

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
6 March 2003 (06.03.2003)(10) International Publication Number
PCT
WO 03/018612 A2(51) International Patent Classification⁷: **C07K**

(21) International Application Number: PCT/US02/27213

(22) International Filing Date: 22 August 2002 (22.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/314,811	24 August 2001 (24.08.2001)	US
60/340,584	14 December 2001 (14.12.2001)	US
60/350,595	18 January 2002 (18.01.2002)	US
60/363,432	11 March 2002 (11.03.2002)	US
60/364,607	15 March 2002 (15.03.2002)	US
60/370,761	5 April 2002 (05.04.2002)	US
60/391,378	24 June 2002 (24.06.2002)	US

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

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **DUGGAN, Brendan, M.** [AU/US]; 243 Buena Vista Avenue # 306, Sunnyvale, CA 94086 (US). **LEE, Sally** [US/US]; 3643 26th Street, San Francisco, CA 94110 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **HAFALIA, April, J.A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street A # 205, San Leandro, CA 94577 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **LEE, Soo, Yeun** [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). **KHARE, Reena** [US/US]; 12650 Orella Court, Saratoga, CA 95070 (US). **WILSON, Amy, D.** [US/US]; 1056 Continentals Way, #27, Belmont, CA 94002 (US). **JIN, Pei** [US/US]; 320 Curtner Avenue #D, Palo Alto, CA 94306 (US). **HAWKINS, Phillip, R.** [US/US]; 750 North Shoreline Boulevard #115, Mountain View, CA 94043 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **DING, Li** [CN/US]; 11036 Ridgecorde Lane, CreveCoeur, MO 63141 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive #117, Castro Valley, CA 94546 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **CHIEN, David** [US/US]; 444 Yerba Buena Avenue, Los Altos, CA 94022 (US). **ZEBARJADIAN, Yeganeh** [IR/US]; 830 Junipero Serra Boulevard, San Francisco, CA 94127 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **KABLE, Amy, E.** [US/US]; 2345 Polk Street #4, San Francisco, CA 94109 (US). **CHANG, Hsin-Ru** [US/US]; 326 Treasure Island Drive, Belmont, CA 94002 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **LEHR-MASON, Patricia, M.** [US/US]; 360 Clarke Lane, Morgan Hill, CA 95014 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US).(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 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, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, 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 03/018612 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).

Interferons

Interferons (IFNs) are a part of the group of intercellular messenger proteins known as cytokines. a-IFN is the product of a multigene family of at least 16 members, whereas b-IFN is the product of a single gene. a- and b-IFNs are also known as type I IFNs. Type I IFNs are produced in a variety of cells types. Biosynthesis of type I IFNs is stimulated by viruses and other pathogens, and by
5 various cytokines and growth factors. g-IFN, also known as type II IFN, is produced in T-cells and natural killer cells. Biosynthesis of type II IFN is stimulated by antigens to which the organism has been sensitized. Both a- and g-IFNs are immunomodulators and anti-inflammatory agents, activating macrophages, T-cells and natural killer cells.

IFNs are part of the body's natural defense to viruses and tumors. They exert these defenses
10 by affecting the function of the immune system and by direct action on pathogens and tumor cells. IFNs mediate these multiple effects in part by inducing the synthesis of many cellular proteins. Some interferon-inducible (IFI) genes are induced equally well by a-,b- and g-IFNs. Other IFI genes are preferentially induced by the type I or by the type II IFNs.

The various proteins produced by IFI genes possess antitumor, antiviral and
15 immunomodulatory functions. The expression of tumor antigens in cancer cells is increased by a-IFN, and renders the cancer cells more susceptible to immune rejection. The IFI proteins synthesized in response to viral infections are known to inhibit viral functions such as cell penetration, uncoating, RNA and protein synthesis, assembly and release (Hardman, J.G. et al. (1996) The Pharmacological Basis of Therapeutics, McGraw-Hill, New York NY pp. 1211-1214). Type II IFN stimulates
20 expression of major histocompatibility complex (MHC) proteins and is thus used in immune response enhancement.

An IFI gene known as 6-16 encodes an mRNA which is highly induced by type I IFNs in a variety of human cells (Kelly, J.M. et al. (1986) EMBO J. 5:1601-1606). After induction, 6-16 mRNA constitutes as much as 0.1% of the total cellular mRNA. The 6-16 mRNA is present at only
25 very low levels in the absence of type I IFN, and is only weakly induced by type II IFN.

The 6-16 mRNA encodes a hydrophobic protein of 130 amino acids. The first 20 to 23 amino acids comprise a putative signal peptide. Protein 6-16 has at least two predicted transmembrane regions culminating in a negatively charged C-terminus.

The p27 gene encodes a protein with 41% amino acid sequence identity to the 6-16 protein.
30 The p27 gene is expressed in some breast tumor cell lines and in a gastric cancer cell line. In other breast tumor cell lines, in the HeLa cervical cancer cell line, and in fetal lung fibroblasts, p27 expression occurs only upon a-IFN induction. In one breast tumor cell line, p27 is independently induced by estradiol and by IFN (Rasmussen, U.B. et al. (1993) Cancer Res. 53:4096-4101).

Expression of p27 was analyzed in 21 primary invasive breast carcinomas, 1 breast cancer
35 bone metastasis, and 3 breast fibroadenomas. High levels of p27 were found in about one-half of the

primary carcinomas and in the bone metastasis, but not in the fibroadenomas. These observations suggest that certain breast tumors may produce high levels of, or have increased sensitivity to, type I IFN as compared to other breast tumors (Rasmussen UB et al, supra). In addition, the p27 gene expressed at significant levels in normal tissues including colon, stomach and lung, but not expressed in placenta, kidney, liver or skin. (Rasmussen UB et al, supra).

The small hydrophobic IFI gene products may contribute to viral resistance. A hepatitis-C virus (HCV)-induced gene, 130-51, was isolated from a cDNA library prepared from chimpanzee liver during the acute phase of the infection. The protein product of this gene has 97% identity to the human 6-16 protein (Kato T et al (1992) Virology 190:856-860). The investigators suggest that HCV infection actively induces IFN expression, which in turn induces expression of IFI genes including 130-51.

The IFI proteins synthesized in response to viral infections are known to inhibit viral functions such as penetration, uncoating, RNA or protein synthesis, assembly or release. The 130-51 protein may inhibit one or more of these functions in HCV. A particular virus may be inhibited in multiple functions by IFI proteins. In addition, the principle inhibitory effect exerted by IFI proteins differs among the virus families (Hardman JG, supra, p 1211).

The hydrophobic IFI proteins of the invention may provide the basis for clinical diagnosis of diseases associated with their induction. These proteins may be useful in the diagnosis and treatment of tumors, viral infections, inflammation, or conditions associated with impaired immunity. Furthermore, these proteins may be used for investigations of the control of gene expression by IFNs and other cytokines in normal and diseased cells.

Tetratricopeptide repeats (TPRs) are conserved 34-amino acid sequence motifs which provide support for protein-protein interactions. For example, the TPR protein P58IPK possesses cochaperone and oncogenic properties (Melville, M.W. et al. (2000) Cell. Mol. Life Sci. 57:311-322). P58IPK is an influenza virus-activated cellular inhibitor of the double-stranded RNA-activated serine/threonine protein kinase, PKR. It also has nine TPR motifs, as well as a C-terminal DNA-J motif, both of which are known to mediate protein-protein interactions. TPR motifs 5 to 7 share some homology to the natural PKR substrate, eIF2- α (Gale, M. et al. (1996) Mol. Cell. Biol. 16:4172-4181). An activated P58IPK-PKR complex functions to inhibit PKR phosphorylation and activity, and the limited homology between P58IPK and eIF2- α may indicate a common mechanism of PKR inhibition by these two types of interferon-associated proteins.

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-g, 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.

Protease inhibitors

Protease inhibitors and other regulators of protease activity control the activity and effects of proteases. Protease inhibitors have been shown to control pathogenesis in animal models of proteolytic disorders (Murphy, G. (1991) Agents Actions Suppl. 35:69-76). Low levels of the cystatins, low molecular weight inhibitors of the cysteine proteases, correlate with malignant progression of tumors (Calkins, C. et al. (1995) Biol. Biochem. Hoppe Seyler 376:71-80). The cystatin superfamily of protease inhibitors is characterized by a particular pattern of linearly arranged and tandemly repeated disulfide loops (Kellermann, J. et al. (1989) J. Biol. Chem. 264:14121-14128). Serpins are inhibitors of mammalian plasma serine proteases. Many serpins serve to regulate the blood clotting cascade and/or the complement cascade in mammals. Sp32 is a positive regulator of the mammalian acrosomal protease, acrosin, that binds the proenzyme, proacrosin, and thereby aides in packaging the enzyme into the acrosomal matrix (Baba, T. et al. (1994) J. Biol. Chem. 269:10133-10140). The Kunitz family of serine protease inhibitors are characterized by one or more "Kunitz domains" containing a series of cysteine residues that are regularly spaced over approximately 50 amino acid residues and form three intrachain disulfide bonds. Members of this family include aprotinin, tissue factor pathway inhibitor (TFPI-1 and TFPI-2), inter- α -trypsin inhibitor, and bikunin (Marlor, C.W. et al. (1997) J. Biol. Chem. 272:12202-12208). Members of this family are potent inhibitors (in the nanomolar range) against serine proteases such as kallikrein and plasmin. Aprotinin has clinical utility in reduction of perioperative blood loss.

A major portion of all proteins synthesized in eukaryotic cells are synthesized on the cytosolic surface of the endoplasmic reticulum (ER). Before these immature proteins are distributed to other organelles in the cell or are secreted, they must be transported into the interior lumen of the ER where post-translational modifications are performed. These modifications include protein folding and the formation of disulfide bonds, and N-linked glycosylations.

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.

The human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of the cells. C3A cells have many characteristics of primary human hepatocytes in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with α -fetoprotein iii) conversion of ammonia to urea and glutamine; iv) metabolize aromatic amino acids; and v) are able to proliferate in glucose-free and insulin-free medium. The C3A cell line is now well established as an *in vitro* model of the mature human liver (Mickelson et al. (1995) Hepatology 22:866-875; Nagendra et al. (1997) Am J Physiol 272:G408-416).

Captopril is an antihypertensive known as an angiotensin converting enzyme (ACE) inhibitor. ACE is a target for treatment of myocardial infarction and hypotension. ACE inhibitors can be classified into three broad groups based on chemical structure: i) sulfhydryl-containing ACE inhibitors, structurally related to captopril (e.g., fentiapril, pivalo-pril, zofenopril, alacepril); ii) dicarboxyl-containing ACE inhibitors, structurally related to enalapril (e.g., lisinopril, benazepril, quinapril, moexipril, ramipril, spirapril, perindopril, indolapril, pentopril, indala-pril, cilazapril); and iii) phosphorus-containing ACE inhibitors, structurally related to fosinopril. Many ACE inhibitors

are ester-containing prodrugs that are 100 to 1000 times less potent as ACE inhibitors than their active metabolites, but have a much better oral bioavailability than the active molecules.

Approximately 16 different ACE inhibitors are used world-wide. In general, ACE inhibitors differ with respect to potency; whether ACE inhibition is due to the drug itself or to activation of a prodrug; and pharmacokinetic properties. With the notable exceptions of fosinopril and spirapril (which display balanced elimination by the liver and kidneys), ACE inhibitors are cleared predominantly by the kidneys. Drugs that interfere with the renin-angiotensin system play a prominent role in the treatment cardiovascular disease and have been used as a therapy for a number of diseases including hypotension, left ventricular systolic dysfunction, myocardial infarction, progressive renal impairment, and scleroderma renal crisis.

Enalapril is a prodrug that is not highly active and, as such, it must be hydrolyzed by esterases in the liver to produce the active parent dicarboxylic acid, enalaprilat. Enalaprilat is a highly potent inhibitor of ACE with a K_i of 0.2 nM but differs from captopril in that it is an analog of a tripeptide rather than a dipeptide. Enalapril is rapidly absorbed when given orally and has an oral bioavailability of about 60% (not reduced by food). Although peak concentrations of plasma enalapril occur within an hour, enalaprilat concentrations do not peak until three to four hours. Enalapril has a half-life of only 1.3 hours. However, because it binds tightly to ACE, enalaprilat has a plasma half-life of about 11 hours. Nearly all the drug is eliminated by the kidneys either as intact enalapril or enalaprilat.

Dexamethasone and its derivatives, dexamethasone sodium phosphate and dexamethasone acetate, are synthetic glucocorticoids used as anti-inflammatory or immunosuppressive agents. Dexamethasone has little to no mineralocorticoid activity and is usually selected for management of cerebral edema because of its superior ability to penetrate the central nervous system. Glucocorticoids are naturally occurring hormones that prevent or suppress inflammation and immune responses when administered at pharmacological doses. Responses can include inhibition of leukocyte infiltration at the site of inflammation, interference in the function of mediators of inflammatory response, and suppression of humoral immune responses. The anti-inflammatory actions of corticosteroids are thought to involve phospholipase A_2 inhibitory proteins, collectively called lipocortins. The numerous adverse effects related to corticosteroid use usually depend on the dose administered and the duration of therapy. Proposed mechanisms of action include decreased IgE synthesis, increased number of β -adrenergic receptors on leukocytes, and decreased arachidonic acid metabolism. During an immediate allergic reaction, such as in chronic bronchial asthma, allergens bridge the IgE antibodies on the surface of mast cells, which triggers these cells to release chemotactic substances. Mast cell influx and activation, therefore, is partially responsible for the inflammation and hyperirritability of the oral mucosa in asthmatic patients. This inflammation can be retarded by

administration of adrenocorticoids. As with other corticosteroids, the effects upon liver metabolism and hormone clearance mechanisms are important to understand the pharmacodynamics of a drug.

CA3 liver cells treated with Steroids

Steroids are a class of lipid-soluble molecules, including cholesterol, bile acids, vitamin D, and hormones, that share a common four-ring structure based on cyclopentanoperhydrophenanthrene and that carry out a wide variety of functions. Cholesterol, for example, is a component of cell membranes that controls membrane fluidity. It is also a precursor for bile acids which solubilize lipids and facilitate absorption in the small intestine during digestion. Vitamin D regulates the absorption of calcium in the small intestine and controls the concentration of calcium in plasma.

Steroid hormones, produced by the adrenal cortex, ovaries, and testes, include glucocorticoids, mineralocorticoids, androgens, and estrogens. They control various biological processes by binding to intracellular receptors that regulate transcription of specific genes in the nucleus. Glucocorticoids, for example, increase blood glucose concentrations by regulation of gluconeogenesis in the liver, increase blood concentrations of fatty acids by promoting lipolysis in adipose tissues, modulate sensitivity to catecholamines in the central nervous system, and reduce inflammation. The principal mineralocorticoid, aldosterone, is produced by the adrenal cortex and acts on cells of the distal tubules of the kidney to enhance sodium ion reabsorption. Androgens, produced by the interstitial cells of Leydig in the testis, include the male sex hormone testosterone, which triggers changes at puberty, the production of sperm and maintenance of secondary sexual characteristics. Female sex hormones, estrogen and progesterone, are produced by the ovaries and also by the placenta and adrenal cortex of the fetus during pregnancy. Estrogen regulates female reproductive processes and secondary sexual characteristics. Progesterone regulates changes in the endometrium during the menstrual cycle and pregnancy.

Steroid hormones are widely used for fertility control and in anti-inflammatory treatments for physical injuries and diseases such as arthritis, asthma, and auto-immune disorders. Progesterone, a naturally occurring progestin, is primarily used to treat amenorrhea, abnormal uterine bleeding, or as a contraceptive. Endogenous progesterone is responsible for inducing secretory activity in the endometrium of the estrogen-primed uterus in preparation for the implantation of a fertilized egg and for the maintenance of pregnancy. It is secreted from the corpus luteum in response to luteinizing hormone (LH). The primary contraceptive effect of exogenous progestins involves the suppression of the midcycle surge of LH. At the cellular level, progestins diffuse freely into target cells and bind to the progesterone receptor. Target cells include the female reproductive tract, the mammary gland, the hypothalamus, and the pituitary. Once bound to the receptor, progestins slow the frequency of release of gonadotropin releasing hormone from the hypothalamus and blunt the pre-ovulatory LH surge, thereby preventing follicular maturation and ovulation. Progesterone has minimal estrogenic

and androgenic activity. Progesterone is metabolized hepatically to pregnanediol and conjugated with glucuronic acid.

Medroxyprogesterone (MAH), also known as 6 α -methyl-17-hydroxyprogesterone, is a synthetic progestin with a pharmacological activity about 15 times greater than progesterone. MAH is used for the treatment of renal and endometrial carcinomas, amenorrhea, abnormal uterine bleeding, and endometriosis associated with hormonal imbalance. MAH has a stimulatory effect on respiratory centers and has been used in cases of low blood oxygenation caused by sleep apnea, chronic obstructive pulmonary disease, or hypercapnia.

Corticosteroids are used to relieve inflammation and to suppress the immune response. They inhibit eosinophil, basophil, and airway epithelial cell function by regulation of cytokines that mediate the inflammatory response. They inhibit leukocyte infiltration at the site of inflammation, interfere in the function of mediators of the inflammatory response, and suppress the humoral immune response. Corticosteroids are used to treat allergies, asthma, arthritis, and skin conditions. Beclomethasone is a synthetic glucocorticoid that is used to treat steroid-dependent asthma, to relieve symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or to prevent recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive effects of intranasal beclomethasone are 5000 times greater than those produced by hydrocortisone. Budesonide is a corticosteroid used to control symptoms associated with allergic rhinitis or asthma. Budesonide has high topical anti-inflammatory activity but low systemic activity. Dexamethasone is a synthetic glucocorticoid used in anti-inflammatory or immunosuppressive compositions. It is also used in inhalants to prevent symptoms of asthma. Due to its greater ability to reach the central nervous system, dexamethasone is usually the treatment of choice to control cerebral edema. Dexamethasone is approximately 20-30 times more potent than hydrocortisone and 5-7 times more potent than prednisone. Prednisone is metabolized in the liver to its active form, prednisolone, a glucocorticoid with anti-inflammatory properties. Prednisone is approximately 4 times more potent than hydrocortisone and the duration of action of prednisone is intermediate between hydrocortisone and dexamethasone. Prednisone is used to treat allograft rejection, asthma, systemic lupus erythematosus, arthritis, ulcerative colitis, and other inflammatory conditions. Betamethasone is a synthetic glucocorticoid with antiinflammatory and immunosuppressive activity and is used to treat psoriasis and fungal infections, such as athlete's foot and ringworm.

Human peripheral blood mononuclear cells (PBMCs) can be classified into discrete cellular populations representing the major cellular components of the immune system. PBMCs contain about 52% lymphocytes (12% B lymphocytes, 40% T lymphocytes {25% CD4+ and 15% CD8+}), 20% NK cells, 25% monocytes, and 3% various cells that include dendritic cells and progenitor cells. The proportions, as well as the biology of these cellular components tend to vary slightly between healthy

individuals, depending on factors such as age, gender, past medical history, and genetic background.

PMA is a broad activator of the protein kinase C-dependent pathways. Ionomycin is a calcium ionophore that permits the entry of calcium in the cell, hence increasing the cytosolic calcium concentration. The combination of PMA and ionomycin activates two of the major signaling pathways used by mammalian cells to interact with their environment. In PBMCs, the combination of PMA and ionomycin mimics the secondary signaling events elicited during activation of lymphocytes, NK cells, and monocytes.

T cells can be subdivided into 2 classes according to their main function and the surface-antigens they express. The main function of CD4 positive T cells, also known as T Helper cells, is to regulate the immune response by producing soluble factors that regulate the activity of effector cells such as B lymphocytes, NK cells, and macrophages. The main function of CD8 positive T cells, also known as cytotoxic T cells, is to kill "abnormal" cells such as tumor cells or cells infected by viruses.

In the blood of healthy adults, CD4 positive T cells and CD8 positive T cells represent 25% and 15% of the mononuclear cells, respectively. These two T cell populations can be readily expanded out of blood by incubating bulk peripheral blood mononuclear cells (PBMCs) in the presence of phytohemagglutinin (PHA) and IL-2. After 8 to 10 days of treatment, both CD4+ and CD8+ T cells usually expand 5 to 10 fold, yielding a cell population consisting of >90% T cells also known as PHA blasts. It is important to note that T cell expansion occurs during the first 5 days of PHA stimulation, therefore after 8 to 10 days in culture, most PHA blasts have returned to a resting state.

T cells require two distinct signals to achieve optimal activation. First, the "antigenic" signal delivered through the binding of the TCR-CD3 complex. Second, the costimulatory signal delivered through the binding of the CD28 molecules. Upon binding of the TCR-CD3 complex alone, T cells only achieve a partial state of activation. However, it is important to note that the signaling requirements of T cells depend greatly on the cycling state of those cells.

PMA is a broad activator of the protein kinase C-dependent pathways. Ionomycin is a calcium ionophore that permits the entry of calcium in the cell, hence increasing the cytosolic calcium concentration. The combination of PMA and ionomycin activates two of the major signaling pathways used by mammalian cells to interact with their environment. In T cells, the combination of PMA and ionomycin mimics the type of secondary signaling events elicited during optimal B cell activation.

There is a need in the art for new compositions, including nucleic acids and proteins, for the 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 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,' 'EXMES-25,' 'EXMES-26,' 'EXMES-27,' and 'EXMES-28' and methods for using these proteins and their encoding 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 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:1-28, 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:1-28, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-28.

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:1-28, 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:1-28, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-28. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:29-56.

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:1-28, 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:1-28, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

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:1-28, 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:1-28, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28. 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 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:1-28, 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:1-28, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28.

Still yet another embodiment provides 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:29-56, 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:29-56, 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:29-56, 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:29-56, 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:29-56, 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:29-56, 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:1-28, 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:1-28, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-28. 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 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:1-28, 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:1-28, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28. The method comprises a) exposing a sample comprising the polypeptide to 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 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:1-28, 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:1-28, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28. The method comprises a) exposing a sample comprising the polypeptide to 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.

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:1-28, 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:1-28, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28. 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:1-28, 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:1-28, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28. 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:29-56, the method comprising a) exposing a sample comprising the target polynucleotide to 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:29-56, 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:29-56, 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:29-56, 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:29-56, 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 (GCG, Madison WI) 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

	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
5	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
10	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
15	Tyr	His, Phe, Trp
	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. 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 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:29-56 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:29-56, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:29-56 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:29-56 from related polynucleotides. The precise length of a fragment of SEQ ID NO:29-56 and the region of SEQ ID NO:29-56 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:1-28 is encoded by a fragment of SEQ ID NO:29-56. A fragment of SEQ ID NO:1-28 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-28. For example, a fragment of SEQ ID NO:1-28 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-28. The precise length of a fragment of SEQ ID NO:1-28 and the region of SEQ ID NO:1-28 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., 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 to the percentage of identical residue 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 computer algorithms or programs known in the art or described herein. For example, percent identity
 5 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 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of
 10 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 Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410),
 15 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 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2
 20 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 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

25 *Matrix: BLOSUM62*

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

30 *Expect: 10*

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined 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 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. et al.

(1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 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%. 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 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 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.

“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 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, 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 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.

“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.

“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.

“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, 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 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, 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 the references, for example Sambrook, J. et al. (1989; Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular
5 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
10 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
15 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
20 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
25 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
30 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, *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.

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,
5 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.

10 “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
15 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
20 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
25 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
30 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 et al. (1989), *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

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 "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 of exons 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 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 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 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.

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.

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 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. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME 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 column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 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:2 is 67% identical, from residue M1 to residue D469, to human interferon-inducible 56 Kd protein (GenBank ID g307041) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.4e-171, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains TPR domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLAST-

PRODOM and BLAST-DOMO analyses provide further corroborative evidence that SEQ ID NO:2 is an interferon-associated protein. In an alternative example, SEQ ID NO:4 is 96% identical, from residue M1 to residue T199, to chicken fibroblast growth factor homologous factor 2 isoform 1Q +1Y' (GenBank ID g6911131) as determined by the Basic Local Alignment Search Tool (BLAST).

5 (See Table 2.) The BLAST probability score is $3.5e-101$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 also has homology to human and mouse fibroblast growth factor 13, which has probable roles in nervous system development and function, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:4 also contains a fibroblast growth factor domain as determined by searching for statistically significant

10 matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:4 is a member of the fibroblast growth factor family. In an alternative example, SEQ ID NO:8 is 95% identical, from residue M1 to residue P169, to human granulocyte colony stimulating factor precursor (GenBank ID g14423573) as determined by the Basic

15 Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $5.3e-45$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 also has homology to a protein that is localized to the extracellular region and is a granulocyte colony stimulating factor (G-CSF) that regulate proliferation, differentiation, and survival of neutrophilic granulocytes, as determined by BLAST analysis using the PROTEOME database.

20 SEQ ID NO:8 also contains an interleukin-6/G-CSF/MGF family 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 PROFILESCAN analyses, and BLAST analyses of the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:8 is a G-CSF family cytokine. In an alternative example, SEQ ID NO:14 is

25 100% identical, from residue M15 to residue G626 and 99% identical from residue T656 to residue G757 to human hepatocyte growth factor-like protein (GenBank ID g1311661) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance.

30 SEQ ID NO:14 also has homology to proteins that are localized extracellularly, have hydrolase and protease activity, regulate macrophage chemotaxis and activation, function in ciliary motility, inhibit apoptosis, may function in blood coagulation and the oxidative stress response and are macrophage stimulating proteins, members of the hepatocyte growth factor family that signal via receptor tyrosine kinases and may play a role in spermatogenesis, sperm motility, and fertilizing capacity, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:14 also contains

Kringle, PAN, and trypsin domains, and a trypsin-like serine protease domain 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 provide further corroborative evidence that SEQ ID NO:14 is a hepatocyte growth factor-like protein. In an alternative example, SEQ ID NO:22 is 99% identical, from residue M1 to residue K1801, and from residue K1800 to residue K2747, to human thyroglobulin (GenBank ID g2707181) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. As determined by BLAST analysis using the PROTEOME database, SEQ ID NO:22 also has homology to extracellular proteins that function as glycoprotein precursors of the thyroid hormones thyroxine and triiodothyronine, which regulate metabolic processes. The mutation of the corresponding genes causes congenital goiter with hypothyroidism and endoplasmic reticulum storage disease. SEQ ID NO:22 also contains a thyroglobulin type I domain 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 and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:22 is a thyroglobulin. SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5-7, SEQ ID NO:9-13, SEQ ID NO:15-21, and SEQ ID NO:23-28 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-28 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:29-56 or that distinguish between SEQ ID NO:29-56 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. A preferred EXMES variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the EXMES amino acid sequence, and which contains 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:29-56, which encodes EXMES. The polynucleotide sequences of SEQ ID NO:29-56, 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:29-56 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:29-56. 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 polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons 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:33, a polynucleotide comprising a sequence of SEQ ID NO:44, and a polynucleotide comprising a sequence of SEQ ID NO:46 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:35 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:38, a polynucleotide comprising a sequence of SEQ ID NO:39, and a polynucleotide comprising a sequence of SEQ ID NO:42 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:41 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:54, a polynucleotide comprising a sequence of SEQ ID NO:55, a polynucleotide comprising a sequence of SEQ ID NO:56, and a polynucleotide comprising a sequence of SEQ ID NO:47 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

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 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 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 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 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:29-56 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.* 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 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). 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).

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, 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 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 (Clontech, Palo Alto CA) to walk 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.

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.

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
5 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
10 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
15 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
20 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
25 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
30 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). Alternatively, EXMES itself or a fragment thereof may be synthesized using chemical methods
5 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). Additionally, the amino acid sequence of EXMES, or any part thereof, may be altered during direct
10 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 composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing
15 (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 a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and
20 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 polynucleotide sequence encoding EXMES and its initiation codon and upstream regulatory
25 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 codons may be of various origins, both natural and synthetic. The efficiency of expression may be
30 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 translational control elements. These methods include *in vitro* recombinant DNA techniques,

synthetic techniques, and *in vivo* genetic recombination (Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express
 5 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
 10 (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook, *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.
 15 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)
 20 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
 25 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
 30 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).

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) *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 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 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 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 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.

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

5 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
10 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; Clontech), β -glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used.
15 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
20 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.

25 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.

30 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
5 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
10 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.
15 (*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
20 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
25 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
30 compounds that bind to a variant of EXMES, but not to EXMES having the exact sequence of a sequence of SEQ ID NO:1-28. 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

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).

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 include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing EXMES or cell membrane fractions which contain EXMES are then contacted with
5 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
10 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 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 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
15 (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 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 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
20 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 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 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 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, 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, giantism, 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.

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.

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.

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.

5 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
10 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
15 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. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide
20 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,
25 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.

30 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 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
5 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
10 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
15 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).

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,
20 G. et al. (1991) Nature 349:293-299).

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
25 easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

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
30 immunoassays typically involve the measurement of complex formation between EXMES and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies 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 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
5 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 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
10 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 Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

15 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,
20 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,
25 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
30 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) 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; 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 (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 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 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 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 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 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 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 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.

5 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
10 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
15 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.

20 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-
25 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
30 associated with decreased EXMES expression or activity, a compound which specifically promotes expression of the polynucleotide encoding EXMES may be therapeutically useful.

At least one, and up to a plurality, of 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 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 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 combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, 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 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).

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 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.

5 The compositions utilized in this invention 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.
10 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,
15 J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of 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.

20 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
25 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
30 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

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
5 therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are 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 ED_{50} 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.

10 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
15 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.

20 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
25 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
30 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:29-56 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

5 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

10 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

15 adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, 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

20 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

25 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

30 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, 5 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 10 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 aspect, polynucleotides encoding EXMES may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides 15 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 in comparison to a control sample then the presence of altered levels of polynucleotides encoding 20 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 EXMES, a normal or standard profile for expression is established. This may be accomplished by 25 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 polynucleotide is used. Standard values obtained in this manner may be compared with values 30 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 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 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 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 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 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 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-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

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 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 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.

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.

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.

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 exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendozze, 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.

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 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 WO95/251116; Shalon, D. et al. (1995) 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 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
5 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
10 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
15 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
20 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
25 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.

30 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.

5 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
10 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
15 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.

20 The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/314,811, U.S. Ser. No. 60/350,595, U.S. Ser. No. 60/364,607, U.S. Ser. No. 60/363,432, U.S. Ser. No. 60/370,761 U.S. Ser. No. 60/340,584, and U.S. Ser. No. 60/391,378, are hereby expressly incorporated by reference.

25 **EXAMPLES**

I. Construction of cDNA Libraries

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

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA

purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed 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 were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were 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 were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was 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 were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal
5 cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were 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
10 were 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 were identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences were selected for extension using the techniques disclosed in
15 Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were 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 were then queried against a selection of public
20 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 Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft,
25 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 were performed using programs based on BLAST, FASTA, BLIMPS, and
30 HMMER. The Incyte cDNA sequences were 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) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on

GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were 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 were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, 5 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 10 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 the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and 15 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, 20 the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:29-56. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

25 IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative extracellular messengers were 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 30 (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 was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode extracellular messengers, the encoded polypeptides were analyzed by querying

against PFAM models for extracellular messengers. Potential extracellular messengers were also identified by homology to Incyte cDNA sequences that had been annotated as extracellular messengers. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr public databases. Where necessary, the Genscan-predicted sequences were 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 was 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 was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were 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 were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were 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) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpr public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

“Stretched” Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was 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 were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore “stretched” or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of EXMES Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:29-56 were 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:29-56 were 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 were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted 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 Généthon 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, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

5 Analogous computer techniques applying BLAST were used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte Genomics). 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:

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

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 were 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; 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 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 GOLD database (Incyte Genomics, Palo Alto CA).

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 was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 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 was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained 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), 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+ were 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 min; Step 7: storage at 4°C.

The concentration of DNA in each well was 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 was scanned in a Fluoroskan II (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 was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the

sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviII cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun
5 sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were 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 into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing
10 media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

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

IX. Identification of Single Nucleotide Polymorphisms in EXMES Encoding Polynucleotides

25 Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:29-56 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were 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 was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of
30 basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse

transcriptase, polymerase, or somatic mutation. Clustering error filters used 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 removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

5 Certain SNPs were 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 comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The
10 Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 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 were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

15 X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:29-56 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
20 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
25 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
30 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 (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested
5 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 (Schena, 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
15 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
20 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
25 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
30 GEMBRIGHT kits (Incyte Genomics). 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 (Clontech,

Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 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 μ l 5X SSC/0.2% SDS.

5 Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element 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
10 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope 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,
15 and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

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 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic
20 apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). 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
25 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g 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
30 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 addition of 140 μ l 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 Genomics). Array elements that exhibited at least about a two-fold change in expression, a signal-to-background ratio of at least 2.5, and an element spot size of at least 40% were identified as

5 differentially expressed.

Expression

For example, SEQ ID NO:30 showed differential expression as determined by microarray analysis. Histological and molecular evaluation of breast tumors reveals that the development of breast cancer evolves through a multi-step process whereby pre-malignant mammary epithelial cells
10 undergo a relatively defined sequence of events leading to tumor formation. In one type of experiment, a cross-comparison protocol was used to evaluate the expression of cDNAs from various human breast tumor cell lines at different stages of tumor progression. These were compared to a non-tumorigenic mammary epithelial cell line, MCF-10A, isolated from a 36-year old female with fibrocystic breast disease and possessing luminal ductal characteristics. The MCF-10A cell line
15 expresses cytoplasmic keratins, epithelial sialomucins, and milkfat globule antigens. This cell line also exhibits three-dimensional growth in collagen and forms domes in confluent culture. All cell cultures were grown to 70-80% confluence prior to RNA isolation.

The expression of SEQ ID NO:30 was decreased at least two-fold in T47D cells, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with
20 infiltrating ductal carcinoma of the breast. The expression of SEQ ID NO:30 was also found to be decreased by at least two-fold in Sk-BR-3 cells, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year old female. It forms poorly differentiated adenocarcinoma when injected into nude mice. These cells were propagated in defined serum-free TCH medium.

In another cross comparison of breast cancer cells lines, the expression of SEQ ID NO:30 was
25 decreased at least two-fold in five human breast cancer cell lines compared to non-tumorigenic MCF-10A cells. In addition to a decreased expression of SEQ ID NO:30 in T47D and Sk-BR-3 cells, expression of SEQ ID NO:30 was similarly decreased in MCF7 cells, a non-malignant breast adenocarcinoma line isolated from the pleural effusion of a 69-year old female. MCF7 has retained characteristics of the mammary epithelium such as the ability to process estradiol via cytoplasmic
30 estrogen receptors and the capacity to form domes in culture. The expression of SEQ ID NO:30 was also found to be decreased by at least two-fold in BT-20 cells, a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of the tumor mass isolated from a 74-year old female; and in MDA-mb-231 cells, a breast tumor cell line isolated from the pleural effusion of a 51-year old female. This cell line forms poorly differentiated adenocarcinoma in nude mice and ALS treated

BALB/c mice. It also expresses the Wnt3 oncogene, EGF, and TGF- α . These cells were propagated in defined serum-free H14 medium.

The expression of SEQ ID NO:30 was increased in a time-course experimental protocol involving MCF7 non-malignant breast adenocarcinoma cells treated with 10 ng/ml tumor necrosis factor alpha (TNF- α). Tumor cells stimulate the formation of stroma that secretes various mediators, such as growth factors, cytokines, and proteases, which are critical for tumor growth. For instance, serum TNF- α increased in the circulation of patients with malignancy. And, it is likely that the serum TNF- α concentration can serve as a biomarker for staging invasiveness of breast cancer. Clinically, TNF- α treatment in combination with interferon-gamma (IFN- γ) may provide a successful approach to overcome the cellular heterogeneity of advanced breast tumors. In this study, treated cells were compared to untreated cells kept in culture for the same amount of time. The expression of SEQ ID NO:30 was increased by at least two-fold following treatment of the MCF7 cells with TNF- α for 48 and 72 hours.

In another time-course study involving human breast tumor cells, MCF7 cells were treated with PD98059, an MEK inhibitor, for 8 days. Treated cells were compared to MCF7 cells treated with DMSO (vehicle) for the same duration. In addition, to assess the possible effect of DMSO on MCF7 cells, untreated cells were compared to DMSO-treated cells. After 8 days, the expression of SEQ ID NO:30 was found to be decreased by at least two-fold in treated cells.

In an experiment designed to compare gene expression profiles of mammary epithelial cells grown in the presence or absence of pituitary extract, MCF-10A cells were grown in either the supplier's recommended medium, a defined serum-free TCH medium, or a chemically defined H14 medium prior to RNA harvest. Culture medium and other growth conditions can influence epithelial cell phenotypes including expression of the cytokeratin markers. The expression of SEQ ID NO:30 was increased by at least two-fold following propagation of cells in pituitary extract-containing medium when compared with cells grown in the two other types of media described above.

As with most tumors, prostate cancer develops through a multistage progression ultimately resulting in an aggressive tumor phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells. Androgen responsive cells become hyperplastic and evolve into early-stage tumors. Although early-stage tumors are often androgen sensitive and respond to androgen ablation, a population of androgen independent cells evolve from the hyperplastic population. These cells represent a more advanced form of prostate tumor that may become invasive and potentially become metastatic to the bone, brain, or lung. The expression of SEQ ID NO:30 was differentially expressed in a cross-comparison between a non-tumorigenic human prostate cell line, PZ-HPV-7, and two human tumorigenic cell lines: PC-3, a prostate

adenocarcinoma cell line isolated from a metastatic site in the bone of a 62-year old male with grade IV prostate adenocarcinoma; and LNCaP, a prostate carcinoma cell line isolated from a lymph node biopsy of a 50-year old male with metastatic prostate carcinoma. The expression of SEQ ID NO:30 was significantly underexpressed by at least two-fold in these experiments.

5 In yet another experiment, the expression of SEQ ID NO:30 was differentially expressed in human prostate tumor cells propagated under various conditions. The CA-HPV-10 cell line was derived from cells from a 63-year old male with prostatic adenocarcinoma of Gleason Grade 4/4. The cells were transformed by transfection with HPV18 DNA. Immunocytochemical analysis showed expression of keratins 5 and 8 and also the early region 6 (E6) oncoprotein of HPV. The cells are
10 negative for prostate specific antigen (PSA). The expression of SEQ ID NO:30 was decreased by at least two-fold in CA-HPV-10 prostate epithelial cells treated with bovine pituitary extract-containing medium, as compared to untreated cells.

In a similar experiment, the expression of SEQ ID NO:30 was differentially expressed in DU145 cells, a line of prostate carcinoma cells isolated from a metastatic site in the brain of a 69-year
15 old male with widespread metastatic prostate carcinoma. DU145 has no detectable sensitivity to hormones; forms colonies in semi-solid medium; is only weakly positive for acid phosphatase; and cells are negative for prostate specific antigen (PSA). The gene expression profiles were compared in the presence or absence of serum. The expression of SEQ ID NO:30 was increased by at least two-fold in the presence of 10% fetal bovine serum.

20 Additional experiments conducted to compare gene expression profiles yielded differential expression of SEQ ID NO:30. PrEC/3 is a primary prostate epithelial cell line isolated from a normal donor. In one experiment, the prostate carcinoma cell lines DU145 and LNCaP were grown under restrictive conditions (starved, e.g. grown in basal media in the absence of growth factors and hormones), and gene expression profiles compared to PrEC/3 cells. Under these conditions, the
25 expression of SEQ ID NO:30 was decreased by at least two-fold in the tumor cell lines.

In a similar experiment, the gene expression profiles of prostate carcinoma cell lines DU145 and LNCaP grown under optimal conditions were compared to those of PrEC/3s grown under restrictive conditions. The expression of SEQ ID NO:30 was decreased by at least two-fold in the tumor cell lines.

30 Colon cancer develops through a multi-step process in which pre-malignant colonocytes undergo a relatively defined sequence of events leading to tumor formation. Factors that contribute to the process of tumor progression and malignant transformation include genetics, mutations, and selection. The expression of SEQ ID NO:30 was decreased by at least two-fold in colon cancer tissue (moderately differentiated adenocarcinoma) compared to normal colon tissue from the same donor.

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 cancer is referred to as small cell lung cancer (SCLC). Collectively, NSCLCs account for approximately 70% of cases while SCLCs account for approximately 18% of all cases. Pair comparisons were performed in which normal lung tissue and lung tumor tissue from the same donor were examined. A lung adenocarcinoma (Dn5799) and three squamous cell carcinomas (Dn5800, Dn5793 and Dn5797) were compared to same-donor normal lung tissue, yielding a decrease in the expression of SEQ ID NO:30 by at least two-fold in all cases.

These experiments indicate that a significant differential expression of SEQ ID NO:30 under various conditions further establishes the utility of SEQ ID NO:30 as a diagnostic marker, for disease staging, or as a therapeutic target for interferon-associated diseases including cancer.

In an alternative example, SEQ ID NO:38, 39, 42, and 45 showed differential expression, as determined by microarray analysis, in liver C3A cells. The human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of the cells. C3A cells have many characteristics of primary human hepatocytes in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with α -fetoprotein iii) conversion of ammonia to urea and glutamine; iv) metabolism of aromatic amino acids; and v) proliferation in glucose-free and insulin-free medium. The C3A cell line is now well established as an in vitro model of the mature human liver (Mickelson et al. (1995) *Hepatology* 22:866-875; Nagendra et al. (1997) *Am J Physiol* 272:G408-G416).

For example, SEQ ID NO:38, SEQ ID NO:39 and SEQ ID NO:42 showed differential expression, as determined by microarray analysis, in liver C3A cells treated with each of the following angiotensin converting enzyme (ACE) inhibitors: captopril and enalapril. Captopril and enalapril are antihypertensives known as an angiotensin converting enzyme (ACE) inhibitors. Approximately 16 different ACE inhibitors are used world-wide. In general, ACE inhibitors differ with respect to potency; whether ACE inhibition is due to the drug itself or to activation of a prodrug; and pharmacokinetic properties.

All ACE inhibitors have similar therapeutic indications, adverse effect profiles, and contraindications. However, the Quality-of-Life Hypertension Study Group reported that, although captopril and enalapril were indistinguishable with regard to antihypertensive efficacy and safety, captopril had a more favorable effect on quality of life (Testa et al., 1993). Since hypertension usually requires lifelong treatment, quality-of-life issues play an increasingly important role when

comparing antihypertensive drugs. ACE inhibitors differ markedly in tissue distribution (Keilani et al., 1995), and it is possible that such difference could be exploited to inhibit some local renin-angiotensin systems while leaving others relatively intact. Most ACE inhibitors are cleared predominantly by the kidneys. Drugs that interfere with the renin-angiotensin system play a prominent role in the treatment cardiovascular disease and have been used as a therapy for a number of diseases including hypotension, left ventricular systolic dysfunction, myocardial infarction, progressive renal impairment, and scleroderma renal crisis.

In the captopril experiment, C3A cells were treated with captopril at 2, 0.5, 48, and 60 $\mu\text{g/ml}$ for 1, 3, and 6 hours. The treated cells were compared to untreated early confluent C3A cells. The experiment showed that at all time points assayed (1, 3, and 6 hours), the expression of SEQ ID NO:38, SEQ ID NO:39 and SEQ ID NO:42 was decreased by at least 2 fold in CA3 cells at all treatment levels when compared to untreated CA3 cells in the absence of captopril. In the enalapril experiment, C3A cells were treated with enalapril at 2, 0.5, 48, and 60 $\mu\text{g/ml}$ for 1, 3, and 6 hours. The treated cells were compared to untreated early confluent C3A cells. The experiment showed that at the 3 and 6 hour time points, the expression of SEQ ID NO:38, SEQ ID NO:39 and SEQ ID NO:42 was decreased by at least 2 fold in CA3 cells at the 2, 48, and 60 $\mu\text{g/ml}$ treatment levels when compared to untreated CA3 cells in the absence of enalapril. The two experiments indicate that SEQ ID NO:38, SEQ ID NO:39 and SEQ ID NO:42 are useful in diagnosis, prognosis, or treatment of hypertension and cardiovascular disorders.

In an alternative example, SEQ ID NO:45 showed differential expression, as determined by microarray analysis, in liver C3A cells treated with one of the following steroids: progesterone, beclomethasone, medroxyprogesterone, budesonide, prednisone, dexamethasone, and betamethasone. C3A cells were treated with 1, 10, and 100 μM progesterone, beclomethasone, medroxyprogesterone, budesonide, prednisone, dexamethasone, and betamethasone for 1hr, 3hr, 6hr and were compared with untreated cells. In one experiment, SEQ ID NO:45 showed at least a two-fold increase in expression at all time points in C3A cells treated with medroxyprogesterone, budesonide, prednisone and betamethasone. In one experiment, SEQ ID NO:45 showed at least a two-fold increase in expression at a minimum of two out of the three time points in C3A cells treated with beclomethasone. SEQ ID NO:45 showed at least a two-fold increase in expression in C3A cells treated with 10 microM progesterone for 1, 3, and 6 hours. In one experiment, SEQ ID NO:45 showed at least a two-fold increase in expression in C3A cells treated with 10 microM beclomethasone for 3 and 6 hours. In a separate experiment, SEQ ID NO:45 showed at least a two-fold increase in expression in C3A cells treated with 1 microM budesonide for 1, 3, and 6 hours. SEQ ID NO:45 also showed at least a two-fold increase in expression in C3A cells treated with 100

microM dexamethasone for 1, 3, and 6 hours. These experiments indicate that SEQ ID NO:45 is useful for monitoring the pharmacodynamics of drugs and effects on liver metabolism upon steroid therapy.

SEQ ID NO:45 also showed differential expression in Alzheimer's Disease (AD), as determined by microarray analysis. AD is a progressive neurodegenerative disorder that is characterized by the formation of senile plaques and neurofibrillary tangles containing amyloid beta peptide. These plaques are found in limbic and association cortices of the brain. The hippocampus is part of the limbic system and plays an important role in learning and memory. In subjects with AD, accumulating plaques damage the neuronal architecture in limbic areas and eventually cripple the memory process. In a comparison of amygdala and anterior hippocampus brain tissue from a 79-year-old female with severe AD to identical tissues from a normal 61-year-old female, the expression of SEQ ID NO:45 was decreased at least two-fold. Therefore, SEQ ID NO:45 is useful in diagnostic assays for AD and as a potential biological marker and therapeutic agent in the treatment of AD.

In an alternative example, SEQ ID NO:51 showed differential expression associated with autoimmune and inflammatory responses, as determined by microarray analysis. The gene expression profiles of treated PHA blasts derived from the PBMCs of 5 healthy volunteer donors were compared to matching untreated PHA blasts. Treated PBMCs were stimulated for 12 days in the presence of PHA and IL-2. These T cell blasts were washed and stimulated for 2 hours in the presence of anti-CD3 monoclonal antibody, anti-CD28 antibody, a combination of both antibodies, PMA, ionomycin, and a combination of PMA and ionomycin. Expression of SEQ ID NO:51 was increased by at least fourfold during the treatment course in treated human T cells from a healthy donor when compared to untreated T cells from the same donor. Therefore, SEQ ID NO:51 is useful in monitoring treatment of, and diagnostic assays for autoimmune and inflammatory disorders.

As another example, SEQ ID NO:51 showed differential expression associated with autoimmune and inflammatory responses, as determined by microarray analysis. The expression of SEQ ID NO:51 was increased by at least eightfold consecutively throughout a time course of PBMCs stimulated *in vitro* with soluble PMA and ionomycin for 1, 2, 4, 8, and 20 hours compared to untreated PBMCs kept in culture in the absence of stimuli. Thus, SEQ ID NO:51 is useful in monitoring treatment of, and diagnostic assays for autoimmune and inflammatory disorders.

XII. Complementary Polynucleotides

Sequences complementary to the EXMES-encoding sequences, or any parts thereof, are used 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' 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 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).

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 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; 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, 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 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

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 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-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.

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 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 ¹²⁵I 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 (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.

5 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
10 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
15 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
20 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
25 proliferative activity. Varying amounts of EXMES are added to cultured leukocytes, such as 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
30 synthesized DNA. A linear dose-response curve over at least a hundred-fold EXMES concentration 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 7:291-301). In this assay, about 10^5 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 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 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 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 ^{125}I -labeled IgG specific for the primary antibody as the secondary antibody.

EXMES protease inhibitor activity for alpha 2-HS-glycoprotein (AHSG) can be measured as a decrease in osteogenic activity in dexamethasone-treated rat bone marrow cell cultures (dex-RBMC). Assays are carried out in 96-well culture plates containing minimal essential medium supplemented with 15% fetal bovine serum, ascorbic acid (50 $\mu\text{g/ml}$), antibiotics (100 $\mu\text{g/ml}$ penicillin G, 50 $\mu\text{g/ml}$ gentamicin, 0.3 $\mu\text{g/ml}$ fungizone), 10 mM B-glycerophosphate, dexamethasone (10^{-8} M) and various concentrations of EXMES for 12-14 days. Mineralized tissue formation in the cultures is quantified by measuring the absorbance at 525 nm using a 96-well plate reader (Binkert, C. et al. *supra*).

EXMES protease inhibitor activity for inter-alpha-trypsin inhibitor (ITI) can be measured by a continuous spectrophotometric rate determination of trypsin activity. The assay is performed at ambient temperature in a quartz cuvette in pH 7.6 assay buffer containing 63 mM sodium phosphate, 0.23 mM N α -benzoyl-L-arginine ethyl ester, 0.06 mM hydrochloric acid, 100 units trypsin, and various concentrations of EXMES. Immediately after mixing by inversion, the increase in $A_{253\text{ nm}}$ is

recorded for approximately 5 minutes and the enzyme activity is calculated (Bergmeyer, H.U. et al. (1974) Meth. Enzym. Anal. 1:515-516).

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. 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 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
7486368	1	7486368CD1	29	7486368CB1	90160611CA2, 90160619CA2, 90160627CA2, 90160635CA2, 90160643CA2, 90160703CA2, 90160727CA2, 90160858CA2, 90160958CA2, 90212936CA2, 90212944CA2, 90212952CA2, 90212960CA2, 90213036CA2, 90213044CA2, 90213052CA2, 95044302CA2, 95044326CA2, 95044334CA2, 95044342CA2, 95044350CA2, 95044358CA2, 95044366CA2, 95044390CA2, 95044402CA2, 95044566CA2, 95044582CA2
7488284	2	7488284CD1	30	7488284CB1	90110995CA2, 90111055CA2, 90111095CA2
7500756	3	7500756CD1	31	7500756CB1	70511573CA2
7506418	4	7506418CD1	32	7506418CB1	4943366CA2
7506942	5	7506942CD1	33	7506942CB1	90118901CA2
7506998	6	7506998CD1	34	7506998CB1	90119844CA2
7506999	7	7506999CD1	35	7506999CB1	90119967CA2
7509150	8	7509150CD1	36	7509150CB1	
1576743	9	1576743CD1	37	1576743CB1	
7503892	10	7503892CD1	38	7503892CB1	
7503895	11	7503895CD1	39	7503895CB1	
7509192	12	7509192CD1	40	7509192CB1	
7509320	13	7509320CD1	41	7509320CB1	90138264CA2
7503898	14	7503898CD1	42	7503898CB1	
7509185	15	7509185CD1	43	7509185CB1	90138132CA2, 90138156CA2, 90138232CA2, 90138248CA2, 90138280CA2, 90138487CA2
7509401	16	7509401CD1	44	7509401CB1	
7510191	17	7510191CD1	45	7510191CB1	4092112CA2, 6568665CA2, 71509586CA2
7506943	18	7506943CD1	46	7506943CB1	90119001CA2, 90130396CA2, 90130472CA2
7500179	19	7500179CD1	47	7500179CB1	90018159CA2, 90018167CA2, 90018175CA2, 90018191CA2, 90018251CA2, 90018259CA2, 90018267CA2
7509149	20	7509149CD1	48	7509149CB1	90119001CA2, 90130396CA2, 90130472CA2

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7509399	21	7509399CD1	49	7509399CB1	90138855CA2
7504789	22	7504789CD1	50	7504789CB1	
7506910	23	7506910CD1	51	7506910CB1	90113323CA2, 90113331CA2, 90113423CA2, 90150101CA2
7507000	24	7507000CD1	52	7507000CB1	90119828CA2, 90119927CA2, 90119951CA2, 90119975CA2,
7510882	25	7510882CD1	53	7510882CB1	90120035CA2
7506575	26	7506575CD1	54	7506575CB1	5951292CA2
7506576	27	7506576CD1	55	7506576CB1	90117477CA2
7512018	28	7512018CD1	56	7512018CB1	90117337CA2
					95005508CA2

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7486368CD1	g198930	2.4E-09	[Mus musculus] differentiation antigen Ly-6E/A (Khan,K.D. et al. (1990) Mol. Cell. Biol. 10 (10), 5150-5159)
2	7488284CD1	g307041	1.4E-171	[Homo sapiens] interferon-inducible 56 Kd protein (Wathelet,M. et al. (1986) Eur. J. Biochem. 155 (1), 11-17)
3	7500756CD1	g8572053	1.2E-15	[Homo sapiens] neurokinin B (Page,N.M. et al. (2000) Nature 405 (6788), 797-800)
3	7500756CD1	606518 TAC3	1E-16	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Tachykinin 3, precursor of the peptide hormone neuromedin K (neurokinin beta, neurokinin B), may be involved in regulating estrogen negative feedback; excessive placental secretion may be a cause of pre-eclampsia and pregnancy-induced hypertension Tian, Y., et al. (1996) J. Biol. Chem. 271, 20250-20257
4	7506418CD1	g6911131	3.5E-101	[Gallus gallus] fibroblast growth factor homologous factor 2 isoform 1Q+1Y' (Munoz-Sanjuan,I. Et al. (2000) J. Biol. Chem. 275 (4), 2589-2597)
4	7506418CD1	752253 FGF13	6.1E-97	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Fibroblast growth factor 13, has probable roles in nervous system development and function Greene, J. M., et al. (1998) Eur J Neurosci. 10, 1911-1925
4	7506418CD1	584781 Fgf13	1.3E-96	[Mus musculus][Ligand][Nuclear] Fibroblast growth factor 13, has probable roles in nervous system development and function
5	7506942CD1	g386824	5.2E-38	[Homo sapiens] interleukin 7 (Brunton,L.L. et al. (1990) Nucleic Acids Res. 18 (5), 1315)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
5	7506942CD1	336016 IL7	4.4E-39	[Homo sapiens][Ligand] Interleukin 7, a hematopoietic growth factor required for normal growth and development of B cells and T cells Goodwin, R. G. et al. (1989) Proc Natl Acad Sci U S A 86, 302-306. Alderson, M. R. et al. (1991) J Exp Med 173, 923-930. Llano, A. et al. (2001) J Virol 75, 10319-10325.
5	7506942CD1	583379 IL7	1.8E-26	[Mus musculus][Ligand] Interleukin 7, a hematopoietic growth factor required for normal growth and development of B cells and T cells, induces T cell-mediated anti-tumor response Namen, A. E. et al. (1988) Nature 333, 571-573.
5	7506942CD1	590275 IL7	1.3E-25	[Rattus norvegicus][Ligand] Interleukin 7, a hematopoietic growth factor that is involved in the growth and development of B cells
6	7506998CD1	g338075 -	7E-98	[Homo sapiens] sex hormone-binding globulin (Hammond, G.L. et al. (1989) Mol. Endocrinol. 3 (11), 1869-1876)
6	7506998CD1	568830 SHBG	5.9E-99	[Homo sapiens][Structural protein; Small molecule-binding protein][Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), a glycoprotein that binds and transports androgens and leads to protein kinase A (PRKACB) activation; plasma levels are associated with obesity, heart disease, and noninsulin-dependent diabetes Avvakumov, G. V. et al. (2000) J Biol Chem 275, 25920-25925. Hammond, G. L. and Bocchinfuso, W. P. (1996) Horm Res 45, 197-201.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
6	7506998CD1	585825 Shbg	5E-63	[Mus musculus][Small molecule-binding protein][Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), binds and transports androgens and is involved in spermatogenesis; plasma levels of human SHBG are associated with obesity, heart disease, and noninsulin-dependent diabetes Wang, Y. M. et al. (1989) Mol Cell Endocrinol 63, 85-92.
6	7506998CD1	590603 Shbg	2.1E-62	[Rattus norvegicus][L-yase; Small molecule-binding protein][Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), binds and transports androgens and is involved in spermatogenesis; plasma levels of human SHBG are associated with obesity, heart disease, and noninsulin-dependent diabetes Joseph, D. R. et al. (1987) Proc Natl Acad Sci U S A 84, 339-343.
7	7506999CD1	g338075	1.2E-63	[Homo sapiens] sex hormone-binding globulin (Hammond, G.L. et al. (1989) Mol. Endocrinol. 3 (11), 1869-1876)
7	7506999CD1	568830 SHBG	1.6E-66	[Homo sapiens][Structural protein; Small molecule-binding protein][Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), a glycoprotein that binds and transports androgens and leads to protein kinase A (PRKACB) activation; plasma levels are associated with obesity, heart disease, and noninsulin-dependent diabetes
7	7506999CD1	585825 Shbg	6.1E-42	[Mus musculus][Small molecule-binding protein][Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), binds and transports androgens and is involved in spermatogenesis; plasma levels of human SHBG are associated with obesity, heart disease, and noninsulin-dependent diabetes Wang, Y. M. et al. (supra)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
7	7506999CD1	590603 Shbg	9.9E-42	[Rattus norvegicus][L-yase; Small molecule-binding protein][Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), binds and transports androgens and is involved in spermatogenesis; plasma levels of human SHBG are associated with obesity, heart disease, and noninsulin-dependent diabetes
8	7509150CD1	g14423573	5.3E-45	[Homo sapiens] colony stimulating factor 3 (granulocyte)
8	7509150CD1	344100 CSF3	4.5E-46	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Granulocyte colony stimulating factor, a glycoprotein that regulates proliferation, differentiation, and survival of neutrophilic granulocytes and recruitment to sites of infection, implicated in neutrophilia in acute respiratory distress
				Souza, L. M. et al. (1986) Science 232, 61-65.
				Saito, S. et al. (1994) Growth Factors 10, 135-143.
9	1576743CD1	g4104963	2.6E-91	[Rattus norvegicus] neutrophil 4 (Missler, M. et al. (1998) J. Neurosci. 18:3630-8)
9	1576743CD1	624406 Nph4	2.2E-92	[Rattus norvegicus][Ligand] Neurexophilin 4, a secreted neuronal glycoprotein that may function as ligands for alpha-neurexins
				Missler, M. et al. 1998 (supra)
9	1576743CD1	427007 NXPH2	1.3E-42	[Homo sapiens][Ligand] Neurexophilin 2, a putative secreted neuronal glycoprotein that may be N-glycosylated and proteolytically processed after synthesis in neuron-like cells, suggested to function as a neuropeptide, may bind to and signal via alpha-neurexins
10	7503892CD1	g1311661	8.7E-242	[Homo sapiens] hepatocyte growth factor-like protein (Waltz, S.E. et al. (1996) J. Biol. Chem. 271 (15), 9024-9032)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
10	7503892CD1	613451 MST1	7.4E-243	[Homo sapiens][Ligand;Hydrolase;Protease (other than proteasomal)][Extracellular (excluding cell wall)] Macrophage stimulating protein, contains four kringle and a serine protease-like domain, regulates macrophage chemotaxis and activation, functions in ciliary motility, inhibits apoptosis, may function in blood coagulation and the oxidative stress response
				Sakamoto, O. et al. J Clin Invest 99; 701-9 (1997).
				Danilkovitch, A. et al. Mol Cell Biol 20, 2218-27 (2000)
10	7503892CD1	688972 Mst1	2.4E-196	[Rattus norvegicus][Ligand] Macrophage stimulating protein 1, a member of the hepatocyte growth factor family that signals via receptor tyrosine kinases and may play a role in spermatogenesis, sperm motility, and fertilizing capacity
				Ohshiro, K. et al., Biochem Biophys Res Commun 227, 273-80 (1996).
11	7503895CD1	g1311661	6.2E-161	[Homo sapiens] hepatocyte growth factor-like protein (Waltz, S.E. et al. (1996) J. Biol. Chem. 271 (15), 9024-9032)
11	7503895CD1	613451 MST1	5.3E-162	[Homo sapiens][Ligand;Hydrolase;Protease (other than proteasomal)][Extracellular (excluding cell wall)] Macrophage stimulating protein, contains four kringle and a serine protease-like domain, regulates macrophage chemotaxis and activation, functions in ciliary motility, inhibits apoptosis, may function in blood coagulation and the oxidative stress response
				Sakamoto, O. et al. (supra)
				Danilkovitch, A. et al. (supra)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
11	7503895CD1	688972 Mst1	7.3E-133	[Rattus norvegicus][Ligand] Macrophage stimulating protein 1, a member of the hepatocyte growth factor family that signals via receptor tyrosine kinases and may play a role in spermatogenesis, sperm motility, and fertilizing capacity
12	7509192CD1	g183096	1.5E-31	Okshiro, K. et al. (supra)
12	7509192CD1	335958 IGF2	1.3E-32	[Homo sapiens] insulin-like growth factor (Dull,T.J. et al. (1984) Nature 310 (5980), 777-781)
				[Homo sapiens][Ligand] Insulin-like growth factor II (somatomedin A), regulates embryonic growth; overproduction due to loss of suppression by WT1 or disruption of genomic imprinting is associated with Wilms' tumour, gigantism, and Beckwith-Wiedemann fetal overgrowth syndrome
12	7509192CD1	711876 Igf2	2.2E-28	Csernus, V. et al. J. Endocrinol. 163:269-80 (1999)
				[Rattus norvegicus][Ligand][Extracellular (excluding cell wall)] Insulin-like growth factor II, a member of the insulin protein family that may have a role in fetal development and in the central nervous system, may play a role in kidney development
13	7509320CD1	g4323579	2.7E-121	Kitraki, E. et al. Int. J. Dev. Neurosci. 11:1-9 (1993)
				[Homo sapiens] interleukin 12; natural killer cell stimulatory factor; cytotoxic lymphocyte maturation factor 35kDa subunit; p35 (Liu, Y. et al. (1994) Mol Cell Endocrinol 104, 173-81) (Wold,S.F. et al. (1991) J. Immunol. 146:3074-81)
13	7509320CD1	335982 IL12A	5E-102	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Interleukin 12A (natural killer cell stimulatory factor 1), involved in bacterial/viral/parasitic defense, inflammation, and antitumor activity; associated with melanomas, Hodgkin's disease, HIV, and recurrent infections

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
13	7509320CD1	697625 III12a	6.8E-27	Haraguchi, S. et al. Proc. Natl. Acad. Sci. U.S.A. 95:13125-9 (1998) [Mus musculus][Ligand][Extracellular (excluding cell wall)] Interleukin 12a, forms bioactive cytokine with IL12b, mediates Th1 cell immune response, NK cell activation, and IFNγ production, has a role in lipopolysaccharide-induced shock, experimental autoimmune encephalomyelitis and mouse model of Crohn's disease
14	7503898CD1	g1311661	0	Klein, T. W. et al. J. Immunol. 16: 6461-6 (2000) [Homo sapiens] hepatocyte growth factor-like protein (Waltz, S.E. et al. (1996) J. Biol. Chem. 271 (15), 9024-9032)
14	7503898CD1	613451 MST1	0.0	[Homo sapiens][Ligand;Hydrolase;Protease (other than proteasomal)] [Extracellular (excluding cell wall)] Macrophage stimulating protein, contains four kringle and a serine protease-like domain, regulates macrophage chemotaxis and activation, functions in ciliary motility, inhibits apoptosis, may function in blood coagulation and the oxidative stress response
				Sakamoto, O. et al. (supra)
				Danilkovitch, A. et al. (supra)
14	7503898CD1	688972 Mst1	0.0	[Rattus norvegicus][Ligand] Macrophage stimulating protein 1, a member of the hepatocyte growth factor family that signals via receptor tyrosine kinases and may play a role in spermatogenesis, sperm motility, and fertilizing capacity
				Ohshiro, K. et al. Biochem. Biophys. Res. Commun. 227:273-80 (1996)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
15	7509185CD1	g4323579	2E-72	[Homo sapiens] interleukin 12; natural killer cell stimulatory factor; cytotoxic lymphocyte maturation factor 35kDa subunit; p35 (Liu, Y. et al. (1994) Mol Cell Endocrinol 104, 173-81)
15	7509185CD1	335982 IL12A	8.2E-54	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Interleukin 12A (natural killer cell stimulatory factor 1), involved in bacterial/viral/parasitic defense, inflammation, and antitumor activity; associated with melanomas, Hodgkin's disease, HIV, and recurrent infections
				Haraguchi, S. et al. (supra)
				Link, A. A. et al. (supra)
15	7509185CD1	697625 II12a	1.5E-29	[Mus musculus][Ligand][Extracellular (excluding cell wall)] Interleukin 12a, forms bioactive cytokine with IL12b, mediates Th1 cell immune response, NK cell activation, and Ifng production, has a role in lipopolysaccharide-induced shock, experimental autoimmune encephalomyelitis and mouse model of Crohn's disease
				Klein, T. W. et al. (supra)
				Rempel, J. D. et al. (supra)
16	7509401CD1	g386824	3.7E-78	[Homo sapiens] interleukin 7 (Brunton, L. L. et al. (1990) Nucleic Acids Res. 18 (5), 1315)
16	7509401CD1	336016 IL7	3E-79	[Homo sapiens][Ligand] Interleukin 7, a hematopoietic growth factor required for normal growth and development of B cells and T cells
				Goodwin, R. G. et al. Proc Natl Acad Sci U S A 86, 302-6 (1989).
16	7509401CD1	583379 II7	2.4E-54	[Mus musculus][Ligand] Interleukin 7, a hematopoietic growth factor required for normal growth and development of B cells and T cells, induces T cell-mediated anti-tumor response
				DeKoter, R. P. et al. Science 288, 1439-41 (2000).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
17	7510191CD1	g206344	1.1E-55	Mertsching, E. et al. <i>Int. Immunol.</i> 7, 401-14 (1995). [Rattus norvegicus] preprotachykinin (Krause, J.E. et al. (1987) <i>Proc. Natl. Acad. Sci. U.S.A.</i> 84 (3), 881-885)
17	7510191CD1	570976 TAC1	6.7E-59	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Preprotachykinin 1, precursor of substance P and neurokinin A (substance K), neuropeptides of the tachykinin family that are involved in inflammatory and neuroendocrine responses and signal through G protein-coupled receptors
17	7510191CD1	430666 Tac1	7.4E-53	Cao, Y.Q. et al. <i>Nature</i> 392, 390-4 (1998). [Rattus norvegicus][Ligand][Axon] Preprotachykinin 1, precursor of substance P and neurokinin A (substance K), neuropeptides of the tachykinin family that are involved in inflammatory and neuroendocrine responses and signal through G protein-coupled receptors
				Cao, Y. Q. et al. (supra)
				van de Witte, S.V. et al. <i>Brain Res Mol Brain Res</i> 61, 219-23 (1998).
18	7506943CD1	g386824	8.1E-12	[Homo sapiens] interleukin 7 (Brunton, L.L. et al. (1990) <i>Nucleic Acids Res.</i> 18 (5), 1315)
19	7500179CD1	g386853	4.5E-43	[Homo sapiens] kininogen
19	7500179CD1	336180 KNG	3.7E-44	[Homo sapiens][Activator; Inhibitor or repressor][Extracellular (excluding cell wall); Cell junction] Bradykinin, a peptide hormone that is cleaved from H-kininogen at the site of infection and mediates inflammatory responses, activates MAP kinase, inhibits cell adhesion; genetic variants are associated with Williams' trait, a blood clotting disorder
				Cheung, P. P. et al., Genetic basis of total kininogen deficiency in Williams' trait, <i>J Biol Chem</i> 268, 23361-5 (1993)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
19	7500179CD1	590821 Kngk	9.5E-29	[Rattus norvegicus][Ligand; Structural protein][Extracellular (excluding cell wall)] Kininogen, precursor of plasma kinins, a peptide mediator released during inflammation, can activate phospholipase C, involved in pain perception, inhibition of activation decreases inflammation in a model of chronic granulomatous intestinal inflammation
				Kageyama, R. et al., Differing utilization of homologous transcription initiation sites of rat K and T kininogen genes under inflammation condition, J Biol Chem 262, 2345-51 (1987)
20	7509149CD1	g386824	8.2E-12	[Homo sapiens] interleukin 7 (Brunton, L.L. et al. (1990) Nucleic Acids Res. 18 (5), 1315)
20	7509149CD1	336016 IL7	6.6E-13	[Homo sapiens][Ligand] Interleukin 7, a hematopoietic growth factor required for normal growth and development of B cells and T cells
				Goodwin, R. G. et al., Human interleukin 7: molecular cloning and growth factor activity on human and murine B-lineage cells, Proc Natl Acad Sci U S A 86, 302-6 (1989)
20	7509149CD1	590275 II7	2.9E-10	[Rattus norvegicus][Ligand] Interleukin 7, a hematopoietic growth factor that is involved in the growth and development of B cells
				Namen, A. E. et al., Stimulation of B-cell progenitors by cloned murine interleukin-7, Nature 333, 571-3 (1988)
21	7509399CD1	g183041	8.3E-83	[Homo sapiens] granulocyte colony stimulating factor precursor (Devlin, J.J. et al. (1987) J. Leukoc. Biol. 41 (4), 302-306)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
21	7509399CD1	344100 CSF3	1.4E-85	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Granulocyte colony stimulating factor, a glycoprotein that regulates proliferation, differentiation, and survival of neutrophilic granulocytes and recruitment to sites of infection, implicated in neutrophilia in acute respiratory distress
				Nagata, S. et al., Molecular cloning and expression of cDNA for human granulocyte colony-stimulating factor, Nature 319, 415-8 (1986)
21	7509399CD1	590981 Csf3	1.9E-53	[Rattus norvegicus][Ligand][Extracellular (excluding cell wall)] Granulocyte colony stimulating factor, stimulates neutrophil recruitment to sites of infection, involved in osteoclastic bone resorption.
				Zhang, P. et al., The effects of granulocyte colony-stimulating factor and neutrophil recruitment on the pulmonary chemokine response to intratracheal endotoxin, J Immunol 166, 458-65. (2001)
22	7504789CD1	g2707181	0.0	[Homo sapiens] thyroglobulin
22	7504789CD1	623574 TG	0.0	[Homo sapiens][Ligand][Extracellular (excluding cell wall)] Thyroglobulin, a glycoprotein precursor of the thyroid hormones thyroxine and triiodothyronine which regulate metabolic processes; mutation of corresponding gene causes congenital goiter with hypothyroidism and endoplasmic reticulum storage disease

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Kim, P. S. et al., A single amino acid change in the acetylcholinesterase-like domain of thyroglobulin causes congenital goiter with hypothyroidism in the cog/cog mouse: a model of human endoplasmic reticulum storage diseases, Proc Natl Acad Sci U S A 95, 9909-13 (1998)
22	7504789CD1	747787 Tg	0.0	[Rattus norvegicus][Cytoplasmic; Plasma membrane] Thyroglobulin, a glycoprotein precursor of the thyroid hormones thyroxine and triiodothyronine which regulate metabolic processes, contributes to repression of thyroid-specific proteins; mutation of human TG causes congenital goiter with hypothyroidism
				Graves, P. N. et al., A second thyroglobulin messenger RNA species (rTg-2) in rat thyrocytes, Mol Endocrinol 4, 155-61 (1990)
23	7506910CD1	g33784	7E-20	[Homo sapiens] interleukin (Holbrook,N.J. et al. (1984) Nucleic Acids Res. 12 (12), 5005-5013)
23	7506910CD1	618366 IL2	6.2E-19	[Homo sapiens][Ligand][Golgi; Cytoplasmic; Extracellular (excluding cell wall)] Interleukin-2 (T-cell growth factor), a T-cell-derived cytokine that promotes activation and proliferation of lymphocytes, involved in the immune response, implicated in Sjorgen's syndrome, autoimmune hemolytic anemia, and multiple sclerosis
				Landay, A. L. et al., Cytokine modulation of HIV-1 chemokine receptor expression, Nat Med 5, 592-3 (1999)
23	7506910CD1	757398 IL2	2.3E-12	[Rattus norvegicus][Ligand] Interleukin-2, a T-cell-derived cytokine that promotes growth of B and T cells, likely involved in the immune response; human IL2 is implicated in Sjorgen's syndrome, autoimmune hemolytic anemia, and multiple sclerosis

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Lowenthal, J. W. et al. (supra)
				Grone, H. J. et al., Met-RANTES reduces vascular and tubular damage during acute renal transplant rejection: blocking monocyte arrest and recruitment, FASEB J 13, 1371-83 (1999)
24	7507000CD1	g296673	7.6E-13	[Homo sapiens] sex hormone-binding globulin (SHBG) (Gershagen, S. et al. (1989) Nucleic Acids Res. 17 (22), 9245-9258)
24	7507000CD1	568830 SHBG	6.5E-14	[Homo sapiens][Structural protein; Small molecule-binding protein][Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), a glycoprotein that binds and transports androgens and leads to protein kinase A (PRKACB) activation; plasma levels are associated with obesity, heart disease, and noninsulin-dependent diabetes
				Misao, R. et al., Expression of sex hormone-binding globulin exon VII splicing variant messenger RNA in human uterine endometrial cancers, Cancer Res 57, 5579-83 (1997)
24	7507000CD1	590603 Shbg	0.094	[Rattus norvegicus][Lyase; Small molecule-binding protein][Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), binds and transports androgens and is involved in spermatogenesis; plasma levels of human SHBG are associated with obesity, heart disease, and noninsulin-dependent diabetes
				Joseph, D. R. et al., Rat androgen-binding protein: evidence for identical subunits and amino acid sequence homology with human sex hormone-binding globulin, Proc Natl Acad Sci U S A 84, 339-43 (1987)
25	7510882CD1	g338075	1.2E-127	[Homo sapiens] sex hormone-binding globulin (Hammond, G.L. et al. (1989) Mol. Endocrinol. 3 (11), 1869-1876)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
25	7510882CD1	568830 SHBG	1E-128	[Homo sapiens][Structural protein; Small molecule-binding protein][Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), a glycoprotein that binds and transports androgens and leads to protein kinase A (PRKACB) activation; plasma levels are associated with obesity, heart disease, and noninsulin-dependent diabetes
				Gershagen, S. et al. (supra), Hammond, G. L. et al. (supra), Misao, R. et al. (supra)
25	7510882CD1	590603 Shbg	2.7E-103	[Rattus norvegicus][L-yase; Small molecule-binding protein][Extracellular (excluding cell wall)] Sex hormone binding globulin (androgen binding protein), binds and transports androgens and is involved in spermatogenesis; plasma levels of human SHBG are associated with obesity, heart disease, and noninsulin-dependent diabetes
				Fazzari, A. et al., The control of progesterone receptor expression in MCF-7 breast cancer cells: effects of estradiol and sex hormone-binding globulin (SHBG), Mol Cell Endocrinol 172, 31-6. (2001)
26	7506575CD1	g177890	4.7E-175	[Homo sapiens] prepro alpha-2-thiol proteinase inhibitor
26	7506575CD1	336180 BDK	4.1E-176	[Homo sapiens][Activator; Inhibitor or repressor][Extracellular (excluding cell wall); Cell junction] Bradykinin, a peptide hormone that is cleaved from H-kininogen at the site of infection and mediates inflammatory responses, activates MAP kinase, inhibits cell adhesion; genetic variants are associated with Williams' trait, a blood clotting disorder

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
26	7506575CD1	664317 Kng	7.2E-110	[Mus musculus] Kininogen, a precursor of plasma kinins including bradykinin, which is a peptide mediator released during inflammation that potently excites sympathetic neurons
27	7506576CD1	g177890	4.8E-195	[Homo sapiens] prepro alpha-2-thiol proteinase inhibitor
27	7506576CD1	336180 BDK	4.2E-196	[Homo sapiens][Activator;Inhibitor or repressor][Extracellular (excluding cell wall);Cell junction] Bradykinin, a peptide hormone that is cleaved from H-kininogen at the site of infection and mediates inflammatory responses, activates MAP kinase, inhibits cell adhesion; genetic variants are associated with Williams' trait, a blood clotting disorder
27	7506576CD1	664317 Kng	1.8E-119	[Mus musculus] Kininogen, a precursor of plasma kinins including bradykinin, which is a peptide mediator released during inflammation that potently excites sympathetic neurons
28	7512018CD1	g386853	1.9E-133	[Homo sapiens] kininogen
28		336180 KNG	1.5E-134	[Homo sapiens][Activator; Inhibitor or repressor] [Extracellular (excluding cell wall); Cell junction] Bradykinin, a peptide hormone that is cleaved from H-kininogen at the site of infection and mediates inflammatory responses, activates MAP kinase, inhibits cell adhesion; genetic variants are associated with Williams' trait, a blood clotting disorder
				Cheung, P. P. et al. Genetic basis of total kininogen deficiency in Williams' trait. J Biol Chem 268, 23361-5 (1993).
28		664317 Kng	3.7E-83	[Mus musculus] Kininogen, a precursor of plasma kinins including bradykinin, may be involved in the inflammatory response
				Pesquero, J. B. et al. Hypoalgesia and altered inflammatory responses in mice lacking kinin B1 receptors. Proc Natl Acad Sci U S A 97, 8140-8145 (2000).

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7486368CD1	94	S43 S47 S62 S86 T2 T67		signal_cleavage: M1-A31	SPSCAN
					Signal Peptide: M1-C36, M14-G33, M14-A31, M14-C36, M1-G33, M8-G33, M8-A31	HMMER
					u-PAR/Ly-6 domain: M8-V51	HMMER_PFAM
					Ly-6 / u-PAR domain proteins BL00983: L19-C27, R30-C39	BLIMPS_BLOCKS
2	7488284CD1	474	S74 S195 S266 T22 T116 T124 T162 T349 Y117 Y218 Y343		TPR Domain: V437-L470, V141-N174, A251-S284, V52-E85, E340-E373, L95-F128	HMMER_PFAM
					PROTEIN INTERFERON INTERFERON-INDUCED INDUCTION INDUCIBLE IFI54K ISG54 K WITH TETRATRICPEPTIDE	BLAST_PRODROM
					PD020984: E340-L470	BLAST_PRODROM
					PD009625: L278-F339	BLAST_PRODROM
					INTERFERON; ALPHA; IFI56; ISG-54K; DM02459	BLAST_DOMO
					IP09914 1-453: M1-V453	BLAST_DOMO
					IB55508 1-464: M1-F474	BLAST_DOMO
					IP09913 1-470: M1-L464	BLAST_DOMO
					IA55508 1-467: M1-G451	BLAST_DOMO
3	7500756CD1	44	S37		signal_cleavage: M1-A16	SPSCAN
					Signal Peptide: M1-A16, M1-S18, M1-G20, M1-C23, M1-K24	HMMER
					NEUROKININ B PRECURSOR NEUROMEDIN K TACHYKININ NEUROPEPTIDE CLEAVAGE ON PAIR PD020370: M1-D40 (P-value = 7.5e-10)	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					NEUROKININ B PEPTIDE DM07934 P08858 1-125: M1-D40 (P-value = 9.0e-10)	BLAST_DOMO
4	7506418CD1	199	S2 S100 S119 S162 S194 T46 Y53	N196	Fibroblast growth factor: I23-P152	HMMER_PFAM
					HBGF/FGF family proteins BL00247: G22-G32, I68-Y114, W123-P150	BLIMPS_BLOCKS
					HBGF/FGF family signature: V65-S119	PROFESCAN
					Heparin binding growth factor signature PR00263: T25-L37, F56-G70, Y76-S88, T92-S111	BLIMPS_PRINTS
					IL1/HBGF family signature PR00262: L63-L90, E94-Y114, K141-P150	BLIMPS_PRINTS
					GROWTH FACTOR FIBROBLAST PRECURSOR MITOGEN SIGNAL HEPARINBINDING GLYCOPROTEIN PROTEIN VASCULARIZATION PD000831: K26-K151	BLAST_PRODROM
					FIBROBLAST GROWTH FACTOR HOMOLOGOUS FACTOR13 FGF13 FHF2 FACTOR 14 FGF14 FHF4 PD006656: P152-T199	BLAST_PRODROM
					HBGF/FGF FAMILY DM00642 P31371 13-190: L20-P152 DM00642 P48804 36-193: P8-K151 DM00642 P48805 29-186: E10-K151 DM00642 P12034 47-218: L27-K151	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): G185-S192	MOTIFS
					HBGF/FGF family signature: G83-Y106	MOTIFS
5	7506942CD1	115	S5 S25 S51 S81	N79	signal_cleavage: M1-C27	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Signal Peptide: M1-S25	HMMER
					Interleukin 7/9 family: C27-G111	HMMER_SMART
					Interleukin 7/9 family: D28-G111	HMMER_PFAM
					Interleukin-7 and -9 proteins	BLIMPS_BLOCKS
					BL00255: M1-M42, N89-G111	BLIMPS_BLOCKS
					Interleukin-7 signature	BLIMPS_PRINTS
					PR00435: F2-S25, D26-L48, S57-E77	BLAST_PRODROM
					Interleukin 7 Precursor IL7 Cytokine Growth Factor Glycoprotein Signal	BLAST_PRODROM
					M1-E100, N75-T112	BLAST_PRODROM
					Interleukin-7	BLAST_PRODROM
					DM07444	BLAST_PRODROM
					P26895 1-175: M1-T112, N75-H115	BLAST_PRODROM
					P10168 1-153: M1-E100, I71-T112	BLAST_PRODROM
					Interleukin-7 and -9 signature: N89-L98	MOTIFS
6	7506998CD1	201	S38 S71 S162 T9 T161		signal_cleavage: M1-R25	SPSCAN
					Signal Peptide: M1-Q26	HMMER
					Signal Peptide: M1-A29	HMMER
					Laminin G domain: T67-L192	HMMER_SMART
					Laminin G domain: V75-S189	HMMER_PFAM
					LAMININ CHAIN EGF-LIKE DOMAIN	BLIMPS_PRODROM
					PD00320: G138-L151	BLIMPS_PRODROM
					Protein Precursor Signal Sex Hormone-Binding Globulin SHBG Steroid-Binding SBP Testis Specific: Q26-E74	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Sex Hormone-Binding Globulin DM01077 P04278 36-198: T36-P185 P15196 1-163: T36-P185 P08689 36-199: P35-P185 A39030 36-218: P35-P185	BLAST_DOMO
7	7506999CD1	181	S38 S71 S135 T9 T172		signal_cleavage: M1-R25	SPSCAN
					Signal Peptide: M1-Q26	HMMER
					Signal Peptide: M1-A29	HMMER
					Laminin G domain: V75-W127	HMMER_PFAM
					Protein Precursor Signal Sex Hormone-Binding Globulin SHBG Steroid-Binding SBP Testis Specific: Q26-E74	BLAST_PRODROM
					Sex Hormone-Binding Globulin DM01077 P04278 36-198: T36-W127 P15196 1-163: T36-W127 P08689 36-199: P35-P129 A39030 36-218: P35-P129	BLAST_DOMO
8	7509150CD1	169	S90 S154 S159 T26		signal_cleavage: M1-A30	SPSCAN
					Signal Peptide: M1-A30	HMMER
					Signal Peptide: M13-A30	HMMER
					Signal Peptide: M10-P32	HMMER
					Signal Peptide: M1-T26	HMMER
					Signal Peptide: M10-A30	HMMER
					Signal Peptide: M10-E29	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Interleukin-6/G-CSF/MGF family A66-R164, V51-L65	HMMER_PFAM
					Interleukin-6 / G-CSF / MGF proteins BL00254: L33-K46, A134-R164	BLIMPS_BLOCKS
					Interleukin-6 / G-CSF / MGF family signature V27-D104	PROFILESAN
					Interleukin-6/G-CSF/MGF family signature PR00433: Q50-L65, V148-R164	BLIMPS_PRINTS
					Interleukin-6 signature PR00434: A134-F155, L147-R164	BLIMPS_PRINTS
					Interleukin-6/G-CSF/MGF family PR00433: Q50-L65, V148-R164	BLIMPS_PRODROM
					FACTOR GROWTH LYCOPROTEIN CYTOKINE PRECURSOR SIGNAL INTERLEUKIN 6 IL6 GRANULOCYTE COLONY STIMULATING PD004356: C68-R164	BLAST_PRODROM
					FACTOR GRANULOCYTE COLONY STIMULATING PRECURSOR GCSF CYTOKINE GROWTH LYCOPROTEIN SIGNAL STIMULATING PD035419: M10-Q50	BLAST_PRODROM
					INTERLEUKIN-6 / G-CSF / MGF DM03308 IP35833 1-173: K64-P169, T31-H73 IA25093 31-204: Q28-Q168, T31-L103 IP09920 37-207: Q28-A167 IP13854 26-200: Q41-Q62, C68-Q168	BLAST_DOMO
9	I576743CD1	254	S104 T186	N79 N89 N95	signal_cleavage: M1-A54	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					NEUREXOPHILIN NEUROPHILIN PD039440: N79-F109, K172-F253	BLAST_PRODOM
10	7503892CD1	512	S109 S149 S199 S222 S260 S466 T161 T359 T376 T399 T453 Y158	N86 N310	signal_cleavage: M1-G29	SPSCAN
					Signal Peptide: M15-G29, M1-G29, M15-G32, M1-G32, M15-S35, M15-R34	HMMER
					PAN domain: G32-Y120	HMMER_PFAM
					Kringle domain: C297-C375, C205-C282, C124-C200, C384-D448	HMMER_PFAM
					Kringle domain: R122-E202, A203-S284, V295-D377, Q382-D448	HMMER_SMART
					PAN_AP, divergent subfamily of APPLE domains: S35-K118	HMMER_SMART
					Kringle domain proteins BL00021: S222-Q227, T267-P288	BLIMPS_BLOCKS
					Type I fibronectin domain proteins BL01253: G283-P326, K335-R373	BLIMPS_BLOCKS
					Kringle domain signature and profile: P153- G196, P234-D278, P326-C369	PROFILESKAN
					Kringle domain signature PR00018: Q247-T267, Q271-C282	BLIMPS_PRINTS
					PROTEIN GROWTH HEPATOCYTE FACTORLIKE PRECURSOR SIGNAL MACROPHAGE STIMULATORY MSP HOMOLOG PD007364: H50-T123	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PRECURSOR SIGNAL SERINE GLYCOPROTEIN PROTEASE KRINGLE HYDROLASE PLASMA GROWTH PLASMINOGEN PD000395: C124-C200, S296-C375, C200-C282	BLAST_PRODOM
					PROTEIN HEPATOCYTE GROWTH FACTORLIKE PRECURSOR SIGNAL MACROPHAGE STIMULATORY MSP KRINGLE PD012913: M15-Q49	BLAST_PRODOM
					KRINGLE DM00069 P26927 96-186: G110-R201, S296-C375, C205-C282, C384-K415 270-358: S284-R373, R201-C277, T123-C195 188-268: E202-G283, E291-C375, C124-C200 P14210 291-379: T294-I372, C205-C277, C124-C195	BLAST_DOMO
					Kringle domain signature: F170-D175, Y253-D258, F345-D350 signal_cleavage: M1-G29	MOTIFS
11	7503895CD1	468	S109 S149 S199 S222 S260 S340 S359 S361 S376 S422 T161 T358 T409 Y158	N86		SPSCAN
					Signal Peptide: M15-G29, M1-G29, M15-G32, M1-G32, M15-S35, M15-R34	HMMER
					PAN domain: G32-Y120	HMMER_PFAM
					Kringle domain: C205-C282, C124-C200	HMMER_PFAM
					Kringle domain: R122-E202, A203-S284	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PAN_AP, divergent subfamily of APPLE domains: S35-K118	HMMER_SMART
					Kringle domain proteins BL00021: S222-Q227, T267-T288	BLIMPS_BLOCKS
					Kringle domain signature and profile: P234-D278, P153-G196	PROFLESCAN
					Kringle domain signature PR00018: C124-T139, E221-H233, Q247-T267, Q271-C282	BLIMPS_PRINTS
					PROTEIN GROWTH HEPATOCYTE FACTOR-LIKE PRECURSOR SIGNAL MACROPHAGE STIMULATOR Y MSP HOMOLOG PD007364: H50-T123	BLAST_PRODROM
					PRECURSOR SIGNAL SERINE GLYCOPROTEIN PROTEASE KRINGLE HYDROLASE PLASMA GROWTH PLASMINOGEN PD000395: C124-C200, C200-C282	BLAST_PRODROM
					PROTEIN HEPATOCYTE GROWTH FACTOR-LIKE PRECURSOR SIGNAL MACROPHAGE STIMULATOR Y MSP KRINGLE PD012913: M15-Q49	BLAST_PRODROM
					KRINGLE DM00069 P26927 96-186: G110-R201, C205-C282 188-268: C124-C200, E202-C282 I59214 208-290: C124-C200, E202-T284 P14210 208-289: C124-E202, E202-C282	BLAST_DOMO
					Kringle domain signature: F170-D175, Y253-D258	MOTIFS
12	7509192CD1	140	S60 S127		signal_cleavage: M1-A24	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Signal Peptide: M5-C21, M1-C22, M5-A25, M1-A25, M5-A24, M1-A24, M1-R27, M1-T31	HMMER
					Insulin/GF/Relaxin family: E30-P82	HMMER_PFAM
					Insulin/insulin-like growth factor/relaxin family: E30-P82	HMMER_SMART
					Insulin family proteins BL00262: L32-G49	BLIMPS_BLOCKS
					Insulin B chain signature PR00277: S29-Q42, F43-P55	BLIMPS_PRINTS
					GROWTH INSULIN LIKE FACTOR PRECURSOR SIGNAL FAMILY INSULIN I PLASMA SOMATOMEDIN PD001048: A24-R58	BLAST_PRODROM
					INSULIN FAMILY DM00149	BLAST_DOMO
					P01344 26-83: Y26-Q68	
					S32558 26-83: Y26-Q68	
					P33717 1-58: Y26-Q68	
					Q02816 49-109: S29-Q68	
13	7509320CD1	239	S114 S133 S137 S146 S201 S202 T99	N127 N237	signal_cleavage: M1-A56	SPSCAN
					Signal Peptide: M35-L55, M35-A56	HMMER
					Interleukin-12 alpha subunit: M35-S239	HMMER_PFAM
					FACTOR CHAIN INTERLEUKIN PD02303: M35-H52, S137-K170, P198-S239	BLIMPS_PRODROM
					INTERLEUKIN12 ALPHA CHAIN PRECURSOR IL12A CYTOTOXIC LYMPHOCYTE MATURATION FACTOR SUBUNIT PD005790: H52-S239	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Interleukin-12 alpha chain precursor, DM07464: P54349 1-220: M35-S239 P43431 1-214: M35-S239	BLAST_DOMO
					Leucine zipper pattern: L72-L93	MOTIFS
					Aminoacyl-transfer RNA synthetases class-II signature 2: A226-Y235	MOTIFS
14	7503898CD1	757	S109 S149 S199 S222 S260 S469 S511 S674 T161 T359 T376 T399 T452 T533 T546 T614 T659 Y158	N86 N310 N661	signal_cleavage: M1-G29	SPSCAN
					Signal Peptide: M15-G29, M1-G29, M1-G32, M15-G32 M15-S35, M15-R34	HMMER
					Kringle domain: Q382-D464, A203-S284, V295-D377, R122-E202	HMMER_SMART
					Pan, divergent subfamily of APPLE domains: S35-K118	HMMER_SMART
					Trypsin-like serine protease: R497-I750	HMMER_SMART
					PAN domain: G32-Y120	HMMER_PFAM
					Kringle domain: C384-C462, C297-C375, C205-C282, C124-C200	HMMER_PFAM
					Trypsin: V498-I750	HMMER_PFAM
					Kringle domain proteins BL00021: S222-Q227, C521-F538, I601-G622, G709-I750	BLIMPS_BLOCKS
					Serine proteases, trypsin family, histidine proteins BL00134: C521-C537, G701-I724, P737-I750	BLIMPS_BLOCKS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Type I fibronectin domain proteins BL01253: G283-P326, K335-R373, C521-A534, V700-C713, W719-V753	BLIMPS_BLOCKS
					Kringle domain signature and profile: T412-A458, P234-D278, P153-G196, P326-C369	PROFILESKAN
					Serine proteases, trypsin family, active sites: N514-E563	PROFILESKAN
					Kringle domain signature PR00018: C297-T312, R400-T412, A426-Y446, R451-C462	BLIMPS_PRINTS
					Chymotrypsin serine protease family (S1) signature PR00722: G522-C537, P578-V592, V700-A712	BLIMPS_PRINTS
					PROTEIN GROWTH HEPATOCYTE FACTORLIKE PRECURSOR SIGNAL MACROPHAGE STIMULATORY MSP HOMOLOG PD007364: H50-T123	BLAST_PRODOM
					PRECURSOR SIGNAL SERINE GLYCOPROTEIN PROTEASE KRINGLE HYDROLASE PLASMA GROWTH PLASMINOGEN PD000395: S296-C375, D383-C462, C200-C282, C124-C200	BLAST_PRODOM
					PROTEASE SERINE PRECURSOR SIGNAL HYDROLASE ZYMOGEN GLYCOPROTEIN FAMILY MULTIGENE FACTOR PD000046: V498-I750	BLAST_PRODOM
					PROTEIN HEPATOCYTE GROWTH FACTORLIKE PRECURSOR SIGNAL MACROPHAGE STIMULATORY MSP KRINGLE PD012913: M15-Q49	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					KRINGLE DM00069 P26927 270-358; S284-R373, R201-C277, C384-Y456, T123-C195	BLAST_DOMO
					KRINGLE DM00069 P26927 360-450; R374-D465, R281-D378, C124-E202, R201-E285	BLAST_DOMO
					KRINGLE DM00069 P26927 96-186; G110-R201, C384-C462, S296-C375, C205-C282	BLAST_DOMO
					TRYPSIN DM00018 P26927 481-707; K495-G626 T656-M754, P153-C171	BLAST_DOMO
					Kringle domain signature: F170-D175, Y253-D258, F345-D350, F432-D437	MOTIFS
15	7509185CD1	164	S114 S142 S149 T99 T125	N127	signal_cleavage: M1-A56	SPSCAN
					Signal Peptide: M35-L55, M35-A56	HMMER
					Interleukin-12 alpha subunit: M35-Y161	HMMER_PFAM
					FACTOR CHAIN INTERLEUKIN PD02303: M35-H52, C71-L122, E123-L150	BLIMPS_PRODOM
					INTERLEUKIN12 ALPHA CHAIN PRECURSOR IL12A CYTOTOXIC LYMPHOCYTE MATURATION FACTOR SUBUNIT PD005790: H52-T140	BLAST_PRODOM
					Interleukin-12 alpha chain precursor DM07464 P54349 1-220: M35-T140 DM07464 P43431 1-214: M35-T140	BLAST_DOMO
					Leucine zipper pattern: L72-L93	MOTIFS
16	7509401CD1	159	S5 S25 S51 S96 S125 T118	N95 N116 N123	signal_cleavage: M1-C27	SPSCAN
					Signal Peptide: M1-S25	HMMER
					Interleukin 7/9 family: D28-G155	HMMER_PFAM
					Interleukin-7 and interleukin-9 family.: C27-G155	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Interleukin-7 and -9 proteins BL00255: M1-M42, K76-Q120, N133-G155	BLIMPS_BLOCKS
					Interleukin-7 signature PR00435: F2-S25, D26-L48, S57-E77, G78-G98, D99-T118, N133-G155	BLIMPS_PRINTS
					INTERLEUKIN7 PRECURSOR IL7 CYTOKINE GROWTH FACTOR GLYCOPROTEIN SIGNAL 3DSTRUCTURE PD013168: M1-M154, E121-T156	BLAST_PRODROM
					INTERLEUKIN-7 DM07444 P10168 1-153: M1-T156	BLAST_DOMO
					INTERLEUKIN-7 DM07444 P26895 1-175: M1-M154, E121-H159	BLAST_DOMO
					Interleukin-7 and -9 signature: N133-L142	MOTIFS
17	7510191CD1	114	S79		signal_cleavage: M1-A19	SPSCAN
					Signal Peptide: M1-T15, M1-A19, M1-E20, M1-G23, M1-A24	HMMER
					Tachykinin family: R58-M68, R82-M92	HMMER_SMART
					Tachykinin family proteins BL00267: P59-M68	BLIMPS_BLOCKS
					PROTACHYKININ PRECURSOR SUBSTANCE P TACHYKININ NEUROPEPTIDE CLEAVAGE ON PAIR OF PD005598: M1-R114	BLAST_PRODROM
					SUBSTANCE P PRECURSOR DM04784 S47039 48-115: F48-R114	BLAST_DOMO
					DM04784 JC2412 1-64: F48-Y110	
					DM04784 P01289 48-111: F48-R114	
18	7506943CD1	37	S5 S25		signal_cleavage: M1-C27	SPSCAN
					Signal Peptide: M1-S25	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Interleukin-7 and -9 proteins BL00255: M1-M38	BLIMPS_BLOCKS
					Interleukin-7 signature PR00435: F2-S25	BLIMPS_PRINTS
					INTERLEUKIN-7 DM07444 P26895 1-175: M1-Q32	BLAST_DOMO
19	7500179CD1	100	S21 S73 S87 T90	N48	signal_cleavage: M1-T18	SPSCAN
			T95		Signal Peptide: M1-S16, M1-T18, M1-E20, M1-S23, M1-E25	HMMER
					Cysteine proteases inhibitors signature: N45-R97	PROFILESCAN
					Cystatin-like domain: T18-M99	HMMER_SMART
					KININOGEN DM02403 P01042 17-127: L17-T90	BLAST_DOMO
					KININOGEN DM02403 P01043 17-127: L17-T90	BLAST_DOMO
					KININOGEN DM02403 P01044 17-126: L17-T90	BLAST_DOMO
					KININOGEN DM02403 P01046 17-126: L17-T90	BLAST_DOMO
20	7509149CD1	37	S5 S25		signal_cleavage: M1-C27	SPSCAN
					Signal Peptide: M1-S25	HMMER
					pfkB family of carbohydrate kinases signatures: M1-G31	PROFILESCAN
					Dihydrodipicolinate synthetase signatures: M1-R35	PROFILESCAN
					N-6 Adenine-specific DNA methylases signature: S5-E34	PROFILESCAN
					Interleukin-7 and -9 proteins BL00255A: M1-W37	BLIMPS_BLOCKS
					Interleukin-7 signature PR00435: F2-S25	BLIMPS_PRINTS
					INTERLEUKIN7 PRECURSOR IL7 CYTOKINE GROWTH FACTOR GLYCOPROTEIN SIGNAL 3DSTRUCTURE PD013168: M1-Q32	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21	7509399CD1	171	S156 S161 T26		INTERLEUKIN-7 DM07444 P26895 1-175: M1-Q32 signal_cleavage: M1-A30	BLAST_DOMO SPSCAN
					Signal Peptide: M13-A30, M10-P32, M1-A30, M1-T26, M10-A30, M10-E29	HMMER
					Interleukin-6/G-CSF/MGF family: V51-R166	HMMER_PFAM
					Interleukin-6 homologues: V51-R166	HMMER_SMART
					Interleukin-6 / G-CSF / MGF proteins BL00254: L33-K46, C47-G91, A136-R166	BLIMPS_BLOCKS
					Interleukin-6 / G-CSF / MGF family signature: V27-D106	PROFILESAN
					Interleukin-6/G-CSF/MGF family signature PR00433: Q50-L65, L61-L86, V150-R166	BLIMPS_PRINTS
					Interleukin-6 signature PR00434: L149-R166	BLIMPS_PRINTS
					FACTOR GROWTH GLYCOPROTEIN	BLAST_PRODROM
					CYTOKINE PRECURSOR SIGNAL	
					INTERLEUKIN6 IL6 GRANULOCYTE	
					COLONYSTIMULATING PD004356: C71-R166	
					FACTOR GRANULOCYTE	BLAST_PRODROM
					COLONYSTIMULATING PRECURSOR GCSF	
					CYTOKINE GROWTH GLYCOPROTEIN SIGNAL	
					STIMULATING PD035419: M10-Q50	
					INTERLEUKIN-6 / G-CSF / MGF	BLAST_DOMO
					DM03308 A25093 31-204: T31-L65, A69-Q170	
					INTERLEUKIN-6 / G-CSF / MGF	BLAST_DOMO
					DM03308 P35833 1-173: T31-T130, G70-P171	
					INTERLEUKIN-6 / G-CSF / MGF	BLAST_DOMO
					DM03308 P09920 37-207: Q28-A169	
					INTERLEUKIN-6 / G-CSF / MGF	BLAST_DOMO
					DM03308 P13854 26-200: Q41-Q62, C71-Q170	

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			T754 T1064 T1075 T1326 T1367 T1421 T1608		CARBOXYLESTERASES TYPE-B DM00175 P01266 2195-2520: A2174-A2500, DM00175 P01267 2197-2522: S2175-A2500	BLAST_DOMO
			T1621 T1817 T2063 T2376 Y234 Y1467		CHOLINESTERASE DM00390 P01266 2522-2716: V2501-W2696, DM00390 P01267 2524-2718: V2501- W2696	BLAST_DOMO
					Leucine zipper pattern: L558-L579	MOTIFS
					Carboxylesterases type-B signature 2: E2257-P2267	MOTIFS
					Thyroglobulin type-1 repeat signature: Y48-G77, Y116-G145, Y315-G343, F616-G643, F683-G710, Y1027-G1056, Y1165-G1195	MOTIFS
23	7506910CD1	64	S26 T27	N49	signal_cleavage: M1-S20	SPSCAN
					Signal Peptide: M1-A15, M4-A21, M1-S20, M4-S25, M1-S25, M4-S20	HMMER
					Interleukin-2 IPB000779: M1-Q42	BLIMPS_BLOCKS
					Interleukin-2 signature PR00265: Y2-T30	BLIMPS_PRINTS
					PRECURSOR SIGNAL INTERLEUKIN2 IL2	BLAST_PRODROM
					TCELL GROWTH FACTOR TCGF CYTOKINE	
					GLYCOPROTEIN PD003649: L7-L60	
					INTERLEUKIN-2 DM01614 P46649 1-153: M1-L60	BLAST_DOMO
					INTERLEUKIN-2 DM01614 Q07885 1-153: M1-L60	BLAST_DOMO
					INTERLEUKIN-2 DM01614 P37997 1-148: M1-L60	BLAST_DOMO
					INTERLEUKIN-2 DM01614 P26891 1-153: M1-I48	BLAST_DOMO
24	7507000CD1	57	T9		signal_cleavage: M1-R25	SPSCAN
					Signal Peptide: M1-Q26, M1-G27, M1-V29	HMMER
					Thyroglobulin type-1 repeat IPB000716: A44-P51	BLIMPS_BLOCKS
					Interleukin-8 signature PR00436: R4-G27	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Tumour necrosis factor alpha (cachectin) signature PR01235: E2-L17	BLIMPS_PRINTS
25	7510882CD1	287	S38 S71 S162 S198 S213 T9 T161 T231	N265 N281	signal_cleavage: M1-R25 Signal Peptide: M1-Q26, M1-A29 Laminin G domain: V75-E205 Laminin G domain: T67-K202 LAMININ CHAIN EGF-LIKE DOMAIN P. PD00320: G138-L151 Laminin G domain proteins. PF00054: P185-G192 SEX HORMONEBINDING GLOBULIN PRECURSOR SHBG STEROIDBINDING PROTEIN SBP TESTISSPECIFIC ANDROGENBINDING PD150870: L34-V58, P228-H287	SPSCAN HMMER HMMER_PFAM HMMER_SMART BLIMPS_PRODOM BLIMPS_PFAM BLAST_PRODOM
					PROTEIN PRECURSOR SIGNAL SEX HORMONEBINDING GLOBULIN SHBG STEROIDBINDING SBP TESTISSPECIFIC PD013322: Q26-E74 SEX HORMONE-BINDING GLOBULIN DM01077/P04278/36-198: T36-W199 SEX HORMONE-BINDING GLOBULIN DM01077/P15196/1-163: T36-W199 SEX HORMONE-BINDING GLOBULIN DM01077/A39030/36-218: P35-D218 SEX HORMONE-BINDING GLOBULIN DM01077/P08689/36-199: P35-W199	BLAST_PRODOM BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO
26	7506575CD1	322	S227 S285 S317 T24 T237	N64 N100 N189	signal_cleavage: M1-T18	SPSCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Signal Peptide: M1-S16, M1-T18, M1-T20, M1-Q22, M1-P25, M1-A26	HMMER
					Cystatin domain: G161-C265, G39-C143	HMMER_PFAM
					Cysteine proteases inhibitors proteins BL00287: E76-T99	BLIMPS_BLOCKS
					Fetuin family proteins BL01254: T46-R82, S138-N188	BLIMPS_BLOCKS
					Cysteine proteases inhibitors signature: N185-T237, N63-L116	PROFILESKAN
					PRECURSOR KININOGEN THIOL INHIBITOR CONTAINS: GLYCOPROTEIN PLASMA DUPLICATION VASODILATOR PROTEASE PD004764: L273-E320	BLAST_PRODOM
					INHIBITOR PROTEASE THIOL PRECURSOR SIGNAL GLYCOPROTEIN PROTEINASE CYSTATIN PLASMA KININOGEN PD001231: D165-C265	BLAST_PRODOM
					CYSTEINE PROTEASES INHIBITORS DM00182 P01042 141-255: D36-F151, C159-V252 P01043 141-255: D36-F151, C159-V252 P01042 257-371: C37-I130, V152-P267 P01043 257-371: V152-P267, C37-I130	BLAST_DOMO
					Cysteine proteases inhibitors signature: R82-S95, V204-V217	MOTIFS
27	7506576CD1	358	S21 S263 S321 S353 T60 T273	N48 N100 N136 N225	signal_cleavage: M1-T18	SPSCAN
					Signal Peptide: M1-S16, M1-T18, M1-E20, M1-S23, M1-E25	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Cystatin domain: G197-C301, G75-C179, S23-T60	HMMER_PFAM
					Cysteine proteases inhibitors proteins BL00287: E112-T135	BLIMPS_BLOCKS
					Cysteine proteases inhibitors signature: N221-T273, N99-L152	PROFILESAN
					PRECURSOR KININOGEN THIOL INHIBITOR CONTAINS: GLYCOPROTEIN PLASMA DUPLICATION VASODILATOR PROTEASE PD004764: L309-E356	BLAST_PRODOM
					INHIBITOR PROTEASE THIOL PRECURSOR SIGNAL GLYCOPROTEIN PROTEINASE CYSTATIN PLASMA KININOGEN PD001231: G75-C301	BLAST_PRODOM
					CYSTEINE PROTEASES INHIBITORS DM00182 P01042 141-255: D72-V288, C195-V288 P01043 141-255: D72-V288, C195-V288 P01042 257-371: C73-P303, C73-I166 P01043 257-371: C73-P303, C73-I166	BLAST_DOMO
					Cysteine proteases inhibitors signature: R118-S131, V240-V253	MOTIFS
28	7512018CD1	304	S21 S209 S267 S299 T219	N48 N82 N171	signal_cleavage: M1-T18	SPSCAN
					Signal Peptide: M1-S16	HMMER
					Signal Peptide: M1-T18	HMMER
					Signal Peptide: M1-E20	HMMER
					Signal Peptide: M1-S23	HMMER
					Signal Peptide: M1-E25	HMMER
					Cystatin-like domain: I140-Q248, T18-D126	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Cystatin domain: G143-C247, S23-C125	HMMER_PFAM
					Cysteine proteases inhibitor IPB000010: Q187-A200	BLIMPS_BLOCKS
					Cysteine proteases inhibitors signature: N167-T219	PROFILES CAN
					Cysteine proteases inhibitors signature: N45-L98	PROFILES CAN
					PRECURSOR KININOGEN THIOL INHIBITOR CONTAINS: GLYCOPROTEIN PLASMA DUPLICATION VASODILATOR PROTEASE PD004764: L255-E302	BLAST_PROD OM
					INHIBITOR PROTEASE THIOL PRECURSOR SIGNAL GLYCOPROTEIN PROTEINASE CYSTATIN PLASMA KININOGEN PD001231: D147-C247	BLAST_PROD OM
					CYSTEINE PROTEASES INHIBITORS DM00182[P01042]257-371: V134-P249	BLAST_DOM O
					CYSTEINE PROTEASES INHIBITORS DM00182[P01043]257-371: V134-P249	BLAST_DOM O
					CYSTEINE PROTEASES INHIBITORS DM00182[S14447]13-108: P153-P249	BLAST_DOM O
					CYSTEINE PROTEASES INHIBITORS DM00182[P01044]256-370: V134-P249	BLAST_DOM O
					Cysteine proteases inhibitors signature: V186-V199	MOTIFS

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
29/7486368CB1/398	1-216, 1-398, 118-271, 190-398
30/7488284CB1/ 1680	1-553, 1-603, 16-242, 16-356, 16-386, 16-437, 16-488, 16-495, 16-508, 16-523, 16-528, 16-541, 16-555, 16-577, 16-579, 16-601, 28-571, 94-1497, 118-493, 123-611, 128-807, 139-677,
	157-620, 157-642, 157-681, 188-739, 221-709, 264-724, 294-766, 299-456, 301-861, 343-807, 345-865, 421-982, 449-919, 457-936, 460-686, 467-982, 485-861, 521-864, 566-1148, 571-861, 575-867, 576-766, 576-1065,
	621-875, 626-1279, 664-1036, 666-1079, 667-1156, 669-1229, 703-1103, 723-1196, 794-1311, 798-1311, 806-1188, 865-1308, 879-1311, 898-1311, 926-1147, 928-1311, 998-1679, 998-1680
31/7500756CB1/789	1-230, 1-789, 27-303, 34-246, 34-303, 36-268, 37-165, 45-303, 49-152, 52-225, 52-262, 52-275, 52-285, 52-299, 52-300, 52-303, 52-481, 52-530, 57-303, 58-287, 59-300, 59-301, 61-257, 67-163, 67-303, 69-279,
	182-303, 230-287, 297-664, 297-768, 297-769, 297-789, 299-789, 309-781, 309-787, 318-779, 345-779, 346-630, 346-637, 346-757, 346-789, 380-781, 414-784, 414-789, 455-789, 486-781, 536-767, 536-779, 571-779
32/7506418CB1/ 1040	1-274, 1-1040, 187-540, 187-668, 187-889, 188-464, 297-531, 322-548, 345-799, 376-1040, 385-824, 402-653, 411-1022, 434-874, 454-735, 532-1017, 537-781, 539-958, 561-904, 576-844, 578-809, 610-1006, 645-804, 680-842, 706-889, 888-1022
33/7506942CB1/ 1087	1-235, 1-280, 1-316, 1-357, 1-1086, 46-203, 123-523, 366-748, 378-822, 387-1087, 392-1004, 432-1087, 443-1087, 483-838, 486-1087, 490-1087, 535-1087, 563-675, 563-1040, 585-1087,
	592-1087, 655-1087, 666-1087, 675-1063, 694-1087, 699-1087, 702-1087, 812-1087, 817-1086
34/7506998CB1/ 1270	1-243, 1-330, 1-460, 1-471, 1-532, 1-544, 1-548, 1-570, 1-597, 1-606, 1-620, 1-625, 1-1243, 45-532, 45-616, 45-625, 51-714, 69-478, 317-541, 335-514, 352-571, 356-503, 361-503, 410-1242, 629-1242,
	632-1243, 636-1263, 637-896, 642-1243, 649-1243, 664-1244, 667-990, 688-796, 737-1263, 759-1238, 871-1014, 876-1056, 909-1270, 966-1242, 967-1233, 967-1243
35/7506999CB1/ 1223	1-243, 1-330, 1-1223, 51-1003, 69-424, 323-1207, 425-1206, 425-1207, 428-1206, 430-1207, 430-1207, 432-1206, 432-1223, 441-698, 442-696, 450-1206, 451-1207, 452-1207, 465-642, 474-1207, 486-1223,
	493-1119, 503-1207, 526-801, 528-1159, 592-1207, 596-1206, 597-856, 602-1207, 609-1220, 624-1208, 627-950, 697-1207, 719-1198, 836-1016, 927-1193, 927-1205

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
36/7509150CB1/928	1-142, 1-167, 1-216, 1-218, 1-292, 1-928, 2-214, 2-219, 7-184, 17-207, 17-485, 73-218, 309-540, 319-588, 330-542, 330-819, 332-604, 340-615, 340-798, 343-799, 349-529, 363-527, 372-618, 388-589, 442-610,
	535-631, 558-816, 563-804, 574-712, 590-870, 594-875, 602-838, 602-884, 615-761, 615-835, 615-879, 615-889, 615-898, 629-841, 647-868, 664-850, 681-902
37/1576743CB1/ 2497	1-718, 201-898, 703-768, 807-1050, 807-1415, 909-1534, 918-983, 1025-1301, 1053-1284, 1082-1379, 1086-1298, 1101-1446, 1255-1438, 1255-1510, 1255-1738, 1286-2013, 1313-1603, 1313-1878, 1316-1592,
	1332-2014, 1359-1596, 1472-1989, 1546-2014, 1555-1916, 1574-2031, 1577-2024, 1674-2029, 1683-1906, 1693-2024, 1731-2031, 1745-1983, 1759-2497, 1767-2032, 1769-2029
38/7503892CB1/ 2573	1-2557, 131-371, 131-610, 131-653, 131-774, 134-749, 142-749, 272-818, 377-601, 377-905, 377-957, 379-1081, 384-694, 386-957, 386-1037, 386-1093, 386-1118, 386-1130, 386-1137, 386-1183, 388-825, 390-1043, 397-1006, 421-1100, 426-1223, 478-1030, 519-1121, 551-799, 554-1226, 576-872, 593-1180, 603-1186, 608-1153, 623-1269, 667-1330, 668-974, 669-940, 679-1222, 703-1327, 707-1292, 714-1031, 714-1270, 748-1354, 768-1343, 785-1317, 788-1359, 844-870, 845-870, 858-1533, 944-1271, 1116-1310, 1351-1509, 1370-1389, 1519-1924, 1593-1619, 1593-1646, 1594-1613, 1594-1725, 1594-1837, 1596-2059, 1603-2080, 1605-2113, 1607-2202, 1613-2218, 1624-2032, 1624-2114, 1632-1958, 1650-2217, 1653-2217, 1667-2241, 1681-2032, 1701-1996, 1709-2509, 1738-2302, 1741-2244, 1757-1999, 1769-2250, 1805-2509, 1818-2400, 1830-2571, 1873-2509, 1877-2077, 1909-2201, 1983-2573, 2025-2185, 2040-2573, 2047-2557, 2049-2413, 2078-2573, 2079-2553, 2097-2557, 2112-2557, 2115-2556, 2122-2529, 2123-2573, 2124-2517, 2132-2397, 2144-2557, 2176-2559, 2177-2573, 2178-2556, 2181-2573, 2210-2570, 2212-2558, 2276-2554, 2308-2559, 2314-2557, 2321-2556, 2336-2557, 2397-2509, 2435-2552, 2435-2553,
	2438-2556, 2455-2544, 2456-2544, 2459-2544, 2464-2544
39/7503895CB1/ 2440	1-2424, 131-371, 131-610, 131-653, 131-774, 134-749, 142-749, 272-818, 377-601, 377-905, 377-957, 379-1081, 384-694, 386-957, 386-1037, 386-1093, 386-1118, 386-1130, 386-1137, 386-1183, 388-825, 390-1043,
	397-1006, 421-1100, 478-1030, 519-1121, 551-799, 576-872, 593-1180, 603-1186, 608-1153, 668-974, 669-940, 714-1031, 839-870, 864-1297, 1200-1220, 1417-1995, 1456-1487, 1456-1514, 1461-1481, 1477-1750, 1686-2268, 1698-2440, 1702-1988
40/7509192CB1/858	1-220, 21-844, 21-849, 28-144, 78-226, 93-427, 306-851, 311-855, 313-851, 336-835, 340-694, 341-849, 353-618, 354-843, 363-844, 370-847, 370-854, 374-696, 375-851, 377-831, 392-677,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
41/7509320CB1/ 1172	413-643, 436-694, 438-823, 443-858, 469-833, 511-836, 577-836, 591-847, 718-813, 734-815 1-307, 1-308, 1-533, 6-1172, 90-807, 90-876, 90-1025, 106-349, 106-567, 107-516, 182-466, 182-467, 338-1025, 569-968, 569-1025, 571-893, 614-684, 725-1172, 876-1172
42/7503898CB1/ 2769	1-2694, 24-2694, 131-371, 131-610, 131-653, 131-774, 134-749, 142-749, 272-818, 377-601, 377-905, 377-957, 379- 1081, 384-694, 386-957, 386-1037, 386-1093, 386-1118, 386-1130, 386-1137, 386-1183, 388-825, 390-1043, 397-1006, 421-1100, 426-1223, 478-1030, 519-1121, 551-799, 554-1226, 576-872, 593-1180, 603-1186, 608-1153, 623-1269, 667-1330, 668-974, 669-940, 679-1222, 703-1327, 707-1292, 714-1031, 714-1270, 748-1354, 768-1343, 785-1317,
	788-1359, 839-870, 845-870, 858-1533, 944-1271, 1116-1310, 1289-1802, 1351-1509, 1360-1389, 1369-1389, 1380-1648, 1433-2008, 1438-1718, 1444-1957, 1454-1726, 1475-1917, 1475-1921, 1485-1993, 1495-1799, 1562- 1789, 1572-2220,
	1573-2084, 1586-2164, 1601-2041, 1607-2222, 1619-1728, 1623-1874, 1625-1656, 1625-1683, 1630-1649, 1630- 1650, 1630-2016, 1633-2096, 1640-2117, 1646-1919, 1669-1995, 1738-2033, 1794-2036, 1807-2007, 1833-2215, 1871-2157, 1941-2192, 2059-2348, 2062-2230, 2199-2690, 2329-2564, 2355-2769, 2530-2642
43/7509185CB1/ 1289	1-307, 1-308, 1-533, 6-1289, 90-664, 90-756, 90-839, 90-845, 90-900, 90-917, 90-924, 90-925, 90-1037, 90-1044, 90-1052, 107-516, 111-1142, 188-1141, 196-1142, 239-867, 242-867, 243-1142, 251-786, 267-645, 267-646, 273-1142, 307-748, 321-1141, 330-1142, 334-1088, 337-866, 347-1141, 425-987, 587-1085, 635-1142, 652-1207, 842-1289, 993-1289
44/7509401CB1/ 1219	1-235, 1-280, 1-316, 1-357, 1-1218, 46-203, 123-655, 498-880, 510-954, 519-1219, 524-1136, 564-1219, 575-1219, 615-970, 618-1219, 622-1219, 667-1219, 695-807, 695-1104, 717-1219, 774-1219,
45/7510191CB1/ 1058	787-1219, 798-1219, 807-1195, 826-1219, 831-1219, 834-1219, 944-1219, 949-1218 1-256, 1-287, 1-314, 1-326, 1-352, 1-403, 1-428, 1-429, 1-445, 1-446, 1-457, 1-464, 1-475, 1-484, 1-491, 1-520, 1- 530, 1-532, 1-561, 1-562, 1-662, 11-1058, 29-262, 40-293, 40-308, 43-293, 44-294, 44-300, 44-317, 44-322, 44-328, 44-334, 44-336, 44-362,
	46-298, 49-326, 50-359, 55-295, 55-378, 56-330, 56-600, 57-335, 69-306, 71-368, 75-525, 79-320, 79-341, 84-368, 84-372, 99-372, 118-372, 125-299, 127-584, 156-586, 217-705, 232-908, 239-429, 357-587, 371-999, 390-648, 428- 693,

Table 4

Polynucleotide SEQ ID NO./ Incye ID/ Sequence Length	Sequence Fragments
	461-1041, 466-1012, 473-726, 473-729, 473-736, 509-944, 536-777, 542-801, 546-842, 551-1058, 553-1015, 573-878, 600-1030, 605-1052, 609-907, 646-902, 646-908, 646-1058, 649-1058, 674-1058, 680-1058, 691-1048, 696-1047,
	707-944, 719-964, 759-1047, 790-1048, 808-915, 823-1022, 881-1047, 972-1058
46/7506943CB1/ 1217	1-217, 1-220, 1-1216, 46-203, 126-653, 423-631, 496-878, 508-952, 517-1217, 522-1134, 562-1217, 573-1217, 613-968, 616-1217, 620-1217, 665-1217, 693-805, 693-1170, 715-1217, 722-1217,
	785-1217, 796-1217, 805-1193, 824-1217, 829-1217, 832-1217, 942-1217, 947-1216
47/7500179CB1/785	1-640, 202-439, 202-785, 204-533, 205-487, 208-560, 224-477, 231-465, 231-583, 231-591, 231-640, 233-480, 234-434, 234-596, 235-641, 237-506, 237-596, 238-478, 238-641, 240-628, 242-584, 243-628, 249-636, 249-637, 257-637,
	258-639, 277-529, 277-582, 278-430, 279-448, 279-512, 279-551, 282-489, 282-512, 283-620, 290-426, 290-563, 297-541, 298-643, 306-580, 315-588, 315-659, 315-767, 315-768, 350-574, 352-639, 355-628, 357-640, 364-612, 569-667
48/7509149CB1/ 1084	1-217, 46-203, 123-520, 299-498, 363-745, 375-819, 389-1001, 480-835, 483-1084, 487-1084, 497-1084, 509-1084, 532-1084, 540-1084, 560-613, 560-969, 582-1084, 639-1084, 652-1084, 663-1084, 672-1060, 691-1084, 696-1084, 699-1084, 809-1084, 814-1083
49/7509399CB1/933	1-142, 1-167, 1-216, 1-218, 1-301, 1-725, 1-847, 1-933, 2-214, 2-226, 7-184, 17-207, 17-491, 17-637, 73-218, 226-518, 227-406, 227-477, 231-804, 233-407, 233-490, 233-858, 234-445, 255-560, 257-406, 257-407, 258-335, 301-805, 308-546,
	323-594, 336-548, 336-825, 338-610, 346-621, 346-804, 355-535, 369-533, 378-624, 394-595, 448-616, 541-637, 564-822, 569-810, 580-718, 596-876, 600-881, 608-844, 608-890, 621-767, 621-841, 621-885, 621-895, 621-904, 635-847, 653-874, 670-856, 687-908

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
50/7504789CB1/ 8378	1-214, 1-246, 1-394, 1-504, 11-293, 11-297, 11-437, 11-458, 11-465, 11-590, 11-643, 11-8378, 83-568, 83-727, 84-729, 102-643, 104-740, 115-631, 115-730, 120-434, 140-731, 145-628, 149-705, 152-653, 183-542, 196-579, 228-762, 231-485, 233-422, 235-400, 271-819, 283-937, 292-828, 292-845, 307-818, 319-861, 322-629, 348-794, 348-810, 348-826, 348-848, 348-862, 348-864, 348-872, 348-884, 348-895, 348-899, 348-902, 348-914, 348-953, 348-962, 348-966, 348-996, 349-978, 355-853, 363-828, 371-807, 371-811, 378-806, 385-831, 387-716, 409-1050, 418-971, 418-1090, 428-995, 433-980, 439-481, 439-487, 476-1094, 480-653, 480-654, 480-743, 481-951, 516-1002, 516-1011, 516-1054, 516-1062, 516-1069, 516-1074, 522-781, 532-873, 537-1276, 556-1029, 564-1011, 564-1138, 583-1090, 594-1232, 598-824, 608-1244, 611-1072, 611-1118, 611-1168, 618-1284, 621-1273, 626-1029, 626-1117, 657-956, 657-1164, 657-1188, 657-1231, 657-1264, 657-1287, 657-1314, 662-1216, 666-1269, 687-1241, 711-1216, 720-1231, 747-1377, 765-1341, 765-1342, 770-1295, 785-1147, 785-1148, 816-1216, 819-1358, 820-1257, 837-1072, 859-1381, 869-1380, 893-1381, 896-1420, 916-1248, 991-1370, 994-1593, 1033-1081, 1033-1089,
	1058-1673, 1101-1365, 1124-1491, 1130-1778, 1218-1802, 1262-2081, 1317-1747, 1346-1967, 1440-2067, 1442-2174, 1457-2063, 1525-2108, 1582-2266, 1596-2067, 1640-2196, 1656-1894, 1661-2013, 1714-2052,
	1714-2053, 1718-2160, 1727-2233, 1744-2380, 1796-2185, 1809-2170, 1840-2426, 1946-2335, 1952-2196, 1958-2592, 1962-2107, 2026-2476, 2034-2349, 2039-2450, 2045-2554, 2194-2592, 2198-2452, 2224-2408,
	2233-2675, 2241-2904, 2254-2906, 2259-2689, 2288-2711, 2321-2824, 2390-3068, 2396-2636, 2406-2836, 2412-2944, 2430-2820, 2514-2833, 2516-2810, 2615-3016, 2618-3255, 2642-2946, 2730-3385, 2731-3129, 2731-3385, 2736-3248, 2803-2957, 2814-3085, 2816-3291, 2825-3225, 2891-3494, 2904-3652, 2938-3167, 2957-3221, 2957-3222, 3036-3154, 3098-3646, 3099-3646, 3109-3435, 3248-3522, 3381-3890, 3417-3662, 3417-3737, 3417-3739, 3417-4008, 3417-4010, 3418-3706, 3420-3913, 3431-3704, 3431-3892, 3432-3706, 3457-3905, 3486-3677, 3495-3753, 3509-3962, 3512-3739, 3515-3895, 3518-4049, 3526-4195, 3598-4176, 3805-4312, 3821-4064, 3833-4349, 3835-4009, 3922-4283, 3926-4281, 3930-4490, 3933-4283, 3953-4283, 3969-4211, 4004-4271, 4006-4257, 4015-4270, 4020-4703, 4031-4286, 4035-4296, 4042-4342, 4042-4343, 4053-4704, 4070-4706, 4086-4489, 4086-4522, 4089-4238, 4123-4394, 4125-4360, 4125-4617, 4139-4701, 4165-4823, 4177-4442, 4277-4940, 4306-4872, 4317-4794, 4329-4947, 4333-4663, 4337-4576, 4337-4688, 4339-4832, 4348-4883, 4409-4716, 4412-4668, 4431-4718, 4439-5005, 4479-4762, 4487-4614, 4489-5135, 4519-5046, 4538-4825, 4551-5157, 4553-4823, 4554-5011,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	4582-4828, 4594-4844, 4599-4849, 4601-4897, 4615-5210, 4620-5151, 4625-5148, 4639-5218, 4641-5155, 4657-4843, 4671-4923, 4673-5144, 4698-4990, 4708-5184, 4739-5073, 4739-5145, 4739-5151, 4739-5159, 4739-5167, 4739-5171, 4739-5173, 4739-5176, 4739-5187, 4740-5283, 4742-5159, 4743-5012, 4743-5090, 4743-5120, 4743-5143, 4743-5144, 4743-5146, 4743-5155, 4743-5156, 4743-5177, 4744-5222, 4744-5246, 4764-4893, 4835-5142, 4835-5412, 4873-5253, 4876-5428, 4879-5422, 4883-5141, 4883-5422, 4892-5422, 4898-5148, 4927-5422, 4932-5422, 4959-5213, 4972-5396, 5000-5430, 5006-5430, 5007-5422, 5012-5428, 5013-5430, 5030-5292, 5042-5423, 5045-5406, 5046-5431, 5050-5430, 5051-5407, 5054-5430, 5055-5430, 5064-5430, 5085-5427, 5094-5342, 5109-5358, 5120-5356, 5160-5413, 5180-5416, 5180-5437, 5189-5428, 5278-5422, 5280-5674, 5338-5594, 5429-5613, 5429-5660, 5429-5672, 5429-5673, 5434-5795, 5446-5663, 5460-5764, 5474-5735, 5479-6161, 5486-5736, 5499-5729, 5500-5766, 5508-5955, 5513-5816, 5524-5760, 5533-6024, 5548-6229, 5549-5885, 5552-5837, 5556-5821, 5559-6111, 5571-6148, 5586-5876, 5601-6007, 5610-5819, 5612-5819, 5619-6010, 5634-5891, 5636-5972, 5637-5877, 5638-5865, 5641-5887, 5641-5931, 5641-6210, 5645-5970, 5648-5900, 5658-5892, 5658-5913, 5662-5914, 5662-6280, 5677-5928, 5678-5906, 5681-6344, 5686-6333, 5693-6244, 5702-5973, 5703-5952, 5705-6263, 5721-5984, 5723-5955, 5736-6364, 5750-6031, 5783-6444, 5790-6252, 5798-6294, 5815-6102, 5834-6014, 5846-6112, 5861-6408, 5867-6324, 5872-6095, 5872-6293, 5898-6074, 5899-6342, 5902-6150, 5902-6171, 5902-6439, 5907-6146, 5907-6518, 5908-6157, 5917-6133, 5917-6421, 5936-6183, 5950-6207, 5953-6200, 5953-6223, 5957-6216, 5958-6236, 5960-6201, 5960-6217, 5965-6155, 5965-6215, 5965-6223, 5971-6613, 5974-6142, 5992-6549, 6002-6188, 6009-6263, 6016-6245, 6016-6265, 6019-6254, 6025-6314, 6031-6352, 6034-6265, 6034-6271, 6045-6271, 6070-6331, 6074-6640, 6096-6362, 6112-6351, 6113-6365, 6118-6375, 6123-6719, 6127-6620, 6146-6695, 6152-6803, 6157-6381, 6162-6410, 6184-6454, 6184-6491, 6187-6461, 6199-6307, 6210-6491, 6215-6490, 6215-6491, 6216-6491, 6225-6450, 6230-6499, 6246-6509, 6249-6491, 6250-6503, 6255-6491, 6256-6499, 6256-6684, 6258-6707, 6259-6564, 6261-6604, 6262-6448, 6262-6491, 6262-6496, 6262-6499, 6263-6491, 6264-6442,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	6268-6513, 6270-6684, 6271-6491, 6271-6515, 6273-6801, 6274-6433, 6278-6962, 6279-6491, 6283-6682, 6283-6705, 6285-6630, 6288-6485, 6288-6542, 6288-6571, 6288-6898, 6295-6472, 6295-6547, 6297-6704, 6297-6750, 6301-6528, 6302-6548, 6302-6665, 6307-6558, 6313-6549, 6313-7009, 6314-6571, 6314-6581, 6316-6948, 6321-6579, 6323-6588, 6341-6584, 6344-6604, 6346-6588, 6350-6592, 6356-6598, 6356-6617, 6356-6934, 6362-6620, 6365-6620, 6367-6649, 6370-6643, 6373-6958, 6377-6838, 6378-6885, 6392-6742, 6396-6644, 6396-6663, 6396-7045, 6402-6571, 6403-6594, 6403-6660, 6405-6796, 6406-6741, 6409-6770, 6409-7109, 6411-6796, 6413-6838, 6413-6917, 6413-6933, 6428-6826, 6428-7027, 6433-6677, 6434-6724, 6438-6559, 6438-7027, 6440-6680, 6440-6691, 6442-6667, 6442-6685, 6444-6723, 6445-6826, 6447-6761, 6447-6903, 6453-6698, 6470-7161, 6471-6719, 6471-6745, 6476-6759, 6476-6769, 6479-6737, 6498-6990, 6499-6868, 6500-6838, 6503-6767, 6504-7136, 6511-7151, 6513-6760, 6515-6755, 6515-6760, 6520-6916, 6523-6913, 6540-6903, 6541-7212, 6550-6802, 6553-6660, 6557-6801, 6568-6825, 6573-7256, 6575-6850, 6575-6854, 6576-7205, 6579-7086, 6579-7154, 6583-6793, 6591-7026, 6597-7205, 6597-7252, 6599-7140, 6600-7221, 6601-6853, 6602-7197, 6617-6856, 6626-6886, 6626-7037, 6634-7208, 6635-7207, 6636-6887, 6636-6888, 6642-6857, 6642-6878, 6646-6880, 6649-6899, 6652-6879, 6652-6947, 6662-7227, 6663-6895, 6664-6935, 6666-6842, 6666-6930, 6667-6842, 6669-6893, 6672-7002, 6675-6911, 6682-6907, 6683-6929, 6697-6848, 6698-6947, 6700-6907, 6700-6925, 6701-6955, 6701-7210, 6704-6935, 6705-6990, 6713-6987, 6718-7189, 6725-6969, 6730-6936, 6733-6977, 6739-7424, 6743-6917, 6743-7167, 6745-6996, 6751-7235, 6752-6999, 6763-7317, 6770-7301, 6774-7052, 6776-7044, 6790-7027, 6796-7029, 6800-7057, 6803-7049, 6804-7077, 6809-7046, 6809-7052, 6811-7047, 6811-7058, 6814-7061, 6814-7076, 6814-7077, 6815-7455, 6816-7064, 6817-7359, 6818-7385, 6827-7065, 6827-7067, 6827-7090, 6827-7096, 6829-7071, 6829-7090, 6829-7092, 6831-7074, 6831-7082, 6831-7099, 6837-7143, 6846-7070, 6846-7081, 6846-7342, 6848-6994, 6849-7469, 6851-7110, 6856-7116, 6859-7109, 6859-7114, 6864-7130, 6868-7128, 6869-7434, 6870-7110, 6879-7124, 6880-7141, 6890-7130, 6890-7552, 6893-7124, 6893-7126, 6893-7153, 6897-7552, 6900-7131,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	6900-7157, 6900-7365, 6901-7130, 6902-7150, 6906-7401, 6909-7301, 6919-7552, 6927-7173, 6928-7425, 6929-7155, 6929-7367, 6935-7248, 6937-7189, 6937-7210, 6941-7196, 6958-7453, 6966-7245, 6987-7422, 7003-7260, 7007-7258, 7014-7257, 7020-7788, 7035-7291, 7037-7666, 7042-7305, 7064-7296, 7079-7306, 7079-7328, 7083-7313, 7083-7331, 7084-7557, 7085-7555, 7085-7782, 7086-7306, 7086-7322, 7087-7299, 7090-7298, 7105-7381, 7109-7377, 7110-7368, 7111-7358, 7111-7957, 7116-7329, 7116-7371, 7121-7393, 7134-7387, 7134-7415, 7135-7762, 7135-7824, 7144-7307, 7144-7399, 7144-7403, 7144-7824, 7152-7413, 7160-7403, 7160-7415, 7171-7391, 7202-7458, 7205-7459, 7208-7423, 7219-7318, 7226-7465, 7233-7528, 7235-7500, 7236-7750, 7241-7328, 7241-7722, 7241-7780, 7243-7794, 7245-7482, 7249-7497, 7251-7493, 7251-7499, 7258-7857, 7259-7570, 7260-7590, 7261-7522, 7265-7542, 7265-7773, 7266-7624, 7272-7801, 7273-7528, 7273-7533, 7273-7568, 7274-7545, 7276-7616, 7279-7525, 7279-7527, 7279-7530, 7279-7532, 7279-7535, 7279-7536, 7279-7559, 7280-7880, 7281-7501, 7282-7527, 7283-7516, 7283-7521, 7283-7523, 7283-7527, 7283-7537, 7283-7709, 7284-7567, 7286-7544,
	7288-7566, 7288-7745, 7294-7506, 7294-7750, 7295-7545, 7295-7630, 7299-7665, 7301-7558, 7303-7536, 7303-7579, 7304-7540, 7304-7953, 7306-7549, 7306-7698, 7306-7907, 7309-7537, 7309-7540, 7309-7577, 7309-7581, 7312-7832, 7314-7579, 7315-7585, 7316-7543, 7316-7557, 7316-7570, 7318-7562, 7319-7558, 7320-7631, 7320-8021, 7325-7576, 7327-7631, 7329-7578, 7329-7611, 7332-7587, 7340-7916, 7342-7577, 7344-7504, 7345-7625, 7350-8003, 7354-7598, 7354-7601, 7356-7588, 7357-7545, 7358-7596, 7359-7608, 7359-7641, 7361-7585, 7361-7637, 7361-7747, 7362-7652, 7362-7794, 7365-7609, 7366-7622, 7366-7632, 7368-7853, 7369-7965, 7370-7653, 7371-7532, 7372-7613, 7373-7715, 7373-7814, 7378-7801, 7379-7655, 7379-7660, 7382-8021, 7383-7649, 7383-7771, 7383-7814, 7390-7619, 7390-7739, 7390-7747, 7393-7823, 7395-7682, 7398-7922, 7400-7753, 7401-7794, 7401-8013, 7403-7533, 7404-7636, 7405-7647, 7405-7671, 7405-7672, 7407-7713, 7409-7633, 7414-7652, 7415-7664, 7416-7750, 7416-7814, 7416-7913, 7419-7665, 7419-7675, 7428-7683, 7431-8248, 7433-7664, 7435-7676, 7436-7681, 7436-7708, 7437-7717, 7438-7670, 7438-7695, 7438-7728, 7438-7733, 7438-7965, 7439-7678,

Table 4

Polynucleotide SEQ ID NO: / Incyte ID/ Sequence Length	Sequence Fragments
	7439-7697, 7439-8012, 7445-7624, 7445-7805, 7449-7801, 7454-7701, 7454-7704, 7454-7706, 7455-7699, 7457-7717, 7464-7691, 7465-7725, 7469-7742, 7470-7677, 7470-7716, 7470-7717, 7470-7718, 7470-7720, 7470-7727, 7470-7751, 7470-7770, 7471-7707, 7471-7729, 7472-7687, 7475-7733, 7475-7736, 7475-7737, 7479-7709, 7483-7757, 7485-7953, 7486-7731, 7488-7574, 7490-7742, 7490-7767, 7491-7868, 7493-7727, 7493-7734, 7493-7738, 7493-7769, 7497-7521, 7497-7739, 7497-7775, 7497-7922, 7497-7953, 7497-7957, 7498-7654, 7498-7922, 7499-7650, 7499-7935, 7499-7950, 7499-7953, 7499-7966, 7499-7983, 7500-7963, 7503-7684, 7503-7830, 7504-7748, 7505-7776, 7507-7741, 7507-8000, 7509-7704, 7514-7942, 7517-7953, 7517-8035, 7522-7965, 7523-7766, 7524-7776, 7524-7779, 7525-7943, 7526-7876, 7527-7817, 7527-7943, 7527-8042, 7530-7943, 7532-7669, 7532-7747, 7532-7775, 7532-7778, 7532-7781, 7532-7801, 7532-7807, 7532-7943, 7532-7953, 7533-7772, 7533-7783, 7533-7789, 7533-7791, 7533-7792, 7534-7943, 7534-8030, 7536-7943, 7537-7942, 7542-7767, 7542-7769, 7543-7909, 7543-7943, 7547-8129, 7549-7769, 7549-7811, 7550-7943, 7552-7696, 7552-7793, 7552-7804, 7552-7813, 7552-7820, 7552-8132, 7554-7833, 7554-7965, 7555-8028, 7556-8028, 7562-7756, 7563-7810, 7563-7831, 7564-7756, 7564-7852, 7565-7839, 7571-7922, 7574-8016, 7582-7832, 7584-7829, 7584-7840, 7584-7849, 7584-7864, 7590-7788, 7590-7808, 7590-7865, 7592-7712, 7597-7753, 7598-7836, 7598-7840, 7602-7906, 7602-7920, 7602-7927, 7602-7934, 7604-7953, 7605-7953, 7606-7848, 7607-7859, 7607-7860, 7608-7841, 7608-7865, 7608-7868, 7608-7908, 7614-7857, 7614-7873, 7614-7896, 7616-7846, 7616-7862, 7617-7837, 7623-7871, 7627-7912, 7628-7929, 7629-7919, 7630-7894, 7634-7917, 7634-8067, 7639-7891, 7655-7905, 7655-7906, 7655-7915, 7655-8062, 7657-7915, 7657-7945, 7659-7918, 7660-7880, 7660-7924, 7661-7918, 7661-7951, 7661-7965, 7662-7948, 7664-8016, 7666-7794, 7667-7896, 7667-7921, 7667-7982, 7668-7817, 7668-7894, 7668-7932, 7668-7953, 7668-7993, 7670-7951, 7671-7905, 7672-7948, 7678-7894, 7680-7937, 7680-7943, 7681-7937, 7681-7973, 7683-7966, 7688-7918, 7690-7716, 7691-7934, 7691-7936, 7691-7940, 7691-7943, 7691-7945, 7692-7960, 7694-7964, 7694-8013, 7699-7901, 7699-7941, 7701-7913, 7701-7943, 7701-7945, 7701-7949, 7701-7954, 7701-7956, 7707-7994, 7708-7979, 7712-7918, 7714-7988, 7721-7993, 7721-7995, 7722-7963, 7722-7964, 7722-7969, 7726-8042, 7729-7984, 7731-8005, 7732-8030, 7736-7943, 7736-8019, 7741-7981, 7741-7982, 7741-7994, 7741-8032, 7742-8009,
	7743-7970, 7743-7973, 7743-7974, 7743-7981, 7744-7989, 7746-8072, 7747-7857, 7747-8008, 7747-8019, 7749-7891, 7749-8029, 7749-8035, 7749-8055, 7750-8004, 7751-8080, 7752-8003, 7754-7936, 7754-7940, 7754-7943,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	7757-7941, 7757-7974, 7757-7988, 7757-8027, 7758-7922, 7758-8087, 7759-7935, 7759-8004, 7759-8016, 7761-8085, 7762-8009, 7763-7998, 7765-7946, 7778-8012, 7780-8016, 7780-8020, 7782-7943, 7797-8036,
	7803-8060, 7808-7943, 7886-8028, 7927-7993, 7929-8327, 7941-8139, 8013-8150
51/7506910CB1/864	1-97, 1-600, 1-603, 1-837, 265-603, 339-859, 416-864, 487-837, 710-864
52/7507000CB1/ 1259	1-1259, 51-725, 51-785, 51-797, 51-855, 51-869, 51-871, 51-885, 51-887, 51-926, 51-941, 51-944, 51-949, 51-954, 51-973, 354-1126, 412-1259, 474-1259, 478-1259, 480-1259, 493-635, 507-1259, 530-1259, 541-1259, 542-1259, 545-1259, 552-1259, 558-1259, 574-827, 574-1259, 578-1259, 585-1259, 590-1256, 628-1259, 648-1259, 674-1259, 675-1259, 676-1259, 698-1259, 710-1259, 717-1259, 727-1259, 750-1025, 752-1259, 816-1259,
	819-939, 820-1259, 821-1080, 826-1259, 833-1259, 848-1259, 851-1174, 872-980, 921-1259, 943-1259, 1055-1198, 1060-1240, 1093-1259, 1113-1259, 1150-1259, 1151-1259
53/7510882CB1/965	1-330, 1-460, 1-471, 1-532, 1-544, 1-548, 1-570, 1-597, 1-606, 1-617, 1-621, 1-739, 1-769, 1-800, 1-821, 1-950, 126-731, 209-777, 245-827, 267-667, 279-965, 295-687, 350-965, 392-960, 399-919, 403-965, 459-962, 462-962, 472-817, 616-965
54/7506575CB1/ 1351	1-121, 1-239, 1-241, 4-1341, 34-234, 78-230, 79-248, 90-226, 135-904, 254-969, 336-725, 347-644, 350-949, 360-1221, 367-668, 401-471, 404-758, 405-1058, 408-1323, 415-646, 415-1026, 432-1298, 447-1298, 452-1324, 462-1302,
	475-758, 522-1298, 537-1012, 539-789, 542-1298, 544-1324, 545-1298, 545-1324, 549-1298, 550-1328, 552-1298, 554-1298, 560-843, 579-1298, 588-1298, 590-1298, 593-1110, 628-957, 637-1324, 655-1258, 687-1324, 706-1298,
	707-1136, 755-1341, 758-992, 771-1021, 782-1318, 799-939, 859-1326, 865-1328, 882-1326, 897-1326, 925-1211, 942-1328, 966-1326, 966-1332, 968-1351, 985-1268, 1001-1324, 1053-1275, 1060-1327,
	1084-1351, 1086-1327, 1094-1325, 1102-1223, 1102-1326, 1106-1326, 1111-1328, 1131-1351
55/7506576CB1/ 1459	1-239, 4-333, 5-287, 8-360, 24-277, 27-258, 31-265, 33-280, 34-234, 37-306, 38-278, 77-329, 78-230, 79-248, 79-351, 90-226, 90-363, 106-376, 135-867, 135-869, 150-364, 444-833, 455-752, 458-1057, 468-1329, 512-866, 513-1166,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
	516-1431, 523-1134, 540-1406, 555-1406, 560-1432, 570-1410, 583-866, 630-1406, 645-1120, 647-897, 652-1432, 653-1406, 653-1432, 657-1406, 658-1436, 660-1406, 662-1406, 668-951, 687-1406, 698-1406, 701-1218, 745-1432,
	763-1366, 795-1432, 814-1406, 815-1244, 866-1100, 879-1129, 890-1426, 907-1047, 967-1434, 973-1436, 990-1434, 1005-1434, 1033-1319, 1050-1436, 1074-1434, 1076-1459, 1093-1376, 1109-1432,
	1161-1383, 1168-1435, 1192-1459, 1194-1435, 1202-1433, 1210-1331, 1210-1434, 1214-1434, 1219-1436, 1239-1459
56/7512018CB1/ 1469	1-121, 1-239, 1-241, 4-1287, 5-287, 5-333, 8-360, 24-277, 30-240, 31-265, 31-391, 32-264, 33-280, 34-234, 34-393, 37-306, 38-278, 42-384, 77-329, 77-382, 78-230, 79-248, 79-312, 79-351, 82-289, 83-393, 90-226, 90-363, 97-341, 106-380,
	115-388, 149-821, 150-374, 164-393, 386-1163, 386-1269, 391-1163, 392-1163, 395-1163, 396-1163, 397-1244, 398-1163, 398-1270, 421-704, 442-1163, 443-1163, 460-1163, 483-958, 485-735, 487-780, 490-1270, 491-1163, 491-1270,
	496-1274, 504-1163, 505-1163, 506-789, 533-1163, 539-1056, 574-903, 583-1270, 601-1204, 633-1270, 653-1082, 701-1287, 703-936, 709-853, 717-967, 731-1228, 745-885, 805-1272, 811-1274, 828-1272, 843-1272, 871-1157,
	888-1274, 912-1272, 912-1278, 914-1469, 931-1214, 947-1270, 999-1221, 1006-1273, 1013-1232, 1028-1274, 1030-1298, 1031-1276, 1032-1273, 1040-1271, 1048-1169, 1048-1272, 1052-1272, 1057-1274, 1077-1297

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
29	7486368CB1	LIVRBC01
30	7488284CB1	UCMCNOT02
31	7500756CB1	ADRETUT05
32	7506418CB1	BRAIFEN05
33	7506942CB1	TYMUNT01
34	7506998CB1	LIVRTUN04
35	7506999CB1	SINTNON02
36	7509150CB1	SMCCNOS01
37	1576743CB1	PROSTUS08
38	7503892CB1	SINTNOR01
39	7503895CB1	SINTNOR01
40	7509192CB1	LIVRDIT06
41	7509320CB1	BRAINOS12
42	7503898CB1	SINTNOR01
43	7509185CB1	KIDNFET02
44	7509401CB1	TYMUNT01
45	7510191CB1	CONTTUT01
46	7506943CB1	TYMUNT01
47	7500179CB1	MIXDTUE02
48	7509149CB1	TYMUNT01
49	7509399CB1	SMCCNOS01
50	7504789CB1	SPLNTUE01
51	7506910CB1	TYMTXT03
52	7507000CB1	SINTNON02
53	7510882CB1	LIVRTUN04
54	7506575CB1	KIDNNOT19
55	7506576CB1	KIDNNOT19
56	7512018CB1	KIDNNOT19

Table 6

Library	Vector	Library Description
ADRETUT05	pINCY	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 52-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated a pheochromocytoma.
BRAIFEN05	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.
BRAINOS12	PSPORT1	Library was constructed from 4.9 million clones from a brain tissue library, by subtraction of abundantly expressed clone pools. RNA was made from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.
CONTTUT01	pINCY	Library was constructed using RNA isolated from tumorous soft tissue of the left lateral thigh removed from a 34-year-old Caucasian female during a soft tissue excision. Pathology indicated metastatic grade 2 myxoid liposarcoma which formed multiple, lobulated, circumscribed masses situated in the subcutaneous adipose tissue. Patient history included a malignant soft tissue neoplasm of the leg. Family history included benign hypertension, acute leukemia, benign hypertension, and type II diabetes.
KIDNFET02	pINCY	Library was constructed using RNA isolated from kidney tissue removed from a Caucasian male fetus, who was stillborn with a hypoplastic left heart and died at 23 weeks' gestation.
KIDNNOT19	pINCY	Library was constructed using RNA isolated from kidney tissue of a 65-year-old Caucasian male during an exploratory laparotomy and nephroureterectomy. Pathology for the associated tumor tissue indicated a grade 1 renal cell carcinoma within the upper pole of the left kidney. Patient history included malignant melanoma of the abdominal skin, benign neoplasm of the colon, cerebrovascular disease, and umbilical hernia. Family history included myocardial infarction, atherosclerotic coronary artery disease, cerebrovascular disease, prostate cancer, myocardial infarction, and atherosclerotic coronary artery disease.
LIVRBC'T01	PBLUESCRIPT	Library was constructed using RNA isolated from the liver tissue of a patient with primary biliary cirrhosis who had a liver transplant.
LIVRDIT06	pINCY	This library was constructed using RNA isolated from diseased liver tissue removed from a 35-year-old Caucasian male during needle biopsy of the liver. Patient history included hepatitis C.

Table 6

Library	Vector	Library Description
LIVRTUN04	pINCY	This normalized liver tumor cell line library was constructed from 1.72 million independent clones from a hepatocyte library. Starting RNA was isolated from an untreated C3A hepatocyte cell line, which is a derivative of a hepatoblastoma removed from a 15-year-old Caucasian male. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, (except that a significantly longer (48 -hours/round) reannealing hybridization was used.
MIXDTUE02	PBK-CMV	This 5' biased random primed library was constructed using pooled cDNA from five donors. cDNA was generated using mRNA isolated from liver tissue removed from a Caucasian male fetus (donor A), who died from fetal demise; from liver tissue removed from a 4-year-old male (donor B), who died from anoxia; from liver tissue removed from an 8-year-old Caucasian female (donor C), who died from a closed head injury due to a motor vehicle accident; from liver tissue removed from a 62-year-old Caucasian female (donor D) during partial hepatectomy and exploratory laparotomy; and from pancreatic tumor tissue removed from a 65-year-old female (donor E). For donor A, serologies were negative. For donor B, serologies were negative except for anti-CMV. Patient history included asthma. For donor C, serologies were negative. Patient history included asthma. For donor D, pathology for the matched tumor tissue indicated metastatic intermediate grade neuroendocrine carcinoma, consistent with islet cell tumor, forming nodules ranging in size, in the lateral and medial left liver lobe.
		Patient history included malignant neoplasm of the pancreas tail, pulmonary embolism, hyperlipidemia, thrombophlebitis, joint pain in multiple joints, type II diabetes, benign hypertension, cerebrovascular disease, and normal delivery. Previous surgeries included blood vessel filter implantation, distal pancreatectomy, total splenectomy, and partial hepatectomy. Patient medications included metoprolol and trimethoprim. Family history included pancreas cancer, and secondary liver cancer in the mother; benign hypertension and hyperlipidemia in the sibling(s). For donor E, pathology indicated well-differentiated neuroendocrine carcinoma (islet cell tumor), nuclear grade 1, forming a dominant mass in the distal pancreas. Multiple smaller tumor nodules were immediately adjacent to the main mass.

Table 6

Library	Vector	Library Description
PROSTUS08	pINCY	This subtracted prostate tumor library was constructed using 2.36 million clones from a prostate tumor library and was subjected to one round of subtractive hybridization with 448,000 clones from a prostate tumor library. The starting library for subtraction was constructed using RNA isolated from a prostate tumor removed from a 59-year-old Caucasian male during a radical prostatectomy with regional lymph node excision. Pathology indicated adenocarcinoma (Gleason grade 3+3) Adenofibromatous hyperplasia was present. The patient presented with elevated prostate-specific antigen (PSA). Patient history included colon diverticuli, asbestosis, and thrombophlebitis. Family history included multiple myeloma, hyperlipidemia, and rheumatoid arthritis. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al. Genome Research (1996) 6:791.
SINTNON02	pINCY	This normalized small intestine tissue library was constructed from 1.84 million independent clones from a pooled small intestine tissue library. Starting RNA was made from pooled cDNA from six different donors. cDNA was generated using mRNA isolated from small intestine tissue removed from a Caucasian male fetus (donor A) who died from fetal demise; from small intestine tissue removed from a 8-year-old Black male (donor B) who died from anoxia; from small intestine tissue removed from a 13-year-old Caucasian male (donor C) who died from a gunshot wound to the head; from jejunum and duodenum tissue removed from the small intestine of a 16-year-old Caucasian male (donor D) who died from head trauma; from ileum tissue removed from an 8-year-old Caucasian female (donor E) who died from head trauma; and from small intestine tissue removed from a 15-year-old Caucasian female who died from a closed head injury. Serologies were negative for donors A, B, C and D. Donors E and F had serologies positive for cytomegalovirus (CMV). Donor B's medications included DDAVP, Versed, and labetalol. Donor C's previous surgeries included a hernia repair.
		The patient was not taking any medications. Family history included diabetes in the grandparent(s). Donor D's history included a kidney infection three years prior to death, marijuana use, and tobacco use. Donor E's history included migraine headaches and urinary tract infection. Previous surgeries included an adenotonsillectomy. Patient medications included Dilantin (phenytoin), Acef (cephalosporin), and Zantac (ranitidine). Donor F's history included seasonal allergies and marijuana use. Patient medications included Dopamine and Neo-Synephrine. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
SINTNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity.

Table 6

Library	Vector	Library Description
SMCCNOS01	pINCY	This subtracted coronary artery smooth muscle cell library was constructed using 7.56 X 10 ⁶ clones from a coronary artery smooth muscle cell library and was subjected to two rounds of subtraction hybridization for 48 hours with 6.12 X 10 ⁶ clones from a second coronary artery smooth muscle cell library. The starting library for subtraction was constructed using RNA isolated from coronary artery smooth muscle cells removed from a 3-year-old Caucasian male. The cells were treated with TNF alpha & IL-1 beta 10ng/ml each for 20 hours. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated coronary artery smooth muscle cells from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., (NAR (1991) 19:1954) and Bonaldo, et al. (Genome Research (1996) 6:791-806).
SPLNTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from spleen tumor tissue removed from a 28-year-old male during total splenectomy. Pathology indicated malignant lymphoma, diffuse large cell type, B-cell phenotype with abundant reactive T-cells and marked granulomatous response involving the spleen, where it formed approximately 45 nodules, liver, and multiple lymph nodes.
TLYMTXT03	pINCY	Library was constructed using RNA isolated from CD4+ T cells obtained from a pool of donors. The cells were treated with CD28 antibodies.
TLYMUNT01	pINCY	Library was constructed using RNA isolated from resting allogenic T-lymphocyte tissue removed from an adult (40-50-year old) Caucasian male.
UCMCNOT02	pINCY	Library was constructed using RNA isolated from mononuclear cells obtained from the umbilical cord blood of nine individuals.

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-6; 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. 37:417-424.	Probability value = 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
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. 1-350.	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.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, 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.
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.	

Table 7

Program	Description	Reference	Parameter Threshold
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
31	7500756	1898396H1	SNP00142701	102	158	C	C	T	noncoding	n/a	n/a	n/a	n/a
32	7506418	4010251H1	SNP00093023	41	496	T	T	C	L85	n/a	n/a	n/a	n/a
32	7506418	7335094H1	SNP00119271	48	16	G	G	A	noncoding	0.90	0.93	0.96	0.93
32	7506418	7371509H1	SNP00119272	377	165	C	C	T	noncoding	n/a	n/a	n/a	n/a
32	7506418	7371509H1	SNP00119273	398	186	A	A	G	noncoding	n/d	n/d	n/d	n/d
32	7506418	7371509H1	SNP00119274	503	291	G	G	A	K16	n/a	n/a	n/a	n/a
34	7506998	4172221H1	SNP00095724	117	1083	G	G	A	noncoding	n/a	n/a	n/a	n/a
35	7506999	4172221H1	SNP00095724	117	1043	G	G	A	noncoding	n/a	n/a	n/a	n/a
35	7506999	7340116H1	SNP00095724	552	1045	A	G	A	noncoding	n/a	n/a	n/a	n/a
36	7509150	1510009H1	SNP00059804	16	780	G	G	A	noncoding	n/a	n/a	n/a	n/a
36	7509150	2724447H1	SNP00000481	93	464	A	G	A	L147	n/a	n/a	n/a	n/a
36	7509150	3142215H1	SNP00000481	212	460	G	G	A	G146	n/a	n/a	n/a	n/a
36	7509150	3512238H1	SNP00059804	13	779	G	G	A	noncoding	n/a	n/a	n/a	n/a
36	7509150	3658288H1	SNP00059804	102	778	G	G	A	noncoding	n/a	n/a	n/a	n/a
36	7509150	3679887H1	SNP00000481	144	459	G	G	A	V146	n/a	n/a	n/a	n/a
36	7509150	3736042H1	SNP00059804	114	777	G	G	A	noncoding	n/a	n/a	n/a	n/a
36	7509150	3736201H1	SNP00000481	162	462	G	G	A	V147	n/a	n/a	n/a	n/a
37	1576743	1576743H1	SNP00131276	138	1223	A	A	G	L220	n/a	n/a	n/a	n/a
37	1576743	1576743H1	SNP00131277	145	1230	T	T	C	F223	n/a	n/a	n/a	n/a
37	1576743	6717148H1	SNP00131276	124	1224	A	A	G	K221	n/a	n/a	n/a	n/a
37	1576743	6717148H1	SNP00131277	131	1231	T	T	C	I223	n/a	n/a	n/a	n/a
37	1576743	8093879H1	SNP00131276	308	1175	C	T	C	F204	n/a	n/a	n/a	n/a
37	1576743	8093879H1	SNP00131277	315	1167	G	A	G	V202	n/a	n/a	n/a	n/a
38	7503892	138107H1	SNP00076457	157	1784	G	G	T	V481	n/d	n/d	n/d	n/d
38	7503892	1516454H1	SNP00062146	112	864	C	C	T	P174	n/a	n/a	n/a	n/a
38	7503892	1516454H1	SNP00062147	186	943	C	C	T	C200	0.81	0.71	0.63	0.73
38	7503892	1516454H1	SNP00093030	105	857	C	C	T	R172	0.75	0.78	0.74	0.77
38	7503892	2128132H1	SNP00054998	58	1871	C	C	G	R510	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
38	7503892	2128132H1	SNP00059376	30	1843	G	G	C	W500	n/a	n/a	n/a	n/a
38	7503892	2157084H1	SNP00076457	29	1785	G	G	T	G481	n/d	n/d	n/d	n/d
38	7503892	2355535H1	SNP00071807	208	2424	C	C	T	noncoding	n/a	n/a	n/a	n/a
38	7503892	2639966H1	SNP00125623	153	1647	A	C	A	Q435	n/a	n/a	n/a	n/a
38	7503892	2676303H1	SNP00040162	12	1124	C	C	G	Q261	n/a	n/a	n/a	n/a
38	7503892	2676303H1	SNP00113217	23	1135	A	G	A	stop264	0.84	0.83	0.74	0.84
38	7503892	2676303H1	SNP00135008	112	1224	G	G	C	S294	n/a	n/a	n/a	n/a
38	7503892	2950202H1	SNP00147627	263	2396	C	C	T	noncoding	n/a	n/a	n/a	n/a
38	7503892	2950319H1	SNP00062146	132	863	C	C	T	P174	n/a	n/a	n/a	n/a
38	7503892	2950319H1	SNP00062147	206	942	T	C	T	F200	0.81	0.71	0.63	0.73
38	7503892	2950319H1	SNP00093030	125	856	C	C	T	C171	0.75	0.78	0.74	0.77
38	7503892	2951733H1	SNP00147627	260	2392	C	C	T	noncoding	n/a	n/a	n/a	n/a
38	7503892	3532905H1	SNP00062147	195	1686	C	C	T	A448	0.81	0.71	0.63	0.73
38	7503892	4044811H1	SNP00072502	185	939	C	A	C	S199	0.20	0.31	n/a	0.20
38	7503892	4301585H1	SNP00071806	119	2403	G	A	G	noncoding	n/a	n/a	n/a	n/a
38	7503892	4894859H1	SNP00071807	163	2422	C	C	T	noncoding	n/a	n/a	n/a	n/a
38	7503892	4896066H1	SNP00076457	29	1786	G	G	T	V481	n/d	n/d	n/d	n/d
38	7503892	5260720H1	SNP00071805	111	2161	C	C	T	noncoding	n/a	n/a	n/a	n/a
38	7503892	6245790H1	SNP00072503	249	2269	T	T	C	noncoding	0.01	n/a	n/a	n/a
38	7503892	6245790H1	SNP00072504	415	2435	G	T	G	noncoding	n/a	n/a	n/a	n/a
38	7503892	6245790H1	SNP00095804	354	2374	A	A	G	noncoding	n/a	n/a	n/a	n/a
38	7503892	6426127H1	SNP00050266	45	2314	C	G	C	noncoding	n/d	n/a	n/a	n/a
38	7503892	6756088H1	SNP00113217	27	1638	G	G	A	G432	0.84	0.83	0.74	0.84
38	7503892	6756088J1	SNP00071805	377	2162	C	C	T	noncoding	n/a	n/a	n/a	n/a
38	7503892	6756088J1	SNP00136015	412	2127	G	G	A	noncoding	0.64	n/a	n/a	n/a
38	7503892	6846592H1	SNP00054998	103	1870	C	C	G	A509	n/a	n/a	n/a	n/a
38	7503892	6846592H1	SNP00059376	75	1842	C	G	C	S500	n/a	n/a	n/a	n/a
38	7503892	6846592H1	SNP00136015	361	2126	G	G	A	noncoding	0.64	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
38	7503892	7011735H1	SNP00093048	380	1671	G	A	G	S443	0.36	n/a	n/a	n/a
38	7503892	7065222H1	SNP00092287	407	2213	C	A	C	noncoding	n/d	n/d	n/d	n/d
38	7503892	7318774H1	SNP00071806	108	2404	A	A	G	noncoding	n/a	n/a	n/a	n/a
38	7503892	7342985H1	SNP00039903	240	1843	C	C	G	C500	n/a	n/a	n/a	n/a
39	7503895	138107H1	SNP00076457	157	1652	G	G	T	V437	n/d	n/d	n/d	n/d
39	7503895	1542405H1	SNP00093048	301	814	A	A	G	K157	0.36	n/a	n/a	n/a
39	7503895	2128132H1	SNP00054998	58	1739	C	C	G	R466	n/a	n/a	n/a	n/a
39	7503895	2128132H1	SNP00059376	30	1711	G	G	C	W456	n/a	n/a	n/a	n/a
39	7503895	2157084H1	SNP00076457	29	1653	G	G	T	G437	n/d	n/d	n/d	n/d
39	7503895	2355535H1	SNP00071807	208	2292	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7503895	2950202H1	SNP00147627	263	2264	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7503895	2951733H1	SNP00147627	260	2260	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7503895	4044811H1	SNP00072502	185	939	C	A	C	S199	0.20	0.31	n/a	0.20
39	7503895	4301585H1	SNP00071806	119	2271	G	A	G	noncoding	n/a	n/a	n/a	n/a
39	7503895	4549911H1	SNP00039902	171	1667	A	G	A	R442	n/a	n/a	n/a	n/a
39	7503895	4894859H1	SNP00071807	163	2290	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7503895	4896066H1	SNP00076457	29	1654	G	G	T	V437	n/d	n/d	n/d	n/d
39	7503895	5260720H1	SNP00071805	111	2029	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7503895	6245790H1	SNP00072503	249	2137	T	T	C	noncoding	0.01	n/a	n/a	n/a
39	7503895	6245790H1	SNP00072504	415	2303	G	T	G	noncoding	n/a	n/a	n/a	n/a
39	7503895	6245790H1	SNP00095804	354	2242	A	A	G	noncoding	n/a	n/a	n/a	n/a
39	7503895	6426127H1	SNP00050266	45	2182	C	G	C	noncoding	n/d	n/a	n/a	n/a
39	7503895	6756088J1	SNP00071805	377	2030	C	C	T	noncoding	n/a	n/a	n/a	n/a
39	7503895	6756088J1	SNP00136015	412	1995	G	G	A	noncoding	0.64	n/a	n/a	n/a
39	7503895	6846592H1	SNP00054998	103	1738	C	C	G	A465	n/a	n/a	n/a	n/a
39	7503895	6846592H1	SNP00059376	75	1710	C	G	C	S456	n/a	n/a	n/a	n/a
39	7503895	6846592H1	SNP00136015	361	1994	G	G	A	noncoding	0.64	n/a	n/a	n/a
39	7503895	7065222H1	SNP00092287	407	2081	C	A	C	noncoding	n/d	n/d	n/d	n/d

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
39	7503895	7081317H1	SNP00054998	153	1374	C	C	G	A344	n/a	n/a	n/a	n/a
39	7503895	7081317H1	SNP00059376	181	1347	G	G	C	S335	n/a	n/a	n/a	n/a
39	7503895	7318774H1	SNP00071806	108	2272	A	A	G	noncoding	n/a	n/a	n/a	n/a
39	7503895	7342985H1	SNP00039903	240	1711	C	C	G	C456	n/a	n/a	n/a	n/a
40	7509192	1455433H1	SNP00113850	197	190	C	C	A	P28	n/a	n/a	n/a	n/a
40	7509192	1623136H1	SNP00065851	172	664	A	A	C	noncoding	0.88	n/a	n/a	n/a
40	7509192	2048475H1	SNP00099850	108	403	G	G	T	V99	n/a	n/a	n/a	n/a
40	7509192	2051836H1	SNP00100662	196	293	G	G	A	R62	n/a	n/a	n/a	n/a
40	7509192	2105883H1	SNP00074197	76	428	C	C	T	P107	0.73	n/a	n/a	n/a
40	7509192	2539134H1	SNP00074197	23	435	C	C	T	G109	0.73	n/a	n/a	n/a
40	7509192	2953727H1	SNP00060431	270	531	A	A	G	noncoding	n/a	n/a	n/a	n/a
40	7509192	3277676H1	SNP00113850	110	187	C	C	A	R27	n/a	n/a	n/a	n/a
40	7509192	3563013H1	SNP00113850	62	189	C	C	A	R27	n/a	n/a	n/a	n/a
40	7509192	3747745H1	SNP00113850	129	191	C	C	A	P28	n/a	n/a	n/a	n/a
40	7509192	3808337H1	SNP00060431	209	529	A	A	G	noncoding	n/a	n/a	n/a	n/a
40	7509192	3808337H1	SNP00099850	81	401	G	G	T	G98	n/a	n/a	n/a	n/a
40	7509192	4861681H1	SNP00060431	176	536	A	A	G	noncoding	n/a	n/a	n/a	n/a
40	7509192	4861681H1	SNP00068654	220	580	C	C	T	noncoding	n/a	n/a	n/a	n/a
40	7509192	5633461H1	SNP00113850	75	184	C	C	A	H26	n/a	n/a	n/a	n/a
40	7509192	5949712H1	SNP00140287	172	725	C	T	C	noncoding	n/a	n/a	n/a	n/a
40	7509192	6838582H1	SNP00100662	219	302	G	G	A	G65	n/a	n/a	n/a	n/a
40	7509192	6969711H1	SNP00105646	387	222	A	A	G	V38	n/a	n/a	n/a	n/a
40	7509192	6969711H1	SNP00139136	388	223	G	G	T	D39	n/a	n/a	n/a	n/a
40	7509192	6971617H1	SNP00068655	220	776	C	C	T	noncoding	n/a	n/a	n/a	n/a
40	7509192	7120550H1	SNP00065851	251	668	A	A	C	noncoding	0.88	n/a	n/a	n/a
40	7509192	7120550H1	SNP00068654	158	575	C	C	T	noncoding	n/a	n/a	n/a	n/a
40	7509192	7165016H1	SNP00105646	140	221	A	A	G	E38	n/a	n/a	n/a	n/a
40	7509192	7165016H1	SNP00105647	13	349	C	C	T	H81	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
40	7509192	7165016H1	SNP00139136	139	222	G	G	T	V38	n/a	n/a	n/a	n/a
40	7509192	7464630H1	SNP00060432	259	733	C	C	T	noncoding	n/a	n/a	n/a	n/a
40	7509192	985748H1	SNP00068655	50	761	C	C	T	noncoding	n/a	n/a	n/a	n/a
42	7503898	138107H1	SNP00076457	157	1821	G	G	T	R493	n/d	n/d	n/d	n/d
42	7503898	1516454H1	SNP00062146	112	864	C	C	T	P174	n/a	n/a	n/a	n/a
42	7503898	1516454H1	SNP00062147	186	943	C	C	T	C200	0.81	0.71	0.63	0.73
42	7503898	1516454H1	SNP00093030	105	857	C	C	T	R172	0.75	0.78	0.74	0.77
42	7503898	1542405H1	SNP00093048	301	814	A	A	G	K157	0.36	n/a	n/a	n/a
42	7503898	2128132H1	SNP00054998	58	1908	C	C	G	A522	n/a	n/a	n/a	n/a
42	7503898	2128132H1	SNP00059376	30	1880	G	G	C	G513	n/a	n/a	n/a	n/a
42	7503898	2157084H1	SNP00076457	29	1822	G	G	T	R493	n/d	n/d	n/d	n/d
42	7503898	2355535H1	SNP00071807	208	2557	C	C	T	A738	n/a	n/a	n/a	n/a
42	7503898	2639966H1	SNP00125623	153	1684	A	C	A	T447	n/a	n/a	n/a	n/a
42	7503898	2676303H1	SNP00040162	12	1124	C	C	G	Q261	n/a	n/a	n/a	n/a
42	7503898	2676303H1	SNP00113217	23	1135	A	G	A	stop264	0.84	0.83	0.74	0.84
42	7503898	2676303H1	SNP00135008	112	1224	G	G	C	S294	n/a	n/a	n/a	n/a
42	7503898	2950319H1	SNP00062146	132	863	C	C	T	P174	n/a	n/a	n/a	n/a
42	7503898	2950319H1	SNP00062147	206	942	T	C	T	F200	0.81	0.71	0.63	0.73
42	7503898	2950319H1	SNP00093030	125	856	C	C	T	C171	0.75	0.78	0.74	0.77
42	7503898	3532905H1	SNP00062147	195	1723	C	C	T	R460	0.81	0.71	0.63	0.73
42	7503898	4044811H1	SNP00072502	185	939	C	A	C	S199	0.20	0.31	n/a	0.20
42	7503898	4301585H1	SNP00071806	119	2536	G	A	G	W731	n/a	n/a	n/a	n/a
42	7503898	4894859H1	SNP00071807	163	2555	C	C	T	P738	n/a	n/a	n/a	n/a
42	7503898	4896066H1	SNP00076457	29	1823	G	G	T	A494	n/d	n/d	n/d	n/d
42	7503898	6426127H1	SNP00050266	45	2447	C	G	C	P702	n/d	n/a	n/a	n/a
42	7503898	6756088H1	SNP00113217	27	1675	G	G	A	W444	0.84	0.83	0.74	0.84
42	7503898	6846592H1	SNP00054998	103	1907	C	C	G	R522	n/a	n/a	n/a	n/a
42	7503898	6846592H1	SNP00059376	75	1879	C	G	C	F512	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
42	7503898	6846592H1	SNP00136015	361	2163	G	G	A	W607	0.64	n/a	n/a	n/a
42	7503898	7065222H1	SNP00092287	407	2251	C	A	C	G636	n/d	n/d	n/d	n/d
42	7503898	7218024H1	SNP00071806	74	2537	A	A	G	T732	n/a	n/a	n/a	n/a
42	7503898	7342985H1	SNP00039903	240	1880	C	C	G	R513	n/a	n/a	n/a	n/a
45	7510191	7060920H1	SNP00051930	174	725	C	G	C	noncoding	n/d	n/a	n/a	n/a
45	7510191	903086H1	SNP00051930	156	728	G	G	C	noncoding	n/d	n/a	n/a	n/a
47	7500179	138105H1	SNP00007021	19	432	G	G	A	R12	0.68	0.52	0.69	0.68
47	7500179	6732880H1	SNP00121923	376	719	G	G	A	noncoding	n/a	n/a	n/a	n/a
47	7500179	6895746H1	SNP00007021	116	435	A	G	A	L13	0.68	0.52	0.69	0.68
49	7509399	1510009H1	SNP00059804	16	786	G	G	A	noncoding	n/a	n/a	n/a	n/a
49	7509399	2724447H1	SNP00000481	93	470	A	G	A	L149	n/a	n/a	n/a	n/a
49	7509399	3142215H1	SNP00000481	212	466	G	G	A	G148	n/a	n/a	n/a	n/a
49	7509399	3512238H1	SNP00059804	13	785	G	G	A	noncoding	n/a	n/a	n/a	n/a
49	7509399	3658288H1	SNP00059804	102	784	G	G	A	noncoding	n/a	n/a	n/a	n/a
49	7509399	3679887H1	SNP00000481	144	465	G	G	A	V148	n/a	n/a	n/a	n/a
49	7509399	3736042H1	SNP00059804	114	783	G	G	A	noncoding	n/a	n/a	n/a	n/a
49	7509399	3736201H1	SNP00000481	162	468	G	G	A	V149	n/a	n/a	n/a	n/a
50	7504789	1391722H1	SNP00124447	34	6733	A	A	G	S2236	n/a	n/a	n/a	n/a
50	7504789	1391722H1	SNP00124713	37	6736	A	A	G	K2237	n/a	n/a	n/a	n/a
50	7504789	1392036H1	SNP00021649	78	7044	G	G	T	A2339	n/a	n/a	n/a	n/a
50	7504789	1392036H1	SNP00061181	94	7060	C	C	T	Q2345	0.97	0.96	0.95	0.93
50	7504789	1392036H1	SNP00124714	223	7189	G	G	A	A2388	n/d	n/a	n/a	n/a
50	7504789	1392059H1	SNP00003243	177	7369	C	C	T	L2448	n/a	n/a	n/a	n/a
50	7504789	1392258H1	SNP00021648	232	6956	C	C	T	T2310	n/a	n/a	n/a	n/a
50	7504789	1392314H1	SNP00021647	28	5473	A	A	G	N1816	0.65	0.25	0.50	0.69
50	7504789	1392323H1	SNP00003243	162	7368	C	C	T	L2447	n/a	n/a	n/a	n/a
50	7504789	1392957H1	SNP00124711	63	6502	A	A	G	T2159	n/a	n/a	n/a	n/a
50	7504789	1392957H1	SNP00124712	117	6556	C	C	G	P2177	n/d	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
50	7504789	1393085H1	SNP00021650	247	7770	C	C	T	S2581	n/a	n/a	n/a	n/a
50	7504789	1393142H1	SNP00021650	242	7774	C	C	T	R2583	n/a	n/a	n/a	n/a
50	7504789	1393437H1	SNP00124710	18	6242	A	A	G	Q2072	n/a	n/a	n/a	n/a
50	7504789	1394761H1	SNP00124708	198	4344	C	C	T	G1439	n/d	n/d	n/d	n/d
50	7504789	1394830H1	SNP00124708	196	4320	C	C	T	G1431	n/d	n/d	n/d	n/d
50	7504789	1396112H1	SNP00124711	60	6501	A	A	G	A2158	n/a	n/a	n/a	n/a
50	7504789	1396112H1	SNP00124712	114	6555	C	C	G	V2176	n/d	n/a	n/a	n/a
50	7504789	1396323H1	SNP00124708	145	4321	C	C	T	P1432	n/d	n/d	n/d	n/d
50	7504789	1396467H1	SNP00124447	108	6734	A	A	G	N2236	n/a	n/a	n/a	n/a
50	7504789	1396467H1	SNP00124713	111	6737	A	A	G	K2237	n/a	n/a	n/a	n/a
50	7504789	1396593H1	SNP00124714	222	7188	G	G	A	R2387	n/d	n/a	n/a	n/a
50	7504789	1440005H1	SNP00060137	192	5868	G	G	A	P1947	n/a	n/a	n/a	n/a
50	7504789	1440388H1	SNP00021649	234	7045	G	G	T	A2340	n/a	n/a	n/a	n/a
50	7504789	1440441H1	SNP00047036	207	5956	C	C	T	R1977	n/a	n/a	n/a	n/a
50	7504789	1440441H1	SNP00061180	133	5882	T	T	C	M1952	0.98	0.80	n/d	n/d
50	7504789	1440625H1	SNP00021648	67	6957	C	C	T	T2310	n/a	n/a	n/a	n/a
50	7504789	1440625H1	SNP00061181	171	7061	C	C	T	P2345	0.97	0.96	0.95	0.93
50	7504789	1443345H1	SNP00021650	166	7773	C	C	T	A2582	n/a	n/a	n/a	n/a
50	7504789	1443885H1	SNP00124709	151	4562	C	C	T	T1512	n/a	n/a	n/a	n/a
50	7504789	2190432H1	SNP00003244	19	7550	A	A	G	Q2508	n/a	n/a	n/a	n/a
50	7504789	2190687H1	SNP00003244	45	7549	A	A	G	R2508	n/a	n/a	n/a	n/a
50	7504789	2191513H1	SNP00124714	255	7191	G	G	A	A2388	n/d	n/a	n/a	n/a
50	7504789	2191734H1	SNP00021651	62	7947	C	C	T	F2640	n/a	n/a	n/a	n/a
50	7504789	2194638H1	SNP00124714	176	7190	G	G	A	G2388	n/d	n/a	n/a	n/a
50	7504789	2462643H1	SNP00124712	116	6557	C	C	G	P2177	n/d	n/a	n/a	n/a
50	7504789	2465856H1	SNP00021649	108	7046	G	G	T	G2340	n/a	n/a	n/a	n/a
50	7504789	2465856H1	SNP00061181	124	7062	C	C	T	H2345	0.97	0.96	0.95	0.93
50	7504789	2868260H1	SNP00003244	70	7548	A	A	G	R2507	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
50	7504789	2868481H1	SNP00021650	146	7772	C	C	T	A2582	n/a	n/a	n/a	n/a
50	7504789	2868951H1	SNP00003243	123	7367	T	C	T	L2447	n/a	n/a	n/a	n/a
50	7504789	2869158H1	SNP00124714	54	7187	G	G	A	R2387	n/d	n/a	n/a	n/a
50	7504789	2870078H1	SNP00124710	26	6240	A	A	G	stop2071	n/a	n/a	n/a	n/a
50	7504789	2870594H1	SNP00021651	126	7943	C	C	T	P2639	n/a	n/a	n/a	n/a
50	7504789	2870733H1	SNP00021651	74	7942	C	C	T	H2639	n/a	n/a	n/a	n/a
50	7504789	2871002H1	SNP00021649	68	7043	G	G	T	G2339	n/a	n/a	n/a	n/a
50	7504789	2871002H1	SNP00061181	84	7059	C	C	T	V2344	0.97	0.96	0.95	0.93
50	7504789	2871369H1	SNP00021651	191	7946	C	C	T	S2640	n/a	n/a	n/a	n/a
50	7504789	2872224H1	SNP00003244	78	7547	A	A	G	N2507	n/a	n/a	n/a	n/a
50	7504789	2872964H1	SNP00060137	283	5869	G	G	A	V1948	n/a	n/a	n/a	n/a
50	7504789	2873753H1	SNP00021647	109	5471	A	A	G	K1815	0.65	0.25	0.50	0.69
50	7504789	2874421H1	SNP00021650	206	7769	C	C	T	T2581	n/a	n/a	n/a	n/a
50	7504789	2874622H1	SNP00021651	121	7944	C	C	T	H2639	n/a	n/a	n/a	n/a
50	7504789	2874873H1	SNP00021651	81	7945	C	C	T	L2640	n/a	n/a	n/a	n/a
50	7504789	2875089H1	SNP00124711	215	6503	A	A	G	N2159	n/a	n/a	n/a	n/a
50	7504789	3487175H1	SNP00060137	52	5866	G	G	A	A1947	n/a	n/a	n/a	n/a
50	7504789	3488439H1	SNP00124710	217	6241	A	A	G	K2072	n/a	n/a	n/a	n/a
50	7504789	3489102H1	SNP00021647	13	5472	A	A	G	K1815	0.65	0.25	0.50	0.69
50	7504789	3490959H1	SNP00124711	98	6499	A	A	G	T2158	n/a	n/a	n/a	n/a
50	7504789	3490959H1	SNP00124712	152	6552	C	C	G	S2175	n/d	n/a	n/a	n/a
50	7504789	7633503H1	SNP00124446	12	2096	A	A	G	E690	n/a	n/a	n/a	n/a
50	7504789	7642168H1	SNP00124445	188	3601	A	A	G	M1192	n/a	n/a	n/a	n/a
50	7504789	7754292J1	SNP00133136	219	3803	C	C	T	P1259	n/a	n/a	n/a	n/a
50	7504789	7754470J1	SNP00003243	227	6382	C	C	T	L2119	n/a	n/a	n/a	n/a
50	7504789	7754542H1	SNP00057619	449	7369	C	C	T	L2448	n/a	n/a	n/a	n/a
50	7504789	7754542H1	SNP000092458	356	7462	T	T	C	W2479	n/d	n/a	n/a	n/a
50	7504789	7756468H1	SNP00124445	131	2732	A	A	G	D902	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
50	7504789	7756707H1	SNP00124446	382	191	A	A	G	D55	n/a	n/a	n/a	n/a
50	7504789	7757409J1	SNP00003243	166	6346	C	C	T	L2107	n/a	n/a	n/a	n/a
50	7504789	7758053J1	SNP00124447	490	6754	T	T	C	W2243	n/a	n/a	n/a	n/a
50	7504789	7758129H1	SNP00124444	279	771	A	A	G	T248	n/d	n/a	n/a	n/a
50	7504789	878687H1	SNP00060137	181	5820	G	G	A	K1931	n/a	n/a	n/a	n/a
50	7504789	879037H1	SNP00124447	91	6732	A	A	G	A2235	n/a	n/a	n/a	n/a
50	7504789	879037H1	SNP00124713	94	6735	A	A	G	R2236	n/a	n/a	n/a	n/a
50	7504789	879174H1	SNP00124714	71	7181	G	G	A	C2385	n/d	n/a	n/a	n/a
50	7504789	879433H1	SNP00021648	128	6955	C	C	T	P2310	n/a	n/a	n/a	n/a
50	7504789	881292H1	SNP00021647	113	5459	G	A	G	R1811	0.65	0.25	0.50	0.69
50	7504789	881808H1	SNP00021648	128	6954	C	C	T	V2309	n/a	n/a	n/a	n/a
50	7504789	881808H1	SNP00021649	216	7042	T	G	T	S2339	n/a	n/a	n/a	n/a
50	7504789	881808H1	SNP00061181	232	7058	C	C	T	A2344	0.97	0.96	0.95	0.93
50	7504789	882442H1	SNP00047036	231	5951	C	C	T	A1975	n/a	n/a	n/a	n/a
50	7504789	882442H1	SNP00061180	157	5877	T	T	C	N1950	0.98	0.80	n/d	n/d
50	7504789	883271H1	SNP00021651	201	7948	C	C	T	L2641	n/a	n/a	n/a	n/a
53	7510882	4172221H1	SNP00095724	117	792	G	G	A	D241	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:1-28,
- 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:1-2, SEQ ID NO:7, SEQ ID NO:9-12, SEQ ID NO:15, SEQ ID NO:18, SEQ ID NO:20, and SEQ ID NO:23-24,
- c) a polypeptide comprising a naturally occurring amino acid sequence at least 92% identical to the amino acid sequence of SEQ ID NO:3,
- d) a polypeptide comprising a naturally occurring amino acid sequence at least 97% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:4 and SEQ ID NO:17,
- e) a polypeptide comprising a naturally occurring amino acid sequence at least 93% identical to the amino acid sequence of SEQ ID NO:6,
- f) a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to the amino acid sequence of SEQ ID NO:8,
- g) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO:14,
- h) a polypeptide comprising a naturally occurring amino acid sequence at least 91% identical to the amino acid sequence of SEQ ID NO:19,
- i) 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:5, SEQ ID NO:13, SEQ ID NO:16, SEQ ID NO:21-22, and SEQ ID NO:25-28,
- j) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, and
- k) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-28.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-28.

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:29-56.

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

10

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

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

15

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

20

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:1-28.

25

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:29-56,

30

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:29-30, SEQ ID NO:32, SEQ ID NO:34-35, SEQ ID NO:37, SEQ ID NO:39-40, SEQ ID NO:42-43, SEQ ID NO:47, and SEQ ID NO:52-56,

c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least

- 92% identical to the polynucleotide sequence of SEQ ID NO:31,
- d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 97% identical to the polynucleotide sequence of SEQ ID NO:36,
- e) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 93% identical to the polynucleotide sequence of SEQ ID NO:41,
- f) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:33, SEQ ID NO:38, SEQ ID NO:44-46, SEQ ID NO:48-51,
- g) a polynucleotide complementary to a polynucleotide of a),
- h) a polynucleotide complementary to a polynucleotide of b),
- i) a polynucleotide complementary to a polynucleotide of c),
- j) a polynucleotide complementary to a polynucleotide of d),
- k) a polynucleotide complementary to a polynucleotide of e),
- l) a polynucleotide complementary to a polynucleotide of f), and
- m) an RNA equivalent of a)-l).

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.

5 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:1-28.

10

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.

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

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

20 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.

25

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

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- 30 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.

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,
- 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.

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:

- a) exposing a sample comprising the target polynucleotide to 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.

29. A method of assessing 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 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,

- 5 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.

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

- 15 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:

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

25 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, wherein the antibody is labeled.

35 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:

- 5 a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, 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
10 polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-28.

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

- 15 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:

- 20 a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-28, 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,
- 25 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:1-28.

- 30 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.

- 35 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:1-28 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- 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:1-28 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-28 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a 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 NO:1-28.

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

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 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 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 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
5 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 completely complementary to said target polynucleotide.

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

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

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

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
20 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.

25 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.

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

30

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.

35 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.

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

5 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.

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

10 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.

15 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.

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.

20 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.

25 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.

76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.

30 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.

35 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.

81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.

5 82. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.

83. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:28.

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

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
25 NO:34.

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.

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

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

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

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.

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

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

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.

106. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
5 NO:51.

107. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:52.

108. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
10 NO:53.

109. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:54.

110. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
15 NO:55.

111. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
20 NO:56.

<110> INCYTE GENOMICS, INC.
DUGGAN, Brendan M.
LEE, Sally
BAUGHN, Mariah R.
HAFALIA, April J.A.
WALIA, Narinder K.
ELLIOTT, Vicki S.
LEE, Soo Yeun
KHARE, Reena
WILSON, Amy D.
JIN, Pei
HAWKINS, Phillip R.
TANG, Y.Tom
BURFORD, Neil
DING, Li
YAO, Monique G.
BECHA, Shanya D.
TRAN, Uyen K.
CHIEN, David
ZEBARJADIAN, Yeganeh
RICHARDSON, Thomas W.
KABLE, Amy E.
CHANG, Hsin-Ru
SWARNAKAR, Anita
LEHR-MASON, Patricia M.
RAMKUMAR, Jayalaxmi

<120> EXTRACELLULAR MESSENGERS

<130> PF-1152 PCT

<140> To Be Assigned
<141> Herewith

<150> US 60/314,811
<151> 2001-08-24

<150> US 60/340,584
<151> 2001-12-14

<150> US 60/350,595
<151> 2002-01-18

<150> US 60/363,432
<151> 2002-03-11

<150> US 60/364,607
<151> 2002-03-15

<150> US 60/370,761
<151> 2002-04-05

<150> US 60/391,378
<151> 2002-06-24

<160> 56
<170> PERL Program

<210> 1
<211> 94
<212> PRT
<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7486368CD1

<400> 1

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

<210> 2

<211> 474

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7488284CD1

<400> 2

Met	Ser	Glu	Glu	Ser	Asp	Gly	Lys	Leu	Ile	Glu	Asp	Ser	Leu	Ile
1				5					10					15
Gln	Leu	Arg	Cys	His	Phe	Thr	Trp	Lys	Leu	Leu	Ile	Glu	Ala	Pro
				20					25					30
Glu	Ile	Pro	Asp	Leu	Glu	Asn	Arg	Ile	Trp	Glu	Glu	Ile	Gln	Phe
				35					40					45
Leu	Asp	Thr	Lys	Tyr	Asn	Val	Gly	Ile	His	Asn	Leu	Leu	Ala	Tyr
				50					55					60
Val	Lys	His	Leu	Lys	Gly	Gln	Asn	Glu	Glu	Ala	Leu	Val	Ser	Leu
				65					70					75
Lys	Lys	Ala	Glu	Asp	Leu	Ile	Gln	Lys	Glu	His	Ala	Asn	Gln	Ala
				80					85					90
Asp	Ile	Arg	Ser	Leu	Val	Thr	Trp	Gly	Asn	Phe	Ala	Trp	Val	Tyr
				95					100					105
Tyr	His	Met	Gly	Arg	Leu	Ala	Glu	Ala	Gln	Thr	Tyr	Leu	Asp	Lys
				110					115					120
Val	Glu	Asn	Thr	Cys	Lys	Lys	Phe	Ala	Asn	Pro	Ser	Arg	Tyr	Arg
				125					130					135
Met	Glu	Cys	Pro	Glu	Val	Asp	Cys	Glu	Glu	Gly	Trp	Ala	Leu	Ala
				140					145					150
Lys	Cys	Gly	Gly	Lys	Asn	Tyr	Glu	Arg	Ala	Lys	Thr	Cys	Phe	Glu
				155					160					165
Lys	Ala	Leu	Glu	Gly	Asn	Pro	Glu	Asn	Pro	Glu	Phe	Asn	Thr	Gly
				170					175					180
Tyr	Ala	Ile	Thr	Val	Tyr	Arg	Leu	Asp	Lys	Phe	Asn	Thr	Ala	Ser
				185					190					195
Gly	Arg	Asn	Lys	Ala	Phe	Ser	Leu	His	Val	Leu	Lys	Arg	Ala	Val
				200					205					210
Arg	Leu	Asn	Pro	Asp	Asp	Val	Tyr	Ile	Arg	Val	Leu	Leu	Ala	Leu
				215					220					225
Lys	Leu	Gln	Asp	Glu	Gly	Gln	Glu	Ala	Glu	Gly	Glu	Lys	Tyr	Ile
				230					235					240
Glu	Glu	Ala	Leu	Thr	Ser	Ile	Ser	Ser	Gln	Ala	Tyr	Val	Phe	Gln

	245		250		255									
Tyr	Ala	Ala	Lys	Phe	Tyr	Arg	Arg	Lys	Gly	Ser	Val	Asp	Lys	Ala
	260		265		270									
Leu	Glu	Leu	Leu	Lys	Met	Ala	Leu	Glu	Thr	Thr	Pro	Thr	Ser	Ala
	275		280		285									
Phe	Leu	His	His	Gln	Met	Gly	Leu	Cys	Tyr	Arg	Ala	Gln	Met	Ile
	290		295		300									
Gln	Ile	Lys	Glu	Ala	Thr	Asn	Trp	Gln	Pro	Arg	Gly	Gln	Asp	Arg
	305		310		315									
Glu	Thr	Val	Asp	Arg	Leu	Val	Gln	Leu	Ala	Ile	Cys	Lys	Phe	Glu
	320		325		330									
Lys	Thr	Ile	Met	Leu	Lys	Arg	Thr	Phe	Glu	Met	Ala	Tyr	Val	Asp
	335		340		345									
Leu	Ala	Glu	Thr	Tyr	Ala	Glu	Ile	Gly	His	His	Arg	Lys	Ala	Glu
	350		355		360									
Glu	His	Phe	Gln	Glu	Gly	Leu	Arg	Met	Lys	Ile	Phe	Glu	Asp	Gln
	365		370		375									
Leu	Lys	Gln	Glu	Ile	His	Tyr	His	Tyr	Gly	Arg	Phe	Gln	Glu	His
	380		385		390									
His	Gly	Lys	Ser	Gln	Asp	Lys	Ala	Ile	Thr	His	Tyr	Leu	Lys	Gly
	395		400		405									
Leu	Lys	Ile	Glu	Lys	Met	Ser	His	Ser	Arg	Glu	Lys	Leu	Leu	Asn
	410		415		420									
Ala	Leu	Glu	Lys	Leu	Ala	Lys	Arg	Cys	Ile	His	Gln	Asn	Val	Arg
	425		430		435									
Val	Val	Glu	Ser	Val	Ser	Leu	Leu	Gly	Leu	Ile	His	Lys	Leu	Lys
	440		445		450									
Gly	Glu	Val	Ser	Asp	Ala	Leu	Leu	Cys	Tyr	Glu	Arg	Ala	Leu	Arg
	455		460		465									
Leu	Ala	Ala	Asp	Leu	Asn	Pro	Ile	Phe						
	470													

<210> 3

<211> 44

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500756CD1

<400> 3

Met	Arg	Ile	Met	Leu	Leu	Phe	Thr	Ala	Ile	Leu	Ala	Phe	Ser	Leu
1				5					10					15
Ala	Gln	Ser	Phe	Gly	Ala	Val	Cys	Lys	Glu	Pro	Gln	Glu	Glu	Val
				20					25					30
Val	Pro	Gly	Gly	Gly	Arg	Ser	Lys	Arg	Asp	Cys	Ser	Lys	Pro	
				35					40					

<210> 4

<211> 199

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7506418CD1

<400> 4

Met	Ser	Gly	Lys	Val	Thr	Lys	Pro	Lys	Glu	Glu	Lys	Asp	Ala	Ser
1				5					10					15
Lys	Glu	Pro	Gln	Leu	Lys	Gly	Ile	Val	Thr	Lys	Leu	Tyr	Ser	Arg
				20					25					30

```

Gln Gly Tyr His Leu Gln Leu Gln Ala Asp Gly Thr Ile Asp Gly
      35              40              45
Thr Lys Asp Glu Asp Ser Thr Tyr Thr Leu Phe Asn Leu Ile Pro
      50              55              60
Val Gly Leu Arg Val Val Ala Ile Gln Gly Val Gln Thr Lys Leu
      65              70              75
Tyr Leu Ala Met Asn Ser Glu Gly Tyr Leu Tyr Thr Ser Glu Leu
      80              85              90
Phe Thr Pro Glu Cys Lys Phe Lys Glu Ser Val Phe Glu Asn Tyr
      95              100             105
Tyr Val Thr Tyr Ser Ser Met Ile Tyr Arg Gln Gln Gln Ser Gly
      110             115             120
Arg Gly Trp Tyr Leu Gly Leu Asn Lys Glu Gly Glu Ile Met Lys
      125             130             135
Gly Asn His Val Lys Lys Asn Lys Pro Ala Ala His Phe Leu Pro
      140             145             150
Lys Pro Leu Lys Val Ala Met Tyr Lys Glu Pro Ser Leu His Asp
      155             160             165
Leu Thr Glu Phe Ser Arg Ser Gly Ser Gly Thr Pro Thr Lys Ser
      170             175             180
Arg Ser Val Ser Gly Val Leu Asn Gly Gly Lys Ser Met Ser His
      185             190             195
Asn Glu Ser Thr

```

```

<210> 5
<211> 115
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7506942CD1

```

```

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

```

```

<210> 6
<211> 201
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7506998CD1

```

```

<400> 6
Met Glu Ser Arg Gly Pro Leu Ala Thr Ser Arg Leu Leu Leu Leu

```

1	5	10	15
Leu Leu Leu Leu Leu	Leu Arg His Thr Arg	Gln Gly Trp Ala Leu	
	20	25	30
Arg Pro Val Leu Pro	Thr Gln Ser Ala His	Asp Pro Pro Ala Val	
	35	40	45
His Leu Ser Asn Gly	Pro Gly Gln Glu Pro	Ile Ala Val Met Thr	
	50	55	60
Phe Asp Leu Thr Lys	Ile Thr Lys Thr Ser	Ser Ser Phe Glu Val	
	65	70	75
Arg Thr Trp Asp Pro	Glu Gly Val Ile Phe	Tyr Gly Asp Thr Asn	
	80	85	90
Pro Lys Asp Asp Trp	Phe Met Leu Gly Leu	Arg Asp Gly Arg Pro	
	95	100	105
Glu Ile Gln Leu His	Asn His Trp Ala Gln	Leu Thr Val Gly Ala	
	110	115	120
Gly Pro Arg Leu Asp	Asp Gly Arg Trp His	Gln Val Glu Val Lys	
	125	130	135
Met Glu Gly Asp Ser	Val Leu Leu Glu Val	Asp Gly Glu Glu Val	
	140	145	150
Leu Arg Leu Arg Gln	Val Ser Gly Pro Leu	Thr Ser Lys Arg His	
	155	160	165
Pro Ile Met Arg Ile	Ala Leu Gly Gly Leu	Leu Phe Pro Ala Ser	
	170	175	180
Asn Leu Arg Leu Pro	Ala Pro Arg Ser Gln	His Leu Pro Pro Leu	
	185	190	195
Ala Ser Glu Ala Val	Met		
	200		

<210> 7

<211> 181

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7506999CD1

<400> 7

Met Glu Ser Arg Gly	Pro Leu Ala Thr Ser	Arg Leu Leu Leu Leu	
1	5	10	15
Leu Leu Leu Leu Leu	Leu Arg His Thr Arg	Gln Gly Trp Ala Leu	
	20	25	30
Arg Pro Val Leu Pro	Thr Gln Ser Ala His	Asp Pro Pro Ala Val	
	35	40	45
His Leu Ser Asn Gly	Pro Gly Gln Glu Pro	Ile Ala Val Met Thr	
	50	55	60
Phe Asp Leu Thr Lys	Ile Thr Lys Thr Ser	Ser Ser Phe Glu Val	
	65	70	75
Arg Thr Trp Asp Pro	Glu Gly Val Ile Phe	Tyr Gly Asp Thr Asn	
	80	85	90
Pro Lys Asp Asp Trp	Phe Met Leu Gly Leu	Arg Asp Gly Arg Pro	
	95	100	105
Glu Ile Gln Leu His	Asn His Trp Ala Gln	Leu Thr Val Gly Ala	
	110	115	120
Ala Pro Glu Thr Gly	Leu Trp Ala Pro Asp	Gln Gln Thr Pro Ser	
	125	130	135
His His Glu Asp Cys	Ala Trp Gly Ala Ala	Leu Pro Arg Phe Gln	
	140	145	150
Pro Ser Val Ala Ala	Gly Ser Cys Pro Gly	Trp Leu Pro Ala Pro	
	155	160	165
Gly Phe Leu Ala Gly	Gln Thr Gly Arg Asp	Leu Ser Ile Cys Pro	
	170	175	180

His

<210> 8
 <211> 169
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7509150CD1

<400> 8
 Met Ala Gly Pro Ala Thr Gln Ser Pro Met Lys Leu Met Ala Leu
 1 5 10 15
 Gln Leu Leu Leu Trp His Ser Ala Leu Trp Thr Val Gln Glu Ala
 20 25 30
 Thr Pro Leu Gly Pro Ala Ser Ser Leu Pro Gln Ser Phe Leu Leu
 35 40 45
 Lys Cys Leu Glu Gln Val Arg Lys Ile Gln Gly Asp Gly Ala Ala
 50 55 60
 Leu Gln Glu Lys Leu Ala Gly Cys Leu Ser Gln Leu His Ser Gly
 65 70 75
 Leu Phe Leu Tyr Gln Gly Leu Leu Gln Ala Leu Glu Trp Ile Ser
 80 85 90
 Leu Arg Val Gly Val Pro Ala Trp His Thr Leu Gln Leu Asp Val
 95 100 105
 Ala Asp Phe Ala Thr Thr Ile Trp Gln Gln Met Glu Glu Leu Gly
 110 115 120
 Met Ala Pro Ala Leu Gln Pro Thr Gln Gly Ala Met Pro Ala Phe
 125 130 135
 Ala Ser Ala Phe Gln Arg Arg Ala Gly Gly Val Leu Val Ala Ser
 140 145 150
 His Leu Gln Ser Phe Leu Glu Val Ser Tyr Arg Val Leu Arg His
 155 160 165
 Leu Ala Gln Pro

<210> 9
 <211> 254
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1576743CD1

<400> 9
 Met Glu Pro Leu Gly Gly Gly Ser Val Thr Phe Phe Gly Leu Leu
 1 5 10 15
 Pro Val Trp Lys Tyr Ser Ala Ser Gly Leu Asn Leu Pro Leu Arg
 20 25 30
 Cys Leu Asp Val Ser Gly Val Trp Thr Ala Cys Pro Leu Met Leu
 35 40 45
 Pro Ala Pro Leu Ser Ala Val Tyr Ala His Pro Gly Gly Gly Ala
 50 55 60
 Ser Trp Ala Gly Ala Pro Ala Gly Gly Asp Arg Gln Asp Ser Trp
 65 70 75
 Thr Thr Val Asn Gly Thr Phe Ser Val Tyr Phe Arg His Asn Ser
 80 85 90
 Ser Ser Leu Gly Asn Leu Ser Val Ser Ile Val Pro Pro Ser Lys
 95 100 105
 Arg Val Glu Phe Gly Gly Val Trp Leu Pro Gly Pro Val Pro His
 110 115 120

Pro	Leu	Gln	Ser	Thr	Leu	Ala	Leu	Glu	Gly	Val	Leu	Pro	Gly	Leu
				125					130					135
Gly	Pro	Pro	Leu	Gly	Met	Ala	Ala	Ala	Ala	Ala	Gly	Pro	Gly	Leu
				140					145					150
Gly	Gly	Ser	Leu	Gly	Gly	Ala	Leu	Ala	Gly	Pro	Leu	Gly	Gly	Ala
				155					160					165
Leu	Gly	Val	Pro	Gly	Ala	Lys	Glu	Ser	Arg	Ala	Phe	Asn	Cys	His
				170					175					180
Val	Glu	Tyr	Glu	Lys	Thr	Asn	Arg	Ala	Arg	Lys	His	Arg	Pro	Cys
				185					190					195
Leu	Tyr	Asp	Pro	Ser	Gln	Val	Cys	Phe	Thr	Glu	His	Thr	Gln	Ser
				200					205					210
Gln	Ala	Ala	Trp	Leu	Cys	Ala	Lys	Pro	Phe	Lys	Val	Ile	Cys	Ile
				215					220					225
Phe	Val	Ser	Phe	Leu	Ser	Phe	Asp	Tyr	Lys	Leu	Val	Gln	Lys	Val
				230					235					240
Cys	Pro	Asp	Tyr	Asn	Phe	Gln	Ser	Glu	His	Pro	Tyr	Phe	Gly	
				245					250					

<210> 10

<211> 512

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503892CD1

<400> 10

Met	Gly	Leu	Trp	Trp	Val	Thr	Val	Gln	Pro	Pro	Ala	Arg	Arg	Met
1				5					10					15
Gly	Trp	Leu	Pro	Leu	Leu	Leu	Leu	Leu	Thr	Gln	Cys	Leu	Gly	Val
				20					25					30
Pro	Gly	Gln	Arg	Ser	Pro	Leu	Asn	Asp	Phe	Gln	Val	Leu	Arg	Gly
				35					40					45
Thr	Glu	Leu	Gln	His	Leu	Leu	His	Ala	Val	Val	Pro	Gly	Pro	Trp
				50					55					60
Gln	Glu	Asp	Val	Ala	Asp	Ala	Glu	Glu	Cys	Ala	Gly	Arg	Cys	Gly
				65					70					75
Pro	Leu	Met	Asp	Cys	Arg	Ala	Phe	His	Tyr	Asn	Val	Ser	Ser	His
				80					85					90
Gly	Cys	Gln	Leu	Leu	Pro	Trp	Thr	Gln	His	Ser	Pro	His	Thr	Arg
				95					100					105
Leu	Arg	Arg	Ser	Gly	Arg	Cys	Asp	Leu	Phe	Gln	Lys	Lys	Asp	Tyr
				110					115					120
Val	Arg	Thr	Cys	Ile	Met	Asn	Asn	Gly	Val	Gly	Tyr	Arg	Gly	Thr
				125					130					135
Met	Ala	Thr	Thr	Val	Gly	Gly	Leu	Pro	Cys	Gln	Ala	Trp	Ser	His
				140					145					150
Lys	Phe	Pro	Asn	Asp	His	Lys	Tyr	Thr	Pro	Thr	Leu	Arg	Asn	Gly
				155					160					165
Leu	Glu	Glu	Asn	Phe	Cys	Arg	Asn	Pro	Asp	Gly	Asp	Pro	Gly	Gly
				170					175					180
Pro	Trp	Cys	Tyr	Thr	Thr	Asp	Pro	Ala	Val	Arg	Phe	Gln	Ser	Cys
				185					190					195
Gly	Ile	Lys	Ser	Cys	Arg	Glu	Ala	Ala	Cys	Val	Trp	Cys	Asn	Gly
				200					205					210
Glu	Glu	Tyr	Arg	Gly	Ala	Val	Asp	Arg	Thr	Glu	Ser	Gly	Arg	Glu
				215					220					225
Cys	Gln	Arg	Trp	Asp	Leu	Gln	His	Pro	His	Gln	His	Pro	Phe	Glu
				230					235					240
Pro	Gly	Lys	Phe	Leu	Asp	Gln	Gly	Leu	Asp	Asp	Asn	Tyr	Cys	Arg
				245					250					255

Asn	Pro	Asp	Gly	Ser	Glu	Arg	Pro	Trp	Cys	Tyr	Thr	Thr	Asp	Pro	
				260					265					270	
Gln	Ile	Glu	Arg	Glu	Phe	Cys	Asp	Leu	Pro	Arg	Cys	Gly	Ser	Glu	
				275					280					285	
Ala	Gln	Pro	Arg	Gln	Glu	Ala	Thr	Thr	Val	Ser	Cys	Phe	Arg	Gly	
				290					295					300	
Lys	Gly	Glu	Gly	Tyr	Arg	Gly	Thr	Ala	Asn	Thr	Thr	Thr	Ala	Gly	
				305					310					315	
Val	Pro	Cys	Gln	Arg	Trp	Asp	Ala	Gln	Ile	Pro	His	Gln	His	Arg	
				320					325					330	
Phe	Thr	Pro	Glu	Lys	Tyr	Ala	Cys	Lys	Asp	Leu	Arg	Glu	Asn	Phe	
				335					340					345	
Cys	Arg	Asn	Pro	Asp	Gly	Ser	Glu	Ala	Pro	Trp	Cys	Phe	Thr	Leu	
				350					355					360	
Arg	Pro	Gly	Met	Arg	Ala	Ala	Phe	Cys	Tyr	Gln	Ile	Arg	Arg	Cys	
				365					370					375	
Thr	Asp	Asp	Val	Arg	Pro	Gln	Asp	Cys	Tyr	His	Gly	Ala	Gly	Glu	
				380					385					390	
Gln	Tyr	Arg	Gly	Thr	Val	Ser	Lys	Thr	Arg	Lys	Gly	Val	Gln	Cys	
				395					400					405	
Gln	Arg	Trp	Ser	Ala	Glu	Thr	Pro	His	Lys	Pro	Gln	Arg	Thr	Ser	
				410					415					420	
Ala	Gly	Thr	Gln	Met	Gly	Ile	Ala	Met	Gly	Pro	Gly	Ala	Thr	Arg	
				425					430					435	
Trp	Thr	Gln	Gly	Pro	His	Ser	Thr	Thr	Val	Pro	Cys	Asp	Ala	Ala	
				440					445					450	
Leu	Met	Thr	Ser	Arg	His	Gln	Ser	Trp	Thr	Pro	Gln	Thr	Arg	Cys	
				455					460					465	
Ser	Leu	Arg	Ser	Val	Ala	Arg	Gly	Trp	Ile	Gly	Trp	Ile	Ser	Gly	
				470					475					480	
Val	Pro	Ser	Cys	Ala	Trp	Leu	Gly	Ala	Ile	Arg	Ala	Thr	His	Pro	
				485					490					495	
Gly	Gln	Ser	Ala	Cys	Gly	Ile	Gly	Arg	Ala	Ser	Ile	Ser	Ala	Gly	
				500					505					510	

Gly Leu

<210> 11
 <211> 468
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7503895CD1

<400> 11

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

Val	Arg	Thr	Cys	Ile	Met	Asn	Asn	Gly	Val	Gly	Tyr	Arg	Gly	Thr
				125					130					135
Met	Ala	Thr	Thr	Val	Gly	Gly	Leu	Pro	Cys	Gln	Ala	Trp	Ser	His
				140					145					150
Lys	Phe	Pro	Asn	Asp	His	Lys	Tyr	Thr	Pro	Thr	Leu	Arg	Asn	Gly
				155					160					165
Leu	Glu	Glu	Asn	Phe	Cys	Arg	Asn	Pro	Asp	Gly	Asp	Pro	Gly	Gly
				170					175					180
Pro	Trp	Cys	Tyr	Thr	Thr	Asp	Pro	Ala	Val	Arg	Phe	Gln	Ser	Cys
				185					190					195
Gly	Ile	Lys	Ser	Cys	Arg	Glu	Ala	Ala	Cys	Val	Trp	Cys	Asn	Gly
				200					205					210
Glu	Glu	Tyr	Arg	Gly	Ala	Val	Asp	Arg	Thr	Glu	Ser	Gly	Arg	Glu
				215					220					225
Cys	Gln	Arg	Trp	Asp	Leu	Gln	His	Pro	His	Gln	His	Pro	Phe	Glu
				230					235					240
Pro	Gly	Lys	Phe	Leu	Asp	Gln	Gly	Leu	Asp	Asp	Asn	Tyr	Cys	Arg
				245					250					255
Asn	Pro	Asp	Gly	Ser	Glu	Arg	Pro	Trp	Cys	Tyr	Thr	Thr	Asp	Pro
				260					265					270
Gln	Ile	Glu	Arg	Glu	Phe	Cys	Asp	Leu	Pro	Arg	Cys	Glu	Thr	Phe
				275					280					285
Gly	Arg	Thr	Ser	Ala	Gly	Thr	Pro	Thr	Ala	Gln	Arg	Arg	Pro	Gly
				290					295					300
Ala	Ser	His	Cys	Gly	Pro	Ala	Cys	Ala	Arg	Pro	Phe	Ala	Thr	Arg
				305					310					315
Ser	Gly	Val	Val	Gln	Thr	Thr	Cys	Gly	Pro	Arg	Thr	Ala	Thr	Thr
				320					325					330
Ala	Gln	Gly	Ser	Ser	Thr	Ala	Ala	Arg	Ser	Ala	Arg	Pro	Ala	Arg
				335					340					345
Val	Ser	Ser	Ala	Ser	Ala	Gly	Pro	Leu	Arg	Arg	Arg	Thr	Ser	Arg
				350					355					360
Ser	Ser	Arg	Leu	Pro	Pro	Asn	Arg	Met	His	Asn	Trp	Arg	Arg	Thr
				365					370					375
Ser	Ala	Gly	Thr	Gln	Met	Gly	Ile	Ala	Met	Gly	Pro	Gly	Ala	Thr
				380					385					390
Arg	Trp	Thr	Gln	Gly	Pro	His	Ser	Thr	Thr	Val	Pro	Cys	Asp	Ala
				395					400					405
Ala	Leu	Met	Thr	Ser	Arg	His	Gln	Ser	Trp	Thr	Pro	Gln	Thr	Arg
				410					415					420
Cys	Ser	Leu	Arg	Ser	Val	Ala	Arg	Gly	Trp	Ile	Gly	Trp	Ile	Ser
				425					430					435
Gly	Val	Pro	Ser	Cys	Ala	Trp	Leu	Gly	Ala	Ile	Arg	Ala	Thr	His
				440					445					450
Pro	Gly	Gln	Ser	Ala	Cys	Gly	Ile	Gly	Arg	Ala	Ser	Ile	Ser	Ala
				455					460					465

Gly Gly Leu

<210> 12

<211> 140

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509192CD1

<400> 12

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

```

Thr Leu Cys Gly Gly Glu Leu Val Asp Thr Leu Gln Phe Val Cys
      35      40      45
Gly Asp Arg Gly Phe Tyr Phe Ser Arg Pro Ala Ser Arg Val Ser
      50      55      60
Arg Arg Ser Arg Gly Ile Val Gln Gly Gly Gln Thr Ser Pro Ser
      65      70      75
Pro Asp Cys Ser Thr His Pro Arg Pro Arg Pro Arg Gly Arg Pro
      80      85      90
Pro Arg Asp Gly Gln Gln Ser Glu Val Ser Lys Thr Ala Ala Ser
      95     100     105
Leu Gln Pro Gly Ala Thr Ile Leu Gln Pro Pro Pro Asp His Gly
     110     115     120
Arg Phe His Gln Val Pro Ser Arg Lys Ser Leu Gly Ser Thr Ser
     125     130     135
Pro Trp Gly Phe Ser
      140

```

<210> 13

<211> 239

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509320CD1

<400> 13

```

Met Trp Pro Pro Gly Ser Ala Ser Gln Pro Pro Pro Ser Pro Ala
  1      5      10      15
Ala Ala Thr Gly Leu His Pro Ala Ala Arg Pro Val Ser Leu Gln
      20      25      30
Cys Arg Leu Ser Met Cys Pro Ala Arg Ser Leu Leu Leu Val Ala
      35      40      45
Thr Leu Val Leu Leu Asp His Leu Ser Leu Ala Arg Asn Leu Pro
      50      55      60
Val Ala Thr Pro Asp Pro Gly Met Phe Pro Cys Leu His His Ser
      65      70      75
Gln Asn Leu Leu Arg Ala Val Ser Asn Met Leu Gln Lys Ala Arg
      80      85      90
Gln Thr Leu Glu Phe Tyr Pro Cys Thr Ser Glu Glu Ile Asp His
      95     100     105
Glu Asp Ile Thr Lys Asp Lys Thr Ser Thr Val Glu Ala Cys Leu
     110     115     120
Pro Leu Glu Leu Thr Lys Asn Gly Ser Cys Leu Ala Ser Arg Lys
     125     130     135
Thr Ser Phe Met Met Ala Leu Cys Leu Ser Ser Ile Tyr Glu Asp
     140     145     150
Leu Lys Met Tyr Gln Val Glu Phe Lys Thr Met Asn Ala Lys Leu
     155     160     165
Leu Met Asp Pro Lys Arg Gln Ile Phe Leu Asp Gln Asn Met Leu
     170     175     180
Ala Val Ile Asp Glu Leu Met Gln Ala Leu Asn Phe Asn Ser Glu
     185     190     195
Thr Val Pro Gln Lys Ser Ser Leu Glu Glu Pro Asp Phe Tyr Lys
     200     205     210
Thr Lys Ile Lys Leu Cys Ile Leu Leu His Ala Phe Arg Ile Arg
     215     220     225
Ala Val Thr Ile Asp Arg Val Met Ser Tyr Leu Asn Ala Ser
     230     235

```

<210> 14

<211> 757

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503898CD1

<400> 14

Met	Gly	Leu	Trp	Trp	Val	Thr	Val	Gln	Pro	Pro	Ala	Arg	Arg	Met	1	5	10	15
Gly	Trp	Leu	Pro	Leu	Leu	Leu	Leu	Leu	Thr	Gln	Cys	Leu	Gly	Val	20	25	30	35
Pro	Gly	Gln	Arg	Ser	Pro	Leu	Asn	Asp	Phe	Gln	Val	Leu	Arg	Gly	40	45	50	55
Thr	Glu	Leu	Gln	His	Leu	Leu	His	Ala	Val	Val	Pro	Gly	Pro	Trp	60	65	70	75
Gln	Glu	Asp	Val	Ala	Asp	Ala	Glu	Glu	Cys	Ala	Gly	Arg	Cys	Gly	80	85	90	95
Pro	Leu	Met	Asp	Cys	Arg	Ala	Phe	His	Tyr	Asn	Val	Ser	Ser	His	100	105	110	115
Gly	Cys	Gln	Leu	Leu	Pro	Trp	Thr	Gln	His	Ser	Pro	His	Thr	Arg	120	125	130	135
Leu	Arg	Arg	Ser	Gly	Arg	Cys	Asp	Leu	Phe	Gln	Lys	Lys	Asp	Tyr	140	145	150	155
Val	Arg	Thr	Cys	Ile	Met	Asn	Asn	Gly	Val	Gly	Tyr	Arg	Gly	Thr	160	165	170	175
Met	Ala	Thr	Thr	Val	Gly	Gly	Leu	Pro	Cys	Gln	Ala	Trp	Ser	His	180	185	190	195
Lys	Phe	Pro	Asn	Asp	His	Lys	Tyr	Thr	Pro	Thr	Leu	Arg	Asn	Gly	200	205	210	215
Leu	Glu	Glu	Asn	Phe	Cys	Arg	Asn	Pro	Asp	Gly	Asp	Pro	Gly	Gly	220	225	230	235
Pro	Trp	Cys	Tyr	Thr	Thr	Asp	Pro	Ala	Val	Arg	Phe	Gln	Ser	Cys	240	245	250	255
Gly	Ile	Lys	Ser	Cys	Arg	Glu	Ala	Ala	Cys	Val	Trp	Cys	Asn	Gly	260	265	270	275
Glu	Glu	Tyr	Arg	Gly	Ala	Val	Asp	Arg	Thr	Glu	Ser	Gly	Arg	Glu	280	285	290	295
Cys	Gln	Arg	Trp	Asp	Leu	Gln	His	Pro	His	Gln	His	Pro	Phe	Glu	300	305	310	315
Pro	Gly	Lys	Phe	Leu	Asp	Gln	Gly	Leu	Asp	Asp	Asn	Tyr	Cys	Arg	320	325	330	335
Asn	Pro	Asp	Gly	Ser	Glu	Arg	Pro	Trp	Cys	Tyr	Thr	Thr	Asp	Pro	340	345	350	355
Gln	Ile	Glu	Arg	Glu	Phe	Cys	Asp	Leu	Pro	Arg	Cys	Gly	Ser	Glu	360	365	370	375
Ala	Gln	Pro	Arg	Gln	Glu	Ala	Thr	Thr	Val	Ser	Cys	Phe	Arg	Gly	380	385	390	395
Lys	Gly	Glu	Gly	Tyr	Arg	Gly	Thr	Ala	Asn	Thr	Thr	Thr	Ala	Gly	400	405	410	415
Val	Pro	Cys	Gln	Arg	Trp	Asp	Ala	Gln	Ile	Pro	His	Gln	His	Arg				
Phe	Thr	Pro	Glu	Lys	Tyr	Ala	Cys	Lys	Asp	Leu	Arg	Glu	Asn	Phe				
Cys	Arg	Asn	Pro	Asp	Gly	Ser	Glu	Ala	Pro	Trp	Cys	Phe	Thr	Leu				
Arg	Pro	Gly	Met	Arg	Ala	Ala	Phe	Cys	Tyr	Gln	Ile	Arg	Arg	Cys				
Thr	Asp	Asp	Val	Arg	Pro	Gln	Asp	Cys	Tyr	His	Gly	Ala	Gly	Glu				
Gln	Tyr	Arg	Gly	Thr	Val	Ser	Lys	Thr	Arg	Lys	Gly	Val	Gln	Cys				
Gln	Arg	Trp	Ser	Ala	Glu	Thr	Pro	His	Lys	Pro	Gln	Phe	Thr	Phe				

<400> 15														
Met	Trp	Pro	Pro	Gly	Ser	Ala	Ser	Gln	Pro	Pro	Pro	Ser	Pro	Ala
1				5					10					15
Ala	Ala	Thr	Gly	Leu	His	Pro	Ala	Ala	Arg	Pro	Val	Ser	Leu	Gln
				20					25					30
Cys	Arg	Leu	Ser	Met	Cys	Pro	Ala	Arg	Ser	Leu	Leu	Leu	Val	Ala
				35					40					45

```

Thr Leu Val Leu Leu Asp His Leu Ser Leu Ala Arg Asn Leu Pro
      50      55      60
Val Ala Thr Pro Asp Pro Gly Met Phe Pro Cys Leu His His Ser
      65      70      75
Gln Asn Leu Leu Arg Ala Val Ser Asn Met Leu Gln Lys Ala Arg
      80      85      90
Gln Thr Leu Glu Phe Tyr Pro Cys Thr Ser Glu Glu Ile Asp His
      95     100     105
Glu Asp Ile Thr Lys Asp Lys Thr Ser Thr Val Glu Ala Cys Leu
     110     115     120
Pro Leu Glu Leu Thr Lys Asn Glu Ser Cys Leu Asn Ser Arg Glu
     125     130     135
Thr Ser Phe Ile Thr Val Ser Gln Lys Met Lys Ser Phe Ser Leu
     140     145     150
Tyr Asp Glu Phe Ile Ser Leu Met Ser Asp Tyr Phe Phe Leu
     155     160

```

<210> 16
 <211> 159
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7509401CD1

```

<400> 16
Met Phe His Val Ser Phe Arg Tyr Ile Phe Gly Leu Pro Pro Leu
  1      5      10
Ile Leu Val Leu Leu Pro Val Ala Ser Ser Asp Cys Asp Ile Glu
     20     25     30
Gly Lys Asp Gly Lys Gln Tyr Glu Ser Val Leu Met Val Ser Ile
     35     40     45
Asp Gln Leu Leu Asp Ser Met Lys Glu Ile Gly Ser Asn Cys Leu
     50     55     60
Asn Asn Glu Phe Asn Phe Phe Lys Arg His Ile Cys Asp Ala Asn
     65     70     75
Lys Glu Gly Met Phe Leu Phe Arg Ala Ala Arg Lys Leu Arg Gln
     80     85     90
Phe Leu Lys Met Asn Ser Thr Gly Asp Phe Asp Leu His Leu Leu
     95    100    105
Lys Val Ser Glu Gly Thr Thr Ile Leu Leu Asn Cys Thr Gly Gln
    110    115    120
Glu Glu Asn Lys Ser Leu Lys Glu Gln Lys Lys Leu Asn Asp Leu
    125    130    135
Cys Phe Leu Lys Arg Leu Leu Gln Glu Ile Lys Thr Cys Trp Asn
    140    145    150
Lys Ile Leu Met Gly Thr Lys Glu His
    155

```

<210> 17
 <211> 114
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7510191CD1

```

<400> 17
Met Lys Ile Leu Val Ala Leu Ala Val Phe Phe Leu Val Ser Thr
  1      5      10
Gln Leu Phe Ala Glu Glu Ile Gly Ala Asn Asp Asp Leu Asn Tyr

```


<223> Incyte ID No: 7509149CD1

<400> 20

Met	Phe	His	Val	Ser	Phe	Arg	Tyr	Ile	Phe	Gly	Leu	Pro	Pro	Leu	
1				5					10					15	
Ile	Leu	Val	Leu	Leu	Pro	Val	Ala	Ser	Ser	Asp	Cys	Asp	Ile	Glu	
				20					25					30	
Gly	Gln	His	Glu	Arg	Asn	Trp									
				35											

<210> 21

<211> 171

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509399CD1

<400> 21

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

<210> 22

<211> 2747

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7504789CD1

<400> 22

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

Val	Gln	Cys	Gln	Asn	Asp	Gly	Arg	Ser	Cys	Trp	Cys	Val	Gly	Ala	
				65					70					75	
Asn	Gly	Ser	Glu	Val	Leu	Gly	Ser	Arg	Gln	Pro	Gly	Arg	Pro	Val	
				80					85					90	
Val	Cys	Leu	Ser	Phe	Cys	Gln	Leu	Gln	Lys	Gln	Gln	Ile	Leu	Leu	
				95					100					105	
Ser	Gly	Tyr	Ile	Asn	Ser	Thr	Asp	Thr	Ser	Tyr	Leu	Pro	Gln	Cys	
				110					115					120	
Gln	Asp	Ser	Gly	Asp	Tyr	Ala	Pro	Val	Gln	Cys	Asp	Val	Gln	Gln	
				125					130					135	
Val	Gln	Cys	Trp	Cys	Val	Asp	Ala	Glu	Gly	Met	Glu	Val	Tyr	Gly	
				140					145					150	
Thr	Arg	Gln	Leu	Gly	Arg	Pro	Lys	Arg	Cys	Pro	Arg	Ser	Cys	Glu	
				155					160					165	
Ile	Arg	Asn	Arg	Arg	Leu	Leu	His	Gly	Val	Gly	Asp	Lys	Ser	Pro	
				170					175					180	
Pro	Gln	Cys	Ser	Ala	Glu	Gly	Glu	Phe	Met	Pro	Val	Gln	Cys	Lys	
				185					190					195	
Phe	Val	Asn	Thr	Thr	Asp	Met	Met	Ile	Phe	Asp	Leu	Val	His	Ser	
				200					205					210	
Tyr	Asn	Arg	Phe	Pro	Asp	Ala	Phe	Val	Thr	Phe	Ser	Ser	Phe	Gln	
				215					220					225	
Arg	Arg	Phe	Pro	Glu	Val	Ser	Gly	Tyr	Cys	His	Cys	Ala	Asp	Ser	
				230					235					240	
Gln	Gly	Arg	Glu	Leu	Ala	Glu	Thr	Gly	Leu	Glu	Leu	Leu	Leu	Asp	
				245					250					255	
Glu	Ile	Tyr	Asp	Thr	Ile	Phe	Ala	Gly	Leu	Asp	Leu	Pro	Ser	Thr	
				260					265					270	
Phe	Thr	Glu	Thr	Thr	Leu	Tyr	Arg	Ile	Leu	Gln	Arg	Arg	Phe	Leu	
				275					280					285	
Ala	Val	Gln	Ser	Val	Ile	Ser	Gly	Arg	Phe	Arg	Cys	Pro	Thr	Lys	
				290					295					300	
Cys	Glu	Val	Glu	Arg	Phe	Thr	Ala	Thr	Ser	Phe	Gly	His	Pro	Tyr	
				305					310					315	
Val	Pro	Ser	Cys	Arg	Arg	Asn	Gly	Asp	Tyr	Gln	Ala	Val	Gln	Cys	
				320					325					330	
Gln	Thr	Glu	Gly	Pro	Cys	Trp	Cys	Val	Asp	Ala	Gln	Gly	Lys	Glu	
				335					340					345	
Met	His	Gly	Thr	Arg	Gln	Gln	Gly	Glu	Pro	Pro	Ser	Cys	Ala	Glu	
				350					355					360	
Gly	Gln	Ser	Cys	Ala	Ser	Glu	Arg	Gln	Gln	Ala	Leu	Ser	Arg	Leu	
				365					370					375	
Tyr	Phe	Gly	Thr	Ser	Gly	Tyr	Phe	Ser	Gln	His	Asp	Leu	Phe	Ser	
				380					385					390	
Ser	Pro	Glu	Lys	Arg	Trp	Ala	Ser	Pro	Arg	Val	Ala	Arg	Phe	Ala	
				395					400					405	
Thr	Ser	Cys	Pro	Pro	Thr	Ile	Lys	Glu	Leu	Phe	Val	Asp	Ser	Gly	
				410					415					420	
Leu	Leu	Arg	Pro	Met	Val	Glu	Gly	Gln	Ser	Gln	Gln	Phe	Ser	Val	
				425					430					435	
Ser	Glu	Asn	Leu	Leu	Lys	Glu	Ala	Ile	Arg	Ala	Ile	Phe	Pro	Ser	
				440					445					450	
Arg	Gly	Leu	Ala	Arg	Leu	Ala	Leu	Gln	Phe	Thr	Thr	Asn	Pro	Lys	
				455					460					465	
Arg	Leu	Gln	Gln	Asn	Leu	Phe	Gly	Gly	Lys	Phe	Leu	Val	Asn	Val	
				470					475					480	
Gly	Gln	Phe	Asn	Leu	Ser	Gly	Ala	Leu	Gly	Thr	Arg	Gly	Thr	Phe	
				485					490					495	
Asn	Phe	Ser	Gln	Phe	Phe	Gln	Gln	Leu	Gly	Leu	Ala	Ser	Phe	Leu	
				500					505					510	
Asn	Gly	Gly	Arg	Gln	Glu	Asp	Leu	Ala	Lys	Pro	Leu	Ser	Val	Gly	
				515					520					525	
Leu	Asp	Ser	Asn	Ser	Ser	Thr	Gly	Thr	Pro	Glu	Ala	Ala	Lys	Lys	

Asp Gly Thr Met	530	Asn Lys Pro Thr Val	535	Gly Ser Phe Gly Phe	540
	545		550		555
Ile Asn Leu Gln	560	Glu Asn Gln Asn Ala	565	Leu Lys Phe Leu Ala	570
Leu Leu Glu Leu	575	Pro Glu Phe Leu Leu	580	Phe Leu Gln His Ala	585
Ser Val Pro Glu	590	Asp Val Ala Arg Asp	595	Leu Gly Asp Val Met	600
Thr Val Leu Ser	605	Ser Gln Thr Cys Glu	610	Gln Thr Pro Glu Arg	615
Phe Val Pro Ser	620	Cys Thr Thr Glu Gly	625	Ser Tyr Glu Asp Val	630
Cys Phe Ser Gly	635	Glu Cys Trp Cys Val	640	Asn Ser Trp Gly Lys	645
Leu Pro Gly Ser	650	Arg Val Arg Gly Gly	655	Gln Pro Arg Cys Pro	660
Asp Cys Glu Lys	665	Gln Arg Ala Arg Met	670	Gln Ser Leu Met Gly	675
Gln Pro Ala Gly	680	Ser Thr Leu Phe Val	685	Pro Ala Cys Thr Ser	690
Gly His Phe Leu	695	Pro Val Gln Cys Phe	700	Asn Ser Glu Cys Tyr	705
Val Asp Ala Glu	710	Gly Gln Ala Ile Pro	715	Gly Thr Arg Ser Ala	720
Gly Lys Pro Lys	725	Lys Cys Pro Thr Pro	730	Cys Gln Leu Gln Ser	735
Gln Ala Phe Leu	740	Arg Thr Val Gln Ala	745	Leu Leu Ser Asn Ser	750
Met Leu Pro Thr	755	Leu Ser Asp Thr Tyr	760	Ile Pro Gln Cys Ser	765
Asp Gly Gln Trp	770	Arg Gln Val Gln Cys	775	Asn Gly Pro Pro Glu	780
Val Phe Glu Leu	785	Tyr Gln Arg Trp Glu	790	Ala Gln Asn Lys Gly	795
Asp Leu Thr Pro	800	Ala Lys Leu Leu Val	805	Lys Ile Met Ser Tyr	810
Glu Ala Ala Ser	815	Gly Asn Phe Ser Leu	820	Phe Ile Gln Ser Leu	825
Glu Ala Gly Gln	830	Gln Asp Val Phe Pro	835	Val Leu Ser Gln Tyr	840
Ser Leu Gln Asp	845	Val Pro Leu Ala Ala	850	Leu Glu Gly Lys Arg	855
Gln Pro Arg Glu	860	Asn Ile Leu Leu Glu	865	Pro Tyr Leu Phe Trp	870
Ile Leu Asn Gly	875	Gln Leu Ser Gln Tyr	880	Pro Gly Ser Tyr Ser	885
Phe Ser Thr Pro	890	Leu Ala His Phe Asp	895	Leu Arg Asn Cys Trp	900
Val Asp Glu Ala	905	Gly Gln Glu Leu Glu	910	Gly Met Arg Ser Glu	915
Ser Lys Leu Pro	920	Thr Cys Pro Gly Ser	925	Glu Glu Ala Lys Leu	930
Arg Val Leu Gln	935	Phe Ile Arg Glu Thr	940	Glu Glu Ile Val Ser	945
Ser Asn Ser Ser	950	Arg Phe Pro Leu Gly	955	Glu Ser Phe Leu Val	960
Lys Gly Ile Arg	965	Leu Arg Asn Glu Asp	970	Leu Gly Leu Pro Pro	975
Phe Pro Pro Arg	980	Glu Ala Phe Ala Glu	985	Phe Leu Arg Gly Ser	990
Asp Tyr Ala Ile	995	Arg Leu Ala Ala Gln	1000	Ser Thr Leu Ser Phe	1005

Gln	Arg	Arg	Arg	Phe	Ser	Pro	Asp	Asp	Ser	Ala	Gly	Ala	Ser	Ala
				1010					1015					1020
Leu	Leu	Arg	Ser	Gly	Pro	Tyr	Met	Pro	Gln	Cys	Asp	Ala	Phe	Gly
				1025					1030					1035
Ser	Trp	Glu	Pro	Val	Gln	Cys	His	Ala	Gly	Thr	Gly	His	Cys	Trp
				1040					1045					1050
Cys	Val	Asp	Glu	Lys	Gly	Gly	Phe	Ile	Pro	Gly	Ser	Leu	Thr	Ala
				1055					1060					1065
Arg	Ser	Leu	Gln	Ile	Pro	Gln	Cys	Pro	Thr	Thr	Cys	Glu	Lys	Ser
				1070					1075					1080
Arg	Thr	Ser	Gly	Leu	Leu	Ser	Ser	Trp	Lys	Gln	Ala	Arg	Ser	Gln
				1085					1090					1095
Glu	Asn	Pro	Ser	Pro	Lys	Asp	Leu	Phe	Val	Pro	Ala	Cys	Leu	Glu
				1100					1105					1110
Thr	Gly	Glu	Tyr	Ala	Arg	Leu	Gln	Ala	Ser	Gly	Ala	Gly	Thr	Trp
				1115					1120					1125
Cys	Val	Asp	Pro	Ala	Ser	Gly	Glu	Glu	Leu	Arg	Pro	Gly	Ser	Ser
				1130					1135					1140
Ser	Ser	Ala	Gln	Cys	Pro	Ser	Leu	Cys	Asn	Val	Leu	Lys	Ser	Gly
				1145					1150					1155
Val	Leu	Ser	Arg	Arg	Val	Ser	Pro	Gly	Tyr	Val	Pro	Ala	Cys	Arg
				1160					1165					1170
Ala	Glu	Asp	Gly	Gly	Phe	Ser	Pro	Val	Gln	Cys	Asp	Gln	Ala	Gln
				1175					1180					1185
Gly	Ser	Cys	Trp	Cys	Val	Met	Asp	Ser	Gly	Glu	Glu	Val	Pro	Gly
				1190					1195					1200
Thr	Arg	Val	Thr	Gly	Gly	Gln	Pro	Ala	Cys	Glu	Ser	Pro	Arg	Cys
				1205					1210					1215
Pro	Leu	Pro	Phe	Asn	Ala	Ser	Glu	Val	Val	Gly	Gly	Thr	Ile	Leu
				1220					1225					1230
Cys	Glu	Thr	Ile	Ser	Gly	Pro	Thr	Gly	Ser	Ala	Met	Gln	Gln	Cys
				1235					1240					1245
Gln	Leu	Leu	Cys	Arg	Gln	Gly	Ser	Trp	Ser	Val	Phe	Pro	Pro	Gly
				1250					1255					1260
Pro	Leu	Ile	Cys	Ser	Leu	Glu	Ser	Gly	Arg	Trp	Glu	Ser	Gln	Leu
				1265					1270					1275
Pro	Gln	Pro	Arg	Ala	Cys	Gln	Arg	Pro	Gln	Leu	Trp	Gln	Thr	Ile
				1280					1285					1290
Gln	Thr	Gln	Gly	His	Phe	Gln	Leu	Gln	Leu	Pro	Pro	Gly	Lys	Met
				1295					1300					1305
Cys	Ser	Ala	Asp	Tyr	Ala	Gly	Leu	Leu	Gln	Thr	Phe	Gln	Val	Phe
				1310					1315					1320
Ile	Leu	Asp	Glu	Leu	Thr	Ala	Arg	Gly	Phe	Cys	Gln	Ile	Gln	Val
				1325					1330					1335
Lys	Thr	Phe	Gly	Thr	Leu	Val	Ser	Ile	Pro	Val	Cys	Asn	Asn	Ser
				1340					1345					1350
Ser	Val	Gln	Val	Gly	Cys	Leu	Thr	Arg	Glu	Arg	Leu	Gly	Val	Asn
				1355					1360					1365
Val	Thr	Trp	Lys	Ser	Arg	Leu	Glu	Asp	Ile	Pro	Val	Ala	Ser	Leu
				1370					1375					1380
Pro	Asp	Leu	His	Asp	Ile	Glu	Arg	Ala	Leu	Val	Gly	Lys	Asp	Leu
				1385					1390					1395
Leu	Gly	Arg	Phe	Thr	Asp	Leu	Ile	Gln	Ser	Gly	Ser	Phe	Gln	Leu
				1400					1405					1410
His	Leu	Asp	Ser	Lys	Thr	Phe	Pro	Ala	Glu	Thr	Ile	Arg	Phe	Leu
				1415					1420					1425
Gln	Gly	Asp	His	Phe	Gly	Thr	Ser	Pro	Arg	Thr	Trp	Phe	Gly	Cys
				1430					1435					1440
Ser	Glu	Gly	Phe	Tyr	Gln	Val	Leu	Thr	Ser	Glu	Ala	Ser	Gln	Asp
				1445					1450					1455
Gly	Leu	Gly	Cys	Val	Lys	Cys	Pro	Glu	Gly	Ser	Tyr	Ser	Gln	Asp
				1460					1465					1470
Glu	Glu	Cys	Ile	Pro	Cys	Pro	Val	Gly	Phe	Tyr	Gln	Glu	Gln	Ala

1475	1480	1485
Gly Ser Leu Ala Cys Val Pro Cys Pro Val Gly Arg Thr Thr Ile		
1490	1495	1500
Ser Ala Gly Ala Phe Ser Gln Thr His Cys Val Thr Asp Cys Gln		
1505	1510	1515
Arg Asn Glu Ala Gly Leu Gln Cys Asp Gln Asn Gly Gln Tyr Arg		
1520	1525	1530
Ala Ser Gln Lys Asp Arg Gly Ser Gly Lys Ala Phe Cys Val Asp		
1535	1540	1545
Gly Glu Gly Arg Arg Leu Pro Trp Trp Glu Thr Glu Ala Pro Leu		
1550	1555	1560
Glu Asp Ser Gln Cys Leu Met Met Gln Lys Phe Glu Lys Val Pro		
1565	1570	1575
Glu Ser Lys Val Ile Phe Asp Ala Asn Ala Pro Val Ala Val Arg		
1580	1585	1590
Ser Lys Val Pro Asp Ser Glu Phe Pro Val Met Gln Cys Leu Thr		
1595	1600	1605
Asp Cys Thr Glu Asp Glu Ala Cys Ser Phe Phe Thr Val Ser Thr		
1610	1615	1620
Thr Glu Pro Glu Ile Ser Cys Asp Phe Tyr Ala Trp Thr Ser Asp		
1625	1630	1635
Asn Val Ala Cys Met Thr Ser Asp Gln Lys Arg Asp Ala Leu Gly		
1640	1645	1650
Asn Ser Lys Ala Thr Ser Phe Gly Ser Leu Arg Cys Gln Val Lys		
1655	1660	1665
Val Arg Ser His Gly Gln Asp Ser Pro Ala Val Tyr Leu Lys Lys		
1670	1675	1680
Gly Gln Gly Ser Thr Thr Thr Leu Gln Lys Arg Phe Glu Pro Thr		
1685	1690	1695
Gly Phe Gln Asn Met Leu Ser Gly Leu Tyr Asn Pro Ile Val Phe		
1700	1705	1710
Ser Ala Ser Gly Ala Asn Leu Thr Asp Ala His Leu Phe Cys Leu		
1715	1720	1725
Leu Ala Cys Asp Arg Asp Leu Cys Cys Asp Gly Phe Val Leu Thr		
1730	1735	1740
Gln Val Gln Gly Gly Ala Ile Ile Cys Gly Leu Leu Ser Ser Pro		
1745	1750	1755
Ser Val Leu Leu Cys Asn Val Lys Asp Trp Met Asp Pro Ser Glu		
1760	1765	1770
Ala Trp Ala Asn Ala Thr Cys Pro Gly Val Thr Tyr Asp Gln Glu		
1775	1780	1785
Ser His Gln Val Ile Leu Arg Leu Gly Asp Gln Glu Phe Ile Lys		
1790	1795	1800
Asn Ser Asp Met Gly Ser Arg Pro Glu Ser Met Gly Cys Arg Lys		
1805	1810	1815
Asn Thr Val Pro Arg Pro Ala Ser Pro Thr Glu Ala Gly Leu Thr		
1820	1825	1830
Thr Glu Leu Phe Ser Pro Val Asp Leu Asn Gln Val Ile Val Asn		
1835	1840	1845
Gly Asn Gln Ser Leu Ser Ser Gln Lys His Trp Leu Phe Lys His		
1850	1855	1860
Leu Phe Ser Ala Gln Gln Ala Asn Leu Trp Cys Leu Ser Arg Cys		
1865	1870	1875
Val Gln Glu His Ser Phe Cys Gln Leu Ala Glu Ile Thr Glu Ser		
1880	1885	1890
Ala Ser Leu Tyr Phe Thr Cys Thr Leu Tyr Pro Glu Ala Gln Val		
1895	1900	1905
Cys Asp Asp Ile Met Glu Ser Asn Ala Gln Gly Cys Arg Leu Ile		
1910	1915	1920
Leu Pro Gln Met Pro Lys Ala Leu Phe Arg Lys Lys Val Ile Leu		
1925	1930	1935
Glu Asp Lys Val Lys Asn Phe Tyr Thr Arg Leu Pro Phe Gln Lys		
1940	1945	1950

Leu	Met	Gly	Ile	Ser	Ile	Arg	Asn	Lys	Val	Pro	Met	Ser	Glu	Lys
				1955					1960					1965
Ser	Ile	Ser	Asn	Gly	Phe	Phe	Glu	Cys	Glu	Arg	Trp	Cys	Asp	Ala
				1970					1975					1980
Asp	Pro	Cys	Cys	Thr	Gly	Phe	Gly	Phe	Leu	Asn	Val	Ser	Gln	Leu
				1985					1990					1995
Lys	Gly	Gly	Glu	Val	Thr	Cys	Leu	Thr	Leu	Asn	Ser	Leu	Gly	Ile
				2000					2005					2010
Gln	Met	Cys	Ser	Glu	Glu	Asn	Gly	Gly	Ala	Trp	Arg	Ile	Leu	Asp
				2015					2020					2025
Cys	Gly	Ser	Pro	Asp	Ile	Glu	Val	His	Thr	Tyr	Pro	Phe	Gly	Trp
				2030					2035					2040
Tyr	Gln	Lys	Pro	Ile	Ala	Gln	Asn	Asn	Ala	Pro	Ser	Phe	Cys	Pro
				2045					2050					2055
Leu	Val	Val	Leu	Pro	Ser	Leu	Thr	Glu	Lys	Val	Ser	Leu	Asp	Ser
				2060					2065					2070
Trp	Gln	Ser	Leu	Ala	Leu	Ser	Ser	Val	Val	Val	Asp	Pro	Ser	Ile
				2075					2080					2085
Arg	His	Phe	Asp	Val	Ala	His	Val	Ser	Thr	Ala	Ala	Thr	Ser	Asn
				2090					2095					2100
Phe	Ser	Ala	Val	Arg	Asp	Leu	Cys	Leu	Ser	Glu	Cys	Ser	Gln	His
				2105					2110					2115
Glu	Ala	Cys	Leu	Ile	Thr	Thr	Leu	Gln	Thr	Gln	Pro	Gly	Ala	Val
				2120					2125					2130
Arg	Cys	Met	Phe	Tyr	Ala	Asp	Thr	Gln	Ser	Cys	Thr	His	Ser	Leu
				2135					2140					2145
Gln	Gly	Gln	Asn	Cys	Arg	Leu	Leu	Leu	Arg	Glu	Glu	Ala	Thr	His
				2150					2155					2160
Ile	Tyr	Arg	Lys	Pro	Gly	Ile	Ser	Leu	Leu	Ser	Tyr	Glu	Ala	Ser
				2165					2170					2175
Val	Pro	Ser	Val	Pro	Ile	Ser	Thr	His	Gly	Arg	Leu	Leu	Gly	Arg
				2180					2185					2190
Ser	Gln	Ala	Ile	Gln	Val	Gly	Thr	Ser	Trp	Lys	Gln	Val	Asp	Gln
				2195					2200					2205
Phe	Leu	Gly	Val	Pro	Tyr	Ala	Ala	Pro	Pro	Leu	Ala	Glu	Arg	Arg
				2210					2215					2220
Phe	Gln	Ala	Pro	Glu	Pro	Leu	Asn	Trp	Thr	Gly	Ser	Trp	Asp	Ala
				2225					2230					2235
Ser	Lys	Pro	Arg	Ala	Ser	Cys	Trp	Gln	Pro	Gly	Thr	Arg	Thr	Ser
				2240					2245					2250
Thr	Ser	Pro	Gly	Val	Ser	Glu	Asp	Cys	Leu	Tyr	Leu	Asn	Val	Phe
				2255					2260					2265
Ile	Pro	Gln	Asn	Val	Ala	Pro	Asn	Ala	Ser	Val	Leu	Val	Phe	Phe
				2270					2275					2280
His	Asn	Thr	Met	Asp	Arg	Glu	Glu	Ser	Glu	Gly	Trp	Pro	Ala	Ile
				2285					2290					2295
Asp	Gly	Ser	Phe	Leu	Ala	Ala	Val	Gly	Asn	Leu	Ile	Val	Val	Thr
				2300					2305					2310
Ala	Ser	Tyr	Arg	Val	Gly	Val	Phe	Gly	Phe	Leu	Ser	Ser	Gly	Ser
				2315					2320					2325
Gly	Glu	Val	Ser	Gly	Asn	Trp	Gly	Leu	Leu	Asp	Gln	Val	Ala	Ala
				2330					2335					2340
Leu	Thr	Trp	Val	Gln	Thr	His	Ile	Arg	Gly	Phe	Gly	Gly	Asp	Pro
				2345					2350					2355
Arg	Arg	Val	Ser	Leu	Ala	Ala	Asp	Arg	Gly	Gly	Ala	Asp	Val	Ala
				2360					2365					2370
Ser	Ile	His	Leu	Leu	Thr	Ala	Arg	Ala	Thr	Asn	Ser	Gln	Leu	Phe
				2375					2380					2385
Arg	Arg	Ala	Val	Leu	Met	Gly	Gly	Ser	Ala	Leu	Ser	Pro	Ala	Ala
				2390					2395					2400
Val	Ile	Ser	His	Glu	Arg	Ala	Gln	Gln	Ala	Ile	Ala	Leu	Ala	
				2405					2410					2415
Lys	Glu	Val	Ser	Cys	Pro	Met	Ser	Ser	Ser	Gln	Glu	Val	Val	Ser

2420	2425	2430
Cys Leu Arg Gln Lys Pro Ala Asn Val Leu Asn Asp Ala Gln Thr		
2435	2440	2445
Lys Leu Leu Ala Val Ser Gly Pro Phe His Tyr Trp Gly Pro Val		
2450	2455	2460
Ile Asp Gly His Phe Leu Arg Glu Pro Pro Ala Arg Ala Leu Lys		
2465	2470	2475
Arg Ser Leu Trp Val Glu Val Asp Leu Leu Ile Gly Ser Ser Gln		
2480	2485	2490
Asp Asp Gly Leu Ile Asn Arg Ala Lys Ala Val Lys Gln Phe Glu		
2495	2500	2505
Glu Ser Arg Gly Arg Thr Ser Ser Lys Thr Ala Phe Tyr Gln Ala		
2510	2515	2520
Leu Gln Asn Ser Leu Gly Gly Glu Asp Ser Asp Ala Arg Val Glu		
2525	2530	2535
Ala Ala Ala Thr Trp Tyr Tyr Ser Leu Glu His Ser Thr Asp Asp		
2540	2545	2550
Tyr Ala Ser Phe Ser Arg Ala Leu Glu Asn Ala Thr Arg Asp Tyr		
2555	2560	2565
Phe Ile Ile Cys Pro Ile Ile Asp Met Ala Ser Ala Trp Ala Lys		
2570	2575	2580
Arg Ala Arg Gly Asn Val Phe Met Tyr His Ala Pro Glu Asn Tyr		
2585	2590	2595
Gly His Gly Ser Leu Glu Leu Leu Ala Asp Val Gln Phe Ala Leu		
2600	2605	2610
Gly Leu Pro Phe Tyr Pro Ala Tyr Glu Gly Gln Phe Ser Leu Glu		
2615	2620	2625
Glu Lys Ser Leu Ser Leu Lys Ile Met Gln Tyr Phe Ser His Phe		
2630	2635	2640
Ile Arg Ser Gly Asn Pro Asn Tyr Pro Tyr Glu Phe Ser Arg Lys		
2645	2650	2655
Val Pro Thr Phe Ala Thr Pro Trp Pro Asp Phe Val Pro Arg Ala		
2660	2665	2670
Gly Gly Glu Asn Tyr Lys Glu Phe Ser Glu Leu Leu Pro Asn Arg		
2675	2680	2685
Gln Gly Leu Lys Lys Ala Asp Cys Ser Phe Trp Ser Lys Tyr Ile		
2690	2695	2700
Ser Ser Leu Lys Thr Ser Ala Gly Asn Gly Ala Lys Gly Gly Gln		
2705	2710	2715
Ser Ala Glu Ser Glu Glu Glu Glu Leu Thr Ala Gly Ser Gly Leu		
2720	2725	2730
Arg Glu Asp Leu Leu Ser Leu Gln Glu Pro Gly Ser Lys Thr Tyr		
2735	2740	2745
Ser Lys		

<210> 23

<211> 64

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7506910CD1

<400> 23

Met Tyr Arg Met Gln Leu Leu Ser Cys Ile Ala Leu Ser Leu Ala		
1	5	10
Leu Val Thr Asn Ser Ala Pro Thr Ser Ser Ser Thr Lys Lys Thr		
	20	25
Gln Leu Gln Leu Glu His Leu Leu Leu Asp Leu Gln Met Ile Leu		
	35	40
Asn Gly Ile Asn Val Ser Ile Phe Pro Phe Leu Leu Lys Leu Leu		

His Leu Val Ile 50 55 60

<210> 24
 <211> 57
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <223> Incyte ID No: 7507000CD1

<400> 24
 Met Glu Ser Arg Gly Pro Leu Ala Thr Ser Arg Leu Leu Leu Leu
 1 5 10 15
 Leu Leu Leu Leu Leu Leu Arg His Thr Arg Gln Gly Trp Val Leu
 20 25 30
 Arg Pro Val Leu Pro Thr Gln Val Gln Glu Arg Asp Arg Ala Leu
 35 40 45
 Ser Ser Cys Ser Leu Pro Phe Ser Pro Leu Ala Leu
 50 55

<210> 25
 <211> 287
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <223> Incyte ID No: 7510882CD1

<400> 25
 Met Glu Ser Arg Gly Pro Leu Ala Thr Ser Arg Leu Leu Leu Leu
 1 5 10 15
 Leu Leu Leu Leu Leu Leu Arg His Thr Arg Gln Gly Trp Ala Leu
 20 25 30
 Arg Pro Val Leu Pro Thr Gln Ser Ala His Asp Pro Pro Ala Val
 35 40 45
 His Leu Ser Asn Gly Pro Gly Gln Glu Pro Ile Ala Val Met Thr
 50 55 60
 Phe Asp Leu Thr Lys Ile Thr Lys Thr Ser Ser Ser Phe Glu Val
 65 70 75
 Arg Thr Trp Asp Pro Glu Gly Val Ile Phe Tyr Gly Asp Thr Asn
 80 85 90
 Pro Lys Asp Asp Trp Phe Met Leu Gly Leu Arg Asp Gly Arg Pro
 95 100 105
 Glu Ile Gln Leu His Asn His Trp Ala Gln Leu Thr Val Gly Ala
 110 115 120
 Gly Pro Arg Leu Asp Asp Gly Arg Trp His Gln Val Glu Val Lys
 125 130 135
 Met Glu Gly Asp Ser Val Leu Leu Glu Val Asp Gly Glu Glu Val
 140 145 150
 Leu Arg Leu Arg Gln Val Ser Gly Pro Leu Thr Ser Lys Arg His
 155 160 165
 Pro Ile Met Arg Ile Ala Leu Gly Gly Leu Leu Phe Pro Ala Ser
 170 175 180
 Asn Leu Arg Leu Pro Leu Val Pro Ala Leu Asp Gly Cys Leu Arg
 185 190 195
 Arg Asp Ser Trp Leu Asp Lys Gln Ala Glu Ile Ser Ala Ser Ala
 200 205 210
 Pro Thr Ser Leu Arg Ser Cys Asp Val Glu Ser Asn Pro Gly Ile
 215 220 225

Phe	Leu	Pro	Pro	Gly	Thr	Gln	Ala	Glu	Phe	Asn	Leu	Arg	Gly	Glu
				230					235					240
Asp	Ser	Ser	Thr	Ser	Phe	Cys	Leu	Asn	Gly	Leu	Trp	Ala	Gln	Gly
				245					250					255
Gln	Arg	Leu	Asp	Val	Asp	Gln	Ala	Leu	Asn	Arg	Ser	His	Glu	Ile
				260					265					270
Trp	Thr	His	Ser	Cys	Pro	Gln	Ser	Pro	Gly	Asn	Gly	Thr	Asp	Ala
				275					280					285
Ser	His													

<210> 26
 <211> 322
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7506575CD1

<400> 26

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

Ala Ser Glu Arg Glu Val Ser
320

<210> 27

<211> 358

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7506576CD1

<400> 27

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

<210> 28

<211> 304
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7512018CD1

<400> 28
 Met Lys Leu Ile Thr Ile Leu Phe Leu Cys Ser Arg Leu Leu Leu
 1 5 10 15
 Ser Leu Thr Gln Glu Ser Gln Ser Glu Glu Ile Asp Cys Asn Asp
 20 25 30
 Lys Asp Leu Phe Lys Ala Val Asp Ala Leu Lys Lys Tyr Asn
 35 40 45
 Ser Gln Asn Gln Ser Asn Asn Gln Phe Val Leu Tyr Arg Ile Thr
 50 55 60
 Glu Ala Thr Lys Thr Val Val Ala Gly Leu Asn Phe Arg Ile Thr
 65 70 75
 Tyr Ser Ile Val Gln Thr Asn Cys Ser Lys Glu Asn Phe Leu Phe
 80 85 90
 Leu Thr Pro Asp Cys Lys Ser Leu Trp Asn Gly Asp Thr Gly Glu
 95 100 105
 Cys Thr Asp Asn Ala Tyr Ile Asp Ile Gln Leu Arg Ile Ala Ser
 110 115 120
 Phe Ser Gln Asn Cys Asp Ile Tyr Pro Gly Lys Asp Phe Val Gln
 125 130 135
 Pro Pro Thr Lys Ile Cys Val Gly Cys Pro Arg Asp Ile Pro Thr
 140 145 150
 Asn Ser Pro Glu Leu Glu Glu Thr Leu Thr His Thr Ile Thr Lys
 155 160 165
 Leu Asn Ala Glu Asn Asn Ala Thr Phe Tyr Phe Lys Ile Asp Asn
 170 175 180
 Val Lys Lys Ala Arg Val Gln Val Val Ala Gly Lys Lys Tyr Phe
 185 190 195
 Ile Asp Phe Val Ala Arg Glu Thr Thr Cys Ser Lys Glu Ser Asn
 200 205 210
 Glu Glu Leu Thr Glu Ser Cys Glu Thr Lys Lys Leu Gly Gln Ser
 215 220 225
 Leu Asp Cys Asn Ala Glu Val Tyr Val Val Pro Trp Glu Lys Lys
 230 235 240
 Ile Tyr Pro Thr Val Asn Cys Gln Pro Leu Gly Met Ile Ser Leu
 245 250 255
 Met Lys Arg Pro Pro Gly Phe Ser Pro Phe Arg Ser Ser Arg Ile
 260 265 270
 Gly Glu Ile Lys Glu Glu Thr Thr Ser His Leu Arg Ser Cys Glu
 275 280 285
 Tyr Lys Gly Arg Pro Pro Lys Ala Gly Ala Glu Pro Ala Ser Glu
 290 295 300
 Arg Glu Val Ser

<210> 29
 <211> 398
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7486368CB1

<400> 29
 gtggaaattg cactggctta tttgtagttt ttttttaagc tacaaaaacta agaacaaatt 60

```

ttagttttatt gaaactaaat gactggcaag agaggaagaa tgagcagttc ctatgccatg 120
aagacctttg ccctgggtcct gctgatggcc ctgctgtgca tgaagagagc tcaggggtctg 180
cactgctaca ggtgcttggc agtctcggaa aggaactcct gccgtgtggg catgggcctc 240
ttccaggagg ggatctgtgt ctcccagaaa gtaagcacct ttcaggagat gaggtcttgc 300
tttgttggca tttatcttgc ttggtgttct ctgagcttcc tggatctgtg gtttgggtgc 360
tgacatgaat ttggggaaat tctcagtcac tattgttt 398

```

<210> 30

<211> 1680

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7488284CB1

<400> 30

```

ggaattcggc tcgagcggct cgaggaacct tgaaggagcc tccaagcctg aaccaaagca 60
ctacagatca cctgctatct tcatagcacc atgagtgaag aatctgatgg aaagcttatt 120
gaagacagcc tgattcagct gagatgtcac ttacatgga agttgttaat tgaagccct 180
gaaattcctg atttagaaaa caggatctgg gaagagattc agttcctgga caccaaatac 240
aatgtgggaa tacacaacct actagcctat gtgaaacacc tgaaaggcca gaatgaggaa 300
gccctgggtca gcttgaaaaa ggctgaagac ttaattcaga aagaacatgc caaccaagca 360
gatattagaa gtctggtgac ctggggcaac tttgcctggg tgtattacca catgggcaga 420
ttggcagaag ccagactta cctggacaag gtggagaaca cttgcaagaa gtttgcaaat 480
ccttcccgtc atagaatgga gtgtccagag gtggactgtg aggaaggatg ggccttggcg 540
aagtgtgggtg gaaagaatta tgaacgggcc aagacctgct ttgaaaaggc tctggaaggg 600
aaccctgaaa accctgaatt caatactggg tacgcaatca ccgtctatcg cctggataaa 660
tttaacacag catcaggagag gaataaggca ttttctctgc acgtcctaaa acgagctgtc 720
aggctaaatc cagatgatgt atatatagg gtctctcctg ccctgaagct tcaggatgaa 780
ggacaggaag ctgaaggaga aaagtacatt gaagaagctc tgaccagtat atcttcacag 840
gcctatgtct ttcaatatgc agccaagttt tatcgaagaa aagggtctgt ggataaagct 900
cttgagctct taaaaatggc cttggagaca acaccactt ctgccttctt gcataccaa 960
atggggcttt gctacagggc acaaatgatc caaatcaagg aagctacaaa ctggcagcct 1020
agagggcaag atagggaaac tgtggacaga ttggttcaat tggctatatg caaatttgaa 1080
aagactataa tggttaaagcg aacatttgag atggcctatg ttgacctggc tgaaacgtat 1140
gcagaaatag gccaccacag aaaggctgag gaacattttc aggaagggtt acgcatgaag 1200
atctttgaag atcagctaaa gcaagagatt cattaccact acggccgttt ccaagaacat 1260
catgggaaat ctcaagataa agcaattacc cattatttaa aaggtttgaa aatagaaaaa 1320
atgtcccatt ccagggaaaa acttctcaat gctttagaga aattggctaa aagatgtatt 1380
caccagaatg tacgggttgt ggaaagtgtc agcctccttg ggcttatcca caaattgaaa 1440
ggagaagtaa gtgatgcttt gctgtgctat gagagggctc tgaggctggc tgctgacctg 1500
aaccctatat tttaacatag aggtcaccat tatccattta atggtcttat aactaaatag 1560
aatgactatg aaattaaata atgcaaactt aaagctgttg gaaatttacc ttatttttgag 1620
ccttgagagg aatgtggctg tgtggcctga gttatgtagc atgcaactgc aacttctgca 1680

```

<210> 31

<211> 789

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500756CB1

<400> 31

```

agagctttat cagggagctg ggactgagtg actgcagcct tcctagatcc cctccactcg 60
gtttctctct ttgcaggagc accggcagca ccagtgtgtg aggggagcag gcagcgggtc 120
tagccagttc cttgatcctg ccagaccacc cagccccggg cacagagctg ctccacaggc 180
accatgagga tcatgctgct attcacagcc atcctggcct tcagcctagc tcagagcttt 240
ggggctgtct gtaaggagcc acaggaggag gtggttcctg gcgggggccc cagcaagagg 300
gattgtctca agccttgagc caggctagca cagatcctaa ggaatcaaca tctcccgaga 360
aacgtgacat gcatgacttc tttgtgggac ttatgggcaa gaggagcgtc cagccagact 420

```

```

ctcctacgga tgtgaatcaa gagaacgtcc ccagctttgg catcctcaag tatccccga 480
gagcagaata ggtactccac ttccggactc ctggactgca ttaggaagac ctctttccct 540
gtcccaatcc ccaggtgccc acgctcctgt taccctttct cttccctgtt cttgtaacat 600
tcttgtgctt tgactccttc tccatctttt ctacctgacc ctggtgtgga aactgcatag 660
tgaatatccc caacccaat gggcattgac tgtagaatac ctagagttc ctgtagtgtc 720
ctacattaaa aatataatgt ctctctctat tcctcaaca taaaggattt ttgcatatga 780
aaaaaaaaa

```

```

<210> 32
<211> 1040
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7506418CB1

```

```

<400> 32
ctttggaagg agagcgcgag cgagggagcg ggcgagctcc gagggggtgt ggggtgtaggy 60
agagagagaa agagagcagg cagcggcgcc ggccggcagcg gtggggaaaa gcggattccg 120
ccccgaacca caccgagggg agctcgtggt cgagacttgc cgccctaagc actctcccaa 180
gtccgacccg ctccggcagg acttccgtct tctgagcgaa ccttgtcaag caagctggga 240
tctatgagtg gaaagggtgac caagcccaaa gaggagaaag atgcttctaa ggagcctcag 300
cttaagggtta tagttaccaa gctatacagc cgacaaggct accacttgca gctgcaggcg 360
gatggaacca ttgatggcac caaagatgag gacagcactt acactctgtt taacctcatc 420
cctgtgggtc tgcgagtggg ggctatccaa ggagttcaaa ccaagctgta cttggcaatg 480
aacagtgagg gatacttgta cacctcggaa cttttcacac ctgagtgcaa attcaaagaa 540
tcagtgtttg aaaattatta tgtgacatat tcatcaatga tataccgtca gcagcagtca 600
ggccgagggg ggtatctggg tctgaacaaa gaaggagaga tcatgaaagg caacctgtg 660
aagaagaaca agcctgcagc tcattttctg cctaaaccac tgaaagtggc catgtacaag 720
gagccatcac tgcacgatct cacggagtct tcccgatctg gaagcgggac cccaaccaag 780
agcagaagtg tctctggcgt gctgaacgga ggcaaatcca tgagccacaa tgaatcaacg 840
tagccagtga gggcaaaaga agggctctgt aacagaacct tacctccaga tttgtggtat 900
tctgggtcaag ttattgtgct gtactttgtg cgtagaaatt gatttgtatt gtcaacccca 960
gtcagtaaag agaacttcaa aaaattatcc tcaagtgtag atttctctta attccatttg 1020
tgtatcatgt taaactattg
1040

```

```

<210> 33
<211> 1087
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7506942CB1

```

```

<400> 33
gatcacccctg cccaacacag actcggcaac tccgcggaag accaggggtcc tgggagtgc 60
tatgggcggt gagagcttgc tctgctcca gttgcggtca tcatgactac gcccgctcc 120
cgcagaccat gttccatgtt tcttttaggt atatctttgg acttctccc ctgatccttg 180
ttctgttgcc agtagcatca tctgattgtg atattgaagg taaagatggc aaacaatatg 240
agagtgttct aatggtcagc atcgatcaat tattggacag catgaaagaa attggttagca 300
attgcctgaa taatgaattt aactttttta aaagacatat ctgtgatgct aataagggaag 360
aaaataaatc tttaaaggaa cagaaaaaac tgaatgactt gtgtttccta aagagactat 420
tacaagagat aaaaacttgt tggataaaaa ttttgatggg cactaaagaa cactgaaaaa 480
tatggagtgg caatatagaa acacgaactt tagctgcac ctccaagaat ctatctgctt 540
atgcagtttt tcagagtggg atgcttccta gaagtactg aatgcaccat ggtcaaaacg 600
gattagggca tttgagaaat gcatattgta ttactagaag atgaatacaa acaatggaaa 660
ctgaatgctc cagtcaacaa actatttctt atatatgtga acatttatca atcagtataa 720
ttctgtactg atttttgtaa gacaatccat gtaaggatc agttgcaata atactttca 780
aacctgttta aatatttcaa gatattaaat ctatgaagta tataatggtt tcaaagattc 840
aaaattgaca ttgctttact gtcaaaataa ttttatggct cactatgaat ctattatact 900
gtattaagag tgaaaattgt cttcttctgt gctggagatg ttttagagtt aacaatgata 960

```

tatggataat	gccggtgaga	ataagagagt	cataaacctt	aagtaagcaa	cagcataaca	1020
aggtccaaga	tacctaaaag	agattttcaag	agattttaatt	aatcatgaat	gtgtaacaca	1080
gtgcctc						1087

<210> 34
 <211> 1270
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7506998CB1

<400> 34

cacgcaaggc	tgctgtctc	tacacattct	cccaagagtt	gtctgagccg	ccgagtggac	60
agtggctgat	tatggagagc	agaggccac	tggctacctc	gcgcctgctg	ctgttgctgc	120
tggtgctact	actgcgtcac	acccgccagg	gatgggccct	gagacctgtt	ctccccaccc	180
agagtgcca	cgacctccg	gctgtccacc	tcagcaatgg	cccaggacaa	gagcctatcg	240
ctgtcatgac	ctttgacctc	accaagatca	caaaaacctc	ctcctccttt	gaggttcgaa	300
cctgggaccc	agagggagtg	atTTTTtTatg	gggataccaa	ccctaaggat	gactggttta	360
tgctgggact	tcgagacggc	aggcctgaga	tccaactgca	caatcactgg	gccagctta	420
cggtgggtgc	tggaccacgg	ctggatgatg	ggagatggca	ccaggtggaa	gtcaagatgg	480
agggggactc	tgtgctgctg	gaggtggatg	gggaggaggt	gctgcgcctg	agacaggctc	540
ctggggccct	gaccagcaaa	cgccatccca	tcatgaggat	tgcgcttggg	gggctgctct	600
tccccgcttc	caaccttcgg	ctgcctgcgc	cgagatctca	gcctctgccc	ccactagcct	660
cagaagctgt	gatgtagaat	caaattcccg	gatatttctc	cctccaggga	ctcaggcaga	720
attcaatctc	cgagacattc	cccagcctca	tgcagagccc	tgggccttct	ctttggacct	780
gggactcaag	caggcagcag	gctcaggcca	cctccttgct	cttggggacac	cagagaaccc	840
atcttggctc	agtctccacc	tccaagatca	aaagtggtg	ttgtcttctg	ggtcggggcc	900
agggctggat	ctgcccctgg	tcttgggact	ccctcttcag	ctgaagctga	gtatgtccag	960
ggtggtcttg	agccaagggt	cgaagatgaa	ggcccttgcc	ctgcctccct	taggcctggc	1020
tccccctctt	aacctctggg	ccaagcctca	agggcgctct	ttcctggggg	ctttaccagg	1080
agaagactct	tccacctctt	tttgctgaa	tggcctttgg	gcacaaggtc	agaggctgga	1140
tgtggaccag	gccctgaaca	gaagccatga	gatctggact	cacagctgcc	cccagagccc	1200
aggcaatggc	actgacgctt	cccattaaag	ctccacctaa	gaaaaaaaaa	aaaaaaaaaa	1260
aaagatcttt						1270

<210> 35
 <211> 1223
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7506999CB1

<400> 35

cacgcaaggc	tgctgtctc	tacacattct	cccaggagtt	gtctgagccg	ccgagtggac	60
agtggctgat	tatggagagc	agaggccac	tggctacctc	gcgcctgctg	ctgttgctgc	120
tggtgctact	actgcgtcac	acccgccagg	gatgggccct	gagacctgtt	ctccccaccc	180
agagtgcca	cgacctccg	gctgtccacc	tcagcaatgg	cccaggacaa	gagcctatcg	240
ctgtcatgac	ctttgacctc	accaagatca	caaaaacctc	ctcctccttt	gaggttcgaa	300
cctgggaccc	agagggagtg	atTTTTtTatg	gggataccaa	ccctaaggat	gactggttta	360
tgctgggact	tcgagacggc	aggcctgaga	tccaactgca	caatcactgg	gccagctta	420
cggtaggtgc	tgcgcccag	acaggtctct	gggcccctga	ccagcaaacg	ccatcccac	480
atgaggattg	cgcttggggg	gctgctcttc	cccgttcca	accttcgggt	gccgctgggt	540
cctgccttgg	atggctgcct	gcgcggggat	tcctggctgg	acaaacaggc	cgagatctca	600
gcctctgccc	ccactagcct	cagaagctgt	gatgtagaat	caaattcccg	gataatttctc	660
cctccaggga	ctcaggcaga	attcaatctc	cgagacattc	cccagcctca	tgcagagccc	720
tgggctcttc	ctttggacct	gggactcaag	caggcagcag	gctcaggcca	cctccttgct	780
cttgggacac	cagagaaccc	atcttggctc	agtctccacc	tccaagatca	aaaggtgggtg	840
ttgtcttctg	ggtcggggcc	agggctggat	ctgcccctgg	tcttgggact	ccctcttcag	900
ctgaagctga	gtatgtccag	ggtggtcttg	agccaagggt	cgaagatgaa	ggcccttgcc	960

```

ctgcctccct taggcctggc tccccctcctt aacctctggg ccaagcctca agggcgtctc 1020
ttcctggggg ctttaccagg agaagactct tccacctctt tttgcctgaa tggccttttg 1080
gcacaagggtc agaggctgga tgtggaccag gccctgaaca gaagccatga gatctggact 1140
cacagctgcc cccagagccc aggcaatggc actgacgctt cccattaaag ctccacctaa 1200
gaacccccctt tgaaagttaa aaa 1223

```

```

<210> 36
<211> 928
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7509150CB1

```

```

<220>
<221> unsure
<222> 903, 915
<223> a, t, c, g, or other

```

```

<400> 36
gcagcccagc cccacccaga cccatggctg gacctgccac ccagagcccc atgaagctga 60
tggccctgca gctgctgctg tggcacagt gactctggac agtgcaggaa gccaccccc 120
tgggccctgc cagctccctg cccagagct tctgctcaa gtgcttagag caagtgagga 180
agatccaggg cgatggcgca gcgctccagg agaagctggc tggctgcttg agccaactcc 240
atagcggcct tttcctctac caggggctcc tgcaggccct ggaatggatc tccctccgag 300
ttggtgtccc agcttggcac aactgcagc tggacgtcgc cgactttgcc accaccatct 360
ggcagcagat ggaagaactg ggaatggccc ctgccctgca gccacccag ggtgccatgc 420
cggccttcgc ctctgcttcc cagcgccggg caggaggggt cctgggttgc tcccatctgc 480
agagcttcct ggaggtgtcg taccgcgttc tacgccacct tgcccagccc tgagccaagc 540
cctccccatc ccatgtatct atctctatct aatatttatg tctattttaag cctcatactt 600
aaagacaggg aagagcagaa cggagcccca ggctctgtg tcttccctg catttctgag 660
tttcattctc ctgcctgtag cagtgcagaa aagctcctgt cctcccatcc cctggactgg 720
gaggtagata ggtaaatacc aagtatttat tactatgact gctccccagc cctggctctg 780
caatgggcac tgggatgagc cgctgtgagc cctgggtcct gaggggtccc acctgggacc 840
cttgagagta tcaggtctcc cagctgggag acaagaaatc cctgtttaat atttaaacag 900
cantgttccc cattnggggtc cttgcaac 928

```

```

<210> 37
<211> 2497
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 1576743CB1

```

```

<400> 37
cgaactgttt caagtaacag atgttccagg cagagccagc cagagtgagc tgttccttct 60
ctggagggtg atctggtatc cctgaacgcc tgttggcctc atctccacca accctgcag 120
tctctgcccc tgagtcccc tcttccatc cgcctccctt tactagagcc tcagccctcc 180
ctcctcgccct ggaagccttg ccccgcgcc cttgtgctgg ctggagctca agcctcgtgt 240
cctttgtcgc agctccgccc agttgaacac accgctggg gaagggtgct ctgttccctc 300
cccacgcact ctggggcctg agctgacaga gatggaccat cgaaaagcca gggtcctccc 360
agctgggcac tactgcccct cgctaggaat atgggcctcg caggtcggca gcgtgaggtc 420
ctctgtgcca ccttccatca ggtagctgtc accgaggagc atgttgagc gccgggtggg 480
ggctgccttg catgcaagga gcctggcagc agcggagggc aaggctttga gtgaggcggc 540
ccggacagcc atagctgagg agcatggagc cactgggagg gggcagtgct accttttttg 600
gccttcttcc tgtgtggaat tacagcgctc cggccttgaa cctgccactc aggtgtcttg 660
atgtgtcggg ggtgtggact gcctgcccc tgatgtctcc tgccccactc tctgcagtct 720
acgctcacc tggagggggt gcttccctgg cttggggccc cgctgggggt gaccggcaag 780
actcgtggac cactgtgaac ggtaccttca gtgtgtatct ccgccacaac tcgtccagcc 840
tgggcaacct cagtgtcagc atcgtgccgc cctccaagcg tgtcgagtcc ggaggagtct 900

```

```

ggctgcccgg gacctgtcccc caccctctgc agtctacgct cgccctggag ggggtgtctc 960
ctgggctggg gccccgcctg gggatggcag cagcagcggc ggggcccggg cttgggggct 1020
ccctcggggg cgcactggcg gggcgcttg gggcgcgctt gggagtgcct ggggccaaag 1080
agtcacgcgc tttcaattgc cagtgaggat atgagaagac aaaccgcgcg cgcaagcacc 1140
gaccgtgcct gtacgacccg tcgcagggtg gcttcaccga gcacacgcag agccaggccg 1200
cctggctctg tgccaagccc ttcaaagtca tctgtatctt cgtctctttc ctcagctttg 1260
actacaaact ggtgcagaag gtgtgcccag actataactt ccagagttag caccctact 1320
tcggatagcg cccctcccca gccagtcttg agcctcccgc caaatcccag cctcactagg 1380
tgggaccccc ttcccagtggt tctgccgctc ctgtggccat gtcgccact ccttccactc 1440
tgggggcgga ggggaatggc ttctcgggac cctcagctag cgtgggtgcc cttttcctta 1500
tgcgaggatgc ccgcaaggct ggggtagccc cctccagtac accccaaagt gaaagggata 1560
agagtgcagc cccagaatag gcggggcttg gaggcgggtc caatgtcccc tgggtccaca 1620
gtgggtcccc ttttcaccct tggcgctagg ctgcgcactc cttttcccg cagctttaat 1680
aactcctggc ctggcaccct caccacccc tgactttccc atcccccagc gcttgtcctg 1740
cttcacccta cccgcctaa gactgtaaag gcctaaaaac ctggcctgt cctcccacca 1800
ttctgcctgc catatgcctg tccccttttc ctccaaaccc tattagggta ccggaagcag 1860
aacccttggg ctgaggccct ggccctgccc ccggcccctg cccctgcccg cccccctcca 1920
gtccaggcag tcgagctcca cctgccctct cctgtgctt cctctcgggtg atattttttc 1980
tacgccaaaa cagacgggaa agggaacaaa ataaagttaa atccaatata aaaaacacaa 2040
acaatccaaa cagcataaat aacaccagca ccaaaaaaaa ccaaaacaac acacacccac 2100
acctcaccct cccccccca ctaccacaa agggcccata ggctgcctgc acccagccc 2160
ccccaccctc ccgcgatgcc ccgcctcca cgtccacac ccaccaacaa cataaaacaa 2220
ctaaccctac cactcaacca cccataaccc cacactcaat cacaccaaca accaactctc 2280
ccacaccccc accacccac cgctacgcta ccacccccca cccgaaccga aaaacaaaga 2340
aaaacagcac accggcaaca cacaccgcaa cacatcaacc ccaaccaca ccaccgcaac 2400
cacaaaacac accacatcat ccagaccccc cgcacacgac caacaccaca gccaccaag 2460
ccatcccacc acccacacc ccaccacccc acacaca 2497

```

<210> 38

<211> 2573

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503892CB1

<400> 38

```

tggaacaaat tccccatcac aggaaccccg aaatcagaaa agttaagtca cccagggctg 60
gaccagact cttgcagctc tcactttcac aatgcccttg ggctgactag gctgcagagg 120
ggtttcaccc ccaaccccag ggcacctcaa gtgtcccac caaaccttcc taacacctgt 180
ccactaagct gtactaggcc ctgcaactg acctatggga cctgaggcct ggcccctcat 240
ggctcctgtc accaggcttc aggtcagggt ccagcaggcc ctgagctgac gtgtggagcc 300
agagccaccc aatcccgtag ggacagggtt cacaacttcc cggatggggc tgtgggtgggt 360
cacagtgcag cctccagcca gaaggatggg gtgggtccca ctctgctgc ttctgactca 420
atgcttaggg gtccctgggc agcgctcgcc attgaatgac ttccaagtgc tccggggcac 480
agagctacag cacctgctac atgcggtggg gcccgggcct tggcaggagg atgtggcaga 540
tgctgaagag tgtgctgggc gctgtgggcc cttaatggac tgccgggcct tccactacaa 600
cgtgagcagc catggttgcc aactgctgcc atggactcaa cactcgcccc acacgaggct 660
gcggcgcttc gggcgctgtg acctcttcca gaagaaagac tacgtacgga cctgcatcat 720
gaacaatggg gttgggtacc ggggcacccat ggccaacgac gtgggtggcc tgccctgcc 780
ggcttggagc cacaagtccc cgaatgatca caagtacacg cccactctcc ggaatggcct 840
ggaagagaac ttctgccgta accctgatgg cgaccccgga ggtccttggg gctacacaa 900
agaccctgct gtgcgcttcc agagctgcgg catcaaatcc tgccgggagg ccgcgtgtgt 960
ctggtgcaat ggcgaggaat accgcggcgc ggtagaccgc acggagttag ggcgcgagt 1020
ccagcgctgg gatcttcagc accgcacca gcaaccttcc gagccgggca agttcctcga 1080
ccaaggctct gacgacaact attgccggaa tctgacggc tccgagcggc catggtgcta 1140
cactacggat ccgcagatcg agcgagagtt ctgtgacctc ccccgctgcg ggtccgaggc 1200
acagccccgc caagaggcca caactgtcag ctgcttccgc ggggaaggggtg agggctaccg 1260
gggcacagcc aataaccacca ctgcgggcgt accttgccag cgttgggacg cgcaaatccc 1320
tcatcagcac cgatttacgc cagaaaataa cgcgtgcaaa gaccttcggg agaacttctg 1380
ccggaacccc gacggctcag aggcgccttg gtgcttcaca ctgcggcccg gcatgcgcgc 1440
ggccttttgc taccagatcc ggcgttgtac agacgacgtg cggccccagg actgctacca 1500

```

```

cggcgccaggg gagcagtacc gcggcacggt cagcaagacc cgcaaggggtg tccagtgcc 1560
gcgctgggtcc gctgagacgc cgcacaagcc gcagagaact tctgccggaa cccagatggg 1620
gatagccatg ggccttggtg ctacacgatg gacccaagga cccattcga ctactgtgcc 1680
ctgcgacgct gcgctgatga ccagccgcca tcaatcctgg acccccaga ccaggtgcag 1740
tttgagaagt gtggcaagag ggtggatcgg ctggatcagc ggcgttccaa gctgcgcgtg 1800
gttgggggcc atccgggcaa ctcaccctgg acagtcagct tgcggaatcg gcagggccag 1860
catttctgcg gggggtctct agtgaaggag cagtggatac tgactgcccg gcagtgcctc 1920
tcctcctgcc atatgcctct cacgggctat gaggtatggg tgggcaccct gttccagaac 1980
ccacagcatg gagagccaag cctacagcgg gtcccagtag ccaagatggg gtgtgggccc 2040
tcaggctccc agcttgctct gctcaagctg gagagatctg tgaccctgaa ccagcgtgtg 2100
gcctgatct gctgcccccc tgaatggtat gtggtgcctc cagggaccaa gtgtgagatt 2160
gcaggctggg gtgagaccaa aggtacgggt aatgacacag tcctaaatgt ggccttgctg 2220
aatgtcatct ccaaccagga gtgtaacatc aagcaccgag gacgtgtgcg ggagagtgcg 2280
atgtgcactg agggactgtt ggcccctgtg ggggcctgtg aggtgacta cgggggccca 2340
cttgctgctt ttaccacaaa ctgctgggtc ctggaaggaa ttataatccc caaccgagta 2400
tgcgcaagggt ccgctgggcc agctgtcttc acgctgtctc ctgtgtttgt ggactggatt 2460
cacaagggtca tgagactggg ttagggccag ccttgatgcc atatgccttg gggaggacaa 2520
aacttcttgt cagacataaa gccatgtttc ctcttataaa aaaaaaaaaa aaa 2573

```

<210> 39

<211> 2440

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503895CB1

<400> 39

```

tggcaaaaat tccccatcac aggaaacccc aaatcagaaa agttaagtca cccagggctg 60
gacccagact cttgcagctc tcactttcac aatgcccttg ggctgactag gctgcagagg 120
ggtttcaccc ccaaccccag ggcacctcaa gtgtccccac caaaccttcc taacacctgt 180
ccactaagct gtactaggcc ctgcaactg acctatggga cctgaggcct ggcccctcat 240
ggctcctgtc accagggtctc aggtcagggt ccagcaggcc ctgagctgac gtgtggagcc 300
agagccaccc aatcccgtag ggacagggtt cacaacttcc cggatggggc tgtggtgggt 360
cacagtgcag cctccagcca gaaggatggg gtggctccca ctctgctgc ttctgactca 420
atgcttaggg gtccctgggc agcgctcgcc attgaatgac ttccaagtgc tccggggcac 480
agagctacag cacctgctac atgcggtggg gcccgccct tggcaggagg atgtggcaga 540
tgctgaagag tgtgctgggc gctgtgggcc cttaatggac tgccgggcct tccactacaa 600
cgtgagcagc catggttgcc aactgctgcc atggactcaa cactcgcccc acacgaggct 660
gcggcgcttc gggcgctgtg acctcttcca gaagaaagac tacgtacgga cctgcatcat 720
gaacaatggg gttgggtacc ggggcacccat ggccacgacc gtgggtggcc tgccctgcca 780
ggcttgagc cacaagttcc cgaatgatca caagtacacg cccactctcc ggaatggcct 840
ggaagagaac ttctgcgta accctgatgg cgaccccgga ggtccttggg gctacacaac 900
agaccctgct gtgcgcttcc agagctgcgg catcaaatcc tgccgggagg ccgctgtgtg 960
ctggtgcaat ggcgaggaat accgcggcgc ggtagaccgc acggagtacg ggcgcgagtg 1020
ccagcgtgg gatcttcagc accgcaccca gcaccccttc gagccgggca agttcctcga 1080
ccaaggctct gacgacaact attgcccga tctgacggc tccgagcggc catggtgcta 1140
cactacggat ccgcagatcg agcgagagtt ctgtgacctc ccccgctgag agaccttcgg 1200
gagaacttct gccggaaccc cgacggctca gaggcgccct ggtgcttcac actgcccgcc 1260
ggcatgcgcg cgcccttttg ctaccagatc cggcgcttgta cagacgacgt gcggccccag 1320
gactgctacc acggcgaggg ggagcagtag cggcgacagg tcagcaagac ccgcaagggg 1380
gtccagtgcc agcgtgggtc cgctgagacg ccgcacaagc cgcagttcac gtttacctcc 1440
gaaccgcatg cacaactgga ggagaacttc tgccgggaacc cagatgggga tagccatggg 1500
ccctgggtgt acacgatgga cccaaggacc ccattcgact actgtgccct gcgacgctgc 1560
gctgatgacc agccgccatc aatcctggac ccccagacc aggtgcagtt tgagaagtgt 1620
ggcaagaggg tggatcggct ggatcagcgg cgttccaagc tgcgcgtggg tgggggccat 1680
ccgggcaact caccctggac agtcagcttg cggaaatcggc agggccagca tttctgcggg 1740
gggtctctag tgaaggagca gtggatactg actgcccggc agtgcttctc ctctgccc 1800
atgcctctca cgggctatga ggtatgggtg ggcacctgt tccagaaccc acagcatgga 1860
gagccaagcc tacagcgggt cccagtagcc aagatgggtg gtgggccctc agggctccag 1920
cttgctcctg tcaagctgga gagatctgtg accctgaacc agcgtgtggc cctgatctgc 1980
ctgccccctg aatggtatgt ggtgcctcca gggaccaagt gtgagattgc agggctgggg 2040

```

```

gagaccaaag gtacgggtaa tgacacagtc ctaaagtgtgg ccttgctgaa tgtcatctcc 2100
aaccaggagt gtaacatcaa gcaccgagga cgtgtgcggg agagtgagat gtgcactgag 2160
ggactgttgg cccctgtggg ggcctgtgag ggtgactacg ggggccact tgcctgcttt 2220
accacaact gctgggtcct ggaaggaatt ataatcccca accgagtatg cgcaagggtcc 2280
cgctggccag ctgtcttcac gcgtgtctct gtgtttgtgg actggattca caagggtcatg 2340
agactgggtt agggccagcc ttgatgccat atgccttggg gaggacaaaa cttcttgtca 2400
gacataaagc catgtttcct cttttaaaaa aaaaaaaaaa 2440

```

<210> 40

<211> 858

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509192CB1

<400> 40

```

gcatctaga actagttagc ttctcctgtg aaagagactt ccagcttcct cctcctcctc 60
ttcctcctcc tctcctgcc ccagcgagcc ttctgctgag ctacaccaat gggaatccca 120
atggggaagt cgatgctggt gcttctcacc ttcttggcct tcgcctcgtg ctgcattgct 180
gcttacggcc ccagtggagc cctgtgcggc ggggagctgg tggacaccct ccagttcgtc 240
tgtggggacc gcggcttcta cttcagcagg ccgcaagcc gtgtgagccg tcgcagccgt 300
ggcatcgttc agggaggcca aacgtcaccg tccccgtgatt gctctaccca cccaagaccc 360
cgccacggg ggcgcccccc cagagatggc cagcaatcgg aagtgagcaa aactgccgca 420
agtctgcagc ccggcgccac catcctgcag cctcctcctg accacggacg tttccatcag 480
gttccatccc gaaaatctct cggttccacg tccccctggg gcttctcctg acccagtcctc 540
cgtgccccgc ctccccgaaa caggctactc tctcggccc cctccatcgg gctgaggaag 600
cacagacgca tcttcaaac tgtacaaaa cgattggctt taaacaccct tcacataccc 660
tcccccaaaa ttatccccaa ttatccccac acataaaaaa tcaaaacatt aaactaaccc 720
ccttcccccc ccacaacaa cctctttaa actaattggc tttttagaaa cccccacaa 780
aagctcagaa attggcttta aaaaaacaa ccaccaaaaa aaatcaattg gctaaaaaaa 840
aaaaaaaaaa agatcggc 858

```

<210> 41

<211> 1172

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509320CB1

<400> 41

```

gcggcggggc cgtcccggaa cggctgcggc cgggcacccc gggagttaat ccgaaagcgc 60
cgcaagcccc gcgggcccgc cgcaccgcac gtgtcaccga gaagctgatg tagagagaga 120
cacagaagga gacagaaagc aagagaccag agtcccggga aagtccctgc gcgcctcggg 180
acaattataa aaatgtggcc ccctgggtca gcctcccagc caccgcctc acctgccgcg 240
gccacaggtc tgcattccagc ggctcgccct gtgtccctgc agtgccggct cagcatgtgt 300
ccagcgcgca gcctcctcct tgtggctacc ctggctcctc tggaccacct cagtttggcc 360
agaaacctcc ccgtggccac tccagaccca ggaatgttcc catgccttca ccaactccaa 420
aacctgctga gggccgtcag caacatgctc cagaaggcca gacaaactct agaattttac 480
ccttgcaact ctgaagagat tgatcatgaa gatatacaaa aagataaaac cagcacagt 540
gaggcctgtt taccattgga attaaccaag aatgggagtt gcctggcctc cagaaagacc 600
tcttttatga tggccctgtg ccttagtagt atttatgaag acttgaagat gtaccagggt 660
gagttcaaga ccatgaatgc aaagcttctg atggatccta agaggcagat ctttctagat 720
caaaacatgc tggcagttat tgatgagctg atgcaggccc tgaatttcaa cagtggagact 780
gtgccacaaa aatcctccct tgaagaaccg gatttttata aaactaaaat caagctctgc 840
atacttcttc atgctttcag aattcgggca gtgactattg atagagtgat gagctatctg 900
aatgcttctc aaaaagcgag gtccctccaa accgttgtca tttttataaa actttgaaat 960
gaggaaactt tgataggatg tggattaaga actagggagg gggaaagaag gatgggacta 1020
ttacatccac atgatacctc tgatcaagta tttttgacat ttactgtgga taaattgttt 1080
ttaagttttc atgaatgaat tgctaagaag ggaaaatatc catcctgaag gtgtttttca 1140

```

ttcacttttaa tagaagggca aatatttata ag

1172

<210> 42

<211> 2769

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503898CB1

<400> 42

```

tggcaaaaat tccccatcac aggaaacccg aaatcagaaa agttaagtca cccagggctg 60
gaccagact cttgcagctc tcactttcac aatgcccttg ggctgactag gctgcagagg 120
ggtttcaccc ccaaccccag ggcacctcaa gtgtccccac caaaccttcc taacacctgt 180
ccactaagct gtactaggcc cttgcaactg acctatggga cctgaggcct ggccccctcat 240
ggctcctgtc accaggtctc aggtcagggg ccagcaggcc ctgagctgac gtgtggagcc 300
agagccaccc aatcccgtag ggacaggttt cacaacttcc cggatggggc tgtggtgggt 360
cacagtgcag cctccagcca gaaggatggg gtggctccca ctctgctgac ttctgactca 420
atgcttaggg gtccctgggc agcgctcgcc attgaatgac ttccaagtgc tccggggcac 480
agagctacag cacctgctac atgcggtggg gcccgggcct tggcaggagg atgtggcaga 540
tgtctaagag tgtgtgggct gctgtgggac cttaatggac tgccgggcct tccactacaa 600
cgtgagcagc catggttgcc aactgctgcc atggactcaa cactcgcccc acacgaggct 660
ggggcgcttct gggcgctgtg acctcttcca gaagaaagac tacgtacgga cctgcatcat 720
gaacaatggg gttgggtacc ggggcaccat ggccacgacc gtgggtggcc tgccctgcca 780
ggcttgagac cacaagttcc cgaatgatca caagtacacg cccactctcc ggaatggcct 840
ggaagagaac ttctgcccga acctgatgg cgaccccgga ggtccttggg gctacacaac 900
agaccctgct gtgcgcttcc agagctgcgg catcaaatcc tgccgggagg ccgcgtgtgt 960
ctggtgcaat ggcgaggaat accgcggcgc ggtagaccgc acggagtcag ggcgcgagtg 1020
ccagcgctgg gatcttcagc accgcacca gcaccttc gagccgggca agttcctcga 1080
ccaaggtctg gacgacaact attgccgga tcttgacggc tccgagcggc catggtgcta 1140
cactacggat ccgcagatcg agcgagagtt ctgtgacctc ccccgctgcg ggtccgaggg 1200
acagccccgc caagaggcca caactgtcag ctgcttccgc ggggaagggtg agggctaccg 1260
gggcacagcc aataccacca ctgcgggcgt acctgcccag cgttgggacg cgcaaatccc 1320
tcatcagcac cgatttacgc cagaaaaata cgcgtgcaaa gaccttcggg agaacttctg 1380
ccggaacccc gacggctcag aggcgccctg gtgcttcaca ctgcggcccg gcatgcgcgc 1440
ggccttttgc taccagatcc ggcgttgtac agacgacgtg cggccccagg actgctacca 1500
cggcgctggg gagcagtacc gcggcacggt cagcaagacc cgcaagggtg tccagtacca 1560
gcgctggtcc gctgagacgc cgcacaagcc gcagttcacg tttacctccg aaccgcatgc 1620
acaactggag gagaacttct gccggaaccc agatggggat agccatgggc cctggtgcta 1680
cacgatggac ccaaggaccc cattcgacta ctgtgccctg cgacgctgcg ctgatgacca 1740
gcgcgatcca atcctggacc ccccagacca ggtgcagttt gagaagtgtg gcaagagggg 1800
ggatcggtct gatcagcggc gttccaagct gcgcgtgggt gggggccatc cgggcaactc 1860
accttggaac gtcagcttgc ggaatcggca gggccagcat ttctgcgggg ggtctctagt 1920
gaaggagcag tggatactga ctgcccggca gtgcttctcc tcttgccata tgcctctcac 1980
gggctatgag gtatggttgg gcacctgtt ccagaaacca cagcatggag agccaagcct 2040
acagcgggtc ccagtagcca agatggtgtg tgggccctca ggctccagc ttgtcctgct 2100
caagctggag agatctgtga ccctgaacca gcgtgtggcc ctgatctgcc tgccccctga 2160
atggtatgtg gtgcctccag ggaccaagt tgagattgca ggctgggggt agaccaaaagg 2220
taagagcaca gtgcacagga ctgctggtgg ccaggaggcc cagccctgga tcttcctgca 2280
ggacctctc cctctcccca ttccccctac tgcaggtacg ggtaatgaca cagtccataa 2340
tgtggccttg ctgaatgtca tctccaacca ggagtgtaac atcaagcacc gaggacgtgt 2400
gcgggagagt gagatgtgca ctgagggact gttggccct gtgggggcct gtgaggggtga 2460
ctacgggggg ccacttgctt gctttaccca caactgctgg gtccctggaag gaattataat 2520
ccccaacgga gtatgcgcaa ggtcccgtg gccagctgtc ttcacgcgtg tctctgtgtt 2580
tgtggactgg attcacaagg tcatgagact ggggttaggc cagccttgat gccatatgcc 2640
ttggggagga caaaacttct tgtcagacat aaagccatgt ttcctcttta tgcctgtaaa 2700
aactcattaa taaaaaaaaa agatctttaa ttaagcgggt caagacttat tccatataat 2760
gagtgttaa 2769

```

<210> 43

<211> 1289

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509185CB1

<400> 43

```

gcggcgggggc cgtcccggaa cggctgcggc cgggcacccc gggagttaat ccgaaagcgc 60
cgcaagcccc gcgggcccggc cgcaccgcac gtgtcaccga gaagctgatg tagagagaga 120
cacagaagga gacagaaagc aagagaccag agtcccggga aagtccctgcc gcgcctcggg 180
acaattataa aaatgtggcc ccttgggtca gcctcccagc caccgcccctc acctgccgcg 240
gccacagggtc tgcattccagc ggctcgcctt gtgtccctgc agtgccgggt cagcatgtgt 300
ccagcgcgca gctcctctct tgtggctacc ctggtcctcc tggaccacct cagtttgggc 360
agaaacctcc ccgtggccac tccagaccca ggaatgttcc catgccttca ccactcccaa 420
aacctgctga gggccgtcag caacatgctc cagaaggcca gacaaactct agaattttac 480
ccttgcactt ctgaagagat tgatcatgaa gatatacaaa aagataaaac cagcacagt 540
gaggcctgtt taccattgga attaaccaag aatgagagtt gcctaaattc cagagagacc 600
tctttcataa ctgtaagtca aaaaatgaaa agtttcagcc tgtatgatga attcatatca 660
ctgatgtctg attatttttt cctctagaat gggagttgcc tggcctccag aaagacctct 720
tttatgatgg cctgtgcct tagtagtatt tatgaagact tgaagatgta ccaggtggag 780
ttcaagacca tgaatgcaaa gcttctgatg gatcctaaga ggcagatctt tctagatcaa 840
aacatgctgg cagttattga tgagctgatg caggccctga atttcaacag tgagactgtg 900
ccacaaaaat cctcccttga agaaccggat ttttataaaa ctaaaatcaa gctctgcata 960
cttcttcatg ctttcagaat tcgggcagtg actattgata gagtgatgag ctatctgaat 1020
gcttcctaaa aagcgagggtc cctccaaacc gttgtcattt ttataaaact ttgaaatgag 1080
gaaactttga taggatgtgg attaagaact agggaggggg aaagaaggat gggactatta 1140
catccacatg atacctctga tcaagtattt ttgacattta ctgtggataa attgttttta 1200
agttttcatg aatgaattgc taagaaggga aaatatccat cctgaagggtg tttttcattc 1260
actttaatag aagggcaaat atttataag 1289

```

<210> 44

<211> 1219

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509401CB1

<400> 44

```

gatcaccctg cccaacacag actcggcaac tccgcggaag accaggggtcc tgggagtgc 60
tatgggcccgt gagagcttgc tctgtctcca gttgcgggtc tcatgactac gccgcctcc 120
cgcagaccat gttccatgtt tcttttaggt atatctttgg acttcctccc ctgatccttg 180
ttctgttgcc agtagcatca tctgattgtg atattgaagg taaagatggc aaacaatatg 240
agagtgttct aatggtcagc atcgatcaat tattggacag catgaaagaa attggtagca 300
attgcctgaa taatgaattt aactttttta aaagacatat ctgtgatgct aataaggaag 360
gtatgttttt attccgtgct gctcgcaagt tgaggcaatt tcttaaaatg aatagcactg 420
gtgattttga tctccactta ttaaaagttt cagaaggcac aacaatactg ttgaactgca 480
ctggccagga agaaaaataa tctttaaagg aacagaaaaa actgaatgac ttgtgtttcc 540
taaagagact attacaagag ataaaaactt gttggaataa aattttgatg ggcactaaag 600
aacactgaaa aatatggagt ggcaatatag aaacacgaac tttagctgca tcctccaaga 660
atctatctgc ttatgcagtt tttcagagtg gaatgcttcc tagaagttac tgaatgcacc 720
atggtcaaaa cggattaggg catttgagaa atgcatattg tattactaga agatgaatac 780
aaacaatgga aactgaatgc tccagtcaac aaactatttc ttatatatgt gaacatttat 840
caatcagtat aattctgtac tgatttttgt aagacaatcc atgtaaggta tcagttgcaa 900
taatacttct caaacctgtt taaatatttc aagacattaa atctatgaag tatataatgg 960
tttcaaagat tcaaaattga cattgcttta ctgtcaaaat aattttatgg ctactatga 1020
atctattata ctgtattaag agtgaaaatt gtcttcttct gtgctggaga tgttttagag 1080
ttaacaatga tatatggata atgccggtga gaataagaga gtcataaacc ttaagtaagc 1140
aacagcataa caagggtcaa gatacctaaa agagatttca agagatttaa ttaatcatga 1200
atgtgtaaca cagtgcctc 1219

```

<210> 45

<211> 1058
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7510191CB1

<400> 45
gggataaata cgcgaaggca ctgagcaggc gaaagagcgc gctcggacct ccttcccggc 60
ggcagctacc gagagtgcgg agcgaccagc gtgcgctcgg aggaaccaga gaaactcagc 120
accccgcggg actgtccgtc gcaaaatcca acatgaaaat cctcgtggcc ttggcagtct 180
tttttcttgt ctccactcag ctgtttgcag aagaaatagg agccaatgat gatctgaatt 240
actgggtccga ctgggtacgac agcgaccaga tcaaggagga actgccggag ccctttgagc 300
atcttctgca gagaatcgcc cggagaccca agcctcagca gttctttgga ttaatgggca 360
aacgggatgc tggacatggc cagatctctc acaaaagaca taaaacagat tcctttgttg 420
gactaatggg caaaagagct ttaaattctg tggcttatga aaggagtgc atgcagaatt 480
atgaaagaag acgttaataa actacctaac attattttatt cagcttcatt tgtgtcaatg 540
ggcaatgaca ggtaaatata gacatgcact atgaggaata attattttatt taataacaat 600
tgtttggggg tgaaaattca aaaagtgttt atttttcata ttgtgccaat atgtattgta 660
aacatgtgtt ttaattccaa tatgatgact cccttaaaat agaaataagt ggttatttct 720
caacaaagca cagtgttaaa tgaaattgta aaacctgtca atgatacagt ccctaaagaa 780
aaaaaatcat tgctttgaag cagttgtgtc agctactgcg gaaaaggaag gaaactcctg 840
acagtcttgt gcttttccta tttgttttca tgggtgaaaat gtactgagat tttgggtatta 900
cactgtattt gtatctctga agcatgtttc atgttttgtg actatataga gatgttttta 960
aaagtttcaa tgtgattcta atgtcttcat ttcattgtat gatgtgttgt gatagctaac 1020
attttaaata aaagaaaaaa tatcttgata aaaaaaaa 1058

<210> 46
<211> 1217
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7506943CB1

<400> 46
gatcacccctg cccaacacag actcggcaac tccgcggaag accaggggtcc tgggagtgc 60
tatgggcggt gagagcttgc tcctgctcca gttgcggtca tcatgactac gcccgctcc 120
cgcagaccat gttccatgtt tcttttaggt atatcttttg acttcctccc ctgatccttg 180
ttctgttgcc agtagcatca tctgattgtg atattgaagg acagcatgaa agaaatttgt 240
agcaattgcc tgaataatga atttaacttt tttaaaagac atatctgtga tgctaataag 300
gaaggatatg ttttattccg tgctgctcgc aagttgaggg aatttcttaa aatgaatagc 360
actgggtgatt ttgatctcca cttattaaaa gtttcagaag gcacaacaat actgttgaac 420
tgcaactggcc aggttaaaagg aagaaaacca gctgccctgg gtgaagccca accaacaag 480
agtttggaag aaaataaatc tttaaaggaa cagaaaaaac tgaatgactt gtgtttccta 540
aagagactat tacaagagat aaaaacttgt tggataaaaa ttttgatggg cactaaagaa 600
cactgaaaaa tatggagtgg caatatagaa acacgaactt tagctgcatc ctccaagaat 660
ctatctgctt atgcagtttt tcagagtggg atgcttccta gaagttactg aatgcacat 720
ggtcaaaacg gattagggca tttgagaaat gcataattgta ttactagaag atgaatacaa 780
acaatggaaa ctgaatgctc cagtcaacaa actatttctt atatatgtga acatttatca 840
atcagtataa ttctgtactg atttttgtaa gacaatccat gtaagggtatc agttgcaata 900
atacttctca aacctgttta aatatttcaa gacattaaat ctatgaagta tataatgggt 960
tcaaagattc aaaattgaca ttgctttact gtcaaaataa ttttatggct cactatgaat 1020
ctattatact gtattaagag tgaaaattgt cttcttctgt gctggagatg ttttagagtt 1080
aacaatgata tatggataat gccggtgaga ataagagagt cataaacctt aagtaagcaa 1140
cagcataaca aggtccaaga tacctaaaag agatttcaag agattttaatt aatcatgaat 1200
gtgtaacaca gtgcctc 1217

<210> 47
<211> 785
<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500179CB1

<400> 47

```
cggctcatca aggggggtgtt gcacaaatca acggacttcc aaaatgtcgt aacacttccg 60
ccccagtaac gcaaattggcg gtaggcgtgt acgctgggag gtctatataa gcagagctgg 120
tttagtgaac cgtcagatcc gctagccgca attactgtga gttagctcac tcattaaggc 180
accccagggt ttcggctcga gccagcagag ccatttaacc tttgggaaca aggcaactag 240
cgtctggcag caggaatcca accagtcccc tgagttctga ggcagagagg aggacagaag 300
aaacaagagg ctggagattg tcaaattcag tatcccagtt ggctcttgat tcttggtgaa 360
accatccctc agtccttaga gggagattgt tagatcatga aactaattac catccttttc 420
ctctgctcca ggctgctact aagtttaacc caggaattcac agtccgagga aattgactgc 480
aatgacaagg atttatttaa agctgtggat gctgctctga agaaatataa cagtcaaaac 540
caaagtaaca accagtttgt attgtaccgc ataactgaag ccactaagac ggttggctct 600
gacacgtttt attccttcaa gtacgaaatc aaggaggggg attgtcctgt tcaaagtggc 660
aaaaccttga gaggaaggac aagaagaaag atgggataga atttaaatag agaagaatgc 720
cattttatca ctctgcctct gggtgaaata aagatcagtc ttgatgttct aaaaaaaaaa 780
aaaaa 785
```

<210> 48

<211> 1084

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509149CB1

<400> 48

```
gatcaccctg cccaacacag actcgggaac tccgcggaag accaggggtcc tgggagtgac 60
tatggggcgt gagagcttgc tcctgctcca gttgcggtca tcatgactac gcccgctcc 120
cgcagaccat gttccatgtt tcttttaggt atatctttgg acttcctccc ctgacccctg 180
ttctgttgcc agtagcatca tctgattgtg atattgaagg acagcatgaa agaaattggt 240
agcaattgcc tgaataatga atttaacttt tttaaaagac atatctgtga tgctaataag 300
gttaaaggaa gaaaaccagc tgcctgggtg aagcccaacc aacaaagagt ttggaagaaa 360
ataaatcttt aaaggaacag aaaaaactga atgacttgtg tttcctaaag agactattac 420
aagagataaa aacttggttg aataaaattt tgatgggcac taaagaacac tgaaaaatat 480
ggagtggcaa tatagaaaca cgaactttag ctgcatcctc caagaatcta tctgcttatg 540
cagtttttca gagtggaatg cttcctagaa gttactgaat gcaccatggg caaaacggat 600
tagggcattt gagaaatgca tattgtatta ctagaagatg aatacaaaac atggaaactg 660
aatgctccag tcaacaaact atttcttata tatgtgaaca tttatcaatc agtataattc 720
tgtactgatt tttgtaagac aatccatgta aggtatcagt tgcaataata cttctcaaac 780
ctgttttaat atttcaagac attaaatcta tgaagtatat aatggtttca aagattcaaa 840
attgacattg ctttactgtc aaaataattt tatggctcac tatgaatcta ttatactgta 900
ttaagagtga aaattgtctt cttctgtgct ggagatgttt tagagttaac aatgatatat 960
ggataatgcc ggtgagaata agagagtcac aaacccttaag taagcaacag cataacaagg 1020
tccaagatac ctaaaagaga tttcaagaga tttaattaat catgaatgtg taacacagtg 1080
cttc 1084
```

<210> 49

<211> 933

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509399CB1

<400> 49

```
gcagcccagc cccaccaga cccatggctg gacctgccac ccagagcccc atgaagctga 60
```

```

tggccctgca gctgctgctg tggcacagtg cactctggac agtgcaggaa gccaccccc 120
tgggcccctgc cagctcccctg cccagagct tccctgctcaa gtgcttagag caagtgagga 180
agatccaggg cgatggcgca gcgctccagg agaagctggg agtgaggga ggctgcttga 240
gccaaactcca tagcggcctt ttcctctacc aggggctcct gcaggccctg gaagggatct 300
cccccgagtt ggggtcccacc ttggacacac tgcagctgga cgtcgccgac ttgcccacca 360
ccatctggga gcagatggaa gaactgggaa tggcccctgc cctgcagccc acccagggtg 420
ccatgcgggc cttcgctctt gctttccagc gccgggcagg aggggtcctg gttgcctccc 480
atctgcagag cttcctggag gtgtcgtacc gcgttctacg ccaccttgcc cagccctgag 540
ccaagccctc cccatcccat gtatttatct ctatttaata tttatgtcta ttttaagctc 600
atattttaaag acagggaaga gcagaacgga gccccaggcc tctgtgtcct tccctgcatt 660
tctgagtttc attctcctgc ctgtagcagt gagaaaaagc tcctgtcctc ccattcccctg 720
gactgggagg tagataggta aataccaagt atttattact atgactgtct cccagccctg 780
gctctgcaat gggcactggg atgagccgct gtgagccctt ggtcctgagg gtccccacct 840
gggacccttg agagtatcag gtctccacg tgggagacaa gaaatccctg tttaatattt 900
aaacagcagt gttccccatc tgggtccttg cac 933

```

<210> 50

<211> 8378

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7504789CB1

<400> 50

```

tccttcctcc caggaagggc caggaaaatg gccctggctc tggagatctt caccctgctg 60
gcctccatct gctgggtgtc ggccaatatc ttcgagtacc aggtggatgc ccagcccctt 120
cgtccctgtg agctgcagag ggaaacggcc tttctgaagc aagcagacta cgtgccccag 180
tgtgcagagg atggcagctt ccagactgtc cagtgcaga acgacggccg ctccctgctg 240
tgtgtgggtg ccaacggcag tgaagtgtc ggcagcaggc agccaggacg gcctgtgggt 300
tgtctgtcat tttgtcagct acagaaacag cagatcttac tgagtggcta cattaacagc 360
acagacacct cctacctccc tcagtgtcag gattcagggg actacgcgcc tgttcagtgt 420
gatgtgcagc aggtccagtg ctggtgtgtg gacgcagagg ggatggagg gtatgggacc 480
cgccagctgg ggaggccaaa gcgatgtcca aggagctgtg aaataagaaa tcgtcgtctt 540
ctccacgggg tgggagataa gtcaccaccc cagtgttctg cggagggaga gtttatgctt 600
gtccagtga aatttgtcaa caccacagac atgatgatt ttgatctggt ccacagctac 660
aacagggttc cagatgctatt tgtgaccttc agttccttcc agaggagggt ccctgaggta 720
tctgggtatt gccactgtgc tgacagccaa gggcggaac tggctgagac aggtttggag 780
ttgttactgg atgaaattta tgacaccatt tttgctggcc tggaccttcc ttcaccttc 840
actgaaacca ccctgtaccg gatactgcag agacgggtcc tcgcagttca atcagtcatt 900
tctggcagat tccgatgccc caaaaaatgt gaagtggagc ggtttacagc aaccagcttt 960
ggtcacccct atgttccaag ctgcccgcga aatggcgact atcaggcggg gcagtgccag 1020
acggaagggc cctgctggtg tgtggacgcc caggggaagg aaatgcatgg aaccgggac 1080
caaggggagc cgccatcttg tgcgaaggc caatcttggt cctccgaaag gcagcaggcc 1140
ttgtccagac ttacttttgg gacctcaggc tacttcagcc agcacgacct gttctcttcc 1200
ccagagaaaa gatgggcctc tccaagagta gccagatttg ccacatcctg cccaccacg 1260
atcaaggagc tctttgtgga ctctgggctt ctccgcccaa tgggtggagg acagagccaa 1320
cagttttctg tctcagaaaa tcttctcaaa gaagccatcc gagcaatttt tccctcccga 1380
gggctggctc gtcttgccct tcagtttacc accaacccaa agagactcca gcaaaacctt 1440
tttggaggga aatttttggg gaatgttggc cagtttaact tgtctggagc ccttggcaca 1500
agaggcacat ttaacttcag tcaatttttc cagcaacttg gtcttgcaag cttcttgaat 1560
ggaggagac aagaagattt ggccaagcca ctctctgtg gattagattc aaattcttcc 1620
acaggaaacc ctgaagctgc taagaaggat ggtactatga ataagccaac tgtgggcagc 1680
tttggctttg aaattaacct acaagagaac caaatgccc tcaaattcct tgcctctctc 1740
ctggagcttc cagaattcct tctcttcttg caacatgcta tctctgtgcc agaagatgtg 1800
gcaagagatt taggtgatgt gatggaaacg gtactcagct cccagacctg tgagcagaca 1860
cctgaaaggc tatttgtccc atcatgcacg acagaaggaa gctatgagga tgtccaatgc 1920
ttttccggag agtgctggtg tgtgaattcc tggggcaaag agcttccagg ctcaagagtc 1980
agagggtgac agccaaggtg cccacagac tgtgaaaagc aaagggtctg catgcaaagc 2040
ctcatgggca gccagcctgc tggctccacc ttgtttgtcc ctgcttgtag tagtgaggga 2100
catttcctgc ctgtccagtg cttcaactca gagtgtact gtgttgatgc tgagggtcag 2160
gccattcctg gaactcgaag tgcaataggg aagcccaaga aatgccccac gccctgtcaa 2220

```

ttacagtctg	agcaagcttt	cctcaggacg	gtgcaggccc	tgctctctaa	ctccagcatg	2280
ctaccacccc	tttccgacac	ctacatccca	cagtgcagca	ccgatgggca	gtggagacaa	2340
gtgcaatgca	atgggcctcc	tgagcagggtc	ttcgagttgt	accaacgatg	ggaggctcag	2400
aacaagggcc	aggatctgac	gcctgccaa	ctgctagtga	agatcatgag	ctacagagaa	2460
gcagcttccg	gaaacttcag	tctctttatt	caaagtctgt	atgaggctgg	ccagcaagat	2520
gtcttcccg	tgctgtcaca	atacccttct	ctgcaagatg	tcccactagc	agcactggaa	2580
gggaaacggc	cccagcccag	ggagaatatc	ctcctggagc	cctacctctt	ctggcagatc	2640
ttaaatggcc	aactcagcca	ataccggggg	tcctactcag	acttcagcac	tcctttggca	2700
cattttgatc	ttcggaactg	ctgggtgtgtg	gatgaggctg	gccaagaact	ggaaggaatg	2760
cggctctgagc	caagcaagct	cccaacatgt	cctggctcct	gtgaggaagc	aaagctccgt	2820
gtactgcagt	tcattaggga	aacggaagag	attgtttcag	cttccaacag	ttctcggttc	2880
cctctggggg	agagtttcct	ggtggccaag	ggaatccggc	tgaggaatga	ggacctcggc	2940
cttcctccgc	tcttcccgcc	ccgggagget	ttcgcgagc	agtttctgcg	tgggagtgat	3000
tacgccattc	gcctggcggc	tcagtctacc	ttaagcttct	atcagagacg	ccgcttttcc	3060
ccggacgact	cggctggagc	atccgcccct	ctgcggtcgg	gcccctacat	gccacagtgt	3120
gatgcgtttg	gaagtgggga	gcctgtgcag	tgccacgctg	ggactgggca	ctgctgggtg	3180
gtagatgaga	aaggagggtt	catccctggc	tcactgactg	cccgtctctt	gcagattcca	3240
cagtgcccg	caacctgcga	gaaatctcga	accagtgggc	tgctttccag	ttggaaacag	3300
gctagatccc	aagaaaaccc	atctccaaaa	gacctgttcg	tcccagcctg	cctagaaaca	3360
ggagagtatg	ccaggtcgca	ggcatcgggg	gctggcacct	ggtgtgtgga	ccctgcatca	3420
ggagaagagt	tgccggcctgg	ctcgagcagc	agtggccagt	gcccagccct	ctgcaatgtg	3480
ctcaagagtg	gagtcctctc	caggagagtc	agccaggct	atgtcccagc	ctgcagggca	3540
gaggatgggg	gcttttcccc	agtgcaatgt	gaccaggccc	agggcagctg	ctgggtgtgc	3600
atggacagcg	gagaagaggt	gcctgggacg	cgcgtgaccg	ggggccagcc	cgctgtgag	3660
agcccgcggt	gtccgctgcc	attcaacgcg	tcggagggtg	ttggtggaac	aatcctgtgt	3720
gagacaatct	cgggccccac	aggctctgcc	atgcagcagt	gccaattgct	gtgccgccag	3780
ggctcctgga	gcgtgtttcc	accagggcca	ttgatattga	gcctggagag	cggacgctgg	3840
gagtcacagc	tgccctcagcc	ccgggcctgc	caacggcccc	agctgtggca	gaccatccag	3900
acccaagggc	actttcagct	ccagctcccg	ccgggcaaga	tgtgcagtgc	tgactacgcg	3960
ggtttgctgc	agactttcca	ggttttccata	ttggatgagc	tgacagcccg	cggcttctgc	4020
cgatccagg	tgaagacttt	tgggaccctg	gtttccattc	ctgtctgcaa	caactcctct	4080
gtgcagggtg	gttgtctgac	cagggagcgt	ttaggagtga	atgttacatg	gaaatcacgg	4140
cttgaggaca	tcccagtggc	ttctcttctt	gacttacatg	acattgagag	agccttggtg	4200
ggcaaggatc	tccttgggcg	cttcacagat	ctgatccaga	gtggctcatt	ccagcttcat	4260
ctggactcca	agacgttccc	agcggaaaacc	atccgcttcc	tccaagggga	ccactttggc	4320
acctctccca	ggacatggtt	tggtgtctcg	gaaggattct	accaagtctt	gacaagttag	4380
gccagtcagg	acggagctggg	atgctgtaag	tgtcctgaag	gaagctattc	ccaagatgag	4440
gaaatgcattc	cttgtctgtg	tggaattctac	caagaacagg	cagggagctt	ggcctgtgtc	4500
ccatgtcctg	tgggcagaac	gaccatttct	gccggagctt	tcagccagac	tcactgtgtc	4560
actgactgtc	agaggaacga	agcaggcctg	caatgtgacc	agaatggcca	gtatcgagcc	4620
agccagaagg	acaggggcag	tgggaaggcc	ttctgtgtgg	acggcgaggg	gcgagggtg	4680
ccatggtggg	aaacagaggc	ccctcttgag	gactcacagt	gtttgatgat	gcagaagttt	4740
gagaagggtt	cagaatcaaa	ggtgatcttc	gacgccaatg	ctcctgtggc	tgtcagatcc	4800
aaagttcctg	attctgagtt	ccccgtgatg	cagtgtctga	cagattgcac	agaggacgag	4860
gcctgcagct	gtccacccgt	gtccacagag	gagccagaga	tttctgtgta	tttctatgtc	4920
tggaacaagt	acaatgttgc	ctgcatgact	ctcgaccaga	aacgagatgc	actgggggaa	4980
tcaaaggcca	ccagctttgg	aagtcttcgc	tgccagggtga	aagtgaggag	ccatggtcaa	5040
gattctccag	ctgtgtattt	gaaaaagggc	caaggatcca	ccacaacact	tcagaaacgc	5100
tttgaaccca	ctgggtttcca	aaacatgctt	tctggattgt	acaaccccat	tgtgttctca	5160
gcctcaggag	ccaatctaac	cgatgtctac	ctcttctgtc	ttcttgcatg	cgaccgtgat	5220
ctgtgttgcg	atggcttcgt	cctcacacag	gttcaaggag	gtgccatcat	ctgtgggttg	5280
ctgagctcac	ccagtgtcct	gctttgtaat	gtcaaagact	ggatggatcc	ctctgaagcc	5340
tgggctaata	ctacatgtcc	tggtgtgaca	tatgaccagg	agagccacca	ggtgatattg	5400
cgtcttggag	accaggagtt	catcaagaat	tctgacatgg	ggtctcggcc	tgagtctatg	5460
ggatgtagaa	aaaacacagt	gccaaaggcca	gcatctccaa	cagaagcagg	tttgacaaca	5520
gaacttttct	cccctgtgga	cctcaaccag	gtcattgtca	atggaaatca	atcactatcc	5580
agccagaagc	actggctttt	caagcacctg	ttttcagccc	agcaggcaaa	cctatggtgc	5640
ctttctcggt	gtgtgcagga	gcactctttc	tgtcagctcg	cagagataac	agagagtgc	5700
tccttgact	tcacctgcac	cctctaccca	gaggcacagg	tgtgtgatga	catcatggag	5760
tccaatgccc	agggctgcag	actgatcctg	cctcagatgc	caaaggccct	gttccggaag	5820
aaagttatac	tggaagataa	agtgaaagac	ttttacactc	gcctgccggt	ccaaaaactg	5880
atggggatat	ccattagaaa	taaagtgcc	atgtctgaaa	aatctatttc	taatgggttc	5940
tttgaatgtg	aacgatgggtg	cgatgcggac	ccatgctgca	ctggcttttg	atttctaaat	6000

gtttccagct	taaaaggagg	agaggtgaca	tgtctcactc	tgaacagctt	gggaattcag	6060
atgtgcagtg	aggagaatgg	aggagcctgg	cgcatttttg	actgtggctc	tcctgacatt	6120
gaagtccaca	cctatccctt	cggatggtac	cagaagccca	ttgctcaaaa	taatgctccc	6180
agtttttgcc	cttttggttg	tctgccttcc	ctcacagaga	aagtgtctct	ggactcgtgg	6240
cagtcacctg	ccctctcttc	agtggttggt	gatccatcca	ttaggcactt	tgatggtgcc	6300
catgtcagca	ctgctgccac	cagcaatttc	tctgctgtcc	gagacctctg	tttgtcgga	6360
tgttcccaac	atgaggcctg	tctcatcacc	actctgcaaa	cccaacctgg	ggctgtgaga	6420
tgtatgttct	atgctgatac	tcaaagctgc	acacatagtc	tgcagggtca	gaactgccga	6480
cttctgcttc	gtgaagaggc	caccacatc	taccggaagc	caggaatctc	tctgctcagc	6540
tatgaggcat	ctgtaccttc	tgtgccatt	tccacctatg	gccggctgct	gggcaggctc	6600
caggccatcc	aggtgggtac	ctcatggaag	caagtggacc	agttccttgg	agttccatat	6660
gctgccccgc	ccctggcaga	gaggcgcttc	caggcaccag	agcccttgaa	ctggacaggc	6720
tcctgggatg	ccagcaagcc	aagggccagc	tgctggcagc	caggcaccag	aacatccacg	6780
tctcctggag	tcagtgaaag	ttgtttgtat	ctcaatgtgt	tcattccctca	gaatgtggcc	6840
cctaacgcgt	ctgtgctggt	gttcttccac	aacaccatgg	acagggagga	gagtgaagga	6900
tggccggcta	tcgacggctc	cttcttggct	gctgttggca	acctcatcgt	ggtcactgcc	6960
agctaccgag	tgggtgtctt	cggcttcctg	agttctgggt	ccggagaggt	gagtggcaac	7020
tgggggctgc	tggaccaggt	ggcggctctg	acctgggtgc	agaccacat	ccgaggattt	7080
ggcggggacc	ctcggcgctg	gtccctggca	gcagaccgtg	gcggggctga	tgtggccagc	7140
atccaccttc	tcacggccag	ggccaccaac	tcccaacttt	tccggagagc	tgtgctgatg	7200
ggaggctccg	cactctcccc	ggcggccgtc	atcagccatg	agagggctca	gcagcaggca	7260
attgcttttg	caaaggaggt	cagttgcccc	atgtcatcca	gccaagaagt	ggtgtcctgc	7320
ctccgccaga	agcctgccaa	tgtcctcaat	gatgcccaga	ccaagctcct	ggccgtgagt	7380
ggcccttttc	actactgggg	tctgtgatc	gatggccact	tcctccgtga	gcctccagcc	7440
agagcactga	agaggtcttt	atgggtagag	gtcgatctgc	tcattgggag	ttctcaggac	7500
gacgggctca	tcaacagagc	aaaggctgtg	aagcaatttg	aggaaagtgc	aggccggacc	7560
agtagcaaaa	cagcctttta	ccaggcactg	cagaattctc	tgggtggcga	ggactcagat	7620
gcccgctcg	agggtgctgc	tacatggtat	tactctctgg	agcactccac	ggatgactat	7680
gctccttct	cccgggctct	ggagaatgcc	acccgggact	actttatcat	ctgccctata	7740
atcgacatgg	ccagtgcctg	ggcaaagagg	gcccaggagaa	acgtcttcat	gtaccatgct	7800
cctgaaaact	acggccatgg	cagcctggag	ctgctggcgg	atgttcagtt	tgccttgggg	7860
cttcccttct	accagccta	tgaggggcag	ttttctctgg	aggagaagag	cctgtcgctg	7920
aaaatcatgc	agtacttttc	ccacttccatc	agatcaggaa	atcccaacta	cccttatgag	7980
ttctcacgga	aagtaccac	atttgcaacc	ccctggcctg	actttgtacc	ccgtgctggt	8040
ggagagaact	acaaggagtt	cagtgaagctg	ctccccaatc	gacagggcct	gaagaaagcc	8100
gactgctcct	tctggtccaa	gtacatctcg	tctctgaaga	catctgcagg	taatggagcc	8160
aagggcgggc	agtcagcaga	gagtgaagag	gaggagttga	cggctggatc	tgggctaaga	8220
gaagatctcc	taagcctcca	ggaaccaggc	tctaagacct	acagcaagtg	accagccctt	8280
gagctcccca	aaaacctcac	ccgaggctgc	ccactatggt	catctttttc	tctaaaatag	8340
ccacttacct	tcaataaagt	atctacatgc	ggtgaagc			8378

<210> 51

<211> 864

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7506910CB1

<400> 51

cctcaactcc	tgccacaatg	tacaggatgc	aactcctgtc	ttgcattgca	ctaagtcttg	60
cacttgtcac	aaacagtgca	cctacttcaa	gttctacaaa	gaaaacacag	ctacaactgg	120
agcattttact	gctggattta	cagatgattt	tgaatggaat	taatgtaagt	atatttcctt	180
tcttactaaa	attattacat	ttagtaatct	agctggagat	catttcttaa	taacaatgca	240
ttatactttc	ttagaattac	aagaatccca	aactcaccag	gatgctcaca	tttaagtttt	300
acatgcccac	gaaggccaca	gaactgaaac	atcttcagtg	tctagaagaa	gaactcaaac	360
ctctggagga	agtgtctaat	ttagctcaaa	gcaaaaactt	tcacttaaga	cccagggaact	420
taatcagcaa	tatcaacgta	atagttcttg	aactaaaggg	atctgaaaca	acattcatgt	480
gtgaatatgc	ttagtagaga	gcaaccattg	tagaatttct	gaacagatgg	attacctttt	540
gtcaaagcat	catctcaaca	ctgacttgat	aattaaagtgc	ttcccaacta	aaacatatca	600
ggccttctat	ttatttaaatt	atttaaattt	tatatattat	gttgaatgta	tggtttgcta	660
cctattgtaa	ctattattct	taattctaaa	actataaata	tggatctttt	atgattcttt	720

```

ttgtaagccc taggggctct aaaatggttt cacttatttta tcccaaaata tttattatta 780
tggtgaatgt taaatatagt atctatgtag attgggttagt aaaactatatt aataaatttg 840
ataaatataa acaaaaaaaaa aaaa 864

```

```

<210> 52
<211> 1259
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7507000CB1

```

```

<400> 52
cacgcaaggc tgctgtgtct tacacattct cccaggagtt gtctgagccg ccgagtggac 60
agtggctgat tatggagagc agaggccac tggctacctc gcgcctgctg ctgttgctgc 120
tggtgctact actgcgtcac acccgccagg gatgggtcct gagacctgtt ctccccaccc 180
aggtgcagga gcgggacagg gcaactcagct catgcagtct tcccttctct cctctggccc 240
tgtagcaggg cctctccctc tgtctgtctc tgacatgtcc ctactcagct ttgtttgttt 300
tctctttctg atagagtgcc cacgacctc cggctgtcca cctcagcaat ggcccaggac 360
aagagcctat cgctgtcatg acctttgacc tcaccaagat cacaaaaacc tcctcctcct 420
ttgagggttcg aacctggggc ccagagggag tgatttttta tggggatacc aaccctaagg 480
atgactgggt tatgctggga cttcgagacg gcaggcctga gatccaactg cacaatcact 540
gggcccagct tacggtgggt gctggaccac ggctggatga tgggagatgg caccaggtgg 600
aagtcaagat ggagggggac tctgtgctgc tggaggtgga tggggaggag gtgctgcgcc 660
tgagacaggt ctctggggcc ctgaccagca aacgccatcc catcatgagg attgcgcttg 720
gggggctgct cttccccgct tccaaccttc ggttgccgct ggttcctgcc ctggatggct 780
gcctgcgccg ggattcctgg ctggacaaac agggccagat ctcagcatct gccccacta 840
gcctcagaag ctgtgatgta gaatcaaata ccgggatatt tctccctcca gggactcagg 900
cagaattcaa tctccgagac attccccagc ctcatgcaga gccctgggcc ttctctttgg 960
acctgggact caagcaggca gcaggctcag gccacctcct tgctcttggg acaccagaga 1020
acccatcttg gctcagtcct cacctccaag atcaaaagggt ggtgttgtct tctgggtcgg 1080
ggccagggct ggatctgccc ctggctcttg gactccctct tcagctgaag ctgagtatgt 1140
ccaggggtgg cttgagccaa gggtcgaaga tgaaggccct tgccctgcct cccttaggcc 1200
tggtccccc ccttaacctc tgggccaagc ctcaagggcg tctcttctg ggggcttta 1259

```

```

<210> 53
<211> 965
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7510882CB1

```

```

<400> 53
cacgcaaggc tgctgtgtct tacacattct cccaagagtt gtctgagccg ccgagtggac 60
agtggctgat tatggagagc agaggccac tggctacctc gcgcctgctg ctgttgctgc 120
tggtgctact actgcgtcac acccgccagg gatggggcct gagacctgtt ctccccaccc 180
agagtgcaca cgacctccg gctgtccacc tcagcaatgg cccaggacaa gagcctatcg 240
ctgtcatgac ctttgacctc accaagatca caaaaacctc ctctccttt gaggttcgaa 300
cctgggacct agagggagtg attttttatg gggataccaa ccctaaggat gactggttta 360
tgctgggact tcgagacggc aggcctgaga tccaactgca caatcactgg gccagctta 420
cgggtgggtgc tggaccacgg ctggatgatg ggagatggca ccaggtggaa gtcaagatgg 480
agggggactc tgtgctgctg gaggtggatg gggaggaggt gctgcgcctg agacaggtct 540
ctggggccct gaccagcaaa cgccatccca tcatgaggat tgcgcttggg gggctgctct 600
tccccgcttc caaccttcgg ttgccgctgg ttcctgccct ggatggctgc ctgcgccggg 660
attcctggct ggacaaacag gccgagatct cagcatctgc cccactagc ctcagaagct 720
gtgatgtaga atcaaatccc gggatatttt tccctccagg gactcaggca gaattcaatc 780
tccgaggaga agactcttcc acctcttttt gectgaatgg cctttgggca caaggtcaga 840
ggctggatgt ggaccaggcc ctgaacagaa gccatgagat ctggactcac agctgcccc 900
agagcccagg caatggcact gacgcttccc attaaagctc cacctaagaa aaaaaaaaaa 960
aaagg 965

```

<210> 54
 <211> 1351
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7506575CB1

<400> 54
 ccagcagag ccatttaacc tttgggaaca aggcaactag cgtctggcag caggaatcca 60
 accagtgcc tgagttctga ggcagagagg aggacagaag aaacaagagg ctggagattg 120
 tcaaattcag tatcccagtt ggctcttgat tcttggtgaa accatccctc agctcctaga 180
 gggagattgt tagatcatga aactaattac catccttttc ctctgctcca ggctgctact 240
 aagtttaacc cagacctgcc agattactcc agccgagggc cctgtggtga cagcccagta 300
 cgactgcctc ggctgtgtgc atcctatata aacgcagagc ccagacctgg agcccattct 360
 gagacacggc attcagtact ttaacaacaa cactcaacat tcctccctct tcatgcttaa 420
 tgaagtaaaa cgggccc aaa gacaggtggt ggctggattg aactttcgaa ttacctactc 480
 aattgtgcaa acgaattggt ccaaagagaa ttttctgttc ttaactccag actgcaagtc 540
 cctttggaat ggtgataccg gtgaatgtac agataatgca tacatcgata ttcagctacg 600
 aattgcttcc ttctcacaga actgtgacat ttatccaggg aaggattttg tacaaccacc 660
 tacciaagatt tgcgtgggct gcccagaga tatacccacc aacagcccag agctggagga 720
 gacactgact cacaccatca caaagcttaa tgcagagaaat aacgcaactt tctatttcaa 780
 gattgacaat gtgaaaaaag caagagtaca ggtgggtggc ggcaagaaat attttattga 840
 cttcgtggcc agggaaacca catgttccaa ggaaagtaat gaagagtga ccgaaagctg 900
 tgagaccaaa aaacttggcc aaagcctaga ttgcaacgct gaagtttatg tggtagcctg 960
 ggagaaaaaa atttacccta ctgtcaactg tcaaccactg ggaatgatct cactgatgaa 1020
 aaggcctcca ggtttttcac ctttccgata atcacgaata ggggaaataa aagaagaaa 1080
 aactagtcac ctaaggctcc gcgagtacaa gggctgaccc ccaaaggcag gggcagagcc 1140
 agcatctgag agggaggtct cttgaccaat gggcagaatc ttcactccag gcacatagcc 1200
 ccaaccacct ctgccagcaa ccttgagagg aaggacaaga agaaagatgg gatagaattt 1260
 aaatagagaa gaatgccatt ttatcactct gcctctgggt gaaataaaga tcagtcttga 1320
 tgttctaact ctaaaaaaa aaaaaaaaaa a 1351

<210> 55
 <211> 1459
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7506576CB1

<400> 55
 ccagcagag ccatttaacc tttgggaaca aggcaactag cgtctggcag caggaatcca 60
 accagtgcc tgagttctga ggcagagagg aggacagaag aaacaagagg ctggagattg 120
 tcaaattcag tatcccagtt ggctcttgat tcttggtgaa accatccctc agctcctaga 180
 gggagattgt tagatcatga aactaattac catccttttc ctctgctcca ggctgctact 240
 aagtttaacc caggaatcac agtccgagga aattgactgc aatgacaagg atttatttaa 300
 agctgtggat gctgctctga agaaatataa cagtcaaaac caaagtaaca accagtttgt 360
 attgtaccgc attactccag ccgagggccc tgtggtgaca gccagtagc actgcctcgg 420
 ctgtgtgcat cctatatcaa cgcagagccc agacctggag ccattctga gacacggcat 480
 tcagtacttt aacaacaaca ctcaacattc ctccctcttc atgcttaatg aagtaaaacg 540
 ggcccaaaga caggtggtgg ctggattgaa ctttcgaatt acctactcaa ttgtgcaaac 600
 gaattgttcc aaagagaatt ttctgttctt aactccagac tgcaagtccc tttggaatgg 660
 tgataccggt gaatgtacag ataatgcata catcgatatt cagctacgaa ttgcttcctt 720
 ctacacagaa tgtgacattt atccagggaa ggattttgta caaccaccta ccaagatttg 780
 cgtgggctgc ccagagata taccaccaa cagcccagag ctggaggaga cactgactca 840
 caccatcaca agctttaatg cagagaataa cgcaactttc tatttcaaga ttgacaatgt 900
 gaaaaaagca agagtacagg tgggtgctgg caagaaatat tttattgact tcgtggccag 960
 ggaaaccaca tgttccaagg aaagtaatga agagttgacc gaaagctgtg agacaaaaa 1020
 acttggccaa agcctagatt gcaacgctga agtttatgtg gtaccctggg agaaaaaaat 1080
 ttaccctact gtcaactgtc aaccactggg aatgatctca ctgatgaaa ggcctccagg 1140

```

tttttcacct ttccgatcat cacgaatagg ggaaataaaa gaagaaacaa ctagtcacct 1200
aaggtcctgc gagtacaagg gtcgaccccc aaaggcaggg gcagagccag catctgagag 1260
ggaggtctct tgaccaatgg gcagaatctt cactccaggg acatagcccc aaccacctct 1320
gccagcaacc ttgagaggaa ggacaagaag aaagatggga tagaatttaa atagagaaga 1380
atgccatttt atcactctgc ctctgggtga aataaagatc agtcttgatg ttctaactct 1440
aaaaaaaaaa aaaaaaaaaa
1459

```

<210> 56

<211> 1469

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7512018CB1

<400> 56

```

cccagcagag ccatttaacc tttgggaaca aggcaactag cgtctggcag caggaatcca 60
accagtcccc tgagttctga ggcagagagg aggacagaag aaacaagagg ctggagattg 120
tcaaattcag tatcccagtt ggctcttgat tcttggtgaa accatccctc agctcctaga 180
gggagattgt tagatcatga aactaattac catccttttc ctctgctcca ggctgctact 240
aagtttaacc caggaatcac agtccgagga aattgactgc aatgacaagg atttatttaa 300
agctgtggat gctgctctga agaaatataa cagtcaaaac caaagtaaca accagtttgt 360
attgtaccgc ataactgaag ccactaagac ggtgggtggct ggattgaact ttcgaattac 420
ctactcaatt gtgcaaacga attgttccaa agagaatttt ctgttcttaa ctccagactg 480
caagtccctt tgggaatggtg ataccggtga atgtacagat aatgcataca tcgatattca 540
gctacgaatt gcttccttct cacagaactg tgacatttat ccagggaagg attttgtaca 600
accacctacc aagatttgcg tgggctgccc cagagatata cccaccaaca gccagagct 660
ggaggagaca ctgactcaca ccatcacaaa gcttaatgca gagaataacg caactttcta 720
tttcaagatt gacaatgtga aaaaagcaag agtacagggtg gtggctggca agaaatattt 780
tattgacttc gtggccaggg aaaccacatg ttccaaggaa agtaatgaag agttgaccga 840
aagctgtgag accaaaaaac ttggccaaag cctagattgc aacgctgaag tttatgtgg 900
accctgggag aaaaaaattt accctactgt caactgtcaa ccactgggaa tgatctcact 960
gatgaaaagg cctccagggt tttcaccttt ccgatcatca cgaatagggg aaataaaaga 1020
agaaacaact agtcacctaa ggtcctgcga gtacaagggt cgacccccaa aggcaggggc 1080
agagccagca tctgagaggg aggtctcttg accaatgggc agaattctta ctccaggcac 1140
atagcccaaa ccacctctgc cagcaacctt gagaggaagg acaagaagaa agatgggata 1200
gaatttaaat agagaagaat gccattttat cactctgcct ctgggtgaaa taaagatcag 1260
tcttgatgtt ctaactctaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaacaaa 1320
cacaacaaac ggggggcccg cataaaagag gagcccgaca cacggaggaa tataatccgg 1380
ggcgatcag gagggaatac accattacca tattggagcg cacattaacc gagggggcaa 1440
aacagggcaa cgagagtcaa cagtgtgca
1469

```