

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
12 February 2004 (12.02.2004)

PCT

(10) International Publication Number
WO 2004/013292 A2

(51) International Patent Classification⁷: **C12N**

(21) International Application Number:
PCT/US2003/024084

(22) International Filing Date: 30 July 2003 (30.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/400,810 2 August 2002 (02.08.2002) US
60/412,197 19 September 2002 (19.09.2002) US
60/416,004 4 October 2002 (04.10.2002) US
60/424,862 8 November 2002 (08.11.2002) US

(71) Applicant (*for all designated States except US*): **INCYTE CORPORATION** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **KHARE, Reena** [IN/US]; 12650 Orella Court, Saratoga, CA 95070 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **MARQUIS, Joseph, P.** [US/US]; 4428 Lazy Lane, San Jose, CA 95135 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **EMERLING, Brooke, M.** [US/US]; 833 W. Buena Avenue, #2108, Chicago, IL 60613 (US). **LINDQUIST, Erika, A.**

[US/US]; 2394 Mariner Square Drive #C-121, Alameda, CA 94501 (US). **CHAWLA, Narinder, K.** [US/US]; 33 Union Square, #712, Union City, CA 94587 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **LEE, Soo Yeun** [US/US]; 40 Westdale Avenue, Daly City, CA 94015 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Corporation, 3160 Porter Drive, Palo Alto, CA 94304 (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, 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, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, 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: EXTRACELLULAR MESSENGERS

(57) Abstract: Various embodiments of the invention provide human extracellular messengers (EXMES) and polynucleotides which identify and encode EXMES. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of EXMES.



WO 2004/013292 A2

EXTRACELLULAR MESSENGERS

TECHNICAL FIELD

The invention relates to novel nucleic acids, extracellular messengers encoded by these
5 nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and
prevention of autoimmune/inflammatory disorders, cell proliferative disorders, and endocrine
disorders. The invention also relates to the assessment of the effects of exogenous compounds on the
expression of nucleic acids and extracellular messengers.

BACKGROUND OF THE INVENTION

10 Intercellular communication is essential for the growth and survival of multicellular
organisms, and in particular, for the function of the endocrine, nervous, and immune systems. In
addition, intercellular communication is critical for developmental processes such as tissue
construction and organogenesis, in which cell proliferation, cell differentiation, and morphogenesis
15 must be spatially and temporally regulated in a precise and coordinated manner. Cells communicate
with one another through the secretion and uptake of diverse types of signaling molecules such as
hormones, growth factors, neuropeptides, and cytokines.

Hormones

Hormones are signaling molecules that coordinately regulate basic physiological processes
20 from embryogenesis throughout adulthood. These processes include metabolism, respiration,
reproduction, excretion, fetal tissue differentiation and organogenesis, growth and development,
homeostasis, and the stress response. Hormonal secretions and the nervous system are tightly
integrated and interdependent. Hormones are secreted by endocrine glands, primarily the
hypothalamus and pituitary, the thyroid and parathyroid, the pancreas, the adrenal glands, and the
25 ovaries and testes.

The secretion of hormones into the circulation is tightly controlled. Hormones are often
secreted in diurnal, pulsatile, and cyclic patterns. Hormone secretion is regulated by perturbations in
blood biochemistry, by other upstream-acting hormones, by neural impulses, and by negative
feedback loops. Blood hormone concentrations are constantly monitored and adjusted to maintain
30 optimal, steady-state levels. Once secreted, hormones act only on those target cells that express
specific receptors.

Most disorders of the endocrine system are caused by either hyposecretion or hypersecretion
of hormones. Hyposecretion often occurs when a hormone's gland of origin is damaged or otherwise
impaired. Hypersecretion often results from the proliferation of tumors derived from hormone-
35 secreting cells. Inappropriate hormone levels may also be caused by defects in regulatory feedback

loops or in the processing of hormone precursors. Endocrine malfunction may also occur when the target cell fails to respond to the hormone.

Hormones can be classified biochemically as polypeptides, steroids, eicosanoids, or amines. Polypeptides, which include diverse hormones such as insulin and growth hormone, vary in size and function and are often synthesized as inactive precursors that are processed intracellularly into mature, active forms. Amines, which include epinephrine and dopamine, are amino acid derivatives that function in neuroendocrine signaling. Steroids, which include the cholesterol-derived hormones estrogen and testosterone, function in sexual development and reproduction. Eicosanoids, which include prostaglandins and prostacyclins, are fatty acid derivatives that function in a variety of processes. Most polypeptides and some amines are soluble in the circulation where they are highly susceptible to proteolytic degradation within seconds after their secretion. Steroids and lipids are insoluble and must be transported in the circulation by carrier proteins. The following discussion will focus primarily on polypeptide hormones.

Hormones secreted by the hypothalamus and pituitary gland play a critical role in endocrine function by regulating hormonal secretions from other endocrine glands in response to neural signals. Hypothalamic hormones include thyrotropin-releasing hormone, gonadotropin-releasing hormone, somatostatin, growth-hormone releasing factor, corticotropin-releasing hormone, substance P, dopamine, and prolactin-releasing hormone. These hormones directly regulate the secretion of hormones from the anterior lobe of the pituitary. Hormones secreted by the anterior pituitary include adrenocorticotrophic hormone (ACTH), melanocyte-stimulating hormone, somatotrophic hormones such as growth hormone and prolactin, glycoprotein hormones such as thyroid-stimulating hormone, luteinizing hormone (LH), and follicle-stimulating hormone (FSH), β -lipotropin, and β endorphins. These hormones regulate hormonal secretions from the thyroid, pancreas, and adrenal glands, and act directly on the reproductive organs to stimulate ovulation and spermatogenesis. The posterior pituitary synthesizes and secretes antidiuretic hormone (ADH, vasopressin) and oxytocin.

Disorders of the hypothalamus and pituitary often result from lesions such as primary brain tumors, adenomas, infarction associated with pregnancy, hypophysectomy, aneurysms, vascular malformations, thrombosis, infections, immunological disorders, and complications due to head trauma. Such disorders have profound effects on the function of other endocrine glands. Disorders associated with hypopituitarism include hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism. Disorders associated with hyperpituitarism include acromegaly, gigantism, and syndrome of inappropriate ADH secretion (SIADH), often caused by benign adenomas.

Hormones secreted by the thyroid and parathyroid primarily control metabolic rates and the regulation of serum calcium levels, respectively. Thyroid hormones include calcitonin, somatostatin,

and thyroid hormone. The parathyroid secretes parathyroid hormone. Disorders associated with hypothyroidism include goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism. Disorders associated with hyperthyroidism include thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease. Disorders associated with hyperparathyroidism include Conn disease (chronic hypercalcemia) leading to bone resorption and parathyroid hyperplasia.

Hormones secreted by the pancreas regulate blood glucose levels by modulating the rates of carbohydrate, fat, and protein metabolism. Pancreatic hormones include insulin, glucagon, amylin, γ -aminobutyric acid, gastrin, somatostatin, and pancreatic polypeptide. The principal disorder associated with pancreatic dysfunction is diabetes mellitus caused by insufficient insulin activity. Diabetes mellitus is generally classified as either Type I (insulin-dependent, juvenile diabetes) or Type II (non-insulin-dependent, adult diabetes). The treatment of both forms by insulin replacement therapy is well known. Diabetes mellitus often leads to acute complications such as hypoglycemia (insulin shock), coma, diabetic ketoacidosis, lactic acidosis, and chronic complications leading to disorders of the eye, kidney, skin, bone, joint, cardiovascular system, nervous system, and to decreased resistance to infection.

The anatomy, physiology, and diseases related to hormonal function are reviewed in McCance, K.L. and S.E. Huether(1994) Pathophysiology: The Biological Basis for Disease in Adults and Children, Mosby-Year Book, Inc., St. Louis, MO; Greenspan, F.S. and J.D. Baxter (1994) Basic and Clinical Endocrinology, Appleton and Lange, East Norwalk, CT.

Growth Factors

Growth factors are secreted proteins that mediate intercellular communication. Unlike hormones, which travel great distances via the circulatory system, most growth factors are primarily local mediators that act on neighboring cells. Most growth factors contain a hydrophobic N-terminal signal peptide sequence which directs the growth factor into the secretory pathway. Most growth factors also undergo post-translational modifications within the secretory pathway. These modifications can include proteolysis, glycosylation, phosphorylation, and intramolecular disulfide bond formation. Once secreted, growth factors bind to specific receptors on the surfaces of neighboring target cells, and the bound receptors trigger intracellular signal transduction pathways. These signal transduction pathways elicit specific cellular responses in the target cells. These responses can include the modulation of gene expression and the stimulation or inhibition of cell division, cell differentiation, and cell motility.

Growth factors fall into at least two broad and overlapping classes. The broadest class includes the large polypeptide growth factors, which are wide-ranging in their effects. These factors

include epidermal growth factor (EGF), fibroblast growth factor (FGF), transforming growth factor- β (TGF- β), insulin-like growth factor (IGF), nerve growth factor (NGF), and platelet-derived growth factor (PDGF), each defining a family of numerous related factors. The large polypeptide growth factors, with the exception of NGF, act as mitogens on diverse cell types to stimulate wound healing, bone synthesis and remodeling, extracellular matrix synthesis, and proliferation of epithelial, epidermal, and connective tissues. Members of the TGF- β , EGF, and FGF families also function as inductive signals in the differentiation of embryonic tissue. NGF functions specifically as a neurotrophic factor, promoting neuronal growth and differentiation.

Another class of growth factors includes the hematopoietic growth factors, which are narrow in their target specificity. These factors stimulate the proliferation and differentiation of blood cells such as B-lymphocytes, T-lymphocytes, erythrocytes, platelets, eosinophils, basophils, neutrophils, macrophages, and their stem cell precursors. These factors include the colony-stimulating factors (G-CSF, M-CSF, GM-CSF, and CSF1-3), erythropoietin, and the cytokines. The cytokines are specialized hematopoietic factors secreted by cells of the immune system and are discussed in detail below.

Growth factors play critical roles in neoplastic transformation of cells *in vitro* and in tumor progression *in vivo*. Overexpression of the large polypeptide growth factors promotes the proliferation and transformation of cells in culture. Inappropriate expression of these growth factors by tumor cells *in vivo* may contribute to tumor vascularization and metastasis. Inappropriate activity of hematopoietic growth factors can result in anemias, leukemias, and lymphomas. Moreover, growth factors are both structurally and functionally related to oncoproteins, the potentially cancer-causing products of proto-oncogenes. Certain FGF and PDGF family members are themselves homologous to oncoproteins, whereas receptors for some members of the EGF, NGF, and FGF families are encoded by proto-oncogenes. Growth factors also affect the transcriptional regulation of both proto-oncogenes and oncosuppressor genes. (Reviewed in Pimentel, E. (1994) Handbook of Growth Factors, CRC Press, Ann Arbor, MI; McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York, NY; Habenicht, A., ed. (1990) Growth Factors, Differentiation Factors, and Cytokines, Springer-Verlag, New York, NY.)

In addition, some of the large polypeptide growth factors play crucial roles in the induction of the primordial germ layers in the developing embryo. This induction ultimately results in the formation of the embryonic mesoderm, ectoderm, and endoderm which in turn provide the framework for the entire adult body plan. Disruption of this inductive process would be catastrophic to embryonic development.

Small Peptide Factors - Neuropeptides and Vasomediators

Neuropeptides and vasomediators (NP/VM) comprise a family of small peptide factors,

typically of 20 amino acids or less. These factors generally function in neuronal excitation and inhibition of vasoconstriction/vasodilation, muscle contraction, and hormonal secretions from the brain and other endocrine tissues. Included in this family are neuropeptides and neuropeptide hormones such as bombesin, neuropeptide Y, neurotensin, neuromedin N, melanocortins, opioids, galanin, somatostatin, tachykinins, urotensin II and related peptides involved in smooth muscle stimulation, vasopressin, vasoactive intestinal peptide, and circulatory system-borne signaling molecules such as angiotensin, complement, calcitonin, endothelins, formyl-methionyl peptides, glucagon, cholecystokinin, gastrin, and many of the peptide hormones discussed above. NP/VMs can transduce signals directly, modulate the activity or release of other neurotransmitters and hormones, and act as catalytic enzymes in signaling cascades. The effects of NP/VMs range from extremely brief to long-lasting. (Reviewed in Martin, C.R. et al. (1985) Endocrine Physiology, Oxford University Press, New York, NY, pp. 57-62.)

Cytokines

Cytokines comprise a family of signaling molecules that modulate the immune system and the inflammatory response. Cytokines are usually secreted by leukocytes, or white blood cells, in response to injury or infection. Cytokines function as growth and differentiation factors that act primarily on cells of the immune system such as B- and T-lymphocytes, monocytes, macrophages, and granulocytes. Like other signaling molecules, cytokines bind to specific plasma membrane receptors and trigger intracellular signal transduction pathways which alter gene expression patterns. There is considerable potential for the use of cytokines in the treatment of inflammation and immune system disorders.

Cytokine structure and function have been extensively characterized *in vitro*. Most cytokines are small polypeptides of about 30 kilodaltons or less. Over 50 cytokines have been identified from human and rodent sources. Examples of cytokine subfamilies include the interferons (IFN- α , - β , and - γ), the interleukins (IL1-IL13), the tumor necrosis factors (TNF- α and - β), and the chemokines. Many cytokines have been produced using recombinant DNA techniques, and the activities of individual cytokines have been determined *in vitro*. These activities include regulation of leukocyte proliferation, differentiation, and motility.

Cytokines interact with a target through receptors expressed on the surface of the responsive cell. Cytokines bind with hemopoietin receptors, receptor kinases, and tumor necrosis factor (TNF)/nerve growth factor (NGF) receptors by bringing together two receptor subunits. This dimerization of receptor subunits transmits a signal through the plasma membrane to the cell cytoplasm. In the case of protein kinase receptors, such as the receptors for epidermal growth factor (EGF) and insulin, the juxtaposition of the two receptor subunit cytoplasmic domains activates their intrinsic tyrosine kinase activity. As a result, the subunits phosphorylate each other. The resulting

phosphorylated tyrosine residues then interact with cytoplasmic proteins containing src homology 2 (SH2) domains. SH2-containing proteins that interact with phosphorylated receptor molecules include phosphatidylinositol 3'-kinase, src kinase family members, GRB2, and shc. These SH2 containing proteins are often associated with other cytoplasmic proteins, such as members of the small, monomeric GTP-binding protein families Ras and Rho, and phosphatases, such as the phosphotyrosine phosphatase SHP-2. The signaling complexes formed by these interactions can initiate signal cascades, such as the kinase cascade involving raf and mitogen activated protein (MAP) kinase, which result in transcriptional regulation and cytoskeleton reorganization. Hemopoietin and TNF/NGF receptors, though they have no intrinsic kinase activity, still activate many of the same signal cascades within responding cells.

Many of the kinases involved in cytokine signaling cascades were first identified as products of oncogenes in cancer cells in which kinase activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode protein kinases. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) *Annu. Rev. Cell Biol.* 8:463-493). Thus, the cell must have regulatory systems which keep the cytokine signaling cascades under appropriate control.

Eps8 is a protein which associates with and is phosphorylated by the EGF receptor. Human tumor cell lines contain high constitutive levels of tyrosine-phosphorylated Eps8, and overexpression of Eps8 in NIH3T3 cells expressing the EGF receptor (EGFR) leads to an enhanced mitogenic response and cell overgrowth (Provenzano, C. et al. (1998) *Exp. Cell Res.* 242:186-200). A family of molecules, which include ABI (Abl interactor protein)-1 and ABI-2/e3B1, interact with tyrosine kinases, such as the src-like kinase Abl, and Eps8. Overexpression of ABI-2/e3B1 in NIH3T3 cells expressing EGFR inhibits the mitogenic response and cell growth. Thus, the ABI family of proteins function as negative regulators of cytokine signaling (Ziemnicka-Kotula, D. et al. (1998) *J. Biol. Chem.* 273:13681-13692).

The SH2-containing phosphotyrosine phosphatases, SHP-1 and SHP-2, are involved in cytokine signaling. SHP-1, the hemopoietic cell phosphatase, is a potent inhibitor of signaling, whereas SHP-2 is a positive signal transducer for several cytokines. A family of transmembrane glycoproteins, called SIRPs (signal regulatory proteins), are substrates of tyrosine kinases.

Phosphorylated SIRPs bind to SHP-2 and have a negative effect on cell response induced by cytokines, including an inhibition of growth factor-induced DNA synthesis. This inhibition correlates with reduced MAP kinase activation in SIRP-transfected NIH3T3 cells stimulated with insulin or EGF. SIRP overexpression also suppressed transformation of NIH3T3 cells by a retrovirus carrying the v-fms oncogene (Kharitononkov, A. et al. (1997) *Nature* 386:181-186).

The activity of an individual cytokine *in vitro* may not reflect the full scope of that cytokine's

activity *in vivo*. Cytokines are not expressed individually *in vivo* but are instead expressed in combination with a multitude of other cytokines when the organism is challenged with a stimulus. Together, these cytokines collectively modulate the immune response in a manner appropriate for that particular stimulus. Therefore, the physiological activity of a cytokine is determined by the stimulus itself and by complex interactive networks among co-expressed cytokines which may demonstrate both synergistic and antagonistic relationships.

Recently, a unique cytokine has been isolated that appears to have anti-tumor activity *in vitro* (Ridge, R.J. and N.J. Sloane (1996) Cytokine 8:1-5). This cytokine, anti-neoplastic urinary protein (ANUP), was originally purified as a dimer from human urine. ANUP was later classified as a cytokine when localization studies demonstrated that it was expressed in human granulocytes. ANUP inhibits the growth of cell lines derived from tumors of the breast, skin, lung, bladder, pancreas, and cervix. However, ANUP does not affect the growth of human non-tumor cell lines. The N-terminal 22 amino acids of ANUP comprise a signal peptide which is cleaved from the mature protein. The first nine amino acids of the mature protein retain about 10% of the anti-tumor activity. In addition, ANUP contains a Ly-6/u-PAR sequence motif that is typical of certain cell surface glycoproteins. This motif is characterized by a distinct pattern of six cysteine residues within a 50-residue consensus sequence. The Ly-6/u-PAR motif is found in the Ly-6 T-lymphocyte surface antigen and in the receptor (u-PAR) for urokinase-type plasminogen activator, an extracellular serine protease.

Chemokines comprise a cytokine subfamily with over 30 members. (Reviewed in Wells, T.N.C. and M.C. Peitsch (1997) J. Leukoc. Biol. 61:545-550.) Chemokines were initially identified as chemotactic proteins that recruit monocytes and macrophages to sites of inflammation. Recent evidence indicates that chemokines may also play key roles in hematopoiesis and HIV-1 infection. Chemokines are small proteins which range from about 6-15 kilodaltons in molecular weight. Chemokines are further classified as C, CC, CXC, or CX₃C based on the number and position of certain cysteine residues. The CC chemokines, for example, each contain a conserved motif consisting of two consecutive cysteines followed by two additional cysteines which occur downstream at 24- and 16-residue intervals, respectively (ExPASy PROSITE database, documents PS00472 and PDOC00434). The presence and spacing of these four cysteine residues are highly conserved, whereas the intervening residues diverge significantly. However, a conserved tyrosine located about 15 residues downstream of the cysteine doublet seems to be important for chemotactic activity. Most of the human genes encoding CC chemokines are clustered on chromosome 17, although there are a few examples of CC chemokine genes that map elsewhere. Other chemokines include lymphotactin (C chemokine); macrophage chemotactic and activating factor (MCAF/MCP-1; CC chemokine); platelet factor 4 and IL-8 (CXC chemokines); and fractalkine and neurotractin (CX₃C chemokines). (Reviewed in Luster, A.D. (1998) N. Engl. J. Med. 338:436-445.)

Human Suppressors of Cytokine Signaling (SOCS) Homologs

Signal transduction is a general process in which cells respond to extracellular signals (hormones, neurotransmitters, growth and differentiation factors, etc.) through a cascade of biochemical reactions beginning with the binding of the signal molecule to a cell membrane receptor and ending with an effect on an intracellular target molecule. Intermediate steps in this process involve the activation of various cytoplasmic proteins by phosphorylation via protein kinases and the translocation of some of these activated proteins to the cell nucleus, where the transcription of specific genes is affected. The signal transduction process regulates all types of cell functions, including cell proliferation, differentiation, and gene transcription.

Many of the cytokine receptors, including those for the growth factors EGF, PDGF, and FGF exhibit intrinsic protein kinase activity. Binding of the cytokine to its receptor triggers the autophosphorylation of a tyrosine residue on the receptor. It is believed that these phosphorylated residues are recognition sites for the binding of other cytoplasmic signaling proteins which link the initial receptor activation at the cell surface to the activation of a specific intracellular target molecule. These signaling proteins contain a src homology 2 (SH2) domain that is a recognition and binding site for the phosphotyrosine residue. SH2 domains are found in a variety of signaling molecules and oncogenic proteins, such as phospholipase C- β , Ras GTP-ase activating protein, and GRB2 (Lowenstein, E.J. et al. (1992) Cell 70:431-442).

While much is known about key events in the activation of signaling pathways, less is known about how they are switched off. Recently, several SH2-containing proteins have been identified that are induced in murine lymphoid cells by various cytokines, including IL-2, IL-3, IL-6, Interferon- γ , and EPO (Yoshimura, A. et al. (1995) EMBO Journal 14:2816-2826; Starr, R. et al. (1997) Nature 387:917-921; and Naka, T. et al. (1997) Nature 387:924-929). A common property of these proteins is the ability to suppress growth and differentiation in murine cells. The induction of these SH2-containing proteins in cytokine stimulated cells suggests that they may function as negative regulators of cytokine signaling. Transcription of the genes encoding four of these proteins, CIS (cytokine-inducible SH2-containing protein), and SOCS-1, -2, and -3 (suppressor of cytokine signaling), is induced by IL-6 both *in vitro* and *in vivo* (Starr et al., *supra*).

The four proteins share little sequence homology in their N-terminal regions, but all contain a central SH2 domain and a conserved C-terminal region designated the "SOCS box." The function of the SOCS box is unknown. However, a conserved core triplet sequence (K/R) (D/E) (Y/F) within the SOCS box is similar to the tyrosine phosphorylation site recognized by the JAK kinase family. This similarity suggests that the SOCS box may provide a site for interaction with, and inhibition of, JAK kinases. The finding that SOCS-1 interacts with the catalytic region of JAK kinases supports this hypothesis (Endo, T.A. et al. (1997) Nature 387:921-24). Constitutive expression of SOCS-1 in

M1 murine lymphoid cells also inhibits the phosphorylation of certain cell signaling components (gp130 and Stat3) in response to IL-6 (Starr et al., *supra*). CIS binds to tyrosine-phosphorylated residues in the beta-chain of the IL-3 and EPO receptors and provides another possible mechanism for suppressing cell signaling by preventing the binding of other signaling proteins (Yoshimura et al., *supra*).

Recently, sixteen additional proteins have been identified containing the SOCS box domain (Hilton, D.J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:114-119). Like the SH2-containing proteins described above, each of the proteins contains a C-terminal SOCS box and a distinctive motif N-terminal of the SOCS box. In addition to four new SOCS proteins containing the SH2 domain, three additional classes of SOCS proteins were found containing WD-40 repeats (WSB-1 and -2), SPRY domains (SSB-1 to -3), or ankyrin repeats (ASB-1 to -3). A class of small GTPases (Rar proteins) that contain the SOCS box were also identified. The function of WSB, SSB, and ASB proteins are as yet unknown. However, like SH2 domains, WD-40 repeats, ankyrin repeats, and SPRY domains have been implicated in protein-protein interactions (Hilton et al., *supra*).

Defects or alterations in the activity of signaling proteins such as CIS may play a role in the development of various proliferative disorders and diseases such as cancer. Loss or rearrangement of the putative human gene encoding CIS is associated with the development of renal cell carcinomas and lung cancer (Yoshimura et al., *supra*). This association suggests that CIS may function as a tumor suppressor gene.

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants.

When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Lung cancer is the leading cause of cancer death for men and the second leading cause of cancer death for women in the U.S. Lung cancers are divided into four histopathologically distinct

groups. Three groups (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) are classified as non-small cell lung cancers (NSCLCs). The fourth group of cancers is referred to as small cell lung cancer (SCLC). Deletions on chromosome 3 are common in this disease and are thought to indicate the presence of a tumor suppressor gene in this region. Activating mutations in K-ras are commonly found in lung cancer and are the basis of one of the mouse models for the disease.

While soft tissue sarcomas are relatively rare, more than 50% of new patients diagnosed with the disease will die from it. The molecular pathways leading to the development of sarcomas are relatively unknown, due to the rarity of the disease and variation in pathology. Colon cancer evolves through a multi-step process whereby pre-malignant colonocytes undergo a relatively defined sequence of events leading to tumor formation. Several factors participate in the process of tumor progression and malignant transformation including genetic factors, mutations, and selection.

To understand the nature of gene alterations in colorectal cancer, a number of studies have focused on the inherited syndromes. Familial adenomatous polyposis (FAP), is caused by mutations in the adenomatous polyposis coli gene (APC), resulting in truncated or inactive forms of the protein. This tumor suppressor gene has been mapped to chromosome 5q. Hereditary nonpolyposis colorectal cancer (HNPCC) is caused by mutations in mis-match repair genes. Although hereditary colon cancer syndromes occur in a small percentage of the population and most colorectal cancers are considered sporadic, knowledge from studies of the hereditary syndromes can be generally applied. For instance, somatic mutations in APC occur in at least 80% of sporadic colon tumors. APC mutations are thought to be the initiating event in the disease. Other mutations occur subsequently. Approximately 50% of colorectal cancers contain activating mutations in ras, while 85% contain inactivating mutations in p53. Changes in all of these genes lead to gene expression changes in colon cancer.

More than 180,000 new cases of breast cancer are diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (Gish, K. (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).

Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, *supra*). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

The relationship between expression of epidermal growth factor (EGF) and its receptor,

EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie, K. et al. (1993) *Cancer and Metastasis Rev.* 12:255-274, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S.S. et al. (1994) *Am. J. Clin. Pathol.* 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down-regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou, Z. et al. (1998) *Int. J. Cancer* 78:95-99; Chen, L. et al. (1990) *Oncogene* 5:1391-1395; Ulrix, W. et al (1999) *FEBS Lett* 455:23-26; Sager, R. et al. (1996) *Curr. Top. Microbiol. Immunol.* 213:51-64; and Lee, S.W. et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba, I.I. et al. (1998) *Clin. Cancer Res.* 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

The most important function of adipose tissue is its ability to store and release fat during periods of feeding and fasting. White adipose tissue is the major energy reserve in periods of excess energy use. Its primary purpose is mobilization during energy deprivation. Understanding how various molecules regulate adiposity and energy balance in physiological and pathophysiological situations may lead to the development of novel therapeutics for human obesity. Adipose tissue is also one of the important target tissues for insulin. Adipogenesis and insulin resistance in type II diabetes are linked and present intriguing relations. Most patients with type II diabetes are obese and obesity in turn causes insulin resistance.

The majority of research in adipocyte biology to date has been done using transformed mouse

preadipocyte cell lines. The culture condition which stimulates mouse preadipocyte differentiation is different from that for inducing human primary preadipocyte differentiation. In addition, primary cells are diploid and may therefore reflect the *in vivo* context better than aneuploid cell lines.

Understanding the gene expression profile during adipogenesis in humans will lead to an

5 understanding of the fundamental mechanism of adiposity regulation. Furthermore, through comparing the gene expression profiles of adipogenesis between donor with normal weight and donor with obesity, identification of crucial genes, potential drug targets for obesity and type II diabetes, will be possible.

Thiazolidinediones (TZDs) act as agonists for the peroxisome-proliferator-activated receptor
10 gamma (PPAR γ), a member of the nuclear hormone receptor superfamily. TZDs reduce hyperglycemia, hyperinsulinemia, and hypertension, in part by promoting glucose metabolism and inhibiting gluconeogenesis. Roles for PPAR γ and its agonists have been demonstrated in a wide range of pathological conditions including diabetes, obesity, hypertension, atherosclerosis, polycystic ovarian syndrome, and cancers such as breast, prostate, liposarcoma, and colon cancer.

15 The mechanism by which TZDs and other PPAR γ agonists enhance insulin sensitivity is not fully understood, but may involve the ability of PPAR γ to promote adipogenesis. When ectopically expressed in cultured preadipocytes, PPAR γ is a potent inducer of adipocyte differentiation. TZDs, in combination with insulin and other factors, can also enhance differentiation of human preadipocytes in culture (Adams et al. (1997) J. Clin. Invest. 100:3149-3153). The relative potency
20 of different TZDs in promoting adipogenesis *in vitro* is proportional to both their insulin sensitizing effects *in vivo*, and their ability to bind and activate PPAR γ *in vitro*. Interestingly, adipocytes derived from omental adipose depots are refractory to the effects of TZDs. It has therefore been suggested that the insulin sensitizing effects of TZDs may result from their ability to promote adipogenesis in subcutaneous adipose depots (Adams et al., *supra*). Further, dominant negative mutations in the
25 PPAR γ gene have been identified in two non-obese subjects with severe insulin resistance, hypertension, and overt non-insulin dependent diabetes mellitus (NIDDM) (Barroso et al. (1998) Nature 402:880-883).

NIDDM is the most common form of diabetes mellitus, a chronic metabolic disease that affects 143 million people worldwide. NIDDM is characterized by abnormal glucose and lipid
30 metabolism that results from a combination of peripheral insulin resistance and defective insulin secretion. NIDDM has a complex, progressive etiology and a high degree of heritability. Numerous complications of diabetes including heart disease, stroke, renal failure, retinopathy, and peripheral neuropathy contribute to the high rate of morbidity and mortality.

At the molecular level, PPAR γ functions as a ligand activated transcription factor. In the
35 presence of ligand, PPAR γ forms a heterodimer with the retinoid X receptor (RXR) which then

activates transcription of target genes containing one or more copies of a PPAR γ response element (PPRE). Many genes important in lipid storage and metabolism contain PPREs and have been identified as PPAR γ targets, including PEPCK, aP2, LPL, ACS, and FAT-P (Auwerx, J. (1999) *Diabetologia* 42:1033-1049). Multiple ligands for PPAR γ have been identified. These include a variety of fatty acid metabolites; synthetic drugs belonging to the TZD class, such as Pioglitazone and Rosiglitazone (BRL49653); and certain non-glitazone tyrosine analogs such as GI262570 and GW1929. The prostaglandin derivative 15-dPGJ2 is a potent endogenous ligand for PPAR γ .

Expression of PPAR γ is very high in adipose but barely detectable in skeletal muscle, the primary site for insulin stimulated glucose disposal in the body. PPAR γ is also moderately expressed in large intestine, kidney, liver, vascular smooth muscle, hematopoietic cells, and macrophages. The high expression of PPAR γ in adipose tissue suggests that the insulin sensitizing effects of TZDs may result from alterations in the expression of one or more PPAR γ regulated genes in adipose tissue. Identification of PPAR γ target genes will contribute to better drug design and the development of novel therapeutic strategies for diabetes, obesity, and other conditions.

Systematic attempts to identify PPAR γ target genes have been made in several rodent models of obesity and diabetes (Suzuki et al. (2000) *Jpn. J. Pharmacol.* 84:113-123; Way et al. (2001) *Endocrinology* 142:1269-1277). However, a serious drawback of the rodent gene expression studies is that significant differences exist between human and rodent models of adipogenesis, diabetes, and obesity (Taylor (1999) *Cell* 97:9-12; Gregoire et al. (1998) *Physiol. Reviews* 78:783-809). Therefore, an unbiased approach to identifying TZD regulated genes in primary cultures of human tissues is necessary to fully elucidate the molecular basis for diseases associated with PPAR γ activity.

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for examining which genes are tissue specific, carrying out housekeeping functions, parts of a signaling cascade, or specifically related to a particular genetic predisposition, condition, disease, or disorder. The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of disease. For example, both the levels and sequences expressed in tissues from subjects with diabetes may be compared with the levels and sequences expressed in normal tissue.

Ovarian cancer

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rate for this disease is very low.

Identification of early-stage markers for ovarian cancer would significantly increase the survival rate. Genetic variations involved in ovarian cancer development include mutation of p53 and microsatellite instability. Gene expression patterns likely vary when normal ovary is compared to ovarian tumors.

There is a need in the art for new compositions, including nucleic acids and proteins, for the
5 diagnosis, prevention, and treatment of autoimmune/inflammatory disorders, cell proliferative disorders, and endocrine disorders.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, extracellular
10 messengers, referred to collectively as 'EXMES' and individually as 'EXMES-1,' 'EXMES-2,'
'EXMES-3,' 'EXMES-4,' 'EXMES-5,' 'EXMES-6,' 'EXMES-7,' 'EXMES-8,' 'EXMES-9,'
'EXMES-10,' 'EXMES-11,' 'EXMES-12,' 'EXMES-13,' 'EXMES-14,' 'EXMES-15,' 'EXMES-16,'
'EXMES-17,' 'EXMES-18,' 'EXMES-19,' 'EXMES-20,' 'EXMES-21,' 'EXMES-22,' 'EXMES-23,'
'EXMES-24,' and 'EXMES-25' and methods for using these proteins and their encoding
15 polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions.
Embodiments also provide methods for utilizing the purified extracellular messengers and/or their
encoding polynucleotides for facilitating the drug discovery process, including determination of
efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the
purified extracellular messengers and/or their encoding polynucleotides for investigating the
20 pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a
polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-
25, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at
least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID
25 NO:1-25, c) a biologically active fragment of a polypeptide having an amino acid sequence selected
from the group consisting of SEQ ID NO:1-25, and d) an immunogenic fragment of a polypeptide
having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. Another
embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID
NO:1-25.

30 Still another embodiment provides an isolated polynucleotide encoding a polypeptide
selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected
from the group consisting of SEQ ID NO:1-25, b) a polypeptide comprising a naturally occurring
amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence
selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of a
35 polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25,

and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-25. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:26-50.

5 Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group
10 consisting of SEQ ID NO:1-25, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

15 Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of a polypeptide having an amino
20 acid sequence selected from the group consisting of SEQ ID NO:1-25, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering
25 the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an
30 amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25.

Still yet another embodiment provides an isolated polynucleotide selected from the group
35 consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group

consisting of SEQ ID NO:26-50, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an

amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, and a pharmaceutically acceptable excipient.

5 In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional EXMES, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an
10 agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ
15 ID NO:1-25, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of
20 treating a disease or condition associated with decreased expression of functional EXMES, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a polypeptide
25 comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. The method comprises a)
30 contacting a sample comprising the polypeptide with a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional EXMES, comprising administering to a patient in need of such treatment the composition.

35 Another embodiment provides a method of screening for a compound that specifically binds

to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, the method comprising a) contacting a sample comprising the target polynucleotide with a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide

comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to “a host cell” includes a plurality of such host cells, and a reference to “an antibody” is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

“EXMES” refers to the amino acid sequences of substantially purified EXMES obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term “agonist” refers to a molecule which intensifies or mimics the biological activity of EXMES. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of EXMES either by directly interacting with EXMES or by acting on components of the biological pathway in which EXMES participates.

An “allelic variant” is an alternative form of the gene encoding EXMES. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of

nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding EXMES include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as EXMES or a polypeptide with at least one functional characteristic of EXMES. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding EXMES, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding EXMES. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent EXMES. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of EXMES is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms “amino acid” and “amino acid sequence” can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where “amino acid sequence” is recited to refer to a sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

“Amplification” relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of EXMES. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of EXMES either by directly interacting with EXMES or by acting on components of the biological pathway in which EXMES participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind EXMES polypeptides can be prepared using intact polypeptides or using

fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH₂), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides

having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic EXMES, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding EXMES or fragments of EXMES may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys, Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
5	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
10	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
15	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
20	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the

evolution of new protein functions.

A "fragment" is a unique portion of EXMES or a polynucleotide encoding EXMES which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue.

5 For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain
10 length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:26-50 can comprise a region of unique polynucleotide sequence
15 that specifically identifies SEQ ID NO:26-50, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:26-50 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:26-50 from related polynucleotides. The precise length of a fragment of SEQ ID NO:26-50 and the region of SEQ ID
20 NO:26-50 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-25 is encoded by a fragment of SEQ ID NO:26-50. A fragment of SEQ ID NO:1-25 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-25. For example, a fragment of SEQ ID NO:1-25 can be used as an immunogenic
25 peptide for the development of antibodies that specifically recognize SEQ ID NO:1-25. The precise length of a fragment of SEQ ID NO:1-25 and the region of SEQ ID NO:1-25 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g.,
30 methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer
35 to the percentage of identical nucleotide matches between at least two polynucleotide sequences

aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more
5 computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-
10 153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic
15 Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2
20 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version
25 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

30 *Gap x drop-off: 50*

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example,
35 as defined by a particular SEQ ID number, or may be measured over a shorter length, for example,

over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

“Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%.

5 Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high
10 stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C₀t or R₀t analysis) or formed between one nucleic acid present in
15 solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

20 "Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of EXMES which is
25 capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of EXMES which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides,
30 polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of EXMES. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other
35 biological, functional, or immunological properties of EXMES.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

5 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

10 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

15 "Post-translational modification" of an EXMES may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of EXMES.

"Probe" refers to nucleic acids encoding EXMES, their complements, or fragments thereof, 20 which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a 25 DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 30 or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, 35 Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in

Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a

vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

5 A “regulatory element” refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

“Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An “RNA equivalent,” in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term “sample” is used in its broadest sense. A sample suspected of containing EXMES, nucleic acids encoding EXMES, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms “specific binding” and “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term “substantially purified” refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A “substitution” refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

“Substrate” refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,

microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

5 "Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or
10 viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to
15 animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a
20 recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods
25 known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of
30 the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an
35 "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have

significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule.

Species variants are polynucleotides that vary from one species to another. The resulting

5 polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a
10 propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for
15 example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

20 THE INVENTION

Various embodiments of the invention include new human extracellular messengers (EXMES), the polynucleotides encoding EXMES, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/inflammatory disorders, cell proliferative disorders, and endocrine disorders.

25 Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide
30 sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME
35 database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ

ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites and potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Accelrys, Burlington MA), and amino acid residues comprising signature sequences, domains, and motifs. Column 5 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are extracellular messengers. For example, SEQ ID NO:1 is 100% identical, from residue M1 to residue S150, to human pre-pro-orphanin FQ (GenBank ID g1185010) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.1e-77$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:1 also has homology to proteins that are extracellular neuropeptides related to opioid peptide precursors as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:1 also contains a vertebrate endogenous opioids neuropeptide domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein families/domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses, and BLAST analyses against the PRODOM database, provide further corroborative evidence that SEQ ID NO:1 is related to an opioid neuropeptide precursor.

For example, SEQ ID NO:6 is a splice variant of human transforming growth factor beta 1 precursor (GenBank ID g1212989) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $6.7E-185$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:6 also has homology to human transforming growth factor beta 1, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:6 also contains a transforming growth factor-beta (TGF-beta)

domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based SMART database of conserved protein families/domains, as well as a transforming growth factor beta like domain and a TGF-beta propeptide domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:6 is a transforming growth factor.

For example, SEQ ID NO:16 is 74% identical, from residue M1 to residue D128, to human preprovasoactive intestinal polypeptide (GenBank ID g340280) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.2×10^{-42} , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:16 also has homology to vasoactive intestinal peptide precursor which produces a neuropeptide hormone that signals via G protein-coupled receptors and may modulate smooth muscle function and electrolyte balance, cell proliferation and survival and inflammatory responses, neuronal cell survival, and oxidative stress response, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:16 also contains a glucagon-like hormone and peptide hormone domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM/SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:16 is a vasoactive intestinal polypeptide (VIP) family member.

For example, SEQ ID NO:22 is 98% identical, from residue M1 to residue S256, to human transforming growth factor-beta (GenBank ID g339564) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 3.7×10^{-199} , which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:22 also has homology to proteins that regulate cell proliferation, differentiation, and apoptosis, and are bone morphogenetic proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:22 also contains a transforming growth factor-beta-like domain and a TGF-beta propeptide, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:22 is a member of the transforming growth factor (TGF)-beta family of extracellular signaling molecules.

SEQ ID NO:2-5, SEQ ID NO:7-15, SEQ ID NO:17-21 and SEQ ID NO:23-25 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID

NO:1-25 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:26-50 or that distinguish between SEQ ID NO:26-50 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄ represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N_{1,2,3...}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FLXXXXXX_gAAAAA_gBBBBB_1_N is a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier

(denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not

available) indicates that the allele frequency was not determined for the population.

The invention also encompasses EXMES variants. Various embodiments of EXMES variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the EXMES amino acid sequence, and can contain at least one functional or structural characteristic of EXMES.

Various embodiments also encompass polynucleotides which encode EXMES. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:26-50, which encodes EXMES. The polynucleotide sequences of SEQ ID NO:26-50, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding EXMES. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding EXMES. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:26-50 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:26-50. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of EXMES.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding EXMES. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding EXMES, but will generally have a greater or lesser number of nucleotides due to additions or deletions of blocks of sequence arising from alternate splicing during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding EXMES over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding EXMES. For example, a polynucleotide comprising a sequence of SEQ ID NO:29 and a polynucleotide comprising a sequence of SEQ ID NO:30 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:34 and a polynucleotide comprising a sequence of SEQ ID NO:43 are splice variants of each other; and a polynucleotide comprising a sequence of SEQ ID NO:32, a polynucleotide comprising a sequence of SEQ ID NO:35, and a polynucleotide comprising a sequence of SEQ ID NO:45 are splice variants of each other. Any one of the splice

variants described above can encode a polypeptide which contains at least one functional or structural characteristic of EXMES.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding EXMES, some bearing minimal
5 similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring EXMES, and all such variations are to be considered
10 as being specifically disclosed.

Although polynucleotides which encode EXMES and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring EXMES under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding EXMES or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring
15 codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding EXMES and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than
20 transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode EXMES and EXMES derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to
25 introduce mutations into a polynucleotide encoding EXMES or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:26-50 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.*
30 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied
35 Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations

of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems).

5 Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

10 The nucleic acids encoding EXMES may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method,
15 inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR
20 Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (BD Clontech, Palo Alto CA) to walk
25 genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

30 When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

35 Capillary electrophoresis systems which are commercially available may be used to analyze

the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode EXMES may be cloned in recombinant DNA molecules that direct expression of EXMES, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express EXMES.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter EXMES-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of EXMES, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable

manner.

In another embodiment, polynucleotides encoding EXMES may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232).

- 5 Alternatively, EXMES itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems).
- 10 Additionally, the amino acid sequence of EXMES, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

- The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The
- 15 composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

- In order to express a biologically active EXMES, the polynucleotides encoding EXMES or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in
- 20 a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding EXMES. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding EXMES. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a
- 25 polynucleotide sequence encoding EXMES and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation
- 30 codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

- Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding EXMES and appropriate transcriptional and
- 35 translational control elements. These methods include *in vitro* recombinant DNA techniques,

synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding EXMES. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding EXMES. For example, routine cloning, subcloning, and propagation of polynucleotides encoding EXMES can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding EXMES into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of EXMES are needed, e.g. for the production of antibodies, vectors which direct high level expression of EXMES may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of EXMES. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such

vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184).

5 Plant systems may also be used for expression of EXMES. Transcription of polynucleotides encoding EXMES may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) 10 *Science* 224:838-843; Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases 15 where an adenovirus is used as an expression vector, polynucleotides encoding EXMES may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses EXMES in host cells (Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659). In addition, transcription enhancers, such as the Rous 20 sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino 25 polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of EXMES in cell lines is preferred. For example, polynucleotides encoding EXMES can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate 30 vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

35 Any number of selection systems may be used to recover transformed cell lines. These

include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; BD Clontech), β -glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding EXMES is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding EXMES can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding EXMES under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding EXMES and that express EXMES may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of EXMES using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on EXMES is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa

NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding EXMES include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding EXMES, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding EXMES may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode EXMES may be designed to contain signal sequences which direct secretion of EXMES through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding EXMES may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric EXMES protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of EXMES activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST),

maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity
5 purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the EXMES encoding sequence and the heterologous protein sequence, so that EXMES may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al.
10 (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled EXMES may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6
15 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

EXMES, fragments of EXMES, or variants of EXMES may be used to screen for compounds that specifically bind to EXMES. One or more test compounds may be screened for specific binding to EXMES. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be
20 screened for specific binding to EXMES. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of EXMES can be used to screen for binding of test compounds, such as antibodies, to EXMES, a variant of EXMES, or a combination of EXMES and/or one or more variants EXMES. In an embodiment, a variant of EXMES can be used to screen for
25 compounds that bind to a variant of EXMES, but not to EXMES having the exact sequence of a sequence of SEQ ID NO:1-25. EXMES variants used to perform such screening can have a range of about 50% to about 99% sequence identity to EXMES, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to EXMES can be
30 closely related to the natural ligand of EXMES, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor EXMES (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

35 In other embodiments, a compound identified in a screen for specific binding to EXMES can

be closely related to the natural receptor to which EXMES binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for EXMES which is capable of propagating a signal, or a decoy receptor for EXMES which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG₁ (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to EXMES, fragments of EXMES, or variants of EXMES. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of EXMES. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of EXMES. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of EXMES.

In an embodiment, anticalins can be screened for specific binding to EXMES, fragments of EXMES, or variants of EXMES. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit EXMES involves producing appropriate cells which express EXMES, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing EXMES or cell membrane fractions which contain EXMES are then contacted with a test compound and binding, stimulation, or inhibition of activity of either EXMES or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is

detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with EXMES, either in solution or affixed to a solid support, and detecting the binding of EXMES to the compound.

Alternatively, the assay may detect or measure binding of a test compound in the presence of a
5 labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include
10 radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to
15 improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

EXMES, fragments of EXMES, or variants of EXMES may be used to screen for compounds that modulate the activity of EXMES. Such compounds may include agonists, antagonists, or partial
20 or inverse agonists. In one embodiment, an assay is performed under conditions permissive for EXMES activity, wherein EXMES is combined with at least one test compound, and the activity of EXMES in the presence of a test compound is compared with the activity of EXMES in the absence of the test compound. A change in the activity of EXMES in the presence of the test compound is indicative of a compound that modulates the activity of EXMES. Alternatively, a test compound is
25 combined with an *in vitro* or cell-free system comprising EXMES under conditions suitable for EXMES activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of EXMES may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding EXMES or their mammalian homologs
30 may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of
35 interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R.

(1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding EXMES may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding EXMES can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding EXMES is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress EXMES, e.g., by secreting EXMES in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of EXMES and extracellular messengers. In addition, examples of tissues expressing EXMES can be found in Table 6 and can also be found in Example XI. Therefore, EXMES appears to play a role in autoimmune/inflammatory disorders, cell proliferative disorders, and endocrine disorders. In the treatment of disorders associated with increased EXMES expression or activity, it is desirable to decrease the expression or activity of EXMES. In the treatment of disorders associated with decreased EXMES expression or activity, it is desirable to increase the expression or activity of EXMES.

Therefore, in one embodiment, EXMES or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of EXMES. Examples of such disorders include, but are not limited to, an

autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an endocrine disorder such as a disorder of the hypothalamus and/or pituitary resulting from lesions such as a primary brain tumor, adenoma, infarction associated with pregnancy, hypophysectomy, aneurysm, vascular malformation, thrombosis, infection, immunological disorder, and complication due to head trauma; a disorder associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism; a disorder associated with hyperpituitarism including acromegaly, gigantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; a disorder associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism; a disorder associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease; a disorder associated with hyperparathyroidism including Conn disease (chronic hypercalcemia); a pancreatic disorder such as Type I or Type II diabetes mellitus and associated complications; a disorder associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal

cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease; a disorder associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbation of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, a hypergonadal disorder associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α -reductase, and gynecomastia.

10 In another embodiment, a vector capable of expressing EXMES or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of EXMES including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified EXMES in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of EXMES including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EXMES may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of EXMES including, but not limited to, those listed above.

20 In a further embodiment, an antagonist of EXMES may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of EXMES. Examples of such disorders include, but are not limited to, those autoimmune/inflammatory disorders, cell proliferative disorders, and endocrine disorders described above. In one aspect, an antibody which specifically binds EXMES may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express EXMES.

25 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding EXMES may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of EXMES including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of EXMES may be produced using methods which are generally known in the art. In particular, purified EXMES may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind EXMES. Antibodies to EXMES may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e., those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with EXMES or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to EXMES have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of EXMES amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to EXMES may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain

antibodies may be adapted, using methods known in the art, to produce EXMES-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

5 Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

10 Antibody fragments which contain specific binding sites for EXMES may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

15 Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between EXMES and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies
20 reactive to two non-interfering EXMES epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

 Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for EXMES. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of EXMES-antibody complex
25 divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple EXMES epitopes, represents the average affinity, or avidity, of the antibodies for EXMES. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular EXMES epitope, represents a true measure of affinity. High-affinity antibody
30 preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the EXMES-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of EXMES, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical
35 Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to

Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of EXMES-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding EXMES, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding EXMES. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding EXMES (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102:469-475; Scanlon, K.J. et al. (1995) *FASEB J.* 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) *Blood* 76:271-278; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) *Br. Med. Bull.* 51:217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87:1308-1315; Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding EXMES may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal,

R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in EXMES expression or regulation causes disease, the expression of EXMES from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in EXMES are treated by constructing mammalian expression vectors encoding EXMES and introducing these vectors by mechanical means into EXMES-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of EXMES include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (BD Clontech, Palo Alto CA). EXMES may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding EXMES from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method

(Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to EXMES expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding EXMES under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding EXMES to cells which have one or more genetic abnormalities with respect to the expression of EXMES. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; *Annu. Rev. Nutr.* 19:511-544) and Verma, I.M. and N. Somia (1997; *Nature* 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding EXMES to target cells which have one or more genetic abnormalities with

respect to the expression of EXMES. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing EXMES to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has
5 been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the
10 control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; *J. Virol.* 73:519-532) and Xu, H. et al. (1994; *Dev. Biol.* 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of
15 the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding EXMES to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based
20 on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for EXMES into the
25 alphavirus genome in place of the capsid-coding region results in the production of a large number of EXMES-coding RNAs and the synthesis of high levels of EXMES in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy
30 application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of EXMES into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the
35 art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding EXMES.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding EXMES. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous

endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. siRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. siRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

siRNA can be generated indirectly by introduction of dsRNA into the targeted cell. Alternatively, siRNA can be synthesized directly and introduced into a cell by transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable siRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected siRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in

selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene can be determined, for example, by northern analysis methods using the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined, for example, by microarray methods; by polyacrylamide gel electrophoresis; and by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding EXMES. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased EXMES expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding EXMES may be therapeutically useful, and in the treatment of disorders associated with decreased EXMES expression or activity, a compound which specifically promotes expression of the polynucleotide encoding EXMES may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a

polynucleotide encoding EXMES is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding EXMES are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is
5 detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding EXMES. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of
10 the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a
15 combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable
20 for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

25 Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable
30 excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of EXMES, antibodies to EXMES, and mimetics, agonists, antagonists, or inhibitors of EXMES.

35 In various embodiments, the compositions described herein, such as pharmaceutical

compositions, may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form.

5 These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton,
10 J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

15 Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising EXMES or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, EXMES or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to
20 transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration
25 range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example EXMES or fragments thereof, antibodies of EXMES, and agonists, antagonists or inhibitors of EXMES, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be
30 determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are
35 used to formulate a range of dosage for human use. The dosage contained in such compositions is

preferably within a range of circulating concentrations that includes the ED₅₀ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. 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.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind EXMES may be used for the diagnosis of disorders characterized by expression of EXMES, or in assays to monitor patients being treated with EXMES or agonists, antagonists, or inhibitors of EXMES. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for EXMES include methods which utilize the antibody and a label to detect EXMES in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring EXMES, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of EXMES expression. Normal or standard values for EXMES expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to EXMES under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of EXMES expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding EXMES may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of EXMES may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of EXMES, and to monitor regulation of EXMES levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding EXMES or closely related molecules may be used to identify nucleic acid sequences which encode EXMES. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding EXMES, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the EXMES encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:26-50 or from genomic sequences including promoters, enhancers, and introns of the EXMES gene.

Means for producing specific hybridization probes for polynucleotides encoding EXMES include the cloning of polynucleotides encoding EXMES or EXMES derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding EXMES may be used for the diagnosis of disorders associated with expression of EXMES. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic

anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and an endocrine disorder such as a disorder of the hypothalamus and/or pituitary resulting from lesions such as a primary brain tumor, adenoma, infarction associated with pregnancy, hypophysectomy, aneurysm, vascular malformation, thrombosis, infection, immunological disorder, and complication due to head trauma; a disorder associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism; a disorder associated with hyperpituitarism including acromegaly, gigantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma; a disorder associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism; a disorder associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease; a disorder associated with hyperparathyroidism including Conn disease (chronic hypercalcemia); a pancreatic disorder such as Type I or Type II diabetes mellitus and associated complications; a disorder associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease; a disorder associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbation of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis; and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, a hypergonadal disorder associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α -reductase, and gynecomastia. Polynucleotides encoding EXMES may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR

technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered EXMES expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding EXMES may be used in assays that
5 detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding EXMES may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered
10 in comparison to a control sample then the presence of altered levels of polynucleotides encoding EXMES in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of
15 EXMES, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding EXMES, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified
20 polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the
25 patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the
30 development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding
35 EXMES may involve the use of PCR. These oligomers may be chemically synthesized, generated

enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding EXMES, or a fragment of a polynucleotide complementary to the polynucleotide encoding EXMES, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or
5 quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding EXMES may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation
10 polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding EXMES are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel
15 electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-
20 based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16
25 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that
30 influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating
35 genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations

and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of EXMES include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, EXMES, fragments of EXMES, or antibodies specific for EXMES may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The

resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

5 Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

25 In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

30 Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and

at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for EXMES to quantify the levels of EXMES expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by contacting the microarray with the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoze, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological

sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

5 In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of
10 protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/25116; Shalon, D. et al. (1995)
15 PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding EXMES may be
20 used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific
25 region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage
30 maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data
35 can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM)

World Wide Web site. Correlation between the location of the gene encoding EXMES on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, EXMES, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between EXMES and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with EXMES, or fragments thereof, and washed. Bound EXMES is then detected by methods well known in the art. Purified EXMES can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding EXMES specifically compete with a test compound for binding EXMES. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with EXMES.

In additional embodiments, the nucleotide sequences which encode EXMES may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding

description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/400,810, U.S. Ser. No. 60/412,197, U.S. Ser. No. 60/416,004, and U.S. Ser. No. 60/424,862, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

10 Incyte cDNAs are derived from cDNA libraries described in the LIFESEQ database (Incyte, Palo Alto CA). Some tissues are homogenized and lysed in guanidinium isothiocyanate, while others are homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates are centrifuged over CsCl cushions or extracted with chloroform. RNA is precipitated from the lysates
15 with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA are repeated as necessary to increase RNA purity. In some cases, RNA is treated with DNase. For most libraries, poly(A)+ RNA is isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA is isolated directly
20 from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene is provided with RNA and constructs the corresponding cDNA libraries. Otherwise, cDNA is synthesized and cDNA libraries are constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Invitrogen), using the recommended
25 procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription is initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters are ligated to double stranded cDNA, and the cDNA is digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA is size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative
30 agarose gel electrophoresis. cDNAs are ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte, Palo Alto CA), pRARE (Incyte), or pINCY (Incyte), or derivatives thereof. Recombinant plasmids are transformed
35 into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α ,

DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I are recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids are purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids are resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA is amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps are carried out in a single reaction mixture. Samples are processed and stored in 384-well plates, and the concentration of amplified plasmid DNA is quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incye cDNA recovered in plasmids as described in Example II are sequenced as follows. Sequencing reactions are processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions are prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides are carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences are identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences are selected for extension using the techniques disclosed in Example VIII.

Polynucleotide sequences derived from Incyte cDNAs are validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof are then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus*

norvegicus, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries are performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences are assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) are used to extend Incyte cDNA assemblages to full length. Assembly is performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages are screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences are translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences are subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences are also used to identify polynucleotide sequence fragments from SEQ ID NO:26-50. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and

amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative extracellular messengers are initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once is set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode extracellular messengers, the encoded polypeptides are analyzed by querying against PFAM models for extracellular messengers. Potential extracellular messengers are also identified by homology to Incyte cDNA sequences that have been annotated as extracellular messengers. These selected Genscan-predicted sequences are then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences are then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis is also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage is available, this information is used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences are obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences are derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

25 "Stitched" Sequences

Partial cDNA sequences are extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III are mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster is analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that are subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval is present on more than one sequence in the cluster are identified, and intervals thus identified are considered to be equivalent by transitivity. For example, if an interval is present on a cDNA and two genomic sequences, then all three intervals are considered to be equivalent. This process allows unrelated but consecutive

genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified are then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) are given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences are translated and compared by BLAST analysis to the genpept and gbpr public databases. Incorrect exons predicted by Genscan are corrected by comparison to the top BLAST hit from genpept. Sequences are further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

10 **"Stretched" Sequences**

Partial DNA sequences are extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III are queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog is then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein is generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both are used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences are therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences are examined to determine whether they contain a complete gene.

VI. Chromosomal Mapping of EXMES Encoding Polynucleotides

The sequences used to assemble SEQ ID NO:26-50 are compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:26-50 are assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon are used to determine if any of the clustered sequences have been previously mapped. Inclusion of a mapped sequence in a cluster results in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between

chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Génethon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST are used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length}(\text{Seq. 1}), \text{length}(\text{Seq. 2})\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding EXMES are analyzed with respect to the tissue

sources from which they are derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system;

5 embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following

10 disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding EXMES. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ database (Incyte, Palo Alto CA).

15 **VIII. Extension of EXMES Encoding Polynucleotides**

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer is synthesized to initiate 5' extension of the known fragment, and the other primer is synthesized to initiate 3' extension of the known fragment. The initial primers are designed using OLIGO 4.06

20 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations is avoided.

Selected human cDNA libraries are used to extend the sequence. If more than one extension

25 is necessary or desired, additional or nested sets of primers are designed.

High fidelity amplification is obtained by PCR using methods well known in the art. PCR is performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contains DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen),

30 and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ are as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5

35 min; Step 7: storage at 4°C.

The concentration of DNA in each well is determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate is scanned in a Fluoroskan II
5 (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture is analyzed by electrophoresis on a 1 % agarose gel to determine which reactions are successful in extending the sequence.

The extended nucleotides are desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and
10 sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides are separated on low concentration (0.6 to 0.8%) agarose gels, fragments are excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected
15 into competent *E. coli* cells. Transformed cells are selected on antibiotic-containing media, and individual colonies are picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells are lysed, and DNA is amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3
20 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA is quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries are reamplified using the same conditions as described above. Samples are diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC
25 DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

30 IX. Identification of Single Nucleotide Polymorphisms in EXMES Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) are identified in SEQ ID NO:26-50 using the LIFESEQ database (Incyte). Sequences from the same gene are clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters is used to distinguish
35 SNPs from other sequence variants. Preliminary filters remove the majority of basecall errors by

requiring a minimum Phred quality score of 15, and remove sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis is applied to the original chromatogram files in the vicinity of the putative SNP. Clone error filters use statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters use statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removes duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs are selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprises 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprises 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprises 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprises 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies are first analyzed in the Caucasian population; in some cases those SNPs which show no allelic variance in this population are not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:26-50 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ - 32 P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate.

Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Skena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one

with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (BD Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1
5 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 µl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element
10 is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope
15 slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma-Aldrich, St. Louis MO) in 95% ethanol. Coated slides are cured in a 110°C oven.

20 Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 µl of the array element DNA, at an average concentration of 100 ng/µl, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).
25 Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

30 Hybridization reactions contain 9 µl of sample mixture consisting of 0.2 µg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the
35 addition of 140 µl of 5X SSC in a corner of the chamber. The chamber containing the arrays is

incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores.

Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each

spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be differentially expressed.

Expression

For example, SEQ ID NO:28 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. The expression of SEQ ID NO:28 was increased by at least two-fold in brain tissues as compared to the reference sample. Therefore, SEQ ID NO:28 can be used as a tissue marker for brain.

In an alternative example, SEQ ID NO:29 and SEQ ID NO:30 showed differential expression associated with breast cancer as determined by microarray analysis. A tumor from the right breast of a 43-year-old female diagnosed with invasive lobular carcinoma was compared to grossly uninvolved breast tissue from the same donor (Huntsman Cancer Institute, Salt Lake City, UT). The tumor is described as well differentiated and metastatic to 2 of 13 lymph nodes. Expression of SEQ ID NO:29 and SEQ ID NO:30 was decreased at least two-fold in cancerous tissue from the patient with breast cancer compared to matched microscopically normal tissue from the same donor. Therefore, in various embodiments, SEQ ID NO:29 and SEQ ID NO:30 can each be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

In an alternative example, SEQ ID NO:29 and SEQ ID NO:30 showed differential expression associated with colon cancer as determined by microarray analysis. Gene expression profiles were obtained by comparing normal sigmoid colon tissue from a 48-year-old female to a sigmoid colon tumor originating from a metastatic gastric sarcoma (stromal tumor) from the same donor (Huntsman Cancer Institute, Salt Lake City, UT). Expression of SEQ ID NO:29 and SEQ ID NO:30 was increased at least two-fold in cancerous tissue from the patient with colon cancer compared to matched microscopically normal tissue from the same donor. Therefore, in various embodiments,

SEQ ID NO:29 and SEQ ID NO:30 can each be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

In an alternative example, SEQ ID NO:29 and SEQ ID NO:30 showed differential expression associated with lung cancer as determined by microarray analysis. Moderately differentiated adenocarcinoma of the right lung was compared to grossly uninvolved lung tissue from a 60 year-old donor (Huntsman Cancer Institute). Expression of SEQ ID NO:29 and SEQ ID NO:30 was decreased at least two-fold in cancerous tissue from the patient with lung cancer compared to matched microscopically normal tissue from the same donor. Therefore, in various embodiments, SEQ ID NO:29 and SEQ ID NO:30 can each be used for one or more of the following: i) monitoring treatment of lung cancer, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

In an alternative example, SEQ ID NO:29 and SEQ ID NO:30 showed differential expression in comparisons of differentiating preadipocytes from adipose tissue of obese and normal donors as determined by microarray analysis. Primary subcutaneous preadipocytes were isolated from adipose tissue of a 28-year-old healthy female with body mass index (BMI) of 23.59. The preadipocytes were cultured and induced to differentiate into adipocytes by culturing them in differentiation medium containing active components PPAR- γ agonist and human insulin (Zen-Bio). Thiazolidinediones or PPAR- γ agonists can bind and activate an orphan nuclear receptor, PPAR- γ , and some of them have been proven to be able to induce human adipocyte differentiation. The preadipocytes were treated with human insulin and PPAR- γ agonist for 3 days and subsequently were switched to medium containing insulin for a variety of time periods ranging from one to 20 days before the cells were collected for analysis. Differentiated adipocytes were compared to untreated preadipocytes maintained in culture in the absence of inducing agents. Between 80% and 90% of the preadipocytes finally differentiated to adipocytes as observed under phase contrast microscope. In addition, primary subcutaneous preadipocytes were isolated from adipose tissue of a 40-year-old healthy female with a body mass index (BMI) of 32.47. The preadipocytes were cultured and induced to differentiate into adipocytes by culturing them in differentiation medium containing active components PPAR- γ agonist and human insulin (Zen-Bio). Thiazolidinediones or PPAR- γ agonists can bind and activate an orphan nuclear receptor, PPAR- γ , and some of them have been proven to be able to induce human adipocyte differentiation. The preadipocytes were treated with human insulin and PPAR- γ agonist for 3 days and subsequently were switched to medium containing insulin alone for a variety of time periods ranging from one to 20 days before the cells were collected for analysis. Differentiated adipocytes were compared to untreated preadipocytes maintained in culture in the absence of inducing agents. Between 80% and 90% of the preadipocytes finally differentiated to adipocytes as

observed under phase contrast microscope. Expression of SEQ ID NO:29 and SEQ ID NO:30 was increased at least two-fold in differentiating adipocytes compared to untreated preadipocytes from the obese donor at 24 hours, 48 hours, 4 days, and 1.7 weeks. In contrast, expression of SEQ ID NO:29 and SEQ ID NO:30 was decreased at least two-fold in differentiating adipocytes compared to untreated preadipocytes from the normal donor at 2.1 weeks and did not show differential expression at earlier time points. Therefore, in various embodiments, SEQ ID NO:29 and SEQ ID NO:30 can be used for one or more of the following: i) monitoring treatment of diabetes mellitus, obesity, hypertension, and atherosclerosis, ii) diagnostic assays for diabetes mellitus, obesity, hypertension, and atherosclerosis, and iii) developing therapeutics and/or other treatments for diabetes mellitus, obesity, hypertension, and atherosclerosis.

In one example, SEQ ID NO:47 showed differential expression in breast cancer cell lines, as determined by microarray analysis. The gene expression profile of a nonmalignant mammary epithelial cell line (HMEC) was compared to the gene expression profiles of breast carcinoma lines at different stages of tumor progression. Cell lines compared included: a) BT-20, a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of tumor mass isolated from a 74-year-old female, b) BT-474, a breast ductal carcinoma cell line that was isolated from a solid, invasive ductal carcinoma of the breast obtained from a 60-year-old woman, c) BT-483, a breast ductal carcinoma cell line that was isolated from a papillary invasive ductal tumor obtained from a 23-year-old normal, menstruating, parous female with a family history of breast cancer, d) Hs 578T, a breast ductal carcinoma cell line isolated from a 74-year-old female with breast carcinoma, e) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, f) MCF-10A, a breast mammary gland (luminal ductal characteristics) cell line isolated from a 36-year-old woman with fibrocystic breast disease, g) MDA-MB-468, a breast adenocarcinoma cell line isolated from the pleural effusion of a 51-year-old female with metastatic adenocarcinoma of the breast, and h) Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female. When compared to starved HMEC cells, expression of SEQ ID NO:47 was increased at least 3.5-fold in MCF7 cells grown under optimal media conditions, and was increased at least 4-fold in BT-474 cells grown under optimal media conditions. Expression of SEQ ID NO:47 was also increased at least 2.5-fold in starved BT-474 cells, when compared to expression levels in starved HMEC cells. Expression levels of SEQ ID NO:47 also increased at least 6-fold in untreated Sk-BR-3 cells, when compared to the levels of gene expression in untreated HMEC cells. In addition, expression of SEQ ID NO:47 was examined in comparison to MCF-10A cells. The expression of SEQ ID NO:47 was increased at least 4.5-fold in Sk-BR-3 cells grown in two different types of defined serum-free media, when compared to the expression levels in MCF-10A cells grown under the same media conditions. Also, expression of SEQ ID NO:47 was increased at least 7.5-fold in

untreated Sk-BR-3 cells, when compared to untreated MCF-10A cells. Therefore, in various embodiments, SEQ ID NO:47 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

5 In another example, SEQ ID NO:47 showed differential expression in ovarian tumor tissue, as determined by microarray analysis. The expression levels of SEQ ID NO:47 were examined in tissue from a normal ovary from a 79 year-old female donor, and compared to expression levels in tissue from an ovarian tumor from the same donor (Huntsman Cancer Institute, Salt Lake City, UT). SEQ ID NO:47 expression was increased at least 3-fold in the tumor tissue, when compared to the
10 expression levels in the normal ovarian tissue. Therefore, in various embodiments, SEQ ID NO:47 can be used for one or more of the following: i) monitoring treatment of ovarian cancer, ii) diagnostic assays for ovarian cancer, and iii) developing therapeutics and/or other treatments for ovarian cancer.

XII. Complementary Polynucleotides

Sequences complementary to the EXMES-encoding sequences, or any parts thereof, are used
15 to detect, decrease, or inhibit expression of naturally occurring EXMES. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of EXMES. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5'
20 sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the EXMES-encoding transcript.

XIII. Expression of EXMES

Expression and purification of EXMES is achieved using bacterial or virus-based expression
25 systems. For expression of EXMES in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).
30 Antibiotic resistant bacteria express EXMES upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of EXMES in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding EXMES by either homologous recombination or bacterial-mediated
35 transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong

polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, EXMES is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from EXMES at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified EXMES obtained by these methods can be used directly in the assays shown in Examples XVII and XVIII, where applicable.

XIV. Functional Assays

EXMES function is assessed by expressing the sequences encoding EXMES at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; BD Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis

as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry,
5 Oxford, New York NY).

The influence of EXMES on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding EXMES and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected
10 cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding EXMES and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of EXMES Specific Antibodies

15 EXMES substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the EXMES amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is
20 synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-
25 Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-EXMES activity by, for example, binding the peptide or EXMES to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

30 XVI. Purification of Naturally Occurring EXMES Using Specific Antibodies

Naturally occurring or recombinant EXMES is substantially purified by immunoaffinity chromatography using antibodies specific for EXMES. An immunoaffinity column is constructed by covalently coupling anti-EXMES antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and
35 washed according to the manufacturer's instructions.

Media containing EXMES are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of EXMES (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/EXMES binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and EXMES is collected.

XVII. Identification of Molecules Which Interact with EXMES

EXMES, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled EXMES, washed, and any wells with labeled EXMES complex are assayed. Data obtained using different concentrations of EXMES are used to calculate values for the number, affinity, and association of EXMES with the candidate molecules.

Alternatively, molecules interacting with EXMES are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; *Nature* 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (BD Clontech).

EXMES may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of EXMES Activity

EXMES activity is measured by one of several methods. Growth factor activity is measured by the stimulation of DNA synthesis in Swiss mouse 3T3 cells. (McKay, I. and I. Leigh, eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York, NY.) Initiation of DNA synthesis indicates the cells' entry into the mitotic cycle and their commitment to undergo later division. 3T3 cells are competent to respond to most growth factors, not only those that are mitogenic, but also those that are involved in embryonic induction. This competence is possible because the *in vivo* specificity demonstrated by some growth factors is not necessarily inherent but is determined by the responding tissue. In this assay, varying amounts of EXMES are added to quiescent 3T3 cultured cells in the presence of [^3H]thymidine, a radioactive DNA precursor. EXMES for this assay can be obtained by recombinant means or from biochemical preparations. Incorporation of [^3H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold EXMES concentration range is indicative of growth factor activity. One unit of activity per milliliter is defined as the concentration of EXMES producing a 50% response level, where 100% represents maximal incorporation of [^3H]thymidine into acid-

precipitable DNA .

Alternatively, an assay for cytokine activity measures the proliferation of leukocytes. In this assay, the amount of tritiated thymidine incorporated into newly synthesized DNA is used to estimate proliferative activity. Varying amounts of EXMES are added to cultured leukocytes, such as
5 granulocytes, monocytes, or lymphocytes, in the presence of [³H]thymidine, a radioactive DNA precursor. EXMES for this assay can be obtained by recombinant means or from biochemical preparations. Incorporation of [³H]thymidine into acid-precipitable DNA is measured over an appropriate time interval, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold EXMES concentration
10 range is indicative of EXMES activity. One unit of activity per milliliter is conventionally defined as the concentration of EXMES producing a 50% response level, where 100% represents maximal incorporation of [³H]thymidine into acid-precipitable DNA.

An alternative assay for EXMES cytokine activity utilizes a Boyden micro chamber (Neuroprobe, Cabin John MD) to measure leukocyte chemotaxis (Vicari, A.P. et al. (1997) Immunity
15 7:291-301). In this assay, about 10⁵ migratory cells such as macrophages or monocytes are placed in cell culture media in the upper compartment of the chamber. Varying dilutions of EXMES are placed in the lower compartment. The two compartments are separated by a 5 or 8 micron pore polycarbonate filter (Nucleopore, Pleasanton CA). After incubation at 37°C for 80 to 120 minutes, the filters are fixed in methanol and stained with appropriate labeling agents. Cells which migrate to
20 the other side of the filter are counted using standard microscopy. The chemotactic index is calculated by dividing the number of migratory cells counted when EXMES is present in the lower compartment by the number of migratory cells counted when only media is present in the lower compartment. The chemotactic index is proportional to the activity of EXMES.

Alternatively, cell lines or tissues transformed with a vector encoding EXMES can be
25 assayed for EXMES activity by immunoblotting. Cells are denatured in SDS in the presence of b-mercaptoethanol, nucleic acids removed by ethanol precipitation, and proteins purified by acetone precipitation. Pellets are resuspended in 20 mM tris buffer at pH 7.5 and incubated with Protein G-Sepharose pre-coated with an antibody specific for EXMES. After washing, the Sepharose beads are boiled in electrophoresis sample buffer, and the eluted
30 proteins subjected to SDS-PAGE. The SDS-PAGE is transferred to a nitrocellulose membrane for immunoblotting, and the EXMES activity is assessed by visualizing and quantifying bands on the blot using the antibody specific for EXMES as the primary antibody and ¹²⁵I-labeled IgG specific for the primary antibody as the secondary antibody.

35 Various modifications and variations of the described compositions, methods, and systems of

the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions.

- 5 Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of
- 10 embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7513276	1	7513276CD1	26	7513276CB1	
7513290	2	7513290CD1	27	7513290CB1	
7513749	3	7513749CD1	28	7513749CB1	90096877CA2, 95119236CA2, 95119260CA2, 95119276CA2, 95119336CA2, 95146747CA2, 95165002CA2, 95165010CA2, 95165034CA2, 95165090CA2, 95165102CA2, 95165190CA2
7513017	4	7513017CD1	29	7513017CB1	
7513018	5	7513018CD1	30	7513018CB1	
7521129	6	7521129CD1	31	7521129CB1	95109330CA2
7521148	7	7521148CD1	32	7521148CB1	95128184CA2
7521157	8	7521157CD1	33	7521157CB1	
7521195	9	7521195CD1	34	7521195CB1	95139264CA2
7521277	10	7521277CD1	35	7521277CB1	95128004CA2
7515701	11	7515701CD1	36	7515701CB1	95039685CA2
7515527	12	7515527CD1	37	7515527CB1	90127316CA2, 90127356CA2, 90127416CA2, 90127432CA2
7519261	13	7519261CD1	38	7519261CB1	95059312CA2
7519923	14	7519923CD1	39	7519923CB1	95110205CA2
7519986	15	7519986CD1	40	7519986CB1	95090363CA2, 95090372CA2
7520799	16	7520799CD1	41	7520799CB1	95120450CA2
7520947	17	7520947CD1	42	7520947CB1	95115016CA2, 95115024CA2, 95115124CA2
7521071	18	7521071CD1	43	7521071CB1	95139288CA2, 95155718CA2
7521082	19	7521082CD1	44	7521082CB1	95114835CA2
7521329	20	7521329CD1	45	7521329CB1	95127905CA2
7522977	21	7522977CD1	46	7522977CB1	95162935CA2, 95163051CA2
7521939	22	7521939CD1	47	7521939CB1	95130211CA2
7522030	23	7522030CD1	48	7522030CB1	95146921CA2
7522059	24	7522059CD1	49	7522059CB1	95139339CA2
7523442	25	7523442CD1	50	7523442CB1	95172947CA2

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7513276CD1	g1185010	1.1E-77	[Homo sapiens] pre-pro-orphanin FQ
				Nothacker, H. P. et al., Primary structure and tissue distribution of the orphanin FQ precursor, Proc. Natl. Acad. Sci. U.S.A. 93, 8677-8682 (1996)
		343672 PNOC	8.3E-79	[Homo sapiens][Ligand] Prepronociceptin, precursor of the neuropeptide hormones nociceptin and nocistatin; nociceptin binds the receptor OPR1.1 and positively regulates pain response and nocistatin also binds an unknown component and negatively regulates pain response
				Okuda-Ashitaka, E. et al., Nocistatin, a peptide that blocks nociceptin action in pain transmission Nature 392, 286-9 (1998).
				Mouldous, L. et al., Direct identification of a peptide binding region in the opioid receptor-like 1 receptor by photoaffinity labeling with [Bpa10, Tyr14]Nociceptin, J Biol Chem 275, 29268-74 (2000).
		580013 Phoc	5.4E-59	[Mus musculus][Ligand][Extracellular (excluding cell wall)] Prepronociceptin, precursor of the neuropeptide hormones nociceptin and nocistatin; nociceptin binds the receptor Opr1 and positively regulates pain response and nocistatin also binds an unknown component and negatively regulates pain
				Reinscheid, R. K. et al., Orphanin FQ: a neuropeptide that activates an opioidlike G protein-coupled receptor, Science 270, 792-4 (1995).
				Koster, A. et al., Targeted disruption of the orphanin FQ/nociceptin gene increases stress susceptibility and impairs stress adaptation in mice, Proc Natl Acad Sci U S A 96, 10444-9 (1999).
2	7513290CD1	g2197069	1.6E-294	[Homo sapiens] Delta
				Gray, G. E. et al., Human ligands of the Notch receptor, Am. J. Pathol. 154, 785-794 (1999)
		747879 D111	1.3E-291	[Rattus norvegicus][Ligand][Plasma membrane] Reticulon 1, member of the reticulon family and a putative Notch ligand, which is involved in somite development and neuronal differentiation; expression of the human protein is associated with neuroendocrine cells in lung tumors
				GrandPre, T. et al., Identification of the Nogo inhibitor of axon regeneration as a Reticulon protein, Nature 403, 439-44 (2000).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Bao, Z. Z. et al., The expression and function of Notch pathway genes in the developing rat eye, J Neurosci 17, 1425-34 (1997).
		605950 DLL4	1.1E-170	[Homo sapiens][Ligand][Plasma membrane] Delta-like ligand 4, a member of the Delta family of Notch ligands, a membrane-bound ligand with EGF-like domains and a DSL domain, may regulate vascular development, vascular endothelial processes, and hematopoiesis
				Shutter, J. R. et al., Dll4, a novel notch ligand expressed in arterial endothelium, Genes And Development 14, 1313-8 (2000).
				Karanu, F. N. et al., Human homologues of Delta-1 and Delta-4 function as mitogenic regulators of primitive human hematopoietic cells, Blood 97, 1960-7 (2001).
3	7513749CD1	g17062032	8.9E-58	[Homo sapiens] WNT10B
				Saitoh, T. et al., Proto-oncogene WNT10B is up-regulated by tumor necrosis factor alpha in human gastric cancer cell line MKN45, Int. J. Oncol. 19, 1187-1192 (2001)
		758844 WNT10B	6.9E-59	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Wingless-type MMTV integration site family member 10b, member of the Wnt family of cell signaling proteins that plays a role in hematopoiesis and perhaps adipogenesis, may also contribute to the tumorigenesis of breast carcinomas
				Schroeder, J. A. et al., Cooperative induction of mammary tumorigenesis by TGFalpha and Wnts, Oncogene 19, 3193-9 (2000).
				Ross, S. E. et al., Inhibition of adipogenesis by Wnt signaling, Science 289, 950-3 (2000).
				Van Den Berg, D. J. et al., Role of members of the Wnt gene family in human hematopoiesis, Blood 92, 3189-202. (1998).
		586151 Wnt10b	1.0E-55	[Mus musculus][Ligand][Extracellular (excluding cell wall)] Wingless-type MMTV integration site family member 10b, member of the Wnt family of cell signaling proteins and a putative physiological ligand for Fzd5, involved in tooth development, mammary gland development, and malignancy, regulates sexual dimorphism
				Lane, T. F. et al., Wnt-10b directs hypermorphic development and transformation in mammary glands of male and female mice, Oncogene 15, 2133-44 (1997).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Hardiman, G. et al., The mouse Wnt-10B gene isolated from helper T cells is widely expressed and a possible oncogene in BR6 mouse mammary tumorigenesis, Gene 172, 199-205 (1996).
4	7513017CD1	g1572721	0.0E+00	[Homo sapiens] megakaryocyte stimulating factor; MSF
				Turner, K. J. et al., Purification, Biochemical Characterization, and Cloning of a Novel Megakaryocyte Stimulating Factor that has Megakaryocyte Colony Stimulating Activity, Blood 78 (Suppl. 1), 279 (1991)
				Merberg, D. M. et al., A Comparison of Vitronectin and Megakaryocyte Stimulating Factor (in) Preissner, K.T., Rosenblatt, S., Kost, C., Wegerhoff, J. and Mosher, D.F. (Eds.); BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS: 45-52; Elsevier Science Publishers B.V. (1993)
		342642 PRG4	0.0	[Homo sapiens][Structural protein][Extracellular (excluding cell wall)] Proteoglycan 4 (megakaryocyte stimulating factor), a secreted proteoglycan with possible roles in regulating cell proliferation, ossification and joint lubrication; mutations in the gene cause camptodactyly-arthropathy-coxa vara-pericarditis syndrome
				Marcelino, J. et al., CACP, encoding a secreted proteoglycan, is mutated in camptodactyly-arthropathy-coxa vara-pericarditis syndrome, Nat Genet 23, 319-22 (1999).
				Flannery, C. R. et al., Articular cartilage superficial zone protein (SZP) is homologous to megakaryocyte stimulating factor precursor and is a multifunctional proteoglycan with potential growth-promoting, cytoprotective, and lubricating properties in cartilage metabolism, Biochem Biophys Res Commun 254, 535-41 (1999).
		618726 Prg4	7.3E-261	[Mus musculus] Protein with high similarity to proteoglycan 4 (megakaryocyte stimulating factor, human PRG4), which is associated with camptodactyly-arthropathy-coxa vara-pericarditis syndrome, contains two hemopexin domains and two somatomedin B domains
				Ikegawa, S. et al., Isolation, characterization and mapping of the mouse and human PRG4 (proteoglycan 4) genes, Cytogenet Cell Genet 90, 291-297 (2000)
5	7513018CD1	g1572721	0.0	[Homo sapiens] megakaryocyte stimulating factor; MSF
				Turner, K. J. et al. (supra)
				Merberg, D. M. et al. (supra)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		342642 PRG4	0.0	[Homo sapiens][Structural protein][Extracellular (excluding cell wall)] Proteoglycan 4 (megakaryocyte stimulating factor), a secreted proteoglycan with possible roles in regulating cell proliferation, ossification and joint lubrication; mutations in the gene cause camptodactyly-arthropathy-coxa vara-pericarditis syndrome
				Marcelino, J. et al., CACP, encoding a secreted proteoglycan, is mutated in camptodactyly-arthropathy-coxa vara-pericarditis syndrome, Nat Genet 23, 319-22 (1999).
				Flannery, C. R. et al., Articular cartilage superficial zone protein (SZP) is homologous to megakaryocyte stimulating factor precursor and is a multifunctional proteoglycan with potential growth-promoting, cytoprotective, and lubricating properties in cartilage metabolism, Biochem Biophys Res Commun 254, 535-41 (1999).
		618726 PrG4	7.3E-261	[Mus musculus] Protein with high similarity to proteoglycan 4 (megakaryocyte stimulating factor, human PRG4), which is associated with camptodactyly-arthropathy-coxa vara-pericarditis syndrome, contains two hemopexin domains and two somatomedin B domains
				Ikegawa, S. et al., Isolation, characterization and mapping of the mouse and human PRG4 (proteoglycan 4) genes, Cytogenet Cell Genet 90, 291-297 (2000).
6	7521129CD1	g1212989	6.7E-185	[Homo sapiens] transforming growth factor beta 1 precursor
				Derynck, R. et al. Intron-exon structure of the human transforming growth factor-beta precursor gene, Nucleic Acids Res. 15, 3188-3189 (1987)
		618424 TGFB1	1.3E-183	[Homo sapiens][Ligand;Activator;DNA-binding protein;Transcription factor] Transforming growth factor-beta 1, an anti-inflammatory cytokine that regulates cell proliferation, differentiation, and apoptosis, associated with tissue fibrosis, allograft rejection, and atopic dermatitis
				Puente, E. et al. Transcriptional Activation of Mouse sst2 Somatostatin Receptor Promoter by Transforming Growth Factor-beta. INVOLVEMENT OF Smad4. J Biol Chem 276, 13461-8. (2001).
		619046 Tgfb1	6.7E-166	[Rattus norvegicus][Ligand] Transforming growth factor-beta 1, cytokine that regulates cell proliferation, differentiation, and apoptosis; human TGFB1 is associated with tissue fibrosis, allograft rejection, and atopic dermatitis

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Ziyadeh, F. N. et al. Long-term prevention of renal insufficiency, excess matrix gene expression, and glomerular mesangial matrix expansion by treatment with monoclonal antitransforming growth factor-beta antibody in db/db diabetic mice, Proc Natl Acad Sci U S A 97, 8015-20 (2000).
7	7521148CD1	g4512044	5.8E-51	[Homo sapiens] angiopoietin 4
				Valenzuela, D. M. et al. Angiopoietins 3 and 4: diverging gene counterparts in mice and humans, Proc. Natl. Acad. Sci. U.S.A. 96, 1904-1909 (1999)
		475297 ANGPT4	4.3E-52	[Homo sapiens][Ligand] Angiopoietin-3, a member of the angiopoietin family of growth factors that are largely specific for the vascular epithelium, and expression in human umbilical vein endothelial cells is decreased by VEGF; may play a role in angiogenesis
				Brown, L. F. et al. Expression of Tie1, Tie2, and angiopoietins 1, 2, and 4 in Kaposi's sarcoma and cutaneous angiosarcoma. Am J Pathol 156, 2179-83. (2000).
		429624 Agpt4	3.0E-13	[Mus musculus][Ligand] Angiopoietin 4, member of the angiopoietin family of growth factors which is specific to the vascular endothelium, functions as a Tie2 receptor (Tek) antagonist
				Kersten, S. et al. Characterization of the fasting-induced adipose factor FIAF, a novel peroxisome proliferator-activated receptor target gene. J Biol Chem 275, 28488-93 (2000).
8	7521157CD1	g642835	6.9E-16	[Homo sapiens] LERK-4
				Kozlosky, C. J. et al. Ligands for the receptor tyrosine kinases hek and elk: isolation of cDNAs encoding a family of proteins, Oncogene 10, 299-306 (1995)
		342012 EFNA4	5.1E-17	[Homo sapiens][Protein kinase;Ligand;Transferase;Receptor (signalling)][Plasma membrane] Ephrin-A4, GPI-linked plasma membrane-associated ligand of the Eph-related receptor tyrosine kinases that may mediate signal transduction; alternative secreted form made by activated tonsillar B cells may mediate communication with dendritic cells
				Aasheim, H. C. et al. A splice variant of human ephrin-A4 encodes a soluble molecule that is secreted by activated human B lymphocytes. Blood 95, 221-30 (2000).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		584077 Efna4	2.9E-14	[Mus musculus][Ligand][Plasma membrane] Ephrin-A4, putative GPI-linked plasma membrane-associated ligand of the Eph-related receptor tyrosine kinases, may have roles in signal transduction, embryogenesis and morphogenesis, brain development, axonogenesis, and cell motility
				Knoll, B. et al. A role for the EphA family in the topographic targeting of vomeronasal axons. Dev Suppl 128, 895-906. (2001).
9	7521195CD1	g1741868	1.0E-110	[Homo sapiens] SDF2
				Hamada, T. et al. Isolation and characterization of a novel secretory protein, stromal cell-derived factor-2 (SDF-2) using the signal sequence trap method, Gene 176, 211-214 (1996)
		742542 SDF2	9.6E-114	[Homo sapiens][Transferase][Extracellular (excluding cell wall)] Stromal cell-derived factor 2, a putative secreted protein with similarity to S. cerevisiae Pmt1p and Pmt2p, which are dolichyl phosphate-D-mannose:protein mannosyltransferases
				Fukuda, S. et al. Murine and Human SDF2L1 Is an Endoplasmic Reticulum Stress-Inducible Gene and Encodes a New Member of the Pmt/rt Protein Family. Biochem Biophys Res Commun 280, 407-414. (2001).
		581301 Sdf2	3.0E-108	[Mus musculus][Extracellular (excluding cell wall)] Stromal cell-derived factor 2, a putative secreted protein with similarity to S. cerevisiae Pmt1p and Pmt2p, which are dolichyl phosphate-D-mannose:protein mannosyltransferases
10	7521277CD1	g4884555	5.0E-132	[Homo sapiens] angiopoietin-3
				Nishimura, M. et al. Angiopoietin-3, a novel member of the angiopoietin family, FEBS Lett. 448, 254-256 (1999)
		475297 ANGPT4	3.7E-133	[Homo sapiens][Ligand] Angiopoietin-3, a member of the angiopoietin family of growth factors that are largely specific for the vascular epithelium, and expression in human umbilical vein endothelial cells is decreased by VEGF; may play a role in angiogenesis
				Brown, L. F. et al. (supra)
		429624 Agpt4	1.4E-69	[Mus musculus][Ligand] Angiopoietin 4, member of the angiopoietin family of growth factors which is specific to the vascular endothelium, functions as a Tie2 receptor (Tek) antagonist
				Kersten, S. et al. (supra)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
11	7515701CD1	g4028583	3.4E-28	[Homo sapiens] connective tissue growth factor related protein WISP-2 Pennica, D. et al., WISP genes are members of the connective tissue growth factor family that are up-regulated in wnt-1-transformed cells and aberrantly expressed in human colon tumors, Proc. Natl. Acad. Sci. U.S.A. 95, 14717-14722 (1998)
		338900 WISP2	2.6E-29	[Homo sapiens][Ligand] WNT1-inducible signaling pathway protein 2, a member of the connective tissue growth factor family, may play a role in regulating bone formation; aberrant transcription is associated with colon tumors and adenocarcinoma
				Kumar, S. et al., Identification and cloning of a connective tissue growth factor-like cDNA from human osteoblasts encoding a novel regulator of osteoblast functions., J Biol Chem 274, 17123-31 (1999).
		587855 Wisp2	3.0E-21	[Mus musculus][Ligand] Wnt1-inducible signaling pathway protein 2, a member of the connective tissue growth factor family; aberrant transcription of human WISP2 is associated with colon tumors and adenocarcinoma
				Kumar, S. et al., (supra)
12	7515527CD1	g4761612	1.5E-142	[Homo sapiens] TNF and Apol. related ligand TALL-1 Shu, H. B. et al., TALL-1 is a novel member of the TNF family that is down-regulated by mitogens, J. Leukoc. Biol. 65, 680-683 (1999)
		432444 TNFSF13B	1.2E-143	[Homo sapiens][Ligand][Plasma membrane] Tumor necrosis factor (ligand) superfamily member 13b, TNF superfamily ligand involved in B-cell proliferation, immunoglobulin secretion, apoptosis, and NFk-B and JNK signaling pathways; may play a role in Sjogren's syndrome and systemic lupus erythematosus

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Moore, P. A. et al., BLYS: member of the tumor necrosis factor family and B lymphocyte stimulator., Science 285, 260-3 (1999).
		749940 Tnfrsf13b	4.8E-87	[Mus musculus] Tumor necrosis factor superfamily member 13b, a ligand of Baffr, BCMA (Tnfrsf17), TACI (Tnfrsf13b) and BR3 receptors, stimulates B cell survival and proliferation and may mediate B cell-associated autoimmune lupus-like diseases
				Schiemann, B. et al., An essential role for BAFF in the normal development of B cells through a BCMA-independent pathway., Science 293, 2111-4. (2001).
13	7519261CD1	g33971	1.8E-11	[Homo sapiens] Irp protein (AA 1-360) Wainwright, B. J. et al., Isolation of a human gene with protein sequence similarity to human and murine int-1 and the Drosophila segment polarity mutant wingless, EMBO J. 7, 1743-1748 (1988)
		338904 WNT2	1.3E-12	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Wingless-type MMTV integration site family member 2, member of the Wnt family of cell signaling proteins, involved in the regulation of development; overexpressed in a variety of cancer types, may contribute to the pathogenesis of autism
				Wainwright, B. J. et al., (supra)
		746577 Wnt2	1.3E-11	[Mus musculus] Wingless-type MMTV integration site family member 2, member of the Wnt family of cell signaling proteins, involved fetal placental vascularization; human WNT2 is overexpressed in a variety of cancers and may contribute to the pathogenesis of autism
				Monkley, S. J. et al., Targeted disruption of the Wnt2 gene results in placental defects., Dev Suppl 122, 3343-53 (1996).
14	7519923CD1	g1841553	6.7E-80	[Homo sapiens] NG3

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		697430 NG3	5.1E-81	[Homo sapiens] Protein containing two epidermal growth factor (EGF)-like domains, has a region of moderate similarity to a region of fibrillin 2 (human FBN2), which is an extracellular matrix protein that regulates formation of connective tissue microfibrils
		368734 AW047140	3.5E-66	[Mus musculus] Protein containing two epidermal growth factor (EGF)-like domains, has a region of moderate similarity to a region of notch Drosophila homolog 3 (mouse Notch3), which functions in cell fate determination
15	7519986CD1	g3650492	1.2E-43	[Homo sapiens] proliferation inducing ligand APRIL
				Hahne, M. et al., APRIL, a new ligand of the tumor necrosis factor family, stimulates tumor cell growth, J. Exp. Med. 188, 1185-1190 (1998)
		338608 TNFSF13	8.8E-45	[Homo sapiens][Ligand] Tumor necrosis factor (ligand) superfamily member 13, binds to TNF receptors, Fas (TNFRSF6), and HVEM (TNFRSF14), stimulates apoptosis, stimulates tumor cell proliferation; overexpressed in lymphoid, thyroid, colon, and gastrointestinal tumors
				Hahne, M. et al., (supra)
16	7520799CD1	g340280	2.2E-42	[Homo sapiens] preprovasoactive intestinal peptide
				Itoh, N. et al., Human preprovasoactive intestinal polypeptide contains a novel PHL-27-like peptide, PHM-27, Nature 304, 547-549 (1983)
		338878 VIP	1.7E-43	[Homo sapiens][Ligand] Vasoactive intestinal peptide precursor, produces a neuropeptide hormone that signals via G protein-coupled receptors and may modulate smooth muscle function, and electrolyte balance, cell proliferation and survival, and inflammatory responses

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Heinz-Erian, P. et al., Deficient vasoactive intestinal peptide innervation in the sweat glands of cystic fibrosis patients., Science 229, 1407-8 (1985).
		625065 Vip	6.7E-31	[Rattus norvegicus][Ligand] Vasoactive intestinal peptide precursor, produces a neuropeptide hormone that signals through G protein-coupled receptors and may modulate cell proliferation, neuronal cell survival, inflammatory and oxidative stress responses, and smooth muscle function
				Ishihara, T. et al., Functional expression and tissue distribution of a novel receptor for vasoactive intestinal polypeptide., Neuron 8, 811-9 (1992).
17	7520947CD1	g14327972	4.1E-34	[Homo sapiens] Similar to mesoderm development candidate 2
		758665 MESDC2	3.0E-35	[Homo sapiens] Mesoderm development candidate 2, a putative integral plasma membrane protein; mouse Mesdc2 gene lies within a chromosomal region critical for the differentiation of mesoderm
				Wines, M. E. et al., Identification of mesoderm development (mesd) candidate genes by comparative mapping and genome sequence analysis., Genomics 72, 88-98. (2001).
		685569 Mesdc2	2.7E-20	[Mus musculus] Mesoderm development candidate 2, a widely expressed protein; gene lies within a chromosomal region critical for the differentiation of mesoderm
				Wines, M. E. et al., (supra)
18	7521071CD1	g1741868	1.3E-89	[Homo sapiens] SDF2
				Hamada, T. et al., Isolation and characterization of a novel secretory protein, stromal cell-derived factor-2 (SDF-2) using the signal sequence trap method, Gene 176, 211-214 (1996)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		742542 SDF2	1.6E-92	[Homo sapiens][Transferase][Extracellular (excluding cell wall)] Stromal cell-derived factor 2, a putative secreted protein with similarity to S. cerevisiae Pmt1p and Pmt2p, which are dolichyl phosphate-D-mannose:protein mannosyltransferases
				Hamada, T. et al. (supra)
		581301 Sdf2	2.6E-88	[Mus musculus][Extracellular (excluding cell wall)] Stromal cell-derived factor 2, a putative secreted protein with similarity to S. cerevisiae Pmt1p and Pmt2p, which are dolichyl phosphate-D-mannose:protein mannosyltransferases
				Hamada, T. et al., (supra)
19	7521082CD1	g183213	1.4E-41	[Homo sapiens] gastric inhibitory polypeptide
				Takeda, J. et al., Sequence of an intestinal cDNA encoding human gastric inhibitory polypeptide precursor, Proc. Natl. Acad. Sci. U.S.A. 84, 7005-7008 (1987)
		340618 GIP	1.0E-42	[Homo sapiens][Ligand] Gastric inhibitory polypeptide (glucose-dependent insulinotropic polypeptide), a peptide hormone involved in the nutritional response pathway, stimulates insulin secretion in a glucose dependent manner
				Takeda, J. et al., (supra)
		609983 Gludins	3.8E-26	[Rattus norvegicus][Ligand] Gastric inhibitory polypeptide (glucose-dependent insulinotropic polypeptide), a putative peptide hormone involved in the nutritional response pathway
				Sharma, S. K. et al., Characterization of rat gastric inhibitory peptide cDNA., J Mol Endocrinol 9, 265-72 (1992).
20	7521329CD1	g4884555	4.9E-182	[Homo sapiens] angiotensinogen-3
				Nishimura, M. et al. (1999) Angiotensinogen-3, a novel member of the angiotensin family. FEBS Lett. 448:254-256.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		475297 ANGPT4	3.6E-183	[Homo sapiens] [Ligand] Angiopoietin-3, a member of the angiopoietin family of growth factors that are largely specific for the vascular epithelium, and expression in human umbilical vein endothelial cells is decreased by VEGF; may play a role in angiogenesis, may be associated with the pathogenesis of Kaposi's sarcoma
				Kim, I. et al. (1999) Molecular cloning and characterization of a novel angiopoietin family protein, angiopoietin-3. FEBS Lett 443:353-356.
				Brown, L.F. et al. (2000) Expression of Tie1, Tie2, and angiopoietins 1, 2, and 4 in Kaposi's sarcoma and cutaneous angiosarcoma. Am. J. Pathol. 156:2179-2183.
		429624 Agpt4	5.7E-105	[Mus musculus] [Ligand] Angiopoietin 4, member of the angiopoietin family of growth factors which is specific to the vascular endothelium, functions as a Tie2 receptor (Tek) antagonist
				Valenzuela, D.M. et al. (1999) Angiopoietins 3 and 4: diverging gene counterparts in mice and humans. Proc. Natl. Acad. Sci. USA 96:1904-1909.
21	7522977CD1	g189411	6.1E-17	[Homo sapiens] oxytocin/neurophysin
				Rehbein, M. et al. (1986) The neurohypophyseal hormones vasopressin and oxytocin. Precursor structure, synthesis and regulation. Biol. Chem. Hoppe-Seyler 367:695-704.
		336720 OXT	4.5E-18	[Homo sapiens] [Ligand] Preprooxytocin-neurophysin I, precursor of the hormone oxytocin and its carrier protein neurophysin I, controls parturition and lactation
				Rao, V.V. et al. (1992) The human gene for oxytocin-neurophysin I (OXT) is physically mapped to chromosome 20p13 by in situ hybridization. Cytogenet. Cell Genet. 61:271-273.
				Copland, J.A. et al. (1999) Demonstration of functional oxytocin receptors in human breast Hs578T cells and their up-regulation through a protein kinase C-dependent pathway. Endocrinology 140:2258-2267.
		590435 Oxt	9.6E-16	[Rattus norvegicus] [Ligand] [Axon;Cell body (soma)] Preprooxytocin-neurophysin I, precursor of the hormone oxytocin and its carrier protein neurophysin I, involved in parturition and lactation, modulates a range of social behavior
				Ivell, R. et al. (1984) Structure and comparison of the oxytocin and vasopressin genes from rat. Proc. Natl. Acad. Sci. USA 81:2006-2010.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
22	7521939CD1	g339564	3.7E-199	[Homo sapiens] transforming growth factor-beta Celeste, A.J. et al. (1990) Identification of transforming growth factor beta family members present in bone-inductive protein purified from bovine bone. Proc. Natl. Acad. Sci. USA 87:9843-9847.
		339852 BMP7	2.7E-200	[Homo sapiens] Bone morphogenetic protein 7, a regulator of cell proliferation, differentiation, and apoptosis that induces bone formation and may play a role in skeletal and nervous system development, prevents loss of kidney function after acute renal failure
				Hahn, G.V. et al. (1992) A bone morphogenetic protein subfamily: chromosomal localization of human genes for BMP5, BMP6, and BMP7. Genomics 14:759-762.
				Hatakeyama, S. et al. (1997) Expression of bone morphogenetic proteins of human neoplastic epithelial cells. Biochem. Mol. Biol. Int. 42:497-505.
		583681 Bmp7	5.2E-195	[Mus musculus] Bone morphogenetic protein 7, a regulator of cell proliferation, differentiation, motility, and apoptosis that plays a role in the development of the skeleton, kidney, brain, eye, heart, and epididymis
				Wawersik, S. et al. (1999) BMP7 acts in murine lens placode development. Dev. Biol. 207:176-188.
				Dudley, A.T. et al. (1997) Overlapping expression domains of bone morphogenetic protein family members potentially account for limited tissue defects in BMP7 deficient embryos. Dev. Dyn. 208:349-362.
23	7522030CD1	g2204347	3.5E-37	[Homo sapiens] manic fringe precursor
				Johnston, S.H. et al. (1997) A family of mammalian Fringe genes implicated in boundary determination and the Notch pathway. Development 124:2245-2254.
		336396 MFNG	2.6E-38	[Homo sapiens] [Extracellular (excluding cell wall)] Manic fringe, member of the Fringe family of proteins, a secreted protein predicted to be in the Notch signaling pathway, may be involved in boundary determination during segmentation
				Cohen, B. et al. (1997) Fringe boundaries coincide with Notch-dependent patterning centres in mammals and alter Notch-dependent development in Drosophila. Nat. Genet. 16:283-288.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		582081 Mfng	4.1E-24	[Mus musculus] [Transferase] [Golgi; Cytoplasmic] Manic fringe, member of the Fringe family of proteins, a secreted glycosyltransferase in the Notch signaling pathway, may be involved in neurogenesis and boundary determination during segmentation
				Moloney, D.J. et al. (2000) Fringe is a glycosyltransferase that modifies Notch. Nature 406:369-375.
				Hicks, C. et al. (2000) Fringe differentially modulates Jagged1 and Delta1 signalling through Notch1 and Notch2. Nat. Cell Biol. 2:515-520.
				Ishii, Y. et al. (2000) Demarcation of early mammalian cortical development by differential expression of fringe genes. Brain Res. Dev. Brain Res. 119:307-320.
				Shimizu, K. et al. (2001) Manic fringe and lunatic fringe modify different sites of the Notch2 extracellular region, resulting in different signaling modulation. J. Biol. Chem. 276:25753-25758.
24	7522059CD1	g3335144	2.3E-90	[Homo sapiens] stanniocalcin-2
				Chang, A.C. et al. (1998) Identification of a second stanniocalcin cDNA in mouse and human: stanniocalcin 2. Mol. Cell. Endocrinol. 141:95-99.
		338312 STC2	1.6E-91	[Homo sapiens] [Ligand] Stanniocalcin 2, a widely expressed putative hormone, may regulate metal ion homeostasis, downregulates expression of a sodium-phosphate transporter and inhibits phosphate uptake
				Ishibashi, K. et al. (1998) Molecular cloning of a second human stanniocalcin homologue (STC2). Biochem. Biophys. Res. Commun. 250:252-258.
				Moore, E.E. et al. (1999) Stanniocalcin 2: characterization of the protein and its localization to human pancreatic alpha cells. Horm. Metab. Res. 31:406-414.
				Jelinek, D.A. et al. (2000) Stanniocalcin 1 and 2 are secreted as phosphoproteins from human fibrosarcoma cells. Biochem. J. 350:453-461.
		585935 Stc2	5.3E-86	[Mus musculus] [Ligand] Stanniocalcin 2, a putative hormone that may regulate metal ion homeostasis, downregulated in hypophosphatemic mice, has strong similarity to human STC2, which downregulates expression of a sodium-phosphate transporter and inhibits phosphate uptake
25	7523442CD1	g183867	1.7E-67	[Homo sapiens] heparin-binding EGF-like growth factor

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Higashiyama, S. et al. (1991) A heparin-binding growth factor secreted by macrophage-like cells that is related to EGF. Science 251:936-939.

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7513276CD1	177	signal_cleavage: M1-C21	SPSCAN
			Signal Peptide: M1-S16	HMMER
			Signal Peptide: M1-S19	HMMER
			Signal Peptide: M1-S20	HMMER
			Signal Peptide: M1-Q22	HMMER
			Signal Peptide: M1-C21	HMMER
			Signal Peptide: M1-R23	HMMER
			Vertebrate endogenous opioids neuropeptide: M1-I169	HMMER_PFAM
			Endogenous opioids neuropeptides precursors IPB000094: L13-L32, L36-T62, K128-N145, Q127-A136	BLIMPS_BLOCKS
			Opioid neuropeptide precursor signature PR01028: S20-K31, D40-C48, C48-C61, Q125-G135	BLIMPS_PRINTS
			Nociceptin precursor signature	BLIMPS_PRINTS
			PR01031: K2-C21, D24-F39, L41-T59, C61-E78, A82-S103, M115-A136, A136-I154	
			PRECURSOR OPIOID PEPTIDE NEUROPEPTIDE CLEAVAGE ON PAIR OF BASIC RESIDUES	BLAST_PRODROM
			PD005389: L10-G132 Q127-F133	
			Endogenous opioids neuropeptides precursors signature: C21-C61	MOTIFS
			Potential Phosphorylation Sites: S139 S150 S174 T27 T164	MOTIFS
2	7513290CD1	524	signal_cleavage: M1-S21	SPSCAN
			Signal Peptide: S3-S21	HMMER
			Signal Peptide: M1-S21	HMMER
			Signal Peptide: M1-G23	HMMER
			Signal Peptide: M1-V24	HMMER
			Signal Peptide: M1-E26	HMMER
			Signal Peptide: M1-S22	HMMER
			delta serrate ligand: W159-C221	HMMER_SMART
			Epidermal growth factor-like domain: E331-E364, D446-S479, Y291-E326, T369-E403, Y408-D441, E256-N286, I225-D255	HMMER_SMART
			Calcium-binding EGF-like domain: N443-S479, D288-E326, G328-E364, K405-D441, C370-E403	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Delta serrate ligand: W159-C221	HMMER_PFAM
			EGF-like domain: C409-C440, C447-C478, C332-C363, C292-C325, C370-C402, C257-C285, C226-C254, C179-C221	HMMER_PFAM
			C-terminal cysteine knot IPB000359: G310-A323	BLIMPS_BLOCKS
			EGF-like domain IPB000561: C467-G475	BLIMPS_BLOCKS
			Calcium-binding EGF-like domain IPB001881: M368-G378, C343-C354	BLIMPS_BLOCKS
			Type I EGF signature PR00009: D386-Y397	BLIMPS_PRINTS
			Type II EGF-like signature PR00010: G328-N339, N348-F358, S398-K404	BLIMPS_PRINTS
			Type III EGF-like signature PR00011: D460-C478, T307-C325, T307-C325	BLIMPS_PRINTS
			GLYCOPROTEIN EGF-LIKE DOMAIN TRANSMEMBRANE PROTEIN PRECURSOR SIGNAL	BLAST_PRODROM
			DELTA-LIKE DELTA1 C DELTA1 PD008515: R476-V524	BLAST_PRODROM
			GLYCOPROTEIN EGF-LIKE DOMAIN TRANSMEMBRANE PROTEIN PRECURSOR SIGNAL	BLAST_PRODROM
			DELTA-LIKE DELTA1 DIFFERENTIATION PD149899: C257-F284	BLAST_PRODROM
			GLYCOPROTEIN EGF-LIKE DOMAIN TRANSMEMBRANE PROTEIN PRECURSOR SIGNAL	BLAST_PRODROM
			DELTA-LIKE JAGGED DELTA1 PD003890: C5-P117 PD149730: T119-C212	BLAST_DOMO
			do EGF; DELTA; CYSTEINE; JAGGED;	BLAST_DOMO
			DM03706 48324 48-162: G48-H164	BLAST_DOMO
			DM03706 P10041 54-175: A55-L171	BLAST_DOMO
			EGF	BLAST_DOMO
			DM00003 48324 235-285: F236-Q287	BLAST_DOMO
			DM00003 P10041 275-331: W272-E326	BLAST_DOMO
			Aspartic acid and asparagine hydroxylation site: C343-C354, C420-C431, C458-C469	MOTIFS
			EGF-like domain signature 1: C243-C254, C274-C285, C314-C325, C352-C363, C391-C402, C429-C440, C467-C478	MOTIFS
			EGF-like domain signature 2: C243-C254, C274-C285, C314-C325, C352-C363, C391-C402, C429-C440, C467-C478	MOTIFS
			Calcium-binding EGF-like domain pattern signature: N443-C467	MOTIFS
			Potential Phosphorylation Sites: S93 S131 S166 S174 S315 S342 S390 S436 S483 S488 T149 T154 T203 T344 T369 T457 T474 T499 Y183 Y260	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	7513749CD1	143	Potential Glycosylation Sites: N477	MOTIFS
			signal_cleavage: M1-S28	SPSCAN
			Signal Peptide: P7-S28	HMIMER
			Signal Peptide: M1-E30	HMIMER
			Signal Peptide: M1-S28	HMIMER
			Signal Peptide: P5-S28	HMIMER
4	7513017CD1	1270	Developmental signaling protein, Wnt-1 family IPB000970: C48-L63, G75-C94	BLIMPS_BLOCKS
			PROTEIN DEVELOPMENTAL GLYCOPROTEIN PRECURSOR SIGNAL WNT1 WNT5A	BLAST_PRODROM
			WNT2 EXTRACELLULAR MATRIX	
			PD000810: V47-G113	
			WNT-1 FAMILY	BLAST_DOMO
			DM00403 49263 30-388: E30-G113	
			DM00403 P48614 30-388: E30-G113	
			DM00403 P43446 80-441: E30-G113	
			Potential Phosphorylation Sites: S55 S95 S126	MOTIFS
			Potential Glycosylation Sites: N93	MOTIFS
4	7513017CD1	1270	signal_cleavage: M1-S24	SPSCAN
			Signal Peptide: M1-V18	HMIMER
			Signal Peptide: M1-Q21	HMIMER
			Signal Peptide: M1-S24	HMIMER
			Signal Peptide: M1-C29	HMIMER
			Signal Peptide: M1-L27	HMIMER
			Signal Peptide: M1-Q20	HMIMER
			Hemopexin-like repeats: I1061-G1104, V1017-S1059	HMIMER_SMART
			Somatomedin B -like domains: Q25-K67	HMIMER_SMART
			Somatomedin B domain: Q25-D68	HMIMER_PFAM
			Hemopexin: I1061-G1107, V1017-S1059	HMIMER_PFAM
			Somatomedin B domain IPB001212: C43-C63, P1058-F1099, Q1108-Q1145, D1232-P1270	BLIMPS_BLOCKS
			Hemopexin domain signature: F1031-D1088	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Somatomedin B domain signature: Q21-T100	PROFILESCAN
			4-disulphide core signature PR00003: E42-C49, C49-D58	BLIMPS_PRINTS
			Somatomedin B signature PR00022: S24-F37, R41-Y52, D53-A64	BLIMPS_PRINTS
			Proline rich extensin signature PR01217: P366-P382, P885-M906, T573-T589, K268-S293	BLIMPS_PRINTS
			MEGAKARYOCYTE STIMULATING FACTOR	BLAST_PRODROM
			PD094009: E862-P1060	
			PD097597: T1106-P1270	
			PD117815: E65-S226	
			ACIDIC SERINE CLUSTER REPEAT	BLAST_DOMO
			DM04746 S57757 1-646: K153-P732, K416-T970	
			do ANONYMOUS; THROMBOSPONDIN	BLAST_DOMO
			DM06934 Q01443 275-825: K409-E812, N69-P388, P543-T970	
			do NEUROFILAMENT; TRIPLET	BLAST_DOMO
			DM04498 P12036 434-1019: A64-K338, E520-P971, K146-A701, T462-P949	BLAST_DOMO
			SALIVARY GLUE PROTEIN	
			DM02055 P02840 17-234: D682-T914, P102-T337, P208-T438, P428-T659, P775-T948	
			Hemopexin domain signature: I1051-F1065	MOTIFS
			Somatomedin B domain signature: C43-C63	MOTIFS
			Potential Phosphorylation Sites: S23 S28 S90 S115 S172 S183 S612 S681 S738 S742 S758 S899 S1059 T74 T79 T100 T120 T136 T148 T158 T162 T190 T198 T204 T212 T220 T228 T236 T243 T251 T258 T266 T274 T282 T290 T297 T305 T313 T321 T329 T336 T344 T360 T368 T376 T384 T392 T399 T407 T414 T422 T430 T438 T446 T454 T462 T463 T478 T486 T491 T493 T501 T509 T510 T517 T518 T525 T526 T534 T542 T550 T558 T566 T574 T582 T590 T603 T610 T619 T627 T635 T643 T651 T659 T672 T679 T688 T696 T704 T711 T719 T728 T751 T766 T773 T780 T786 T792 T809 T811 T830 T841 T853 T854 T859 T870 T884 T898 T900 T917 T922 T924 T931 T942 T1022 T1052 T1115 T1181 T1250 T1255 Y1126 Y1233	
			Potential Glycosylation Sites: N72 N1025	MOTIFS
5	7513018CD1	1311	signal_cleavage: M1-S24	SPSCAN
			Signal Peptide: M1-V18	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Signal Peptide: M1-Q21	HMMER
			Signal Peptide: M1-S24	HMMER
			Signal Peptide: M1-S29	HMMER
			Signal Peptide: M1-C30	HMMER
			Signal Peptide: M1-Q20	HMMER
			Hemopexin-like repeats.: I1102-G1145, V1058-S1100	HMMER_SMART
			Somatomedin B -like domains: D26-L68, S69-K108	HMMER_SMART
			Somatomedin B domain: D26-S69, C70-D109	HMMER_PFAM
			Hemopexin: I1102-G1148, V1058-S1100	HMMER_PFAM
			Somatomedin B domain IPB001212: C84-C104, P1099-F1140, Q1149-Q1186, D1273-P1311	BLIMPS_BLOCKS
			Hemopexin domain signature: F1072-D1129	PROFILES SCAN
			Somatomedin B domain signature: R62-T141	PROFILES SCAN
			Somatomedin B domain signature: Q25-R82	PROFILES SCAN
			4-disulphide core signature PR00003: E83-C90, C90-D99	BLIMPS_PRINTS
			Somatomedin B signature PR00022: T65-F78, R82-Y93, D94-A105	BLIMPS_PRINTS
			Proline rich extensin signature PR01217: P407-P423, P926-M947, T614-T630, K309-S334	BLIMPS_PRINTS
			MEGAKARYOCYTE STIMULATING FACTOR	BLAST_PRODROM
			PD094009: E903-P1101	
			PD097597: T1147-P1311	
			PD117815: E106-S267	
			ACIDIC SERINE CLUSTER REPEAT	BLAST_DOMO
			DM04746[S57757]1-646: K194-P773, K92-K707, S295-K913, T434-P1018	
			do ANONYMOUS; THROMBOSPONDIN;	BLAST_DOMO
			DM06934[Q01443]275-825: K450-E853, N110-P429, P230-P658, P584-T1011	
			do NEUROFILAMENT; TRIPLET	BLAST_DOMO
			DM04498[P12036]434-1019: A105-K379, E106-P605, E561-A1004, E561-P1012, K187-A742	
			SALIVARY GLUE PROTEIN DM02055[P02840]17-234: D723-T955	BLAST_DOMO
			SALIVARY GLUE PROTEIN	BLAST_DOMO
			DM02055[P02840]17-234: P143-T378, P249-T479, P360-P592, P533-T753, P816-T989	

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Hemopexin domain signature: I1092-F1106	MOTIFS
			Somatomedin B domain signature: C44-C64, C84-C104	MOTIFS
			Potential Phosphorylation Sites: S23 S69 S131 S156 S213 S224 S653 S722 S779 S783 S799 S940 S1100 T115 T120 T141 T161 T177 T189 T199 T203 T231 T239 T245 T253 T261 T269 T277 T284 T292 T299 T307 T315 T323 T331 T338 T346 T354 T362 T370 T377 T385 T401 T409 T417 T425 T433 T440 T448 T455 T463 T471 T479 T487 T495 T503 T504 T519 T527 T532 T534 T542 T550 T551 T558 T559 T566 T567 T575 T583 T591 T599 T607 T615 T623 T631 T644 T651 T660 T668 T676 T684 T692 T700 T713 T720 T729 T737 T745 T752 T760 T769 T792 T807 T814 T821 T827 T833 T850 T852 T871 T882 T894 T895 T900 T911 T925 T939 T941 T958 T963 T965 T972 T983 T1063 T1093 T1156 T1222 T1291 T1296 Y1167 Y1274	MOTIFS
			Potential Glycosylation Sites: N113 N1066	MOTIFS
6	7521129CD1	363	Signal cleavage: M1-G24	SPSCAN
			Signal Peptide: M1-T22	HMMER
			Signal Peptide: M1-G24	HMMER
			Signal Peptide: M1-G29	HMMER
			Transforming growth factor-beta (TGF-beta): C293-G351	HMMER_SMART
			Transforming growth factor beta like: E290-E362	HMMER_PFAM
			TGF-beta propeptide: C33-G252	HMMER_PFAM
			Transforming growth factor (TGF) beta family IPB001839: W310-C326	BLIMPS_BLOCKS
			TGF-beta family signature: K291-T346	PROFILESAN
			Transforming growth factor beta precursor signature PR01423: W17-I36, P72-T84, K163-N177, S178-P193, W195-S209	BLIMPS_PRINTS
			Transforming growth factor beta 1 precursor signature PR01424: L12-S31, K34-R43, F135-E154, Q166-N177, W207-L219, P264-R275	BLIMPS_PRINTS
			GLYCOPROTEIN PRECURSOR SIGNAL GROWTH FACTOR PROTEIN CYTOKINE BETA BONE MORPHOGENETIC PD000853: L30-L187	BLAST_PRODROM
			TRANSFORMING GROWTH FACTOR BETA 1 PRECURSOR TGF-beta 1 GROWTH FACTOR MITOGEN SIGNAL PD162517: M261-N292	BLAST_PRODROM
			TGF-BETA FAMILY DM00245[P50414]59-390: L59-V339	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			TGF-BETA FAMILY DM00245P16047 51-412: L57-Q269, R244-V339	BLAST_DOMO
			TGF-BETA FAMILY DM00245P16176 50-382: L73-V339	BLAST_DOMO
			TGF-BETA FAMILY DM00245P30371 49-412: R58-C225, P256-V339	BLAST_DOMO
			Potential Phosphorylation Sites: S64 S83 S151 S178 S183 S273 S287 T32 T116 T241 T263 T289 Y284 Y317	MOTIFS
			Potential Glycosylation Sites: N82 N136 N176	MOTIFS
			Cell attachment sequence: R244-D246	MOTIFS
			TGF-beta family signature: I311-C326	MOTIFS
7	7521148CD1	180	Signal_cleavage: M1-A22	SPSCAN
			Fibrinogen beta and gamma chains C-terminal globular domain IPB002181: Q39-L49	BLIMPS_BLOCKS
			Potential Phosphorylation Sites: T25	MOTIFS
			Potential Glycosylation Sites: N96	MOTIFS
8	7521157CD1	106	Signal_cleavage: M1-G22	SPSCAN
			Signal Peptide: M1-G17	HMMER
			Signal Peptide: M1-L20	HMMER
			Signal Peptide: M1-G22	HMMER
			Signal Peptide: M1-S25	HMMER
			Ephrin IPB001799: V30-V39	BLIMPS_BLOCKS
			Interleukin-1 signature: L11-L69	PROFILESAN
			Ephrin signature PR01347: V30-T44	BLIMPS_PRINTS
			EPLG2; STRA1; KINASE; LIGAND; DM01154 P52798 1-178: M1-V40	BLAST_DOMO
			Potential Phosphorylation Sites: S25 S64 T54	MOTIFS
			Potential Glycosylation Sites: N33	MOTIFS
9	7521195CD1	230	Signal_cleavage: M1-A18	SPSCAN
			Signal Peptide: M1-A18	HMMER
			Signal Peptide: M1-S20	HMMER
			Signal Peptide: M1-G22	HMMER
			Signal Peptide: M1-C26	HMMER
			Domain in ryanodine and inositol triphosphate receptor: L21-K75, P159-G212, G83-N157	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			MIR domain (protein mannosyltransferase, IP3R and RyR): L21-K75, P159-G212, G83-N157	HMMER_PFAM
			SDF2 R12E2.13 PROTEIN PD152269: M1-H96	BLAST_PRODROM
			DOLICHYLPHOSPHATEMANNOSE PROTEIN MANNOSYLTRANSFERASE TRANSFERASE GLYCOSYLTRANSFERASE GLYCOPROTEIN TRANSMEMBRANE ENDOPLASMIC RETICULUM MULTIGENE FAMILY Y PD005330: V97-E219	BLAST_PRODROM
			Potential Phosphorylation Sites: S62 S172 T61 T78 T99	MOTIFS
10	7521277CD1	358	Signal_cleavage: M1-A22	SPSCAN
			Fibrinogen-related domains (FReDs): Q234-V358	HMMER_SMART
			Fibrinogen beta and gamma chains, C-term: V235-V358	HMMER_PFAM
			Fibrinogen beta and gamma chains C-terminal globul IPB002181: Q39-L49, Q243-V257, V266-G302, E307-T319, L326-Y342	BLIMPS_BLOCKS
			PRECURSOR GLYCOPROTEIN SIGNAL FIBRINOGEN BLOOD COAGULATION CHAIN PLASMA PROTEIN PLATELET PD001241: Q237-V356	BLAST_PRODROM
			ANGIOPOIETIN1 ANGIOPOIETIN2 ANGIOPOIETIN PD016415: E28-L206	BLAST_PRODROM
			FIBRINOGEN BETA/GAMMA DM00531 JN0596 27-305: V192-V356	BLAST_DOMO
			FIBRINOGEN BETA/GAMMA DM00531 P02671 574-863: Q237-V358	BLAST_DOMO
			FIBRINOGEN BETA/GAMMA DM00531 P06399 493-779: Q234-V358	BLAST_DOMO
			FIBRINOGEN BETA/GAMMA DM00531 P04115 105-411: G191-V356	BLAST_DOMO
			Potential Phosphorylation Sites: S109 S144 S166 S205 S325 T25 T108 T122 T183 T319	MOTIFS
			Potential Glycosylation Sites: N96 N106 N195 N222 N259 N285	MOTIFS
11	7515701CD1	63	signal_cleavage: M1-T23	SPSCAN
			Signal Peptide: M1-T23	HMMER
			Signal Peptide: M1-C30	HMMER
			Signal Peptide: M1-L25	HMMER
			Signal Peptide: P5-T23	HMMER
			Insulin-like growth factor-binding protein IPB000867: Q24-W34	BLIMPS_BLOCKS
			Potential Phosphorylation Sites: T4	MOTIFS
12	7515527CD1	284	signal_cleavage: M1-S65	SPSCAN
			Signal Peptide: G45-T62	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Signal Peptide: K43-F66	HMMER
			Cytosolic domain: M1-L47 Transmembrane domain: L48-A70 Non-cytosolic domain: A71-L284	TMHMMER
			TNF family signature and profile: K173-V217	PROFILESCAN
			PROLIFERATION INDUCING LIGAND APRIL TRDL-1 FACTOR INDUCING LEUKOCYTE 2310026N09RIK TALL-2 NECROSIS TNF-RELATED PD180968: Q73-L283	BLAST_PRODROM
			Potential Phosphorylation Sites: S36 S38 S41 S78 S171 S176 T5 T202 T263 T268	MOTIFS
			Potential Glycosylation Sites: N124 N242	MOTIFS
13	7519261CD1	91	signal_cleavage: M1-S25	SPSCAN
			Signal Peptide: M1-S25	HMMER
			Potential Phosphorylation Sites: S67 S81 T72	MOTIFS
			Potential Glycosylation Sites: N24	MOTIFS
14	7519923CD1	273	signal_cleavage: M1-A26	SPSCAN
			Signal Peptide: M1-A26	HMMER
			Signal Peptide: M1-G28	HMMER
			Signal Peptide: M1-G25	HMMER
			Epidermal growth factor-like domain: I113-H142	HMMER_SMART
			EGF-like domain: C114-C141	HMMER_PFAM
			Calcium-binding EGF-like domain IPB001881: C130-H140	BLIMPS_BLOCKS
			Type III EGF-like signature PR00011: V123-C141	BLIMPS_PRINTS
			PROTEIN GLYCOPROTEIN EGFLIKE DOMAIN PD031267: M1-I113	BLAST_PRODROM
			PROTEIN GLYCOPROTEIN EGFLIKE DOMAIN PD030385: G166-G267	BLAST_PRODROM
			Potential Phosphorylation Sites: S3 S30 S244 S260 T73 T162	MOTIFS
			Potential Glycosylation Sites: N50	MOTIFS
			Egf_1: C130-C141	MOTIFS
			Egf_2: C130-C141	MOTIFS
15	7519986CD1	106	signal_cleavage: M1-G40	SPSCAN
			Signal Peptide: M18-G40	HMMER
			Signal Peptide: M18-A43	HMMER
			Signal Peptide: M18-V42	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Cytosolic domain: Q51-M106 Transmembrane domain: S28-T50 Non-cytosolic domain: M1-L27	TMHMMER
			PROLIFERATION INDUCING LIGAND APRIL PD180968: M1-W94	BLAST_PRODOM
			Potential Phosphorylation Sites: S57 S82	MOTIFS
16	7520799CD1	135	signal_cleavage: M1-S21	SPSCAN
			Signal Peptide: N5-S21	HMMER
			Signal Peptide: A7-A25	HMMER
			Signal Peptide: M1-S21	HMMER
			Signal Peptide: M1-A25	HMMER
			Signal Peptide: M1-P27	HMMER
			Glucagon like hormones: H90-L116	HMMER_SMART
			Peptide hormone: H90-N117	HMMER_PFAM
			Cytosolic domain: M1-K6 Transmembrane domain: A7-Y29 Non-cytosolic domain: R30-K135	TMHMMER
			Glucagon / GIP / secretin / VIP family IPB000532: H90-L116	BLIMPS_BLOCKS
			Glucagon polypeptide hormone superfamily signature PR00275: H90-T100, R101-L112	BLIMPS_PRINTS
			VASOACTIVE INTESTINAL PEPTIDE PRECURSOR VIP GLUCAGON FAMILY CLEAVAGE	BLAST_PRODOM
			ON PAIR PD012348: F20-S76	
			do GLUCAGON; PEPTIDE; INTESTINAL; VASOACTIVE; DM04807 P01282 I-79: M1-S76	BLAST_DOMO
			do GLUCAGON; PEPTIDE; INTESTINAL; VASOACTIVE; DM04807 P32648 I-79: M1-S76	BLAST_DOMO
			GLUCAGON / GIP / SECRETIN / VIP FAMILY DM00283 P01282 I16-150: S81-L116	BLAST_DOMO
			GLUCAGON / GIP / SECRETIN / VIP FAMILY DM00283 P48143 I22-154: P86-L116	BLAST_DOMO
			Potential Phosphorylation Sites: S53 S91 S122	MOTIFS
			Potential Glycosylation Sites: N79 N98	MOTIFS
			Glucagon / GIP / secretin / VIP family signature: H90-L112	MOTIFS
17	7520947CD1	132	signal_cleavage: M1-A33	SPSCAN
			Signal Peptide: C15-C32	HMMER
			Signal Peptide: K9-C32	HMMER
			Signal Peptide: M1-A34	HMMER
			Signal Peptide: C15-A34	HMMER
			Signal Peptide: M1-A33	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Signal Peptide: M1-C32	HMMER
			Potential Phosphorylation Sites: S85 S100 S125 T40 T96 Y58	MOTIFS
18	7521071CD1	175	signal_cleavage: M1-A18	SPSCAN
			Signal Peptide: M1-A18	HMMER
			Signal Peptide: M1-S20	HMMER
			Signal Peptide: M1-G22	HMMER
			Signal Peptide: M1-C26	HMMER
			Domain in ryanodine and inositol trisphosphate receptors and protein o-mannosyl transferases: G83-G157, L21-K75	HMMER_SMART
			MIR domain: G83-G157, L21-K75	HMMER_PFAM
			SDF2 R12E2.13 PROTEIN PD152269: M1-H96	BLAST_PRODOME
			LUMENAL DOMAIN DM02305[P42934]30-714: L41-Q129	BLAST_DOMO
			Potential Phosphorylation Sites: S62 T61 T78 T99	MOTIFS
19	7521082CD1	95	signal_cleavage: M1-G20	SPSCAN
			Signal Peptide: T4-G20	HMMER
			Signal Peptide: M1-L19	HMMER
			Signal Peptide: M1-G20	HMMER
			Signal Peptide: M1-K22	HMMER
			Signal Peptide: M1-G25	HMMER
			GASTRIC INHIBITORY POLYPEPTIDE PRECURSOR GIP GLUCOSE DEPENDENT	BLAST_PRODOME
			INSULINOTROPIC GLUCAGON FAMILY HORMONE PD019203: W29-R95	MOTIFS
			Potential Phosphorylation Sites: S28 S66 T34	MOTIFS
			Potential Glycosylation Sites: N32	SPSCAN
20	7521329CD1	444	signal_cleavage: M1-A22	HMMER_SMART
			Fibrinogen-related domains (FReDs): Q234-L442	HMMER_PFAM
			Fibrinogen beta and gamma chains, C-terminal globular domain: V235-L442	BLIMPS_BLOCKS
			Fibrinogen beta and gamma chains C-terminal globular domain	
			IPB002181: Q39-L49, Q243-V257, V266-G302, E307-T319, L326-Y342, L347-Q361, K384-N413, M417-P441	

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Fibrinogen beta and gamma chains C-terminal domain signature: G348-S429	PROFESCAN
			PRECURSOR GLYCOPROTEIN SIGNAL FIBRINOGEN BLOOD COAGULATION CHAIN	BLAST_PRODOM
			PLASMA PROTEIN PLATELET	
			PD001241: Q237-Y354, F371-P441	
			ANGIOPOIETIN1 ANGIOPOIETIN2 ANGIOPOIETIN	BLAST_PRODOM
			PD016415: E28-L206	
			FIBRINOGEN BETA/GAMMA	BLAST_DOMO
			DM00531 JN0596 27-305: K81-K103, Q128-K168, V192-S362, Q367-D443	
			DM00531 S61517 64-326: Q237-P441	
			DM00531 P10039 1554-1796: Q237-P441	
			DM00531 JH0675 1096-1338: V217-P441	
			Potential Phosphorylation Sites:	MOTIFS
			S109 S144 S166 S205 S325 S372 S376 S431 T25 T108 T122 T183 T319	
			Potential Glycosylation Sites:	MOTIFS
			N96 N106 N195 N222 N259 N285 N368	
			Fibrinogen beta and gamma chains C-terminal domain signature: W393-G405	MOTIFS
21	7522977CD1	60	signal_cleavage: M1-A19	SPSCAN
			Signal Peptide: M1-A19, M1-C20, M1-I22, M1-N24, M1-A32	HMMER
			Neurohypophysial hormone	BLIMPS_BLOCKS
			IPB000981: C20-L47	
			Neurohypophysial hormones signature: M1-T60	PROFESCAN
			Neurohypophysial hormone signature PR00831: Y21-D35	BLIMPS_PRINTS
			NEUROHYPOPHYSIAL HORMONES DM00829 P01175 1-29: M1-K30	BLAST_DOMO
			Potential Phosphorylation Sites: T50	MOTIFS
			Neurohypophysial hormones signature: C20-G28	MOTIFS
22	7521939CD1	365	signal_cleavage: M1-A29	SPSCAN
			Signal Peptide: M1-A29	HMMER
			Transforming growth factor-beta (TGF-beta): C264-H365	HMMER_SMART
			Transforming growth factor beta like domain: R261-H365	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			TGF-beta propeptide: V37-R274	HMMER_PFAM
			Transforming growth factor (TGF) beta family	BLIMPS_BLOCKS
			IPB001839: W281-C297, P328-Y342, S346-H365	
			TGF-beta family signature: Q262-H318	PROFILESAN
			Growth factor cysteine knot superfamily signature	BLIMPS_PRINTS
			PR00438: A289-A298, R360-C364	
			Inhibin alpha chain signature PR00669: C264-W281, W281-A298	BLIMPS_PRINTS
			GLYCOPROTEIN PRECURSOR SIGNAL GROWTH FACTOR PROTEIN CYTOKINE BETA	BLAST_PRODROM
			BONE MORPHOGENETIC	
			PD000357: C264-C364	
			PD000853: Q49-L252	
			BONE MORPHOGENETIC PROTEIN PRECURSOR BMP7 OSTEOGENIC OP1 SIGNAL	BLAST_PRODROM
			GROWTH FACTOR PD027962: M1-R45	
			TGF-BETA FAMILY	BLAST_DOMO
			DM00245 P18075 I07-431: Y107-S256, P239-H365	
			DM00245 P22003 I32-454: S112-S256, V249-H365	
			DM00245 I49542 I30-452: S112-S256, S256-H365	
			DM00245 P22004 I90-513: K75-G95, P117-D253, V249-H365	
			Potential Phosphorylation Sites: S5 S120 S256 S272 T128 T170 T189 T251	MOTIFS
			Potential Glycosylation Sites: N187 N255 N306	MOTIFS
			Transforming growth factor beta like domain: I282-C297	MOTIFS
23	7522030CD1	100	signal_cleavage: M1-R25	SPSCAN
			Signal Peptide: M1-G20, M1-C23, M1-R25, M1-L28	HMMER
			MANIAC FRINGE HOMOLOG PRECURSOR MANIC	BLAST_PRODROM
			FRINGE SIGNAL PD127820: M1-P48	
			Potential Phosphorylation Sites: T62	MOTIFS
24	7522059CD1	187	signal_cleavage: M1-G24	SPSCAN
			Signal Peptide: M1-G24, M1-A27	HMMER
			Stanniocalcin family: L6-D187	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			STANNIOCALCIN HORMONE GLYCOPROTEIN PRECURSOR SIGNAL STC CORPUSCLES OF STANNIUS PROTEIN PD011156: S46-D164	BLAST_PRODOME
			STANNIOCALCIN2 PD155582: M1-L45	BLAST_PRODOME
			STANNIOCALCIN CHAIN	BLAST_DOMO
			DM04598 P52823 1-246: A27-I160	
			DM04598 P18301 1-262: G7-M159	
			DM04598 Q08264 1-255: G24-I160	
			Potential Phosphorylation Sites: S122	MOTIFS
			Potential Glycosylation Sites: N73	MOTIFS
25	7523442CD1	156	signal_cleavage: M1-G23	SPSCAN
			Signal Peptide: M1-S18, M1-A19, M1-G23, M1-S25	HMMER
			HEPARIN-BINDING EGF-LIKE GROWTH FACTOR PRECURSOR SIGNAL HB-EGF HB-EGF DOMAIN TRANSMEMBRANE PD012465: G32-D106	BLAST_PRODOME
			AMPHIREGULIN	BLAST_DOMO
			DM01746 Q09118 58-150: D58-E143	
			DM01746 Q06186 58-150: V61-C132	
			HEPARIN-BINDING EGF-LIKE GROWTH FACTOR DM04849 Q06186 1-56: M1-R57	BLAST_DOMO
			DM04849 Q09118 1-56: M1-R57	
			Potential Phosphorylation Sites: S38 S77	MOTIFS

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
26/7513276CB1/1014	1-481, 1-1014, 91-717, 95-438, 95-601, 95-605, 97-381, 113-411, 113-428, 129-389, 129-598, 307-497, 323-605, 365-605, 603-1014, 604-1012, 605-1012, 613-1012, 616-1014, 618-1012, 641-1014, 645-1012, 654-1004, 674-1007, 704-1004, 727-1012, 732-1014, 832-1014
27/7513290CB1/2367	1-433, 3-2360, 67-486, 86-551, 94-551, 95-547, 98-551, 101-581, 102-500, 108-563, 108-581, 114-581, 139-551, 142-581, 148-551, 151-547, 153-552, 269-551, 270-572, 270-576, 342-572, 511-706, 559-1009, 559-1135, 574-1052, 582-1240, 583-927, 592-1024, 653-797, 701-1108, 779-963, 884-1130, 904-1395, 944-1290, 959-1291, 965-1528, 1022-1718, 1083-1306, 1172-1424, 1176-1450, 1181-1440, 1193-1449, 1346-1620, 1517-2247, 1527-1796, 1527-2041, 1540-1790, 1549-2250, 1586-1822, 1603-1885, 1610-2074, 1610-2121, 1643-2332, 1651-1883, 1651-2068, 1655-1889, 1656-2332, 1662-1913, 1667-1922, 1700-1955, 1719-2261, 1741-2330, 1745-2254, 1786-2266, 1791-2278, 1816-2059, 1838-2307, 1842-2321, 1846-2329, 1854-2318, 1855-2333, 1857-2319, 1863-2312, 1869-2331, 1870-2125, 1870-2312, 1870-2333, 1872-2318, 1889-2144, 1895-2318, 1896-2166, 1896-2261, 1896-2312, 1911-2333, 1924-2320, 1942-2224, 1950-2091, 1961-2212, 1961-2240, 1961-2280, 1961-2318, 1995-2155, 1998-2259, 1998-2279, 2038-2318, 2069-2318, 2138-2367, 2188-2333
28/7513749CB1/1874	1-875, 211-1196, 741-1648, 1111-1874
29/7513017CB1/4140	1-88, 1-149, 1-176, 1-228, 2-228, 2-229, 3-227, 3-373, 3-4140, 6-397, 7-367, 8-216, 9-185, 9-228, 11-228, 12-186, 12-227, 12-360, 16-228, 33-175, 66-717, 117-583, 117-709, 118-359, 127-692, 144-329, 144-640, 159-385, 171-820, 194-715, 194-717, 225-418, 257-628, 313-614, 313-632, 313-700, 313-714, 330-600, 357-956, 373-600, 378-635, 387-987, 416-670, 501-735, 549-946, 645-884, 653-901, 653-1038, 653-1395, 654-745, 661-891, 683-706, 683-714, 733-892, 747-867, 749-890, 749-901, 759-790, 759-949, 759-1171, 763-963, 766-1171, 767-963, 792-852, 792-864, 792-872, 794-1024, 794-1074, 794-1207, 798-959, 806-840, 806-852, 806-997, 806-1013, 808-1013, 812-1013, 813-963, 815-862, 815-868, 815-876, 823-877, 831-1185, 832-867, 833-963, 833-1171, 835-1185, 838-900, 842-1104, 843-1104, 854-888, 854-996, 854-1058, 856-1058, 857-963, 858-901, 858-1096, 866-1182, 909-952, 909-960, 909-1030, 909-1185, 910-1104, 911-1025, 912-1324, 916-957, 925-959, 925-969, 927-1165, 928-1030, 930-1020, 931-1020, 932-969, 935-1251, 936-1396, 939-1104, 962-1213, 962-1252, 963-1104, 964-1019, 964-1027, 964-1041, 969-1005, 969-1110, 969-1178, 973-1017, 973-1020, 979-1041, 980-1033, 988-1042, 998-1104, 1008-1320, 1008-1468, 1011-1171, 1021-1213, 1029-1178, 1035-1213, 1035-1429, 1043-1171, 1043-1182, 1052-1316, 1053-1495, 1056-1190, 1066-1100, 1066-1110, 1070-1171, 1070-1407, 1073-1171, 1074-1335, 1076-1207, 1091-1149, 1094-1185, 1098-1149, 1103-1362, 1105-1168, 1105-1194, 1105-1285, 1110-1146, 1110-1159, 1114-1149, 1114-1159, 1114-1197, 1116-1354, 1116-1499, 1117-1149, 1120-1182, 1124-1388, 1136-1170, 1143-1213, 1145-1198, 1145-1206, 1153-1207, 1162-1362, 1163-1213, 1165-1499, 1166-1362, 1173-1436, 1184-1213, 1186-1213, 1188-1213, 1217-1358, 1217-1640, 1218-1426, 1230-1258, 1231-1436, 1260-1320, 1260-1332, 1260-1362, 1260-1388, 1260-1479, 1260-1499, 1262-1498, 1262-1499, 1265-1409, 1266-1430, 1274-1308, 1274-1365, 1274-1484, 1276-1484, 1277-1362, 1279-1693, 1280-1484, 1281-1388, 1291-1345, 1299-1431, 1301-1368, 1301-1429, 1301-1436, 1306-1359,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1307-1351, 1307-1359, 1310-1495, 1324-1362, 1324-1368, 1325-1436, 1326-1495, 1327-1429, 1329-1436, 1331-1368, 1338-1431, 1352-1431, 1356-1407, 1361-1431, 1363-1432, 1368-1401, 1372-1461, 1387-1609, 1394-1428, 1394-1478, 1396-1431, 1396-1436, 1396-1499, 1405-1499, 1410-1499, 1424-1499, 1427-1729, 1427-1729, 1448-1499, 1452-1499, 1457-1499, 1476-1640, 1476-1729, 1477-1623, 1488-1670, 1526-1671, 1548-1868, 1551-2011, 1561-1810, 1566-1613, 1574-1719, 1614-1841, 1614-1921, 1615-1963, 1616-1841, 1642-1793, 1658-1732, 1663-1753, 1665-1756, 1665-1829, 1666-1756, 1667-1725, 1670-1908, 1673-1963, 1685-1749, 1709-1837, 1767-2128, 1780-1815, 1783-1815, 1787-1815, 1788-1820, 1794-2038, 1813-1963, 1815-1846, 1815-2227, 1817-1851, 1817-1863, 1817-1873, 1818-1847, 1818-1863, 1818-1864, 1818-2169, 1819-2048, 1821-1868, 1821-1925, 1821-2017, 1821-2075, 1822-2538, 1823-2048, 1827-1974, 1850-2036, 1851-1963, 1861-1963, 1871-1963, 1872-2000, 1877-1963, 1877-2115, 1880-2475, 1889-1963, 1899-1996, 1916-1980, 1916-2044,
	1974-2170, 1987-2022, 1994-2022, 2001-2080, 2020-2560, 2022-2053, 2024-2058, 2024-2080, 2024-2169, 2025-2053, 2025-2054, 2025-2070, 2026-2170, 2029-2170, 2030-2170, 2034-2157, 2035-2139, 2035-2156, 2035-2157, 2036-2231, 2057-2170, 2058-2347, 2068-2416, 2078-2307, 2078-2476, 2079-2157, 2082-2157, 2084-2142, 2084-2270, 2087-2173, 2095-2363, 2099-2128, 2099-2157, 2104-2157, 2106-2149, 2109-2157, 2122-2303, 2125-2373, 2127-2166, 2129-2713, 2202-2716, 2203-2490, 2208-2469, 2208-2475, 2218-2450, 2221-2408, 2230-2674, 2250-2512, 2252-2540, 2254-2673, 2256-2470, 2256-2477, 2273-2505, 2279-2544, 2287-2484, 2300-2552, 2304-2482, 2349-2548, 2364-2647, 2368-2642, 2370-2640, 2370-2651, 2380-2624, 2382-2633, 2392-2596, 2404-2705, 2405-2988, 2412-2633, 2434-2649, 2437-2577, 2442-2696, 2444-2815, 2453-2665, 2453-2722, 2465-2714, 2465-2734, 2481-2757, 2488-2731, 2492-2785, 2493-2780, 2496-2728, 2502-2741, 2514-2634, 2515-2993, 2522-2751, 2526-2788, 2534-2775, 2534-2815, 2539-2797, 2562-2756, 2568-2805, 2568-2840,
	2570-2800, 2576-2862, 2577-2878, 2581-2800, 2586-2882, 2592-2898, 2598-3376, 2600-2950, 2603-2826, 2603-2968, 2613-2888, 2618-2845, 2618-2854, 2618-2857, 2623-2866, 2624-2882, 2626-2901, 2627-2938, 2633-2884, 2633-2905, 2641-2875, 2641-2941, 2643-2909, 2646-2882, 2648-2874, 2651-2944, 2667-2879, 2681-2934, 2682-2927, 2682-2929, 2682-2955, 2684-2946, 2684-2980, 2691-2948, 2696-2979, 2703-3009, 2704-3150, 2706-3023, 2713-2903, 2720-2963, 2720-3002, 2722-3003, 2727-3018, 2731-2886, 2733-2974, 2739-3000, 2741-2963, 2750-3014, 2750-3101, 2758-2962, 2758-2965, 2758-3007, 2760-2915, 2760-2982, 2760-2988, 2760-3005, 2761-3071, 2776-3022, 2778-3045, 2779-3010, 2788-3132, 2792-3085, 2793-3047, 2794-3048, 2797-3070, 2799-2922, 2812-3047, 2812-3098, 2826-3027, 2828-3119, 2828-3129, 2835-3103, 2843-3061, 2848-3067, 2849-2980, 2851-3146, 2870-3119, 2887-3315, 2897-3148, 2909-3177, 2911-3203, 2912-3169, 2916-3161, 2916-3184, 2918-3158, 2921-3184, 2922-3124, 2922-3153, 2922-3194, 2941-3121, 2961-3194, 2968-3187, 2978-3242,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	2980-3205, 2980-3237, 2995-3232, 2996-3260, 3006-3217, 3008-3255, 3014-3255, 3015-3248, 3020-3180, 3023-3325, 3024-3286, 3026-3594, 3029-3288, 3032-3274, 3032-3293, 3043-3291, 3043-3341, 3044-3267, 3049-3270, 3051-3167, 3065-3305, 3065-3333, 3066-3318, 3067-3657, 3070-3171, 3078-3368, 3091-3309, 3107-3362, 3109-3378, 3109-3467, 3111-3346, 3128-3301, 3131-3387, 3131-3395, 3139-3394, 3146-3339, 3146-3365, 3148-3358, 3149-3445, 3168-3330, 3168-3714, 3169-3416, 3173-3390, 3198-3429, 3198-3434, 3221-3309, 3221-3391, 3223-3483, 3228-3476, 3238-3435, 3245-3472, 3245-3517, 3246-3521, 3247-3467, 3247-3513, 3247-3550, 3259-3554, 3272-3519, 3273-3516, 3278-3535, 3279-3540, 3279-3651, 3291-3873, 3296-3580, 3298-3529, 3303-3504, 3303-3546, 3304-3822, 3308-3487, 3308-3499, 3309-3557, 3312-3562, 3315-3576, 3317-3599, 3319-3626, 3321-3612, 3321-3621, 3322-3454, 3322-3895, 3323-3568, 3331-3549, 3334-3605, 3337-3562, 3343-3573, 3343-3605, 3345-3558, 3348-3592, 3350-3611, 3353-3935, 3359-3592, 3365-3612.
	3377-3527, 3384-3939, 3391-3625, 3396-3650, 3399-3628, 3404-3938, 3406-3641, 3406-3720, 3408-3603, 3410-3640, 3422-3658, 3425-3654, 3425-3679, 3427-3648, 3428-3645, 3428-3889, 3429-3615, 3429-3676, 3431-3894, 3448-3994, 3449-3724, 3451-3733, 3452-3701, 3452-3865, 3454-3685, 3461-3703, 3462-3725, 3469-3692, 3469-3758, 3476-3736, 3477-3734, 3478-3725, 3491-3931, 3509-3804, 3512-3653, 3519-3808, 3527-3958, 3534-3765, 3534-3813, 3554-3856, 3555-3786, 3555-3788, 3560-3771, 3561-3820, 3561-3826, 3561-3887, 3561-3916, 3561-3956, 3563-3795, 3563-3934, 3575-3819, 3581-3869, 3582-3814, 3589-3848, 3592-4012, 3604-3777, 3604-3839, 3626-3855, 3626-3875, 3632-3862, 3634-3852, 3634-3853, 3635-3869, 3636-3881, 3648-3880, 3652-3812, 3652-3938, 3655-3939, 3658-3894, 3660-3926, 3661-3956, 3665-3923, 3680-3939, 3682-3893, 3694-3811, 3698-3902, 3699-3992, 3718-3954, 3724-3939, 3741-3936, 3745-4087, 3746-3939, 3752-4140, 3757-3982, 3760-3939, 3774-3937, 3788-3918, 3798-3937, 3807-3937, 3807-3939, 3814-3958, 3824-3939, 3854-4091
30/7513018CB1/4289	1-520, 3-524, 17-235, 27-114, 27-131, 27-134, 27-233, 27-235, 27-246, 27-287, 28-185, 28-257, 28-258, 28-277, 28-295, 28-522, 29-4289, 30-293, 32-716, 34-277, 39-307, 40-669, 43-300, 84-573, 115-649, 117-651, 133-693, 140-682, 142-517, 255-377, 256-377, 266-732, 266-858, 267-508, 276-841, 293-478, 293-789, 308-534, 320-969, 343-864, 343-866, 374-567, 406-777, 479-749, 506-1105, 522-749, 527-784, 536-1136, 565-819, 698-1095, 794-1033, 802-1050, 802-1187, 802-1544, 803-894, 810-1040, 832-855, 832-863, 882-1041, 896-1016, 898-1039, 898-1050, 908-939, 908-1098, 908-1320, 912-1112, 915-1320, 916-1112, 941-1001, 941-1021, 943-1173, 943-1223, 943-1356, 947-1108, 955-989, 955-1001, 955-1146, 955-1162, 957-1162, 961-1162, 962-1112, 964-1017, 964-1025, 980-1334, 981-1016, 982-1112, 982-1320, 984-1334, 987-1049, 991-1253, 992-1253, 1003-1037, 1003-1145, 1003-1207, 1005-1207, 1006-1112, 1007-1050, 1007-1245, 1015-1331, 1058-1109, 1058-1179, 1058-1334, 1059-1253, 1060-1174, 1061-1473, 1074-1108, 1074-1118, 1076-1314,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1077-1179, 1079-1169, 1080-1118, 1084-1400, 1085-1545, 1088-1253, 1111-1362, 1111-1401, 1112-1253, 1113-1168, 1113-1190, 1118-1154, 1118-1259, 1118-1327, 1122-1166, 1122-1169, 1128-1190, 1129-1182, 1147-1253, 1157-1469, 1157-1617, 1160-1320, 1170-1362, 1178-1327, 1184-1362, 1184-1578, 1192-1320, 1192-1331, 1201-1465, 1202-1644, 1205-1339, 1215-1249, 1215-1259, 1219-1320, 1219-1556, 1222-1320, 1223-1484, 1225-1356, 1240-1298, 1243-1334, 1247-1298, 1252-1511, 1254-1434, 1259-1295, 1259-1308, 1263-1298, 1263-1308, 1263-1346, 1265-1503, 1265-1648, 1266-1298, 1269-1331, 1273-1537, 1285-1319, 1292-1362, 1294-1347, 1294-1355, 1311-1511, 1312-1362, 1314-1648, 1315-1511, 1322-1585, 1333-1362, 1337-1362, 1366-1507, 1366-1789, 1367-1575, 1379-1407, 1380-1585, 1409-1469, 1409-1511, 1409-1537, 1409-1628, 1409-1648, 1411-1647, 1411-1648, 1414-1558, 1415-1579, 1423-1457, 1423-1514, 1423-1633, 1425-1633, 1426-1511, 1428-1842, 1429-1633, 1430-1537, 1448-1580, 1450-1517, 1450-1578, 1450-1585, 1455-1508,
	1456-1508, 1459-1644, 1473-1511, 1473-1517, 1474-1585, 1475-1644, 1476-1578, 1478-1585, 1480-1517, 1487-1580, 1501-1580, 1505-1556, 1510-1580, 1517-1550, 1521-1610, 1536-1758, 1543-1577, 1543-1627, 1545-1580, 1545-1585, 1545-1648, 1554-1648, 1559-1648, 1573-1648, 1576-1878, 1576-2318, 1597-1648, 1601-1648, 1606-1648, 1625-1789, 1625-1878, 1626-1772, 1637-1819, 1675-1820, 1697-2017, 1700-2160, 1710-1959, 1715-1762, 1723-1868, 1763-1990, 1763-2070, 1764-2112, 1765-1990, 1791-1942, 1807-1881, 1812-1902, 1814-1905, 1814-1978, 1815-1905, 1816-1874, 1819-1905, 1819-2057, 1822-2112, 1834-1898, 1858-1986, 1916-2277, 1929-1964, 1932-1964, 1936-1964, 1937-1969, 1943-2187, 1962-2112, 1964-1995, 1964-2376, 1966-2000, 1966-2012, 1967-1996, 1967-2012, 1967-2013, 1967-2318, 1968-2197, 1970-2017, 1970-2074, 1970-2166, 1970-2224, 1971-2687, 1972-2197, 1976-2123, 1999-2185, 2000-2112, 2010-2112, 2020-2112, 2021-2149, 2026-2112, 2026-2264, 2029-2624, 2038-2112, 2048-2145, 2065-2129, 2065-2193, 2123-2319, 2136-2171, 2139-2171,
	2143-2171, 2150-2229, 2169-2709, 2171-2202, 2173-2207, 2173-2318, 2174-2202, 2174-2203, 2174-2219, 2175-2319, 2178-2319, 2179-2319, 2183-2306, 2184-2288, 2184-2305, 2184-2306, 2185-2380, 2206-2319, 2207-2496, 2217-2565, 2227-2456, 2227-2625, 2228-2306, 2231-2306, 2233-2291, 2233-2419, 2236-2322, 2244-2512, 2248-2277, 2248-2306, 2253-2306, 2255-2298, 2258-2306, 2271-2452, 2274-2522, 2276-2315, 2278-2862, 2351-2865, 2352-2639, 2357-2618, 2357-2624, 2367-2599, 2370-2557, 2379-2823, 2399-2661, 2401-2689, 2403-2822, 2405-2619, 2405-2626, 2422-2654, 2428-2693, 2436-2633, 2449-2701, 2453-2631, 2498-2697, 2513-2796, 2517-2791, 2519-2789, 2519-2800, 2529-2773, 2531-2782, 2541-2745, 2553-2854, 2554-3137, 2561-2782, 2583-2798, 2586-2726, 2591-2845, 2593-2964, 2602-2814, 2602-2871, 2614-2863, 2614-2883, 2630-2906, 2637-2880, 2641-2934, 2642-2929, 2645-2877, 2651-2890, 2663-2783, 2664-3142, 2671-2900, 2675-2937, 2683-2924, 2683-2964, 2688-2946, 2711-2905, 2717-2954, 2717-2989, 2719-2949, 2725-3011, 2726-3027, 2730-2949,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	2735-3031, 2741-3047, 2747-3525, 2749-3099, 2752-2975, 2752-3117, 2762-3037, 2767-2994, 2767-3003, 2767-3006, 2772-3015, 2773-3031, 2775-3050, 2776-3087, 2782-3033, 2782-3054, 2790-3024, 2790-3090, 2792-3058, 2795-3031, 2797-3023, 2800-3093, 2816-3028, 2830-3083, 2831-3076, 2831-3078, 2831-3104, 2833-3095, 2833-3129, 2840-3097, 2845-3128, 2852-3158, 2853-3299, 2855-3172, 2862-3052, 2869-3112, 2869-3151, 2871-3152, 2876-3167, 2880-3035, 2882-3123, 2888-3149, 2890-3112, 2899-3163, 2899-3250, 2907-3111, 2907-3114, 2907-3156, 2909-3064, 2909-3131, 2909-3137, 2909-3154, 2910-3220, 2925-3171, 2927-3194, 2928-3159, 2937-3281, 2941-3234, 2942-3196, 2943-3197, 2946-3219, 2948-3071, 2961-3196, 2961-3247, 2975-3176, 2977-3268, 2977-3278, 2984-3252, 2992-3210, 2997-3216, 2998-3129, 3000-3295, 3019-3268, 3036-3464, 3046-3297, 3058-3326, 3060-3352, 3061-3318, 3065-3310, 3065-3333, 3067-3307, 3070-3333, 3071-3273, 3071-3302, 3071-3343, 3090-3270, 3110-3343, 3117-3336, 3127-3391, 3129-3354, 3129-3386, 3144-3381, 3145-3409,
	3155-3366, 3157-3404, 3163-3404, 3164-3397, 3169-3329, 3172-3474, 3173-3435, 3175-3743, 3178-3437, 3181-3423, 3181-3442, 3192-3440, 3192-3490, 3193-3416, 3198-3419, 3200-3316, 3214-3454, 3214-3482, 3215-3467, 3216-3806, 3219-3320, 3227-3517, 3240-3458, 3256-3511, 3258-3380, 3258-3527, 3258-3616, 3260-3495, 3277-3450, 3280-3536, 3288-3543, 3295-3488, 3295-3514, 3297-3507, 3298-3594, 3317-3479, 3317-3863, 3318-3565, 3322-3539, 3347-3578, 3347-3583, 3370-3458, 3370-3540, 3372-3632, 3377-3625, 3387-3584, 3394-3621, 3394-3666, 3395-3670, 3396-3616, 3396-3662, 3396-3699, 3408-3703, 3421-3668, 3422-3665, 3427-3684, 3428-3689, 3428-3800, 3440-4022, 3445-3729, 3447-3678, 3452-3653, 3452-3695, 3453-3971, 3457-3636, 3457-3648, 3458-3706, 3461-3711, 3464-3725, 3464-3759, 3466-3748, 3468-3775, 3470-3761, 3470-3770, 3471-3603, 3471-4044, 3472-3717, 3480-3698, 3483-3754, 3486-3711, 3492-3722, 3492-3754, 3494-3707, 3497-3741, 3499-3760, 3502-4084, 3508-3741, 3514-3761, 3526-3676, 3533-4088, 3540-3774, 3545-3799,
	3548-3777, 3553-4087, 3555-3790, 3555-3869, 3557-3752, 3559-3789, 3571-3807, 3574-3803, 3574-3828, 3576-3797, 3577-3794, 3577-4038, 3578-3764, 3578-3825, 3580-4043, 3597-4143, 3598-3873, 3600-3882, 3601-3850, 3601-4014, 3603-3834, 3610-3852, 3611-3874, 3618-3841, 3618-3907, 3625-3885, 3626-3883, 3627-3874, 3640-4080, 3658-3953, 3661-3802, 3668-3957, 3676-4107, 3683-3914, 3683-3962, 3703-4005, 3704-3935, 3704-3937, 3709-3920, 3710-3969, 3710-3975, 3710-4036, 3710-4065, 3710-4105, 3712-3944, 3712-4083, 3724-3968, 3730-4018, 3731-3963, 3738-3997, 3741-4161, 3753-3926, 3753-3988, 3775-4004, 3775-4024, 3781-4011, 3783-4001, 3783-4002, 3784-4018, 3785-4030, 3797-4029, 3801-3961, 3801-4087, 3804-4088, 3807-4043, 3809-4075, 3810-4105, 3814-4072, 3829-4088, 3831-4042, 3843-3960, 3847-4051, 3848-4141, 3867-4103, 3873-4088, 3890-4085, 3894-4236, 3895-4088, 3901-4289, 3906-4131, 3909-4088, 3923-4086, 3937-4067, 3947-4086, 3956-4086, 3956-4088, 3963-4107, 3973-4088, 4003-4240, 4038-4156
317521129CB1/1406	1-893, 398-1406, 555-1405

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
32/7521148CB1/1149	1-715, 2-1149, 376-1149, 377-1149, 603-1149, 1081-1149
33/7521157CB1/808	1-807, 56-808, 300-808, 306-808, 310-808
34/7521195CB1/826	1-639, 18-826
35/7521277CB1/1259	1-793, 1-798, 2-752, 2-765, 355-1259, 387-1259, 568-1259, 573-1259, 616-1259
36/7515701CB1/754	1-717, 2-753, 151-754, 191-754, 254-753, 311-753, 514-754
37/7515527CB1/1090	1-877, 191-1090
38/7519261CB1/1285	1-649, 2-632, 523-1285
39/7519923CB1/891	1-136, 1-295, 1-759, 5-890, 347-891
40/7519986CB1/733	1-728, 1-729, 6-732, 6-733, 199-733
41/7520799CB1/467	1-467
42/7520947CB1/641	1-632, 1-641
43/7521071CB1/661	1-661
44/7521082CB1/588	1-588, 207-588
45/7521329CB1/1368	1-792, 1-797, 2-751, 2-764, 2-880, 445-1368, 569-1368, 625-1368, 1299-1368
46/7522977CB1/217	1-217, 2-216
47/7521939CB1/1152	1-574, 467-1151, 467-1152
48/7522030CB1/918	1-915, 2-917, 125-918, 243-918, 246-918, 364-918
49/7522059CB1/1094	1-659, 2-643, 229-1094
50/7523442CB1/490	1-490, 2-489

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
26	7513276CB1	BRAIFEN03
27	7513290CB1	PROSNON01
29	7513017CB1	SYNOOAT01
30	7513018CB1	SYNOOAT01

Table 6

Library	Vector	Library Description
BRAIFEN03	pINCY	This normalized fetal brain tissue library was constructed from 3.26 million independent clones from a fetal brain library. Starting RNA was made from brain tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart at 23 weeks' gestation. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
PROSNON01	PSPORT1	This normalized prostate library was constructed from 4.4 M independent clones from a prostate library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.
SYNOOAT01	PSPORT1	Library was constructed using RNA isolated from the knee synovial membrane tissue of an 82-year-old female with osteoarthritis.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-	Probability value = 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribkov, M. et al. (1988) CABIOS 4:61-66; Gribkov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.

Table 7

Program	Description	Reference	Parameter Threshold
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
29	7513017	078325H1	SNP00031171	110	125	C	C	A	R32	n/d	n/a	n/a	n/a
29	7513017	1261306H1	SNP00061665	203	3783	C	C	T	A1251	n/d	n/d	n/d	1
29	7513017	1262049H1	SNP00009620	101	2767	G	A	G	P912	0.61	0.85	0.92	0.5
29	7513017	1264949R1	SNP00108532	169	2246	C	C	T	P739	n/d	n/a	n/a	n/a
29	7513017	1297702T6	SNP00061665	76	3820	C	C	T	S1263	n/d	n/d	n/d	1
29	7513017	190325F1	SNP00064566	340	607	A	A	G	K192	n/a	n/a	n/a	n/a
29	7513017	1986384T6	SNP00061665	111	3784	C	C	T	A1251	n/d	n/d	n/d	1
29	7513017	2583613T6	SNP00061665	152	3807	C	C	T	A1259	n/d	n/d	n/d	1
29	7513017	2586205T6	SNP00061665	103	3787	C	C	T	S1252	n/d	n/d	n/d	1
29	7513017	7751923H1	SNP00064566	144	677	A	A	G	T216	n/a	n/a	n/a	n/a
29	7513017	7753336H1	SNP00108532	238	2254	T	C	T	D741	n/d	n/a	n/a	n/a
30	7513018	1261306H1	SNP00061665	203	3932	C	C	T	A1292	n/d	n/d	n/d	1
30	7513018	1262049H1	SNP00009620	101	2916	G	A	G	P953	0.61	0.85	0.92	0.5
30	7513018	1264949R1	SNP00108532	169	2395	C	C	T	P780	n/d	n/a	n/a	n/a
30	7513018	1297702T6	SNP00061665	76	3969	C	C	T	S1304	n/d	n/d	n/d	1
30	7513018	190325F1	SNP00064566	340	756	A	A	G	K233	n/a	n/a	n/a	n/a
30	7513018	1986384T6	SNP00061665	111	3933	C	C	T	A1292	n/d	n/d	n/d	1
30	7513018	2583613T6	SNP00061665	152	3956	C	C	T	A1300	n/d	n/d	n/d	1
30	7513018	2586205T6	SNP00061665	103	3936	C	C	T	S1293	n/d	n/d	n/d	1
30	7513018	7751923H1	SNP00064566	144	826	A	A	G	T257	n/a	n/a	n/a	n/a
30	7513018	7753336H1	SNP00108532	238	2403	T	C	T	D782	n/d	n/a	n/a	n/a
31	7521129	1234366H1	SNP00045824	242	921	C	C	T	T263	n/d	n/d	n/d	n/d
31	7521129	1253806H1	SNP00128439	138	724	T	T	C	S197	n/a	n/a	n/a	n/a
31	7521129	1450242H1	SNP00045823	216	475	G	G	C	V114	n/a	n/a	n/a	n/a
31	7521129	7619032J1	SNP00045823	501	476	G	G	C	E115	n/a	n/a	n/a	n/a
31	7521129	7619032J1	SNP00128439	252	725	T	T	C	F198	n/a	n/a	n/a	n/a
31	7521129	8610819H1	SNP00128438	197	175	C	C	T	P14	n/a	n/a	n/a	n/a
31	7521129	8610819J1	SNP00128438	304	162	C	C	T	P10	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
34	7521195	1313357H1	SNP00061075	9	76	C	C	T	noncoding	n/a	n/a	n/a	n/a
34	7521195	623354R6	SNP00061075	236	78	C	C	T	noncoding	n/a	n/a	n/a	n/a
36	7515701	3200745F6	SNP00002356	36	167	C	C	A	L47	0.36	n/a	n/a	n/a
36	7515701	3704691H1	SNP00116759	10	430	C	C	T	noncoding	n/a	n/a	n/a	n/a
37	7515527	1339038H1	SNP00149747	17	619	C	C	T	T155	n/a	n/a	n/a	n/a
37	7515527	476685H1	SNP00149746	268	563	G	G	A	Q136	n/a	n/a	n/a	n/a
39	7519923	1623933T6	SNP00108625	316	793	G	G	T	G253	n/a	n/a	n/a	n/a
39	7519923	1723428H1	SNP00108625	88	817	G	G	T	C261	n/a	n/a	n/a	n/a
39	7519923	2212164H1	SNP00069043	221	213	C	C	T	P60	n/d	n/d	n/d	n/d
39	7519923	3984515H1	SNP00037991	164	292	A	A	G	K86	n/a	n/a	n/a	n/a
39	7519923	4509565H1	SNP00141815	268	733	C	C	G	A233	n/a	n/a	n/a	n/a
39	7519923	608210T6	SNP00108625	323	813	G	G	T	A260	n/a	n/a	n/a	n/a
39	7519923	609033T6	SNP00108625	326	816	G	G	T	G261	n/a	n/a	n/a	n/a
40	7519986	2987970H1	SNP00000986	254	228	A	G	A	R67	0.83	n/a	n/a	n/a
40	7519986	4145262H1	SNP00135659	48	327	C	C	A	P100	n/a	n/a	n/a	n/a
40	7519986	6207907H1	SNP00132276	47	298	T	T	C	L90	n/a	n/a	n/a	n/a
42	7520947	1255387H1	SNP00001124	194	602	G	G	T	noncoding	0.86	n/a	n/a	n/a
42	7520947	1321153H1	SNP00054045	165	399	A	A	G	R119	n/a	n/a	n/a	n/a
42	7520947	2520143H1	SNP00016888	93	33	G	G	C	noncoding	n/a	n/a	n/a	n/a
43	7521071	1313357H1	SNP00061075	9	77	C	C	T	noncoding	n/a	n/a	n/a	n/a
43	7521071	623354R6	SNP00061075	236	79	C	C	T	noncoding	n/a	n/a	n/a	n/a
44	7521082	1631043F6	SNP00064280	195	379	A	A	G	S45	0.47	0.82	0.25	0.51
48	7522030	1379161F1	SNP00107952	363	166	C	C	G	Q52	n/d	n/d	n/d	n/d
49	7522059	2474784H1	SNP00123695	162	456	C	C	T	A148	0.95	0.92	0.91	0.91
49	7522059	2474827H1	SNP00123694	114	132	A	A	G	Q40	n/d	n/d	n/d	n/d
49	7522059	7019852H1	SNP00123694	80	133	G	A	G	Q40	n/d	n/d	n/d	n/d
49	7522059	7019852H1	SNP00123695	403	457	T	C	T	A148	0.95	0.92	0.91	0.91
50	7523442	1601359H1	SNP00020851	1	277	C	C	T	H91	0.89	0.87	0.84	0.95

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
50	7523442	1862257F7	SNP00020851	63	290	T	C	T	stop96	0.89	0.87	0.84	0.95
50	7523442	5979622H1	SNP00128770	46	475	A	A	G	noncoding	n/a	n/a	n/a	n/a

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

- a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25,
- b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:7-8, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:21, and SEQ ID NO:23,
- c) a polypeptide comprising a naturally occurring amino acid sequence at least 91% identical to the amino acid sequence of SEQ ID NO:24,
- d) a polypeptide comprising a naturally occurring amino acid sequence at least 92% identical to the amino acid sequence of SEQ ID NO:9,
- e) a polypeptide comprising a naturally occurring amino acid sequence at least 94% identical to the amino acid sequence of SEQ ID NO:6,
- f) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:12 and SEQ ID NO:19,
- g) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to the amino acid sequence of SEQ ID NO:4,
- h) a polypeptide consisting essentially of a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:5, SEQ ID NO:10, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22, and SEQ ID NO:25,
- i) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, and
- j) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-25.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a
5 polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

10

9. A method of producing a polypeptide of claim 1, the method comprising:

- a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
15 b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-25.

20

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:26-50,
25 b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:33 and SEQ ID NO:44,
c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 91% identical to the polynucleotide sequence of SEQ ID NO:38,
30 d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 94% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:31, SEQ ID NO:34, and SEQ ID NO:49,
e) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 97% identical to the polynucleotide sequence of SEQ ID NO:37,
35

- f) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 98% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:27-29,
- g) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 99% identical to the polynucleotide sequence of SEQ ID NO:30,
- h) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence selected from the group consisting of SEQ ID NO:26, SEQ ID NO:32, SEQ ID NO:35-36, SEQ ID NO:39-43, SEQ ID NO:45-48 and SEQ ID NO:50,
- i) a polynucleotide complementary to a polynucleotide of a),
- j) a polynucleotide complementary to a polynucleotide of b),
- k) a polynucleotide complementary to a polynucleotide of c),
- l) a polynucleotide complementary to a polynucleotide of d),
- m) a polynucleotide complementary to a polynucleotide of e),
- n) a polynucleotide complementary to a polynucleotide of f),
- o) a polynucleotide complementary to a polynucleotide of g),
- p) a polynucleotide complementary to a polynucleotide of h), and
- q) an RNA equivalent of a)-p).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide

having a sequence of a polynucleotide of claim 12, the method comprising:

- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-25.

19. A method for treating a disease or condition associated with decreased expression of functional EXMES, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) contacting a sample comprising a polypeptide of claim 1 with a compound, and
- b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional EXMES, comprising administering to a patient in need of such treatment a composition of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) contacting a sample comprising a polypeptide of claim 1 with a compound, and
- b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional EXMES, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:
- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
 - b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

10

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- 15 b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

20

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

25

- a) contacting a sample comprising the target polynucleotide with a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

30

29. A method of screening for potential toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions

35

whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample indicates potential toxicity of the test compound.

30. A method for a diagnostic test for a condition or disease associated with the expression of EXMES in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')₂ fragment, or
- e) a humanized antibody.

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of EXMES in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

34. A composition of claim 32, further comprising a label.

35. A method of diagnosing a condition or disease associated with the expression of EXMES in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from the animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-25, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and
- 5 b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-25 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID
15 NO:1-25.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

20 47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a
25 hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first
30 oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is
35 completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is
5 completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a
10 nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

15 55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

20

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

25 58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

30

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

35 63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.
65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.
- 5 66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.
67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
- 10 69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
- 15 71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
- 20 74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.
75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.
- 25 76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.
77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.
78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.
- 30 79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.
80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.
- 35 81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:26.

82. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:27.

5 83. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:28.

84. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:29.

10 85. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:30.

15 86. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:31.

87. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:32.

20 88. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:33.

89. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:34.

25 90. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:35.

30 91. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:36.

92. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:37.

35 93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.

94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.

5 95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.

96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.

10 97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.

15 98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.

99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.

20 100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:45.

101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:46.

25 102. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:47.

30 103. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:48.

104. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:49.

35 105. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:50.

<110> INCYTE CORPORATION; ELLIOTT, Vicki S.;
 KHARE, Reena; TRAN, Uyen K.;
 SWARNAKAR, Anita; MARQUIS, Joseph P.;
 RICHARDSON, Thomas W.; EMERLING, Brooke M.;
 LINDQUIST, Erika A.; CHAWLA, Narinder K.;
 RAMKUMAR, Jayalaxmi; LEE, Soo Yeun

<120> EXTRACELLULAR MESSENGERS

<130> PF-1501 PCT

<140> To Be Assigned

<141> Herewith

<150> US 60/400,810

<151> 2002-08-02

<150> US 60/412,197

<151> 2002-09-19

<150> US 60/416,004

<151> 2002-10-04

<150> US 60/424,862

<151> 2002-11-08

<160> 10

<170> PERL Program

<210> 1

<211> 177

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7513276CD1

<400> 1

Met	Lys	Val	Leu	Leu	Cys	Asp	Leu	Leu	Leu	Leu	Ser	Leu	Phe	Ser
1				5					10					15
Ser	Val	Phe	Ser	Ser	Cys	Gln	Arg	Asp	Cys	Leu	Thr	Cys	Gln	Glu
				20					25					30
Lys	Leu	His	Pro	Ala	Leu	Asp	Ser	Phe	Asp	Leu	Glu	Val	Cys	Ile
				35					40					45
Leu	Glu	Cys	Glu	Glu	Lys	Val	Phe	Pro	Ser	Pro	Leu	Trp	Thr	Pro
				50					55					60
Cys	Thr	Lys	Val	Met	Ala	Arg	Ser	Ser	Trp	Gln	Leu	Ser	Pro	Ala
				65					70					75
Ala	Pro	Glu	His	Val	Ala	Ala	Ala	Leu	Tyr	Gln	Pro	Arg	Ala	Ser
				80					85					90
Glu	Met	Gln	His	Leu	Arg	Arg	Met	Pro	Arg	Val	Arg	Ser	Leu	Phe
				95					100					105
Gln	Glu	Gln	Glu	Glu	Pro	Glu	Pro	Gly	Met	Glu	Glu	Ala	Gly	Glu
				110					115					120
Met	Glu	Gln	Lys	Gln	Leu	Gln	Lys	Arg	Phe	Gly	Gly	Phe	Thr	Gly
				125					130					135
Ala	Arg	Lys	Ser	Ala	Arg	Lys	Leu	Ala	Asn	Gln	Lys	Arg	Phe	Ser
				140					145					150
Val	Gln	Val	Ile	Pro	Gln	Thr	Ala	Cys	Ala	Gln	Thr	Gln	Thr	Cys
				155					160					165
Arg	Leu	Gly	Ile	Arg	Ile	Pro	Ser	Ser	Pro	Arg	His			
				170					175					

<210> 2
 <211> 524
 <212> PRT
 <213> Homo sapiens

 <220>
 <221> misc_feature
 <223> Incyte ID No: 7513290CD1

<400> 2
 Met Gly Ser Arg Cys Ala Leu Ala Leu Ala Val Leu Ser Ala Leu
 1 5 10 15
 Leu Cys Gln Val Trp Ser Ser Gly Val Phe Glu Leu Lys Leu Gln
 20 25 30
 Glu Phe Val Asn Lys Lys Gly Leu Leu Gly Asn Arg Asn Cys Cys
 35 40 45
 Arg Gly Gly Ala Gly Pro Pro Pro Cys Ala Cys Arg Thr Phe Phe
 50 55 60
 Arg Val Cys Leu Lys His Tyr Gln Ala Ser Val Ser Pro Glu Pro
 65 70 75
 Pro Cys Thr Tyr Gly Ser Ala Val Thr Pro Val Leu Gly Val Asp
 80 85 90
 Ser Phe Ser Leu Pro Asp Gly Gly Gly Ala Asp Ser Ala Phe Ser
 95 100 105
 Asn Pro Ile Arg Phe Pro Phe Gly Phe Thr Trp Pro Gly Thr Phe
 110 115 120
 Ser Leu Ile Ile Glu Ala Leu His Thr Asp Ser Pro Asp Asp Leu
 125 130 135
 Ala Thr Glu Asn Pro Glu Arg Leu Ile Ser Arg Leu Ala Thr Gln
 140 145 150
 Arg His Leu Thr Val Gly Glu Glu Trp Ser Gln Asp Leu His Ser
 155 160 165
 Ser Gly Arg Thr Asp Leu Lys Tyr Ser Tyr Arg Phe Val Cys Asp
 170 175 180
 Glu His Tyr Tyr Gly Glu Gly Cys Ser Val Phe Cys Arg Pro Arg
 185 190 195
 Asp Asp Ala Phe Gly His Phe Thr Cys Gly Glu Arg Gly Glu Lys
 200 205 210
 Val Cys Asn Pro Gly Trp Lys Gly Pro Tyr Cys Thr Glu Pro Ile
 215 220 225
 Cys Leu Pro Gly Cys Asp Glu Gln His Gly Phe Cys Asp Lys Pro
 230 235 240
 Gly Glu Cys Lys Cys Arg Val Gly Trp Gln Gly Arg Tyr Cys Asp
 245 250 255
 Glu Cys Ile Arg Tyr Pro Gly Cys Leu His Gly Thr Cys Gln Gln
 260 265 270
 Pro Trp Gln Cys Asn Cys Gln Glu Gly Trp Gly Gly Leu Phe Cys
 275 280 285
 Asn Gln Asp Leu Asn Tyr Cys Thr His His Lys Pro Cys Lys Asn
 290 295 300
 Gly Ala Thr Cys Thr Asn Thr Gly Gln Gly Ser Tyr Thr Cys Ser
 305 310 315
 Cys Arg Pro Gly Tyr Thr Gly Ala Thr Cys Glu Leu Gly Ile Asp
 320 325 330
 Glu Cys Asp Pro Ser Pro Cys Lys Asn Gly Gly Ser Cys Thr Asp
 335 340 345
 Leu Glu Asn Ser Tyr Ser Cys Thr Cys Pro Pro Gly Phe Tyr Gly
 350 355 360
 Lys Ile Cys Glu Leu Ser Ala Met Thr Cys Ala Asp Gly Pro Cys
 365 370 375
 Phe Asn Gly Gly Arg Cys Ser Asp Ser Pro Asp Gly Gly Tyr Ser
 380 385 390
 Cys Arg Cys Pro Val Gly Tyr Ser Gly Phe Asn Cys Glu Lys Lys

	395		400		405
Ile Asp Tyr Cys	Ser Ser Ser Pro Cys	Ser Asn Gly Ala Lys Cys			
	410		415		420
Val Asp Leu Gly	Asp Ala Tyr Leu Cys	Arg Cys Gln Ala Gly Phe			
	425		430		435
Ser Gly Arg His	Cys Asp Asp Asn Val	Asp Asp Cys Ala Ser Ser			
	440		445		450
Pro Cys Ala Asn	Gly Gly Thr Cys Arg	Asp Gly Val Asn Asp Phe			
	455		460		465
Ser Cys Thr Cys	Pro Pro Gly Tyr Thr	Gly Arg Asn Cys Ser Ala			
	470		475		480
Pro Val Ser Ser	Gly Glu Ala Ser Glu	Arg Lys Arg Pro Asp Ser			
	485		490		495
Gly Cys Ser Thr	Ser Lys Asp Thr Lys	Tyr Gln Ser Val Tyr Val			
	500		505		510
Ile Ser Glu Glu	Lys Asp Glu Cys Val	Ile Ala Thr Glu Val			
	515		520		

<210> 3
 <211> 143
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7513749CD1

<400> 3

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

<210> 4
 <211> 1270
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7513017CD1

<400> 4

Met Ala Trp Lys Thr Leu Pro Ile Tyr Leu Leu Leu Leu Leu Ser	
1 5 10 15	
Val Phe Val Ile Gln Gln Val Ser Ser Gln Glu Leu Ser Cys Lys	
20 25 30	

Gly	Arg	Cys	Phe	Glu	Ser	Phe	Glu	Arg	Gly	Arg	Glu	Cys	Asp	Cys	
				35					40					45	
Asp	Ala	Gln	Cys	Lys	Lys	Tyr	Asp	Lys	Cys	Cys	Pro	Asp	Tyr	Glu	
				50					55					60	
Ser	Phe	Cys	Ala	Glu	Val	Lys	Asp	Asn	Lys	Lys	Asn	Arg	Thr	Lys	
				65					70					75	
Lys	Lys	Pro	Thr	Pro	Lys	Pro	Pro	Val	Val	Asp	Glu	Ala	Gly	Ser	
				80					85					90	
Gly	Leu	Asp	Asn	Gly	Asp	Phe	Lys	Val	Thr	Thr	Pro	Asp	Thr	Ser	
				95					100					105	
Thr	Thr	Gln	His	Asn	Lys	Val	Ser	Thr	Ser	Pro	Lys	Ile	Thr	Thr	
				110					115					120	
Ala	Lys	Pro	Ile	Asn	Pro	Arg	Pro	Ser	Leu	Pro	Pro	Asn	Ser	Asp	
				125					130					135	
Thr	Ser	Lys	Glu	Thr	Ser	Leu	Thr	Val	Asn	Lys	Glu	Thr	Thr	Val	
				140					145					150	
Glu	Thr	Lys	Glu	Thr	Thr	Thr	Thr	Asn	Lys	Gln	Thr	Ser	Thr	Asp	
				155					160					165	
Gly	Lys	Glu	Lys	Thr	Thr	Ser	Ala	Lys	Glu	Thr	Gln	Ser	Ile	Glu	
				170					175					180	
Lys	Thr	Ser	Ala	Lys	Asp	Leu	Ala	Pro	Thr	Ser	Lys	Val	Leu	Ala	
				185					190					195	
Lys	Pro	Thr	Pro	Lys	Ala	Glu	Thr	Thr	Thr	Lys	Gly	Pro	Ala	Leu	
				200					205					210	
Thr	Thr	Pro	Lys	Glu	Pro	Thr	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	
				215					220					225	
Ser	Thr	Thr	Pro	Lys	Glu	Pro	Thr	Pro	Thr	Thr	Ile	Lys	Ser	Ala	
				230					235					240	
Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Thr	Lys	Ser	Ala	
				245					250					255	
Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Thr	Lys	Glu	Pro	
				260					265					270	
Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Thr	Lys	Glu	
				275					280					285	
Pro	Ala	Pro	Thr	Thr	Thr	Lys	Ser	Ala	Pro	Thr	Thr	Pro	Lys	Glu	
				290					295					300	
Pro	Ala	Pro	Thr	Thr	Pro	Lys	Lys	Pro	Ala	Pro	Thr	Thr	Pro	Lys	
				305					310					315	
Glu	Pro	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Thr	Pro	Thr	Thr	Pro	
				320					325					330	
Lys	Glu	Pro	Ala	Pro	Thr	Thr	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Pro	
				335					340					345	
Lys	Glu	Pro	Ala	Pro	Thr	Ala	Pro	Lys	Lys	Pro	Ala	Pro	Thr	Thr	
				350					355					360	
Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro	Thr	
				365					370					375	
Thr	Thr	Lys	Glu	Pro	Ser	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro	
				380					385					390	
Thr	Thr	Thr	Lys	Ser	Ala	Pro	Thr	Thr	Thr	Lys	Glu	Pro	Ala	Pro	
				395					400					405	
Thr	Thr	Thr	Lys	Ser	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ser	Pro	
				410					415					420	
Thr	Thr	Thr	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	
				425					430					435	
Pro	Thr	Thr	Pro	Lys	Lys	Pro	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	
				440					445					450	
Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Thr	Lys	Lys	
				455					460					465	
Pro	Ala	Pro	Thr	Ala	Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Pro	Lys	
				470					475					480	
Glu	Thr	Ala	Pro	Thr	Thr	Pro	Lys	Lys	Leu	Thr	Pro	Thr	Thr	Pro	
				485					490					495	
Glu	Lys	Leu	Ala	Pro	Thr	Thr	Pro	Glu	Lys	Pro	Ala	Pro	Thr	Thr	

5/14

```

Arg Pro His Val Phe Met Pro Glu Val Thr Pro Asp Met Asp Tyr
          980          985          990
Leu Pro Arg Val Pro Asn Gln Gly Ile Ile Ile Asn Pro Met Leu
          995          1000          1005
Ser Asp Glu Thr Asn Ile Cys Asn Gly Lys Pro Val Asp Gly Leu
          1010          1015          1020
Thr Thr Leu Arg Asn Gly Thr Leu Val Ala Phe Arg Gly His Tyr
          1025          1030          1035
Phe Trp Met Leu Ser Pro Phe Ser Pro Pro Ser Pro Ala Arg Arg
          1040          1045          1050
Ile Thr Glu Val Trp Gly Ile Pro Ser Pro Ile Asp Thr Val Phe
          1055          1060          1065
Thr Arg Cys Asn Cys Glu Gly Lys Thr Phe Phe Phe Lys Asp Ser
          1070          1075          1080
Gln Tyr Trp Arg Phe Thr Asn Asp Ile Lys Asp Ala Gly Tyr Pro
          1085          1090          1095
Lys Pro Ile Phe Lys Gly Phe Gly Gly Leu Thr Gly Gln Ile Val
          1100          1105          1110
Ala Ala Leu Ser Thr Ala Lys Tyr Lys Asn Trp Pro Glu Ser Val
          1115          1120          1125
Tyr Phe Phe Lys Arg Gly Gly Ser Ile Gln Gln Tyr Ile Tyr Lys
          1130          1135          1140
Gln Glu Pro Val Gln Lys Cys Pro Gly Arg Arg Pro Ala Leu Asn
          1145          1150          1155
Tyr Pro Val Tyr Gly Glu Thr Thr Gln Val Arg Arg Arg Arg Phe
          1160          1165          1170
Glu Arg Ala Ile Gly Pro Ser Gln Thr His Thr Ile Arg Ile Gln
          1175          1180          1185
Tyr Ser Pro Ala Arg Leu Ala Tyr Gln Asp Lys Gly Val Leu His
          1190          1195          1200
Asn Glu Val Lys Val Ser Ile Leu Trp Arg Gly Leu Pro Asn Val
          1205          1210          1215
Val Thr Ser Ala Ile Ser Leu Pro Asn Ile Arg Lys Pro Asp Gly
          1220          1225          1230
Tyr Asp Tyr Tyr Ala Phe Ser Lys Asp Gln Tyr Tyr Asn Ile Asp
          1235          1240          1245
Val Pro Ser Arg Thr Ala Arg Ala Ile Thr Thr Arg Ser Gly Gln
          1250          1255          1260
Thr Leu Ser Lys Val Trp Tyr Asn Cys Pro
          1265          1270

```

<210> 5

<211> 1311

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7513018CD1

<400> 5

```

Met Ala Trp Lys Thr Leu Pro Ile Tyr Leu Leu Leu Leu Leu Ser
  1          5          10          15
Val Phe Val Ile Gln Gln Val Ser Ser Gln Asp Leu Ser Ser Cys
          20          25          30
Ala Gly Arg Cys Gly Glu Gly Tyr Ser Arg Asp Ala Thr Cys Asn
          35          40          45
Cys Asp Tyr Asn Cys Gln His Tyr Met Glu Cys Cys Pro Asp Phe
          50          55          60
Lys Arg Val Cys Thr Ala Glu Leu Ser Cys Lys Gly Arg Cys Phe
          65          70          75
Glu Ser Phe Glu Arg Gly Arg Glu Cys Asp Cys Asp Ala Gln Cys
          80          85          90

```

Lys	Lys	Tyr	Asp	Lys	Cys	Cys	Pro	Asp	Tyr	Glu	Ser	Phe	Cys	Ala
				95					100					105
Glu	Val	Lys	Asp	Asn	Lys	Lys	Asn	Arg	Thr	Lys	Lys	Lys	Pro	Thr
				110					115					120
Pro	Lys	Pro	Pro	Val	Val	Asp	Glu	Ala	Gly	Ser	Gly	Leu	Asp	Asn
				125					130					135
Gly	Asp	Phe	Lys	Val	Thr	Thr	Pro	Asp	Thr	Ser	Thr	Thr	Gln	His
				140					145					150
Asn	Lys	Val	Ser	Thr	Ser	Pro	Lys	Ile	Thr	Thr	Ala	Lys	Pro	Ile
				155					160					165
Asn	Pro	Arg	Pro	Ser	Leu	Pro	Pro	Asn	Ser	Asp	Thr	Ser	Lys	Glu
				170					175					180
Thr	Ser	Leu	Thr	Val	Asn	Lys	Glu	Thr	Thr	Val	Glu	Thr	Lys	Glu
				185					190					195
Thr	Thr	Thr	Thr	Asn	Lys	Gln	Thr	Ser	Thr	Asp	Gly	Lys	Glu	Lys
				200					205					210
Thr	Thr	Ser	Ala	Lys	Glu	Thr	Gln	Ser	Ile	Glu	Lys	Thr	Ser	Ala
				215					220					225
Lys	Asp	Leu	Ala	Pro	Thr	Ser	Lys	Val	Leu	Ala	Lys	Pro	Thr	Pro
				230					235					240
Lys	Ala	Glu	Thr	Thr	Thr	Lys	Gly	Pro	Ala	Leu	Thr	Thr	Pro	Lys
				245					250					255
Glu	Pro	Thr	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Ser	Thr	Thr	Pro
				260					265					270
Lys	Glu	Pro	Thr	Pro	Thr	Thr	Ile	Lys	Ser	Ala	Pro	Thr	Thr	Pro
				275					280					285
Lys	Glu	Pro	Ala	Pro	Thr	Thr	Thr	Lys	Ser	Ala	Pro	Thr	Thr	Pro
				290					295					300
Lys	Glu	Pro	Ala	Pro	Thr	Thr	Thr	Lys	Glu	Pro	Ala	Pro	Thr	Thr
				305					310					315
Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Thr	Lys	Glu	Pro	Ala	Pro	Thr
				320					325					330
Thr	Thr	Lys	Ser	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro	Thr
				335					340					345
Thr	Pro	Lys	Lys	Pro	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro
				350					355					360
Thr	Thr	Pro	Lys	Glu	Pro	Thr	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala
				365					370					375
Pro	Thr	Thr	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala
				380					385					390
Pro	Thr	Ala	Pro	Lys	Lys	Pro	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro
				395					400					405
Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Thr	Lys	Glu
				410					415					420
Pro	Ser	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Thr	Lys
				425					430					435
Ser	Ala	Pro	Thr	Thr	Thr	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Thr	Lys
				440					445					450
Ser	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ser	Pro	Thr	Thr	Thr	Lys
				455					460					465
Glu	Pro	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Pro
				470					475					480
Lys	Lys	Pro	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr
				485					490					495
Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Thr	Lys	Lys	Pro	Ala	Pro	Thr
				500					505					510
Ala	Pro	Lys	Glu	Pro	Ala	Pro	Thr	Thr	Pro	Lys	Glu	Thr	Ala	Pro
				515					520					525
Thr	Thr	Pro	Lys	Lys	Leu	Thr	Pro	Thr	Thr	Pro	Glu	Lys	Leu	Ala
				530					535					540
Pro	Thr	Thr	Pro	Glu	Lys	Pro	Ala	Pro	Thr	Thr	Pro	Glu	Glu	Leu
				545					550					555
Ala	Pro	Thr	Thr	Pro	Glu	Glu	Pro	Thr	Pro	Thr	Thr	Pro	Glu	Glu

Pro Ala Pro Thr	560	Ala Pro Asn Thr Pro	565	570
Thr Pro Lys Ala Ala	575	Ala Pro Asn Thr Pro	580	Lys
Glu Pro Ala Pro Thr	590	Pro Ala Pro Thr Thr	595	Pro
Lys Glu Pro Ala Pro	605	Glu Thr Ala Pro Thr	610	Thr
Pro Lys Gly Thr Ala	620	Lys Glu Pro Ala Pro	625	Thr
Thr Pro Lys Lys Pro	635	Leu Ala Pro Thr Thr	640	Thr
Lys Glu Pro Thr Ser	650	Lys Pro Ala Pro Thr	655	Thr
Pro Lys Gly Thr Ala	665	Lys Glu Pro Ala Pro	670	Thr
Thr Pro Lys Glu Pro	680	Pro Lys Gly Thr Ala	685	Pro
Thr Thr Leu Lys Glu	695	Thr Pro Lys Lys Pro	700	Ala
Pro Lys Glu Leu Ala	710	Lys Gly Pro Thr Ser	715	Thr
Thr Ser Asp Lys Pro	725	Pro Lys Glu Thr Ala	730	Pro
Thr Thr Pro Lys Glu	740	Thr Pro Lys Lys Pro	745	Ala
Pro Thr Thr Pro Glu	755	Thr Thr Ser Glu Val	760	Ser
Thr Pro Thr Thr Thr	770	Thr Ile His Lys Ser	775	Pro
Asp Glu Ser Thr Pro	785	Glu Pro Thr Pro Lys	790	Ala
Leu Glu Asn Ser Pro	800	Val Pro Thr Thr Lys	805	Thr
Pro Ala Ala Thr Lys	815	Thr Thr Ala Lys Asp	820	Lys
Thr Thr Glu Arg Asp	830	Pro Glu Thr Thr Thr	835	Ala
Ala Pro Lys Met Thr	845	Thr Thr Thr Glu Lys	850	Thr
Thr Glu Ser Lys Ile	860	Thr Gln Val Thr Ser	865	Thr
Thr Thr Gln Asp Thr	875	Ile Thr Thr Leu Lys	880	Thr
Thr Thr Leu Ala Pro	890	Thr Lys Lys Thr Ile	895	Thr
Thr Thr Glu Ile Met	905	Glu Thr Ala Lys Pro	910	Lys
Asp Arg Ala Thr Asn	920	Thr Pro Lys Pro Gln	925	Lys
Pro Thr Lys Ala Pro	935	Ser Thr Lys Lys Pro	940	Lys
Thr Met Pro Arg Val	950	Thr Thr Pro Thr Pro	955	Arg
Lys Met Thr Ser Thr	965	Asn Pro Thr Ser Arg	970	Ile
Ala Glu Ala Met Leu	980	Arg Pro Asn Gln Thr	985	Pro
Asn Ser Lys Leu Val	995	Ser Glu Asp Ala Gly	1000	
Gly Ala Glu Gly Glu	1010	Leu Leu Arg Pro His	1015	Val
Phe Met Pro Glu Val	1025	Asp Tyr Leu Pro Arg	1030	Val
				1035

Pro Asn Gln Gly Ile Ile Ile Asn Pro Met Leu Ser Asp Glu Thr
 1040 1045 1050
 Asn Ile Cys Asn Gly Lys Pro Val Asp Gly Leu Thr Thr Leu Arg
 1055 1060 1065
 Asn Gly Thr Leu Val Ala Phe Arg Gly His Tyr Phe Trp Met Leu
 1070 1075 1080
 Ser Pro Phe Ser Pro Pro Ser Pro Ala Arg Arg Ile Thr Glu Val
 1085 1090 1095
 Trp Gly Ile Pro Ser Pro Ile Asp Thr Val Phe Thr Arg Cys Asn
 1100 1105 1110
 Cys Glu Gly Lys Thr Phe Phe Phe Lys Asp Ser Gln Tyr Trp Arg
 1115 1120 1125
 Phe Thr Asn Asp Ile Lys Asp Ala Gly Tyr Pro Lys Pro Ile Phe
 1130 1135 1140
 Lys Gly Phe Gly Gly Leu Thr Gly Gln Ile Val Ala Ala Leu Ser
 1145 1150 1155
 Thr Ala Lys Tyr Lys Asn Trp Pro Glu Ser Val Tyr Phe Phe Lys
 1160 1165 1170
 Arg Gly Gly Ser Ile Gln Gln Tyr Ile Tyr Lys Gln Glu Pro Val
 1175 1180 1185
 Gln Lys Cys Pro Gly Arg Arg Pro Ala Leu Asn Tyr Pro Val Tyr
 1190 1195 1200
 Gly Glu Thr Thr Gln Val Arg Arg Arg Arg Phe Glu Arg Ala Ile
 1205 1210 1215
 Gly Pro Ser Gln Thr His Thr Ile Arg Ile Gln Tyr Ser Pro Ala
 1220 1225 1230
 Arg Leu Ala Tyr Gln Asp Lys Gly Val Leu His Asn Glu Val Lys
 1235 1240 1245
 Val Ser Ile Leu Trp Arg Gly Leu Pro Asn Val Val Thr Ser Ala
 1250 1255 1260
 Ile Ser Leu Pro Asn Ile Arg Lys Pro Asp Gly Tyr Asp Tyr Tyr
 1265 1270 1275
 Ala Phe Ser Lys Asp Gln Tyr Tyr Asn Ile Asp Val Pro Ser Arg
 1280 1285 1290
 Thr Ala Arg Ala Ile Thr Thr Arg Ser Gly Gln Thr Leu Ser Lys
 1295 1300 1305
 Val Trp Tyr Asn Cys Pro
 1310

<210> 6
 <211> 1014
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7513276CB1

<400> 6
 cacgagggtc atggaggtgg gcaggaagg ggagggctgt gcgtggggga gaggatatat 60
 atgctgggtg ggctgagaaa gcggaaccga gcctcgcatc catcgaggag agccggggac 120
 tgacagctct cagcacctgc ttctgtctcc tgcacatga aagtcctgct ttgtgacctg 180
 ctgctgctca gtctcttctc cagtgtgttc agcagttgtc agagggactg tctcacatgc 240
 caggagaagc tccaccagc cctggacagc ttcgacctgg aggtgtgcat cctcgagtgt 300
 gaagagaagg tcttccccag cccctctctg actccatgca ccaagggtcat ggccaggagc 360
 tcttggcagc tcagccctgc cgccccagag catgtggcgg ctgctctcta ccagccgaga 420
 gcttcggaga tgcagcatct gcggcgaatg ccccgagtcc ggagcttggt ccaggagcag 480
 gaagagccc agcctggcat ggaggaggct ggtgagatgg agcagaagca gctgcagaag 540
 agatttgggg gcttcaccgg ggcccgggaag tcggccagga agttggccaa tcagaagcgg 600
 ttcagcgtcc aggtgatccc ccaaacagca tgtgtcaga cccagacctg ccgcctggga 660
 atcaggattc cttcttcccc aaggcactga gcgcctgcag atcccgcagg cttcgtttgc 720
 ctccagaacc ttcccgctctg attgttctc cccagcccc tggcatgttt caccacaacc 780
 ctgttgctac atcagagtgt atttttgtaa ttcctctagc taccatttca atagccccat 840

```

ctctcctgct caccgcctc ttgccccttc taggggcagg tgaaaggaat aggaaattga 900
acctgggggt ttgacttgcc actgccataa cttgtttgta aaagagctgt tctttttgac 960
tgattgtttt aaacaacgat ttctccatta aacttctact gagcaaatgg ttaa 1014

```

```

<210> 7
<211> 2367
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7513290CB1

```

```

<400> 7
tcgatcccc gcgcgtccgc cgctgttcta aggagagaag tgggggcccc ccaggctcgc 60
gcgtggagcg aagcagcatg ggcagtcggt gcgcgtcggc cctggcggtg ctctcggcct 120
tgctgtgtca ggtctggagc tctgggggtg tcgaactgaa gctgcaggag ttcgtcaaca 180
agaaggggct gctgggggaa cgcaactgct gccgcggggg cgcggggcca ccgccgtgcg 240
cctgccggac cttcttccgc gtgtgcctca agcactacca ggccagcgtg tcccccgagc 300
cgccctgcac ctacggcagc gccgtcacc ccgtgctggg cgtcgactcc ttcagtctgc 360
ccgacggcgg gggcgccgac tccgcgttca gcaaccccat ccgcttcccc ttcggcttca 420
cctggccggg cacttctct ctgattattg aagctctcca cacagattct cctgatgacc 480
tcgcaacaga aaaccagaa agactcatca gccgcctggc caccagaggg cacctgacgg 540
tgggcgagga gtggtcccag gacctgcaca gcagcgccg cacggacctc aagtactcct 600
accgcttcgt gtgtgacgaa cactactacg gagaggctg tccggttttc tgccgtcccc 660
gggacgatgc cttcggccac ttcacctgtg gggagcgtgg ggagaaagtg tgcaaccctg 720
gctggaaagg gccctactgc acagagccga tctgcctgcc tggatgtgat gagcagcatg 780
gattttgtga caaacaggg gaatgcaagt gcagagtggg ctggcagggc cgggtactgtg 840
acgagtgtat ccgctatcca ggctgtctcc atggcacctg ccagcagccc tggcagtgca 900
actgccagga aggctggggg ggccttttct gcaaccagga cctgaactac tgcacacacc 960
ataagccctg caagaatgga gccacctgca ccaacacggg ccagggggagc tacacttgct 1020
cttgccggcc tgggtacaca ggtgccacct gcgagctggg gattgacgag tgtgacccc 1080
gcccttgtaa gaacggaggg agctgcacgg atctcgagaa cagctactcc tgtacctgcc 1140
caccggctt ctacggcaaa atctgtgaat tgagtgccat gacctgtgcg gacggccctt 1200
gctttaacgg gggctcgtgc tcagacagcc ccgatggagg gtacagctgc cgctgccccg 1260
tgggctactc cggcttcaac tgtgagaaga aaattgacta ctgcagctct tcacctgtt 1320
ctaattggtg caagtgtgtg gacctcgtg atgcctacct gtgccgctgc caggccggct 1380
tctcggggag gcaactgtgac gacaacgtgg acgactgcgc ctctcccccg tgcgccaacg 1440
ggggcacctg ccgggatggc gtgaacgact tctcctgcac ctgcccgcct ggctacacgg 1500
gcaggaactg cagtgcctcc gtcagcagtg gagaagcatc tgaaagaaaa aggccggact 1560
cgggctgttc aacttcaaaa gacaccaagt accagtcggt gtacgtcata tccgaggaga 1620
aggatgagtg cgtcatagca actgaggtgt aaaatggaag tgagatggca agactcccgt 1680
ttctcttaaa ataagtaaaa ttccaaggat atatgcccc acaaatgctg ctgaagagga 1740
gggaggcctc gtggactgct gctgagaaac cgagttcaga ccgagcaggt tctcctcctg 1800
aggctctcga cgcctgccga cagcctgtcg cggcccgcc gcctgcggca ctgccttccg 1860
tgacgtcgcc gttgcactat ggacagttgc tcttaagaga atatatattt aaatgggtga 1920
actgaattac gcataagaag catgcactgc ctgagtgtat attttgatt cttatgagcc 1980
agtcttttct tgaattagaa acacaaacac tgcccttatt gtcctttttg atacgaagat 2040
gtgcttttct tagatggaaa agatgtgtgt tttttttgg atttgtaaaa atatttttca 2100
tgatatctgt aaagcttgag ttttttgtag tgttcgtttt ttataattta aatttttggt 2160
aatatgtaca aaggcacttc gggctctatg gactatattt ttttgtatat aaatgtattt 2220
atggaatatt gtgcaaatgt tttttgagtt ttttactgtt ttgttaatga agaaattcct 2280
ttttaaaata tttttccaaa ataaatttta tgaatgacaa aaaaaaaaaa aaaccacgcg 2340
tccgcccacg cgtccgccca ccgtccg 2367

```

```

<210> 8
<211> 1874
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7513749CB1

```

<400> 8

```

ttgctatctc tctcgatcac tccctccctt cctcccgcgc ttcctcccgg cggcgcgggc 60
gcgctgggga agcgggtgaag aggagtggcc cggccctgga agaattgcggc tctgacaagg 120
ggacagaacc cagcgcagtc tccccacggt ttaagcagca ctagtgaagc ccaggcaacc 180
caaccgtgcc tgtctcggac cccgcaccca aaccactgga ggtcctgac gatctgcca 240
ccggagcctc cgggcttcga catgctggag gagccccggc cgcggcctcc gccctcgggc 300
ctcgcgggtc tcctgttcct ggcgttgtgc agtcgggctc taagcaatga gattctgggc 360
ctgaagtgtc ctggcgagcc gccgtgacg gccaacaccg tgtgcttgac gctgtccggc 420
ctgagcaagc ggcagctagg cctgtgcctg cgcaaccccg acgtgacggc gtccgcgctt 480
cagggctcgc acatcgcggt ccacgagtgt cagcaccagc tgcgcgacca gcgctggaac 540
tgctccgcgc ttgagggcgg cggccgcctg ccgcaccaca gcgccatcct caagcgcggg 600
cctgggtgtg gggaaaggctt ggacacccaa atgggtttcc agaaagtgtc ttttcttct 660
ccatgctggc tgctggggtc atgcacgcag tagccacggc ctgcagcctg ggcaagctgg 720
tgagctgtgg ctgtggctgg aagggcagtg gtgagcagga tcggctgagg gccaaactgc 780
tgacgctgca ggcactgtcc cgaggcaaga gtttccccca ctctctgccc agccctggcc 840
ctggctcaag cccagccctt ggcctccagg acacatggga atgggggtggc tghtaaccatg 900
acatggactt tggagagaag ttctctcggg atttcttggg ttccaggga gctccccggg 960
acatccaggc acgaatgcga atccacaaca acaggggtgg gcgccagggt gtaactgaaa 1020
acctgaagcg gaaatgcaag tgtcatggca catcaggcag ctgccagttc aagacatgct 1080
ggagggcggc cccagagttc cgggcagtg gggcggcggt gagggagcgg ctgggcccgg 1140
ccatcttcat tgataccac aaccgcaatt ctggagcctt ccagccccgt ctgcgtcccc 1200
gtcgcctctc aggagagctg gtctactttg agaagtctcc tgacttctgt gagcgagacc 1260
ccactatggg gtcctccagg gacaagggcc agggcctgca acaagaccag ccgcctgttg 1320
gatggctgtg gcagcctgtg ctgtggccgt gggcacaacg tgctccggca gacacgagtt 1380
gagcgtgcc attgccgctt cactgtgtgc tgctatgtgc tgtgtgatga gtgcaaggtt 1440
acagatggg tgaatgtgtg taagttaggg cttggggact ggggaagagg 1500
actgtgtgag agggcgctt tttcagcct ttgctctgat ttcttccaa ggtcactctt 1560
gggtcctgga agcttaaagt atctacctgg aaacagcttt aggggtggtg ggggtcaggt 1620
ggactctggg atgtgtagcc ttctcccaa caattggagg gtcttgaggg gaagctgcca 1680
cccctcttca tgctccttag acacctgaat ggactaagat gaaatgcact gtattgctcc 1740
tcccacttct caactccagg gcccctttaa gcctgattca tactcctttt ggctggggag 1800
tccttatagt ttcaccactc ctctcccttg agggataacc ccaggcactg tttggagcca 1860
taagatctgt atct 1874

```

<210> 9

<211> 4140

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7513017CB1

<400> 9

```

gcaagggtac ctacgggtacc tgaaaacaac gatggcatgg aaaacacttc ccatttacct 60
gttggtgtctg ctgtctgttt tcgtgattca gcaagtttca tctcaagagc tttcctgtaa 120
aggccgctgc tttgagtcct tcgagagagg gagggagtgt gactgcgacg cccaatgtaa 180
gaagtatgac aagtgtgtgc ccgattatga gagtttctgt gcagaagtaa aagataacaa 240
gaagaacaga actaaaaaga aacctacccc caaacacca gttgtagatg aagctggaag 300
tggaattggac aatgggtgact tcaagggtcac aactcctgac acgtctacca cccaacacaa 360
taaagtgcgc acatctccca agatcacaa agcaaaacca ataaatccca gaccaggtct 420
tccacctaat tctgatacat cttaaagagac gtctttgaca gtgaataaag agacaacagt 480
tgaaactaaa gaaactacta caacaaataa acagacttca actgatggaa aagagaagac 540
tacttccgct aaagagacac aaagtataga gaaaacatct gctaaagatt tagcaccac 600
atctaaagtg ctggctaaac ctacacccaa agctgaaact acaaccaaag gccctgctct 660
caccactccc aaggagccca cgcaccacc tcccaaggag cctgcatcta ccacacccaa 720
agagcccaca cctaccacca tcaagtctgc accaccacc cccaaggagc ctgcaccac 780
caccaccaag tctgcaccca cactcccaa cccaccacca ccaaggagcc ccaaggagcc 840
tgacccacc actcccaagg agcctgcacc caccaccacc aaggagcctg caccaccac 900
caccaagtct gcaccacca ctcccaagga gcctgcacc accaccccca agaagcctgc 960
cccaactacc ccaaggagc ctgcaccac cactcccaag gagcctacac ccactactcc 1020
caaggagcct gcaccacca ccaaggagcc tgcaccacc actcccaaag agcctgcacc 1080
cactgcccc aagaagcctg cccaactac cccaaggag cctgcaccca ccactcccaa 1140

```

```

ggagcctgca cccaccacca ccaaggagcc ttcaccacc actcccaagg agcctgcacc 1200
caccaccacc aagtctgcac ccaccactac caaggagcct gacccacca ctaccaagtc 1260
tgcaccacc actcccaagg agccttcacc caccaccacc aaggagcctg caccaccacc 1320
tcccaaggag cctgcaccca ccacccccaa gaagcctgcc ccaactaccc ccaaggagcc 1380
tgcaccacc actcccaagg aacctgcacc caccaccacc aagaagcctg caccaccgcc 1440
tcccaagag cctgccccaa ctacccccaa ggagactgca cccaccaccc ccaagaagct 1500
cacgcccacc acccccgaga agctcgcacc caccaccacct gagaagcccc caccaccacc 1560
ccctgaggag ctgcaccca ccacccctga ggagcccaca cccaccaccc ctgaggagcc 1620
tgctcccacc actcccaagg cagcggctcc caacacccct aaggagcctg ctccaactac 1680
ccctaaggag cctgctccaa ctacccctaa ggagcctgct ccaactaccc ctaaggagac 1740
tgctccaact acccctaaag ggactgctcc aactacccct aaggaacctg caccactac 1800
tcccaagaag cctgccccca aggagcttgc acccaccacc accaaggagc ccacatccac 1860
cacctctgac aagcccgctc caactacccc taaggggact gctccaacta cccctaagga 1920
gcctgctcca actaccccta aggagcctgc tccaactacc cctaagggga ctgctccaac 1980
taccctcaag gaacctgcac ccactactcc caagaagcct gcccccgaag agcttgcaac 2040
caccaccacc aaggggcccc catccaccac ctctgacaag cctgctccaa ctacacctaa 2100
ggagactgct ccaactaccc ccaaggagcc tgcaccact acccccgaag agcctgctcc 2160
aactactcct gagacacctc ctccaaccac ttcagagggt tctactccaa ctaccacca 2220
ggagcctacc actatccaca aaagccctga tgaatcaact cctgagcttt ctgcagaacc 2280
cacaccaaaa gctcttgaaa acagtcccaa ggaacctggt gtacctacaa ctaagactcc 2340
tgcagcgact aaacctgaaa tgactacaac agctaaagac aagacaacag aaagagactt 2400
acgtactaca cctgaaacta caactgctgc acctaagatg acaaaagaga cagcaactac 2460
aacagaaaaa actaccgaat ccaaaataac agctacaacc acacaagtaa catctaccac 2520
aactcaagat accacaccat tcaaaattac tactcttaaa acaactactc ttgcacccaa 2580
agtaactaca acaaaaaaga caattactac cactgagatt atgaacaaac ctgaagaaac 2640
agctaaacca aaagacagag ctactaattc taaagcgaca actcctaaac ctcaaaagcc 2700
aaccaaagca ccaaaaaaac ccacttctac caaaaagcca aaaacaatgc ctagagttag 2760
aaaaccaaag acgacaccaa ctccccgcaa gatgacatca acaatgccag aattgaaccc 2820
tacctcaaga atagcagaag ccatgctcca aaccaccacc agacctaac aaactccaaa 2880
ctccaaacta gttgaagtaa atccaaagag tgaagatgca ggtggtgctg aaggagaaac 2940
acctcatatg cttctcaggc cccatgtgtt catgctgaa gttactcccg acatggatta 3000
cttaccgaga gtacccaatc aaggcattat catcaatccc atgctttccg atgagacca 3060
tatatgcaat ggtaagccag tagatggact gactactttg cgcaatggga cattagttag 3120
attccgaggt cattatttct ggatgctaag tccattcagt ccaccatctc cagctcgag 3180
aattactgaa gtttggggta ttccttcccc cattgatact gtttttacta ggtgcaactg 3240
tgaaggaaaa acttttcttct ttaaggattc tcagtactgg cgttttacca atgatataaa 3300
agatgcaggg taccacaaac caattttcaa aggatttga ggactaactg gacaaatagt 3360
ggcagcgctt tcaacagcta aatataagaa ctggcctgaa tctgtgtatt ttttcaagag 3420
aggtggcagc attcagcagt atatttataa acaggaacct gtacagaagt gccctggaag 3480
aaggcctgct ctaaattatc cagtgtatgg agaaacgaca caggttagga gacgtcgctt 3540
tgaacgtgct ataggacctt ctcaaacaca caccatcaga attcaatatt cacctgccag 3600
actggcttat caagacaaag gtgtccttca taatgaagtt aaagttagta tactgtggag 3660
aggacttcca aatgtggtta cctcagctat atcactgccc aacatcagaa aacctgacgg 3720
ctatgattac tatgcctttt cttaaagatca atactataac attgatgtgc ctagtagaac 3780
agcaagagca attactactc gttctgggca gaccttatcc aaagtctggt acaactgtcc 3840
ttagactgat gagcaaagga ggagtcaact aatgaagaaa tgaataataa attttgacac 3900
tgaaaaacat tttattaata aagaatattg acatgagtat accagtttat atataaaaa 3960
gtttttaaac ttgacaatca ttacactaaa acagatttga taatcttatt cacagttgtt 4020
attgtttaca gaccatttaa ttaatatatt ctctgtttat tctcctctc cctccattg 4080
catggctcac acctctctgt ctgttaaact gacgaccacg gaccctcgtg ccgaattctt 4140

```

<210> 10

<211> 4289

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7513018CB1

<400> 10

```

agagagagag agagaactag tctcgagcaa ggtacctac ggtacctgaa aacaacgatg 60
gcatggaaaa cacttcccat ttacctgttg ttgctgctgt ctgttttctg gattcagcaa 120

```

gtttcatctc	aagattttatc	aagctgtgca	gggagatgtg	gggaagggtg	ttctagagat	180
gccacctgca	actgtgatta	taactgtcaa	cactacatgg	agtgtgtccc	tgatttcaag	240
agagtctgca	ctgctggagct	ttcctgtaaa	ggcctgtgct	ttgagtcctt	cgagagaggg	300
agggagtgtg	actgctgacgc	ccaatgtaag	aagtatgaca	agtgtgtccc	cgattatgag	360
agtttctgtg	cagaagtaaa	agataacaag	aagaacagaa	ctaaaaagaa	acctaccccc	420
aaaccaccag	ttgtagatga	agctggaagt	ggattggaca	atgggtgactt	caaggtcaca	480
actcctgaca	cgtctaccac	ccaacacaat	aaagtcagca	catctcccaa	gatcacaaca	540
gcaaaaccaa	taaatcccag	accagtcctt	ccacctaat	ctgatacatc	taaagagacg	600
tctttgacag	tgaataaaga	gacaacagtt	gaaactaaag	aaactactac	aacaaataaa	660
cagacttcaa	ctgatggaaa	agagaaaaact	acttccgcta	aagagacaca	aagtatagag	720
aaaacatctg	ctaaagattt	agcaccacaca	tctaaagtgc	tggctaaacc	tacacccaaa	780
gctgaaacta	caaccaaagg	ccctgtctct	accactccca	aggagccccc	gcccaccact	840
cccaaggagc	ctgcatctac	cacacccaaa	gagcccacac	ctaccaccat	caagtctgca	900
cccaccacc	ccaaggagcc	tgcacccacc	accaccaagt	ctgcacccac	cactcccaag	960
gagcctgcac	ccaccaccac	caaggagcct	gcacccacca	ctcccaagga	gcctgcaccc	1020
accaccacca	aggagcctgc	accaccacc	accaagtctg	caccaccacc	tcccaaggag	1080
ccctgaccca	ccacccccaa	gaagcctgcc	ccaactaccc	ccaaggagcc	tgcacccacc	1140
actcccaagg	agcctacacc	caccactccc	aaggagcctg	caccaccacc	caaggagcct	1200
gcacccacca	ctcccaaaaga	gcctgcaccc	actgccccca	agaagcctgc	cccaactacc	1260
cccaaggagc	ctgcacccac	cactcccaag	gagcctgcac	ccaccaccac	caaggagcct	1320
tacccaccaca	ctcccaagga	gcctgcaccc	accaccacca	agtctgcacc	caccactacc	1380
aaggagcctg	caccaccacc	taccaagtct	gcacccacca	ctcccaagga	gccttcaccc	1440
accaccacca	aggagcctgc	accaccacc	ccaaggagc	ctgcacccac	caccccccaag	1500
aagcctgccc	caactacccc	caaggagcct	gcacccacca	ctcccaagga	acctgcaccc	1560
accaccacca	agaagcctgc	accaccgct	cccaagagc	ctgccccaac	taccccccaag	1620
gagactgcac	ccaccacccc	caagaagctc	acgccaccaca	cccccgagaa	gctcgcaccc	1680
accaccctg	agaagccccc	accaccacc	cctgaggagc	tgcacccac	caccctgag	1740
gagcccacac	ccaccacccc	tgaggagcct	gctccacca	ctcccaaggc	agcggctccc	1800
aacaccctta	aggagcctgc	tccaactacc	cctaaggagc	ctgctccaac	tacccttaag	1860
gagcctgctc	caactacccc	taaggagact	gctccaacta	cccctaaagg	gactgctcca	1920
actaccctca	aggaacctgc	accactact	cccaagaagc	ctgcccccaa	ggagcttgca	1980
cccaccacca	ccaaggagcc	cacatccacc	acctctgaca	agcccgcctc	aactaccctt	2040
aaggggagctg	ctccaactac	ccctaaggag	actgtcccaa	ctacccttaa	ggagcctgct	2100
ccaactaccc	ctaaggggac	tgctccaact	acctcaagg	aacctgcacc	cactactccc	2160
aagaagcctg	cccccaagga	gcttgacccc	accaccacca	agggggccac	atccaccacc	2220
tctgacaagc	ctgctccaac	tacacctaag	gagactgctc	caactacccc	caaggagcct	2280
gcacccacta	cccccaagaa	gcctgctcca	actactcctg	agacacctcc	tccaaccact	2340
tcagaggtct	ctactccaac	taccaccaag	gagcctacca	ctatccacaa	aagccctgat	2400
gaatcaactc	ctgagctttc	tgcagaaccc	acaccaaaaag	ctcttgaaaa	cagtcccaag	2460
gaacctgggtg	tacctacaac	taagactcct	gcagcgacta	aacctgaaat	gactacaaca	2520
gctaaagaca	agacaacaga	aagagactta	cgtactacac	ctgaaactac	aactgctgca	2580
cctaagatga	caaaagagac	agcaactaca	acagaaaaaa	ctaccgaatc	caaaataaca	2640
gctacaacca	cacaagtaac	atctaccaca	actcaagata	ccacaccatt	caaaattact	2700
actcttaaaa	caactactct	tgcacccaaa	gtaactacaa	caaaaaagac	aattactacc	2760
actgagatta	tgaacaaaac	tgaagaaaca	gctaaaccac	aagacagagc	tactaattct	2820
aaagcgacaa	ctcctaaacc	tcaaaagcca	accaaagcac	ccaaaaaacc	cacttctacc	2880
aaaaagccaa	aaacaatgcc	tagagtgaga	aaaccaaaga	cgacaccaac	tccccgcaag	2940
atgacatcaa	caatgccaga	attgaaccct	acctcaagaa	tagcagaagc	catgctccaa	3000
accaccacca	gacctaacca	aactccaaac	tccaaactag	ttgaagttaa	tccaaagagt	3060
gaagatgcag	gtgggtgctga	aggagaaaca	cctcatatgc	ttctcaggcc	ccatgtgttc	3120
atgcctgaag	ttactcccga	catggattac	ttaccgagag	tacccaatca	aggcattatc	3180
atcaatccca	tgctttccga	tgagaccaat	atatgcaatg	gtaagccagt	agatggactg	3240
actactttgc	gcaatgggac	attagttgca	ttccgaggtc	attattttctg	gatgctaagt	3300
ccattcagtc	caccatctcc	agctcgacga	attactgaag	tttgggggtat	tccttcccc	3360
attgatactg	tttttactag	gtgcaactgt	gaaggaaaaa	ctttcttctt	taaggattct	3420
cagtaactggc	gtttttaccaa	tgatataaaa	gatgcagggt	accccaaacc	aattttcaaa	3480
ggatttggag	gactaactgg	acaaatagt	gcagcgcttt	caacagctaa	atataagaac	3540
tggcctgaat	ctgtgtatct	tttcaagaga	gggtggcagca	ttcagcagta	tatttataaa	3600
caggaacctg	tacagaagt	ccctggaaga	aggcctgtct	taaattatcc	agtgtatgga	3660
gaaacgacac	aggttaggag	acgtcgcttt	gaacgtgcta	taggaccttc	tcaaacacac	3720
accatcagaa	ttcaatatct	acctgccaga	ctggcttatc	aagacaaagg	tgctcttcat	3780
aatgaagtta	aagttagtat	actgtggaga	ggacttccaa	atgtggttac	ctcagctata	3840
tcactgcccc	acatcagaaa	acctgacggc	tatgattact	atgccttttc	ttaagatcaa	3900

```
tactataaca ttgatgtgcc tagtagaaca gcaagagcaa ttactactcg ttctgggcag 3960
accttatcca aagtctggta caactgtcct tagactgatg agcaaaggag gagtcaacta 4020
atgaagaaat gaataataaa ttttgacact gaaaaacatt ttattaataa agaattattga 4080
catgagtata ccagtttata tataaaaaatg tttttaaact tgacaatcat tacactaaaa 4140
cagat ttgat aatcttattc acagttgtta ttgtttacag accattttaat taatatttcc 4200
tctgtttatt cctcctctcc ctcccattgc atggctcaca cctctctgtc tgttaaactg 4260
acgaccacgg accctcgtgc cgaattctt 4289
```