

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
13 March 2003 (13.03.2003)

PCT

(10) International Publication Number
WO 03/020897 A2

- (51) International Patent Classification⁷: **C12N**
- (21) International Application Number: PCT/US02/27727
- (22) International Filing Date: 29 August 2002 (29.08.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/316,307 31 August 2001 (31.08.2001) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/316,307 (CIP)
Filed on 31 August 2001 (31.08.2001)
- (71) Applicant (*for all designated States except US*): **DI-ADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, San Francisco, CA 94080 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **SUN, Yongming** [CN/US]; 551 Shoal Drive, Redwood City, CA 94065 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US). **SALCEDA, Susana** [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US).
- (74) Agents: **LICATA, Jane, Massey et al.**; Licata & Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (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: COMPOSITIONS AND METHODS RELATING TO OVARY SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic ovarian cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions containing the nucleic acid molecules, polypeptides, antibodies, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in ovarian, identifying ovarian tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered ovarian tissue for treatment and research.



WO 03/020897 A2

COMPOSITIONS AND METHODS RELATING TO OVARY SPECIFIC GENES AND PROTEINS

5

INTRODUCTION

This application claims the benefit of priority from U.S. Provisional Application No. 60/316,307, filed August 31, 2001, which is herein incorporated by reference in its entirety.

10

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic ovary cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating ovary cancer and non-cancerous disease states in ovary, identifying ovary tissue and monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered ovary tissue for treatment and research.

15

20

BACKGROUND OF THE INVENTION

25

Cancer of the ovaries is the fourth-most cause of cancer death in women in the United States, with more than 23,000 new cases and roughly 14,000 deaths predicted for the year 2001. Shridhar, V. et al., *Cancer Res.* 61(15): 5895-904 (2001); Memarzadeh, S. & Berek, J. S., *J. Reprod. Med.* 46(7): 621-29 (2001). The incidence of ovarian cancer is of serious concern worldwide, with an estimated 191,000 new cases predicted annually. Runnebaum, I. B. & Stickeler, E., *J. Cancer Res. Clin. Oncol.* 127(2): 73-79 (2001). Unfortunately, women with ovarian cancer are typically asymptomatic until the disease has metastasized. Because effective screening for ovarian cancer is not available, roughly 70% of women diagnosed have an advanced stage of the cancer with a five-year survival

30

-2-

rate of ~25-30%. Memarzadeh, S. & Berek, J. S., *supra*; Nunns, D. et al., *Obstet. Gynecol. Surv.* 55(12): 746-51. Conversely, women diagnosed with early stage ovarian cancer enjoy considerably higher survival rates. Werness, B. A. & Eltabbakh, G. H., *Int'l. J. Gynecol. Pathol.* 20(1): 48-63 (2001). Although our understanding of the etiology of ovarian cancer is incomplete, the results of extensive research in this area point to a combination of age, genetics, reproductive, and dietary/environmental factors. Age is a key risk factor in the development of ovarian cancer: while the risk for developing ovarian cancer before the age of 30 is slim, the incidence of ovarian cancer rises linearly between ages 30 to 50, increasing at a slower rate thereafter, with the highest incidence being among septagenarian women. Jeanne M. Schilder et al., *Hereditary Ovarian Cancer: Clinical Syndromes and Management*, in Ovarian Cancer 182 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001).

With respect to genetic factors, a family history of ovarian cancer is the most significant risk factor in the development of the disease, with that risk depending on the number of affected family members, the degree of their relationship to the woman, and which particular first degree relatives are affected by the disease. *Id.* Mutations in several genes have been associated with ovarian cancer, including BRCA1 and BRCA2, both of which play a key role in the development of breast cancer, as well as hMSH2 and hMLH1, both of which are associated with hereditary non-polyposis colon cancer. Katherine Y. Look, *Epidemiology, Etiology, and Screening of Ovarian Cancer*, in Ovarian Cancer 169, 171-73 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). BRCA1, located on chromosome 17, and BRCA2, located on chromosome 13, are tumor suppressor genes implicated in DNA repair; mutations in these genes are linked to roughly 10% of ovarian cancers. *Id.* at 171-72; Schilder et al., *supra* at 185-86. hMSH2 and hMLH1 are associated with DNA mismatch repair, and are located on chromosomes 2 and 3, respectively; it has been reported that roughly 3% of hereditary ovarian carcinomas are due to mutations in these genes. Look, *supra* at 173; Schilder et al., *supra* at 184, 188-89. Reproductive factors have also been associated with an increased or reduced risk of ovarian cancer. Late menopause, nulliparity, and early age at menarche have all been linked with an elevated risk of ovarian cancer. Schilder et al., *supra* at 182. One theory hypothesizes that these factors increase the number of ovulatory cycles over the course of a woman's life, leading to "incessant ovulation," which is thought to be the primary cause of mutations to the ovarian epithelium. *Id.*; Laura J. Havrilesky & Andrew Berchuck,

-3-

Molecular Alterations in Sporadic Ovarian Cancer, in Ovarian Cancer 25 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). The mutations may be explained by the fact that ovulation results in the destruction and repair of that epithelium, necessitating increased cell division, thereby increasing the possibility that an undetected mutation will occur. *Id.* Support for this theory may be found in the fact pregnancy, lactation, and the use of oral contraceptives, all of which suppress ovulation, confer a protective effect with respect to developing ovarian cancer. *Id.*

Among dietary/environmental factors, there would appear to be an association between high intake of animal fat or red meat and ovarian cancer, while the antioxidant Vitamin A, which prevents free radical formation and also assists in maintaining normal cellular differentiation, may offer a protective effect. Look, *supra* at 169. Reports have also associated asbestos and hydrous magnesium trisilicate (talc), the latter of which may be present in diaphragms and sanitary napkins. *Id.* at 169-70.

Current screening procedures for ovarian cancer, while of some utility, are quite limited in their diagnostic ability, a problem that is particularly acute at early stages of cancer progression when the disease is typically asymptomatic yet is most readily treated. Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 166 (1998); Memarzadeh & Berek, *supra*; Runnebaum & Stickeler, *supra*; Werness & Eltabbakh, *supra*. Commonly used screening tests include biannual rectovaginal pelvic examination, radioimmunoassay to detect the CA-125 serum tumor marker, and transvaginal ultrasonography. Burdette, *supra* at 166.

Pelvic examination has failed to yield adequate numbers of early diagnoses, and the other methods are not sufficiently accurate. *Id.* One study reported that only 15% of patients who suffered from ovarian cancer were diagnosed with the disease at the time of their pelvic examination. Look, *supra* at 174. Moreover, the CA-125 test is prone to giving false positives in pre-menopausal women and has been reported to be of low predictive value in post-menopausal women. *Id.* at 174-75. Although transvaginal ultrasonography is now the preferred procedure for screening for ovarian cancer, it is unable to distinguish reliably between benign and malignant tumors, and also cannot locate primary peritoneal malignancies or ovarian cancer if the ovary size is normal. Schilder et al., *supra* at 194-95. While genetic testing for mutations of the BRCA1, BRCA2, hMSH2, and hMLH1 genes is now available, these tests may be too costly for

-4-

some patients and may also yield false negative or indeterminate results. Schilder et al., *supra* at 191-94.

The staging of ovarian cancer, which is accomplished through surgical exploration, is crucial in determining the course of treatment and management of the disease. AJCC
5 Cancer Staging Handbook 187 (Irvin D. Fleming et al. eds., 5th ed. 1998); Burdette, *supra* at 170; Memarzadeh & Berek, *supra*; Shridhar et al., *supra*. Staging is performed by reference to the classification system developed by the International Federation of Gynecology and Obstetrics. David H. Moore, *Primary Surgical Management of Early Epithelial Ovarian Carcinoma*, in Ovarian Cancer 203 (Stephen C. Rubin & Gregory P.
10 Sutton eds., 2d ed. 2001); Fleming et al. eds., *supra* at 188. Stage I ovarian cancer is characterized by tumor growth that is limited to the ovaries and is comprised of three substages. *Id.* In substage IA, tumor growth is limited to one ovary, there is no tumor on the external surface of the ovary, the ovarian capsule is intact, and no malignant cells are present in ascites or peritoneal washings. *Id.* Substage IB is identical to A1, except that
15 tumor growth is limited to both ovaries. *Id.* Substage IC refers to the presence of tumor growth limited to one or both ovaries, and also includes one or more of the following characteristics: capsule rupture, tumor growth on the surface of one or both ovaries, and malignant cells present in ascites or peritoneal washings. *Id.*

Stage II ovarian cancer refers to tumor growth involving one or both ovaries, along
20 with pelvic extension. *Id.* Substage IIA involves extension and/or implants on the uterus and/or fallopian tubes, with no malignant cells in the ascites or peritoneal washings, while substage IIB involves extension into other pelvic organs and tissues, again with no malignant cells in the ascites or peritoneal washings. *Id.* Substage IIC involves pelvic extension as in IIA or IIB, but with malignant cells in the ascites or peritoneal washings.
25 *Id.*

Stage III ovarian cancer involves tumor growth in one or both ovaries, with peritoneal metastasis beyond the pelvis confirmed by microscope and/or metastasis in the regional lymph nodes. *Id.* Substage IIIA is characterized by microscopic peritoneal metastasis outside the pelvis, with substage IIIB involving macroscopic peritoneal
30 metastasis outside the pelvis 2 cm or less in greatest dimension. *Id.* Substage IIIC is identical to IIIB, except that the metastasis is greater than 2 cm in greatest dimension and may include regional lymph node metastasis. *Id.* Lastly, Stage IV refers to the presence distant metastasis, excluding peritoneal metastasis. *Id.*

-5-

While surgical staging is currently the benchmark for assessing the management and treatment of ovarian cancer, it suffers from considerable drawbacks, including the invasiveness of the procedure, the potential for complications, as well as the potential for inaccuracy. Moore, *supra* at 206-208, 213. In view of these limitations, attention has
5 turned to developing alternative staging methodologies through understanding differential gene expression in various stages of ovarian cancer and by obtaining various biomarkers to help better assess the progression of the disease. Vartiainen, J. et al., *Int'l J. Cancer*, 95(5): 313-16 (2001); Shridhar et al. *supra*; Baekelandt, M. et al., *J. Clin. Oncol.* 18(22): 3775-81.

10 The treatment of ovarian cancer typically involves a multiprong attack, with surgical intervention serving as the foundation of treatment. Dennis S. Chi & William J. Hoskins, *Primary Surgical Management of Advanced Epithelial Ovarian Cancer*, in Ovarian Cancer 241 (Stephen C. Rubin & Gregory P. Sutton eds., 2d ed. 2001). For example, in the case of epithelial ovarian cancer, which accounts for ~90% of cases of
15 ovarian cancer, treatment typically consists of: (1) cytoreductive surgery, including total abdominal hysterectomy, bilateral salpingo-oophorectomy, omentectomy, and lymphadenectomy, followed by (2) adjuvant chemotherapy with paclitaxel and either cisplatin or carboplatin. Eltabbakh, G.H. & Awtrey, C.S., *Expert Op. Pharmacother.* 2(10): 109-24. Despite a clinical response rate of 80% to the adjuvant therapy, most
20 patients experience tumor recurrence within three years of treatment. *Id.* Certain patients may undergo a second cytoreductive surgery and/or second-line chemotherapy. Memarzadeh & Berek, *supra*.

From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of ovarian cancer are
25 of critical importance to the outcome of the patient. Moreover, current procedures, while helpful in each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

30 Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop ovarian cancer, for diagnosing ovarian cancer, for monitoring the progression of the disease, for staging the ovarian cancer, for

determining whether the ovarian cancer has metastasized, and for imaging the ovarian cancer. There is also a need for better treatment of ovarian cancer.

SUMMARY OF THE INVENTION

5 The present invention solves many needs in the art by providing nucleic acid molecules, polypeptides and antibodies thereto, variants and derivatives of the nucleic acids and polypeptides, agonists and antagonists that may be used to identify, diagnose, monitor, stage, image and treat ovarian cancer and non-cancerous disease states in ovarian; identify and monitor ovarian tissue; and identify and design agonists and
10 antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered ovarian tissue for treatment and research.

 One aspect of the present invention relates to nucleic acid molecules that are specific to ovarian cells, ovarian tissue and/or the ovarian organ. These ovarian specific
15 nucleic acids (OSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the OSNA is genomic DNA, then the OSNA is an ovarian specific gene (OSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to ovarian. More preferred is a nucleic acid molecule that encodes a
20 polypeptide comprising an amino acid sequence of SEQ ID NO: 62-116. In another preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-61. For the sequences listed herein, DEX0237_1 corresponds to SEQ ID NO: 1, DEX0237_62 corresponds to SEQ ID NO: 62, etc.

 This aspect of the present invention also relates to nucleic acid molecules that
25 selectively hybridize or exhibit substantial sequence similarity to nucleic acid molecules encoding a Ovarian Specific Protein (OSP), or that selectively hybridize or exhibit substantial sequence similarity to an OSNA. In one embodiment of the present invention the nucleic acid molecule comprises an allelic variant of a nucleic acid molecule encoding an OSP, or an allelic variant of an OSNA. In another embodiment, the nucleic acid
30 molecule comprises a part of a nucleic acid sequence that encodes an OSP or a part of a nucleic acid sequence of an OSNA.

 In addition, this aspect of the present invention relates to a nucleic acid molecule further comprising one or more expression control sequences controlling the transcription

-7-

and/or translation of all or a part of an OSNA or the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of an OSP.

Another aspect of the present invention relates to vectors and/or host cells comprising a nucleic acid molecule of this invention. In a preferred embodiment, the nucleic acid molecule of the vector and/or host cell encodes all or a fragment of an OSP. In another preferred embodiment, the nucleic acid molecule of the vector and/or host cell comprises all or a part of an OSNA. Vectors and host cells of the present invention are useful in the recombinant production of polypeptides, particularly OSPs of the present invention.

Another aspect of the present invention relates to polypeptides encoded by a nucleic acid molecule of this invention. The polypeptide may comprise either a fragment or a full-length protein. In a preferred embodiment, the polypeptide is an OSP. However, this aspect of the present invention also relates to mutant proteins (muteins) of OSPs, fusion proteins of which a portion is an OSP, and proteins and polypeptides encoded by allelic variants of an OSNA as provided herein.

Another aspect of the present invention relates to antibodies and other binders that specifically binds to a polypeptide of the instant invention. Accordingly antibodies or binders of the present specifically bind to OSPs, muteins, fusion proteins, and/or homologous proteins or a polypeptides encoded by allelic variants of an OSNA as provided herein.

Another aspect of the present invention relates to agonists and antagonists of the nucleic acid molecules and polypeptides of this invention. The agonists and antagonists of the instant invention may be used to treat ovarian cancer and non-cancerous disease states in the ovary and to produce engineered ovarian tissue.

Another aspect of the present invention relates to methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. Such methods are useful in identifying, diagnosing, monitoring, staging, imaging and treating ovarian cancer and non-cancerous disease states in the ovary. Such methods are also useful in identifying and/or monitoring ovarian tissue. In addition, measurement of levels of the nucleic acid molecules of this invention may be useful for diagnostics as part of panel in combination with other markers.

-8-

Another aspect of the present invention relates to use of the nucleic acid molecules of this invention in gene therapy, for producing transgenic animals and cells, and for producing engineered ovarian tissue for treatment and research.

Another aspect of the present invention relates to methods for detecting polypeptides this invention, preferably using antibodies thereto. Such methods are useful to identify, diagnose, monitor, stage, image and treat ovarian cancer and non-cancerous disease states in the ovary. In addition, measurement of levels of the polypeptides of this invention may be useful to identify, diagnose, monitor, stage, image ovarian cancer in combination with other ovarian cancer markers. The polypeptides of the present invention can also be used to identify and/or monitor ovarian tissue, and to produce engineered ovarian tissue.

Yet another aspect of the present invention relates to a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences. In addition, the computer records regarding the nucleic acid and/or amino acid sequences and/or measurements of their levels may be used alone or in combination with other markers to diagnose ovarian related diseases.

DETAILED DESCRIPTION OF THE INVENTION

20 Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.,* Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et*

al., Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999).

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single and double stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides),

intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term “nucleic acid molecule” also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A “gene” is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term “exon” refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute contiguous sequence to a mature mRNA transcript. The term “intron” refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be “spliced out” during processing of the transcript.

A nucleic acid molecule or polypeptide is “derived” from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An “isolated” or “substantially pure” nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the “isolated polynucleotide” is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are

not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that
5 are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid
10 molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a
15 nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the
20 disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic
25 acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single-or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably,
30 oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene.

-12-

Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including in vitro recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), and United States Patent No. 5,151,510, the disclosure of which is hereby incorporated by reference in its entirety.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as

the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends
5 ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical
10 biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least
15 about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer
20 Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998)). Unless otherwise
25 specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1.

A reference to a nucleic acid sequence encompasses its complement unless
30 otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms “percent sequence identity”, “percent sequence similarity” and “percent sequence homology” interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

5 The term “substantial similarity” or “substantial sequence similarity,” when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least
10 about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

 Alternatively, substantial similarity exists between a first and second nucleic acid sequence when the first nucleic acid sequence or fragment thereof hybridizes to an
15 antisense strand of the second nucleic acid, , under selective hybridization conditions. Typically, selective hybridization will occur between the first nucleic acid sequence and an antisense strand of the second nucleic acid sequence when there is at least about 55% sequence identity between the first and second nucleic acid sequences—preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% —
20 over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

 Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches
25 between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. “Stringent hybridization conditions” and “stringent wash conditions” in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One
30 having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, “stringent hybridization” is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. “Stringent washing” is performed at temperatures about 5°C lower than

-15-

the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

5
$$T_m = 81.5^{\circ}\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l) \text{ where } l \text{ is the length of the hybrid in base pairs.}$$

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$$T_m = 79.8^{\circ}\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

10 The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^{\circ}\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50 (\% \text{ formamide}) - (820/l).$$

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid

sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. See Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58. See also Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (see Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acids that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid is created synthetically or recombinantly using a high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (e.g., for oligonucleotide probes) may be calculated by the formula:

$$T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/N),$$
 wherein N is change length and the $[\text{Na}^+]$ is 1 M or less. See Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or “guessmers,” as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. See, e.g., Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term “digestion” or “digestion of DNA” refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The

-17-

various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μ g of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μ l of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μ g of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies. In another aspect, the invention is directed to single exon probes based on the OSNAs disclosed herein.

-18-

In one embodiment, the term "microarray" refers to a "nucleic acid microarray" having a substrate-bound plurality of nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). Additionally, these nucleic acid microarrays include substrate-bound plurality of nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000). Examples of nucleic acid microarrays may be found in U.S. Patent Nos. 6,391,623, 6,383,754, 6,383,749, 6,380,377, 6,379,897, 6,376,191, 6,372,431, 6,351,712, 6,344,316, 6,316,193, 6,312,906, 6,309,828, 6,309,824, 6,306,643, 6,300,063, 6,287,850, 6,284,497, 6,284,465, 6,280,954, 6,262,216, 6,251,601, 6,245,518, 6,263,287, 6,251,601, 6,238,866, 6,228,575, 6,214,587, 6,203,989, 6,171,797, 6,103,474, 6,083,726, 6,054,274, 6,040,138, 6,083,726, 6,004,755, 6,001,309, 5,958,342, 5,952,180, 5,936,731, 5,843,655, 5,814,454, 5,837,196, 5,436,327, 5,412,087, 5,405,783, the disclosures of which are incorporated herein by reference in their entireties.

In an alternative embodiment, a "microarray" may also refer to a "peptide microarray" or "protein microarray" having a substrate-bound collection of plurality of polypeptides, the binding to each of the plurality of bound polypeptides being separately detectable. Alternatively, the peptide microarray, may have a plurality of binders, including but not limited to monoclonal antibodies, polyclonal antibodies, phage display binders, yeast 2 hybrid binders, aptamers, which can specifically detect the binding of the polypeptides of this invention. The array may be based on autoantibody detection to the polypeptides of this invention, see Robinson *et al.*, *Nature Medicine* 8(3):295-301 (2002). Examples of peptide arrays may be found in WO 02/31463, WO 02/25288, WO 01/94946, WO 01/88162, WO 01/68671, WO 01/57259, WO 00/61806, WO 00/54046, WO 00/47774, WO 99/40434, WO 99/39210, WO 97/42507 and U.S. Patent Nos. 6,268,210, 5,766,960, 5,143,854, the disclosures of which are incorporated herein by reference in their entireties.

In addition, determination of the levels of the OSNA or OSP may be made in a multiplex manner using techniques described in WO 02/29109, WO 02/24959, WO 01/83502, WO01/73113, WO 01/59432, WO 01/57269, WO 99/67641, the disclosures of which are incorporated herein by reference in their entireties.

5 The term “mutant”, “mutated”, or “mutation” when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci
10 within a nucleic acid sequence. In a preferred embodiment of the present invention, the nucleic acid sequence is the wild type nucleic acid sequence encoding an OSP or is an OSNA. The nucleic acid sequence may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

15 The term “error-prone PCR” refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. See, *e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

20 The term “oligonucleotide-directed mutagenesis” refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. See, *e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

25 The term “assembly PCR” refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

30 The term “sexual PCR mutagenesis” or “DNA shuffling” refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. See, *e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes (“Family shuffling”).

-20-

The term “*in vivo* mutagenesis” refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These “mutator” strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term “cassette mutagenesis” refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide “cassette” that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term “recursive ensemble mutagenesis” refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. See, *e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term “exponential ensemble mutagenesis” refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. See, *e.g.*, Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993).

“Operatively linked” expression control sequences refers to a linkage in which the expression control sequence is either contiguous with the gene of interest to control the gene of interest, or acts in *trans* or at a distance to control the gene of interest.

The term “expression control sequence” as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature

of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refers to that portion of a transcript-derived nucleic acid that can be translated in its

-22-

entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence is meant to be inclusive of all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, as well as polypeptide fragments and polypeptide mutants, derivatives and analogs thereof. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises an OSP encoded by a nucleic acid molecule of the instant invention, or a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be determined by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well-

known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "fragment" when used herein with respect to polypeptides of the present invention refers to a polypeptide that has an amino-terminal and/or
5 carboxy-terminal deletion compared to a full-length OSP. In a preferred embodiment, the fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring polypeptide. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or
10 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" when used herein with respect to polypeptides of the present invention refers to a polypeptide which is substantially similar in primary structural sequence to an OSP but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical
15 modifications that are not found in the OSP. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation,
20 formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.
25 Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ¹²⁵I, ³²P, ³⁵S, ¹⁴C and ³H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores,
30 chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation.

Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*.

The term "fusion protein" refers to polypeptides of the present invention coupled to a heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence that encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH₂NH--, --CH₂S--, --CH₂-CH₂--, --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂--, and --CH₂SO--, by methods well-known in the art. Systematic

substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known
5 in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992)). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

The term "mutant" or "mutein" when referring to a polypeptide of the present invention relates to an amino acid sequence containing substitutions, insertions or
10 deletions of one or more amino acids compared to the amino acid sequence of an OSP. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence
15 at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90%
20 sequence similarity to an OSP. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to
25 proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion
30 of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not

substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991).

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991). Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α -, α -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate, ϵ -N,N,N-trimethyllysine, ϵ -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

By "homology" or "homologous" when referring to a polypeptide of the present invention it is meant polypeptides from different organisms with a similar sequence to the encoded amino acid sequence of an OSP and a similar biological activity or function. Although two polypeptides are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the polypeptides. Instead, the term "homologous" is defined to mean that the two polypeptides have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous polypeptide is one that exhibits 50% sequence similarity to OSP, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous polypeptides that exhibit 80%, 85% or 90% sequence similarity to an OSP. In a yet more preferred embodiment, a homologous polypeptide exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to polypeptides, it is recognized that residue positions that are not identical often differ by conservative amino acid

-27-

substitutions. In a preferred embodiment, a polypeptide that has “sequence similarity” comprises conservative or moderately conservative amino acid substitutions. A “conservative amino acid substitution” is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. See, *e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994).

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992). A “moderately conservative” replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as “Gap” and “Bestfit” which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. See, *e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

-28-

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. See, *e.g.*, Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997). Preferred parameters for

5 blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default)
10 Max. alignments:	100 (default)
Word size:	11 (default)
No. of descriptions:	100 (default)
Penalty Matrix:	BLOSUM62

15 The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Algorithms other than blastp for database searching using amino acid sequences
20 are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA
25 with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by
30 recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')₂, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an

immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. A Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; a F(ab')₂ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; a Fd fragment consists of the VH and CH1 domains; a Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. See, e.g., Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" as used herein it is meant the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which VL and VH regions are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. See, e.g., Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. See e.g., Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

-30-

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified
5 proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that
10 normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains
15 and usually have specific three dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than $1\ \mu\text{M}$, preferably less than $100\ \text{nM}$ and most preferably less than $10\ \text{nM}$.

The term "patient" includes human and veterinary subjects.

20 Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "ovarian specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the ovary as compared to other tissues in the body. In a
25 preferred embodiment, an "ovarian specific" nucleic acid molecule or polypeptide is detected at a level that is 2-fold higher than any other tissue in the body. In a more preferred embodiment, the "ovarian specific" nucleic acid molecule or polypeptide is detected at a level that is 5-fold higher than any other tissue in the body, more preferably at least 10-fold, 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue
30 in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

5 One aspect of the invention provides isolated nucleic acid molecules that are specific to the ovary or to ovarian cells or tissue or that are derived from such nucleic acid molecules. These isolated ovarian specific nucleic acids (OSNAs) may comprise cDNA genomic DNA, RNA, or a combination thereof, a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. An OSNA may be derived
10 from an animal. In a preferred embodiment, the OSNA is derived from a human or other mammal. In a more preferred embodiment, the OSNA is derived from a human or other primate. In an even more preferred embodiment, the OSNA is derived from a human.

 In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to the ovary, an ovarian-specific polypeptide (OSP). In a more preferred
15 embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 62-116. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-61. Nucleotide sequences of the instantly-described nucleic acid molecules were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one
20 enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Amersham Biosciences, Sunnyvale, CA, USA).

 Nucleic acid molecules of the present invention may also comprise sequences that selectively hybridizes to a nucleic acid molecule encoding an OSNA or a complement or
25 antisense thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may or may not encode an OSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes an OSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule that encodes a polypeptide
30 comprising an amino acid sequence of SEQ ID NO: 62-116. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1-61 or the antisense sequence thereof. Preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding an

-32-

OSP under low stringency conditions. More preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding an OSP under moderate stringency conditions. Most preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding an OSP under high stringency conditions. In a preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 62-116. In a more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1-61.

Nucleic acid molecules of the present invention may also comprise nucleic acid sequences that exhibit substantial sequence similarity to a nucleic acid encoding an OSP or a complement of the encoding nucleic acid molecule. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid molecule encoding human OSP. More preferred is a nucleic acid molecule exhibiting substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 62-116. By substantial sequence similarity it is meant a nucleic acid molecule having at least 60% sequence identity with a nucleic acid molecule encoding an OSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 62-116, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding an OSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. Most preferred in this embodiment is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding an OSP.

The nucleic acid molecules of the present invention are also inclusive of those exhibiting substantial sequence similarity to an OSNA or its complement. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid molecule having a nucleic acid sequence of SEQ ID NO: 1-61. By substantial sequence similarity it is meant a nucleic acid molecule that has at least 60%

sequence identity with an OSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1-61, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. More preferred is a nucleic acid molecule that has at least 90% sequence identity with an OSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. Most preferred is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with an OSNA.

Nucleic acid molecules that exhibit substantial sequence similarity are inclusive of sequences that exhibit sequence identity over their entire length to an OSNA or to a nucleic acid molecule encoding an OSP, as well as sequences that are similar over only a part of its length. In this case, the part is at least 50 nucleotides of the OSNA or the nucleic acid molecule encoding an OSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 62-116 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1-61. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the OSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of an OSNA. In a preferred embodiment, the substantially similar nucleic acid molecule is an OSNA.

-34-

The nucleic acid molecules of the present invention are also inclusive of allelic variants of an OSNA or a nucleic acid encoding an OSP. For example, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes and the sequence determined from one individual of a species may differ from other allelic forms present within the population. More than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001) – Variants with small deletions and insertions of more than a single nucleotide are also found in the general population, and often do not alter the function of the protein. In addition, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes an OSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes an OSP comprising an amino acid sequence of SEQ ID NO: 62-116. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is an OSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1-61. Also preferred is that the allelic variant is a naturally-occurring allelic variant in the species of interest, particularly human.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences comprising a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is an OSP. In a preferred embodiment, the part encodes an OSP. In one embodiment, the nucleic acid molecule comprises a part of an OSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to an OSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that is an allelic variant of an OSNA. In yet another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that encodes an OSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

-35-

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences that encode fusion proteins, homologous proteins, polypeptide fragments, muteins and polypeptide analogs, as described *infra*.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences containing modifications of the native nucleic acid molecule. Examples of such modifications include, but are not limited to, nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

Accordingly, in one embodiment, a nucleic acid molecule may include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. The labeled nucleic acid molecules are particularly useful as hybridization probes.

Common radiolabeled analogues include those labeled with ^{33}P , ^{32}P , and ^{35}S , such as α - ^{32}P -dATP, α - ^{32}P -dCTP, α - ^{32}P -dGTP, α - ^{32}P -dTTP, α - ^{32}P -3'dATP, α - ^{32}P -ATP, α - ^{32}P -CTP, α - ^{32}P -GTP, α - ^{32}P -UTP, α - ^{35}S -dATP, γ - ^{35}S -GTP, γ - ^{33}P -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Biosciences, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas

-36-

Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature*

5 *Biotechnol.* 18: 345-348 (2000).

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules of the present invention can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and Peptide Nucleic Acids (PNA) to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); see Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301-305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994). Alternatively, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

-37-

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. See, e.g., Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); and United States Patent Nos. 5,846,726, 5,925,517, 5,925,517, 5,723,591 and 5,538,848, the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the present invention may also be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant internucleoside bonds. See Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997). Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction, Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000).

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, United States Patent Nos. 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361;

-38-

and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, United States Patent Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred nucleic acid molecules, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, United States Patent Nos. 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference in its entirety. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA). PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and

PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The T_m of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the T_m of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T_m by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the T_m by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. See, e.g., Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999).

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), and United States Patent Nos. 5,760,012 and 5,731,181, the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acid molecules of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); and Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994). Triplex and quadruplex conformations, and

their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997).

5 *Methods for Using Nucleic Acid Molecules as Probes and Primers*

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably
10 labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect and characterize gross alterations in the gene of an OSNA, such as deletions, insertions, translocations, and duplications of the OSNA genomic locus
15 through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999). The isolated nucleic acid molecules of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated
20 nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include a nucleic acid molecule of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

The isolated nucleic acid molecules of the present invention can be also be used as
25 probes to detect, characterize, and quantify OSNA in, and isolate OSNA from, transcript-derived nucleic acid samples. In one embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A⁺- selected RNA samples. In another embodiment, the isolated nucleic acid molecules of the present invention can be
30 used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarczacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000). In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization

probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to OSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000).

In another embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify and/or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In this embodiment, it is preferred that the probe or primer be derived from a nucleic acid molecule encoding an OSP. More preferably, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 62-116. Also preferred are probes or primers derived from an OSNA. More preferred are probes or primers derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-61.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et*

al. (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to
5 Genetic Engineering, Vol. 67, Humana Press (1996); and McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995). Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; and Siebert (ed.), PCR Technique: RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995).

10 PCR and hybridization methods may be used to identify and/or isolate nucleic acid molecules of the present invention including allelic variants, homologous nucleic acid molecules and fragments. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules of the present invention that encode homologous proteins, analogs, fusion protein or muteins of the invention. Nucleic acid
15 primers as described herein can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

These nucleic acid primers can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Pat. No. 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

20 Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); international patent publications WO 97/19193 and WO 00/15779, and United States Patent Nos. 5,854,033 and 5,714,320, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with
25 other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization
30 probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or

-43-

positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that comprise one or more of the nucleic acid molecules of the present invention.

In yet another embodiment, the invention is directed to single exon probes based on the OSNAs disclosed herein.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention provides vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acid molecules of the present invention in host cells (cloning vectors), for shuttling the nucleic acid

-44-

molecules of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acid molecules of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acid molecules of the present invention *in vitro* or within a host cell, and for
5 expressing polypeptides encoded by the nucleic acid molecules of the present invention, alone or as fusion proteins with heterologous polypeptides (expression vectors). Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential
10 Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof are well within
15 the skill in the art. Thus, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and
20 translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in
25 expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred
30 embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*,

-45-

including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λ GT10 and λ GT11, and other phages, *e.g.*, M13 and filamentous single stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEplac series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2 μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences

-46-

resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

The host cells may also be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include, but are not limited to, resistance to neomycin (G418), blasticidin, hygromycin and zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

-47-

Any of a wide variety of expression control sequences may be used in these vectors to express the nucleic acid molecules of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the *trc* promoter, a hybrid derived from the *trp* and *lac* promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of *fd* coat protein, and the *araBAD* operon. Prokaryotic expression vectors may further include transcription terminators, such as the *aspA* terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the *CYC1* promoter, the *GAL1* promoter, the *GAL10* promoter, *ADH1* promoter, the promoters of the yeast α -mating system, or the *GPD* promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the *CYC1* or *ADH1* gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include, but are not limited to, those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 and the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic

enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the OSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PLtetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone

receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one embodiment of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization.

5 Such tags include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal

10 of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega,

15 Madison, WI, USA). As another useful alternative, the polypeptides of the present invention can be expressed as a fusion to glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the

20 Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope, detectable by anti-HA antibody.

25 For secretion of expressed polypeptides, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

30 Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful protein fusions include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such

as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusions for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999). A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see*,

-51-

e.g., Cormack *et al.*, *Gene* 173: 33–38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.*, Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in United States Patent Nos. 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See also* Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the polypeptides of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The *bsd* gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus.

-52-

The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA) allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid molecules of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed polypeptide in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide OSPs with such post-translational modifications.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid molecules of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid molecules of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as

-53-

recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid molecules according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

5 Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

10 Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*).

15 Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the
20 polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

 A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and
25 prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*,
30 *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda* — *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA) — *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include

BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and
5 BW5147 cells. Other mammalian cell lines are well-known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from the ovary are particularly preferred because they may provide a more native post-translational
10 processing. Particularly preferred are human ovarian cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number
15 of texts and laboratory manuals in the art. See, e.g., Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*.

Methods for introducing the vectors and nucleic acid molecules of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

20 Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (e.g., Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or
25 electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, e.g., with CaCl_2 , or a solution of Mg^{2+} , Mn^{2+} , Ca^{2+} , Rb^+ or K^+ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (e.g., Epicurian Coli® XL10-Gold®
30 Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5α competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent *E. coli* Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent to take up exogenous DNA by electroporation by various pre-pulse treatments; vectors are

introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided by BioRad (Richmond, CA, USA).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action
5 of hydrolytic enzymes such as a snail-gut extract, usually denoted Glusulase or Zymolyase, or an enzyme from *Arthrobacter luteus* to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+} . Subsequently, the cells are resuspended in a solution of sorbitol, mixed
10 with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate to permeabilize the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased
15 frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension
20 pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

25 Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with CaPO_4 or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for CaPO_4 transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated
30 transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis,

-56-

IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA).

Protocols for electroporating mammalian cells can be found in, for example, ; Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms,

BioTechniques Books, Eaton Publishing Co. (2000). Other transfection techniques

- 5 include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

- 10 Purification of recombinantly expressed proteins is now well within the skill in the art and thus need not be detailed here. *See, e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale,
15 Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001).

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tag, purification can be effected, at least in part, by means

- 20 appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides, including Fragments Muteins, Homologous Proteins, Allelic Variants,

- 25 Analogues and Derivatives

Another aspect of the invention relates to polypeptides encoded by the nucleic acid molecules described herein. In a preferred embodiment, the polypeptide is an ovarian specific polypeptide (OSP). In an even more preferred embodiment, the polypeptide comprises an amino acid sequence of SEQ ID NO:62-116 or is derived from a polypeptide
30 having the amino acid sequence of SEQ ID NO: 62-116. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings

-57-

of the specification and using methods well-known to those having ordinary skill in the art.

Polypeptides of the present invention may also comprise a part or fragment of an OSP. In a preferred embodiment, the fragment is derived from a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO: 62-116. Polypeptides of the present invention comprising a part or fragment of an entire OSP may or may not be OSPs. For example, a full-length polypeptide may be ovarian-specific, while a fragment thereof may be found in other tissues as well as in the ovary. A polypeptide that is not an OSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-OSP antibodies. In a preferred embodiment, the part or fragment is an OSP. Methods of determining whether a polypeptide of the present invention is an OSP are described *infra*.

Polypeptides of the present invention comprising fragments of at least 6 contiguous amino acids are also useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of a polypeptide of the present invention have utility in such a study.

Polypeptides of the present invention comprising fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize polypeptides of the present invention. *See, e.g., Lerner, Nature* 299: 592-596 (1982); Shinnick *et al., Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al., Science* 219: 660-6 (1983). As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic and are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the polypeptides of the present invention have utility as immunogens.

Polypeptides comprising fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire polypeptide, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners,

such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the polypeptide of interest. See United States Patent Nos. 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

5 The polypeptide of the present invention thus preferably is at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the polypeptide of the present invention is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger polypeptides having at least 75 amino acids, 100 amino acids,
10 or even 150 amino acids are also useful, and at times preferred.

 One having ordinary skill in the art can produce fragments by truncating the nucleic acid molecule, *e.g.*, an OSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically
15 cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment, preferably a fragment of an OSP, may be produced by chemical or enzymatic cleavage of an OSP
20 polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule of the present invention encoding a fragment, preferably of an OSP, in a host cell.

 Polypeptides of the present invention are also inclusive of mutants, fusion proteins, homologous proteins and allelic variants.

25 A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native polypeptide. Small deletions and insertions can often be found that do not alter the function of a protein. Muteins may or may not be ovarian-specific. Preferably, the
30 mutein is ovarian-specific. More preferably the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 62-116. Accordingly, in a preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more

preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 62-116. In a yet more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 62-116.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein is produced from a host cell comprising a mutated nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid molecule of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is ovarian-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), as well as United States Patent No. 5,223,408, which is herein incorporated by reference in its entirety.

The invention also contemplates polypeptides that are homologous to a polypeptide of the invention. In a preferred embodiment, the polypeptide is homologous to an OSP. In an even more preferred embodiment, the polypeptide is homologous to an OSP selected from the group having an amino acid sequence of SEQ ID NO: 62-116. By homologous polypeptide it is meant one that exhibits significant sequence identity to an OSP, preferably an OSP having an amino acid sequence of SEQ ID NO: 62-116. By

-60-

significant sequence identity it is meant that the homologous polypeptide exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 62-116. More preferred are
5 homologous polypeptides exhibiting at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to an OSP comprising an amino acid sequence of SEQ ID NO: 62-116. Most preferably, the homologous polypeptide exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to an OSP comprising an
10 amino acid sequence of SEQ ID NO: 62-116. In a preferred embodiment, the amino acid substitutions of the homologous polypeptide are conservative amino acid substitutions as discussed above.

Homologous polypeptides of the present invention also comprise polypeptide encoded by a nucleic acid molecule that selectively hybridizes to an OSNA or an antisense
15 sequence thereof. In this embodiment, it is preferred that the homologous polypeptide be encoded by a nucleic acid molecule that hybridizes to an OSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. More preferred is a homologous polypeptide encoded by a nucleic acid sequence which hybridizes to an OSNA selected from the group consisting of SEQ ID NO: 1-61 or a homologous
20 polypeptide encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes an OSP, preferably an OSP of SEQ ID NO: 62-116 under low stringency, moderate stringency or high stringency conditions, as defined herein.

Homologous polypeptides of the present invention may be naturally-occurring and derived from another species, especially one derived from another primate, such as
25 chimpanzee, gorilla, rhesus macaque, or baboon, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 62-116. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the OSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a
30 non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be

-61-

isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. The homologous polypeptide may also be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. Alternatively, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of an OSP. In a preferred embodiment, the homologous polypeptide encodes a polypeptide that is an OSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated polypeptide not only identical in sequence to those described with particularity herein, but also to provide isolated polypeptide ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, polypeptides of the present invention are also inclusive of those encoded by an allelic variant of a nucleic acid molecule encoding an OSP. In this embodiment, it is preferred that the polypeptide be encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 62-116. More preferred is that the polypeptide be encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1-61.

Polypeptides of the present invention are also inclusive of derivative polypeptides encoded by a nucleic acid molecule according to the instant invention. In this embodiment, it is preferred that the polypeptide be an OSP. Also preferred are derivative polypeptides having an amino acid sequence selected from the group consisting of SEQ ID NO: 62-116 and which has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ¹²⁵I, ³²P, ³⁵S, and ³H. In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents,

-62-

enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common
 5 modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.),
 10 Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there
 15 are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting
 20 sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

25 General examples of types of post-translational modifications include, but are not limited to: (Z)-dehydrobutyrine; 1-chondroitin sulfate-L-aspartic acid ester; 1'-glycosyl-L-tryptophan; 1'-phospho-L-histidine; 1-thioglycine; 2'-(S-L-cysteinyl)-L-histidine; 2'-[3-carboxamido (trimethylammonio)propyl]-L-histidine; 2'-alpha-mannosyl-L-tryptophan; 2-methyl-L-glutamine; 2-oxobutanoic acid; 2-pyrrolidone carboxylic acid; 3'-(1'-L-histidyl)-
 30 L-tyrosine; 3'-(8alpha-FAD)-L-histidine; 3'-(S-L-cysteinyl)-L-tyrosine; 3', 3'', 5'-triiodo-L-thyronine; 3'-4'-phospho-L-tyrosine; 3-hydroxy-L-proline; 3'-methyl-L-histidine; 3-methyl-L-lanthionine; 3'-phospho-L-histidine; 4'-(L-tryptophan)-L-tryptophyl quinone; 42 N-cysteinyl-glycosylphosphatidylinositoethanolamine; 43 -(T-L-histidyl)-L-tyrosine; 4-

-63-

hydroxy-L-arginine; 4-hydroxy-L-lysine; 4-hydroxy-L-proline; 5'-(N6-L-lysine)-L-topaquinone; 5-hydroxy-L-lysine; 5-methyl-L-arginine; alpha-l-microglobulin-Ig alpha complex chromophore; bis-L-cysteinyl bis-L-histidino dfiron disulfide; bis-L--cysteinyl-L-N3'-histidino-L-serinyI tetrairon' tetrasulfide; chondroitin sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-alanine; D-allo-isoleucine; D-asparagine; dehydroalanine; dehydrotyrosine; dermatan 4-sulfate D-glueuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-glucuronyl-N-glycine; dipyrrolylmethanemethyl-L-cysteine; D-leucine; D-methionine; D-phenylalanine; D-serine; D-tryptophan; glycine amide; glycine oxazolecarboxylic acid; glycine thiazolecarboxylic acid; heme P450-bis-L-cysteine-L-tyrosine; heme-bis-L-cysteine; hemediol-L-aspartyl ester-L-glutamyl ester; hemediol-L-aspartyl ester-L-glutamyl ester-L-methionine sulfonium; heme-L-cysteine; heme-L-histidine; heparan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; herne P450-bis-L-cysteine-L-lysine; hexakis-L-cysteinyl hexairon hexasulfide; keratan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-threonine; L-oxoalanine- lactic acid; L phenyllactic acid; l'-(8alpha-FAD)-L-histidine; L-2'.4',5'-topaquinone; L-3',4'-dihydroxyphenylalanine; L-3'.4'.5'-trihydroxyphenylalanine; L-4'-bromophenylalanine; L-6'-bromotryptophan; L-alanine amide; L-alanyl imidazolinone glycine; L-allysine; L-arginine amide; L-asparagine amide; L-aspartic 4-phosphoric anhydride; L-aspartic acid 1-amide; L-beta-methylthioaspartic acid; L-bromohistidine; L-citrulline; L-cysteine amide; L-cysteine glutathione disulfide; L-cysteine methyl disulfide; L-cysteine methyl ester; L-cysteine oxazolecarboxylic acid; L-cysteine oxazolinecarboxylic acid; L-cysteine persulfide; L-cysteine sulfenic acid; L-cysteine sulfinic acid; L-cysteine thiazolecarboxylic acid; L-cysteinyl homocitryl molybdenum-heptairon-nonasulfide; L-cysteinyl imidazolinone glycine; L-cysteinyl molybdopterin; L-cysteinyl molybdopterin guanine dinucleotide; L-cystine; L-erythro-beta-hydroxyasparagine; L-erythro-beta-hydroxyaspartic acid; L-gamma-carboxyglutamic acid; L-glutamic acid 1-amide; L-glutamic acid 5-methyl ester; L-glutamine amide; L-glutamyl 5-glycerolphosphorylethanolarnine; L-histidine amide; L-isoglutamyl-polyglutamic acid; L-isoglutamyl-polyglycine; L-isoleucine amide; L-lanthionine; L-leucine amide; L-lysine amide; L-lysine thiazolecarboxylic acid; L-lysinoalanine; L-methionine amide; L-methionine sulfone; L-phenylalanine thiazolecarboxylic acid; L-phenylalanine amide; L-proline amide; L-selenocysteine; L-selenocysteinyl molybdopterin guanine dinucleotide; L-serine amide; L-serine thiazolecarboxylic acid; L-seryl imidazolinone glycine; L-T-

bromophenylalanine; L-T-bromophenylalanine; L-threonine amide; L-thyroxine; L-tryptophan amide; L-tryptophyl quinone; L-tyrosine amide; L-valine amide; meso-lanthionine; N-(L-glutamyl)-L-tyrosine; N-(L-isoaspartyl)-glycine; N-(L-isoaspartyl)-L-cysteine; N,N,N-trimethyl-L-alanine; N,N-dimethyl-L-proline; N2-acetyl-L-lysine; N2-succinyl-L-tryptophan; N4-(ADP-ribosyl)-L-asparagine; N4-glycosyl-L-asparagine; N4-hydroxymethyl-L-asparagine; N4-methyl-L-asparagine; N5-methyl-L-glutamine; N6-1-carboxyethyl-L-lysine; N6-(4-amino hydroxybutyl)-L-lysine; N6-(L-isoglutamyl)-L-lysine; N6-(phospho-5'-adenosine)-L-lysine; N6-(phospho-5'-guanosine)-L-lysine; N6,N6,N6-trimethyl-L-lysine; N6,N6-dimethyl-L-lysine; N6-acetyl-L-lysine; N6-biotinyl-L-lysine; N6-carboxy-L-lysine; N6-formyl-L-lysine; N6-glycyl-L-lysine; N6-lipoyl-L-lysine; N6-methyl-L-lysine; N6-methyl-N6-poly(N-methyl-propylamine)-L-lysine; N6-mureinyl-L-lysine; N6-myristoyl-L-lysine; N6-palmitoyl-L-lysine; N6-pyridoxal phosphate-L-lysine; N6-pyruvic acid 2-iminyl-L-lysine; N6-retinal-L-lysine; N-acetyl-glycine; N-acetyl-L-glutamine; N-acetyl-L-alanine; N-acetyl-L-aspartic acid; N-acetyl-L-cysteine; N-acetyl-L-glutamic acid; N-acetyl-L-isoleucine; N-acetyl-L-methionine; N-acetyl-L-proline; N-acetyl-L-serine; N-acetyl-L-threonine; N-acetyl-L-tyrosine; N-acetyl-L-valine; N-alanyl-glycosylphosphatidylinositoethanolamine; N-asparaginyl-glycosylphosphatidylinositoethanolamine; N-aspartyl-glycosylphosphatidylinositoethanolamine; N-formylglycine; N-formyl-L-methionine; N-glycyl-glycosylphosphatidylinositoethanolamine; N-L-glutamyl-poly-L-glutamic acid; N-methylglycine; N-methyl-L-alanine; N-methyl-L-methionine; N-methyl-L-phenylalanine; N-myristoyl-glycine; N-palmitoyl-L-cysteine; N-pyruvic acid 2-iminyl-L-cysteine; N-pyruvic acid 2-iminyl-L-valine; N-seryl-glycosylphosphatidylinositoethanolamine; N-seryl-glycosylsphingolipidinositoethanolamine; O-(ADP-ribosyl)-L-serine; O-(phospho-5'-adenosine)-L-threonine; O-(phospho-5'-DNA)-L-serine; O-(phospho-5'-DNA)-L-threonine; O-(phospho-5'rRNA)-L-serine; O-(phosphoribosyl dephospho-coenzyme A)-L-serine; O-(sn-1-glycerophosphoryl)-L-serine; O4'-(8alpha-FAD)-L-tyrosine; O4'-(phospho-5'-adenosine)-L-tyrosine; O4'-(phospho-5'-DNA)-L-tyrosine; O4'-(phospho-5'-RNA)-L-tyrosine; O4'-(phospho-5'-uridine)-L-tyrosine; O4-glycosyl-L-hydroxyproline; O4'-glycosyl-L-tyrosine; O4'-sulfo-L-tyrosine; O5-glycosyl-L-hydroxylysine; O-glycosyl-L-serine; O-glycosyl-L-threonine; omega-N-(ADP-ribosyl)-L-arginine; omega-N,omega-N'-dimethyl-L-arginine; omega-N-methyl-L-arginine; omega-N,omega-N'-dimethyl-L-arginine; omega-N-phospho-L-arginine; O-octanoyl-L-serine. ; O-palmitoyl-L-serine; O-

palmitoyl-L-threonine; O-phospho-L-serine; O-phospho-L-threonine; O-phosphopantetheine-L-serine; phycoerythrobilin-bis-L-cysteine; phycourobilin-bis-L-cysteine; pyrroloquinoline quinone; pyruvic acid; S hydroxycinnamyl-L-cysteine; S-(2-aminovinyl) methyl-D-cysteine; S-(2-aminovinyl)-D-cysteine; S-(6-FW-L-cysteine; S-(8 α -FAD)-L-cysteine; S-(ADP-ribosyl)-L-cysteine; S-(L-isoglutamyl)-L-cysteine; S-12-hydroxyfarnesyl-L-cysteine; S-acetyl-L-cysteine; S-diacylglycerol-L-cysteine; S-diphytanylglycerol diether-L-cysteine; S-farnesyl-L-cysteine; S-geranylgeranyl-L-cysteine; S-glycosyl-L-cysteine; S-glycyl-L-cysteine; S-methyl-L-cysteine; S-nitrosyl-L-cysteine; S-palmitoyl-L-cysteine; S-phospho-L-cysteine; S-phycobiliviolin-L-cysteine; S-phycocyanobilin-L-cysteine; S-phycoerythrobilin-L-cysteine; S-phytochromobilin-L-cysteine; S-selenyl-L-cysteine; S-sulfo-L-cysteine; tetrakis-L-cysteiny diiron disulfide; tetrakis-L-cysteiny iron; tetrakis-L-cysteiny tetrairon tetrasulfide; trans-2,3-cis 4-dihydroxy-L-proline; tris-L-cysteiny triiron tetrasulfide; tris-L-cysteiny triiron trisulfide; tris-L-cysteiny-L-aspartato tetrairon tetrasulfide; tris-L-cysteiny-L-cysteine persulfido-bis-L-glutamato-L-histidino tetrairon disulfide trioxide; tris-L-cysteiny-L-N3'-histidino tetrairon tetrasulfide; tris-L-cysteiny-L-N1'-histidino tetrairon tetrasulfide; and tris-L-cysteiny-L-seriny tetrairon tetrasulfide.

Additional examples of PTMs may be found in web sites such as the Delta Mass database based on Krishna, R. G. and F. Wold (1998). R. H. Angeletti., Posttranslational Modifications. Proteins - Analysis and Design. San Diego, Academic Press. 1: 121-206.; Methods in Enzymology, 193, J.A. McClosky (ed), pages 647-660(1990); Methods in Protein Sequence Analysis edited by Kazutomo Imahori and Fumio Sakiyama, Plenum Press, (1993); R.G. Krishna and F. Wold Post-translational modifications of proteins pages 167-172; GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources Cooper et al. *Nucleic Acids Res.* 29:332-335 (2001) O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins Gupta et al. *Nucleic Acids Research*, 27:370-372 (1999); and PhosphoBase, a database of phosphorylation sites: release 2.0., Kreegipuu et al. *Nucleic Acids Res* 27(1):237-239 (1999) see also, WO 02/21139A2.

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from

normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6:485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14:151-162 (1994) and Dennis et al., *Bioessays* 5:412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signalling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10:443-452 (2000) and Khwaja et al., *Lancet* 355:741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein

interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

5 Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for
10 different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the
15 particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

20 In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP
25 kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-
30 translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-

translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule may also be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), e.g., offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa

Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568,
 5 BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

10 The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS,
 15 Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMFA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB,
 20 SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

Polypeptides of the present invention, including full length polypeptides,
 25 fragments and fusion proteins, can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to polypeptides of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

Polypeptides of the present invention, including full length polypeptide, fragments
 30 and fusion proteins, can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-OSP antibodies.

Polypeptides of the present invention, including full length polypeptide, fragments and fusion proteins, can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999). PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

10 Polypeptides of the present invention are also inclusive of analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, this polypeptide is an OSP. In a more preferred embodiment, this polypeptide is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 62-116. Also preferred is an analog polypeptide comprising one or more
15 substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In one embodiment, the analog is structurally similar to an OSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH₂NH--, --CH₂S--, --CH₂-CH₂--, --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂-- and --CH₂SO--. In another embodiment, the analog comprises
20 substitution of one or more amino acids of an OSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three dimensional conformations on the peptide. Other
25 amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Krole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

30 Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are

described, *inter alia*, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993).

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinyl--(9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The FMOC and *t*BOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS--FMOC-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)--TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-

-72-

cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- β -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

Another aspect of the present invention relates to the fusion of a polypeptide of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide of the present invention is an OSP. In a more preferred embodiment, the

-73-

polypeptide of the present invention that is fused to a heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 62-116, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the fusion protein is encoded by a nucleic acid molecule comprising all or part of the nucleic acid sequence of SEQ ID NO: 1-61, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-61.

The fusion proteins of the present invention will include at least one fragment of a polypeptide of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the polypeptide of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of a polypeptide of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and preferably at least 15, 20, or 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particularly useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g., Ausubel, Chapter 16, (1992), supra.*

Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins into the periplasmic space or extracellular milieu for

prokaryotic hosts or into the culture medium for eukaryotic cells through incorporation of secretion signals and/or leader sequences. For example, a His⁶ tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful fusion proteins of the present invention include those that permit use of the polypeptide of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); Colas *et al.*, Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380:548-550(1996); Norman, T. *et al.*, Genetic selection of peptide inhibitors of biological pathways. *Science* 285:591-595(1999), Fabbri *et al.*, Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18:4357-4363(1999); Xu *et al.*, Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94:12473-12478(1997); Yang, *et al.*, Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23:1152-1156(1995); Kolonin *et al.*, Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95:14266-14271(1998); Cohen *et al.*, An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95:14272-14277(1998); Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403:623-627(2000); Ito, *et al.*, A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98:4569-4574(2001). Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains.

Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded polypeptide on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β -galactosidase, biotin *trpE*, protein A, β -lactamase, α -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), *lacZ*, green fluorescent protein (GFP), yeast α mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the OSP.

As further described below, the polypeptides of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize polypeptides of the present invention including OSPs and their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly OSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or

-76-

purification of OSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of OSPs.

One may determine whether polypeptides of the present invention including OSPs, muteins, homologous proteins or allelic variants or fusion proteins of the present invention are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the polypeptide at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides or fusion proteins of the present invention is well-known and within the skill of one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated polypeptides or fusion proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated polypeptide or fusion protein of the present invention are used as therapeutic agents, such as in vaccines and replacement therapy, the isolated polypeptides of the present invention are also useful at lower purity. For example, partially purified polypeptides of the present invention can be used as immunogens to raise antibodies in laboratory animals.

-77-

In a preferred embodiment, the purified and substantially purified polypeptides of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides or fusion proteins of the present invention can usefully be
5 attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides or fusion proteins of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic
10 PVDF; so bound, the polypeptides or fusion proteins of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized polypeptide or fusion protein of the present invention.

As another example, the polypeptides or fusion proteins of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and
15 quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter
20 dish, the plastic is typically polystyrene.

The polypeptides and fusion proteins of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the
25 surface-bound polypeptide or fusion protein to indicate biologic interaction there between. The polypeptides or fusion proteins of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-
30 bound polypeptide or fusion protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention. In a preferred embodiment, the antibodies are specific for a polypeptide that is an OSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 62-116, or a fragment, mutein, derivative, analog or fusion protein thereof.

10 The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on an OSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or vis versa. In addition, alternative splice forms of an OSP may be indicative of cancer. Differential degradation of the C or N-terminus of an OSP may also be a marker or target for anticancer therapy. For example, an OSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may

15 selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-OSP polypeptides by at least two-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the polypeptide of the present invention in samples derived from human

20 ovarian.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7}

-79-

M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

5 Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In such case, antibodies to the polypeptides of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the polypeptide of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated
10 and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies
15 therefrom upon specific immunization are described, *inter alia*, in United States Patent Nos. 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced
20 using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived
25 from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention are also usefully obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster), lagomorphs, (typically rabbits), and also larger mammals, such as sheep, goats, cows, and horses; or egg laying birds or reptiles
30 such as chickens or alligators. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization

-80-

protocols, with the polypeptide of the present invention. One form of avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of a polypeptide of the present invention can be used effectively as immunogens when
5 conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide of the present invention to other moieties. For example, polypeptides of the present invention can be
10 produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-
15 established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996). Immunization protocols often include multiple
20 immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990)).

Antibodies from non-human mammals and avian species can be polyclonal or
25 monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the polypeptides of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the polypeptides of the present invention. Antibodies from avian species may have particular advantage in detection of the polypeptides of the present invention, in
30 human serum or tissues (Vikinge *et al.*, *Biosens. Bioelectron.* 13: 1257-1262 (1998). Following immunization, the antibodies of the present invention can be obtained using any art-accepted technique. Such techniques are well-known in the art and are described in detail in references such as Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic

Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); and Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997).

5 Briefly, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the polypeptides of the present invention can be cloned from hybridomas and thereafter
10 expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the polypeptides of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in United States Patent No. 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

15 Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant antibody production of whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies
20 of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. See, *e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to
30 generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. See, *e.g.*, Barbas (2001), *supra*; Kay, *supra*; and Abelson, *supra*.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable

regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell. Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention. For example, antibody fragments of the present invention can be
5 produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. See, e.g., Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); and
10 Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998).

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. See, e.g., Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992).
15

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*,
20 *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995).

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. See, e.g. Pollock *et al.*,
25 *J. Immunol Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); and Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995).

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells. Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81
30 (1998) review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies. Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of

transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999).

The invention further provides antibody fragments that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention. Among such useful fragments are Fab, Fab', Fv, F(ab)'₂, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

The present invention also relates to antibody derivatives that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus are more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985); and United States Patent No. 5,807,715 the disclosure of which is incorporated herein by reference in its entirety. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); and United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are

-84-

incorporated herein by reference in their entireties. Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. Accordingly, the present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention. The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label can usefully be an enzyme that catalyzes production and local deposition of a detectable product. Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-

9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN);
5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium
(INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS);
phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue
5 tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are
luminescent. For example, in the presence of hydrogen peroxide (H₂O₂), horseradish
peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol.
Immediately following the oxidation, the luminol is in an excited state (intermediate
10 reaction product), which decays to the ground state by emitting light. Strong enhancement
of the light emission is produced by enhancers, such as phenolic compounds. Advantages
include high sensitivity, high resolution, and rapid detection without radioactivity and
requiring only small amounts of antibody. *See, e.g., Thorpe et al., Methods Enzymol.* 133:
331-53 (1986); Kricka *et al., J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al., J.*
15 *Biolumin. Chemilumin.* 10(6): 353-9 (1995). Kits for such enhanced chemiluminescent
detection (ECL) are available commercially. The antibodies can also be labeled using
colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.,* for
flow cytometric detection, for scanning laser cytometric detection, or for fluorescent
20 immunoassay, they can usefully be labeled with fluorophores. There are a wide variety of
fluorophore labels that can usefully be attached to the antibodies of the present invention.
For flow cytometric applications, both for extracellular detection and for intracellular
detection, common useful fluorophores can be fluorescein isothiocyanate (FITC),
allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP),
25 Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-
Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488,
Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa
Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc.,
30 Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY
R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY
564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650,
BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B,

-86-

Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

- 5 For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P , ^{32}P , ^{35}S , ^3H , and ^{125}I . As another example, when the antibodies of the present invention are used for
10 radioimmuno-therapy, the label can usefully be ^{228}Th , ^{227}Ac , ^{225}Ac , ^{223}Ra , ^{213}Bi , ^{212}Pb , ^{212}Bi , ^{211}At , ^{203}Pb , ^{194}Os , ^{188}Re , ^{186}Re , ^{153}Sm , ^{149}Tb , ^{131}I , ^{125}I , ^{111}In , ^{105}Rh , $^{99\text{m}}\text{Tc}$, ^{97}Ru , ^{90}Y , ^{90}Sr , ^{88}Y , ^{72}Se , ^{67}Cu , or ^{47}Sc .

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast
15 agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application as for which they were mentioned.

The antibodies of the present invention, including fragments and derivatives
20 thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the polypeptides of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000);
25 and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998).

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or
30 the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, attached to a substrate. Substrates can be porous or nonporous, planar or nonplanar. For example, the antibodies of the present invention

-87-

can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography. For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microsphere can then be used for
5 isolation of cells that express or display the polypeptides of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to
10 provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the polypeptides of the present invention, to one or more of
15 the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has
20 available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

25 Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid
30 molecule encoding an OSP. In a preferred embodiment, the OSP comprises an amino acid sequence selected from SEQ ID NO: 62-116, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise an OSNA of the invention, preferably an OSNA comprising a

nucleotide sequence selected from the group consisting of SEQ ID NO: 1-61, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a
5 targeted disruption or replacement of the endogenous orthologue of the human OSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.,* Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor
10 Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such
15 techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.,* Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and United States Patent No. 4,873,191, herein incorporated by reference in its entirety); retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.,* Van der Putten *et al.*, *Proc.*
20 *Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.,* Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.,* Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.,* Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst;
25 and sperm-mediated gene transfer (*see, e.g.,* Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.,* Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e.,* a
30 nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.,* mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, *e.g.,* head-to-head tandems or head-to-tail tandems. The transgene

may also be selectively introduced into and activated in a particular cell type by following, *e.g.*, the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232- 6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

5 Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques
10 which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

15 Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because
20 of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is
25 appropriate for an experimental model of interest.

 Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions
30 and/or disorders.

 Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a

-90-

cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265: 103-106
5 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g., Smithies et al., Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the
10 invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that
15 contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g., Thomas, supra* and Thompson, *supra*. However this approach can be
20 routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered
25 not to express the polypeptides of the invention (*e.g., knockouts*) are administered to a patient in vivo. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g., lymphocytes*), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the
30 coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g., by transduction* (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures,

-91-

including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, United States Patent Nos. 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention is a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 62-116 and SEQ ID NO: 1-61 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria,

-92-

the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms “nucleic acid sequences of the invention” and “amino acid sequences of the invention” mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis,

sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence. In addition, the invention includes a method of using patterns of expression associated with either the nucleic acids or proteins in a computer-based method to diagnose disease.

Diagnostic Methods for Ovarian Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of an OSNA or an OSP in a human patient that has or may have ovarian cancer, or who is at risk of developing ovarian cancer, with the expression of an OSNA or an OSP in a normal human control. For purposes of the present invention, "expression of an OSNA" or "OSNA expression" means the quantity of OSNA mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of an OSP" or "OSP expression" means the amount of OSP that can be measured by any method known in the art or the level of translation of an OSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing ovarian cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of OSNA or OSP in cells, tissues, organs or bodily fluids compared with levels of OSNA or OSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of an OSNA or OSP in the patient versus the normal human control is associated with the presence of ovarian cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in the structure of the mRNA of an OSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing ovarian cancer in a patient by analyzing changes in an OSP compared to an OSP from a normal patient. These changes include, *e.g.*, alterations, including post translational modifications such as glycosylation and/or phosphorylation of the OSP or changes in the subcellular OSP localization.

For purposes of the present invention, diagnosing means that OSNA or OSP levels are used to determine the presence or absence of disease in a patient. As will be understood by those of skill in the art, measurement of other diagnostic parameters may be required for definitive diagnosis or determination of the appropriate treatment for the disease. The determination may be made by a clinician, a doctor, a testing laboratory, or a patient using an over the counter test. The patient may have symptoms of disease or may be asymptomatic. In addition, the OSNA or OSP levels of the present invention may be used as screening marker to determine whether further tests or biopsies are warranted. In addition, the OSNA or OSP levels may be used to determine the vulnerability or susceptibility to disease.

In a preferred embodiment, the expression of an OSNA is measured by determining the amount of a mRNA that encodes an amino acid sequence selected from SEQ ID NO: 62-116, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the OSNA expression that is measured is the level of expression of an OSNA mRNA selected from SEQ ID NO: 1-61, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acid molecules. OSNA expression may be measured by any method known in the art, such as

those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. OSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of an OSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, OSNA expression may be compared to a known control, such as normal ovarian nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of an OSP is measured by determining the level of an OSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 62-116, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of an OSNA or OSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of ovarian cancer. The expression level of an OSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the OSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See, e.g.*, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the OSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to an OSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-OSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a

protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the OSP will bind to the anti-OSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-OSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the OSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of an OSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure OSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-OSP antibody is attached to a solid support and an allocated amount of a labeled OSP and a sample of interest are incubated with the solid support. The amount of labeled OSP detected which is attached to the solid support can be correlated to the quantity of an OSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of an OSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA)

-97-

with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more OSNAs of interest. In this approach, all or a portion of one or more OSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. As used herein "blood" includes whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of OSNA or OSP includes, without limitation, ovarian tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, ovarian cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary ovarian cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in OSNAs or OSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples are disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the

-98-

expression level of an OSNA or OSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other OSNA or OSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and
5 are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular OSNA or OSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

10 *Diagnosing*

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more OSNA and/or OSP in a sample from a patient suspected of having ovarian cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural
15 alterations of an OSNA and/or OSP and then ascertaining whether the patient has ovarian cancer from the expression level of the OSNA or OSP. In general, if high expression relative to a control of an OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten
20 times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of ovarian cancer, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same
25 cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether ovarian cancer has metastasized in a patient. One may identify whether the ovarian cancer has metastasized by measuring the expression levels and/or structural alterations of one or
30 more OSNAs and/or OSPs in a variety of tissues. The presence of an OSNA or OSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of

-99-

an OSNA or OSP is associated with ovarian cancer. Similarly, the presence of an OSNA or OSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of an OSNA or OSP is associated with ovarian cancer. Further, the presence of a structurally altered OSNA or OSP that is
5 associated with ovarian cancer is also indicative of metastasis.

In general, if high expression relative to a control of an OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells,
10 tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of an OSNA or OSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human
15 control.

Staging

The invention also provides a method of staging ovarian cancer in a human patient. The method comprises identifying a human patient having ovarian cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or
20 structural alterations of one or more OSNAs or OSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression levels of one or more OSNAs or OSPs is determined for each stage to obtain a standard expression level for each OSNA and OSP. Then, the OSNA or OSP expression levels of the OSNA or OSP are determined in a biological sample from a patient whose
25 stage of cancer is not known. The OSNA or OSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the OSNAs and OSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of an OSNA or OSP to determine the stage of an ovarian cancer.

Monitoring

Further provided is a method of monitoring ovarian cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if

-100-

there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the ovarian cancer. The monitoring may determine if there has been a reoccurrence and, if so, determine its nature. The method comprises identifying a human patient that one wants to monitor for ovarian cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more OSNAs or OSPs, and comparing the OSNA or OSP levels over time to those OSNA or OSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in an OSNA or OSP that are associated with ovarian cancer.

If increased expression of an OSNA or OSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of an OSNA or OSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of an OSNA or OSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of an OSNA or OSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of OSNAs or OSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of ovarian cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of an OSNA and/or OSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more OSNAs and/or OSPs are detected. The presence of higher (or lower) OSNA or OSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly ovarian cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more OSNAs and/or OSPs

-101-

of the invention can also be monitored by analyzing levels of expression of the OSNAs and/or OSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent
5 being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in an OSG, thereby determining if a human with the genetic lesion is susceptible to developing ovarian cancer or to determine what genetic lesions are
10 responsible, or are partly responsible, for a person's existing ovarian cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the OSGs of this invention, a chromosomal rearrangement of an OSG, an aberrant modification of an OSG (such as of the methylation pattern of the genomic DNA), or allelic loss of an OSG. Methods to
15 detect such lesions in the OSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous Ovarian Diseases

The present invention also provides methods for determining the expression levels and/or structural alterations of one or more OSNAs and/or OSPs in a sample from a
20 patient suspected of having or known to have a noncancerous ovarian disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of an OSNA and/or OSP, comparing the expression level or structural alteration of the OSNA or OSP to a normal ovarian control, and then ascertaining whether the patient has a noncancerous ovarian disease. In general,
25 if high expression relative to a control of an OSNA or OSP is indicative of a particular noncancerous ovarian disease, a diagnostic assay is considered positive if the level of expression of the OSNA or OSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low
30 expression relative to a control of an OSNA or OSP is indicative of a noncancerous ovarian disease, a diagnostic assay is considered positive if the level of expression of the

-102-

OSNA or OSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

- 5 One having ordinary skill in the art may determine whether an OSNA and/or OSP is associated with a particular noncancerous ovarian disease by obtaining ovarian tissue from a patient having a noncancerous ovarian disease of interest and determining which OSNAs and/or OSPs are expressed in the tissue at either a higher or a lower level than in normal ovarian tissue. In another embodiment, one may determine whether an OSNA or
- 10 OSP exhibits structural alterations in a particular noncancerous ovarian disease state by obtaining ovarian tissue from a patient having a noncancerous ovarian disease of interest and determining the structural alterations in one or more OSNAs and/or OSPs relative to normal ovarian tissue.

Methods for Identifying Ovarian Tissue

15

In another aspect, the invention provides methods for identifying ovarian tissue. These methods are particularly useful in, *e.g.*, forensic science, ovary cell differentiation and development, and in tissue engineering.

- In one embodiment, the invention provides a method for determining whether a
- 20 sample is ovarian tissue or has ovarian tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising ovarian tissue or having ovarian tissue-like characteristics, determining whether the sample expresses one or more OSNAs and/or OSPs, and, if the sample expresses one or more OSNAs and/or OSPs, concluding that the sample comprises ovarian tissue. In a preferred embodiment, the OSNA encodes
- 25 a polypeptide having an amino acid sequence selected from SEQ ID NO: 62-116, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the OSNA has a nucleotide sequence selected from SEQ ID NO: 1-61, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses an OSNA can be accomplished by any method known in the art. Preferred methods include
- 30 hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether an OSP is expressed. Determining whether a sample expresses an OSP can be

-103-

accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the OSP has an amino acid sequence selected from SEQ ID NO: 62-116, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two OSNAs and/or OSPs is
5 determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five OSNAs and/or OSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is ovarian tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are
10 recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into ovarian tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new ovarian tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture
15 medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Ovarian Tissue

In another aspect, the invention provides methods for producing engineered
20 ovarian tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing an OSNA or an OSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of ovarian tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal ovarian tissue comprises a large number of different cell types. Thus, in one embodiment,
25 the engineered ovarian tissue or cells comprises one of these cell types. In another embodiment, the engineered ovarian tissue or cells comprises more than one ovarian cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the ovarian cell tissue. Methods for manipulating culture conditions are well-known in the art.

30 Nucleic acid molecules encoding one or more OSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules

-104-

encode OSPs having amino acid sequences selected from SEQ ID NO: 62-116, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1-61, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, an OSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial ovarian tissue may be used to treat patients who have lost some or all of their ovarian function.

10 Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, fusion proteins, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, or inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises an OSNA or part thereof. In a more preferred embodiment, the OSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-61, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises an OSP or fragment thereof. In a more preferred embodiment, the pharmaceutical composition comprises an OSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 62-116, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-OSP antibody, preferably an antibody that specifically binds to an OSP having an amino acid that is selected from the group consisting of SEQ ID NO: 62-116, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art that is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.),
5 Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000) and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and
10 parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

15 Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and
20 tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium
25 alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (PovidoneTM), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone
30 fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

-106-

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and

-107-

administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as
5 vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran.

10 Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming
15 microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that
20 are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically. For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures,
25 lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid
30 ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

-108-

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example OSP polypeptide, fusion protein, or fragments thereof, antibodies specific for OSP, agonists, antagonists or inhibitors of OSP, which ameliorates the signs or symptoms of the disease or prevent progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model

can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

-110-

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

5 Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of ovarian function. As used herein,
10 “treating” includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term “treating” encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

15 The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for the purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid
20 or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of “naked” nucleic acid vaccination, as further described in United States Patent Nos. 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that
25 the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid molecule of the present invention is administered. The nucleic acid molecule can be delivered in a vector that drives expression of an OSP, fusion protein, or fragment thereof,
30 or without such vector. Nucleic acid compositions that can drive expression of an OSP are administered, for example, to complement a deficiency in the native OSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses,

-111-

adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g., Cid-Arregui, supra.* In a preferred embodiment, the nucleic acid molecule encodes an OSP having the amino acid sequence of SEQ ID NO: 62-116, or a fragment, fusion protein, allelic variant or homolog thereof.

5 In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express an OSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in OSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells
10 encode an OSP having the amino acid sequence of SEQ ID NO: 62-116, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

 Antisense nucleic acid compositions, or vectors that drive expression of an OSG antisense nucleic acid, are administered to downregulate transcription and/or translation of
15 an OSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

 Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of an OSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.,* between positions -10
20 and +10 from the start site, are preferred.

 Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to OSG transcripts, are also useful in therapy. *See, e.g.,* Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al., Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and
25 Sigurdsson *et al., Trends Biotechnol.* 13(8): 286-9 (1995).

 Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the OSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g., Intody et al., Nucleic Acids Res.* 28(21): 4283-90 (2000); and McGuffie *et al., Cancer Res.* 60(14): 3790-9
30 (2000). Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

-112-

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding an OSP, preferably an OSP comprising an amino acid sequence of SEQ ID NO: 62-116, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-61, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an OSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant OSP defect.

Protein compositions are administered, for example, to complement a deficiency in native OSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to OSP. The immune response can be used to modulate activity of OSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate OSP.

In a preferred embodiment, the polypeptide administered is an OSP comprising an amino acid sequence of SEQ ID NO: 62-116, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-61, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of OSP, or to target therapeutic agents to sites of OSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to an OSP comprising an amino acid sequence of SEQ ID NO: 62-116, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to

-113-

an OSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-61, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to an OSP or have a modulatory effect on the expression or activity of an OSP.

5 Modulators which decrease the expression or activity of OSP (antagonists) are believed to be useful in treating ovarian cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of an OSP can also be designed, synthesized and tested for use in the imaging and treatment of ovarian
10 cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the OSPs identified herein. Molecules identified in the library as being capable of binding to an OSP are key candidates for further evaluation for use in the treatment of ovarian cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of an OSP in cells.

15 In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of OSP is administered. Antagonists of OSP can be produced using methods generally known in the art. In particular, purified OSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind
20 and antagonize at least one activity of an OSP.

In other embodiments a pharmaceutical composition comprising an agonist of an OSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and
25 antagonizes or agonizes, respectively, an OSP comprising an amino acid sequence of SEQ ID NO: 62-116, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, an OSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-61, or a part, allelic variant, substantially similar
30 or hybridizing nucleic acid thereof.

Targeting Ovarian Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the ovary or to specific cells in the ovary. In a preferred embodiment, an anti-OSP antibody is
5 linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if ovarian tissue needs to be selectively destroyed. This would be useful for targeting and killing ovarian cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting ovarian cell function.

10 In another embodiment, an anti-OSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring ovarian function, identifying ovarian cancer tumors, and identifying noncancerous ovarian diseases.

EXAMPLES

15 **Example 1: Gene Expression analysis**

Identification of OSGs was carried out by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc, Palo Alto, CA, using the data mining software package CLASP™.

20 The CLASP target gene identification process is focused on, but not limited to, the following 4 CLASP profiles: tissue specific expression, cancer specific expression, differentially expressed in cancer, maximum tissue differential expression.

(1) For these profiles: cDNA libraries were divided into 48 unique
tissue organs. The genes were grouped into gene bins, each bin is a
sequence based cluster grouped together with a common contig.
25 The expression levels for each gene bin were calculated in each
organ. Differential expression significance was calculated with
rigorous statistical significant test considering the influence of
sequence random fluctuations and sampling size of cDNA libraries
from concept published by Audic S and Claverie JM (Genome Res
30 1997 7(10): 986-995: The significance of digital gene expression
profiles).

- (2) Highly expressed organ specific genes were selected based on the percentage abundance level in the targeted organ versus all the other organs (organ-specificity).
- (3) The expression levels of each highly expressed organ-specific gene in the tumor tissue libraries were compared with normal tissue libraries and tissue libraries associated with tumor or disease (cancer-specificity) and analyzed for statistical significance.

5

(4) Target genes exhibiting each CLASP profile criteria were selected

CLASP 1 tissue specific expression profile: In order to meet the organ-specificity

10 criteria, the expression level of the component clones which the gene is composed of must exhibit 3 or more occurrences regardless the total number of genes isolated for the target organ. The percentage abundance level in each organ was calculated to identify the organ with the highest expression percentage level.

CLASP 2 cancer specific expression profile: In order to fulfill the cancer specific

15 criteria, genes must exhibit 0 expression in normal and libraries associated with tumor and disease but not tumor per se. If the gene then exhibited organ-specificity, the gene was selected as a CLASP target for this profile.

CLASP 3 maximum tissue differential expression profile: CLASP targets were selected based on ratio of expression in tumor libraries compared to expression in normal

20 libraries (including normal libraries associated with tumor or disease) for each organ regardless of whether the gene exhibited organ-specificity. This profile was divided into 2 sub-profiles, since the ratio of expression cannot be obtained if no expression is present in normal libraries(including normal libraries associated with tumor or disease). In this case, the maximum expression percentage of the gene, as calculated by the occurrence of the

25 gene divided by the occurrence of all genes in the target organ, was used. CLASP selects the top 50 targets for each sub-profile.

CLASP 4 maximum tissue differential expression profile with negligible expression in normal tissues: CLASP targets were selected based on ratio of expression in tumor libraries compared to expression in normal libraries (including normal libraries

30 associated with tumor or disease) for each organ regardless of whether the gene exhibited organ-specificity.

CLASP 5 differentially expressed in cancer profile: Expression levels in tumor libraries in each organ and normal libraries (including normal libraries associated with

-116-

cancer or disease) for all organs were obtained and statistically analyzed. If the gene exhibited 90% of confidence that it is over-expressed in tumor libraries in the target organ than normal libraries for all organs, it was selected as a CLASP target for this profile.

- Accordingly, CLASP allows the identification of highly expressed organ and cancer specific genes based on the gene expression levels in each tissue organ. CLASP scores for a portion of the OSG of this invention are listed below.

The CLASP scores for SEQ ID NO:1-61 are listed below:

SEQ ID	Organ	Exp.	Organ	Exp.	Organ	Exp.	Organ	Exp.
NO: 1	OVR	3 - .002	EYE	2 - .0125				
NO: 2	OVR	3 - .002	EYE	2 - .0125				
NO: 3	OVR	3 - .002	FTS	1 - .0001				
NO: 4	OVR	5 - .0033	BRN	1 - .0001	INS	1 - .0009	FTS	1 - .0001
NO: 5	OVR	3 - .0031	FTS	4 - .0005	BRN	2 - .0003	BLD	1 - .0016
NO: 6	OVR	2 - .0017						
NO: 7	OVR	3 - .002	BRN	1 - .0001				
NO: 8	OVR	3 - .002						
NO: 9	OVR	3 - .002	UNC	1 - .0011	FTS	1 - .0001		
NO: 10	OVR	2 - .0046						
NO: 11	OVR	2 - .0046						
NO: 12	OVR	2 - .0046						
NO: 13	OVR	3 - .002	UTR	1 - .0004				
NO: 14	OVR	3 - .002	PRO	7 - .0024	LNG	2 - .0007	INT	2 - .0228
NO: 15	OVR	3 - .002	KID	1 - .0006	TST	1 - .0011		
NO: 16	OVR	3 - .0046						
NO: 17	OVR	3 - .002	UTR	2 - .0007	BRN	1 - .0001	PRO	1 - .0003
NO: 18	OVR	2 - .0068						
NO: 19	OVR	2 - .0047						
NO: 20	OVR	2 - .0055						
NO: 21	OVR	3 - .002	FTS	5 - .0007	BRN	2 - .0003	UNC	2 - .0022
NO: 22	OVR	2 - .0047						
NO: 23	OVR	2 -						

-117-

		.0055						
NO: 24	OVR	3 - .002						
NO: 25	OVR	3 - .002						
NO: 26	OVR	3 - .002						
NO: 27	OVR	3 - .002						
NO: 28	OVR	2 - .0046						
NO: 29	OVR	3 - .002						
NO: 30	OVR	2 - .0047						
NO: 31	OVR	3 - .002	UNC	1 - .0011	FTS	1 - .0001	PIB	1 - .0045
NO: 32	OVR	4 - .0026	MAM	3 - .0012	CRD	1 - .002	FTS	1 - .0001
NO: 33	OVR	4 - .0026	MAM	3 - .0012	CRD	1 - .002	FTS	1 - .0001
NO: 34	OVR	3 - .002	KID	1 - .0006	KID	1 - .0013	PIT	1 - .0029
NO: 35	OVR	3 - .002	KID	1 - .0006	KID	1 - .0013	PIT	1 - .0029
NO: 36	OVR	2 - .0055	KID	1 - .0057				
NO: 37	OVR	3 - .002	THY	1 - .0019				
NO: 38	OVR	2 - .0032	PTH	1 - .004				
NO: 39	OVR	5 - .0139	KID	1 - .0006	ABD	1 - .0241		
NO: 40	OVR	3 - .002						
NO: 41	OVR	3 - .002	UTR	1 - .0004				
NO: 42	OVR	3 - .002	CON	1 - .0007	TNS	1 - .0016		
NO: 43	OVR	3 - .002	LNG	1 - .0003				
NO: 44	OVR	3 - .002						
NO: 45	OVR	3 - .002						
NO: 46	OVR	3 - .002						
NO: 47	OVR	4 - .0026	FTS	1 - .0001				
NO: 48	OVR	3 - .002	FTS	1 - .0001				
NO: 49	OVR	3 - .002						
NO: 50	OVR	3 - .002						
NO: 51	OVR	3 - .002						
NO: 52	OVR	3 - .002						

-118-

NO: 53	OVR	3 - .002						
NO: 54	OVR	3 - .002	UNC	1 - .0011				
NO: 55	OVR	5 - .0033	INS	1 - .0009				
NO: 56	OVR	4 - .0026						
NO: 57	OVR	4 - .0026						
NO: 58	OVR	4 - .0026						
NO: 59	OVR	3 - .002						
NO: 60	OVR	3 - .002						
NO: 61	OVR	3 - .002						

- OVR Ovary, CON Connective Tissue, UNC Mixed Tissues, INS Intestine, Small, EYE Eye, UTR Uterus, FTS Fetus, KID Kidney, PTH Parathyroid Glands, ABD Abdomen, INT Intestine, LNG Lung, TNS Tonsil / Adenoids, TST Testis, THY Thymus Gland, PIT Pituitary Gland, PRO Prostate, PIB Pineal Body, BLD Bladder, BRN Brain, MAM Breast, CRD Heart.

- There are 2 values for each organ in the format 9 - 0.9999. The first represent the number of occurrences of the gene in the given organ. The 2nd number represents the percentage of the expression of the gene in the given organ.

In addition the expression values for each organ in the format 9 - 0.9999 are listed. Each box first lists the given organ, then it lists a number representing the percentage of the expression of the gene in the given organ.

- Based on sequence alignment with the human genome, the following chromosomal locations were assigned and predicted peptide sequences for each of the targets are listed below:

Nucleotide Name	Chromosomal Location	ORF location	Amino acid Name
DEX0237_1	1p36.33b	308-457	DEX0237_62
DEX0237_2	1p36.33b	308-457	DEX0237_63
DEX0237_3	15p13e	227-24	DEX0237_64
DEX0237_4	19p13.3j	175-489	DEX0237_65
DEX0237_5	17p13.3g	309-1	DEX0237_66
DEX0237_6		276-1	DEX0237_67
DEX0237_7	15p13e	54-602	DEX0237_68
DEX0237_8	2p25.3g	181-53	DEX0237_69
DEX0237_9	8p23.3b	805-659	DEX0237_70
DEX0237_10	15p13e	248-90	DEX0237_71
DEX0237_11	10p15.3d	1-189	DEX0237_72
DEX0237_12	1p36.33b		
DEX0237_13	17p13.3g	1749-1465	DEX0237_73

-119-

DEX0237_14	4p16.3d		
DEX0237_15	5p15.33e	497-1645	DEX0237_74
DEX0237_16	Xp22.33f	4-771	DEX0237_75
DEX0237_17	1p36.33b	62-175	DEX0237_76
DEX0237_18	19p13.3j	393-647	DEX0237_77
DEX0237_19	8p23.3b	201-434	DEX0237_78
DEX0237_20	6p25.3b	137-355	DEX0237_79
DEX0237_21	15p13e	309-3458	DEX0237_80
DEX0237_22	2p25.3g	1108-926	DEX0237_81
DEX0237_23	4p16.3d		
DEX0237_24	6p25.3b	273-82	DEX0237_82
DEX0237_25	20p13f	139-83	DEX0237_83
DEX0237_26	10p15.3d	333-166	DEX0237_84
DEX0237_27	10p15.3d	153-317	DEX0237_85
DEX0237_28	18p11.32c	40-72	DEX0237_86
DEX0237_29	7p22.3d	114-290	DEX0237_87
DEX0237_30	10p15.3d	270-401	DEX0237_88
DEX0237_31	3p26.3c	634-407	DEX0237_89
DEX0237_32	9p24.3b	115-294	DEX0237_90
DEX0237_33	9p24.3b	456-181	DEX0237_91
DEX0237_34	10p15.3d		
DEX0237_35	7p22.3d	526-383	DEX0237_92
DEX0237_36	10p15.3d	1107-1481	DEX0237_93
DEX0237_37	17p13.3g	250-2	DEX0237_94
DEX0237_38	2p25.3g		
DEX0237_39	10p15.3d	276-70	DEX0237_95
DEX0237_40	1p36.33b	359-183	DEX0237_96
DEX0237_41	3p26.3c	730-185	DEX0237_97
DEX0237_42	12p13.33d	640-798	DEX0237_98
DEX0237_43	17p13.3g	484-936	DEX0237_99
DEX0237_44	7p22.3d	323-234	DEX0237_100
DEX0237_45	7p22.3d	49-309	DEX0237_101
DEX0237_46	16p13.3f	102-22	DEX0237_102
DEX0237_47	2p25.3g	500-300	DEX0237_103
DEX0237_48	16p13.3f	375-497	DEX0237_104
DEX0237_49	15p13e	217-14	DEX0237_105
DEX0237_50	13p13e	144-28	DEX0237_106
DEX0237_51	1p36.33b	17-97	DEX0237_107
DEX0237_52	7p22.3d	170-1	DEX0237_108
DEX0237_53	6p25.3b	24-149	DEX0237_109
DEX0237_54	17p13.3g	1124-1	DEX0237_110
DEX0237_55	16p13.3f	558-662	DEX0237_111
DEX0237_56	4p16.3d	643-915	DEX0237_112
DEX0237_57	19p13.3j	65-1	DEX0237_113
DEX0237_58	19p13.3j	181-468	DEX0237_114
DEX0237_59	9p24.3b		
DEX0237_60	11p15.5d	394-47	DEX0237_115
DEX0237_61	7p22.3d	713-1060	DEX0237_116

Example 2A: Custom Microarray Experiment—Ovarian Cancer

Custom oligonucleotide microarrays were provided by Agilent Technologies, Inc. (Palo Alto, CA). The microarrays were fabricated by Agilent using their technology for the *in-situ* synthesis of 60mer oligonucleotides (Hughes, et al. 2001, Nature Biotechnology

19:342-347). The 60mer microarray probes were designed by Agilent, from gene sequences provided by diaDexus, using Agilent proprietary algorithms. Whenever possible two different 60mers were designed for each gene of interest.

All microarray experiments were two-color experiments and were preformed using
5 Agilent-recommended protocols and reagents. Briefly, each microarray was hybridized with cRNAs synthesized from total RNA, isolated from cancer and normal tissues, labeled with fluorescent dyes Cyanine3 and Cyanine5 (NEN Life Science Products, Inc., Boston, MA) using a linear amplification method (Agilent). In each experiment the experimental sample was total RNA isolated from cancer tissue from a single individual and the
10 reference sample was a pool of total RNA isolated from normal tissues of the same organ as the cancerous tissue (*i.e.* normal ovarian tissue in experiments with ovarian cancer samples). Hybridizations were carried out at 60°C, overnight using Agilent *in-situ* hybridization buffer. Following washing, arrays were scanned with a GenePix 4000B Microarray Scanner (Axon Instruments, Inc., Union City, CA). The resulting images were
15 analyzed with GenePix Pro 3.0 Microarray Acquisition and Analysis Software (Axon). Two different chip designs were evaluated with overlapping sets of a total of 19 samples, comparing the expression patterns of ovarian cancer derived total RNA to total RNA isolated from a pool of 9 normal ovarian tissues were analyzed. For Chip2 all 19 samples (14 invasive carcinomas (INV), 5 low malignant potential (LMP) samples) were analyzed
20 and for Chip1 a subset of 17 of these samples (13 invasive carcinomas, 4 low malignant potential samples) were assessed.

Data normalization and expression profiling were done with Expressionist software from GeneData Inc. (Daly City, CA/Basel, Switzerland). Gene expression analysis was performed using only experiments that meet certain quality criteria. The
25 quality criteria that experiments must meet are a combination of evaluations performed by the Expressionist software and evaluations performed manually using raw and normalized data. To evaluate raw data quality, detection limits (the mean signal for a replicated negative control + 2 Standard Deviations (SD)) for each channel were calculated. The detection limit is a measure of non-specific hybridization. Arrays with poor detection
30 limits were not analyzed and the experiments were repeated. To evaluate normalized data quality, positive control elements included in the array were utilized. These array features should have a mean ratio of 1 (no differential expression). If these features have a mean ratio of greater than 1.5-fold up or down, the experiments were not analyzed further and

-121-

were repeated. In addition to traditional scatter plots demonstrating the distribution of signal in each experiment, the Expressionist software also has minimum thresholding criteria that employ user defined parameters to identify quality data. Only those features that meet the threshold criteria were included in the filtering and analyses carried out by Expressionist. The thresholding settings employed require a minimum area percentage of 60% [(% pixels > background + 2SD)-(% pixels saturated)], and a minimum signal to noise ratio of 2.0 in both channels. By these criteria, very low expressors and saturated features were not included in analysis.

Relative expression data was collected from Expressionist based on filtering and clustering analyses. Up- and down- regulated genes were identified using criteria for percentage of valid values obtained, and the percentage of experiments in which the gene is up- or down-regulated. These criteria were set independently for each data set, depending on the size and the nature of the data set. The results for the significantly upregulated and downregulated genes are shown in Table 1 and Table2. The first three columns of each table contain information about the sequence itself (Oligo ID, Parent ID, and Patent#), the next 3 columns show the results obtained. '%valid' indicates the percentage of unique experiments total (n=17 for Chip1, n=19 for Chip2) in which a valid expression value was obtained, '%up' indicates the percentage of experiments in which up-regulation of at least 2-fold was observed, and '%down' indicates the percentage of the experiments in which down-regulation of at least 2-fold was observed. The last column in each table describes the location of the microarray probe (oligo) relative to the parent sequence. For genes that the parent sequence was extended using databases of public and proprietary sequences, the both the parent DEX number and the extended DEX number are listed. In these cases, the site of the 60mer probe is listed in the last column.

Table 1. Ovarian Microarray up-regulated expression data.

Patent	ParentID	OligoID	%valid n=17	%up n=17	%valid up n=17	valid up Invasive n=13	Oligo Start Pos in Par Seq	Oligo Stop Pos in Par Seq
DEX0237_5	33438	99065.01	100	41.2	41.2	46.2	493	552
DEX0237_5	33438	99065.02	100	35.3	35.3	38.5	493	552
DEX0237_12	33414	98977.01	76.5	29.4	38.5	50	569	628
DEX0237_12	33414	98977.02	88.2	35.3	40	50	569	628

Table 2. Ovarian Microarray tissue descriptions.

Samples	Tissue ID	Cancer Subtype	Age	Tumor Description	Grade	Stage
OV.L.MU.478A1B	17478A1B	LMP	33	benign mucinous cystadenoma with focal proliferation	n/a	no
OV.I.SE370V	2370V	INV	47	invasive papillary serous adenocarcinoma	1-2	IV
OV.L1.SE105	S9822105	LMP	29	papillary serous carcinoma LMP	0	IC
OV.L.SE604A2B	14604A2B	LMP	64	serous cystadenofibroma of low malignant potential(Patient w/transitional cell carcinoma, in-situ and invasive of the bladder grade III/III)	n/a	no
OV.I.EN360	9410C360	INV	52	Endometrioid Adenocarcinoma, Stage Tx, Grade I/III	1-3	IV
OV.I.SE/EN0050	10050	INV	60	Papillary serous and endometrioid ovarian carcinoma. Metastatic breast cancer, probable ovarian cancer	3	IV
OV.INV.SE.10400	10400	INV	67	papillary serous adenocarcinoma (Stage IV)	2	IV
OV.INV.SE.1071C	1071C	INV	51	papillary cystadenocarcinoma (Stage IV)	2	IV
OV.INV.MU.009	9507H009	INV	87	mucinous cystadenoma, multiloculated	N/A	
OV.LMP.MU.084	9808A084	LMP	47	mucinous borderline (LMP)	GB	
OV.INV.SE.291D01	VNM00291D01	INV	28	serous cystadenocarcinoma (Stage II)	1	II

-123-

Samples	Tissue ID	Cancer Subtype	Age	Tumor Description	Grade	Stage
OV.I.MU329D01	VNM00329D01	INV	40	mucinous cystadenocarcinoma	3	n/a
OV.INV.MX.010SP1	9803G010SP1	INV	71	poorly diff. clear cell, endometrioid & serous papillary types (Stage IV)	4	III
OV.I.EN/MU.608A	95017608A	INV	56	endometrioid and mucinous adenocarcinoma	2	IV
OV.I.SE814V	1814V	INV	66	Papillary serous adenocarcinoma,	3-4	IV
OV.L.MU638A1C	14638A1C	LMP	26	mucinous cystic neoplasm of low malignant potential.	n/a	no
OV.I.SE/EN471A1B	14471A1B	INV	56	poorly differentiated adenocarcinoma with serous and endometrioid features	3-4	IV
OV.INV.SE.693C	S995693A	INV	54	serous papillary carcinoma (Stage IV)	N/A	
OV.INV.SE.116D04	VNM00116D04	INV	47	cystadenocarcinoma	1	

Example 2B: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels

-124-

for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every
5 sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are
10 relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the OSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal thymus (calibrator). These RNA
15 samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the OSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the
20 cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the OSNAs that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples. Further, the level of mRNA expression in cancer
25 samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1-61 being a diagnostic marker for
30 cancer.

Example 3: Protein Expression

-125-

The OSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the OSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the OSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH₂-terminus of the coding sequence of OSNA, and six histidines, flanking the COOH-terminus of the coding sequence of OSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of OSP is achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that are separated from total cell lysate were incubated with a nickel chelating resin. The column is packed and washed with five column volumes of wash buffer. OSP is eluted stepwise with various concentration imidazole buffers.

Example 4: Fusion Proteins

The human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to

-126-

culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any
5 suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*,
Gastroenterology 80: 225-232 (1981).

10 The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody
15 which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic
20 antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

For the polypeptides of the invention, the following attributes were found, epitopes, post translational modifications, signal peptides and transmembrane domains. Specifically, the Jameson-Wolf methods were used to predict epitopes were predicted.
25 (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988). Examples of post-translational modifications (PTMs) and other motifs of the LSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. The PTMs and other motifs were predicted by using the ProSite Dictionary of Proteins Sites and Patterns (Bairoch et al., Nucleic Acids
30 Res. 25(1):217-221 (1997)), the following motifs, including PTMs, were predicted for the OSPs of the invention. The signal peptides were detected by using the SignalP 2.0, see "Machine learning approaches to the prediction of signal peptides and other protein sorting signals" Henrik Nielsen, Søren Brunak, and Gunnar von Heijne, Protein

- Engineering 12, 3-9 (1999). Prediction of transmembrane helices in proteins was performed by the application TMHMM 2.0, "currently the best performing transmembrane prediction program", according to authors (A. Krogh, B. Larsson, G. von Heijne, and E. L. L. Sonnhammer. Predicting transmembrane protein topology with a hidden Markov
- 5 model: Application to complete genomes. Journal of Molecular Biology, 305(3):567-580, January 2001). S. Moller, M.D.R. Croning, R. Apweiler "Evaluation of methods for the prediction of membrane spanning regions" Bioinformatics, 17(7):646-653, July 2001. E. L.L. Sonnhammer, G. von Heijne, and A. Krogh. A hidden Markov model for predicting transmembrane helices in protein sequences in J. Glasgow, T. Littlejohn, F. Major, R.
- 10 Lathrop, D. Sankoff, and C. Sensen, editors, Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, pages 175-182, Menlo Park, CA, 1998. AAAI Press. Using the PSORT II program, the following cellular localizations (.pst column) (Paul Horton and Kenta Nakai, Better Prediction of Protein Cellular Localization Sites with the k Nearest Neighbors Classifier, Intelligent Systems for
- 15 Molecular Biology 5 147-152 (1997). In the table below are the following: PTM and other motifs (type, amino acid residue locations); Amino acid location and antigenic index (location, AI score, length); TM (number of membrane domain, topology in orientation and position).

name	DEX0237.aa. ai	DEX0237 .aa.pst	DEX0237.aa.ptm	DEX0237. aa.sp
DEX0237_62		cyt	Pkc_Phospho_Site 39-41;	
DEX0237_63		cyt	Pkc_Phospho_Site 39-41;	
DEX0237_64		exc	Ck2_Phospho_Site 18-21;56-59; Pkc_Phospho_Site 23-25;	
DEX0237_65	76-93 1.08 18 3-37 1.05 35	nuc	Ck2_Phospho_Site 8-11;9-12; Pkc_Phospho_Site 76-78;	
DEX0237_66	3-56 1.12 54	nuc	Asn_Glycosylation 49-52; Glycosaminoglycan 58-61; Myristyl 59-64;61-66; Pkc_Phospho_Site 50-52;	
DEX0237_67		pla	Pkc_Phospho_Site 53-55; Prokar_Lipoprotein 26-36;	
DEX0237_68		nuc	Ck2_Phospho_Site 3-6;60-63;87- 90;146-149; Glycosaminoglycan 155-158; Myristyl 56-61;91- 96;116-121;126-131;158-163; Pkc_Phospho_Site 43-45;163- 165;170-172;	

-128-

DEX0237_69	18-31 1.04 14	cyt	Pkc_Phospho_Site 26-28; Tyr_Phospho_Site 27-35;28-35;	
DEX0237_70		exc	Prokar_Lipoprotein 14-24;	
DEX0237_71		nuc	Ck2_Phospho_Site 6-9;15-18; Pkc_Phospho_Site 6-8;43-45;	
DEX0237_72	47-60 1.12 14	pla	Camp_Phospho_Site 56-59; Ck2_Phospho_Site 20-23; Crystallin_Betagamma 22-37; Myristyl 16-21;	
DEX0237_73	15-27 1.00 13	nuc	Ck2_Phospho_Site 21-24; Myristyl 17-22; Pkc_Phospho_Site 18-20;41- 43;59-61;79-81;85-87;	
DEX0237_74	64-92 1.25 29 311-349 1.06 39	nuc	Asn_Glycosylation 113-116;134- 137;159-162;327-330; Camp_Phospho_Site 88-91;199- 202; Ck2_Phospho_Site 51- 54;100-103;115-118;155- 158;161-164;211-214;379-382; Glycosaminoglycan 283-286; Myristyl 18-23;285-290;289- 294;323-328;352-357; Pkc_Phospho_Site 22-24;51- 53;57-59;100-102;197-199;198- 200;202-204;230-232;304- 306;343-345;364-366;371-373;	
DEX0237_75	93-104 1.10 12 121-135 1.09 15 153-166 1.01 14	pla	Asn_Glycosylation 32-35;38- 41;134-137;169-172;236-239; Ck2_Phospho_Site 25-28; Myristyl 36-41;114-119;148- 153;162-167;235-240;	21 0.997 0.948
DEX0237_76		cyt	Ck2_Phospho_Site 10-13;	
DEX0237_77		cyt	Ck2_Phospho_Site 15-18;24-27; Myristyl 48-53;61-66; Pkc_Phospho_Site 7-9;62-64;	
DEX0237_78		exc	Ck2_Phospho_Site 59-62; Myristyl 22-27;23-28;32-37;46- 51;	
DEX0237_79	32-43 1.36 12	nuc	Asn_Glycosylation 63-66; Myristyl 39-44;43-48; Pkc_Phospho_Site 40-42;	

-129-

DEX0237_80	612-624 1.22 13	nuc	Asn_Glycosylation 48-51;136- 139;168-171;298-301;577- 580;589-592;677-680;693- 696;892-895; Camp_Phospho_Site 67-70;140- 143;698-701;934-937; Ck2_Phospho_Site 6-9;90- 93;231-234;232-235;236- 239;243-246;251-254;255- 258;312-315;342-345;347- 350;352-355;432-435;476- 479;647-650;652-655;664- 667;758-761;779-782;792- 795;903-906;937-940; Glycosaminoglycan 553-556; Myristyl 41-46;82-87;134- 139;167-172;200-205;222- 227;331-336;368-373;445- 450;517-522;554-559;747- 752;755-760;913-918;954-959; Pkc_Phospho_Site 112-114;138- 140;361-363;406-408;412- 414;432-434;495-497;540- 542;579-581;619-621;652- 654;679-681;758-760;809- 811;839-841;847-849;914- 916;927-929;982-984; Tyr_Phospho_Site 381-387;	
	775-789 1.21 15			
	487-510 1.16 24			
	308-325 1.12 18			
	287-302 1.11 16			
	704-743 1.10 40			
	104-149 1.07 46			
	6-67 1.07 62			
	663-702 1.06 40			
	586-603 1.05 18			
	454-463 1.01 10			
	337-414 1.01 78			
	163-195 1.01 33			
DEX0237_81		pla	Pkc_Phospho_Site 4-6;	
DEX0237_82	37-46 1.10 10	pla	Ck2_Phospho_Site 41-44; Myristyl 2-7;6-11; Pkc_Phospho_Site 41-43;	18 0.909 0.659

-130-

DEX0237_83			Pkc_Phospho_Site 8-10;	
DEX0237_84	12-31 1.11 20	nuc	Asn_Glycosylation 13-16; Ck2_Phospho_Site 40-43; Myristyl 9-14;	
DEX0237_85		nuc	Camp_Phospho_Site 35-38; Ck2_Phospho_Site 4-7; Myristyl 42-47; Pkc_Phospho_Site 4-6;	
DEX0237_86			Glycosaminoglycan 6-9; Pkc_Phospho_Site 6-8;	
DEX0237_87		pla	Pkc_Phospho_Site 2-4;8-10;	
DEX0237_88		pla	Myristyl 8-13;	20 0.983 0.864
DEX0237_89	49-66 1.08 18	mit	Myristyl 11-16;	
DEX0237_90	44-58 1.01 15	cyt	Amidation 50-53; Pkc_Phospho_Site 23-25;	
DEX0237_91		nuc	Camp_Phospho_Site 59-62; Myristyl 25-30;66-71; Pkc_Phospho_Site 19-21;43- 45;57-59;	
DEX0237_92		nuc	Ck2_Phospho_Site 3-6; Pkc_Phospho_Site 14-16;	
DEX0237_93		cyt	Ck2_Phospho_Site 22-25;53- 56;118-121; Myristyl 104-109; Pkc_Phospho_Site 7-9;46-48;	26 0.958 0.748
DEX0237_94	11-23 1.03 13	nuc		
DEX0237_95		nuc	Asn_Glycosylation 64-67; Ck2_Phospho_Site 12-15;43-46; Pkc_Phospho_Site 43-45;	
DEX0237_96	23-33 1.10 11	nuc	Amidation 7-10; Asn_Glycosylation 38-41; Camp_Phospho_Site 9-12; Myristyl 29-34; Pkc_Phospho_Site 3-5;12-14;	
DEX0237_97	121-139 1.07 19 8-20 1.00 13	nuc	Ck2_Phospho_Site 14-17;67- 70;134-137; Myristyl 8-13;88- 93;99-104; Pkc_Phospho_Site 38-40;46-48;93-95;130-132;	
DEX0237_98		nuc	Ck2_Phospho_Site 18-21; Tyr_Phospho_Site 17-24;	
DEX0237_99	110-121 1.19 12	mit	Camp_Phospho_Site 102-105; Ck2_Phospho_Site 81-84;111- 114;	
DEX0237_100		mit		18 0.967 0.869
DEX0237_101		nuc	Asn_Glycosylation 30-33; Myristyl 43-48;	

-131-

DEX0237_102		nuc	Myristyl 10-15;	
DEX0237_103	48-60 1.18 13 28-45 1.09 18 17-26 1.03 10	nuc	Amidation 39-42;52-55; Asn_Glycosylation 57-60; Pkc_Phospho_Site 46-48;	18 0.894 0.598
DEX0237_104		cyt	Ck2_Phospho_Site 20-23; Myristyl 4-9;	
DEX0237_105		pla	Pkc_Phospho_Site 11-13;	
DEX0237_106		cyt	Pkc_Phospho_Site 13-15;	
DEX0237_107		nuc	Amidation 8-11;	
DEX0237_108	15-36 1.13 22 38-50 1.09 13	nuc	Myristyl 35-40;50-55; Pkc_Phospho_Site 11-13;40-42;	
DEX0237_109	25-36 1.06 12	nuc	Ck2_Phospho_Site 11-14;13-16; 38-41; Myristyl 5-10; Pkc_Phospho_Site 31-33;	
DEX0237_110	171-187 1.26 17 354-367 1.18 14 333-349 1.14 17 25-41 1.11 17 297-325 1.06 29	pla	Amidation 65-68;283-286; Ck2_Phospho_Site 78-81;148-151; 161-164; Leucine_Zipper 43-64; Myristyl 48-53;65-70;75-80; 107-112;136-141;159-164; 245-250;249-254;268-273; 275-280;278-283;303-308; 340-345;347-352; Pkc_Phospho_Site 2-4;26-28;34-36; 224-226;362-364; Prokar_Lipoprotein 152-162; Tyr_Phospho_Site 193-199;	
DEX0237_111		cyt		22 0.912 0.652
DEX0237_112		nuc	Camp_Phospho_Site 55-58;	
DEX0237_113		nuc	Pkc_Phospho_Site 4-6;7-9;	
DEX0237_114		nuc	Camp_Phospho_Site 7-10; Ck2_Phospho_Site 70-73; Myristyl 48-53; Pkc_Phospho_Site 5-7;70-72;	

-132-

DEX0237_115	34-46 1.10 13	nuc	Ck2_Phospho_Site 31-34; Myristyl 15-20;29-34;67-72;86-91;	
DEX0237_116	34-46 1.10 13	nuc	Ck2_Phospho_Site 31-34; Myristyl 15-20;29-34;67-72;86-91;	

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

5 RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1-61. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; 10 and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products 15 analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

20 Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

25 Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal 30 fractional length measurements are performed using the ISee Graphical Program System.

-133-

(Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological

5 Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that
10 non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 µl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is
15 added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 µl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared,
20 using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion
25 consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

30 As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 , µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to

-134-

therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μ g/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous
5 subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally,
10 topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and
15 intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of
20 L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., *Biopolymers* 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., *J. Biomed. Mater. Res.* 15: 167-277 (1981), and R. Langer, *Chem. Tech.* 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes
25 containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., *Proc. Natl. Acad. Sci. USA* 82: 3688-3692 (1985); Hwang et al., *Proc. Natl. Acad. Sci. USA* 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in
30 which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage

-135-

injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and
5 other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such
10 carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the
15 dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid,
20 aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration
25 of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron
30 membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

-136-

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

15 **Example 9: Method of Treating Decreased Levels of the Polypeptide**

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0

-137-

and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

5 **Example 11: Method of Treatment Using Gene Therapy**

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture
10 flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

15 At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., *DNA*, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf
20 intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a
25 HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector
30 has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to

-138-

the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

5 Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

10 If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after
15 having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-In Vivo

Another aspect of the present invention is using in vivo gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the
20 introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the
25 art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent No. 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. *Cardiovasc. Res.* 35(3):470-479(1997), Chao J et al. *Pharmacol. Res.* 35(6):517-522(1997), Wolff J. A. *Neuromuscul. Disord.* 7(5):314-318(1997), Schwartz B. et al. *Gene Ther.* 3(5):405-411(1996), and Tsurumi Y. et al. *Circulation* 94(12):3281-3290(1996).

30 The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, ovarian, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. *Ann. NY Acad. Sci.* 772:126-139(1995), and Abdallah B. et al. *Biol. Cell* 85(1):1-7(1995)) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, ovarian, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. In vivo muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 µg/kg body weight to about 50 mg/kg body

-140-

weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined
5 by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked
10 polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle in vivo is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard
15 recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the
20 anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

25 After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 μ m cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection
30 may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (I.e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., *Appl. Microbiol. Biotechnol.* 40:691-698 (1994); Carver et al., *Biotechnology (NY)* 11:1263-1270 (1993); Wright et al., *Biotechnology (NY)* 9:830-834 (1991); and Hoppe et al., U. S. Pat. No. 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (Lo, *Mol Cell. Biol.* 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., *Science* 259: 1745 (1993)); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., *Cell* 57: 717-723 (1989). For a review of such techniques, see Gordon, "Transgenic Animals," *Intl. Rev. Cytol.* 115: 171-229 (1989).

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., *Nature* 380: 64-66 (1996); Wilmut et al., *Nature* 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e.g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992)). The regulatory sequences required for

-142-

such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some
5 nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., *Science* 265: 103-106 (1994)). The regulatory sequences required for such a
10 cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished
15 by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR).
20 Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding
25 strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to
30 both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E.g., see Smithies et al., *Nature* 317: 230-234 (1985); Thomas & Capecchi, *Cell* 51: 503-512 (1987); Thompson et al., *Cell* 5: 313-321 (1989)). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e.g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient in vivo. Such cells may be obtained from the patient (I.e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence

associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

5 The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

10 Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent No. 5,399,349; and Mulligan & Wilson, U. S. Patent No. 5,460,959 each of which is incorporated by reference
15 herein in its entirety).

 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of
20 components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

 Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with
25 aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

 While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not
30 by way of limitation. The present invention is limited only by the claims that follow.

-145-

What is Claimed is:

1. An isolated nucleic acid molecule comprising:
 - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 68, 70, 73, 75, 80, 93, 96, 97 or 110;
 - 5 (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 7, 9, 13, 16, 21, 34, 36, 40, 41 or 54;
 - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or
 - (d) a nucleic acid molecule having at least 90% sequence identity to the nucleic acid molecule of (a) or (b).
- 10 2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.
- 15 3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.
- 20 5. The nucleic acid molecule according to claim 4, wherein the nucleic acid molecule is a human nucleic acid molecule.
- 25 6. A method for determining the presence of an ovarian specific nucleic acid (OSNA) in a sample, comprising the steps of:
 - (a) contacting the sample with the nucleic acid molecule of SEQ ID NO: 7, 9, 13, 16, 18, 21, 34, 36, 40, 41 or 54 under conditions in which the nucleic acid molecule will selectively hybridize to an ovarian specific nucleic acid; and
 - (b) detecting hybridization of the nucleic acid molecule to an OSNA in the sample, wherein the detection of the hybridization indicates the presence of an OSNA in the sample.
- 30 7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.
9. A method for producing a polypeptide encoded by the nucleic acid molecule
5 according to claim 1, comprising the steps of:
- (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and
 - (b) incubating the host cell under conditions in which the polypeptide is produced.
- 10
10. A polypeptide encoded by the nucleic acid molecule according to claim 1.
11. An isolated polypeptide selected from the group consisting of:
- (a) a polypeptide comprising an amino acid sequence with at least 60%
15 sequence identity to of SEQ ID NO: 68, 70, 73, 75, 80, 93, 96, 97 or 110 ; or
 - (b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 90% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 7, 9, 13, 16, 21, 34, 36, 40, 41 or 54.
- 20
12. An antibody or fragment thereof that specifically binds to
- (a) a polypeptide comprising an amino acid sequence with at least 60%
sequence identity to of SEQ ID NO: 68, 70, 73, 75, 77, 80, 93, 96, 97 or 110 ; or
 - (b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 90% sequence identity to a nucleic acid molecule comprising
25 a nucleic acid sequence of SEQ ID NO: 7, 9, 13, 16, 18, 21, 34, 36, 40, 41 or 54.
13. A method for determining the presence of an ovarian specific protein in a sample, comprising the steps of:
- (a) contacting the sample with a suitable reagent under conditions in which the
30 reagent will selectively interact with the ovarian specific protein comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 68, 70, 73, 75, 77, 80, 93, 96, 97 or 110; and

-147-

(b) detecting the interaction of the reagent with an ovarian specific protein in the sample, wherein the detection of binding indicates the presence of an ovarian specific protein in the sample.

5 14. A method for diagnosing or monitoring the presence and metastases of ovarian cancer in a patient, comprising the steps of:

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient; and

10 (b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the ovarian specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of ovarian cancer.

15 15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient.

20 16. A method of treating a patient with ovarian cancer, comprising the step of administering a composition according to claim 11 or 12 to a patient in need thereof, wherein said administration induces an immune response against the ovarian cancer cell expressing the nucleic acid molecule or polypeptide.

25 17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

30

1

SEQUENCE LISTING

<110> diaDexus, Inc.
 Sun, Yongming
 Liu, Chenghua
 Salceda, Susana

<120> Compositions and Methods Relating to Ovary Specific Genes and Proteins

<130> DEX-0345

<150> US 60/316,307

<151> 2001-08-31

<160> 116

<170> PatentIn version 3.1

<210> 1

<211> 971

<212> DNA

<213> Homo sapien

<400> 1

```

acatTTTTtct gtatgtagtc agattataag tgcttgctaa aattgagtga aaaaatagga      60
cagagggtatc tttacccaat aactgtctgg attgtgagat gaaaatacag ctctctagcc      120
agttggctat ttttcaagac tttactgtgt acatctttcc gtttactttc atcttcctac      180
cgTTTTtagat gtctcatgtc cttctgttgt tttctctaga ggacaaaaaa gaggctctga      240
ggagggttaca taattaactc aaacttatgt ggtaacaata aaattcttgg aaatacaagg      300
atggctaattg acctatattg ttttaatcac ctattttctc tataattatc ctctgcccaa      360
ggcaacagaa tggagtgcac tttacccgtc cttccagata aacaaaaaaa cacttgataa      420
aagtgaaaaa aagggcccta tgatacaatt aaaatgattg ctattacact aaaagaaaaa      480
aaatattgct taacctgtca caacatagtt ttttttagcat tgtttatcaa ggccagacta      540
tagcaagtta tgtccatggc aatagttgct aacttggtc attctgctac ctgaggatat      600
ttttctatct aacctcttta aggaaggaga atatactaac ttaaattggat taaaatgtag      660
catatacttg gtttgtagtc ttgattgtaa attaacttac cagtcaaagc tgctgctgct      720
tttttttttt tttggacaga attctggcct ctgtcaaccc aagcttgaag ttcaatgggt      780
gtgaatctca agttcacttg caaacttccg gccttcccgg gggttcaaag ttgaattcct      840
ccttgcccct tagccttttc cggagtaagc tggggaattt acaggggaaa ggaaccaaac      900
caaggcccaa gataaatttg tttttttggg atttttttgg ttagaggaat ggggggtttc      960
aaccaatatt g                                     971

```

<210> 2

2

<211> 972
 <212> DNA
 <213> Homo sapien

<400> 2
 acatttttct gtatgtagtc agattataag tgcttgctaa aattgagtga aaaaataggā 60
 cagaggtatc tttacccaat aactgtctgg attgtgagat gaaaatacag ctctctagcc 120
 agttggctat ttttcaagac tttactgtgt acatctttcc gtttactttc atcttcctac 180
 cgttttagat gtctcatgtc cttctgttgt tttctctaga ggacaaaaaa gaggctctga 240
 ggaggttaca taattaactc aaacttatgt ggtaacaata aaattcttgg aaatacaagg 300
 atggctaata acctatattg ttttaatcac ctattttctc tataattatc ctctgcccaa 360
 ggcaacagaa tggagtgcac tttaccgctc cttccagata aacaaaaaaa cacttgataa 420
 aagtgaaaaa aaggggcccta tgatacaatt aaaatgattg ctattacact aaaagaaaaa 480
 aaatattgct taacctgtca caacatagtt ttttttagcat tgtttatcaa ggccagacta 540
 tagcaagtta tgtccatggt aatagttgct aacttggtc attctgctac ctgaggatat 600
 ttttctatct aacctcttta aggaaggaga atatactaac ttaaattggat taaaatgtag 660
 catatacttg gtttgtagtc ttgattcgta aattaactta ccagtcaaag ctgctgctgc 720
 tttttttttt ttttggacag aattctggct tctgtcaacc caagcttgaa gttcaatggg 780
 tgtgaatctc aagttcactt gcaaacttcc ggccctcccg ggggttcaaa gttgaattcc 840
 tccttgcccc ttagcctttt ccggagtaag ctggggaatt tacaggggaa aggaacccaa 900
 ccaaggccca agataaattt gtttttttgg gatTTTTTTTg gttagaggaa tgggggggttt 960
 caaccaatat tg 972

<210> 3
 <211> 566
 <212> DNA
 <213> Homo sapien

<400> 3
 gtctctgttt tctgtaaagt ctgctaaacc gcagcagtat cttccaagtt ctctgacaat 60
 gttataaatt tcattaagaa gatgaagcct tctggggcaa ttaactccaa aagtgttaca 120
 aagccatagc aaggagatac tggcctgtag tacttagcac tcagctctga aatggagata 180
 agccctatca acccgggggt gtgtttaatg ttcaactgcag cttccatacc accgctacaa 240
 gctagcggtc accgtgttaa cactccccaa tatctattac aattagtttt tgattaagag 300
 gcctaacatg gcttttgaag ctttttatgt tgtgtcaaat ggctagtgat ggagtgagat 360
 gcacagacac ctcccggtaa caaagcatat tgtaaaagg ttcgagtga ctattttatc 420
 atcctaattg aaaccagatg actcgaactc ccccttaatt ttcaactatt tactgaagag 480

tccttgtttg gaattactga aagattttatt cctatatgac taactcatgc attaagaatc 540
gctaataaca gagccatttg gaattg 566

<210> 4
<211> 552
<212> DNA
<213> Homo sapien

<400> 4
ccaccaccac cgccaatgtc tgattaaatc taagccacgg ttccattttg gggcaagggg 60
caggggtgtgg gtgtggtgga cagagaacgg gctccagcca ccatcagctg cacgtcctgg 120
tagcacctga gccattccca gtctcctgtg ctgggagatc atcatctaata cctgatgatg 180
acgccctcct ctgagagctc gggggatgag ctgccaaaac ccatggaggt tcaaaacccg 240
ggctctggag acagaagaca ctgggctcag agacatgggg ccactgtcca ggctttccct 300
gggcctggac ctgggctcct tccactgtct gagcagggtc agcctgggct ggctgtgctg 360
gttcaagaga tgagggagga ggacacagcg gttgccagga ccctgaggca ctggactggg 420
cactgccccca aggggcagag gatggaaccc ggtgttgata ttgtgatgct gacccgtctg 480
ctcacctgat ttgggtgctg aagggcagct gaccccccta atcttttata gacttgtgaa 540
acgaaaataa at 552

<210> 5
<211> 589
<212> DNA
<213> Homo sapien

<400> 5
gcggcgctga tgggtgttcaa cgtcaagtag cgcccgcgca ggacctggca ccgcccctcc 60
cgcgcggggcc cgcacccagt gtgcgcgccc cggggggccc cggggggccct cgcgcgtagc 120
tcccagcccc agggccggaga gggggccagg ggccttggag ctgttccgtg gttgggtgctg 180
ggtcgtgtct tgcaagggac cctgtgtctg cctccaagcc ggggaggagg ggggggtccag 240
acccccacgg ggaggtgccg gggaggaggg ctggggggaa gcagtgcctg tgtagacgtg 300
gctgtacata gaaagctcta ttgtgggaaa gcaggagggg tgaccagacc tcagctgcta 360
gtgtgagggc tgggaaggga gggggcgctt tttgtggacc cctgggaggg gagggggtgc 420
tgcttctga ctagcccatc tctgggctac gatgtcacag ggctggtggc tgggacagct 480
ctggggacta gggggtaatg tgatacagcc gatgtgggca gcagcaaata taaaactggt 540
gaggctcaaa aaaatctatg tcgggtgctg agaaagaggt aatgaaatg 589

<210> 6

4

<211> 504
 <212> DNA
 <213> Homo sapien

<400> 6
 gtgcaggatg accgtgaagt ctctgacctg ggtaggtggt ggtgccaggc cttgggaggg 60
 ggagcccagg ccaggagtgg cgaagaacag atttgaggaa aagtaatgta ttttgatcgt 120
 tttgatttgg acacaatcgc aacaacagtg ctctccattt ttggagaggc agcttgccca 180
 gaggtgagtg gtgccgactg cccggctcgc cagctgtgtg accttgagca ggtcacatgc 240
 ctctgagata atgataatca gagccactcc atccatagag ttggtgtgag gactacatga 300
 gttcctgcac gaatcgaagg ccacgtctgg ctggcggctg cagtgtccgg cagcatcggt 360
 ctgtggtggg tgtcatgtgg cttgagggtt agttggggga gccttggggac agggccagca 420
 gacaggcggg gggcagggtcc agcagacaga cggagggcag gtccagcaga caggcggagg 480
 gcagggtccag cagacaggcg gagg 504

<210> 7
 <211> 1165
 <212> DNA
 <213> Homo sapien

<400> 7
 aagggttaac cgaagctctg gtttgtatth gagaaacatc agtctggccc aacatgagga 60
 ctaaactgga ggggtgagact ggagaaaaag gctgggaaaa ggtggttggt gacaagggga 120
 aaagtggga cagcaggcag tcggcatgga ctggagaagg tgggagagaa ccaggaaaca 180
 gctgcaagag ctttgacaca gatggaatth tgagggaagg gagcagttcc acgtgtgctg 240
 aatggaagat aagtaggaat ggctattgcc tggtgatact gatgggtggg gagcagctga 300
 cacgtggttg gagtagggaa gagggtgtg ggtggtgcct aatgagcgag atcccagaag 360
 tcaggctgaa gagctggaat gcagactcta ggtggtgtgg gtgcatggga ggaaagccgg 420
 ttctccaggg tgtctggcgg gcaagtgcgg ggcggtggcg tgcagccctg gacctaactc 480
 atgcgtgtac tgcccatgat cttacaacgc ccatcagcgg ccagggtgcc tgcccagcat 540
 cttgcaaggt cagcagcctc agccaacggc ctctgtgttc catccacccc atcgggcctt 600
 gatgaagggg aggctggcga ggcgggggct caccctctc tgagagctcc tctaggcctt 660
 tcttgctgca ccttggttgt ctctgacaa gctcctggct gaggagtcaa atgcaggaga 720
 cggttggtgaa gaccagttc aggggtgggct gccaccacta atttacctaa tttatggaga 780
 gaaggaaaag gaagtcctgt atctcctcca gggaaggggt cctgaatgca ctaaataact 840
 caacaatctc taaccaaata cccctttttc atcactttct ttcattatga ttgccccgcc 900
 ccaaccatt cttctctggt tacctctatc taccctcgag tcaacaagcc ctgcctgatt 960

```

acgcagcagc agtttctcct ggagagtata tgcccttccc taccagagtg gctgtgctct 1020
gtggaccaac ggcatTTgtg ccgtgggtgg tgtttccacc attccagtgg gttggctgca 1080
gagttatcct ttgtgggtgg gagagagcac caggcctcag gaatctccct gctgggccca 1140
gcctccatct cctccccccc cctct 1165

```

```

<210> 8
<211> 202
<212> DNA
<213> Homo sapien

```

```

<400> 8
tagactctaa tagtaataga atctaaaatt attctaaaat aaaatattta ctttaagaaa 60
aggtaataga ggaaatgtaa accaaaccgt caaagttttt tctactcaga ggctttgctt 120
ctgtttcatg cagggccatc acattcttag tcatagtga aattagcctc tttctcccca 180
tcaacttgta ggatctatga gg 202

```

```

<210> 9
<211> 1005
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (176)..(176)
<223> n=a, c, g or t

```

```

<400> 9
attgaattta ccttatagag ctgttttgag gactgtatgg taaatattga atctactact 60
tactgttggt atttccttct agtctttttt tctaagcata tttttttatt tgtagttaa 120
aacatactgt atatttggtt cttttttttt atttaacatc acattataaa catttncttg 180
tgcaatttcc aggtgatgaa cagcctcctt ctaaatcatt gaccacattt cttattcttt 240
cctgaggata aatttcaaga ttcctgaac caactgatga cagcctattt agctacagga 300
atactctgga aggaaatggt agcttctgga gttttaactg gagtgttaat cttgccacca 360
tgacctaatg gaatatcaca taacctctg gtgtgaccat taagcttagc aggatctcag 420
agaatagaca gaggaatatg caagctgggt atcagagaag gctctggagc ctgtcagtca 480
gtttgaacct tggttctgca actttggcaa tatatatatt ttttatcaact ctgagccctt 540
tctgtacaat aggataaaac taatacctat ttcattgtgg tgtttggaag attaaaatag 600
ttgatatgta atcctggcag gtggtaacta cttaatgaat gttagctcct attgttcatt 660
atattagata cctaaagttt gttttatgat taaatgagag agaattggcag tgatgttcgc 720

```

6

atgtaatgtg agtacaagct gtttaacatat aaaatagggc aatacaaata ttttttactc	780
cgaaatggat tgatttttctt cccattttaa ataaataaca gaaatattct gttttatgtt	840
ttgtggccag aagctagcat gggaagaggc atttgattta ataatcctta tagagaattt	900
ttaacaattt tcttcagtaa ttcaaagggt gaaagtgatt tttgtccaga ttattcttgt	960
tgaaagacaa aaatgcttta tgttagtttt ttaaaaatta aattt	1005

<210> 10
 <211> 296
 <212> DNA
 <213> Homo sapien

<400> 10	
ggcagagtag gggcattcat ttatgtcatt gcaggtttgg tgagatccag agctggcagc	60
cgtaggggat tccttatctt ggcagcctcc tattgctgca gcagcactgt ggttctagag	120
gaggtcactt tctggttgca gaggaggcaa cagctctctt ggtggcccag ttctgcactg	180
tgcttcggga gtctttccca acaactcagc ctagggttta tttctttcca agtgattttg	240
taaaccattt aatatcctga aaatatcact aactacctag taagaaaaga ttctgt	296

<210> 11
 <211> 749
 <212> DNA
 <213> Homo sapien

<400> 11	
atgctaaacg ccttatttgc aattctgttg gatattctag agactggtat atttttatct	60
gacatggata attttgaggg gacgttatat tttggggaag tttaaaagt gattctttct	120
gcctttctaa cttgggcgaa atatatcgac tcataccacc ctttcaaaaa agatagcttt	180
ctctactagg tgtttgggtt cgaagacagt ttaaaaatta cctcgtgttc agtacctaag	240
attgctgttt tgccctgagg gttttccaag tagaagtggg tagttttcat gtactttgag	300
aattctgagg tttttaaccc cccctttttt tttaaataa aatcttttct attttgaagt	360
aatgagactt taaaggatct tgtttttgta atcacttgat ttcaaaggat acttcttaca	420
gcaattataa agtaatagtc aaaatgctaa ttagacattt tcagtcatag ctgctttaat	480
gttgaaacag aatttttttc ttttagcagag actgtatcgt agagttgaaa cactaaatat	540
tcttttgtct ggcatcagat ctctgaactg ctctgacatt tctcctgaaa aacccttagt	600
gcccttttta attgccccca aaagaaaagt tattgagatg gtagatttca agcaaaaaac	660
caggccctct ttctggcaaa ataggcttac agtggcaatg gaattggaat gtttaccctg	720
tagaatggca ttggccttta aatggaaac	749

<210> 12
<211> 660
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (398)..(471)
<223> n=a, c, g or t

<400> 12
gtgaacagga caggaatctg caaaatgaga taaattgtac attggaagaa agacttaaatt 60
agggtggcacc tgtgttcctg ccctattga gggtagatga ttttctgagt gttatgcttt 120
tcccccatct ccacctgttg caatgctagt catttctttc aatattcagc tcaaatgcca 180
cctccacttt gaagtcttac aggatctttc tagttgccat caattgcca tcttatgttc 240
atagtcagcc atgtttatga atatctgttc ctgtgcctgt ttccctcacc agggttcgct 300
ccttgagggt tgggactctg tcatgcacct tgttgtaggg aggcaatgta gcactatggc 360
tgtgcatctg gattcaactc cttgtgccac catgaacnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn natgaatgat 480
atgtaataat tatgttcctg gcacatccta agaactccat aaatgttagc tattgctata 540
ctcattactt gcaactccct ctacaatcca tgggtcttgg ttggctctcc ataaaaattt 600
gttaacttca atttcacgac tccactggca cttctgatcc tgaaacagat gttgcagtga 660

<210> 13
<211> 1803
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (26)..(26)
<223> n=a, c, g or t

<220>
<221> misc_feature
<222> (69)..(69)
<223> n=a, c, g or t

<220>
<221> misc_feature
<222> (82)..(82)
<223> n=a, c, g or t

<400> 13
aggataaggt ggctaggcaa acaganaaaag agctcaggat ctaagagaga catcatgggg 60

tagagatana acactcctaa tntctatggc agaagataga aaaccttaag aactctttta	120
caaagaatgt ccaaggggaag tacagaacag aatataaggc agatgcctgt gaaactgtac	180
cgtgggaaag gggctagtag tcaagccaag aaaggaaaag acagcttgga atgctattga	240
caaatactcc actaatgaaa gaagagaaaa ccgatattca caatccctct gtagccacag	300
tgagagcaga tgttcatgaa ggttcctctc tgctgggtggc ttggccaact ctaatttagt	360
ccttaacttc aatcatggac accctccaac tcagtcacca cctgggcca aagtcagtat	420
tttatttctc attaagcatg ggaggaaagg agataactgc tcattaatac attggaatga	480
aatcttgtaa aaggttaagt tgacctgaag tagaaagcca gccatttggg tgtcacaaaa	540
taagcttaat tggatcagcc agcgtgagtt tttttaatcc atgcattctc ctgattaaat	600
actttcagtg aacgcttgct gctctatggc tggaaggaca catccctgct tgtcttgcaa	660
gcagctgcag gaactggctt ccctgtccct cttgccacat atttgtcctc cccgtcaagc	720
ttccacatca agcttcatcc ccacacaggg ctttggtctc gctcttgctc tgccccagat	780
tccttgctct tgcctttttt ttccatgact agcatcttct cctctttggt gaaatggccc	840
tcagaagaca ggctctttct aattacccta cctactgcca cttccccagt tattcttacc	900
ccagcctcct gaatttccat aacagcacca cttttgatct gtctttatto tgcattgtca	960
tttctctact tgtttattgc ctgattcatt cagtagaatg aaaactccag ataacgaagc	1020
actagttatt ccaccctccg ctgtaccac aggcaacaca gcacttgta cataggaggt	1080
actcaaaga tgttgatcca tgagcgaagt taaatagtaa ttcaatcaat atgcttcatt	1140
ttgacagtga tttaaaaaaa atggatataa actggtggcc aggatggaca ggcaccacat	1200
ttatgggata ttggattttt cagctgagat gggatagtag gctgtggaca cacttatttt	1260
cctgtatttg ctctctgaca acctttctcc aatgcaaaat gattacagat tattaaacct	1320
aatgagggtt agagtttctc tagtttttct aggttcgtga agagaatgat gtcatggaaa	1380
tactcaaaga agcagtgcag aaggtaggtt ctgagaggga actttcatct gcagggtgcta	1440
ctgaaaggga gcggtggagg ggatgccctg gagttcctgg ggtactgccc tcaggagaac	1500
gcgctgtggc ccaacctgac agtgaggcag cacctggagg tgtacgccgc cgtgaaaggg	1560
ctgaggaaag gggatgctga ggttgccatc acacgggttag tggatgcgct caagctgcag	1620
gaccagctga agtctccctg gaagaccttg tcagagggaa taaagagaaa gctgtgcttt	1680
gtcctgagca tactggggaa cccgtcagtg gtgcttctgg atgagccgtc gaccgggatg	1740
gaccccgagg ggcagcagca aatgtggcag gccatccggg ccaccttag aaacacggaa	1800
agc	1803

<210> 14
 <211> 2032
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (875)..(897)
 <223> n=a, c, g or t

<220>
 <221> misc_feature
 <222> (1155)..(1267)
 <223> n=a, c, g or t

<220>
 <221> misc_feature
 <222> (1628)..(1705)
 <223> n=a, c, g or t

<400> 14
 atttttcggag acacaattca acttggttaca ccctgtaaa tccgaaaaag aattccctga 60
 aaggaaccta aagtcactta gagtttattt aataaactca aaatttttag acgttagatc 120
 agtcctatac ataaatagct gatctcataa gtggttttta ggatgctggc gggagggtga 180
 aattgtattg gtggagtctc ctgccaaaga attcaaagtt ggaagatgaa gaattcaggc 240
 caaagtatga taaatatcaa tgaaaataag agagatttca cagataaatg agaggaagta 300
 ttacagaatg ggtaaatatg caagccctta atcaagtctg ggtttcaatt ctagctccgt 360
 cactgtattt gcagatgac ttggacaagt tacttaaaact ttctaagcat agatttcatt 420
 atttgtaaaa tatagtctta gagtagctag ccaagacctt tagatttcca agtataaatt 480
 aaaagactgg tcaaatatgt ctcccaaact acagttttca catagcgagc tttccaagtg 540
 gtactgttga agccctttcc ctaggacctt ggtcagaggg agcacacccc acccatcctt 600
 cttgagggta aatgaaactc ataataacc aaaacctctc tccaaaatgc tcttctaattg 660
 tataactcct ctcagtctct ccaatcccct ttattggcta gagatgggta gttagcaaag 720
 gtcacaaaac cagttatcta gtctcacacc tctctgggg gacctccacc aaaattgaga 780
 ctaacctagt tccattttat cagcctgcta tactagcgag actgcagtgg gatgggcatt 840
 tcatatggca ataataggag ggcaaagtaa ttggnnnnnn nnnnnnnnnn nnnnnnnatt 900
 gaaaatactc atgctctttg atctgtcagt tccactttta ggaatttatc ctaaggaaat 960
 aagttgaaat gtgaacaaat atttatttat aagtatgtta actgtgaggt tatttagcagc 1020
 catggaaaat tggagacatt ccaaatttct aacaatgggg gaatgtgttt aatgacatat 1080
 tgtacaacca tttaataataa tgtataggat taagatggta atttttatgt tatgtctttt 1140

10

```

tcactacaat ttttntnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 1200
nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 1260
nnnnnnnnata ttacaatata ataaaaagca taggatacaa aagtatatgt acactatgat 1320
ttcaactagg ttaaaatatg catagaaaat ggtctagatg aaatatgaaa aaattatcac 1380
tgcttgccta caactgggtg ttgtttttgt ttttatactt ttctgtactt gaaaaattta 1440
caacaaacaa ataaggaaat ggggcttagg agacaaaaaa taggtctctc cagtgaaca 1500
atagagaata aataaagggt ataaagctgg gctttgtaaa ccggcatgct gggattaaat 1560
cctggctcaa tttctctgta gaagagaaaa gagagggcac aaatcctctg tccatagatg 1620
aattcctnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn 1680
nnnnnnnnnn nnnnnnnnnnn nnnnnnaatcg gtggtacaat gtgtgtaact gacttatctc 1740
agggtcatgg ctattcactc tagaggtggc agacacagtg gggaagagaa gcatctcgga 1800
cctccacctg aattttgctt cttccacctt cttctagtga ctttgggtgg ttatgtactc 1860
cttctctgct ttattactgt ggagacgtaa aatgagagta atggaaatac ctgtctcata 1920
ctattgttac gaggattaag tgagttaata catataatgt ctatagaaca gttcctggca 1980
ttaagcaaaa gctcaataaa tgtagaggtt attattatgt cattatcatc aa 2032

```

```

<210> 15
<211> 1645
<212> DNA
<213> Homo sapien

```

```

<400> 15
aaggaattcg gcacgagggg aaatgcaagt ttctaccctt tgaccgacat catgtcatgt 60
ttttctcccc tcctagcccc cagcaaccac ggttctaatt ctactttgtt tccatgagtt 120
ttgtggcagg agaatagggg attagaggaa caggggttca gcagggtgcag ccagttocta 180
caggcaggag agcagcaggt gtggcttaca ggtcacgtct tccctcgtga taacaagcca 240
catcagcctc tgattgactg cgggccgggt ctccactacg gccactgatt ggtcctgagc 300
caatccttca tagcatctaa ccaattggaa gcctctgaag ggcgtctagg ggtgttgccg 360
ggttccttct agcttattac aaacctaata tggggaggct cttgagctgc tggctcgatc 420
ccgctoccac tctgaattgt cttcaataaa tctgtgcttt tctattcaca atagcaaaga 480
cctggaatca acccaaatgc ccatcaatga tagactggat aaagaaaatg tggtaatat 540
acaccatgga atactatgca ccataaaaa agaatgcgat catgtccata aaaaagaatg 600
cagggacatg gatggagctg caagcatta tcctcagcaa actaacacag gaaaagaaaa 660
ccaaacacca cgtgttctct cttttgggca cagaacttct tctaakatat accaaaagaa 720

```

11

```

aaaaaagggg gtcgggggag attcaaggtg gccaaatagg aagagcacca gtctgcagct      780
cccggtgaga tcaacgcaga aggatcacaa cacctcacca gcaagggaac aaaactggac      840
agagaatgag ttgacaaaac tgacagaagt aggcttcaga aggtgggtaa taacaaactc      900
ctctgagcta aaggagcatg ttccaaccca atgcaagaaa tctaagaacc ttgaaaaaag      960
tgatggggag aatggaacca agttggaaaa cactcttcag ggtattatcc aggagaactt     1020
ccccaaccta gcaagacagg ccaacattca aattcaagaa atacagagaa caccacaaag     1080
atactcctca agaagagcaa ccctaagaca caaaatcatc agattgacca aggttgaaat     1140
gaaggaaaaa agattaaggg cagccagaga gaaatgtcgg gttaccaca gaggggaagcc     1200
catcagagta atagcagatc tctctgcaga aatcctacaa gccagaagag agtggggggc     1260
aatattcaac gttctgaaag aaaaggatth tcaaccata atttcatatc cagccaaact     1320
aagtctgacc cagcacagtc ccagtgggtg tggtcacaga ggtgcttggt tcaccataac     1380
cccagcttca ggcagctcag aacagagtga aagactcatt tgtctgagag aaagtaaggg     1440
aagtaacaa gtgtctcttc ctggtaatcc agagaattgt actggatctt atccaagacc     1500
accaaggcag tatctctaca agtctgcaag aaatggtgtt actgagcttg ggggtgtccc     1560
taatgtagat atgactgaga tcacaacacc caagtccatt caaatatcta gaaagttagg     1620
atgggtacaa acaagcccag agtaa                                           1645

```

```

<210> 16
<211> 2073
<212> DNA
<213> Homo sapien

```

```

<400> 16
acaatggtgt tcgcattttg gaaggtcttt ctgatacctaa gctgccttgc aggtcaggtt      60
agtggtgtgc aagtgaccat ccagacggtt ttcgtgaacg tgactgttgg atctaattgc     120
actctcatct gcatctacac caccactgtg gcctcccag aacagctttc catccagtgg     180
tctttcttcc ataagaagga gatggagcca atttctcaca gctcgtgcct cagtactgag     240
ggtatggagg aaaaggcagt cagtcagtgt ctaaaaatga cgcacgcaag agacgctcgg     300
ggaagatgta gctggacctc tgagatttac ttttctcaag gtggacaagc tgtagccatc     360
gggcaattta aagatcgaat tacagggtcc aacgatccag gtaatgcac tatcactatc     420
tcgcatatgc agccagcaga cagtggaatt tacatctgcg atgttaacaa cccccagac     480
tttctcggcc aaaaccaagg catcctcaac gtcagtgtgt tagtgaaacc ttctaagccc     540
ctttgtagcg ttcaaggaag accagaaact ggccacacta tttcccttcc ctgtctctct     600
gcgcttggaa caccttcccc tgtgtactac tggcataaac ttgaggggaag agacatcgtg     660

```

```

ccagtgaaag aaaacttcaa cccaaccacc gggattttgg tcattggaaa tctgacaaat 720
tgttggacaa gggtattacc agtgtactgc catcaacaga cttggcaata gttcctgcga 780
aatcgatctc acttcttcac atccagaagt tggaatcatt gttggggcct tgattggtag 840
cctggtaggt gccgccatca tcatctctgt tgtgtgcttc gcaaggaata aggcaaaagc 900
aaaggcaaaa gaaagaaatt ctaagaccat cgcggaactt gagccaatga caaagataaa 960
cccaagggga gaaggcgaag caatgccaaag agaagacgct acccaactag aagtaactct 1020
accatcttcc attcatgaga ctggccctga taccatccaa gaaccagact atgagccaaa 1080
gcctactcag gagcctgccc cagagcctgc ccaggatca gagcctatgg cagtgcctga 1140
ccttgacatc gagctggagc tggagccaga aacgcagtcg gaattggagc cagagccaga 1200
gccagagcca gagtcagagc ctggggttgt agttgagccc ttaagtgaag atgaaaaggg 1260
agtggttaag gcataggctg gtggcctaag tacagcatta atcattaagg aaccattac 1320
tgccatttgg aattcaaata acctaacca cctccacctc ctcttccat tttgaccaac 1380
cttcttctaa caaggtgctc attcctacta tgaatccaga ataaacacgc caagataaca 1440
gctaaatcag caagggttcc tgtattacca atatagaata ctaacaattt tactaacacg 1500
taagcataac aaatgacagg gcaagtgatt tctaacttag ttgagttttg caacagtacc 1560
tgtgttgtaa tttcagaaaa tattatttct ctctttttaa ctactctttt tttttatttt 1620
agacggagtc ttgctccgtc gcgcaggctg tgatcgtagt ggtgcgatct cggctcactg 1680
cagcctccgc tccctgggtt caggagaatc gcttgaacct aggaggtgga ggttgcagtg 1740
agccaagatc tcgccactgc actccagcct gggcgacaga gcgagactct gtctgagggg 1800
gaaaaaaaaa aattttgaac cccccaaaat atcacatatc tagcagttct aggttccagg 1860
gtggatatatt ttaacatatg tttgcatttg gttctgggat agtaaccgtt atggataatt 1920
gtttgtttct gtcaagtcgg tgagatttta attgtttgaa ttatttctat ttaaaattac 1980
tctcaattta actgtttaa aaactaaaaa tttaaaacat aatggcacia catacgacac 2040
agaataaaac acatgttttc aaggaaaaaa aaa 2073

```

```

<210> 17
<211> 175
<212> DNA
<213> Homo sapien

```

```

<400> 17
ctcgggtctca aaaaacaaaa caaacaaaa caaaattaga gattgggtct tcccaggca 60
tatggtattc tataaacag acttactctc cttggaggat atattttgga gaatgcttca 120
taaaatctat gaatactgta caatgctgat aataaaaact ttttatacct gtgtc 175

```

<210> 18
 <211> 994
 <212> DNA
 <213> Homo sapien

<400> 18
 ggagaggtct gtagccactg agggccccgg tcggggccgc tttgcaggtc cctagtcagg 60
 accgagcagg ggagtaggat aggaatcccc gccgcacctt tgtacgagcc tgaccccttc 120
 cgtgggtttg ttcctgggtc gccgtcaagc tgcggtctct cctccccgc ccttcagccc 180
 cgcggtctcc aggggcggcg ccctgggtct ggaacgcggt tgccaccgag gaggcggcgg 240
 ccctgcgtct ggaacgccgt tgccaccgag gaggcggcgg ccccgagcgc gcctggaagc 300
 cccgggcaac cggccagggt cgggcacagg tgggtccgt caggccgccc ggggctcctc 360
 tgtcccagct ctgcggccca gggggtgacg tgatggcggc agcggtgctg acggaccggg 420
 cccaggtgtc tgtgaccttt gatgatgtgg ctgtgacttt caccaaggag gagtgggggc 480
 agctggacct agctcagcgg accctgtacc aggaggtgat gctggaaaac tgtgggctcc 540
 tgggtgtctc ggcaccttta cagacacacc caggatcaat acgttgcatc cttcaatcca 600
 atcaagttga cacattaata aaaatcccag agcacactga gcattgaata ctttctcttt 660
 aaatttaaca actgttactg gtgggaagta ttattgtatt gatccagaaa taaaggctca 720
 gagaaatgaa atgatgtttt tgaggtaaca gctgtcaagt gtccgtttca agtggttgga 780
 gctagggact tctggtgact gtgttaacac aggatgggaa gccttcaaca tgggccaggc 840
 acctagctgg acctagctgt tgaacgcata actaactcca tcccttgtca acttgaacct 900
 atgacacgat ctcttgaga tcgatacata atgctgtcaa atatatagac taatgaagg 960
 acgatgaatt atgcactaac ataatacaag tatic 994

<210> 19
 <211> 872
 <212> DNA
 <213> Homo sapien

<400> 19
 atttcatggg ataatatatc aaggcctcag gaggttaaca cgtggtctga gatggcacag 60
 ctagttagtg gaagacccgg ggctacctgt cagttcttct gctccccgc agagcttctc 120
 ccatcagctc agaaactgcc atcctcccta gccctttgga ctcatctaact gaccctgtac 180
 agagttaacc ttgatcagga atgagggtctg cccctctgat cttcagtttg tattgtgtag 240
 agcatcttgc cgggcacttt gctggtggtt tggcaacgct ctcattacat tctggttgt 300
 gtttatgtca gaggcacctg tcaccatgtc accttggcgc cttgatcaca tatttacctg 360

14

```

tcagatccac aagctcaggg gctgacagtc tgccattttt acagaacaca ccattcttcg      420
tggtgagtct ctgatcctga cccaatctt  tcctttgttc tttgtgtttc tacacctgga      480
agtgtaata  tctgcagcaa gtcccacttc ttgtgcccc  tgcctctctc ccctagaaa      540
gaaagatgga taagaccag  gagaaagatc tctggctata actagtgtgg gaagggacag      600
ggagcctcag ccctgggaag ctggacaagg cagatctgtg tggagtgggtg gtggctgcag      660
ctgctgtctg cgtggagtct tagagttggc agcagctgtg gccagcagtg gctaggagaa      720
gggctgggg  tgcaggcagg ggggtggttg ccaagcttcc cacatgctct agctgggagg      780
gtactcagga ggtgtctctg agtagccttc ctgattcctt cttccaagca agaaacgaaa      840
ccaaggggag cttattctga tcatttcact ca                                     872

```

```

<210> 20
<211> 1258
<212> DNA
<213> Homo sapien

```

```

<400> 20
tatgcaaadc aaaaccaccg tgagatatca actctcacct gtcaggatgg ctatcattga      60
agaacaaaaa cacaacaaat attgrtgaag aagtggacaa attggaaccc ttgcatactg      120
ttgatgggaa tgcaaaatgg tgcagccact aaggaaaaca gtatgggtgt tcctaaaaaa      180
tgtaaaaata gaactgccat gtgacctagc aacctatctt ctggatattt atccaaaaaa      240
attgaaatca ggatcccaaa gaggtattag catttgcagg ttcattgccg cattattcca      300
caatatccca agatacacia acaatctaag cgtccatcgg cggagaaatg aataaagaaa      360
atgtatatac acatacaatg gaatactatt tagcttttaa aaagaaggaa attctgcaat      420
tattcaacia catggataaa tcttaaggac attacgctaa gtgaaataaa ctattaacag      480
aaagactaat atggcaggat tcaatttatg tgggatattt aaaatagtcc ccttcataga      540
accaaagagt gaaatgatgg tggccagggg ctgaggggaa gtggggagtt actaatcaat      600
aggcataaag tttcagtga  gtcaggtgaa taagctctag agatttgtgg tacaatattg      660
cacctatagt caacaatact gcagtgtaca ctgaacaatt tgttaagata gatctcacct      720
tgtgttctta ctgaaaatgt aaataagaaa gaaaatagaa cagaaaagca attggatttt      780
aattctggaa actcctttct cttccttaca tccaggaaat ttgctgttta ttttgaaaag      840
caaatttaaa cctatttaag ggagagagag ctcttgtaaa aattcattta ttagttctgg      900
accaatgtta tttataagct attatttcaa atgataaaaa ataaatgcat aatacatattg      960
atgatagaac atttttcttt ttaaataagg cctgattttt caaaaagcat tgaatatggt      1020
tgtttcta  ttccaaagct tgacagtttc ttatgtttca ttaaaagcag aagaattggt      1080

```

15

atgtatgcat gttcatatca cttttgaatg actcatgtgg ctacaaatgc agcaatctgt 1140
 gggtatcatg gaccatgaac tagataaaca ttcttgcttt tcttaacatt cctttcctgt 1200
 gcctcattaa ctaatgtaca taataaaaca tctttatttc atattctttg caaaaaaa 1258

<210> 21
 <211> 3841
 <212> DNA
 <213> Homo sapien

<400> 21
 cctgcctcag cctcccaagc agctgggtct acagtaaaca ttgctatctt cagaaattaa 60
 ggaatgagag gcaaggaaga aaggacattc atcttggaag gacccaaagc accttgaaat 120
 ttggaaggaa tgttttgcta gctgctaaaa ccacctttgt caccagaatg aatgaagtgc 180
 ttcagattta aataattgat cattcttaga agagctgctg taacacatag gggaagtga 240
 aactaacaga tgaaagtttt gtcatgcact gaaagaaaag cattgtctgt gaagttctat 300
 ttttaaaaat gtctgcttgt aacaccttta ctgaacatgt ttggaaacct ggtgaatgca 360
 agaattgctt taaacctaaa agtttgacc agcttcccc agaccctgag aaggcacc 420
 tcacccatgg caatgtgaaa actaatgcca atcacagtaa caaccaccgc atcaggaaca 480
 cgggcaattt ccggcctcct gtggctaaaa aaccactat agctgtgaag ccactatga 540
 tagtggcaga tgggcaaagt atatgtggtg agcttagcat ccaagaacac tgtgagaaca 600
 aacctgtcat cataggggtg aaccgaaaca gagctgcctt ggtcagaaa ccacttaaca 660
 ataataatga agatgatgaa ggaattagcc atgttcctaa gccttatggc aataatgata 720
 gtgcaaagaa gatgtcagat aacaataatg gactaactga agtggttaaag gagatagcag 780
 gcttgatgac tgcccctcag ataagaggaa atgaaacaaa ctccagagaa acattcttgg 840
 gaagaataaa tgattgctat aaacgatcat tggaaagaaa gcttccacca agttgcatga 900
 taggtgggat aaaggaaact cagggaagc atgttattct gagtgggagc acagaagtga 960
 ttagtaatga agggggccgg ttctgttacc cagagttttc cagtggcgag gagagtgaag 1020
 aggatgtact tttcagtaac atggaggagg agcacgagag ttgggatgag agtgatgaag 1080
 agctgttggc catggagatt cgcagagag ggcaacctcg ctttgccaac ttcagagcaa 1140
 acacattgtc tcctgttcga ttctttgtgg acaaaaaatg gaataccatc cccctgcgaa 1200
 acaagtctct gcagagaatc tgtgctgtgg actatgatga cagctatgat gaaatcctga 1260
 atggttatga ggaaaattct gtggtctctt atggacaagg aagcattcag agcatgggtg 1320
 catctgactc cacatcacca gattcttctt taacagaaga atcacgttct gagacagcca 1380
 gtagtttatc ccagaagatt tgtaatgggg gattatctcc tggtaaacca ggagattcta 1440

16

aggacatgaa ggaaattgag cccaattatg aaagtccctc tagtaataat caggataaag	1500
atcatcaca ggcttccaaa agctcaataa aagttccaga gaccacaaa gcagtccttg	1560
ctctccgatt agaagagaaa gatggcaaga ttgctgtaca aactgagaag gaagaaagta	1620
aagcctctac agatgttgct gggcaagcag taaccataaa ccttgcccc acagaagagc	1680
aagcaaaacc ttaccgagtt gtgaacctgg aacagccatt gtgcaagcca tatactgtcg	1740
tggatgtgtc agcagccatg gccagtgagc acctcgaggg ccctgttaac agccccaaga	1800
caaaaagctc atcctctact ccaaactctc cagttacatc atcttcattg acaccaggac	1860
aaataagtgc ccatttccaa aaatccagtg caattcgata ccaagaagta tggacttcta	1920
gcaccagtcc acgacaaaag atacctaaag tagaactaat tactagtgga actggacca	1980
atgttcctcc aaggaaaaac tgtcaciaat cagcacctac atcacccaca gctacaaaca	2040
tttcctccaa aaccatccct gttaagtca ctaatttgct tgaaattaaa tttaatagtt	2100
ataacaatgc tggatatgcca ccttttccaa ttatcattca tgacgagcca acttatgtc	2160
ggagttccaa aaatgctatc aaagtccca ttgttatcaa tccaaatgca tatgacaatc	2220
tagctatcta caaaagtttt ctgggaacaa gtggagaact ctcagtgaag gaaaaacca	2280
caagtgtaat aagccatact tatgaagaaa tagaaacaga aagcaaagtg cctgataaca	2340
ccactagcaa aaccactgac tgtcttcaaa ctaaagggtt ttcaaacagc acagagcata	2400
aaaggggctc agtggctcag aaggttcaag agtttaacaa ctgtctcaac agaggctcagt	2460
cttcaccaca gagaagctat agttccagcc acagctcccc agcaaagatc cagagagcca	2520
ctcaagagcc tgtggccaaa atagaaggca ctcaggagtc tcagatgggtg ggcagcagca	2580
gcaccagaga gaaagcaagc acagtgcttt ctcagattgt ggcttcaatc caacccccac	2640
agtctcctcc agaaacacct caatctggcc cttaaagcttg cagtgtggaa gagctttatg	2700
ccattcctcc agatgctgat gttgctaaga gcacacctaa gagtacgcca gtccggccca	2760
aatctctctt tacatctcag cctagtgggtg aggtgaagc acctcagacc acagacagtc	2820
ctaccaccaa agtacagaaa gaccatcca taaagccagt cccccctct ccctccaaat	2880
tagtgactag ccccaaaagt gagccaccag ctccctttcc ccgcccgc tctacttctt	2940
ctccttacca tgcaggtaac cttttgcaga ggcatttcac caactggacc aagccaacca	3000
gccctaccag gtcaacagaa gctgaatcag ttttgcactc tgaaggcagc aggcgggcag	3060
ctgatgcaaa acctaagcgc tggatatcat ttaaaagctt cttccgcctg cggaaaacag	3120
atgaggagga tgacaaagag aaagagcgag agaaaggga actgggtgggc ctggatggca	3180
cagtcattca catgctgcct cctcctccag ttcagcgcca tcttggttc acagaggcga	3240
aaggagagtc cagtgagaaa ccagccattg tcttcattga caggtgcgac cctgctcaag	3300

gccagctcag tgtggatcag agcaaggcta ggacagacca ggcagcagtc atggagaagg 3360
 gtagagcaga gaatgcatta ctacaggact cagagaagaa gaggagtcac tcttctccat 3420
 cacagattcc taaaaagatt ctcaggtatg taaaatagac tggaagaata tgggaatttt 3480
 ttaacacatt ttcaggtgct aagtttttta tatcttttaa gattgcataa gaaccatttc 3540
 tgtttgtaag tagcactctg tggccatcat ctatctttat aactgcattc caagattaga 3600
 atagtattaa agtgatctta ctgcactgaa caaggtttat agccctggat ctttggctta 3660
 agtagcctac tgcattctgg tgaatgcaga ggaacttttag catagaaaag caaatccata 3720
 acaccttaat agcaataaat aatgtgttgt tttggaatgt gatggtagtg tgaggaagaa 3780
 acaggatcaa atgaataaat ggtggtactg ttctttccct cactgtaa at gaaataagct 3840
 c 3841

<210> 22
 <211> 1199
 <212> DNA
 <213> Homo sapien

<400> 22
 gaagtatgta actggtgaag aatacttata taggttctct agtaacattt tatcttttat 60
 tgcaaagatg ctctgattgt ggcggcaatg cttctttaat catagggagg ataattgtagt 120
 gggaaaagaa tgaggaatca aatgagtgat ttcacttaag ctctcagagc ctggagccct 180
 catataaggt agtgataatc cgtaacaagg ctgtcgtggg aaacagatga taggggtgcaa 240
 gtgttttgta gtactttaaa gtttgtaaaa ctgtcttaag gcacagtaac atatatgtaa 300
 gaatgtacca gtgtgactac ttgtgttatt atgctattat atataatgtg tatatttgca 360
 ataaaaagtt ataatttttt cctgtggaca gagatttagt ttgggcaagg gtggtaggca 420
 agtttttaaat tttagctgta aaatgctgat aatatgttca tgataggggt attgtaagta 480
 tcaaatgtgt taaaagtgt taaccactg ccaaatttc tgtcggaggt ggttattacc 540
 ataattgggtg gttatttaag atgagacttt gaactaaaaa gtaattttca aaataggcaa 600
 cttggaaggt gttgtttcag ttattaacat aagatgtgat cattcatgct tcttagctat 660
 cttcatttta ataaatgtgc attttagaca atcctgtggc aaagtagttc ttctggaatt 720
 gtttatttca attgcaggat ttcatatgag ttaatttcac agtttaaggt tatggcagtt 780
 tccaagtaca gttgtccctt ggtatccatg ggaaattagt tccagaacct ccctctgatg 840
 acagaattca aggatgttcc agtccttgga acatggtata gtatttgcat ataacctatg 900
 tgcacctcc cgtacacttg aatctctaca ctataatacc taatacaatg taaatcattg 960
 ttatactgta tgcttgggga ataattgacaa gaaaaaattg tacatgttca gtacagatgc 1020

18

agttaaataa tttttctatt cctggttggt tgaattcatg gatgtagaac ccatagatat 1080
 ggagggctga cttgactgta tagtgattg gcttttcaaa cagtaaactct gcaaaatcta 1140
 acttggtgtc gttatgcaca ggattgccaa aaaaaaaaaa aaaaaaaaaa ttggcggtc 1199

<210> 23
 <211> 57
 <212> DNA
 <213> Homo sapien

<400> 23
 agacccaag tagcaaaagc aatcttgagt aaagggaaga aaactagagg cctcaca 57

<210> 24
 <211> 532
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (420)..(463)
 <223> n=a, c, g or t

<400> 24
 ctgggattac aggcgtgagc actgttcccg gcctcaaagg aagtatcttt aagaagggtc 60
 tacttagaga tatgctggtg gttaatggaa taaatacaag atagcaagat acccagagac 120
 caatagcaga acgctactac cttctctagg actggaagaa tacagggaag aaatgctgta 180
 actggagtgg aggggtgctg gagtggtagt tttggagatg gagcaagatt taagcagggg 240
 ggagcacaaa aaaagacctc tctttactcc cattctctaa tatctagtgt ttaatatctc 300
 tcatggggca gtcttaccag aaaccagttg gtagtgagc ccatagaaac agcctatcaa 360
 aaatgggctg agaaaagggc aagtggatac cagcacagat caaatctgtg attttataan 420
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnaggaaag agctaaatct 480
 aagacgcgtg agtagagaaa aactttaaat ccaagtgagt ggactccaga ga 532

<210> 25
 <211> 352
 <212> DNA
 <213> Homo sapien

<400> 25
 gaatttccta agcctgacat tcaagactct tctaaattta gatgaaagtt accttcccaa 60
 cctcatgaat cattatgcta cottaagaca atttcaacca agatagggtt ttccttgcta 120
 cctccatgac ttccttcatt cagctgctct aaccatgcag tgacctctct aatcctcccc 180
 ttctcttaag gaaccgtgtg gtagaaatta tatagactta ggcccatgac agactagaat 240

tctatgttgg aaaagccact gcttccctct ggggctcagt tcctgactca gtaaagaggt 300
atcagttact ggcacagtaa tgccccataa caaacaacca caaaatttca gt 352

<210> 26
<211> 601
<212> DNA
<213> Homo sapien

<400> 26
caacaattca aaaatataaa caaggggaatt tgtggttgct ctcaacaatta atttggacca 60
gcagcccact ggagacattg ctgggaaccc caaggaatgt agggatttaa aaaaaaagtc 120
attgaaactt agaccacaat cactgggtcc caacaggctc tagtgттаат tcaaattgct 180
atctattggg tagagaaaga gaatttcagt gccagaagct accatagaga taccttgtag 240
ttccacctta tactgccgag aaggaaaccg aggcctggag ggagcccatg atgtgttcat 300
gtttacaccg caagttagtg ttaaataattc catttttttg aaggaatttt acttacatgc 360
atattttcca ttatgtttat aaatttcccc ctatttttaa ttatccccat tccctactaa 420
ttgtatcttt cattttttct ttataaacct tataattctc taatatacta tatatgcttt 480
atttatttta cttatcatct atcttccctc gtctccacta aaatccaaac taactccata 540
aaagcagggc tttttgtctg ggtcactcct gtatggttga cacttatttc ctggaacaca 600
g 601

<210> 27
<211> 565
<212> DNA
<213> Homo sapien

<400> 27
gagagagcag atatttcagg ggaagggatg gcagaagcaa ggctgagaag tcaggcgtga 60
tcattcatct gagctcctac tatgtgctgg ggctcagaac cagatgtgga tctacagctt 120
atttatgggt gcatttgagg gagtaagcgt ttatgagtag gacaatgagg gagagaagga 180
ttaacatttt ctgcccgttt cctgaaatgt gtcttaattc cttcctaagg aaggctgacc 240
aagaaaataa tgcaagaaga gaaagtctct tcaaaggagc ccagctaggg gctgtagatg 300
aaaggattcc tttgtgaata aaatgaagga aggaaactga agtggttcag tggttgaacc 360
actggctgaa tgaagacctg aaataagga aatgcttctc aaaattagag accaggctgg 420
gcttctataa gacgactcat gcggaggagg gcccttgag ctggagggtc cagtggaaac 480
ctagagacag aaatgagatg tagcagtcaa gcagaaattg aagaatgttc tttatattga 540
taccacccat gtgaagaggg tgaca 565

20

<210> 28
 <211> 183
 <212> DNA
 <213> Homo sapien

<400> 28
 tgtgttggtt aagagtctgc tgtaaaccac actctcacta tgacgtttct tttttctggc 60
 agaggcagat aggtgtcct actcctatct attccagtgt tgacttggaa atattattac 120
 agagttctta ctgtgcttgg tggaaatttg tttggttcta aggactatga aaatctggct 180
 tta 183

<210> 29
 <211> 679
 <212> DNA
 <213> Homo sapien

<400> 29
 gccaggcagt tctggcatct caacctcatt gaaagtcagc caatgctgag cccaagggtc 60
 atgttgactg ttggagctac tctgcctgca cagggttcaca cattcaattc agaatgtctg 120
 gttaatgtac ccatagcaat aagtatagac atcgtcagtg ttatattctt ttctccatg 180
 gtcccacacg tatatccaag attgctgcta acgtaaatta ccaaagtctt atcattcttt 240
 tagaggctat ttcttctat gttcgacttg taattggccc cagaacatga ttagatagga 300
 ctcaaattaa gggtttacag ataacaaggc gtgattgggg taggggttga ggaaaataaa 360
 cttctacttg taaggcagct gcccaggtt aggtatttaa atgggtctaac agggagtctg 420
 ctagaaccag gaaaacatag tatgcagagc ttcattctctg catgagaaaag gagagtaaaa 480
 aaagggtgact ttgaagctgg agacaatagt ttatgaactg gccaatgatt gaggccgtga 540
 atctttcttc cctagcttgt gggagccttt tgaaagcagt actgcacctt gctctagcat 600
 tttgaaaagt gacctcagtg taaagaaatt gtgatgtgct ctctctgtaa cagatgtatg 660
 tttttaaaac tgaatctgt 679

<210> 30
 <211> 402
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (92)..(132)
 <223> n=a, c, g or t

<220>
 <221> misc_feature
 <222> (184)..(220)

21

<223> n=a, c, g or t

<400> 30

```

atagagggtta acaaggactg gtagcagaag ggaagagtga ggacttgcca gggagagaag      60
agtaggcagt tatagtttaa taggtacaga annnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn      120
nnnnnnnnnn nngtaatggt tacacaaaat tttaaatgta cctaacacca ctgaattatg      180
cacnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn aaccacaata aaaataatta      240
actctaaaaa agcaattata taaaataata tgtatgtaat tatattattg ggcttatatg      300
taaagtgaat atatttgaca gtaacagcac aatggagatg ggtgggagga aacctgtatt      360
ggtgtaagga aatggcgcca gatggtaacc caaatctata gg                                402

```

<210> 31

<211> 1536

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (1078)..(1227)

<223> n=a, c, g or t

<400> 31

```

caccaagttc catagattct gcttcttaac gtctcttcct gtttttcttc attcccactg      60
ctctgaatat ctctgtccag taataataat aacaattaac atctctctgtg cttactgtgt      120
gccaggctca gcgctttcca tggattatth catttaatta ctgcagtatc ttcacattca      180
tctccatttt accatgctca acacttcagc ccatccctcc ctttctaaaa agggaatctg      240
atcatgctac tccccagctc aatgcatgat gggagcacgg tgggctgggg agcatacagc      300
tgggagtagg aaagcagtgg tcagggaag tcccagagga ggcgacagct gagagggggc      360
ttaaaggact aattagggca gatgtggaga aaagtgataa agagctttac cagtaagcgg      420
gaaccacatg tacaacgtgc ccagggtcag aaaagcctgg ctgggagggg gtttcagagg      480
tgagagagga gtagtctgcg aagacttcag catgcctgga gtgcaagatg tggctgggag      540
agaagagagc caggagaacc tggcaggagg acttgatgtg accctgacag gcctctgtga      600
gtccttgctg ggagagggtg ggccaggctg acatccctcc ccaaagggca gaggctctca      660
ctggctgaga gatctctgat cctgcctctt ttccatggga accttatctg tattctagaa      720
ctgcagccag agggtgaaac ccagcctgca gatgtgtttt gtttggcctg tacaatgttt      780
gaaaagcttc agagtcagtt gccataatt gccaaagatt gtaagatttc acagaaaatc      840
cgtattttctg gctgttctta acaaaaacaa caagaaaagg aaaggccctg gaaacactgg      900

```

22

```

gcacctgttt attcctgcac tgcgtggatc cctggaatat gttgcagctg ctccctttag      960
atgttacatt tcctagttta ccacagtctt cagcccatcc tgttacctga agtcaaggat    1020
caaatggcca tttgtcgtg tttgcactgt gggttccttt agagctgtgg atcccaannn    1080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    1140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn    1200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnact aaagcttaag aatcactgct ctggagtaaa    1260
gaaacatact ctgttcctgt gtctctataa aaaaaatgga aaactgtaaa attaggtgta    1320
cagggccatg ggcattgtaga gggccagttt ggcccatcac ctgagtccta tggacattgc    1380
agtccatgcc caggaaccac ctgaaacagc tctgtgttct ctcccaggt ggaggggtgg    1440
agggactcag atctgggcgg ctctgcaatc actaggctca catttccagt ctgatctaatt    1500
cagattttat gtggaccaca aacaggagggc actatt                                1536

```

<210> 32
 <211> 459
 <212> DNA
 <213> Homo sapien

```

<400> 32
catctctcta ctgacctttt agtggcttag cattttgttg caatgcacaa gtgaagttgg      60
gaaacagagc caagtgacaa gctgtgaagt aacagctcag ccttcacaaa gcaaattgtgc    120
tttgtgcact ttcttgccag agcagatgtg gccttggtgg actgtgggga gcatcatctc    180
tcaggtaaaa ataaattact aagattagga ccaaattgatg caaatccttt catttgcaaa    240
cagatcatga actggcagcc atgtggaaag aagactagac tgogaatcac ttaggatccc    300
gccctcgggt tagtctagaa tactctgcca atcagctttt tggccctagg tgagtgattt    360
taaaaaacat taaatgtgca aatgcataaa atgaactcac gtagttcatg gacacagggtg    420
acataggaag ttaatcaaca aaaaaaaaaa agtcgacgc                                459

```

<210> 33
 <211> 1857
 <212> DNA
 <213> Homo sapien

```

<400> 33
catctctcta ctgacctttt agtggcttag cattttgttg caatgcacaa gtgaagttgg      60
gaaacagagc caagtgacaa gctgtgaagt aacagctcag ccttcacaaa gcaaattgtgc    120
tttgtgcact ttcttgccag agcagatgtg gccttggtgg actgtgggga gcatcatctc    180
tcaggtagga aagagacaaa tggaagcaaa atagcaaaat ggtgaagagg agggttctag    240
aggcagagag ccctgcagtc ctgagtcctg agtccaccgt ttaacagagt gacttttctc    300

```

```

agcctgtgga ttataatgaa acttgccgtg tataggatcc agtttttttag gtggtagcaa      360
catgctcaag gaagatgcag ccccttcctc cagcctgaga gaataaagca tcattggctt      420
atgccaaacca tggcaatccc attccccctt tctcatgatt gatttgggaa tgggcaagtg      480
accagttctt aaccaatgag acgaaagaaa aagtctgttg gtcagcttat ctcagaagat      540
ttttgtttcc tgctataaag agaaacagta agttccttcc tgcctgactt aaatgcaatt      600
gtgtgcaaac atgttacttg gagctacagc aacgattctg tgaccatgag gcaataagcc      660
aaggacaaag gatagaatag tgaaaagaca gagtattgga ttttttttat gatatcattg      720
aactgctgga ccaatgtgga accatctact tttagtcttt catgatttgt gagatgatta      780
aacatttctt ttttttagcc attatttctt ttctgttact tgcaactgaa gacatcctac      840
ataatacaaa cactatccaa cttggattgt aatttatcta ttttcagatc tgtgtgttct      900
gtcttactga tttttgtatt cttagcactg agtataaggc tgctacttag taagacttca      960
ataaatgact gtttaaaaag gtcttagagg atttaaaaaa aattctgaag ttgctgaaat     1020
atgaagacca ggaaccaatt tcagggtota taaatttctc gcttccttca cttctggagt     1080
agtcttcatt tctgtctgat tctaccttta ctaaaatgaa tgaatgattg aggaatatga     1140
aaaatggctg gtatagtggc tactgctgag aagtatctct tggaaattcc cggaaattcc     1200
atcagtcaca ttcagacctc actctctgat gatcaatgtt tctgaactca aaccagtgtc     1260
cataagcagt ggactattta aagtgttatt agcctagaat atctaggaca catgttttat     1320
gcattcagta agcgtttgtt taatgtctgt tatctaccag gcacattatt atgcattggg     1380
catatagaga tgaccaacca cacggcctca atccttgaaa agcttacagt aaccaaatag     1440
ggttaacggc ctaaagagag gtaattacaa gtttacaccc agtcttctta gcatagtatg     1500
gtgctacctc tgagttaatt aagataacaa aattcctaga tatatatctt tgatgaaacg     1560
gtttccatga aaactttcca agtttgggtc ctccccggct gagttcattt tggaagggtg     1620
ctttctgctc tcatcaccta aggtccttaa cacgtatcct gtgttctccg gaggcacccc     1680
ttgccagag tgtgccctct gcatgggtcc ccgagaacac ctatttcaaa ggaaaagaag     1740
ttggatcatt cagtgaagat ttggggggcc acttttctac gtgggaggat tatgagtatt     1800
accttcctcc cctatgtatt agtaaataa aagaaaacat aaattaaccc cctttct      1857

```

```

<210> 34
<211> 1076
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature

```

24

<222> (816)..(844)
 <223> n=a, c, g or t

<400> 34
 tggagcggga gttgagcggg aagtcggggg gggggatatgt tgggaggcgt ggcgggatgg 60
 gagcgttggg aaaagggcag tgctggggga tggggacgag gatgaagcct cctcgaccca 120
 gctcacccat tctcctcaca gggcggtctg caaactcggc tgaagcaaag gttttgagct 180
 tcctgatggg cctgcctcc atttctagaa actgcttagg cgaacactcc aagtgatcag 240
 cctgtccgat tctcctccc tctggcggtt gctttctctg tggcttccag tttctggaga 300
 tcccatgcta agaagtttct ctccagcgaat cccctcaaac ctttgagatg gaaataattc 360
 ccatttcaca gatataaagc aggcctccaa actctttgaa gttcctttgc taatcatgac 420
 tagccaagcc tctggcggtc aaacgcacaa attccggcct ccagtgcctg ggttctcagc 480
 catggctaca atttctgtat tgaaacttcc acctccctc gctgttctct gtgacttttc 540
 ctcttggtg cattgggttt ggctggtaact atccttactc tttcaaggtc acgggtccct 600
 gagggtgata accagaacct tatgcttctt ttgtgttccc aggcggggac ttagccactg 660
 cctctgatca gatttttcag ttgtctgatg ggcagaccg aatgattgct tgtgctctcc 720
 agaaccacac tgcctcatcc cgcacaggcc agcgctttcc cttggctgcc cagtcaccct 780
 gcctgggaat catctctaata gctactatct ctctgnnnnn nnnnnnnnnn nnnnnnnnnn 840
 nnnncttcac tggcttccca ttgcctctct tgagtgtcag cgcctcggtc ggtcattcaa 900
 ggctccgtga tgtgacttca agactttcag gctgtgtctc cacacatctc cccatatgct 960
 tctcaccacc ctggtagttc tttatatatc accttgaaat gtctctcttc tcatctccta 1020
 tctccgcctg atgcaatcct tctgtgcctc atggcccttc ctggtagttt tttatt 1076

<210> 35
 <211> 585
 <212> DNA
 <213> Homo sapien

<400> 35
 aactttgtat ttttggcttc aattttcatt tgccgggtata ttgttgact tgcatttgat 60
 taaccactt ttgtatttta gtcctagatg tgtttagttt tcatgagcct tcagtaatca 120
 gcatattggt gcatttgatt taccacagat ggagaatttt cttctcttaa tcagaccaat 180
 taggttgctt ttatctaggg atgtgactgt tagatttcaa ataccttata tcattctggt 240
 tttcttttta aggttggtgt tcttacctct ataagtcatt ccctcctttt tatttatttt 300
 taaagaagta cctaagtcatt gtacatagtc gttgttaaaa aaattcaggc tgagccaccg 360
 cacctggcct tttgattaag cttcacagac acccttcaat atgaagatta ggttccttgt 420

tgcatacatc atttgctccc ctgcattcat tgaattctac ttgattctct gtaacaagtt 480
 tcattgattt gatgttttgc agaaccaatt ctgttttggga ctgcatttct gggatgttct 540
 taattgtctt atcattttat cctaagcttc ttttaaacc cagtt 585

<210> 36
 <211> 2168
 <212> DNA
 <213> Homo sapien
 <220>
 <221> misc_feature
 <222> (2154)..(2154)
 <223> n=a, c, g or t,

<400> 36
 tttttttttt tgggtggaggt ggggtttcgc cacgttggcc aggttggttt tggactgctg 60
 acctcaggtg gtccaccac ctctgcctcc cgaagtgctg ggattacagg cgtgagccac 120
 cacaaccagc ctcccttttg tcagcttacc accccaccct tatgacctta tttaacctta 180
 attacctccc aaaggccctg tctccaaata cagtcctctgt tggaggtcag agcttcagca 240
 tatgaattta gtcattggga ggtgacacaa ttcagtccat aacagatggc caaagtctga 300
 cccatagtgt tcaacagctg ctgttagtta ctctccttgg ctagtagtat gccttgaggc 360
 attaaagaac aaacacacaa agctagtgtg gtgtaggagg caccctgaga agttctgtga 420
 tgggctctgt ctcttggtgc tgacagaggaa ggcatttgtg tgtaagtga tgggtggcctc 480
 aacctggtga ggggtctcaga tttcaacatc ctccctaaatc ccttgggggt gtttttgtgg 540
 aattcagatt ctgcctaaca ggctggggag gtacaggcta agggcctaca tttcttttct 600
 ttttcttttt tgagacacag tctcactttg tcaccaggt gaggcgatca tggctcaatg 660
 tagcctcaac ctccccaggt ctcaagcgat cctccacct cagcctcctg ggactacagg 720
 cacataccac cacagctggc taatgtattt tttttagtag acgaggtttc actatgttgc 780
 ccaggctggc cttgaactcc tgggctcagg tgatcctcct gccttggcct cccaaggctg 840
 tgggggttata ggtgtgagcc accacacctg atcaaggacc tatgttccta aggagctccc 900
 tggtaatgct aatactgctg gtccacggac cacacttttg agtagtgagg gcctgtgttc 960
 agcagcacag tagccaggag ccacgtgtgg ctgtttcagt ctccgtctta gttaaataaa 1020
 actagaaaat ccagtgcctc agtgactgta attgcattgc atggaggaaa gttctaattg 1080
 ggcagtgcc catcttaaag gaccgaatga tccacttttt ctttactttt aaatcttctt 1140
 tctttatatt tttagagaca gggctcttgc ctgtagctga agctggagtg cagtggcact 1200
 atcttagctc actgcagcct aaacctccta agctcaagct gtcctccgc atcagcctcc 1260

tgagtggctg ggactgcagg cacatgctac cacacccagc tattttttttt gtttttttttg 1320
 tagagatggg gtctctccat gttgcccagg ctggctctga actcctggcc tccagcaatc 1380
 ctctgcctt ggctcccaa agtactgggg ttataggcat gggccactgt gccagatatt 1440
 ttttgaaatc catattttct tggcatgagg tagcttttta gcagcatttc ggggtgctgtt 1500
 ccccaaacat agtggtgttt ggcatagac cagattggag gatatactag tcatctagga 1560
 gaattatatt ttgttgcatg tgaagggtgt gttctagata tactaatgca gtgggagaga 1620
 aataaagcaa ttgagaaga ggcagcactt ttataggaaa gaaaatagat acccaatccc 1680
 attgctgatc aatttcctcc tacttaagca acattaggct gattttaatt atcaataagc 1740
 tattaattac caaaatttag agaactaatt caaatttact ggatgttctt aaattattgt 1800
 tttgttatta tcacctggtt ggggtgacaaa ttttaattta ttactgtaa agaattagga 1860
 agcattcaaa taagaaatta cagaaaaatt gatagtgatg cagaaggat gaaggattta 1920
 tatagggtgt aatttcctaa gtctaaaaat ataaacctat gcaaaattag cctttgggtc 1980
 taccagcaat ttaaaaaagc ccgttgtctt aataatttta aacttatggg agcaggactt 2040
 gctctatcta cctaacttgg ggaaagtaaa aataaacata tggcatggta gacacattaa 2100
 cactttctat gaaggatttt taatggaatt cctcgtgccg aattacttgc ctcnaggcca 2160
 aatcctat 2168

<210> 37
 <211> 843
 <212> DNA
 <213> Homo sapien

<400> 37
 cggtcctaag gtagcgacac ttagatgggg gatgggggtg gaggcaggtg agatgagagc 60
 tccagagtgc agcctggggc acggtcattg ggttcaagat gccttcagat gttggggctt 120
 gacagtgaca gtggctgcag ataccagga cccagcagag ggcctggcac agcggggccc 180
 agggacacgt ctgtgagtca ggctggcact caagtgtgaa tctggacaaa gaagggtggca 240
 ttgttggcat tgaatgtca ggggtgggt gaagtcagct agggagaggg gctctcaagg 300
 caccaggact gactccgtgg gtgagggcgg aggaggagac ggcagaggct gagagaatag 360
 ccagagacag gaggaaggtc agcagagcca tgcacagaag cctggagaag atttgaggaa 420
 aggaagggtt ggggtgctgt cattccctgc agctaggagg caagggtgac ccctgtgtgt 480
 ttcatcttc agaggattgg tggctccatc tggctcagaac cccaagtcca gagatgcata 540
 tgagcttcat gccatggtgc aggcagcagg agccatggag caggcctgcg gggtcacggc 600
 ctgcctgctg cttcctgggt tctctaccct gagctcaggg gttgcttgga tgcctgtgc 660

27

```

tgcaccccc caggcgggtgt tccagagagg atgaaagtgg ctatctcaag gtgtgaccag      720
gtagcgggtcc atgcacactg ctttttcctg gggagagtct gggatggaag ggtatgtgtg      780
tgtttctgag actccttgaa taggctccat gtacctatcc acccatctta ccaccacccc      840
ccg                                                                           843

```

```

<210> 38
<211> 632
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (132)..(255)
<223> n=a, c, g or t

```

```

<400> 38
cttataaaca cagaagagtt gaaaaaatac tacagtgaac acctatatgc tattcaattg      60
gattcaccag ttaccatttt gttatatattg ctttctctcc ctagcatatg taagattgta      120
tatgtgcctt tnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn      180
nnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn      240
nnnnnnnnnn nnnnnnacgt aatgatacct aatgtatagt ttgtaatcag ctttcttgta      300
gtaatctggt ttgaaggctt gcatcctgtc atggctccact gattaaattc tgaacttcaa      360
cctgaaactg tggtcactctc attgacttctt ttcttgtacc cattacatca aaatgagttt      420
tcatttttatt tcaatggaag agagaatgga aaagagggga agagtccctg ggtaccttgg      480
gtaaagtatt agcacttact accatatgta tcctagtttt gtgggttttca aacttcaggg      540
agcatcacia ggcttggtta agcacagata gctgtccctc tccccactt tctgattcag      600
gaagtatggg ggtggccagg aatttgcat t c                                     632

```

```

<210> 39
<211> 834
<212> DNA
<213> Homo sapien

```

```

<400> 39
ggaaaattgg gcacccact acccattcta tcaactatcat gaatgcacag ccatgaacaa      60
agatgccctt tagaaaccag tgtagtcca acgggggtcc aagagcctgg gctctggagc      120
cattctgcct aggtttgagt ccttactact tcctggctgt gtgactgaag gtaagtccaa      180
tcccctctc cattttctca tctgaacact aggggcgatt gtagtactgc tctcatctct      240
ggttatgatg atatactggc aatgcacata aaacatttgg aatggagcct agtgctacag      300

```

28

```

atgaatgttt gtgtcctccc caaattcaca tgttccaatc ctaaccccca acttgatggt 360
attaggaggt tggatgtttg ggaggtgatt agatcatgag agtagaacc ccatgaatgt 420
gattagtaac cttataaaaag agtcaacaga gtcctctcac ccctgtgcca tgtgagaaca 480
cagtgagaag atggccctct gtcaaccagg aagcagaccc tcagcacaca ccaaatctgc 540
tgcaaccttg atcttggact tccagcctct agaaccacag tcacactctc atctccctgt 600
ggcctagtca gttacggtac cacggagcca gggattaatg tgggggcaat aacttcccgg 660
tattggaagt ggagctgtcc tggacaatgg ggactgtttg acctcctcaa gtgccgtaac 720
tgcagcaacg tatggagcaa cagcgacagc aatgcccaat gggagaggca actgaaaaac 780
aggaggcaac agctccatgg actggaggaa gacgtgctga aaggaaacct ctga 834

```

```

<210> 40
<211> 696
<212> DNA
<213> Homo sapien

```

```

<400> 40
ttatgaacat gtacttgac cctgtgtgaa taaccaaata agcaaaggca gaatgaggtc 60
aaagagcacg caaattagaa atttttatgt catactgctc tctaaaacta ttttaagtcc 120
tactgatcat gaagaaaaac gtctttcaaa acctcatcaa atactgggta gcactggctc 180
tgttaatttg gtagtccaaa aataactaagt tgctttaatt tgcatttctt tagacaacta 240
gtgagattga acatagtttc acaggataat tggccattag tatttcttcc gtaaactatg 300
tttatgatcc attgaaatct attggttctc ctcttccta cagacttgta agagctcatt 360
ttatgttagg atatcagccc tctgtgtgtc atttctcac caagcttttt ataaaaaact 420
gtatttgtac catatcttag ttttataatt tttatgtaat caaatttatc aatcttttcc 480
tttatggatg ctgggttgca gttcttaaaa aaatcttttc cttctttcct tatttaacaa 540
ctaaaaacaa acaaacaaaa atcctcttaa ctactctgtg gtttcttgcc ttctaaata 600
tttaaactct ttatacatct agagttcatt ttagtataaa gattgtggta agaatccagc 660
tctatttttc aaataggtga cagttatttt catcat 696

```

```

<210> 41
<211> 1431
<212> DNA
<213> Homo sapien

```

```

<400> 41
tctatcctat aagccacagt tcagactcca attcctgtgg gagcggtccc tgagctctca 60
gaccgagcag ggagtgcctc tctgcctgta tcagcacagt gtccactacg tcgtgattgg 120
ctgcttcaat ttgcctctc cttacactga actccatgag ggcaggaatc gtgtgtgatt 180

```

29

```

tgtctcattg acacaggggt tgcaaagata gtggaagcac catgagatgc ttgtggtgtc      240
cagagaggaa tgaaaaaaat agacgaatga cccgatcccaa gtgagaatgg gactcctgta      300
atatgctggg aacaaactcc tctgggccac tgctcctggc tgatgggtgc ccagtgtgtg      360
aggcagaggg ggtagcagtg aaagtgatct ttggcaacca ccttgccag gaccctggc      420
ggccctcaga gagccagct ggtggcctct gggtagccag ctccctoca ttcacaagtc      480
tcaggcctga cagctgctct aattctatca tcctaccccc atccagccat gtccttccgg      540
cagtccaaag cagcatacca cccagcggcc agcccagaga gaagggttt tctgtactgc      600
atggggctag gcgaaggctt gagcagagca gacaggctgt tccatctccc aacaggccac      660
aacaggtag cgaatggcta tctacttgag agtgggcata cctgagtcct ggggcagtga      720
tagtaaccat gcctcctaga ccagtttcac tggattcgct cagtggctga atctgcacgg      780
tctgctcgac tactgcttga tgctcgattg acgctgatgc tgtagtgaga agaaaagggtg      840
gcccttgtcc tttcctttgt atccaaaacc ctcccttcta cctccctgct ttattttaca      900
atgaagcaca gaagtcacct ctttcagaaa ttcttctggg atccccttct ctgggtagtc      960
cccatccct cactcccacc ctgtgctttt ctttgccaca ctgtgttgcg agtggtggtg     1020
tgctgcctc cactgcaaag actttgcttt cctggacata agggcctgtg acatgctctg     1080
ctgtcctgtg cctgccgtgg tgaccagccc gtgagggaag cagtatgttc agggcagtg     1140
ttaactaaaa tcctccctct gcacttcccc aaggtcacac agcaagaaat cctgcagtgg     1200
ttttcttatc tgtaaaatga ggcttgaaag ggtgttgtga agattaagtc agctaagtga     1260
acgtttgaat agcatctggc atgacagtaa gcacttgatg agtactgcat ttgctagtat     1320
tactaggtga tcaataaatg tctcatgtgt taatggatga agaaaggaag gaaagagtga     1380
agatattagt acaagccaga catgccagat ttgtggtgac ccaaataaaa g              1431

```

```

<210> 42
<211> 877
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (94)..(139)
<223> n=a, c, g or t

```

```

<400> 42
tgctcccttt taaaggcaga aactccaggc tttcttccta ccttcttgac tactccacag      60
tgttgtgctt attcctcagg agttattatc cacnnnnnnn nnnnnnnnnn nnnnnnnnnn      120
nnnnnnnnnn nnnnnnnnna tgattttttt gccctaaata tccaagtaca tgctggata      180

```

30

```

atctcactca ctgccacagc ttttattaac agctattttag ttgggaatct aatctcacc 240
ctccactcca ggcccttatt taaattatgt actctgcatt tccatccaaa tgtcctacaa 300
gcatctcaga ctgtgtccca aactgaagtc atcacctact ccgtcccacc tccaaaaca 360
tccagcacca ccataaccct gttcctacac ctattttgta tcactatttc cctgctgccc 420
aaataggaaa cctatgcttc cctottatac tccataacca ccagtcattg tgttgatca 480
attccacctt ctggatatct caaatcctgc tcctcttctc catcccgct ctcatggcct 540
ctgttcagat cctcatttct cattttgtgc tgagtcctga ctctgctgcg ccctctccaa 600
tccatcccca cactacttct agagttagct ttctaagaga tgaaatggac tacaccattc 660
ctctgcttaa aaactctatg cctccttaga acctacctcg aaataaagta caaacctcaa 720
attgtatgca aggccctcca cactccagct gcagccacac tcacaacgtg ccattctcca 780
aatcattttt cttcatgacg tcgagcttac tgcttctgct tagaataact aaactgttaa 840
ggactatgtc ttgccttaac tattaccttt ccagcac 877

```

```

<210> 43
<211> 937
<212> DNA
<213> Homo sapien

```

```

<400> 43
ctgtatccct agattgctga gaggtaaatg aaatgtatgt agaacatcca gcagaggtgc 60
agcctcagac actggctgac catttggtga acctgagaga ggcgtgctct tgcagcttga 120
gttccaacca tgccccacac tccactccca ttaggctaca ctgcaacctc tcacttgat 180
gtttcctata ttagaccatc agttcactgc atgttaagaa ccaaaagtta ccatttacat 240
gacaactctg ttgccaggta cagacatagc catctcatga gatctaaaag gatgtctgag 300
agcaacactg tagcccttta acccacgtga ggctgaacag gattcaatca gctgcccagt 360
gtagactgct ccttagatct ccgtgcacac cagccagaaa ccagagcact gtgggagcgt 420
gtgcacgctc acatccgaga ccctctccag ggacctcagg ctgaagacgg agctgggcag 480
accatgagcc accctggaag agctccaagt cccacaggaa tgctgcctct cctgcccggg 540
gctggccagt tcctactctc agcccatgcc ggccagggaa agggaccagg aagaacaggg 600
cttccctttt ctccccacc tgagtgtaat ttgcttgctc gaccttaag tcatgtgtcc 660
aacctgggcc aatccctgtg ctttcttttt ttgccccgag atggagtgtg gctcattttt 720
cacaccaggc tggagtgcag atggcagcag atctcaggct cactgcaacc tccacctcct 780
gggttcaagc gattctcctg cctcagcctc cccagtagct gggattacag gcaccagcca 840
ccatgcctgg ctaatttttt tgtattttta gtagagatgt ggtttctcca tgttgccag 900

```

gctggtctca aaatcctgac ctcaagcagt cttcttg 937

<210> 44
 <211> 546
 <212> DNA
 <213> Homo sapien

<400> 44
 gggcctgtag ccatgtggca ggggaggtga caggccagat cattttcagg ctcccagatt 60
 aattgtagat tgtcagaggg ccccaaaagt tcatgtagtg taaacagtga aatatctgtg 120
 tgagttgtca ggtgacctgg gcttaaacctc ctccagtgac aagacactca actccttcac 180
 tacttctgcc actggactga tgggaacact ggaaatagtg gatacacctg gtactatgcc 240
 aagtgcataat actgggtact gtgccgagtc cacacactaa atgttatgct aagtgcataat 300
 accaggtacc atgctaagtg catacactgg atactatgct aagtgtatat accaggtcct 360
 atgctaagtg catacactgg atactgtgct aagtgcataat accagttact atgctaagtg 420
 catacactgg atactatgct aagtgcataat accaggtact atgctaagtg catataccag 480
 atactatgct aagtgcataat accagatact atgctaaatg cttataccag gtactatggt 540
 aagtgt 546

<210> 45
 <211> 628
 <212> DNA
 <213> Homo sapien

<400> 45
 ctattattct tattottatt agcatttcac aggcaaggaa gctgagtcac ggtgcttgca 60
 acaggaggag ccagatttga actcaagcag tgtgaccag caggttcaat taactaccat 120
 gcgatactac cctataattt atctcactca agctttaata caatgcttcc ctgtggctct 180
 tatacagggc attggttctg ttcgctggag ctttatagca caagcccaag tcctcttctg 240
 tacaacatcc ttcagataac tgaccacagt tgcattgtct ctcctccttc agagtcttct 300
 acagggtaaa tgtaccaagt tcctcactgg tcctctaata taatgcttca agtctcctgc 360
 tgttgagggtt cttctcctgt cttgctcaga aaagaacaca cccatcatat gtggcaaaga 420
 gcagagtgcc ccattctag gacaaggggtg tgcctatcc atttttctgg aagctacact 480
 taaagcctga aatcgtatca cattgtggac agagtgaatc actgcacctg ttttccaca 540
 aacaaataaa tgctttgggg ctgtcccttc agaaactacc agtgaaacta cattgagaac 600
 acttaattca aggatatctt gaaagctg 628

<210> 46

32

<211> 148
 <212> DNA
 <213> Homo sapien

<400> 46
 tgaccaggggt caccaggtta gtttaagtagc aaggggttgga ttggatctca gcactgtgaa 60
 ggccttcaca gcccctttaa cttccgggct ctgctgcctc atagaagcat catgggaatt 120
 aaaataaaga ttgtggccag gccggtct 148

<210> 47
 <211> 811
 <212> DNA
 <213> Homo sapien

<400> 47
 gtgtcctgaa attactggaa acttttgaac atgaaaccac taaatatatg aacttggttag 60
 tggtaagaaa cctcacagtt aaaaccctt tgttaggggt attgttttta tccatcaggc 120
 acaacctcaa aatctccttt tagatctcca aggacaggaa cgtggcttca gaagttatag 180
 ctctggtttg ggagaagaaa tgtctagctc ttatttggga ggaagcccat ctttatacat 240
 atctatattt ttactcttcc tcaatttaca tgggatctag gcataaaaga caaagtgtcc 300
 tactttgaca gacctggaaa agaagaggaa ttcattctct tcccagggtg ccagagcttc 360
 tggctcagca gcgttttccct accaggagct ggtgtaggct gtgaattcca tgtggaagag 420
 atgggtttta ctttgtcttt ctgcagaggc tggacaagat gtaagataca cctggtcagg 480
 ctcaggaagg tgtttctcat ctttctocta cactggacta aagacagtgg cagaacacac 540
 tcaacttaaa gattgcatgc ttttctcacc cagtgtctcc atgccattag aatagaaaag 600
 gaatatagtg aggagtgggt ttggtaggca gatctttggc agtcggaagt aaaatgggtt 660
 taacaaggac ctccatgaag gtagagtcag aaatagatac ttgtcaattt tcttgctgtt 720
 gacatacatt ctgcagtatt gttgacactg gagtcacttg acatgtacta ttgttgaact 780
 tttttaattt ttaattttta tttcagtagg t 811

<210> 48
 <211> 497
 <212> DNA
 <213> Homo sapien

<400> 48
 aagaccatga tctcatttgg gtggagttaa tgattcccc tccttgctaa aaaatcataa 60
 agtacgtggg agagaaattt gactatacta cggctcatgag ttctgatgag ggaaaaaatt 120
 ccacctaggg gtaggtcttt cctgaaataa atggaaaggg gtgtttaact tcctggggct 180
 tgaaaatcag ggtaaatcc tcttgcttt gtccatttc ttatttgtac tgcaacacat 240

33

```

tttctttaga aggcaaagcc tcatgagagt ccagcatttg gaatttggac attcctgaga      300
ccagagtcac acttctataa atctcagctg gggaaaggga attttcatct ctgctcttaa      360
attcctaagt gttcatgctg ccaggagtgc acatagccgg gacactgacc cagccccagt      420
gtgaaaaggg caccctcca gaagtgatgg aagtcaaaag gtggaagcag caaataggca      480
aagctttcca gcacact                                     497

```

<210> 49
 <211> 328
 <212> DNA
 <213> Homo sapien

```

<400> 49
gtttcacggg gttttaggga aaggagggt ggaataggca cccctgggta acccgagcaa      60
caccactgtt ctccaaaaat ataaatatat aaaaccaaataaatttaaaa tgtaaagcca      120
aggaaagaat ctcagctttg ggtatgatcc tgaggatatg gccttggtgt tgactcattt      180
ttaatgatga aattatgata tcttggttctt gacacattag aatcaccaga cccaagaatg      240
acccttttga tgagtttggg ggtccttagg ccttcatgct gcctgctcct gccttcccag      300
atccacacgc tttttctggt ttttttgt                                     328

```

<210> 50
 <211> 437
 <212> DNA
 <213> Homo sapien

```

<400> 50
ctcaagcagt cctccttcct tggcctotca aatccctgga attacaggca tgagccactg      60
cgcccagcaa gtgaagctgt tttggaaagg actggtcaac ctgcagcttc gccaaaggga      120
gtgagcaagg gtccctgagc acatcccagg ttgggacggc tcccagtggc cgtcagtgcc      180
tctcacttca gagtttggtg tggttactgc agctcctctc tgccatcttg tgctctgtcc      240
agaaggggct ttacaccgg ggtgcggggg cactcctgtg tgtgtctcag aacccttgg      300
ggtgcagtca taggcgacgt tgctgtctc tttcctgccc tgagtagatg ggatatgaag      360
gtgggaattg gccactgaat gttgagatgg aagcggaagc agcaggtgac tgatggagtt      420
atgagtgcag attccct                                     437

```

<210> 51
 <211> 307
 <212> DNA
 <213> Homo sapien

```

<400> 51
tgaaaaatga aataaaatgc acaaaactaa ctggaaagaa ggcaggaaaa acagaaaagg      60

```

34

```

gaggcaagaa acaggacata tagaaaacaa atcataatat ggtaaaactt attatatgca      120
ttacatttta aattatttaa gccagaatgt cagaatggat tcaaatgagt caattataaa      180
tatctaaaag aaatcacatt taaagacaca ggctggaaaa taaaaaaata gaaaatgata      240
cttcatgtaa atactaggca taagagagcc agaataataat attaatatca aggtagattt      300
attttttc                                     307

```

```

<210> 52
<211> 423
<212> DNA
<213> Homo sapien

```

```

<400> 52
gaaaatttgg ctaatggttg gcccaagggc tggtttgtga aactcctggg gctogaagtg      60
aatccgcctg ccttgtcccc gcaaaagtgc tgggaattac aagggtgtgca ccacctgtgc      120
ctggccaata cottacaggt taagcagcta tgtggccttg tgcatttcac ttaactgtgt      180
ttcatttgtc ctttcatgga gatagacttt gttgatatcc tttcatcggt gatgtgcaga      240
acaaagtatg ctgcttggga aattataaac atactttcat ttgaaatatt gggtaatatc      300
aggagtcctt tgctttaata gaagtagatc tggaattgct tttttttttt tcttgggaata      360
tggtaacacc aaagcatctg agctactcac acccatttcc agaggaaaac atgatcattt      420
cca                                             423

```

```

<210> 53
<211> 190
<212> DNA
<213> Homo sapien

```

```

<400> 53
cttaggggtg ggctgctgtc ctcatgggtc aaaatgggtg ccgtgggttc acatctcaca      60
gtgagataga agaagacaga agaatgatgt ctacacctcc ttctaaggag acttcagcaa      120
agcctgcctc taccactttc tgtgactgag ggaaagttaa ttcataactca gagccaatta      180
acctcgtgta                                     190

```

```

<210> 54
<211> 3969
<212> DNA
<213> Homo sapien

```

```

<400> 54
cgcagcaaac acatccgtag aaggcagcgc ggccgccgag agccgcagcg ccgctcgccc      60
gccgcccccc accccgccgc cccgcccggc gaattgcgcc ccgcgcccct cccctcgcgc      120
ccccgagaca aagaggagag aaagtttgcg cggccgagcg gggcaggtga ggagggtgag      180

```

35

ccgcgcggga ggggcccgcg tgggccccgg ctcagccccc gccgcgccc ccagcccgcg 240
 gccgcgagca gcgcccggac cccccagcgg cggcccccg cgcgccagcc ccccggcccg 300
 ccatgggcgc cgcggcccg accctgcggc tggcgctcgg cctcctgctg ctggcgacgc 360
 tgcttcgccc ggccgacgcc tgcagctgct ccccggtgca cccgcaacag gcgttttgca 420
 atgcagatgt agtgatcagg gccaaagcgg tcagtgagaa ggaagtggac tctggaaacg 480
 acatttatgg caaccctatc aagaggatcc agtatgagat caagcagata aagatgttca 540
 aagggcctga gaaggatata gagtttatct acacggcccc ctccctcggca gtgtgtgggg 600
 tctcgctgga cgttggagga aagaaggaat atctcattgc aggaaaggcc gagggggacg 660
 gcaagatgca catcacccctc tgtgacttca tcgtgccctg ggacaccctg agcaccaccc 720
 agaagaagag cctgaaccac aggtaccaga tgggctgcga gtgcaagatc acgcgctgcc 780
 ccatgatccc gtgctacatc tcctccccgg acgagtgcct ctggatggac tgggtcacag 840
 agaagaacat caacgggcac caggccaagt tcttcgcctg catcaagaga agtgacggct 900
 cctgtgcgtg gtaccgccc gcggcgcccc ccaagcagga gtttctcgac atcgaggacc 960
 cataagcagg cctccaacgc cctgtggcc aactgcaaaa aaagcctcca agggtttcga 1020
 ctgggtccagc tctgacatcc ctccctggaa acagcatgaa taaaacactc atcccatggg 1080
 tccaaattaa tatgattctg ctccccctt ctcccttttag acatggttgt gggctctggag 1140
 ggagacgtgg gtccaaggtc ctcatcccat cctccctctg ccaggcacta tgtgtctggg 1200
 gcttcgatcc ttgggtgcag gcagggtcgg gacacgcggc ttccctccca gtccctgcct 1260
 tggcacccgc acagatgcca agcaggcagc acttagggat ctcccagctg ggtaggggca 1320
 gggcctggaa atgtgcattt tgcagaaact tttgagggtc gttgcaagac tgtgtagcag 1380
 gcctaccagg tccctttcat cttgagaggg acatggccct tgttttctgc agcttccacg 1440
 cctctgcact cctgcccct ggcaagtgc cccatcgccc cggcgccac catgagctcc 1500
 cagcacctga ctccccccac atccaagggc agcctggaac cagtggctag ttcttgaagg 1560
 agcccatca atcctattaa tcctcagaat tcagtgga gcctccctct gaccttgta 1620
 gaaatgggag cgagaaaccc cagctgagct gcgttcagc ctgagctgag tctttttggg 1680
 ctgcaaccac cccccaccc ccccccccc gccacatgc tcccagctt gcaggaggaa 1740
 tcggtgaggt cctgtcctga ggctgctgtc cggggccggg ggctgccctc aaggccctt 1800
 ccctagctgc tgcggttgcc attgcttctt gcctgttctg gcatcaggca cctggattga 1860
 gttgcacagc tttgctttat cggggcttgt gtgcagggcc cggctgggct ccccatctgc 1920
 acatcctgag gacagaaaaa gctgggtctt gctgtgccct ccaggctta gtgttccctc 1980
 cctcaaagac tgacagccat cgttctgcac ggggctttct gcatgtgacg ccagctaagc 2040

atagtaagaa gtccagccta ggaagggag gattttggag gtaggtggct ttggtgacac 2100
actcacttct ttctcagcct ccaggacact atggcctggt ttaagagaca tcttattttt 2160
ctaaaggtga attctcagat gataggtgaa cctgagttgc agatatacca acttctgctt 2220
gtattttctta aatgacaaaag attacctagc taagaaactt cctaggggac taggggaacct 2280
atgtgttccc tcagtgtggt ttctgaagc cagtgatatg ggggttagga taggaagaac 2340
tttctcggta atgataagga gaatctcttg ttctctcca cctgtgttgt aaagataaac 2400
tgacgatata caggcacatt atgtaaacat acacacgcaa tgaaaccgaa gcttggcggc 2460
ctgggcgtgg tcttgcaaaa tgcttccaaa gccaccttag cctgttctat tcagcggcaa 2520
ccccaaagca cctgttaaga ctctgaccc ccaagtggca tgcagcccc atgccaccg 2580
ggacctggtc agcacagatc ttgatgactt cccttcttag ggcagactgg gagggtatcc 2640
aggaatcggc ccctgcccc cgggcgtttt catgctgtac agtgacctaa agttggtaag 2700
atgtcataat ggaccagtcc atgtgatttc agtatataca actccaccag accctccaa 2760
cccatataac accccacccc tgctcgttc ctgtatggtg atatcatatg taacatttac 2820
tcctgtttct gctgattgtt tttttaatgt ttgggttgt ttttgacatc agctgtaatc 2880
attcctgtgc tgtgtttttt attacccttg gtaggtatta gacttgact ttttttaaaa 2940
aaaggtttct gcatcgtgga agcatttgac ccagagtgga acgcgtggcc tatgcaggtg 3000
gattccttca ggtctttcct ttggttcttt gagcatcttt gctttcattc gtctcccgtc 3060
tttggttctc cagttcaaatt tattgcaaag taaaggatct ttgagtaggt tcggtctgaa 3120
aggtgtggcc tttatatttg atccacacac gttggtcttt taaccgtgct gagcagaaaa 3180
caaaacaggt taagaagagc cgggtggcag ctgacagagg aagccgctca aataccttca 3240
caataaatag tggcaatata tatatagttt aagaaggctc tccatttggc atcgtttaat 3300
ttatatgtta tgttctaagc acagctctct tctcctatct tcatcctgca agcaactcaa 3360
aatattttaa ataaagttaa cattgtagtt attttcaa atcttgcctga taagtattaa 3420
gaaatattgg acttgctgcc gtaattttaa gctctgttga ttttggttcc gtttggtttt 3480
ttgggggagg ggagcactgt gtttatgctg gaatatgaag tctgagacct tccggtgctg 3540
ggaacacaca agagtgtgtg aaagtgtgaca agcagactgc gcatgtctct gatgctttgt 3600
atcattcttg agcaatcgct cgggtccgtgg acaataaaca gtattatcaa agagaaaaaa 3660
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaagaagtgg 3720
gcaaagacaa cggaccacag aaacaagtat aaaacaacta ttgggacgga ggcacaaaag 3780
aaaaaagaga agatgtgaca gaaagcgaca caaaaagaaa ctccgaggac agaaaacgga 3840

37

gggtagcaat accacacaga gaaaaagccg cgggggaaac agaaggatta taaaaaaaaa 3900
 cacgaaggga gcacggagac aaaagagagt acacaaaaag gaggatgcgg gcacacgtat 3960
 aaaataccg 3969

<210> 55
 <211> 814
 <212> DNA
 <213> Homo sapien

<400> 55
 actataacgg tcctaaggta gcgatgtccc tgtgcacata agaaatgata agtcaccttc 60
 agtgggtgta gcacaggctc tggagtcagg cttccctgct tccaggcgtg accatgtcac 120
 taaccagctg tgcaagtctc tttatctgct tctcaatttc cccatcgcac tgggggtaat 180
 gtcagtatct gccttctaga gtagctctga tattaaatga tattaatcta taaaattggt 240
 taaaacagct catggcatgg agttgctggt attcttggtg ctgttacttt tcatattgct 300
 gagtttggtc aattgatcca gtctactgaa atattttaaa atttaaactct tgagtctggt 360
 attcagcata tctgtcatcc ctccctagtct catgctatgg gttgggttcc tcagttccaa 420
 ccttgccctta taaaaacaga acaaagcaat ggcaatgaaa ataccctgat gtgagcaggg 480
 acagagtctt gttaaaccat atgatataata tacacacttc tatttaagtg attcaaaacc 540
 atggaagaca acattccatg tgttactcat tccatttagc acatgcttgt cattttggtt 600
 gcttggttcaa ttgtttggct gcgtggcctc agaacacacc ttctatgtg ttaggagaat 660
 aacacccaat aagcaccttt ataggaggaa gagagttcac ttctactcac aggggcttaa 720
 agggccagag agttgattcc ctaactttct ctggaagcta gaatgccaac tcagcaccaa 780
 atcccctgtg ttttatattt ggattcagta tctt 814

<210> 56
 <211> 979
 <212> DNA
 <213> Homo sapien

<400> 56
 cagaattcat gctctcaacc accaagctat actgcctctc atcaacaaaa ctgtatttgg 60
 cacataaatg gagcataatg gtgctatgca agaaactcag atttgggatt tctttctgag 120
 cttatattga aattgttaaa agtagataga tcatcttcta aagctttgca gagatgttaa 180
 agatagttaa atgttttaga tatcggaggc atattttacat ttaaagcaac tgattttact 240
 tagtttactt aagatacccc tgtattgcaa caacaaacat tccacaaagc cagtcttaag 300
 tgaaatttca ttttctgcta gcacacattg aatattttta atcatcggct acttgtctgc 360
 atgaaaaaac atcttctaga tcttattggt catcagacat tgagtattaa tatgtgcagc 420

38

cttctgctgg tcagtggctt tgtggaaagc cattctctgt actgaaagta tcaggaaagc 480
 agttaacata aaccttagtg cctgacattc tgggagtgat gggagacca aggtggaaac 540
 aagagagcag acgaagctga taaccagata catgctagcc cttatgcaca cttctaaata 600
 gacaggctgt agtcattcaa aaatggaagc atctgaataa gaatgacaac acaccattct 660
 tcaggcgtga gtcagtatag agaggttggg aagacacgta atacatggcc caactcaata 720
 ctttttttgt taatttacct ttctctcggt catcatacta tattcctctt ggataccgag 780
 gagctctggg tagtagagaa ccataaaaga caaagtgtc ctataatttc tgtaattagt 840
 tggggaagaa gttctaatat cctgcctgtc cggctgccct tgcccccttg gggtagggct 900
 cctgtttgta gctagataag gttaactttg tttttctttt tttttgagac ggagttgcac 960
 tccagccttg agacagagc 979

<210> 57
 <211> 231
 <212> DNA
 <213> Homo sapien

<400> 57
 cgtagacagg aggtctcact atgtggccca tgctgggatt tcaggctctt tgagctttct 60
 tccatattgg ttagggctgg agcaggactc cactgtggcg ggaggatattt gggcgcaagg 120
 agagcagttg tgaggactgg tggtgacccc cagttactgt ctgttggtgt ttggatgctg 180
 ggcacgggaa gcctggctgt aagtgtgtgt ttccgggccg ggcgtgagcc a 231

<210> 58
 <211> 1446
 <212> DNA
 <213> Homo sapien

<400> 58
 ttctcccctc tgtaatgtcc cccaaattaa tctatactgg gttcgggttc ctcctttctgt 60
 tttttattac accctaatta tttctgtccc agttgccaca gccattatta tgacctcta 120
 acttcccac ctagtagaat ttccgcatag cccaagtgt cgggaacctt ttaaccagaa 180
 atgtccaagc actctttaaa aagatacacc agtaaccaat tttgtaattt atattgcaag 240
 ttaattttta gagtgcccat aaacaggga cctgttactg acataaactt ttcaaactgt 300
 ggtcaaccct taacaccctt gggcattaat aacaactatg ttagacttta ctttaagcta 360
 gacatatcac tacagtgtgt ttttccatca gcgagggaat gtcactgat ttacatactg 420
 tctttacctt gtcttaaaga gacaaaaatt tataaagaga ggagttagaa caaagctaaa 480
 ataattaagt atgaagcccc caaatactct ttggcggaata actataatca tatgtcctgt 540

39

```

aaccacttga ggccaaatta ttgtttttta ttattttatt tattttatttt tgagacagtc      600
tctccctctg gcttcctagg ctggatgtgc ttggagtgc atggcgacga tcttggtca      660
ctgcaacctc cacctccagg gtccaaagcg attctcctgc ctcagcctcc tgagtagctg      720
ggattacagg cgtgcatcac caagcccggc taatttttgt attttttagta gagacgggggt      780
tttgccatgt tggtcaggct ggtctcgaac tcctgacctc gtgatctgcc caccttggcc      840
tcccaagtgc tggaattaca ggtgtgagcc acttgtgccc ggcccagaga cttttgaaag      900
tctaaaatgg gttgtatcat agctcacatc actcctcttc ccactgcagg ataaaacca      960
ggcctctcac tgcaacacac aaggccccga atgggtccgc tccctctcc gttcgttcct     1020
ctttctcttg ccaccacgc tgcagcctgg ggtctgtctc tgcctttcac acatccaact     1080
tcattcctgc ccggtggacc ttgcccttgt caccgaacct ttgtggaacg ctttcacccc     1140
agctctccac atgggtgggc tccttctccc gctggagagc caccgtcaca tcctcagaga     1200
ggttctgcct gagtctccc tccaacgcg ccccaaagc cattctgcct tcggccagtt     1260
tcattgcctc cgtagaacct agaaccgtgt gaaatgatct gaggcttcct ccgtgcactg     1320
gttctgcccg tctccctccc cacttcttc tctgccatcc cccaaggcc tggcgagtag     1380
taggtggtca gtaaaccact gttccgtgag tgccgggtgt gcgttgacct ccagtcacc     1440
aacggc                                             1446

```

```

<210> 59
<211> 684
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (93)..(191)
<223> n=a, c, g or t

```

```

<220>
<221> misc_feature
<222> (360)..(528)
<223> n=a, c, g or t

```

```

<400> 59
aataaaaata aatgcaaaaa ttcatccaca aactaatcaa ctcaaataca tccaagggtg      60
tattgtttct atttttcctt atatattgca gcnnnnnnnn nnnnnnnnnn nnnnnnnnnn     120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     180
nnnnnnnnnn naagttatc cctccgtca aagataattt tatatctcat tgcctccttt     240
ggtaaggcta atcagatact caaagaatgc aacctgtgtc tottaaccta actgtcacct     300

```

40

```

ggaagactgc ctgcctgctt caaagttggt cagcgcctttt ctggacgtga accgatgtgn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnta ctaaacacat 540
actatatgtc aggccttgaa ctaaacatctt caaatgaatt gtcttttgta ttatctctta 600
taatttaaaa aattattttac aatctgtatt ctgctgaaat atctcataat ttttaatggc 660
tatatatgta gttttttgta aatt 684

```

```

<210> 60
<211> 1859
<212> DNA
<213> Homo sapien

```

```

<400> 60
ctgggggtac agaacagcag cacctgctct acagagcgat gttgactcat gagattaaaa 60

ggcctgcagc gggcaggaag aagggaccag caggggaagc atgttgaaat gtatcatgct 120
gtaagcactg gagaacacct gacacagaaa cacagccaca gtgcgggtga ggaaggtgc 180
aagatctgcc ttgcccgtag acccaggtgc ggacaggagg ctgggctcgc aaagcaactc 240
gggtaaaccc cataagccag tctgaggagg gcaagtttcg gtgacggcaa gcatcacaca 300
gggtctgccc acaaaggaaa tgaggagacg aggaatgaca ggagcttcct cctgcacacg 360
gaattcgagg agctgcaaaa ctccctaata gcatggctcc ctaatagctg gcaactgcct 420
gacagttcta agagtgatga aatccagaga tgaaagggct tagcatccct gagcagagag 480
tgctgtggg tgagccgact gctcattaac ctgggctgaa caccctgcat ggtttcttgg 540
gatcaacaaa cacagctgta ggaagtttaa aaaattgaca cctttcttgt gaaacataga 600
tccacccatg gtggcccaca gggttgtgaa tcaaagctga aatgcagagg caatgtgagg 660
aaaagcaaaa gcaaacagta ttattattac ttaggaatgt atttgtagat gactagttta 720
agttagtaat tatataaaga aaagcaaac aatgtttata atacaagtca ggagagtggg 780
ttcctctaag aaggaaggga gagggttgga ctgggaagac acacatgcag cacttcttgg 840
gggttaacaa ggatttactt cttgacctgg gtgggtgcttt ctcaactatt ggctttatat 900
gtatttgcta actcatatct tcccacctgg ttcaattttc cagggtactaa tacacatcta 960
atactcaata cagaaaagaa agaagagcca cgctaggaag atgtgagaac atatagaata 1020
aagaacacag gaaaatggat taaccttagt cagaagtggg cacctttttt ggagactagg 1080
gatgagaaag aatgccaggg gaggatgcat ataaatgcag aggcatttag aagagttaaa 1140
agaattcagg ttatatactat tgtaaaatgg gaagcactat tactgtatct gctgaaagca 1200

```

41

tagctaaaga	gaagagagga	gctaaagaga	agagaggagc	taaagattat	gaggcaaaat	1260
tacagcattt	gttcagggtt	caccagatat	ttaaaatttc	aaggatccat	tcagaattaa	1320
aattgatata	ggacttctga	tactgggttat	ttaataaaagt	taagcagact	cttaactttt	1380
cctcagaagt	tagcatgatt	tagtgattta	acacatttgt	ttaacattct	attctttaaa	1440
atgtagtgaa	ggataataaa	tatttacatt	taattgggaa	ggctattttg	gtaggcaaaa	1500
tagacttcct	ttgccaaagca	ataatgattg	tactaacctc	attttccgtt	cctatttagg	1560
tcatgtatgg	aatgttggtc	tttacattag	tacttcgatc	tatttatatt	gttacatggg	1620
aagtagtttg	gatcatctgc	accatgactg	aagtatattc	agacctataa	cagtaaactct	1680
atagcctaaa	tcaaaataac	acataaaaagc	tcacttcagt	agagatgttg	gaagaaaact	1740
ataacccaac	atattattgtt	atattataag	gaatactggg	tcttcagaca	agttgttaat	1800
gggcatcatg	tgaaaactag	ggttccaaaa	ttaaagacat	gggaaaatta	aagagagct	1859

<210> 61

<211> 1106

<212> DNA

<213> Homo sapien

<400> 61

tgccctctgga	aaacctgtcg	tttgggatgg	gtgcgggaag	gtaggaagtc	tgatcgtgac	60
ctcagatact	agccacattc	tctgggtccag	gtttactagt	aataaagatg	atagattgat	120
ccctaactga	ctcctgcata	agcttctctg	taagttctga	aactatagga	ggggggaaag	180
atggcggagg	tttcctggga	gaaactctga	caaataaatc	cttattcctt	ggcagttcat	240
tattgatttc	ccccatatag	cgaacaccct	gtctcctgat	cttgatctgt	cttaaagtgc	300
cttccatgtc	cagattaaga	tcctgtcccc	ttgaatgttc	tctctcctta	ctgcctctct	360
ggttctgggc	tgtgcccaca	gctaccctc	aatcactgct	ttctggatta	tggagccact	420
cactactgat	catgcttttg	cttttcctca	cattgcctct	gcatttcagc	tttgattcac	480
aacctgtgg	gccaccatgg	gtggatctat	gtttcacaag	aaagggtgtca	attttttaaa	540
cttcctacag	ctgtgttttg	ttgatcccaa	gaaaccatgc	aggggtgttca	gccaggtta	600
atgagcagtc	ggctcaccca	caggcactct	ctgctcaggg	atgctaagcc	ctttcatctc	660
tggatttcat	cactcttaga	actgtcaggc	agttgccagc	tattagggag	ccatgctatt	720
agggagtttg	gcagctcctc	gaattccgtg	tgcaggagga	agctcctgtc	attcctcgtc	780
tcctcatttc	ctttgtgggc	agaccctgtg	tgatgcttgc	cgtcaccgaa	acttgccctc	840
ctcagactgg	cttatggggg	ttacccgagt	tgctttgcga	gccagcctc	ctgtccgcac	900
ctgggtccac	gggcaaggca	gatcttgcac	cttcocctcaa	ccgcaactgtg	gctgtgtttc	960

42

tgtgtcaggt gttctccagt gcttacagca tgatacatTT caacatgctt cccctgctgg 1020
 tcccttcttc ctgccgctg caggcctttt aatctcatga gtcaacatcg ctctgtagag 1080
 caggtgctgc tgttctgtac ccccag 1106

<210> 62
 <211> 49
 <212> PRT
 <213> Homo sapien

<400> 62

Met Thr Tyr Ile Val Leu Ile Thr Tyr Phe Leu Tyr Asn Tyr Pro Leu
 1 5 10 15

Pro Lys Ala Thr Glu Trp Ser Ala Phe Tyr Pro Ser Phe Gln Ile Asn
 20 25 30

Lys Lys Thr Leu Asp Lys Ser Glu Lys Lys Gly Pro Met Ile Gln Leu
 35 40 45

Lys

<210> 63
 <211> 49
 <212> PRT
 <213> Homo sapien

<400> 63

Met Thr Tyr Ile Val Leu Ile Thr Tyr Phe Leu Tyr Asn Tyr Pro Leu
 1 5 10 15

Pro Lys Ala Thr Glu Trp Ser Ala Phe Tyr Pro Ser Phe Gln Ile Asn
 20 25 30

Lys Lys Thr Leu Asp Lys Ser Glu Lys Lys Gly Pro Met Ile Gln Leu
 35 40 45

Lys

<210> 64
 <211> 67
 <212> PRT
 <213> Homo sapien

<400> 64

Met Glu Ala Ala Val Asn Ile Lys His Asn Pro Gly Leu Ile Gly Leu

```
<210> 66
<211> 103
<212> PRT
<213> Homo sapien

<400> 66
```

44

Met Tyr Ser His Val Tyr Thr Gly Thr Ala Ser Pro Gln Pro Ser Ser
 1 5 10 15

Pro Ala Pro Pro Arg Gly Gly Leu Asp Pro Pro Ser Ser Pro Ala Trp
 20 25 30

Arg Gln Thr Gln Gly Pro Leu Gln Asp Thr Thr Gln His Gln Pro Arg
 35 40 45

Asn Ser Ser Lys Ala Pro Gly Pro Leu Ser Gly Leu Gly Leu Gly Ala
 50 55 60

Thr Arg Glu Gly Pro Arg Gly Pro Pro Gly Arg Ala His Trp Val Arg
 65 70 75 80

Ala Arg Ala Gly Gly Ala Val Pro Gly Pro Ala Arg Ala Leu Leu Asp
 85 90 95

Val Glu His His Gln Arg Arg
 100

<210> 67
 <211> 92
 <212> PRT
 <213> Homo sapien

<400> 67

Met Asp Gly Val Ala Leu Ile Ile Ile Ile Ser Glu Ala Cys Asp Leu
 1 5 10 15

Leu Lys Val Thr Gln Leu Ala Ser Arg Ala Val Gly Thr Thr His Leu
 20 25 30

Trp Ala Ser Cys Leu Ser Lys Asn Gly Glu His Cys Cys Cys Asp Cys
 35 40 45

Val Gln Ile Lys Thr Ile Lys Ile His Tyr Phe Ser Ser Asn Leu Phe
 50 55 60

Phe Ala Thr Pro Gly Leu Gly Ser Pro Ser Gln Gly Leu Ala Pro Pro
 65 70 75 80

Pro Thr Gln Val Arg Asp Phe Thr Val Ile Leu His
 85 90

<210> 68
 <211> 182

45

<212> PRT
 <213> Homo sapien

<400> 68

Met Arg Thr Lys Leu Glu Gly Glu Thr Gly Glu Lys Gly Trp Glu Lys
 1 5 10 15

Val Val Gly Asp Lys Gly Lys Ser Gly Asp Ser Arg Gln Ser Ala Trp
 20 25 30

Thr Gly Glu Gly Gly Arg Glu Pro Gly Asn Ser Cys Lys Ser Phe Asp
 35 40 45

Thr Asp Gly Ile Leu Arg Asp Gly Ser Ser Ser Thr Cys Ala Glu Trp
 50 55 60

Lys Ile Ser Arg Asn Gly Tyr Cys Leu Val Ile Leu Met Gly Gly Glu
 65 70 75 80

Gln Leu Thr Arg Gly Trp Ser Arg Glu Glu Gly Cys Gly Trp Cys Leu
 85 90 95

Met Ser Glu Ile Pro Glu Val Arg Leu Lys Ser Trp Asn Ala Asp Ser
 100 105 110

Arg Trp Cys Gly Cys Met Gly Gly Lys Pro Val Leu Gln Gly Val Trp
 115 120 125

Arg Ala Ser Ala Gly Arg Trp Arg Ala Ala Leu Asp Leu Thr His Ala
 130 135 140

Cys Thr Ala His Asp Leu Thr Thr Pro Ile Ser Gly Gln Gly Ala Cys
 145 150 155 160

Pro Ala Ser Cys Lys Val Ser Ser Leu Ser Gln Arg Ser Leu Cys Ser
 165 170 175

Ile His Pro Ile Gly Pro
 180

<210> 69
 <211> 42
 <212> PRT
 <213> Homo sapien

<400> 69

Met Gly Arg Lys Arg Leu Ile Phe Thr Met Thr Lys Asn Val Met Ala

46
 1 5 10 15

 Leu His Glu Thr Glu Ala Lys Pro Leu Ser Arg Lys Asn Phe Asp Gly
 20 25 30

 Leu Val Tyr Ile Ser Ser Ile Thr Phe Ser
 35 40

 <210> 70
 <211> 48
 <212> PRT
 <213> Homo sapien

 <400> 70

 Met Gly Gly Lys Ser Ile His Phe Gly Val Lys Asn Ile Cys Ile Ala
 1 5 10 15

 Leu Phe Tyr Met Leu Thr Ala Cys Thr His Ile Thr Cys Glu His His
 20 25 30

 Cys His Ser Leu Ala Phe Asn His Lys Thr Asn Phe Arg Tyr Leu Ile
 35 40 45

 <210> 71
 <211> 52
 <212> PRT
 <213> Homo sapien

 <400> 71

 Met Val Tyr Lys Ile Thr Trp Lys Glu Ile Asn Pro Arg Leu Ser Cys
 1 5 10 15

 Trp Glu Arg Leu Pro Lys His Ser Ala Glu Leu Gly His Gln Glu Ser
 20 25 30

 Cys Cys Leu Leu Cys Asn Gln Lys Val Thr Ser Ser Arg Thr Thr Val
 35 40 45

 Leu Leu Gln Gln
 50

 <210> 72
 <211> 62
 <212> PRT
 <213> Homo sapien

 <400> 72

 Met Leu Asn Ala Leu Phe Ala Ile Leu Leu Asp Ile Leu Glu Thr Gly

```

<210> 74
<211> 382
<212> PRT
<213> Homo sapien

<400> 74

Met  Pro  Ile  Asn  Asp  Arg  Leu  Asp  Lys  Glu  Asn  Val  Val  His  Ile  His
1              5                      10                      15

His  Gly  Ile  Leu  Cys  Thr  His  Lys  Lys  Glu  Cys  Asp  His  Val  His  Lys
                20                      25                      30

```

48

Lys Glu Cys Arg Asp Met Asp Gly Ala Ala Ser His Tyr Pro Gln Gln
 35 40 45

Thr Asn Thr Gly Lys Glu Asn Gln Thr Pro Arg Val Leu Ser Phe Gly
 50 55 60

His Arg Thr Ser Ser Asn Ile Tyr Gln Lys Lys Lys Lys Gly Val Gly
 65 70 75 80

Gly Asp Ser Arg Trp Pro Asn Arg Lys Ser Thr Ser Leu Gln Leu Pro
 85 90 95

Val Arg Ser Thr Gln Lys Asp His Asn Thr Ser Pro Ala Arg Glu Gln
 100 105 110

Asn Trp Thr Glu Asn Glu Phe Asp Lys Leu Thr Glu Val Gly Phe Arg
 115 120 125

Arg Trp Val Ile Thr Asn Ser Ser Glu Leu Lys Glu His Val Pro Thr
 130 135 140

Gln Cys Lys Lys Ser Lys Asn Leu Glu Lys Ser Asp Gly Glu Asn Gly
 145 150 155 160

Thr Lys Leu Glu Asn Thr Leu Gln Gly Ile Ile Gln Glu Asn Phe Pro
 165 170 175

Asn Leu Ala Arg Gln Ala Asn Ile Gln Ile Gln Glu Ile Gln Arg Thr
 180 185 190

Pro Gln Arg Tyr Ser Ser Arg Arg Ala Thr Leu Arg His Lys Ile Ile
 195 200 205

Arg Leu Thr Lys Val Glu Met Lys Glu Lys Arg Leu Arg Ala Ala Arg
 210 215 220

Glu Lys Cys Arg Val Thr His Arg Gly Lys Pro Ile Arg Val Ile Ala
 225 230 235 240

Asp Leu Ser Ala Glu Ile Leu Gln Ala Arg Arg Glu Trp Gly Pro Ile
 245 250 255

Phe Asn Val Leu Lys Glu Lys Asp Phe Gln Pro Ile Ile Ser Tyr Pro
 260 265 270

Ala Lys Leu Ser Leu Thr Gln His Ser Pro Ser Gly Gly Gly His Arg

49

275 280 285
 Gly Ala Cys Val Thr Ile Thr Pro Ala Ser Gly Ser Ser Glu Gln Ser
 290 295 300
 Glu Arg Leu Ile Cys Leu Arg Glu Ser Lys Gly Ser Lys Gln Val Ser
 305 310 315 320
 Leu Pro Gly Asn Pro Glu Asn Cys Thr Gly Ser Tyr Pro Arg Pro Pro
 325 330 335
 Arg Gln Tyr Leu Tyr Lys Ser Ala Arg Asn Gly Val Thr Glu Leu Gly
 340 345 350
 Val Ser Pro Asn Val Asp Met Thr Glu Ile Thr Thr Pro Lys Ser Ile
 355 360 365
 Gln Ile Ser Arg Lys Leu Gly Trp Val Gln Thr Ser Pro Glu
 370 375 380

 <210> 75
 <211> 255
 <212> PRT
 <213> Homo sapien

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

50

Gly Gly Gln Ala Val Ala Ile Gly Gln Phe Lys Asp Arg Ile Thr Gly
 115 120 125

Ser Asn Asp Pro Gly Asn Ala Ser Ile Thr Ile Ser His Met Gln Pro
 130 135 140

Ala Asp Ser Gly Ile Tyr Ile Cys Asp Val Asn Asn Pro Pro Asp Phe
 145 150 155 160

Leu Gly Gln Asn Gln Gly Ile Leu Asn Val Ser Val Leu Val Lys Pro
 165 170 175

Ser Lys Pro Leu Cys Ser Val Gln Gly Arg Pro Glu Thr Gly His Thr
 180 185 190

Ile Ser Leu Ser Cys Leu Ser Ala Leu Gly Thr Pro Ser Pro Val Tyr
 195 200 205

Tyr Trp His Lys Leu Glu Gly Arg Asp Ile Val Pro Val Lys Glu Asn
 210 215 220

Phe Asn Pro Thr Thr Gly Ile Leu Val Ile Gly Asn Leu Thr Asn Cys
 225 230 235 240

Trp Thr Arg Leu Leu Pro Val Tyr Cys His Gln Gln Thr Trp Gln
 245 250 255

<210> 76
 <211> 38
 <212> PRT
 <213> Homo sapien

<400> 76

Met Val Phe Tyr Lys Thr Asp Leu Leu Ser Leu Glu Asp Ile Phe Trp
 1 5 10 15

Arg Met Leu His Lys Ile Tyr Glu Tyr Cys Thr Met Leu Ile Ile Lys
 20 25 30

Thr Phe Tyr Thr Cys Val
 35

<210> 77
 <211> 84
 <212> PRT
 <213> Homo sapien

51

<400> 77

Met Ala Ala Ala Val Leu Thr Asp Arg Ala Gln Val Ser Val Thr Phe
 1 5 10 15

Asp Asp Val Ala Val Thr Phe Thr Lys Glu Glu Trp Gly Gln Leu Asp
 20 25 30

Leu Ala Gln Arg Thr Leu Tyr Gln Glu Val Met Leu Glu Asn Cys Gly
 35 40 45

Leu Leu Val Ser Leu Ala Pro Leu Gln Thr His Pro Gly Ser Ile Arg
 50 55 60

Cys Ile Leu Gln Ser Asn Gln Val Asp Thr Leu Ile Lys Ile Pro Glu
 65 70 75 80

His Thr Glu His

<210> 78

<211> 77

<212> PRT

<213> Homo sapien

<400> 78

Met Arg Ala Ala Pro Leu Ile Phe Ser Leu Tyr Cys Val Glu His Leu
 1 5 10 15

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

Leu Cys Leu Cys Gln Arg His Leu Ser Pro Cys His Leu Gly Ala Leu
 35 40 45

Ile Thr Tyr Leu Pro Val Arg Ser Thr Ser Ser Gly Ala Asp Ser Leu
 50 55 60

Pro Phe Leu Gln Asn Thr Pro Phe Phe Val Val Ser Leu
 65 70 75

<210> 79

<211> 72

<212> PRT

<213> Homo sapien

<400> 79

Met Val Gln Pro Leu Arg Lys Thr Val Trp Val Phe Leu Lys Asn Val

1	5														10														15													
Lys	Ile	Glu	Leu	Pro	Cys	Asp	Leu	Ala	Thr	Leu	Phe	Leu	Asp	Ile	Tyr																											
			20						25						30																											
Pro	Lys	Lys	Leu	Lys	Ser	Gly	Ser	Gln	Arg	Gly	Ile	Ser	Ile	Cys	Arg																											
		35						40						45																												
Phe	Ile	Ala	Ala	Leu	Phe	His	Asn	Ile	Pro	Arg	Tyr	Thr	Asn	Asn	Leu																											
		50						55						60																												
Ser	Val	His	Arg	Arg	Arg	Asn	Glu																																			
65					70																																					
<210> 80																																										
<211> 1049																																										
<212> PRT																																										
<213> Homo sapien																																										
.																																										
<400> 80																																										
Met	Ser	Ala	Cys	Asn	Thr	Phe	Thr	Glu	His	Val	Trp	Lys	Pro	Gly	Glu																											
1				5						10						15																										
Cys	Lys	Asn	Cys	Phe	Lys	Pro	Lys	Ser	Leu	His	Gln	Leu	Pro	Pro	Asp																											
		20						25						30																												
Pro	Glu	Lys	Ala	Pro	Ile	Thr	His	Gly	Asn	Val	Lys	Thr	Asn	Ala	Asn																											
		35						40						45																												
His	Ser	Asn	Asn	His	Arg	Ile	Arg	Asn	Thr	Gly	Asn	Phe	Arg	Pro	Pro																											
		50						55						60																												
Val	Ala	Lys	Lys	Pro	Thr	Ile	Ala	Val	Lys	Pro	Thr	Met	Ile	Val	Ala																											
65					70						75						80																									
Asp	Gly	Gln	Ser	Ile	Cys	Gly	Glu	Leu	Ser	Ile	Gln	Glu	His	Cys	Glu																											
			85						90						95																											
Asn	Lys	Pro	Val	Ile	Ile	Gly	Trp	Asn	Arg	Asn	Arg	Ala	Ala	Leu	Ser																											
			100						105						110																											
Gln	Lys	Pro	Leu	Asn	Asn	Asn	Asn	Glu	Asp	Asp	Glu	Gly	Ile	Ser	His																											
		115						120						125																												
Val	Pro	Lys	Pro	Tyr	Gly	Asn	Asn	Asp	Ser	Ala	Lys	Lys	Met	Ser	Asp																											
		130						135						140																												

53

Asn Asn Asn Gly Leu Thr Glu Val Leu Lys Glu Ile Ala Gly Leu Asp
 145 150 155 160

Thr Ala Pro Gln Ile Arg Gly Asn Glu Thr Asn Ser Arg Glu Thr Phe
 165 170 175

Leu Gly Arg Ile Asn Asp Cys Tyr Lys Arg Ser Leu Glu Arg Lys Leu
 180 185 190

Pro Pro Ser Cys Met Ile Gly Gly Ile Lys Glu Thr Gln Gly Lys His
 195 200 205

Val Ile Leu Ser Gly Ser Thr Glu Val Ile Ser Asn Glu Gly Gly Arg
 210 215 220

Phe Cys Tyr Pro Glu Phe Ser Ser Gly Glu Glu Ser Glu Glu Asp Val
 225 230 235 240

Leu Phe Ser Asn Met Glu Glu Glu His Glu Ser Trp Asp Glu Ser Asp
 245 250 255

Glu Glu Leu Leu Ala Met Glu Ile Arg Met Arg Gly Gln Pro Arg Phe
 260 265 270

Ala Asn Phe Arg Ala Asn Thr Leu Ser Pro Val Arg Phe Phe Val Asp
 275 280 285

Lys Lys Trp Asn Thr Ile Pro Leu Arg Asn Lys Ser Leu Gln Arg Ile
 290 295 300

Cys Ala Val Asp Tyr Asp Asp Ser Tyr Asp Glu Ile Leu Asn Gly Tyr
 305 310 315 320

Glu Glu Asn Ser Val Val Ser Tyr Gly Gln Gly Ser Ile Gln Ser Met
 325 330 335

Val Ser Ser Asp Ser Thr Ser Pro Asp Ser Ser Leu Thr Glu Glu Ser
 340 345 350

Arg Ser Glu Thr Ala Ser Ser Leu Ser Gln Lys Ile Cys Asn Gly Gly
 355 360 365

Leu Ser Pro Gly Asn Pro Gly Asp Ser Lys Asp Met Lys Glu Ile Glu
 370 375 380

54

Pro Asn Tyr Glu Ser Pro Ser Ser Asn Asn Gln Asp Lys Asp Ser Ser
 385 390 395 400

Gln Ala Ser Lys Ser Ser Ile Lys Val Pro Glu Thr His Lys Ala Val
 405 410 415

Leu Ala Leu Arg Leu Glu Glu Lys Asp Gly Lys Ile Ala Val Gln Thr
 420 425 430

Glu Lys Glu Glu Ser Lys Ala Ser Thr Asp Val Ala Gly Gln Ala Val
 435 440 445

Thr Ile Asn Leu Val Pro Thr Glu Glu Gln Ala Lys Pro Tyr Arg Val
 450 455 460

Val Asn Leu Glu Gln Pro Leu Cys Lys Pro Tyr Thr Val Val Asp Val
 465 470 475 480

Ser Ala Ala Met Ala Ser Glu His Leu Glu Gly Pro Val Asn Ser Pro
 485 490 495

Lys Thr Lys Ser Ser Ser Ser Thr Pro Asn Ser Pro Val Thr Ser Ser
 500 505 510

Ser Leu Thr Pro Gly Gln Ile Ser Ala His Phe Gln Lys Ser Ser Ala
 515 520 525

Ile Arg Tyr Gln Glu Val Trp Thr Ser Ser Thr Ser Pro Arg Gln Lys
 530 535 540

Ile Pro Lys Val Glu Leu Ile Thr Ser Gly Thr Gly Pro Asn Val Pro
 545 550 555 560

Pro Arg Lys Asn Cys His Lys Ser Ala Pro Thr Ser Pro Thr Ala Thr
 565 570 575

Asn Ile Ser Ser Lys Thr Ile Pro Val Lys Ser Pro Asn Leu Ser Glu
 580 585 590

Ile Lys Phe Asn Ser Tyr Asn Asn Ala Gly Met Pro Pro Phe Pro Ile
 595 600 605

Ile Ile His Asp Glu Pro Thr Tyr Ala Arg Ser Ser Lys Asn Ala Ile
 610 615 620

Lys Val Pro Ile Val Ile Asn Pro Asn Ala Tyr Asp Asn Leu Ala Ile

55

625					630						635				640
Tyr	Lys	Ser	Phe	Leu	Gly	Thr	Ser	Gly	Glu	Leu	Ser	Val	Lys	Glu	Lys
				645					650					655	
Thr	Thr	Ser	Val	Ile	Ser	His	Thr	Tyr	Glu	Glu	Ile	Glu	Thr	Glu	Ser
			660					665					670		
Lys	Val	Pro	Asp	Asn	Thr	Thr	Ser	Lys	Thr	Thr	Asp	Cys	Leu	Gln	Thr
		675					680					685			
Lys	Gly	Phe	Ser	Asn	Ser	Thr	Glu	His	Lys	Arg	Gly	Ser	Val	Ala	Gln
	690					695					700				
Lys	Val	Gln	Glu	Phe	Asn	Asn	Cys	Leu	Asn	Arg	Gly	Gln	Ser	Ser	Pro
705					710					715					720
Gln	Arg	Ser	Tyr	Ser	Ser	Ser	His	Ser	Ser	Pro	Ala	Lys	Ile	Gln	Arg
				725					730					735	
Ala	Thr	Gln	Glu	Pro	Val	Ala	Lys	Ile	Glu	Gly	Thr	Gln	Glu	Ser	Gln
			740					745					750		
Met	Val	Gly	Ser	Ser	Ser	Thr	Arg	Glu	Lys	Ala	Ser	Thr	Val	Leu	Ser
		755					760					765			
Gln	Ile	Val	Ala	Ser	Ile	Gln	Pro	Pro	Gln	Ser	Pro	Pro	Glu	Thr	Pro
	770					775					780				
Gln	Ser	Gly	Pro	Lys	Ala	Cys	Ser	Val	Glu	Glu	Leu	Tyr	Ala	Ile	Pro
785					790					795					800
Pro	Asp	Ala	Asp	Val	Ala	Lys	Ser	Thr	Pro	Lys	Ser	Thr	Pro	Val	Arg
				805					810					815	
Pro	Lys	Ser	Leu	Phe	Thr	Ser	Gln	Pro	Ser	Gly	Glu	Ala	Glu	Ala	Pro
			820					825					830		
Gln	Thr	Thr	Asp	Ser	Pro	Thr	Thr	Lys	Val	Gln	Lys	Asp	Pro	Ser	Ile
		835					840					845			
Lys	Pro	Val	Thr	Pro	Ser	Pro	Ser	Lys	Leu	Val	Thr	Ser	Pro	Gln	Ser
	850					855					860				
Glu	Pro	Pro	Ala	Pro	Phe	Pro	Pro	Pro	Arg	Ser	Thr	Ser	Ser	Pro	Tyr
865					870					875					880

56

His Ala Gly Asn Leu Leu Gln Arg His Phe Thr Asn Trp Thr Lys Pro
 885 890 895

Thr Ser Pro Thr Arg Ser Thr Glu Ala Glu Ser Val Leu His Ser Glu
 900 905 910

Gly Ser Arg Arg Ala Ala Asp Ala Lys Pro Lys Arg Trp Ile Ser Phe
 915 920 925

Lys Ser Phe Phe Arg Arg Arg Lys Thr Asp Glu Glu Asp Asp Lys Glu
 930 935 940

Lys Glu Arg Glu Lys Gly Lys Leu Val Gly Leu Asp Gly Thr Val Ile
 945 950 955 960

His Met Leu Pro Pro Pro Pro Val Gln Arg His His Trp Phe Thr Glu
 965 970 975

Ala Lys Gly Glu Ser Ser Glu Lys Pro Ala Ile Val Phe Met Tyr Arg
 980 985 990

Cys Asp Pro Ala Gln Gly Gln Leu Ser Val Asp Gln Ser Lys Ala Arg
 995 1000 1005

Thr Asp Gln Ala Ala Val Met Glu Lys Gly Arg Ala Glu Asn Ala
 1010 1015 1020

Leu Leu Gln Asp Ser Glu Lys Lys Arg Ser His Ser Ser Pro Ser
 1025 1030 1035

Gln Ile Pro Lys Lys Ile Leu Arg Tyr Val Lys
 1040 1045

<210> 81
 <211> 60
 <212> PRT
 <213> Homo sapien

<400> 81

Met His Tyr Thr Val Lys Ser Ala Leu His Ile Tyr Gly Phe Tyr Ile
 1 5 10 15

His Glu Phe Asn Gln Pro Gly Ile Glu Lys Leu Phe Asn Cys Ile Cys
 20 25 30

Thr Glu His Val Gln Phe Phe Leu Val Ile Ile Pro Gln Ala Tyr Ser
35 40 45

Ile Thr Met Ile Tyr Ile Val Leu Gly Ile Ile Val
50 55 60

```
<210> 82
<211> 63
<212> PRT
<213> Homo sapien
```

Met Gly Val Lys Arg Gly Leu Phe Leu Cys Ser Ser Leu Leu Lys Ser
1 5 10 15

Cys Ser Ile Ser Lys Thr Thr Thr Pro Ala Pro Leu His Ser Ser Tyr
20 25 30

Ser Ile Ser Ser Leu Tyr Ser Ser Ser Pro Arg Glu Gly Ser Ser Val
35 40 45

Leu Leu Leu Val Ser Gly Tyr Leu Ala Ile Leu Tyr Leu Phe His
50 55 60

<210>	83
<211>	18
<212>	PRT
<213>	Homo sapien

Met Lys Glu Val Met Glu Val Thr Arg Lys Ser Leu Ser Trp Leu Lys
1 5 10 15

Leu Ser

```
<210> 84
<211> 55
<212> PRT
<213> Homo sapien
```

Met Glu Tyr Leu Thr Leu Thr Cys Gly Val Asn Met Asn Thr Ser Trp
1 5 10 15

Ala Pro Ser Arg Pro Arg Phe Pro Ser Arg Gln Tyr Lys Val Glu Leu
20 25 30

58

Gln Gly Ile Ser Met Val Ala Ser Gly Thr Glu Ile Leu Phe Leu Tyr
 35 40 45

Pro Ile Asp Ser Asn Leu Asn
 50 55

<210> 85
 <211> 54
 <212> PRT
 <213> Homo sapien

<400> 85

Met Ser Arg Thr Met Arg Glu Arg Arg Ile Asn Ile Phe Cys Pro Phe
 1 5 10 15

Pro Glu Met Cys Leu Asn Ser Phe Leu Arg Lys Ala Asp Gln Glu Asn
 20 25 30

Asn Ala Arg Arg Glu Ser Leu Phe Lys Gly Ala Gln Leu Gly Ala Val
 35 40 45

Asp Glu Arg Ile Pro Leu
 50

<210> 86
 <211> 10
 <212> PRT
 <213> Homo sapien

<400> 86

Met Thr Phe Leu Phe Ser Gly Arg Gly Arg
 1 5 10

<210> 87
 <211> 58
 <212> PRT
 <213> Homo sapien

<400> 87

Met Ser Gly Lys Cys Thr His Ser Asn Lys Tyr Arg His Arg Gln Cys
 1 5 10 15

Tyr Ile Leu Phe Leu His Gly Pro Thr Arg Ile Ser Lys Ile Ala Ala
 20 25 30

Asn Val Asn Tyr Gln Ser Leu Ile Ile Leu Leu Glu Ala Ile Ser Ser
 35 40 45

59

Tyr Val Arg Leu Val Ile Gly Pro Arg Thr
 50 55

<210> 88
 <211> 43
 <212> PRT
 <213> Homo sapien

<400> 88

Met Tyr Val Ile Ile Leu Leu Gly Leu Tyr Val Asn Val Ile Tyr Leu
 1 5 10 15

Thr Val Thr Ala Gln Trp Arg Trp Val Gly Gly Asn Leu Tyr Trp Cys
 20 25 30

Lys Glu Met Ala Pro Asp Gly Asn Pro Asn Leu
 35 40

<210> 89
 <211> 75
 <212> PRT
 <213> Homo sapien

<400> 89

Met Ser Ala Trp Pro His Leu Ser Gln Gln Gly Leu Thr Glu Ala Cys
 1 5 10 15

Gln Gly His Ile Lys Ser Pro Cys Gln Val Leu Leu Ala Leu Phe Ser
 20 25 30

Pro Ser His Ile Leu His Ser Arg His Ala Glu Val Phe Ala Asp Tyr
 35 40 45

Ser Ser Leu Thr Ser Glu Thr Pro Ser Gln Pro Gly Phe Ser Asp Pro
 50 55 60

Gly His Val Val His Val Val Pro Ala Tyr Trp
 65 70 75

<210> 90
 <211> 59
 <212> PRT
 <213> Homo sapien

<400> 90

Met Cys Phe Val His Phe Leu Ala Arg Ala Asp Val Ala Leu Val Asp
 1 5 10 15

Cys Gly Glu His His Leu Ser Gly Lys Asn Lys Leu Leu Arg Leu Gly
20 25 30

Pro Asn Asp Ala Asn Pro Phe Ile Cys Lys Gln Ile Met Asn Trp Gln
35 40 45

Pro Cys Gly Lys Lys Thr Arg Leu Arg Ile Thr
50 55

```
<210> 91
<211> 91
<212> PRT
<213> Homo sapien
```

<400> 91

Met Arg Lys Gly Glu Trp Asp Cys His Gly Trp His Lys Pro Met Met
1 5 10 15

Leu Tyr Ser Leu Arg Leu Glu Glu Gly Ala Ala Ser Ser Leu Ser Met
20 25 30

Leu Leu Pro Pro Lys Lys Leu Asp Pro Ile Thr Gly Lys Phe His Tyr
35 40 45

Asn Pro Gln Ala Glu Lys Ser His Ser Val Lys Arg Trp Thr Gln Asp
50 55 60

Ser Gly Leu Gln Gly Ser Leu Pro Leu Glu Pro Ser Ser Ser Pro Phe
65 70 75 80

Cys Tyr Phe Ala Ser Ile Cys Leu Phe Pro Thr
85 90

```
<210> 92
<211> 47
<212> PRT
<213> Homo sapien
```

<400> 92

Met Gln Ser Lys Thr Glu Leu Val Leu Gln Asn Ile Lys Ser Met Lys
1 5 10 15

Leu Val Thr Glu Asn Gln Val Glu Phe Asn Glu Cys Arg Gly Ala Asn
20 25 30

Asp Val Cys Asn Lys Glu Pro Asn Leu His Ile Glu Gly Cys Leu
35 40 45

61

<210> 93
 <211> 124
 <212> PRT
 <213> Homo sapien

<400> 93

Met Ile His Phe Phe Phe Thr Phe Lys Phe Ser Phe Phe Ile Phe Leu
 1 5 10 15

Glu Thr Gly Ser Cys Ser Val Ala Glu Ala Gly Val Gln Trp His Tyr
 20 25 30

Leu Ser Ser Leu Gln Pro Lys Pro Pro Lys Leu Lys Leu Ser Ser Arg
 35 40 45

Ile Ser Leu Leu Ser Gly Trp Asp Cys Arg His Met Leu Pro His Pro
 50 55 60

Ala Ile Phe Phe Val Phe Phe Val Glu Met Gly Ser Leu His Val Ala
 65 70 75 80

Gln Ala Gly Leu Glu Leu Leu Ala Ser Ser Asn Pro Pro Ala Leu Ala
 85 90 95

Ser Gln Ser Thr Gly Val Ile Gly Met Gly His Cys Ala Arg Tyr Phe
 100 105 110

Leu Lys Ser Ile Phe Ser Trp His Glu Val Ala Phe
 115 120

<210> 94
 <211> 83
 <212> PRT
 <213> Homo sapien

<400> 94

Met Pro Thr Met Pro Pro Ser Leu Ser Arg Phe Thr Leu Glu Cys Gln
 1 5 10 15

Pro Asp Ser Gln Thr Cys Pro Trp Ala Pro Leu Cys Gln Ala Leu Cys
 20 25 30

Trp Val Leu Gly Ile Cys Ser His Cys His Cys Gln Ala Pro Thr Ser
 35 40 45

Glu Gly Ile Leu Asn Pro Met Thr Val Pro Gln Ala Glu Leu Trp Ser
 50 55 60

62

Ser His Leu Thr Cys Leu Pro Pro His Pro Pro Ser Lys Cys Arg Tyr
 65 70 75 80

Leu Arg Thr

<210> 95
 <211> 68
 <212> PRT
 <213> Homo sapien

<400> 95

Met Phe Tyr Val His Cys Gln Tyr Ile Ile Ile Thr Arg Asp Glu Ser
 1 5 10 15

Ser Thr Thr Ile Ala Pro Ser Val Gln Met Arg Lys Trp Arg Arg Gly
 20 25 30

Leu Asp Leu Pro Ser Val Thr Gln Pro Gly Ser Ser Lys Asp Ser Asn
 35 40 45

Leu Gly Arg Met Ala Pro Glu Pro Arg Leu Leu Glu Pro Arg Trp Asn
 50 55 60

Tyr Thr Gly Phe
 65

<210> 96
 <211> 58
 <212> PRT
 <213> Homo sapien

<400> 96

Met Ser Ser Tyr Lys Ser Val Gly Arg Arg Gly Thr Asn Arg Phe Gln
 1 5 10 15

Trp Ile Ile Asn Ile Val Tyr Gly Arg Asn Thr Asn Gly Gln Leu Ser
 20 25 30

Cys Glu Thr Met Phe Asn Leu Thr Ser Cys Leu Lys Lys Cys Lys Leu
 35 40 45

Lys Gln Leu Ser Ile Phe Gly Leu Pro Asn
 50 55

<210> 97

63

<211> 181
 <212> PRT
 <213> Homo sapien

<400> 97

Met Val Thr Ile Thr Ala Pro Gly Leu Arg Asp Ala His Ser Gln Val
 1 5 10 15

Asp Ser His Ser Leu Thr Cys Cys Gly Leu Leu Gly Asp Gly Thr Ala
 20 25 30

Cys Leu Leu Cys Ser Ser Leu Arg Leu Ala Pro Cys Ser Thr Glu Lys
 35 40 45

Pro Phe Ser Leu Gly Trp Pro Leu Gly Gly Met Leu Leu Trp Thr Ala
 50 55 60

Gly Arg Thr Trp Leu Asp Gly Gly Arg Met Ile Glu Leu Glu Gln Leu
 65 70 75 80

Ser Gly Leu Arg Leu Val Asn Gly Gly Glu Leu Gly Thr Gln Arg Pro
 85 90 95

Pro Ala Gly Leu Ser Glu Gly Arg Gln Gly Ser Trp Thr Arg Trp Leu
 100 105 110

Pro Lys Ile Thr Phe Thr Ala Thr Pro Ser Ala Ser His Thr Gly His
 115 120 125

Pro Ser Ala Arg Ser Ser Gly Pro Glu Glu Phe Val Pro Ser Ile Leu
 130 135 140

Gln Glu Ser His Ser His Leu Gly Ser Val Ile Arg Leu Phe Phe Ser
 145 150 155 160

Phe Leu Ser Gly His His Lys His Leu Met Val Leu Pro Leu Ser Leu
 165 170 175

Gln Pro Leu Cys Gln
 180

<210> 98
 <211> 52
 <212> PRT
 <213> Homo sapien

<400> 98

64

Met Lys Trp Thr Thr Pro Phe Leu Cys Leu Lys Thr Leu Cys Leu Leu
 1 5 10 15

Arg Thr Tyr Leu Glu Ile Lys Tyr Lys Pro Gln Ile Val Cys Lys Ala
 20 25 30

Leu His Thr Pro Ala Ala Ala Thr Leu Thr Thr Cys His Ser Pro Asn
 35 40 45

His Phe Ser Ser
 50

<210> 99
 <211> 151
 <212> PRT
 <213> Homo sapien

<400> 99

Met Ser His Pro Gly Arg Ala Pro Ser Pro Thr Gly Met Leu Pro Leu
 1 5 10 15

Leu Pro Gly Ala Gly Gln Phe Leu Leu Ser Ala His Ala Gly Gln Gly
 20 25 30

Lys Gly Pro Gly Arg Thr Gly Leu Pro Phe Ser Pro His Pro Glu Cys
 35 40 45

Asn Phe Ala Cys Arg His Leu Ser His Val Ser Asn Leu Gly Gln Ser
 50 55 60

Leu Cys Phe Leu Phe Leu Pro Arg Asp Gly Val Val Leu Ile Phe His
 65 70 75 80

Thr Arg Leu Glu Cys Arg Trp Gln Gln Ile Ser Gly Ser Leu Gln Pro
 85 90 95

Pro Pro Pro Gly Phe Lys Arg Phe Ser Cys Leu Ser Leu Pro Ser Ser
 100 105 110

Trp Asp Tyr Arg His Gln Pro Pro Cys Leu Ala Asn Phe Phe Val Phe
 115 120 125

Leu Val Glu Met Trp Phe Leu His Val Gly Gln Ala Gly Leu Lys Ile
 130 135 140

Leu Thr Ser Ser Ser Leu Leu
 145 150

65

<210> 100
 <211> 29
 <212> PRT
 <213> Homo sapien

<400> 100

Met His Leu Ala Trp Tyr Leu Val Tyr Ala Leu Ser Ile Thr Phe Ser
 1 5 10 15

Val Trp Thr Arg His Ser Thr Gln Tyr Met His Leu Ala
 20 25

<210> 101
 <211> 86
 <212> PRT
 <213> Homo sapien

<400> 101

Met Val Leu Ala Thr Gly Gly Ala Arg Phe Glu Leu Lys Gln Cys Asp
 1 5 10 15

Pro Ala Gly Ser Ile Asn Tyr His Ala Ile Leu Pro Tyr Asn Leu Ser
 20 25 30

His Ser Ser Phe Asn Thr Met Leu Pro Cys Gly Ser Tyr Thr Gly His
 35 40 45

Trp Phe Cys Ser Leu Glu Leu Tyr Ser Thr Ser Pro Ser Pro Leu Leu
 50 55 60

Tyr Asn Ile Leu Gln Ile Thr Asp His Ser Cys Met Ser Pro Pro Pro
 65 70 75 80

Ser Glu Ser Ser Thr Gly
 85

<210> 102
 <211> 26
 <212> PRT
 <213> Homo sapien

<400> 102

Met Arg Gln Gln Ser Pro Glu Val Lys Gly Ala Val Lys Ala Phe Thr
 1 5 10 15

Val Leu Arg Ser Asn Pro Thr Leu Ala Thr
 20 25

66

<210> 103
<211> 66
<212> PRT
<213> Homo sapien

<400> 103

Met Arg Asn Thr Phe Leu Ser Leu Thr Arg Cys Ile Leu His Leu Val
1 5 10 15

Gln Pro Leu Gln Lys Asp Lys Val Lys Pro Ile Ser Ser Thr Trp Asn
20 25 30

Ser Gln Pro Thr Pro Ala Pro Gly Arg Lys Thr Leu Leu Ser Gln Lys
35 40 45

Leu Trp His Pro Gly Lys Arg Met Asn Ser Ser Ser Phe Pro Gly Leu
50 55 60

Ser Lys
65

<210> 104
<211> 41
<212> PRT
<213> Homo sapien

<400> 104

Met Leu Pro Gly Val His Ile Ala Gly Thr Leu Thr Gln Pro Gln Cys
1 5 10 15

Glu Lys Gly Thr Arg Pro Glu Val Met Glu Val Lys Arg Trp Lys Gln
20 25 30

Gln Ile Gly Lys Ala Phe Gln His Thr
35 40

<210> 105
<211> 67
<212> PRT
<213> Homo sapien

<400> 105

Met Cys Gln Glu Gln Asp Ile Ile Ile Ser Ser Leu Lys Met Ser Gln
1 5 10 15

Thr Gln Gly His Ile Leu Arg Ile Ile Pro Lys Ala Glu Ile Leu Ser
20 25 30

Leu Ala Leu His Phe Lys Phe Ile Trp Phe Tyr Ile Phe Ile Phe Leu
35 40 45

Glu Asn Ser Gly Val Ala Arg Val Thr Gln Gly Cys Leu Phe His Pro
50 55 60

Pro Phe Pro
65

```
<210> 106
<211> 38
<212> PRT
<213> Homo sapien
```

<400> 106

Met Cys Ser Gly Thr Leu Ala His Cys Pro Trp Arg Ser Cys Arg Leu
1 5 10 15

Thr Ser Pro Phe Gln Asn Ser Phe Thr Cys Trp Ala Gln Trp Leu Met
20 25 30

Pro Val Ile Pro Gly Ile
35

```
<210> 107
<211> 26
<212> PRT
<213> Homo sapien
```

<400> 107

Met His Lys Thr Asn Trp Lys Glu Gly Arg Lys Asn Arg Lys Gly Arg
1 5 10 15

Gln Glu Thr Gly His Ile Glu Asn Lys Ser
20 25

```
<210> 108
<211> 57
<212> PRT
<213> Homo sapien
```

<400> 108

Met Lys Cys Thr Arg Pro His Ser Cys Leu Thr Cys Lys Val Leu Ala
1 5 10 15

Arg His Arg Trp Cys Thr Pro Cys Asn Ser Gln His Phe Cys Gly Asp
20 25 30

68

Lys Ala Gly Gly Phe Thr Ser Ser Pro Arg Ser Phe Thr Asn Gln Pro
 35 40 45

Leu Gly Gln Pro Leu Ala Lys Phe Ser
 50 55

<210> 109
 <211> 41
 <212> PRT
 <213> Homo sapien

<400> 109

Met Val Gln Asn Gly Cys Arg Gly Ser Thr Ser His Ser Glu Ile Glu
 1 5 10 15

Glu Asp Arg Arg Met Met Ser His Leu Pro Ser Lys Glu Thr Ser Ala
 20 25 30

Lys Pro Ala Ser Thr Thr Phe Cys Asp
 35 40

<210> 110
 <211> 375
 <212> PRT
 <213> Homo sapien

<400> 110

Met Ser Lys Arg Arg Arg Gly Glu Gln Asn His Ile Asn Leu Asp Pro
 1 5 10 15

Trp Asp Glu Cys Phe Ile His Ala Val Ser Arg Lys Gly Cys Gln Ser
 20 25 30

Trp Thr Ser Arg Asn Pro Trp Arg Leu Phe Leu Gln Leu Ala Thr Gly
 35 40 45

Ala Leu Glu Ala Cys Leu Trp Val Leu Asp Val Glu Lys Leu Leu Leu
 50 55 60

Gly Gly Arg Arg Ala Ala Val Pro Arg Thr Gly Ala Val Thr Ser Leu
 65 70 75 80

Asp Ala Gly Glu Glu Leu Gly Leu Val Pro Val Asp Val Leu Leu Cys
 85 90 95

Asp Pro Val His Pro Glu Ala Leu Val Arg Gly Gly Asp Val Ala Arg

69

100	105	110
Asp His Gly Ala Ala Arg Asp Leu Ala Leu Ala Ala His Leu Val Pro		
115	120	125
Val Val Gln Ala Leu Leu Leu Gly Gly Ala Gln Gly Val Pro Gly His		
130	135	140
Asp Glu Val Thr Glu Gly Asp Val His Leu Ala Val Pro Leu Gly Leu		
145	150	155
Ser Cys Asn Glu Ile Phe Leu Leu Ser Ser Asn Val Gln Arg Asp Pro		
165	170	175
Thr His Cys Arg Gly Gly Gly Arg Val Asp Lys Leu Tyr Ile Leu Leu		
180	185	190
Arg Pro Phe Glu His Leu Tyr Leu Leu Asp Leu Ile Leu Asp Pro Leu		
195	200	205
Asp Arg Val Ala Ile Asn Val Val Ser Arg Val His Phe Leu Leu Thr		
210	215	220
Asp Arg Phe Gly Pro Asp His Tyr Ile Cys Ile Ala Lys Arg Leu Leu		
225	230	235
Arg Val His Arg Gly Ala Ala Ala Gly Val Gly Arg Ala Lys Gln Arg		
245	250	255
Arg Gln Gln Gln Glu Ala Glu Arg Gln Pro Gln Gly Ala Gly Arg Gly		
260	265	270
Ala His Gly Gly Pro Gly Gly Trp Ala Gly Gly Gly Arg Arg Trp Gly		
275	280	285
Val Arg Ala Leu Leu Ala Ala Ala Gly Trp Gly Arg Gly Arg Gly Leu		
290	295	300
Ser Arg Gly Arg Gly Gly Pro Leu Pro Arg Gly Ser Pro Ser Ser Pro		
305	310	315
Ala Pro Leu Gly Arg Ala Asn Phe Leu Ser Ser Leu Ser Arg Gly Arg		
325	330	335
Glu Gly Arg Gly Ala Gly Arg Asn Ser Pro Gly Gly Ala Ala Gly Trp		
340	345	350

70

Gly Ala Ala Gly Glu Arg Arg Cys Gly Ser Arg Arg Pro Arg Cys Leu
 355 360 365

Leu Arg Met Cys Leu Leu Arg
 370 375

<210> 111
 <211> 34
 <212> PRT
 <213> Homo sapien

<400> 111

Met Cys Tyr Ser Phe His Leu Ala His Ala Cys His Phe Val Cys Leu
 1 5 10 15

Phe Asn Cys Leu Ala Ala Trp Pro Gln Asn Thr Pro Ser Tyr Val Leu
 20 25 30

Gly Glu

<210> 112
 <211> 90
 <212> PRT
 <213> Homo sapien

<400> 112

Met Thr Thr His His Ser Ser Gly Val Ser Gln Tyr Arg Glu Val Gly
 1 5 10 15

Lys Thr Arg Asn Thr Trp Pro Asn Ser Ile His Phe Leu Leu Ile Tyr
 20 25 30

Pro Ser Leu Gly His His Thr Ile Phe Leu Leu Asp Thr Glu Glu Leu
 35 40 45

Trp Val Val Glu Asn His Lys Arg Gln Ser Ala Pro Ile Ile Ser Val
 50 55 60

Ile Ser Trp Gly Arg Ser Ser Asn Ile Leu Pro Val Arg Leu Pro Leu
 65 70 75 80

Pro Pro Trp Gly Arg Ala Pro Val Cys Ser
 85 90

<210> 113

71

<211> 22
 <212> PRT
 <213> Homo sapien

<400> 113

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

Glu Thr Ser Cys Leu Arg
 20

<210> 114
 <211> 95
 <212> PRT
 <213> Homo sapien

<400> 114

Met Ser Lys His Ser Leu Lys Arg Tyr Thr Ser Asn Gln Phe Cys Asn
 1 5 10 15

Leu Tyr Cys Lys Leu Ile Phe Arg Val Pro Ile Asn Arg Glu Pro Val
 20 25 30

Thr Asp Ile Thr Leu Ser Asn Cys Gly Gln Pro Leu Thr Pro Leu Gly
 35 40 45

Ile Asn Asn Asn Tyr Val Arg Leu Tyr Phe Lys Leu Asp Ile Ser Leu
 50 55 60

Gln Cys Cys Phe Pro Ser Ala Arg Glu Cys Ser Leu Ile Tyr Ile Leu
 65 70 75 80

Ser Leu Pro Cys Leu Lys Glu Thr Lys Ile Tyr Lys Glu Arg Ser
 85 90 95

<210> 115
 <211> 115
 <212> PRT
 <213> Homo sapien

<400> 115

Met Leu Leu Gly Ser Leu Ala Ala Pro Arg Ile Pro Cys Ala Gly Gly
 1 5 10 15

Ser Ser Cys His Ser Ser Ser Pro His Phe Leu Cys Gly Gln Thr Leu
 20 25 30

Cys Asp Ala Cys Arg His Arg Asn Leu Pro Ser Ser Asp Trp Leu Met

72

35

40

45

Gly Phe Thr Arg Val Ala Leu Arg Ala Gln Pro Pro Val Arg Thr Trp
 50 55 60

Val His Gly Gln Gly Arg Ser Cys Thr Phe Pro Gln Pro His Cys Gly
 65 70 75 80

Cys Val Ser Val Ser Gly Val Leu Gln Cys Leu Gln His Asp Thr Phe
 85 90 95

Gln His Ala Ser Pro Ala Gly Pro Phe Phe Leu Pro Ala Ala Gly Leu
 100 105 110

Leu Ile Ser
 115

<210> 116
 <211> 115
 <212> PRT
 <213> Homo sapien

<400> 116

Met Leu Leu Gly Ser Leu Ala Ala Pro Arg Ile Pro Cys Ala Gly Gly
 1 5 10 15

Ser Ser Cys His Ser Ser Ser Pro His Phe Leu Cys Gly Gln Thr Leu
 20 25 30

Cys Asp Ala Cys Arg His Arg Asn Leu Pro Ser Ser Asp Trp Leu Met
 35 40 45

Gly Phe Thr Arg Val Ala Leu Arg Ala Gln Pro Pro Val Arg Thr Trp
 50 55 60

Val His Gly Gln Gly Arg Ser Cys Thr Phe Pro Gln Pro His Cys Gly
 65 70 75 80

Cys Val Ser Val Ser Gly Val Leu Gln Cys Leu Gln His Asp Thr Phe
 85 90 95

Gln His Ala Ser Pro Ala Gly Pro Phe Phe Leu Pro Ala Ala Gly Leu
 100 105 110

Leu Ile Ser
 115