thr Ser Thr Ser Thr Val
65 70 75 80
Tyr Met Glu Leu Arg Asn Leu Gly Ser Asp Asp Thr Ala Val Tyr Tyr
85 90 95
Cys Ala Arg Gly Ser Ser Gly Arg Glu Leu Leu Pro Phe Asp His Trp
100 105 110
Gly Gln Gly Thr Leu Val Thr Val Ser Ser
115 120

<210> 7
<211> 152
<212> PRT
<213> Homo sapiens

<400> 7
Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly
1 5 10 15
Ala Ser Val Lys Val Ser Cys Thr Ala Ser Gly Tyr Thr Phe Ile Asn
20 25 30
Tyr Gly Ile Ile Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu Trp
35 40 45
Met Gly Trp Ile Asn Gly Tyr Ser Gly His Thr Ser Tyr Ala Gln Lys
50 55 60
Phe Gln Asp Lys Val Thr Met Thr Thr Asp Thr Ser Thr Ser Thr Val
65 70 75 80

Tyr	Met	Glu	Leu	Arg 85	Asn	Leu	Gly	Ser	Asp 90	Asp	Thr	Ala	Val	Tyr 95	Tyr
Cys	Ala	Arg	Gly 100	Ser	Ser	Gly	Arg	Glu 105	Leu	Leu	Pro	Phe	Asp 110	His	Trp
Gly	Gln	Gly 115	Thr	Leu	Val	Thr	Val 120	Ser	Ser	Lys	Gly	Asn 125	Ser	Lys	Leu
Glu	Gly 130	Lys	Pro	Ile	Pro	Asn 135	Pro	Leu	Leu	Gly	Leu 140	Asp	Ser	Thr	Arg
Thr 145	Gly	His	His	His 150	His	His	His								

```
<210> 8
<211> 120
<212> PRT
<213> Homo sapiens
```

```
<220>  
<221> MOD_RES  
<222> (47)  
<223> Variable amino acid
```

```

<400> 8
Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly
  1             5             10             15

Ala Ser Val Arg Val Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Ser
          20             25             30

Tyr Gly Ile Thr Trp Val Arg Gln Ala Pro Gly Gln Lys Leu Xaa Trp
      35             40             45

Leu Gly Trp Ile Ser Ala Tyr Asn Gly Asn Thr His Tyr Ala Gln Lys
  50             55             60

Leu Gln Gly Arg Val Thr Met Thr Thr Asp Thr Ser Thr Thr Thr Ala
  65             70             75             80

Tyr Met Glu Leu Arg Ser Leu Arg Ser Asp Asp Thr Ala Val Tyr Tyr
          85             90             95

Cys Ala Arg Val Glu Ala Ala Phe Asp Leu Phe Asp Tyr Trp Gly Gln
          100             105             110

Gly Thr Met Val Thr Val Ser Ser
      115             120

```

```
<210> 9
<211> 150
<212> PRT
<213> Homo sapiens
```

7/22

<220>

<221> MOD_RES

<222> (47)

<223> Variable amino acid

<400> 9

Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly
 1 5 10 15

Ala Ser Val Arg Val Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Ser
 20 25 30

Tyr Gly Ile Thr Trp Val Arg Gln Ala Pro Gly Gln Lys Leu Xaa Trp
 35 40 45

Leu Gly Trp Ile Ser Ala Tyr Asn Gly Asn Thr His Tyr Ala Gln Lys
 50 55 60

Leu Gln Gly Arg Val Thr Met Thr Thr Asp Thr Ser Thr Thr Thr Ala
 65 70 75 80

Tyr Met Glu Leu Arg Ser Leu Arg Ser Asp Asp Thr Ala Val Tyr Tyr
 85 90 95

Cys Ala Arg Val Glu Ala Ala Phe Asp Leu Phe Asp Tyr Trp Gly Gln
 100 105 110

Gly Thr Met Val Thr Val Ser Ser Lys Gly Asn Ser Lys Leu Glu Gly
 115 120 125

Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr Arg Thr Gly
 130 135 140

His His His His His His
 145 150

<210> 10

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 primer

<400> 10

caccatggc aaattccatg

20

<210> 11

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 primer

8/22

<400> 11
gcttgacaaa gtggtcgttg 20

<210> 12
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 12
ggacacaagt gcgatatcac c 21

<210> 13
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 13
atttctctct catgatcgtc 20

<210> 14
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 14
gcctaacaatg cttcgagatc 20

<210> 15
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 15
ctcatggctt thtagatgcc 20

9/22

<210> 16
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 16
tgaggagacg gtgaccakgg t

21

<210> 17
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 17
ccctgggtct taagtgaaag

20

<210> 18
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 18
caagttatat cttggctttt cagc

24

<210> 19
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 19
ctgggatgct cttcgacctc g

21

<210> 20
<211> 26
<212> DNA
<213> Artificial Sequence

10/22

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 20

atgcaggtgc agctggtgsa gtctgg

26

<210> 21

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 21

tgaggagacg gtgaccakgg t

21

<210> 22

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 22

cagtctgtgc tgackcagcc rcc

23

<210> 23

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 23

kaggacggtc agctkggtsc

20

<210> 24

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 24

gamattgtgm tgacscagtc tcc

23

11/22

<210> 25
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 25
tttgatytcc accttggtcc 20

<210> 26
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 26
cgcgaaatta atacgactca ctataggg 28

<210> 27
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 27
cctaaattgt aagcgtaat ccgg 24

<210> 28
<211> 119
<212> PRT
<213> Homo sapiens

<400> 28
Met Gln Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly
1 5 10 15
Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser
20 25 30
Tyr Ala Met Thr Trp Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp
35 40 45
Val Ser Gly Ile Ser Val Ser Gly Gly Ser Arg Asn Tyr Ala Asp Ser
50 55 60

12/22

Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu
65 70 75 80

Tyr Leu Gln Leu Ile Thr Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr
85 90 95

Cys Ala Lys Val Ala Ala Arg Asp Ala Leu Asp Ile Trp Gly Gln Gly
100 105 110

Thr Met Val Thr Val Ser Ser
115

<210> 29
<211> 123
<212> PRT
<213> Homo sapiens

<400> 29
Met Gln Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val Gln Pro Gly
1 5 10 15

Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser
20 25 30

Tyr Ala Met Ser Gln Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp
35 40 45

Val Ser Ala Ile Ser Gly Ser Gly Gly Ser Thr Tyr Tyr Ala Asp Ser
50 55 60

Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu
65 70 75 80

Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala Val Tyr Tyr
85 90 95

Cys Ala Lys Ser Val Pro Tyr Gly Thr Gly Trp Tyr Gly Trp Val Ser
100 105 110

Trp Gly Gln Val Gln Leu Val Thr Val Ser Ser
115 120

<210> 30
<211> 398
<212> PRT
<213> Homo sapiens

<400> 30
Met Ser Gly Ser Ser Leu Pro Ser Ala Leu Ala Leu Ser Leu Leu Leu
1 5 10 15

Val Ser Gly Ser Leu Leu Pro Gly Pro Gly Ala Ala Gln Asn Ala Gly
20 25 30

Phe	Val	Lys	Ser	Pro	Met	Ser	Glu	Thr	Lys	Leu	Thr	Gly	Asp	Ala	Phe
35						40						45			
Glu	Leu	Tyr	Cys	Asp	Val	Val	Gly	Ser	Pro	Thr	Pro	Glu	Ile	Gln	Trp
50						55						60			
Trp	Tyr	Ala	Glu	Val	Asn	Arg	Ala	Glu	Ser	Phe	Arg	Gln	Leu	Trp	Asp
65						70						75			
Gly	Ala	Arg	Lys	Arg	Arg	Val	Thr	Val	Asn	Thr	Ala	Tyr	Gly	Ser	Asn
			85						90			95			
Gly	Val	Ser	Val	Leu	Arg	Ile	Thr	Arg	Leu	Thr	Leu	Glu	Asp	Ser	Gly
			100						105			110			
Thr	Tyr	Glu	Cys	Arg	Ala	Ser	Asn	Asp	Pro	Lys	Arg	Asn	Asp	Leu	Arg
115						120						125			
Gln	Asn	Pro	Ser	Ile	Thr	Trp	Ile	Arg	Ala	Gln	Ala	Thr	Ile	Ser	Val
130						135						140			
Leu	Gln	Lys	Pro	Arg	Ile	Val	Thr	Ser	Glu	Glu	Val	Ile	Ile	Arg	Asp
145						150						155			
Ser	Pro	Val	Leu	Pro	Val	Thr	Leu	Gln	Cys	Asn	Leu	Thr	Ser	Ser	Ser
			165						170			175			
His	Thr	Leu	Thr	Tyr	Ser	Tyr	Trp	Thr	Lys	Asn	Gly	Val	Glu	Leu	Ser
			180						185			190			
Ala	Thr	Arg	Lys	Asn	Ala	Ser	Asn	Met	Glu	Tyr	Arg	Ile	Asn	Lys	Pro
195						200						205			
Arg	Ala	Glu	Asp	Ser	Gly	Glu	Tyr	His	Cys	Val	Tyr	His	Phe	Val	Ser
210						215						220			
Ala	Pro	Lys	Ala	Asn	Ala	Thr	Ile	Glu	Val	Lys	Ala	Ala	Pro	Asp	Ile
225						230						235			
Thr	Gly	His	Lys	Arg	Ser	Glu	Asn	Lys	Asn	Glu	Gly	Gln	Asp	Ala	Thr
			245						250			255			
Met	Tyr	Cys	Lys	Ser	Val	Gly	Tyr	Pro	His	Pro	Asp	Trp	Ile	Trp	Arg
			260						265			270			
Lys	Lys	Glu	Asn	Gly	Met	Pro	Met	Asp	Ile	Val	Asn	Thr	Ser	Gly	Arg
275						280						285			
Phe	Phe	Ile	Ile	Asn	Lys	Glu	Asn	Tyr	Thr	Glu	Leu	Asn	Ile	Val	Asn
290						295						300			
Leu	Gln	Ile	Thr	Glu	Asp	Pro	Gly	Glu	Tyr	Glu	Cys	Asn	Ala	Thr	Asn
305						310						315			
Ala	Ile	Gly	Ser	Ala	Ser	Val	Val	Thr	Val	Leu	Arg	Val	Arg	Ser	His
			325						330			335			

14/22

Leu Ala Pro Leu Trp Pro Phe Leu Gly Ile Leu Ala Glu Ile Ile Ile
 340 345 350

Leu Val Val Ile Ile Val Val Tyr Glu Lys Arg Lys Arg Pro Asp Glu
 355 360 365

Val Pro Asp Asp Asp Glu Pro Ala Gly Pro Met Lys Thr Asn Ser Thr
 370 375 380

Asn Asn His Lys Asp Lys Asn Leu Arg Gln Arg Asn Thr Asn
 385 390 395

<210> 31

<211> 2388

<212> DNA

<213> Homo sapiens

<400> 31

```

gaaggacgga gccgagccgc ggctgcctcc ctgcgtcact ccctcgcgca ctgcgccgcc 60
ccctccctcc ctccctccctc ttcccccgggc ccgggtcttg ccccggccca ttgcgtgttg 120
ggtctttctgc tagggaggat gtccgggttcg tcgctgcccc gcgccctggc cctctcgctg 180
ttgctggtct ctggctccct cctcccaggg ccaggcgccg ctcagaacgc tgggtttgtc 240
aagtcgcccc taagctcacg ggggacgcct ttgagctgta ctgtgacgtg 300
gtccggagcc ccacgccaga gatccagtgg tggtagcgag aagtcaaccg ggcagagtct 360
ttcagacagc tgtgggacgg tgctcggaag cgccgtgtca ccgtaaacac cgcctacggg 420
tcaaacggcg tgagtgtgct gagaataacc cggctcacct tggaggactc tgggacttac 480
gagtgcaggg ccagcaacga cccaagagg aatgacttga ggcaaaaccc ctccataaca 540
tggattcgag ccaggccac cataagcgctc cttcagaagc caaggattgt caccagtga 600
gaggtcatta ttcgagacag ccctgttctc cctgtcacc tgcagtgtaa cctcacctcc 660
agctctcaca cccttaacata cagctactgg acaaagaatg gggtggaact gagtgcact 720
cgtaagaatg ccagcaacat ggagtacagg atcaataagc cgagagctga ggattcaggc 780
gaataccact gcgtatatca ctttgtcagc gctcctaaag caaacgccac cattgaagt 840
aaagccgctc ctgacatcac tggccataaa cggagtgaga acaagaatga agggcaggat 900
gccactatgt attgcaagtc agttggctac cccacccag actggatatg gcgcaagaag 960
gagaacggga tgcccatgga cattgtcaat acctctggcc gcttcttcat catcaacaag 1020
gaaaattaca ctgagttgaa cattgtgaac ctgcagatca cggaagaccc tggcgagtat 1080
gaatgtaatg ccaccaacgc cattggctcg ccctctgttg tcaactgtcc cagggtgcgg 1140
agccacctgg cccactctg gcctttcttg ggaattctgg ctgaaattat catccttgtg 1200
gtgatcattg ttgtgtatga gaagaggaag aggccagatg aggttctctga cgatgatgaa 1260
ccagctggac caatgaaaac caactctacc aacaatcaca aagataaaaa cttgcgccag 1320
agaaacacaa attaatgact gcttacaata tctttagggt cctgaaactg gtggcaacat 1380
gacctgctaa aattttctgc ttggacctct ttggttctct cccctttcaa gtgagcaaca 1440
ccacaatgac tgtctaaagc atgccttatt tagcctctcc tgtaagggtg atctagccag 1500
gtacatttta aacaatgctt cagtgtagaa ggtgtaaact attttgggct tgatgtgctg 1560
tgaatgttgc tttttttttt cctttgttaa aatatttaaa tagaagtga aaggctctct 1620
gaggatcaga tcatgcatgc gccatttttt acttaatgca gctgttaaat tggcaaagct 1680
ctaaaatgca ctgctgccat ctagtgtatc acttttgtaa agtacagcaa aacctacagg 1740
tatatacagc atataaatat atatatatat atatttatat ttttgggggt gggagaaatc 1800
caaaataaag taaatgcttg tttcattttt aagctgctga tattcattcc ttattgtatg 1860
ttgtcagatg aggaaattgt gcagttctgg tacataaaga tgagtaatat aaactgaaat 1920
ctataatttt aagggtctaa cctgtgactt taataagctg gaacagtcca ctgaatgggt 1980
ataatgaatt gcagtatata cgtatgattg ctttttaagt gattatcttt tcttctgtta 2040
agtcatgtaa attcataaat ccttttgca cgtatgtttg aaccttattc ttgtacattc 2100
attcaatcaa ggcaaacctt tataattttt cttttgtttc caatgacctt gaaatgttat 2160
agcatggtaa tattctatgc aactatagtt atactttttg gtttgacact gtattttttc 2220
acattgatth actggttgat gatagatttt ataacctaac ggttctcatg cgggtgcgta 2280
ttgtagatgc atgtacttgt gtgttttgtg taactattga agtgcaatga tgtataaaaa 2340

```

15/22

agtggtattca cctgttttta aaaataaaac attgataaaa aaaaaaaaa

2388

<210> 32

<211> 375

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)..(375)

<400> 32

atg	gcg	cag	gtg	cag	ctg	gtg	gag	tct	gga	gct	gag	gtg	aag	aag	cct	48
Met	Ala	Gln	Val	Gln	Leu	Val	Glu	Ser	Gly	Ala	Glu	Val	Lys	Lys	Pro	
1				5					10					15		

ggg	gcc	tca	gtg	aag	gtc	tcc	tgc	acg	gct	tct	ggt	tac	acc	ttt	atc	96
Gly	Ala	Ser	Val	Lys	Val	Ser	Cys	Thr	Ala	Ser	Gly	Tyr	Thr	Phe	Ile	
		20						25					30			

aat	tac	ggt	atc	atc	tgg	gtg	cga	cag	gcc	cct	gga	caa	gga	ctt	gag	144
Asn	Tyr	Gly	Ile	Ile	Trp	Val	Arg	Gln	Ala	Pro	Gly	Gln	Gly	Leu	Glu	
		35					40					45				

tgg	atg	gga	tgg	atc	aac	ggt	tac	agt	ggt	cac	aca	agc	tat	gca	cag	192
Trp	Met	Gly	Trp	Ile	Asn	Gly	Tyr	Ser	Gly	His	Thr	Ser	Tyr	Ala	Gln	
	50					55					60					

aaa	ttc	cag	gac	aaa	gtc	acc	atg	acc	aca	gac	aca	tcc	acg	agc	aca	240
Lys	Phe	Gln	Asp	Lys	Val	Thr	Met	Thr	Thr	Asp	Thr	Ser	Thr	Ser	Thr	
65					70					75					80	

gtc	tac	atg	gag	ctg	agg	aac	ctg	gga	tct	gac	gac	acg	gcc	gtc	tat	288
Val	Tyr	Met	Glu	Leu	Arg	Asn	Leu	Gly	Ser	Asp	Asp	Thr	Ala	Val	Tyr	
				85					90					95		

tac	tgt	gcg	aga	gga	tca	tcg	ggg	cgg	gaa	cta	ctc	ccg	ttt	gac	cac	336
Tyr	Cys	Ala	Arg	Gly	Ser	Ser	Gly	Arg	Glu	Leu	Leu	Pro	Phe	Asp	His	
		100						105						110		

tgg	ggc	cag	gga	acc	ctg	gtc	acc	gtc	tcc	tca	aga	tct				375
Trp	Gly	Gln	Gly	Thr	Leu	Val	Thr	Val	Ser	Ser	Arg	Ser				
	115						120					125				

<210> 33

<211> 477

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)..(474)

16/22

<400> 33

```

atg gcg cag gtg cag ctg gtg gag tct gga gct gag gtg aag aag cct 48
Met Ala Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro
  1           5           10           15

ggg gcc tca gtg aag gtc tcc tgc acg gct tct ggt tac acc ttt atc 96
Gly Ala Ser Val Lys Val Ser Cys Thr Ala Ser Gly Tyr Thr Phe Ile
          20           25           30

aat tac ggt atc atc tgg gtg cga cag gcc cct gga caa gga ctt gag 144
Asn Tyr Gly Ile Ile Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu
          35           40           45

tgg atg gga tgg atc aac ggt tac agt ggt cac aca agc tat gca cag 192
Trp Met Gly Trp Ile Asn Gly Tyr Ser Gly His Thr Ser Tyr Ala Gln
          50           55           60

aaa ttc cag gac aaa gtc acc atg acc aca gac aca tcc acg agc aca 240
Lys Phe Gln Asp Lys Val Thr Met Thr Thr Asp Thr Ser Thr Ser Thr
          65           70           75           80

gtc tac atg gag ctg agg aac ctg gga tct gac gac acg gcc gtc tat 288
Val Tyr Met Glu Leu Arg Asn Leu Gly Ser Asp Asp Thr Ala Val Tyr
          85           90           95

tac tgt gcg aga gga tca tcg ggg cgg gaa cta ctc ccg ttt gac cac 336
Tyr Cys Ala Arg Gly Ser Ser Gly Arg Glu Leu Leu Pro Phe Asp His
          100          105          110

tgg ggc cag gga acc ctg gtc acc gtc tcc tca aga tct gca gct ggt 384
Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ser Arg Ser Ala Ala Gly
          115          120          125

acc ata tgg gaa ttc gaa gct tac gta gaa caa aaa ctc atc tca gaa 432
Thr Ile Trp Glu Phe Glu Ala Tyr Val Glu Gln Lys Leu Ile Ser Glu
          130          135          140

gag gat ctg aat agc gcc gtc gac cat cat cat cat cat cat tga 477
Glu Asp Leu Asn Ser Ala Val Asp His His His His His His
          145          150          155

```

<210> 34

<211> 366

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)..(366)

<400> 34

```

atg cag gtg cag ctg gtg gag tct gga gct gag gtg aag aag cct ggg 48
Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly
  1           5           10           15

```

17/22

```

gcc tca gtg aag gtc tcc tgc acg gct tct ggt tac acc ttt atc aat 96
Ala Ser Val Lys Val Ser Cys Thr Ala Ser Gly Tyr Thr Phe Ile Asn
          20                25                30

tac ggt atc atc tgg gtg cga cag gcc cct gga caa gga ctt gag tgg 144
Tyr Gly Ile Ile Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu Trp
          35                40                45

atg gga tgg atc aac ggt tac agt ggt cac aca agc tat gca cag aaa 192
Met Gly Trp Ile Asn Gly Tyr Ser Gly His Thr Ser Tyr Ala Gln Lys
          50                55                60

ttc cag gac aaa gtc acc atg acc aca gac aca tcc acg agc aca gtc 240
Phe Gln Asp Lys Val Thr Met Thr Thr Asp Thr Ser Thr Ser Thr Val
          65                70                75                80

tac atg gag ctg agg aac ctg gga tct gac gac acg gcc gtc tat tac 288
Tyr Met Glu Leu Arg Asn Leu Gly Ser Asp Asp Thr Ala Val Tyr Tyr
          85                90                95

tgt gcg aga gga tca tcg ggg cgg gaa cta ctc ccg ttt gac cac tgg 336
Cys Ala Arg Gly Ser Ser Gly Arg Glu Leu Leu Pro Phe Asp His Trp
          100                105                110

ggc cag gga acc atg gtc acc gtc tcc tca 366
Gly Gln Gly Thr Met Val Thr Val Ser Ser
          115                120

```

<210> 35

<211> 459

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)..(456)

<400> 35

```

atg cag gtg cag ctg gtg gag tct gga gct gag gtg aag aag cct ggg 48
Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly
          1                5                10                15

gcc tca gtg aag gtc tcc tgc acg gct tct ggt tac acc ttt atc aat 96
Ala Ser Val Lys Val Ser Cys Thr Ala Ser Gly Tyr Thr Phe Ile Asn
          20                25                30

tac ggt atc atc tgg gtg cga cag gcc cct gga caa gga ctt gag tgg 144
Tyr Gly Ile Ile Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu Trp
          35                40                45

atg gga tgg atc aac ggt tac agt ggt cac aca agc tat gca cag aaa 192
Met Gly Trp Ile Asn Gly Tyr Ser Gly His Thr Ser Tyr Ala Gln Lys
          50                55                60

ttc cag gac aaa gtc acc atg acc aca gac aca tcc acg agc aca gtc 240
Phe Gln Asp Lys Val Thr Met Thr Thr Asp Thr Ser Thr Ser Thr Val
          65                70                75                80

```

18/22

tac atg gag ctg agg aac ctg gga tct gac gac acg gcc gtc tat tac 288
 Tyr Met Glu Leu Arg Asn Leu Gly Ser Asp Asp Thr Ala Val Tyr Tyr
 85 90 95

tgt gcg aga gga tca tcg ggg cgg gaa cta ctc ccg ttt gac cac tgg 336
 Cys Ala Arg Gly Ser Ser Gly Arg Glu Leu Leu Pro Phe Asp His Trp
 100 105 110

ggc cag gga acc atg gtc acc gtc tcc tca aag ggc aat tcg aag ctt 384
 Gly Gln Gly Thr Met Val Thr Val Ser Ser Lys Gly Asn Ser Lys Leu
 115 120 125

gaa ggt aag cct atc cct aac cct ctc ctc ggt ctc gat tct acg cgt 432
 Glu Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr Arg
 130 135 140

acc ggt cat cat cac cat cac cat tga 459
 Thr Gly His His His His His His
 145 150

<210> 36

<211> 366

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)..(366)

<400> 36

atg cag gtg cag ctg gtg gag tct gga gct gag gtg aag aag cct ggg 48
 Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly
 1 5 10 15

gcc tca gtg aag gtc tcc tgc acg gct tct ggt tac acc ttt atc aat 96
 Ala Ser Val Lys Val Ser Cys Thr Ala Ser Gly Tyr Thr Phe Ile Asn
 20 25 30

tac ggt atc atc tgg gtg cga cag gcc cct gga caa gga ctt gag tgg 144
 Tyr Gly Ile Ile Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu Trp
 35 40 45

atg gga tgg atc aac ggt tac agt ggt cac aca agc tat gca cag aaa 192
 Met Gly Trp Ile Asn Gly Tyr Ser Gly His Thr Ser Tyr Ala Gln Lys
 50 55 60

ttc cag gac aaa gtc acc atg acc aca gac aca tcc acg agc aca gtc 240
 Phe Gln Asp Lys Val Thr Met Thr Thr Asp Thr Ser Thr Ser Thr Val
 65 70 75 80

tac atg gag ctg agg aac ctg gga tct gac gac acg gcc gtc tat tac 288
 Tyr Met Glu Leu Arg Asn Leu Gly Ser Asp Asp Thr Ala Val Tyr Tyr
 85 90 95

19/22

```

tgt gcg aga gga tca tcg ggg cgg gaa cta ctc ccg ttt gac cac tgg 336
Cys Ala Arg Gly Ser Ser Gly Arg Glu Leu Leu Pro Phe Asp His Trp
      100              105              110

ggc cag gga acc ctg gtc acc gtc tcc tca 366
Gly Gln Gly Thr Leu Val Thr Val Ser Ser
      115              120

<210> 37
<211> 459
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(456)

<400> 37
atg cag gtg cag ctg gtg gag tct gga gct gag gtg aag aag cct ggg 48
Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly
      1              5              10              15

gcc tca gtg aag gtc tcc tgc acg gct tct ggt tac acc ttt atc aat 96
Ala Ser Val Lys Val Ser Cys Thr Ala Ser Gly Tyr Thr Phe Ile Asn
      20              25              30

tac ggt atc atc tgg gtg cga cag gcc cct gga caa gga ctt gag tgg 144
Tyr Gly Ile Ile Trp Val Arg Gln Ala Pro Gly Gln Gly Leu Glu Trp
      35              40              45

atg gga tgg atc aac ggt tac agt ggt cac aca agc tat gca cag aaa 192
Met Gly Trp Ile Asn Gly Tyr Ser Gly His Thr Ser Tyr Ala Gln Lys
      50              55              60

ttc cag gac aaa gtc acc atg acc aca gac aca tcc acg agc aca gtc 240
Phe Gln Asp Lys Val Thr Met Thr Thr Asp Thr Ser Thr Ser Thr Val
      65              70              75              80

tac atg gag ctg agg aac ctg gga tct gac gac acg gcc gtc tat tac 288
Tyr Met Glu Leu Arg Asn Leu Gly Ser Asp Asp Thr Ala Val Tyr Tyr
      85              90              95

tgt gcg aga gga tca tcg ggg cgg gaa cta ctc ccg ttt gac cac tgg 336
Cys Ala Arg Gly Ser Ser Gly Arg Glu Leu Leu Pro Phe Asp His Trp
      100              105              110

ggc cag gga acc ctg gtc acc gtc tcc tca aag ggc aat tcg aag ctt 384
Gly Gln Gly Thr Leu Val Thr Val Ser Ser Lys Gly Asn Ser Lys Leu
      115              120              125

gaa ggt aag cct atc cct aac cct ctc ctc ggt ctc gat tct acg cgt 432
Glu Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr Arg
      130              135              140

acc ggt cat cat cac cat cac cat tga 459
Thr Gly His His His His His His
      145              150

```

20/22

<210> 38
 <211> 360
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (1)..(360)

<220>
 <221> modified_base
 <222> (140)..(141)
 <223> a, c, t, g, unknown or other

<400> 38
 atg cag gtg cag ctg gtg gag tct gga gct gag gtg aag aag cct ggg 48
 Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly
 1 5 10 15
 gcc tca gtg agg gtc tcc tgc aag gct tct ggt tac acc ttt acc agc 96
 Ala Ser Val Arg Val Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Ser
 20 25 30
 tac ggt atc act tgg gtg cga cag gcc cct gga caa aaa ctt gnn tgg 144
 Tyr Gly Ile Thr Trp Val Arg Gln Ala Pro Gly Gln Lys Leu Xaa Trp
 35 40 45
 ctg gga tgg atc agc gct tac aat ggt aac aca cat tat gca cag aag 192
 Leu Gly Trp Ile Ser Ala Tyr Asn Gly Asn Thr His Tyr Ala Gln Lys
 50 55 60
 ctc cag ggc aga gtc acc atg acc aca gac aca tcc acg acc aca gcc 240
 Leu Gln Gly Arg Val Thr Met Thr Thr Asp Thr Ser Thr Thr Thr Ala
 65 70 75 80
 tac atg gag ctt agg agc ctg aga tct gac gac acg gcc gtg tat tac 288
 Tyr Met Glu Leu Arg Ser Leu Arg Ser Asp Asp Thr Ala Val Tyr Tyr
 85 90 95
 tgt gcg agg gtc gag gct gcg ttc gac ctc ttt gac tac tgg ggc cag 336
 Cys Ala Arg Val Glu Ala Ala Phe Asp Leu Phe Asp Tyr Trp Gly Gln
 100 105 110
 gga acc atg gtc acc gtc tcc tca 360
 Gly Thr Met Val Thr Val Ser Ser
 115 120

<210> 39
 <211> 453
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (1)..(450)

21/22

<220>

<221> modified_base

<222> (140)..(141)

<223> a, c, t, g, unknown or other

<400> 39

atg cag gtg cag ctg gtg gag tct gga gct gag gtg aag aag cct ggg	48
Met Gln Val Gln Leu Val Glu Ser Gly Ala Glu Val Lys Lys Pro Gly	
1 5 10 15	
gcc tca gtg agg gtc tcc tgc aag gct tct ggt tac acc ttt acc agc	96
Ala Ser Val Arg Val Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Ser	
20 25 30	
tac ggt atc act tgg gtg cga cag gcc cct gga caa aaa ctt gnn tgg	144
Tyr Gly Ile Thr Trp Val Arg Gln Ala Pro Gly Gln Lys Leu Xaa Trp	
35 40 45	
ctg gga tgg atc agc gct tac aat ggt aac aca cat tat gca cag aag	192
Leu Gly Trp Ile Ser Ala Tyr Asn Gly Asn Thr His Tyr Ala Gln Lys	
50 55 60	
ctc cag ggc aga gtc acc atg acc aca gac aca tcc acg acc aca gcc	240
Leu Gln Gly Arg Val Thr Met Thr Thr Asp Thr Ser Thr Thr Thr Ala	
65 70 75 80	
tac atg gag ctt agg agc ctg aga tct gac gac acg gcc gtg tat tac	288
Tyr Met Glu Leu Arg Ser Leu Arg Ser Asp Asp Thr Ala Val Tyr Tyr	
85 90 95	
tgt gcg agg gtc gag gct gcg ttc gac ctc ttt gac tac tgg ggc cag	336
Cys Ala Arg Val Glu Ala Ala Phe Asp Leu Phe Asp Tyr Trp Gly Gln	
100 105 110	
gga acc atg gtc acc gtc tcc tca aag ggc aat tcg aag ctt gaa ggt	384
Gly Thr Met Val Thr Val Ser Ser Lys Gly Asn Ser Lys Leu Glu Gly	
115 120 125	
aag cct atc cct aac cct ctc ctc ggt ctc gat tct acg cgt acc ggt	432
Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr Arg Thr Gly	
130 135 140	
cat cat cac cat cac cat tga	453
His His His His His	
145 150	

<210> 40

<211> 6

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
6xHis tag

22/22

<400> 40
His His His His His His
1 5