

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/42330 A2

(51) International Patent Classification⁷: **C07K 14/47**

(21) International Application Number: PCT/US01/50983

(22) International Filing Date: 26 October 2001 (26.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/244,022	27 October 2000 (27.10.2000)	US
60/247,370	8 November 2000 (08.11.2000)	US
60/251,831	7 December 2000 (07.12.2000)	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*): **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, Sunnyvale, CA 94577 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **WALIA, Narinder, K.** [US/US]; 890 David Street #205, San Leandro, CA 94577 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94043 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). **DING, Li** [CN/US]; 3353 Alma Street #146, Palo Alto, CA 94506 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **LAL, Preeti, G.** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **BATRA, Sajeev** [US/US]; 555 El Camino Real #709, San Leandro, CA 94087 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **SANJANWALA, Madhusudan, M.** [US/US]; 210 Silvia Court, Los Altos, CA

94024 (US). **ARVIZU, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **GURURAJAN, Rajagopal** [IN/US]; 5591 Dent Avenue, San Jose, CA 95118 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94522 (US). **XU, Yum-ing** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **BURFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (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, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, 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: CYSTOSKELETON-ASSOCIATED PROTEINS

(57) Abstract: The invention provides human cystoskeleton-associated proteins (CSAP) and polynucleotides which identity and encode CSAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of CSAP.



WO 02/42330 A2

CYTOSKELETON-ASSOCIATED PROTEINS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of cytoskeleton-associated proteins and to the use of these sequences in the diagnosis, treatment, and prevention of cell proliferative disorders, viral infections, and neurological disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of cytoskeleton-associated proteins.

BACKGROUND OF THE INVENTION

The cytoskeleton is a cytoplasmic network of protein fibers that mediate cell shape, structure, and movement. The cytoskeleton supports the cell membrane and forms tracks along which organelles and other elements move in the cytosol. The cytoskeleton is a dynamic structure that allows cells to adopt various shapes and to carry out directed movements. Major cytoskeletal fibers include the microtubules, the microfilaments, and the intermediate filaments. Motor proteins, including myosin, dynein, and kinesin, drive movement of or along the fibers. The motor protein dynamin drives the formation of membrane vesicles. Accessory or associated proteins modify the structure or activity of the fibers while cytoskeletal membrane anchors connect the fibers to the cell membrane.

Microtubules and Associated Proteins

Tubulins

Microtubules, cytoskeletal fibers with a diameter of about 24 nm, have multiple roles in the cell. Bundles of microtubules form cilia and flagella, which are whip-like extensions of the cell membrane that are necessary for sweeping materials across an epithelium and for swimming of sperm, respectively. Marginal bands of microtubules in red blood cells and platelets are important for these cells' pliability. Organelles, membrane vesicles, and proteins are transported in the cell along tracks of microtubules. For example, microtubules run through nerve cell axons, allowing bi-directional transport of materials and membrane vesicles between the cell body and the nerve terminal. Failure to supply the nerve terminal with these vesicles blocks the transmission of neural signals. Microtubules are also critical to chromosomal movement during cell division. Both stable and short-lived populations of microtubules exist in the cell.

Microtubules are polymers of GTP-binding tubulin protein subunits. Each subunit is a heterodimer of α - and β - tubulin, multiple isoforms of which exist. The hydrolysis of GTP is linked to the addition of tubulin subunits at the end of a microtubule. The subunits interact head to tail to form protofilaments; the protofilaments interact side to side to form a microtubule. A microtubule is

polarized, one end ringed with α -tubulin and the other with β -tubulin, and the two ends differ in their rates of assembly. Generally, each microtubule is composed of 13 protofilaments although 11 or 15 protofilament-microtubules are sometimes found. Cilia and flagella contain doublet microtubules. Microtubules grow from specialized structures known as centrosomes or microtubule-organizing centers (MTOCs). MTOCs may contain one or two centrioles, which are pinwheel arrays of triplet microtubules. The basal body, the organizing center located at the base of a cilium or flagellum, contains one centriole. Gamma tubulin present in the MTOC is important for nucleating the polymerization of α - and β - tubulin heterodimers but does not polymerize into microtubules.

Microtubule-Associated Proteins

Microtubule-associated proteins (MAPs) have roles in the assembly and stabilization of microtubules. One major family of MAPs, assembly MAPs, can be identified in neurons as well as non-neuronal cells. Assembly MAPs are responsible for cross-linking microtubules in the cytosol. These MAPs are organized into two domains: a basic microtubule-binding domain and an acidic projection domain. The projection domain is the binding site for membranes, intermediate filaments, or other microtubules. Based on sequence analysis, assembly MAPs can be further grouped into two types: Type I and Type II. Type I MAPs, which include MAP1A and MAP1B, are large, filamentous molecules that co-purify with microtubules and are abundantly expressed in brain and testes. Type I MAPs contain several repeats of a positively-charged amino acid sequence motif that binds and neutralizes negatively charged tubulin, leading to stabilization of microtubules. MAP1A and MAP1B are each derived from a single precursor polypeptide that is subsequently proteolytically processed to generate one heavy chain and one light chain.

Another light chain, LC3, is a 16.4 kDa molecule that binds MAP1A, MAP1B, and microtubules. It is suggested that LC3 is synthesized from a source other than the MAP1A or MAP1B transcripts, and that the expression of LC3 may be important in regulating the microtubule binding activity of MAP1A and MAP1B during cell proliferation (Mann, S.S. et al. (1994) J. Biol. Chem. 269:11492-11497).

Type II MAPs, which include MAP2a, MAP2b, MAP2c, MAP4, and Tau, are characterized by three to four copies of an 18-residue sequence in the microtubule-binding domain. MAP2a, MAP2b, and MAP2c are found only in dendrites, MAP4 is found in non-neuronal cells, and Tau is found in axons and dendrites of nerve cells. Alternative splicing of the Tau mRNA leads to the existence of multiple forms of Tau protein. Tau phosphorylation is altered in neurodegenerative disorders such as Alzheimer's disease, Pick's disease, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia and Parkinsonism linked to chromosome 17. The altered Tau phosphorylation leads to a collapse of the microtubule network and the formation of

intraneuronal Tau aggregates (Spillantini, M.G. and M. Goedert (1998) Trends Neurosci. 21:428-433).

The protein pericentrin is found in the MTOC and has a role in microtubule assembly.

Another microtubule associated protein, STOP (stable tubule only polypeptide), is a calmodulin-
5 regulated protein that regulates stability (Denariier, E. et al. (1998) Biochem. Biophys. Res. Commun. 24:791-796). In order for neurons to maintain conductive connections over great distances, they rely upon axodendritic extensions, which in turn are supported by microtubules. STOP proteins function to stabilize the microtubular network. STOP proteins are associated with axonal microtubules, and are also abundant in neurons (Guillaud, L. et al. (1998) J. Cell Biol. 142:167-179). STOP proteins are
10 necessary for normal neurite formation, and have been observed to stabilize microtubules, *in vitro*, against cold-, calcium-, or drug-induced disassembly (Margolis, R.L. et al. (1990) EMBO 9:4095-502).

Microfilaments and Associated Proteins

15 Actins

Microfilaments, cytoskeletal filaments with a diameter of about 7-9 nm, are vital to cell locomotion, cell shape, cell adhesion, cell division, and muscle contraction. Assembly and disassembly of the microfilaments allow cells to change their morphology. Microfilaments are the polymerized form of actin, the most abundant intracellular protein in the eukaryotic cell. Human cells contain six
20 isoforms of actin. The three α -actins are found in different kinds of muscle, nonmuscle β -actin and nonmuscle γ -actin are found in nonmuscle cells, and another γ -actin is found in intestinal smooth muscle cells. G-actin, the monomeric form of actin, polymerizes into polarized, helical F-actin filaments, accompanied by the hydrolysis of ATP to ADP. Actin filaments associate to form bundles and networks, providing a framework to support the plasma membrane and determine cell shape.
25 These bundles and networks are connected to the cell membrane. In muscle cells, thin filaments containing actin slide past thick filaments containing the motor protein myosin during contraction. A family of actin-related proteins exist that are not part of the actin cytoskeleton, but rather associate with microtubules and dynein.

Actin-Associated Proteins

30 Actin-associated proteins have roles in cross-linking, severing, and stabilization of actin filaments and in sequestering actin monomers. Several of the actin-associated proteins have multiple functions. Bundles and networks of actin filaments are held together by actin cross-linking proteins. These proteins have two actin-binding sites, one for each filament. Short cross-linking proteins promote bundle formation while longer, more flexible cross-linking proteins promote network

formation. Actin-interacting proteins (AIPs) participate in the regulation of actin filament organization. Other actin-associated proteins such as TARA, a novel F-actin binding protein, function in a similar capacity by regulating actin cytoskeletal organization. Calmodulin-like calcium-binding domains in actin cross-linking proteins allow calcium regulation of cross-linking. Group I cross-linking proteins have unique actin-binding domains and include the 30 kD protein, EF-1a, fascin, and scruin. Group II cross-linking proteins have a 7,000-MW actin-binding domain and include villin and dematin. Group III cross-linking proteins have pairs of a 26,000-MW actin-binding domain and include fimbrin, spectrin, dystrophin, ABP 120, and filamin.

The Rho family of low molecular weight GTP-binding proteins regulates actin organization, and controls signal transduction pathways that link extracellular and intracellular signals to the rearrangement of the actin cytoskeleton. This affects such diverse processes as cell shape and motility, cell adhesion, and proliferation. LMW GTP-binding proteins cycle between the active GTP-bound form and the inactive GDP-bound form, and this cycling is regulated by additional proteins. The intrinsic rate of GTP hydrolysis of the LMW GTP-binding proteins is typically very slow, but it can be stimulated by several orders of magnitude by GTPase-activating proteins (GAPs) (Geyer, M. and Wittinghofer, A. (1997) *Curr. Opin. Struct. Biol.* 7:786-792) while guanine-nucleotide exchange factors (GEFs) promote GDP dissociation and facilitate GTP binding. In the active GTP-bound state, Rho proteins interact with and activate downstream effectors to control the assembly of actin filaments (for a review, see Schmidt A. and Hall, M.N. (1998) *Annu. Rev. Cell. Dev. Biol.* 14:305-38).

Severing proteins regulate the length of actin filaments by breaking them into short pieces or by blocking their ends. Severing proteins include gCAP39, severin (fragmin), gelsolin, and villin. Capping proteins can cap the ends of actin filaments, but cannot break filaments. Capping proteins include CapZ and tropomodulin. The proteins thymosin and profilin sequester actin monomers in the cytosol, allowing a pool of unpolymerized actin to exist. The actin-associated proteins tropomyosin, troponin, and caldesmon regulate muscle contraction in response to calcium.

Microtubule and actin filament networks cooperate in processes such as vesicle and organelle transport, cleavage furrow placement, directed cell migration, spindle rotation, and nuclear migration. Microtubules and actin may coordinate to transport vesicles, organelles, and cell fate determinants, or transport may involve targeting and capture of microtubule ends at cortical actin sites. These cytoskeletal systems may be bridged by myosin-kinesin complexes, myosin-CLIP170 complexes, formin-homology (FH) proteins, dynein, the dynactin complex, Kar9p, coronin, ERM proteins, and kelch repeat-containing proteins (for a review, see Goode, B.L. et al. (2000) *Curr. Opin. Cell Biol.* 12:63-71). The kelch repeat is a motif originally observed in the kelch protein, which is involved in

formation of cytoplasmic bridges called ring canals. A variety of mammalian and other kelch family proteins have been identified. The kelch repeat domain is believed to mediate interaction with actin (Robinson, D.N. and L. Cooley (1997) *J. Cell Biol.* 138:799-810).

ADF/cofilins are a family of conserved 15-18 kDa actin-binding proteins that play a role in cytokinesis, endocytosis, and in development of embryonic tissues, as well as in tissue regeneration and in pathologies such as ischemia, oxidative or osmotic stress. LIM kinase 1 downregulates ADF (Carlier, M.F. et al. (1999) *J. Biol. Chem.* 274:33827-33830).

LIM is an acronym of three transcription factors, Lin-11, Isl-1, and Mec-3, in which the motif was first identified. The LIM domain is a double zinc-finger motif that mediates the protein-protein interactions of transcription factors, signaling, and cytoskeleton-associated proteins (Roof, D.J. et al. (1997) *J. Cell Biol.* 138:575-588). These proteins are distributed in the nucleus, cytoplasm, or both (Brown, S. et al. (1999) *J. Biol. Chem.* 274:27083-27091). Recently, ALP (actinin-associated LIM protein) has been shown to bind alpha-actinin-2 (Bouju, S. et al. (1999) *Neuromuscul. Disord.* 9:3-10).

The Frabin protein is another example of an actin-filament binding protein (Obaishi, H. et al. (1998) *J. Biol. Chem.* 273:18697-18700). Frabin (EGD1-related F-actin-binding protein) possesses one actin-filament binding (FAB) domain, one Dbp1 homology (DH) domain, two pleckstrin homology (PH) domains, and a single cysteine-rich FYVE (Fab1p, YOTB, Vac1p, and EEA1 (early endosomal antigen 1)) domain. Frabin has shown GDP/GTP exchange activity for Cdc42 small G protein (Cdc42), and indirectly induces activation of Rac small G protein (Rac) in intact cells. Through the activation of Cdc42 and Rac, Frabin is able to induce formation of both filopodia- and lamellipodia-like processes (Ono, Y. et al. (2000) *Oncogene* 19:3050-3058). The Rho family small GTP-binding proteins are important regulators of actin-dependent cell functions including cell shape change, adhesion, and motility. The Rho family consists of three major subfamilies: Cdc42, Rac, and Rho. Rho family members cycle between GDP-bound inactive and GTP-bound active forms by means of a GDP/GTP exchange factor (GEF) (Umikawa, M. et al. (1999) *J. Biol. Chem.* 274:25197-25200). The Rho GEF family is crucial for microfilament organization.

Intermediate Filaments and Associated Proteins

Intermediate filaments (IFs) are cytoskeletal fibers with a diameter of about 10 nm, intermediate between that of microfilaments and microtubules. IFs serve structural roles in the cell, reinforcing cells and organizing cells into tissues. IFs are particularly abundant in epidermal cells and in neurons. IFs are extremely stable, and, in contrast to microfilaments and microtubules, do not function in cell motility.

Five types of IF proteins are known in mammals. Type I and Type II proteins are the acidic

and basic keratins, respectively. Heterodimers of the acidic and basic keratins are the building blocks of keratin IFs. Keratins are abundant in soft epithelia such as skin and cornea, hard epithelia such as nails and hair, and in epithelia that line internal body cavities. Mutations in keratin genes lead to epithelial diseases including epidermolysis bullosa simplex, bullous congenital ichthyosiform erythroderma (epidermolytic hyperkeratosis), non-epidermolytic and epidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, pachyonychia congenita, and white sponge nevus. Some of these diseases result in severe skin blistering. (See, e.g., Wawersik, M. et al. (1997) *J. Biol. Chem.* 272:32557-32565; and Corden L.D. and W.H. McLean (1996) *Exp. Dermatol.* 5:297-307.)

Type III IF proteins include desmin, glial fibrillary acidic protein, vimentin, and peripherin.

Desmin filaments in muscle cells link myofibrils into bundles and stabilize sarcomeres in contracting muscle. Glial fibrillary acidic protein filaments are found in the glial cells that surround neurons and astrocytes. Vimentin filaments are found in blood vessel endothelial cells, some epithelial cells, and mesenchymal cells such as fibroblasts, and are commonly associated with microtubules. Vimentin filaments may have roles in keeping the nucleus and other organelles in place in the cell. Type IV IFs include the neurofilaments and nestin. Neurofilaments, composed of three polypeptides NF-L, NF-M, and NF-H, are frequently associated with microtubules in axons. Neurofilaments are responsible for the radial growth and diameter of an axon, and ultimately for the speed of nerve impulse transmission. Changes in phosphorylation and metabolism of neurofilaments are observed in neurodegenerative diseases including amyotrophic lateral sclerosis, Parkinson's disease, and Alzheimer's disease (Julien, J.P. and W.E. Mushynski (1998) *Prog. Nucleic Acid Res. Mol. Biol.* 61:1-23). Type V IFs, the lamins, are found in the nucleus where they support the nuclear membrane.

IFs have a central α -helical rod region interrupted by short nonhelical linker segments. The rod region is bracketed, in most cases, by non-helical head and tail domains. The rod regions of intermediate filament proteins associate to form a coiled-coil dimer. A highly ordered assembly process leads from the dimers to the IFs. Neither ATP nor GTP is needed for IF assembly, unlike that of microfilaments and microtubules.

IF-associated proteins (IFAPs) mediate the interactions of IFs with one another and with other cell structures. IFAPs cross-link IFs into a bundle, into a network, or to the plasma membrane, and may cross-link IFs to the microfilament and microtubule cytoskeleton. Microtubules and IFs are particularly closely associated. IFAPs include BPAG1, plakoglobin, desmoplakin I, desmoplakin II, plectin, ankyrin, filaggrin, and lamin B receptor.

Cytoskeletal-Membrane Anchors

Cytoskeletal fibers are attached to the plasma membrane by specific proteins. These attachments are important for maintaining cell shape and for muscle contraction. In erythrocytes, the

spectrin-actin cytoskeleton is attached to the cell membrane by three proteins, band 4.1, ankyrin, and adducin. Defects in this attachment result in abnormally shaped cells which are more rapidly degraded by the spleen, leading to anemia. In platelets, the spectrin-actin cytoskeleton is also linked to the membrane by ankyrin; a second actin network is anchored to the membrane by filamin. In muscle cells the protein dystrophin links actin filaments to the plasma membrane; mutations in the dystrophin gene lead to Duchenne muscular dystrophy. In adherens junctions and adhesion plaques the peripheral membrane proteins α -actinin and vinculin attach actin filaments to the cell membrane.

Focal adhesions

Focal adhesions are specialized structures in the plasma membrane involved in the adhesion of a cell to a substrate, such as the extracellular matrix. Focal adhesions form the connection between an extracellular substrate and the cytoskeleton, and affect such functions as cell shape, cell motility and cell proliferation. Transmembrane integrin molecules form the basis of focal adhesions. Upon ligand binding, integrins cluster in the plane of the plasma membrane. Cytoskeletal linker proteins such as the actin binding proteins α -actinin, talin, tensin, vinculin, paxillin, and filamin are recruited to the clustering site. Key regulatory proteins, such as Rho and Ras family proteins, focal adhesion kinase, and Src family members are also recruited. These events lead to the reorganization of actin filaments and the formation of stress fibers. These intracellular rearrangements promote further integrin-ECM interactions and integrin clustering. Thus, integrins mediate aggregation of protein complexes on both the cytosolic and extracellular faces of the plasma membrane, leading to the assembly of the focal adhesion. Many signal transduction responses are mediated via various adhesion complex proteins, including Src, FAK, paxillin, and tensin. (For a review, see Yamada, K.M. and B. Geiger, (1997) Curr. Opin. Cell Biol. 9:76-85.)

IFs are also attached to membranes by cytoskeletal-membrane anchors. The nuclear lamina is attached to the inner surface of the nuclear membrane by the lamin B receptor. Vimentin IFs are attached to the plasma membrane by ankyrin and plectin. Desmosome and hemidesmosome membrane junctions hold together epithelial cells of organs and skin. These membrane junctions allow shear forces to be distributed across the entire epithelial cell layer, thus providing strength and rigidity to the epithelium. IFs in epithelial cells are attached to the desmosome by plakoglobin and desmoplakins. The proteins that link IFs to hemidesmosomes are not known. Desmin IFs surround the sarcomere in muscle and are linked to the plasma membrane by paranemin, synemin, and ankyrin.

Motor Proteins

Myosin-related Motor Proteins

Myosins are actin-activated ATPases, found in eukaryotic cells, that couple hydrolysis of ATP with motion. Myosin provides the motor function for muscle contraction and intracellular movements

such as phagocytosis and rearrangement of cell contents during mitotic cell division (cytokinesis). The contractile unit of skeletal muscle, termed the sarcomere, consists of highly ordered arrays of thin actin-containing filaments and thick myosin-containing filaments. Crossbridges form between the thick and thin filaments, and the ATP-dependent movement of myosin heads within the thick filaments pulls the thin filaments, shortening the sarcomere and thus the muscle fiber.

Myosins are composed of one or two heavy chains and associated light chains. Myosin heavy chains contain an amino-terminal motor or head domain, a neck that is the site of light-chain binding, and a carboxy-terminal tail domain. The tail domains may associate to form an α -helical coiled coil. Conventional myosins, such as those found in muscle tissue, are composed of two myosin heavy-chain subunits, each associated with two light-chain subunits that bind at the neck region and play a regulatory role. Unconventional myosins, believed to function in intracellular motion, may contain either one or two heavy chains and associated light chains. There is evidence for about 25 myosin heavy chain genes in vertebrates, more than half of them unconventional.

Dynein-related Motor Proteins

Dyneins are (-) end-directed motor proteins which act on microtubules. Two classes of dyneins, cytosolic and axonemal, have been identified. Cytosolic dyneins are responsible for translocation of materials along cytoplasmic microtubules, for example, transport from the nerve terminal to the cell body and transport of endocytic vesicles to lysosomes. As well, viruses often take advantage of cytoplasmic dyneins to be transported to the nucleus and establish a successful infection (Sodeik, B. et al. (1997) J. Cell Biol. 136:1007-1021). Virion proteins of herpes simplex virus 1, for example, interact with the cytoplasmic dynein intermediate chain (Ye, G.J. et al. (2000) J. Virol. 74:1355-1363). Cytoplasmic dyneins are also reported to play a role in mitosis. Axonemal dyneins are responsible for the beating of flagella and cilia. Dynein on one microtubule doublet walks along the adjacent microtubule doublet. This sliding force produces bending that causes the flagellum or cilium to beat. Dyneins have a native mass between 1000 and 2000 kDa and contain either two or three force-producing heads driven by the hydrolysis of ATP. The heads are linked via stalks to a basal domain which is composed of a highly variable number of accessory intermediate and light chains. Cytoplasmic dynein is the largest and most complex of the motor proteins.

Kinesin-related Motor Proteins

Kinesins are (+) end-directed motor proteins which act on microtubules. The prototypical kinesin molecule is involved in the transport of membrane-bound vesicles and organelles. This function is particularly important for axonal transport in neurons. Kinesin is also important in all cell types for the transport of vesicles from the Golgi complex to the endoplasmic reticulum. This role is critical for maintaining the identity and functionality of these secretory organelles.

Kinesins define a ubiquitous, conserved family of over 50 proteins that can be classified into at least 8 subfamilies based on primary amino acid sequence, domain structure, velocity of movement, and cellular function. (Reviewed in Moore, J.D. and S.A. Endow (1996) *Bioessays* 18:207-219; and Hoyt, A.M. (1994) *Curr. Opin. Cell Biol.* 6:63-68.) The prototypical kinesin molecule is a

5 heterotetramer comprised of two heavy polypeptide chains (KHCs) and two light polypeptide chains (KLCs). The KHC subunits are typically referred to as “kinesin.” KHC is about 1000 amino acids in length, and KLC is about 550 amino acids in length. Two KHCs dimerize to form a rod-shaped molecule with three distinct regions of secondary structure. At one end of the molecule is a globular motor domain that functions in ATP hydrolysis and microtubule binding. Kinesin motor domains are

10 highly conserved and share over 70% identity. Beyond the motor domain is an α -helical coiled-coil region which mediates dimerization. At the other end of the molecule is a fan-shaped tail that associates with molecular cargo. The tail is formed by the interaction of the KHC C-termini with the two KLCs.

Members of the more divergent subfamilies of kinesins are called kinesin-related proteins

15 (KRPs), many of which function during mitosis in eukaryotes (Hoyt, *supra*). Some KRPs are required for assembly of the mitotic spindle. *In vivo* and *in vitro* analyses suggest that these KRPs exert force on microtubules that comprise the mitotic spindle, resulting in the separation of spindle poles. Phosphorylation of KRP is required for this activity. Failure to assemble the mitotic spindle results in abortive mitosis and chromosomal aneuploidy, the latter condition being characteristic of

20 cancer cells. In addition, a unique KRP, centromere protein E, localizes to the kinetochore of human mitotic chromosomes and may play a role in their segregation to opposite spindle poles.

Dynamamin-related Motor Proteins

Dynamamin is a large GTPase motor protein that functions as a “molecular pinchase,” generating a mechanochemical force used to sever membranes. This activity is important in forming clathrin-

25 coated vesicles from coated pits in endocytosis and in the biogenesis of synaptic vesicles in neurons. Binding of dynamamin to a membrane leads to dynamamin’s self-assembly into spirals that may act to constrict a flat membrane surface into a tubule. GTP hydrolysis induces a change in conformation of the dynamamin polymer that pinches the membrane tubule, leading to severing of the membrane tubule and formation of a membrane vesicle. Release of GDP and inorganic phosphate leads to dynamamin

30 disassembly. Following disassembly the dynamamin may either dissociate from the membrane or remain associated to the vesicle and be transported to another region of the cell. Three homologous dynamamin genes have been discovered, in addition to several dynamamin-related proteins. Conserved dynamamin regions are the N-terminal GTP-binding domain, a central pleckstrin homology domain that binds membranes, a central coiled-coil region that may activate dynamamin’s GTPase activity, and a C-

terminal proline-rich domain that contains several motifs that bind SH3 domains on other proteins. Some dynamin-related proteins do not contain the pleckstrin homology domain or the proline-rich domain. (See McNiven, M.A. (1998) *Cell* 94:151-154; Scaife, R.M. and R.L. Margolis (1997) *Cell. Signal.* 9:395-401.)

5 The cytoskeleton is reviewed in Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York NY.

 The discovery of new cytoskeleton-associated proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative disorders, viral infections, and neurological disorders, and
10 in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of cytoskeleton-associated proteins.

SUMMARY OF THE INVENTION

 The invention features purified polypeptides, cytoskeleton-associated proteins, referred to
15 collectively as "CSAP" and individually as "CSAP-1," "CSAP-2," "CSAP-3," "CSAP-4," "CSAP-5," "CSAP-6," "CSAP-7," "CSAP-8," "CSAP-9," "CSAP-10," "CSAP-11," "CSAP-12," "CSAP-13," and "CSAP-14." In one aspect, the invention 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:1-14, b) a polypeptide comprising a naturally occurring amino acid sequence at least
20 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-14.

25 The invention further 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:1-14, 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:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the
30 group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-14. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:15-28.

 Additionally, the invention 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:1-14, 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:1-14, c) a
5 biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

10 The invention also 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:1-14, 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:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group
15 consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. 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.

20 Additionally, the invention 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:1-14, 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:1-14, c) a biologically active fragment of a polypeptide having an
25 amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ
30 ID NO:15-28, 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:15-28, 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 one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, 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:15-28, 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, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28, 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:15-28, 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, and, optionally, if present, the amount thereof.

The invention further 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:1-14, 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:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional CSAP, comprising administering to a patient in need of such treatment the composition.

The invention also 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:1-14, 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:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CSAP, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention 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:1-14, 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:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional CSAP, comprising administering to a patient in need of such treatment the composition.

The invention further 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:1-14, 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:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. 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.

The invention further 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:1-14, 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:1-14, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14. 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.

The invention further 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:15-28, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further 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:15-28, ii) 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:15-28, 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:15-28, ii) 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:15-28, 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 comprises a fragment of a polynucleotide sequence 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 the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides 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 sequences of the invention, 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 sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

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 the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, 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 present invention which will be limited only by the appended claims.

It must be noted that 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 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

“CSAP” refers to the amino acid sequences of substantially purified CSAP 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 CSAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CSAP either by directly interacting with CSAP or by acting on components of the biological pathway in which CSAP participates.

An “allelic variant” is an alternative form of the gene encoding CSAP. 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 CSAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CSAP or a polypeptide with at least one functional characteristic of CSAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CSAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CSAP. 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 CSAP. Deliberate

amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CSAP 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" refer to an oligopeptide, peptide, polypeptide, or 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 sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

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

10 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. (See, e.g., 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
15 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

20 substrates containing right-handed nucleotides.

The term “antisense” refers to any composition capable of base-pairing with the “sense” (coding) strand of 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
25 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
30 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 CSAP, 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,

5 3'-TCA-5'.

A “composition comprising a given polynucleotide sequence” and a “composition comprising a given amino acid sequence” refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CSAP or fragments of CSAP may be
10 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
15 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
20 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
25 conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
30	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
35	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu

5	Met Phe Ser Thr Trp Tyr Val	Leu, Ile His, Met, Leu, Trp, Tyr Cys, Thr Ser, Val Phe, Tyr His, Phe, Trp 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 CSAP or the polynucleotide encoding CSAP which is 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 5 to 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
5 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:15-28 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:15-28, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:15-28 is useful, for
10 example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:15-28 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:15-28 and the region of SEQ ID NO:15-28 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:1-14 is encoded by a fragment of SEQ ID NO:15-28. A
15 fragment of SEQ ID NO:1-14 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-14. For example, a fragment of SEQ ID NO:1-14 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-14. The precise length of a fragment of SEQ ID NO:1-14 and the region of SEQ ID NO:1-14 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the
20 intended purpose for the fragment.

A "full length" polynucleotide sequence 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, interchangeably, sequence identity, between two
25 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 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
30 achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may 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 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. Percent identity is reported by CLUSTAL V as the "percent
 5 similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms 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), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at
 10 <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 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The
 15 "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:

Matrix: BLOSUM62

20 *Reward for match: 1*

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

25 *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
 30 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 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.

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. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

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 acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A
5 hybridization complex may be formed in solution (e.g., C₀t or R₀t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence 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 nucleotide
10 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.

15 An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CSAP 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 CSAP 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,
20 polypeptides, or other chemical compounds on a substrate.

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

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

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.

30 "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
5 may be pegylated to extend their lifespan in the cell.

“Post-translational modification” of an CSAP 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 CSAP.

10 “Probe” refers to nucleic acid sequences encoding CSAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. 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
15 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 sequence, 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
20 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
25 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. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; 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
30 purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000

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

A "recombinant nucleic acid" is a sequence 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.

5 An “RNA equivalent,” in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence 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.

10 The term “sample” is used in its broadest sense. A sample suspected of containing CSAP, nucleic acids encoding CSAP, 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
15 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.

20 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 60% free, preferably at least 75% free, and most preferably at least 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
25 by different amino acid residues or nucleotides, respectively.

“Substrate” refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

30 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.

“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 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
5 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 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
10 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. 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,
15 fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods 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
20 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
25 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
30 reference molecule. Species variants are polynucleotide sequences 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 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 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 of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human cytoskeleton-associated proteins (CSAP), the polynucleotides encoding CSAP, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative disorders, viral infections, and neurological disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences 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.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) 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. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank 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.

5 Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are cytoskeleton-associated proteins. For example, SEQ ID NO:1 is 93% identical to mouse NBL4, a Band 4.1 family cytoskeletal protein (GenBank ID g466548) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.5e-287$, which indicates the probability of obtaining the observed

10 polypeptide sequence alignment by chance. SEQ ID NO:1 also contains a FERM/Band 4.1 family domain 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 BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:1 is an Band 4.1 family cytoskeletal protein. In an alternative example, SEQ ID NO:8 is 84%

15 identical to *Rattus norvegicus* nadrin, an actin-filament regulating protein (GenBank ID g9971185) 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. SEQ ID NO:8 also contains a Rho-GAP (GTPase activating) site domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-

20 based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:8 is a nadrin. In an alternative example, SEQ ID NO:11 is 68% identical to sea urchin dynein, intermediate chain (GenBank ID g927639) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $1.5e-222$, which indicates the probability of obtaining the

25 observed polypeptide sequence alignment by chance. SEQ ID NO:11 also contains a WD repeat domain characteristic of dynein intermediate chains, 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 BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:11 is a cytoplasmic dynein intermediate chain. SEQ ID

30 NO:2-7, SEQ ID NO:9-10, and SEQ ID NO:12-14 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-14 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence

identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:15-28 or that distinguish between SEQ ID NO:15-28 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7011045F8 is the identification number of an Incyte cDNA sequence, and KIDNNOC01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71108830V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g1548017) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the identification numbers in column 5 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 identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, 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 identification numbers in column 5 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, FLXXXXXX_gAAAAA_gBBBBB_1_N is the identification number of 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).

5 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.

15

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

20 Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences 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 polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

25 The invention also encompasses CSAP variants. A preferred CSAP 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 CSAP amino acid sequence, and which contains at least one functional or structural characteristic of CSAP.

30 The invention also encompasses polynucleotides which encode CSAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:15-28, which encodes CSAP. The polynucleotide

sequences of SEQ ID NO:15-28, 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 a variant of a polynucleotide sequence encoding CSAP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CSAP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:15-28 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:15-28. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CSAP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide sequence encoding CSAP. A splice variant may have portions which have significant sequence identity to the polynucleotide sequence encoding CSAP, 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 the polynucleotide sequence encoding CSAP 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 sequence encoding CSAP. Any one of the splice variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CSAP.

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 CSAP, 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 CSAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CSAP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CSAP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CSAP 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 CSAP 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 DNA sequences which encode CSAP and CSAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence 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 sequence encoding CSAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:15-28 and fragments thereof under various conditions of stringency. (See, e.g., 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 Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). 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 (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CSAP 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. (See, e.g., 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. (See, e.g., 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. (See, e.g., 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. (See, e.g., 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 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, polynucleotide sequences or fragments thereof which encode CSAP may be cloned in recombinant DNA molecules that direct expression of CSAP, or

fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CSAP.

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

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

In another embodiment, sequences encoding CSAP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, CSAP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and 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 CSAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences

from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.)

5 The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, *supra*, pp. 28-53.)

In order to express a biologically active CSAP, the nucleotide sequences encoding CSAP 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
10 a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CSAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CSAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where
15 sequences encoding CSAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both
20 natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., 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 sequences encoding CSAP and appropriate transcriptional and translational control
25 elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences
30 encoding CSAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or

animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and 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 nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and 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 polynucleotide sequences encoding CSAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CSAP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPO11 plasmid (Life Technologies). Ligation of sequences encoding CSAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CSAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of CSAP 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 CSAP. 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 sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of CSAP. Transcription of sequences encoding CSAP 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. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and 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
5 transfection. (See, e.g., 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, sequences encoding CSAP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader
10 sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CSAP in host cells. (See, e.g., 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.

15 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. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

20 For long term production of recombinant proteins in mammalian systems, stable expression of CSAP in cell lines is preferred. For example, sequences encoding CSAP 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
25 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
30 include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap^r* cells, respectively. (See, e.g., 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. (See, e.g., Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci. USA* 77:3567-3570; Colbere-Garapin, F. et al. (1981) *J. Mol. Biol.* 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., 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. 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. (See, e.g., Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131.)

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

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

Immunological methods for detecting and measuring the expression of CSAP 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 CSAP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., 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; and 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 CSAP include

oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide.

Alternatively, the sequences encoding CSAP, 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
5 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 Pharmacia Biotech, 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.

10 Host cells transformed with nucleotide sequences encoding CSAP 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 CSAP may be designed to contain signal sequences which direct
15 secretion of CSAP 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 sequences 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
20 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.

25 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CSAP 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 CSAP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CSAP activity. Heterologous protein and
30 peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins,

respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity 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 CSAP encoding sequence and the heterologous protein sequence, so that CSAP
5 may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

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

CSAP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to CSAP. At least one and up to a plurality of test compounds may be screened
15 for specific binding to CSAP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of CSAP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):
20 Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which CSAP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express CSAP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E.
25 coli. Cells expressing CSAP or cell membrane fractions which contain CSAP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either CSAP 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
30 assay may comprise the steps of combining at least one test compound with CSAP, either in solution or affixed to a solid support, and detecting the binding of CSAP 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.

CSAP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of CSAP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for CSAP activity, wherein CSAP is combined with at least one test compound, and the activity of CSAP in the presence of a test compound is compared with the activity of CSAP in the absence of the test compound. A change in the activity of CSAP in the presence of the test compound is indicative of a compound that modulates the activity of CSAP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising CSAP under conditions suitable for CSAP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CSAP 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 CSAP 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 CSAP 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 CSAP 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 CSAP 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 CSAP, e.g., by secreting CSAP 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 CSAP and cytoskeleton-associated proteins. In addition, the expression of CSAP is closely associated with brain and neurological tissues, cardiovascular tissues, digestive tissues, and endocrine tissues. Therefore, CSAP appears to play a role in cell proliferative disorders, viral infections, and neurological disorders. In the treatment of disorders associated with increased CSAP expression or activity, it is desirable to decrease the expression or activity of CSAP. In the treatment of disorders associated with decreased CSAP expression or activity, it is desirable to increase the expression or activity of CSAP.

Therefore, in one embodiment, CSAP 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 CSAP. Examples of such disorders include, but are not limited to, 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 a cancer 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; a viral infection such as those caused by adenoviruses (acute respiratory disease, pneumonia), arenaviruses (lymphocytic choriomeningitis), bunyaviruses (Hantavirus), coronaviruses (pneumonia, chronic bronchitis), hepadnaviruses (hepatitis), herpesviruses (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), flaviviruses (yellow fever), orthomyxoviruses (influenza), papillomaviruses (cancer), paramyxoviruses (measles, mumps), picornaviruses (rhinovirus, poliovirus, coxsackie-virus), polyomaviruses (BK virus, JC virus), poxviruses (smallpox), reovirus (Colorado tick fever), retroviruses (human immunodeficiency virus, human T lymphotropic virus), rhabdoviruses (rabies), rotaviruses (gastroenteritis), and togaviruses (encephalitis, rubella); and a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor

neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, a prion disease including kuru, Creutzfeldt-Jakob disease, and

5 Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders,

10 peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder.

15 In another embodiment, a vector capable of expressing CSAP 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 CSAP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified CSAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent

20 a disorder associated with decreased expression or activity of CSAP including, but not limited to, those provided above.

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

25 In a further embodiment, an antagonist of CSAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CSAP. Examples of such disorders include, but are not limited to, those cell proliferative disorders, viral infections, and neurological disorders described above. In one aspect, an antibody which specifically binds CSAP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a

30 pharmaceutical agent to cells or tissues which express CSAP.

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

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary

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

An antagonist of CSAP may be produced using methods which are generally known in the art. In particular, purified CSAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CSAP. Antibodies to CSAP may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

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

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CSAP 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 CSAP 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 CSAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., 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; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate

antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and 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 CSAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., 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. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for CSAP may also be generated. For example, such fragments include, but are not limited to, $F(ab)_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the $F(ab)_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., 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 immunoassays typically involve the measurement of complex formation between CSAP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CSAP 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 CSAP. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of CSAP-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 affinities for multiple CSAP epitopes, represents the average affinity, or avidity, of the antibodies for CSAP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular CSAP 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 CSAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a

ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CSAP, 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).

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 CSAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding CSAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding CSAP. 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 CSAP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding CSAP 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), thalassamias, 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 CSAP expression or regulation causes disease, the expression of CSAP 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 CSAP are treated by constructing mammalian expression vectors encoding CSAP and introducing these vectors by mechanical means into CSAP-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 CSAP 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). CSAP 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 CSAP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID
5 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
10 standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to CSAP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding CSAP 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
15 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
20 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
25 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).
30

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding CSAP to cells which have one or more genetic abnormalities with respect to the expression of CSAP. 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, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CSAP to target cells which have one or more genetic abnormalities with respect to the expression of CSAP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CSAP 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, hereby incorporated by reference. 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 alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding CSAP 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 CSAP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of CSAP-

coding RNAs and the synthesis of high levels of CSAP 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 CSAP 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. (See, e.g., 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 sequences encoding CSAP.

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

Complementary ribonucleic acid molecules and ribozymes of the invention 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

sequences encoding CSAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

5 RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine,
10 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.

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

At least one, and up to a plurality, of test compounds may be screened for effectiveness in
25 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
30 library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CSAP 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 CSAP 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 CSAP. 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 (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable 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. (See, e.g., 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 CSAP, antibodies to CSAP, and mimetics, agonists, antagonists, or inhibitors of CSAP.

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. 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, 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.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising CSAP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CSAP 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 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 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 CSAP or fragments thereof, antibodies of CSAP, and agonists, antagonists or inhibitors of CSAP, 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 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.

The exact dosage will be determined by the practitioner, in light of factors related to the

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

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

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

A variety of protocols for measuring CSAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CSAP expression. Normal or standard values for CSAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to CSAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CSAP 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, the polynucleotides encoding CSAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, 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 CSAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of

CSAP, and to monitor regulation of CSAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CSAP or closely related molecules may be used to identify nucleic acid sequences which encode CSAP. The specificity of the probe, whether it is made
 5 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 CSAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50%
 10 sequence identity to any of the CSAP 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:15-28 or from genomic sequences including promoters, enhancers, and introns of the CSAP gene.

Means for producing specific hybridization probes for DNAs encoding CSAP include the cloning of polynucleotide sequences encoding CSAP or CSAP derivatives into vectors for the
 15 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 ³²P or ³⁵S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CSAP may be used for the diagnosis of disorders
 20 associated with expression of CSAP. Examples of such disorders include, but are not limited to, 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 a cancer including
 25 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; a viral infection such as those caused by adenoviruses (acute respiratory disease, pneumonia), arenaviruses (lymphocytic
 30 choriomeningitis), bunyaviruses (Hantavirus), coronaviruses (pneumonia, chronic bronchitis), hepadnaviruses (hepatitis), herpesviruses (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), flaviviruses (yellow fever), orthomyxoviruses (influenza), papillomaviruses (cancer), paramyxoviruses (measles, mumps), picornaviruses (rhinovirus, poliovirus, coxsackie-virus), polyomaviruses (BK virus, JC virus), poxviruses (smallpox), reovirus (Colorado tick fever),

retroviruses (human immunodeficiency virus, human T lymphotropic virus), rhabdoviruses (rabies), rotaviruses (gastroenteritis), and togaviruses (encephalitis, rubella); and a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, a prion disease including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder. The polynucleotide sequences encoding CSAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CSAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CSAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CSAP 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 nucleotide sequences encoding CSAP 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 CSAP, a normal or standard profile for expression is established. This may be accomplished by combining

body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CSAP, 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.

- 5 Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the
10 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
15 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 CSAP
20 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 CSAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding CSAP, 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
25 quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding CSAP 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
30 polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding CSAP 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 fSCCP, 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

5 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).

10 Methods which may also be used to quantify the expression of CSAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., 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
15 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 polynucleotide sequences 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
20 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
25 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, CSAP, fragments of CSAP, or antibodies specific for CSAP may be
30 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. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, 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
5 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,
10 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
15 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, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties.
20 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
25 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
30 important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is 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.

5 Another particular embodiment relates to the use of the polypeptide sequences of the present invention 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
10 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
15 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
20 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 the present invention. In some cases, further sequence data may be obtained for definitive protein
25 identification.

A proteomic profile may also be generated using antibodies specific for CSAP to quantify the levels of CSAP 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-
30 111; Mendoz, 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
5 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
10 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.

15 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
20 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. (See, e.g., 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*
25 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding CSAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either
30 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 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. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; and Trask, B.J. (1991) *Trends Genet.* 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage 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). (See, for example, 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. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data 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 CSAP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

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

In another embodiment of the invention, CSAP, 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 CSAP 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. (See, e.g., 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 CSAP, or fragments thereof, and washed. Bound CSAP is then detected by methods well known in the art. Purified CSAP 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
5 antibodies capable of binding CSAP specifically compete with a test compound for binding CSAP. In
this manner, antibodies can be used to detect the presence of any peptide which shares one or more
antigenic determinants with CSAP.

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

Without further elaboration, it is believed that one skilled in the art can, using the preceding
description, utilize the present invention to its fullest extent. The following embodiments are, therefore,
to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way
15 whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below,
including U.S. Ser. No. 60/244,022, U.S. Ser. No. 60/247,370, and U.S. Ser. No. 60/251,831, are
hereby expressly incorporated by reference.

20

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD
database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were
homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in
25 phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), 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
30 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 (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) 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 Pharmacia Biotech) 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 (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), 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 Life Technologies.

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 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 Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

- 5 Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); 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 (reviewed in Ausubel, 10 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in 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 15 Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from Homo sapiens, Rattus norvegicus, Mus musculus, Caenorhabditis elegans, Saccharomyces cerevisiae, Schizosaccharomyces pombe, and Candida albicans (Incyte Genomics, Palo Alto CA); and hidden 20 Markov model (HMM)-based protein family databases such as PFAM. (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 HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched 25 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 30 sequences. Alternatively, a polypeptide of the invention 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, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences

are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also

5 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 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
 10 which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and
 15 polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:15-28. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative cytoskeleton-associated proteins were initially identified by running the Genscan gene
 20 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 (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of
 25 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 cytoskeleton-associated proteins, the encoded polypeptides were analyzed by querying against PFAM models for cytoskeleton-associated proteins. Potential cytoskeleton-associated proteins were also identified by homology to Incyte cDNA sequences that
 30 had been annotated as cytoskeleton-associated proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri 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
5 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
10 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
15 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
20 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
25 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
30 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 CSAP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:15-28 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:15-28 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. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA 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:

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

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

Alternatively, polynucleotide sequences encoding CSAP 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 CSAP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of CSAP Encoding Polynucleotides

Full length polynucleotide sequences were also 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 Pharmacia Biotech), ELONGASE enzyme (Life Technologies), 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 CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun 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 Pharmacia Biotech), 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 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 Pharmacia Biotech) 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, 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 Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences 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. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:15-28 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), 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 Pharmacia Biotech). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

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

compared.

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

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

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and

incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (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 µl 5X SSC/0.2% SDS.

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 µg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

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

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

Hybridization

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

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

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an
5 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
10 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
15 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
20 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
25 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
30 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).

XI. Complementary Polynucleotides

Sequences complementary to the CSAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CSAP. 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 CSAP. 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 CSAP-encoding transcript.

XII. Expression of CSAP

Expression and purification of CSAP is achieved using bacterial or virus-based expression systems. For expression of CSAP 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 CSAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CSAP 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 CSAP 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. (See 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, CSAP 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 Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CSAP 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 (1995, supra, ch. 10 and 16). Purified CSAP obtained by these methods can be used directly in the assays shown in Examples XVI and XVII, etc. where applicable.

XIII. Functional Assays

CSAP function is assessed by expressing the sequences encoding CSAP 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 (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), 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 CSAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CSAP 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 CSAP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of CSAP Specific Antibodies

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

Alternatively, the CSAP 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
10 selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, *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
15 increase immunogenicity. (See, e.g., Ausubel, 1995, *supra*.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CSAP activity by, for example, binding the peptide or CSAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

20 XV. Purification of Naturally Occurring CSAP Using Specific Antibodies

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

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

XVI. Identification of Molecules Which Interact with CSAP

CSAP, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., 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 CSAP, washed, and

any wells with labeled CSAP complex are assayed. Data obtained using different concentrations of CSAP are used to calculate values for the number, affinity, and association of CSAP with the candidate molecules.

Alternatively, molecules interacting with CSAP are analyzed using the yeast two-hybrid
5 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).

CSAP 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.
10 Patent No. 6,057,101).

XVII. Demonstration of CSAP Activity

A microtubule motility assay for CSAP measures motor protein activity. In this assay, recombinant CSAP is immobilized onto a glass slide or similar substrate. Taxol-stabilized bovine brain microtubules (commercially available) in a solution containing ATP and cytosolic extract are perfused
15 onto the slide. Movement of microtubules as driven by CSAP motor activity can be visualized and quantified using video-enhanced light microscopy and image analysis techniques. CSAP activity is directly proportional to the frequency and velocity of microtubule movement.

Alternatively, an assay for CSAP measures the formation of protein filaments in vitro. A solution of CSAP at a concentration greater than the "critical concentration" for polymer assembly is
20 applied to carbon-coated grids. Appropriate nucleation sites may be supplied in the solution. The grids are negative stained with 0.7% (w/v) aqueous uranyl acetate and examined by electron microscopy. The appearance of filaments of approximately 25 nm (microtubules), 8 nm (actin), or 10 nm (intermediate filaments) is a demonstration of protein activity.

In another alternative, CSAP activity is measured by the binding of CSAP to protein
25 filaments. ³⁵S-Met labeled CSAP sample is incubated with the appropriate filament protein (actin, tubulin, or intermediate filament protein) and complexed protein is collected by immunoprecipitation using an antibody against the filament protein. The immunoprecipitate is then run out on SDS-PAGE and the amount of CSAP bound is measured by autoradiography.

CSAP activity is demonstrated by measuring the effect of CSAP on the activity of a GTPase
30 such as rac or rho. The GTPase is combined with (³²P)GTP for 30 min at 30 °C in the presence and in the absence of CSAP (+CSAP and -CSAP). Aliquots are removed from the +CSAP and -CSAP reaction solutions at intervals, until the reactions are stopped by addition of Norit activated charcoal in NaH₂PO₄ and charcoal is removed by centrifugation. ³²P_i release in both +CSAP and -CSAP solutions is monitored by scintillation count, and the difference is proportional to CSAP activity (Ogier-

Denis, E. et al. (2000) J. Biol. Chem. 275:39090-39095).

Various modifications and variations of the described 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.

- 5 Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
1806450	1	1806450CD1	15	1806450CB1
959690	2	959690CD1	16	959690CB1
7091536	3	7091536CD1	17	7091536CB1
7472724	4	7472724CD1	18	7472724CB1
5844189	5	5844189CD1	19	5844189CB1
7472720	6	7472720CD1	20	7472720CB1
7583990	7	7583990CD1	21	7583990CB1
2058182	8	2058182CD1	22	2058182CB1
3564377	9	3564377CD1	23	3564377CB1
1568689	10	1568689CD1	24	1568689CB1
1393767	11	1393767CD1	25	1393767CB1
3029343	12	3029343CD1	26	3029343CB1
5507629	13	5507629CD1	27	5507629CB1
5607780	14	5607780CD1	28	5607780CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
1	1806450CD1	g466548	1.5e-287	NBL4 [Mus musculus] (Takeuchi, K. et al. (1994) J. Cell. Sci. 107:1921-1928)
2	959690CD1	g3885834	2.2e-11	lin-7-C [Rattus norvegicus] (Irie, M. et al. (1999) Oncogene 18:2811-2817)
3	7091536CD1	g2739096	3.0e-64	Protein 4.1-G [Homo sapiens] (Parra, M. et al. (1998) Genomics 49:298-306)
4	7472724CD1	g6624055	1.9e-19	Similar to ankyrin motif [Homo sapiens]
5	5844189CD1	g1790878	0.0	[Homo sapiens] microtubule-associated protein 1a
6	7472720CD1	g7274242	1.0e-26	Fink, J.K. et al. (1996) Genomics 35:577-585 [Homo sapiens] novel retinal pigment epithelial-cell cytoskeletal thread protein
7	7583990CD1	g3108195	6.4e-57	Kutty, R.K. et al. J. Biol. Chem. published October 19, 2000 as 10.1074/jbc.M007421200 [Homo sapiens] Duo (binds huntingtin-associated protein 1)
8	2058182CD1	g9971185	0.0	Colomer, V. et al. (1997) Hum. Mol. Genet. 6:1519-1525 [Rattus norvegicus] Nadrin, actin-filament regulating protein
9	3564377CD1	g4826478	0.0	Harada, A. et al. J. Biol. Chem. published August 30, 2000 as 10.1074/jbc.M004069200 [Homo sapiens] SH3 domain binding protein 1, a Rac GTPase activating protein
10	1568689CD1	g4742003	2.5e-17	Cicchetti, P. (1995) EMBO J. 14:3127-3135 [Takifugu rubripes] kelch (associated with protein-protein interaction) actin organizing protein
11	1393767CD1	g927639	1.5e-222	Adams, S. et al. (2000) Trends Cell Biol. 10:17-24 dynein intermediate chain 3 [Anthracidaris crassispina] (Ye, GJ. et al. (2000) J. Virol. 74:1355-1363)

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability Score	GenBank Homolog
11		g11493148	0	[Homo sapiens] intermediate dynein chain Bartoloni, L. et al. No deleterious mutations were found in three genes (HFH4, LC8, DNAI2) on human chromosome 17q in patients with Primary Ciliary Dyskinesia Eur. J. Hum. Genet. 8, 126-126 (2000)
12	3029343CD1	g29865	1.6e-14	CENP-E [Homo sapiens] (Yen, T.J. et al. (1992) Nature 359:536-539)
13	5507629CD1	g1098579	4.4e-104	actin [Diphylllobothrium dendriticum]
14	5607780CD1	g11275669	0	[Homo sapiens] (AF225896) tensin Chen, H. et al. Molecular characterization of human tensin Biochem. J. 351 (Pt 2), 403-411 (2000)
		g212752	6.3e-263	tensin [Gallus gallus] (Katz, B.Z. et al. (2000) Biochem. Biophys. Res. Commun. 272:717-720)

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	1806450CD1	580	S190 S306 S310 S314 S322 S35 S393 S426 S428 S433 S448 S464 S476 S484 S486 S491 S532 T199 T255 T36 T382 T405 T439 T577 T76 Y113	N217 N267 N300 N343 N450 N481	FERM domain (Band 4.1 family): C13-H211 Band 4.1 family motif: H181-L210 Band 4.1 family domain signatures: A78-D122, G186-K234 Band 4.1 family domains: BL00660A: L25-L77 BL00660B: R112-D151 BL00660C: E191-K234 BL00660D: Y242-D265 BL00660E: F273-F295 Band 4.1 protein family PR00935: V49-Y61, L117-C130, C130-Y150, E191-G207 ERM family signature PR00661: L83-D102, G126-L147 Band 4 DM00609 P52963 2-423:G2-S426 DM00609 P29074 19-463: E10-S368 DM00609 P11171 200-623: Y12-Q406 DM00609 P11434 183-612: C13-K435 NEL4, structural cytoskeleton, Band 4.1-like PD040496: S402-N531 Cytoskeleton structural protein, phosphatase, protein tyrosine phosphorylation, Moesin, Band PD000961: F11-D209	HMMER-PFAM MOTIFS ProfileScan BLIMPS-BLOCKS BLIMPS-PRINTS BLIMPS-PRINTS BLAST-DOMO BLAST-PRODOR BLAST-PRODOR

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Cytoskeleton structural protein, phosphatase, protein tyrosine phosphorylation, Band PD014063: L210-D404	BLAST-PRODOM
2	959690CD1	541	S153 S194 S226 S248 S263 S273 S295 S436 T113 T158 T288 T356 T363 T364 T479 Y400	N213 N294 N345	Band 4.1-like protein 4 PD129254:R533-E580 PDZ domain (Also known as DHR or GLGF): V43-V135 PX domain: D156-S265 PDZ domain protein: V95-N105 GLGF domain DM00224 A54971 1358-1454: V44-R122 (P-value = 8.2e-10) F25H2.2 protein PD136546: D274-T541	BLAST-PRODOM HMMER-PFAM HMMER-PFAM BLIMPS-PFAM BLAST-DOMO
3	7091536CD1	570	S150 S158 S168 S261 S320 S331 S375 S382 S396 S407 S408 S425 S455 S487 S525 T367 T468 T470 T523 Y17 Y29 T20 S9		Transmembrane domains: S216-Q233, L504-L522 FERM domain (Band 4.1 family): C19-H210 Band 4.1 family motif: W71-D100 Band 4.1 family domain signatures: K76-D120, G185-Q233 Band 4.1 family domain: BL00660A: D26-V78 BL00660B: R110-D149 BL00660C: T190-Q233 BL00660D: I241-E264 BL00660E: Y272-Y294	HMMER HMMER-PFAM MOTIFS ProfileScan BLIMPS-BLOCKS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	7472724CD1	163	S145 S86 T11 T155 T54	N125	Band 4.1 protein family PR00935:L50-F62, L115-C128, C128-Y148, T190-G206	BLIMPS-PRINTS
					ERM family signature PR00661:T30-H49, Q81-D100, G124-I145	BLIMPS-PRINTS
					Band 4 DM00609 P11171 200-623: R15-G379	BLAST-DOMO
					Band 4 DM00609 P52963 2-423: S12-V327	BLAST-DOMO
					Band 4 DM00609 P11434 183-612: C19-G345	BLAST-DOMO
					Band 4 DM00609 P29074 19-463: E16-E337	BLAST-DOMO
					Cytoskeleton structural protein, phosphatase, protein tyrosine phosphorylation, Moesin, Band PD000961:Y17-D208	BLAST-PRODOM
					Cytoskeleton structural protein, phosphatase, protein tyrosine phosphorylation, Band PD014063:H210-P339	BLAST-PRODOM
					Ank repeat: Y31-R63, K64-R96, E97-F129, F130-K162	HMMER-PFAM
					Ank repeat proteins PF00023:L69-L84, G131-Y140	BLIMPS-PFAM

Table 3 (cont.)

[illegible]

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
		S2509 S2511 S2533 S2589 S2682 S2703 S2723 S322 S367 S410 S460 S526 S527 S612 S644 T1944 T1949 T2011 T2212 T2344 T259 T2655 T295 T326 T359 T452 T563 T616 T622 T675 T692 T749 T833 T862 T991 Y1640 Y291 S66 S667 S744 S751 S771 S798 T1590 T1656 T1674 T1728 T1420 T1462			MICROTUBULE ASSOCIATED PROTEIN 1A CONTAINS: MAP1 LIGHT CHAIN LC2 MICROTUBULES REPEAT PD043025: T452-V938 PD042764: V948-Y1388 PD040324: R2158-E2470 PD014346: M1-S410	BLAST_PRODOM
6	7472720CD1	1029	S147 S268 S311 S333 S367 S377 S393 S597 S622 S683 S688 S770 S781 S8 S823 S953 S959 S967 T338 T356 T390 T440 T491 T529 T625 T672 T723 T913 T926 T933 Y892	N388 N391 N720 N721 N949 N997	Ank repeat ank: Q66-K98, E99-I131, Y132-K164, D165-R197, L198-M230 Ank repeat proteins PD00078B: D130-Y142 Ank repeat proteins PF00023: L71-L86, G133-Y142 TRICHOHYALIN DM03839 P22793 921-1475:A359-Q898	HMMER_PFAM BLIMPS_PRODOM BLIMPS_PFAM BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	7583990CD1	696	S172 S184 S235 S253 S269 S305 S368 S55 S550 S568 S575 S609 S640 S661 S667 T109 T212 T272 T38 T416 T572	N182 N226	Signal_cleavage: M1-M40 DBL; ONCOGENE; TRANSFORMING; DM05391 S51620 1-290: V124-Q388 GUANINE NUCLEOTIDE EXCHANGE FACTOR PD006893: P89-Q279 PD009732: M1-Q85	SPSCAN BLAST_DOMO BLAST_PRODOM
8	2058182CD1	803	S109 S150 S171 S261 S308 S349 S46 S497 S589 S624 S742 T175 T214 T233 T252 T38 T415 T468 T550 T60 T83 T96 Y124	N13 N449 N463 N470 N515 N796	Signal_cleavage: M1-A62 RhoGAP GTPase activator domain RhoGAP: A266-T415 Rho-GAP PF00620B: D316-P332 PH (pleckstrin homology) DOMAIN DM00470 P55194 113-387: E178-D439 P11274 973-1254:F250-L408, H541-P568, E118-T175 A49307 566-842: F250-H437, E118-T175 P98171 405-693: F250-L408 GTPASE ACTIVATION PD109560: D87-P248 PD000780: I265-L408 GTPASE DOMAIN AC PD00930B: L367-L407	SPSCAN HMMER_PFAM BLIMPS_PFAM BLAST_DOMO
9	3564377CD1	701	S131 S243 S333 S373 S491 S544 S550 S61 S613	N395 N476	RhoGAP: A290-Q441 Rho-GAP GTPase activator PF00620B: D316-P332 GTPASE DOMAIN AC PD00930A: L391-L431	BLAST_PRODOM BLIMPS_PRODOM HMMER_PFAM BLIMPS_PFAM BLIMPS_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9			S635 S84 T154 T190 T218 T485 T489 T534 T565 T80		GTPASE ACTIVATION PD109561: Q441-G680, PD109560: I88-P292, PD000780: I289-D440 PH (pleckstrin homology) DOMAIN DM00470 P55194 113-387: E193-T467 A49307 566-842: V273-I462 P15882 109-331: H269-D466 A43953 74-296: H269-D466	BLAST_PRODUM BLAST_DOMO
10	1568689CD1	354	T112 T190 T19 T67 S210 S264 T317		Kelch motif Kelch: C20-P66, A68-P114, A116-P162, P164-P209, R211-L265, K270-P316 Kelch repeat signature PR00501 P313-L325, G174-L187, T247-L261	HMMER_PFAM BLIMPS_PRINTS
11	1393767CD1	605	S12 S21 S56 S58 S64 S115 S148 S183 S250 S256 S261 S298 S447 S498 S546 S578 T93 T246 T276 T309 T351 T382 T387 T425 T426 T477 T504 T533 Y106	N19	WD domain, G-beta repeat: A164-D202, P209-D245, L255-D293, T356-S392, S399-D436 WD-40 repeat proteins BL00678: S282-W292 G-PROTEIN BETA WD-40 REPEAT PR00320A: C280-I294 BETA H-PROTEIN TRANSDUCIN PR00319B: C280-I294 DYNEIN INTERMEDIATE CHAIN MOTOR PROTEIN MICROTUBULES FLAGELLA REPEAT WD CILIARY PD040521: E354-F553 PD018332: S12-E315 PD037574: M1-W81 do DYNEIN; CYTOSOLIC; INTERMEDIATE; 74K; DM08083 P54703 101-535: D84-R347	HMMER_PFAM BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_PRINTS BLAST_PRODUM BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
12	3029343CD1	1179	S117 S124 S133 S218 S221 S223 S266 S299 S309 S424 S438 S593 S608 S641 S694 S706 S832 S839 S931 S939 S986 S1014 S959 S83 S1125 T131 T138 T219 S1132 T240 T250 T481 T657 T841 T901 T914 T1011 S77 T1030 T1107 T1148	N591 N655 N695 N899 N1083 N1139 N1146		MOTIFS
13	5507629CD1	372	S77 S201 S203 S253 S335 T76 T355	N12	Actin: M1-F372 Actin proteins BL00406: P7-Q41, E82-Q136, L141-H195, L266-A320, A323-F372 Actins signatures actins_2.prf: K333-F372 Actin signature PR00190: I83-P101, N114-G127, L139-V158, I233-D249, G48-D59, W60-E82 PROTEIN STRUCTURAL ACTIN MULTIGENE FAMILY ACETYLATION MUSCLE CYTOSKELETON CYTOPLASMIC PD0000056: V8-F372	HMMER_PFAM BLIMPS_BLOCKS PROFILES SCAN BLIMPS_PRINTS BLAST_PRODROM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ACTINS AND ACTIN-RELATED PROTEINS DM00167 A03001 1-269: V8-L266 DM00167 P02578 1-268: V8-L266 DM00167 P12432 1-268: Q5-L266 DM00167 P14883 1-269: M1-P261 Actins & actin-related proteins signature: L103-R115	BLAST_DOMO MOTIFS
14	5607780CD1	1561	S63 S171 S271 S286 S399 S477 S478 S515 S534 S556 S613 S687 S754 S859 S878 S892 S927 S952 S982 S1003 S1057 S1059 S1069 S1079 S1098 S1119 S1221 S1270 S1317 S1416 S1512 S1525 S1557 T116 T129 T518 T529 T873 T1094 T1284 T1358 T1475	N169 N242 N397 N438 N451 N844	Phorbol esters/diacylglycerol binding domain: H37-C83 Src homology domain 2: W1288-H1383 Phorbol esters/DAG binding domain proteins BL00479: H37-G59, Q60-C75 Phorbol esters/DAG binding domain dag_pe_binding_domain.prf: C50-P110 TENSIN ACTIN BINDING CYTOSKELETON SH2 DOMAIN PD148069: E457-P1212, Y361-E457 PD034457: S1416-I1555 PD147924: T1188-F1287 PROTEIN PHOSPHORYLATION AUXILIN COAT REPEAT CYCLIN GASSOCIATED KINASE TRANSFERASE PD025411: S171-C360 SRC HOMOMOLOGY 2 (SH2) DOMAIN DM00048 Q04205 1455-1573: Q1282-L1402 Phorbol esters/DAG binding domain: H37-C83	HMMER_PFAM HMMER_PFAM BLIMPS_BLOCKS PROFILESCAN BLAST_PRODUM BLAST_PRODUM BLAST_DOMO MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
15	1806450CB1	2066	1-53, 914-980	7011045F8 (KIDNNOC01)	397	975
				4334891T6 (KIDCTMT01)	1504	2066
				934453T6 (CERVNOT01)	1	424
				6898687R9 (LIVRTMR01)	905	1659
				6768027J1 (BRAUNOR01)	449	1155
16	959690CB1	1912	1532-1912, 65-127	6618442H1 (BRAUTDR04)	720	1226
				959690T6 (BRSTTUT03)	1092	1698
				6618442J1 (BRAUTDR04)	1242	1912
				5603145T6 (MONOTXN03)	663	1212
				7758631H1 (THYMN0E02)	69	689
				GNN.g9188381_000026_002	1	312
17	7091536CB1	2846	1-83, 2277- 2846, 2197- 2224	7190549T9 (BRATDIC01)	1858	2475
				5970463H1 (BRAZNOT01)	2322	2846
				7274095H1 (PROSUNJ01)	672	1260
				5972128H1 (BRAZNOT01)	233	897
				7190549H1 (BRATDIC01)	1	588
				5899281F6 (BRAYDIN03)	1040	1619
				6055275H1 (BRAENOT04)	1461	2059
18	7472724CB1	1200	1-119	6577872H1 (BRANDIT04)	676	1200
				71989128V1	578	1200
				7651017H1 (STOMTDE01)	1	610
				8008388H2 (NOSEDIC02)	866	1200
19	5844189CB1	10253	1-737	3450803H1 (UTRSNON03)	8896	9160
				6435007H1 (LUNGNON07)	1151	1719
				3825305H1 (BRAIHT01)	9385	9667
				702684R6 (SYNORAT03)	8605	9118
				7226621H1 (BRAXTDR15)	4608	5235
				71108830V1	1700	2266
				1412637F6 (BRAINOT12)	9965	10253
				3470805H1 (BRAIDIT01)	8546	8818
				1290463F6 (BRAINOT11)	9109	9655

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
				7231177H1 (BRAXTDR15)	5266	5828
				70837321V1	1197	1729
				71105936V1	2226	2809
				70484333V1	551	1173
				6438073H1 (BRAENOT02)	7710	8269
				2492760F6 (ADRETUT05)	1	560
				3147658H1 (PENCNOT05)	8386	8699
				6983064H1 (BRAIFER05)	2974	3510
				6762815J1 (BRAUNOR01)	6048	6610
				70485476V1	468	1133
				6950596H1 (BRAITDR02)	7485	8068
				6885353J1 (BRAHTDR03)	7282	7924
				6774466H1 (OVARDIR01)	3835	4509
				7628765H1 (GBLADIE01)	4768	5274
				1290463T6 (BRAINOT11)	9551	10232
				6985284H1 (BRAIFER05)	3295	3911
				6894193J1 (BRAITDR03)	5058	5669
				6463879H2 (OSTEUNC01)	8071	8466
				6887148J1 (BRAITDR03)	2357	3056
				6908894H1 (PITUDIR01)	4241	4709
				6870888H1 (BRAGNON02)	5734	6143
				6888951J1 (BRAITDR03)	6675	7305
				6992849H1 (BRAQTD02)	6439	7061
				71262033V1	1725	2293
20	7472720CB1	3851	165-225, 1040-1205, 2715-2955, 1662-1724	55136611T1	1388	1849
				7948427H1 (BRAENOE02)	1994	2535
				3071429H1 (UTRSNOR01)	2421	2714
				7117454H1 (BRAHNOE01)	1719	1873
				4110728F9 (PROSBPT07)	1	509
				71996691V1	3226	3851
				71998563V1	2974	3633
				6394784F8 (UTRENOT10)	242	971

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
				5963992T9 (BRATNOT05)	934	1430
				71722971V1	1757	2433
				GNN.g6425557_010.edit	1610	3386
21	7583990CB1	3100	1-300, 2838- 3100	5948913H1 (LIVRTUN04)	2802	3100
				6401188T8 (UTRENOT10)	1763	2526
				7153096H1 (BONEUNR01)	431	956
				3728642H1 (SMCCNON03)	2574	2858
				5632906R8 (PLACFER01)	1133	1558
				1458392T6 (COLNFET02)	1974	2558
				6808658H1 (SKIRNOR01)	627	1212
				7430535H1 (UTRMTMR02)	1	566
				412504F1 (BRSTNOT01)	2250	2833
				7076127H1 (BRAUTDR04)	1241	1823
22	2058182CB1	3248	1-313	g1548017	2790	3248
				70790685V1	2155	2752
				3853541F6 (BRAITUT12)	625	1155
				7977047H1 (LSUBDMC01)	1	644
				6936022H1 (SINTTMR02)	1206	1783
				70791096V1	1952	2567
				8013460H1 (HEARNOC04)	2688	3247
				70789630V1	1402	2003
				7039568H1 (UTRSTMR02)	772	1261
23	3564377CB1	2592	1-216, 1741- 2592	1807042F6 (SINTNOT13)	2257	2592
				7083969H1 (STOMTMR02)	1461	1994
				996489R6 (KIDNTUT01)	283	776
				1390744T6 (EOSINOT01)	1823	2469
				7720436J1 (THYRDIE01)	97	625
				6821354J1 (SINTNOR01)	1020	1719
				6178475H1 (BMARUNT02)	1	270
				7765524J1 (URETTUE01)	702	1359
24	1568689CB1	2004	1-51	7665791H1 (SPLNFEC01)	1	547
				1863491F6 (PROSNOT19)	1188	1758
				71955811V1	505	1093

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
				2782052F6 (OVRTUT03)	1685	2004
				71952924V1	658	1312
				1831510F6 (THP1AZT01)	1320	1927
25	1393767CB1	2250	1776-1842, 418-538	6318344F6 (LJUNGDI02)	1888	2249
				71413002V1	1266	1900
				3679624F9 (LJUNGNOT3)	1	608
				6431635H1 (LJUNGNO07)	2003	2250
				71412762V1	1193	1784
				7674549J1 (NOSETUE01)	666	1230
				1849430T6 (LJUNGFET03)	1621	2228
25	1393767CB1	2250	1776-1842, 418-538	71411963V1	575	1172
26	3029343CB1	3728	1807-1863, 2618-2675, 1- 76, 1614- 1694, 247- 912, 1510- 1554	GBI.g8072612_000002_000007_ 000006.regenscan	1	1481
				7594615F8 (LIIVRNOC07)	1901	2790
				72435607D1	2796	3343
				72435382D1	2662	3317
				7594495F8 (LIIVRNOC07)	1602	2444
				72435855D1	3255	3728
				GNN.g8072612_000006_004.edi t	595	1581
				6427666H1 (LJUNGNO07)	1212	1807
27	5507629CB1	2241	968-1091	4138803H1 (ADRENOT15)	1	282
				8133382H1 (SCOMD1C01)	1277	1989
				4249162F6 (BRADDIR01)	1934	2241
				8031936J1 (TESTNOF01)	615	1324
				4919304F6 (TESTNOT11)	878	1492
				5508394R6 (BRADDIR01)	1705	2226
				g6569543	1	442
				3595464H1 (FIBPNOT01)	319	636

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
28	5607780CB1	5203	3850-3946, 1- 305, 976- 1733, 2409- 3166	55148770J1	4451	5203
				6430080F8 (LUNGNON07)	3124	3882
				71136406V1	546	1194
				8246406H1 (BONEUNR01)	4205	4816
				71135385V1	704	1221
				7199141H1 (LUNGFER04)	2787	3349
				6916264H1 (PLACFER06)	1	631
				55105469H1	1613	2516
				5626906R8 (PLACFER01)	2037	2542
				7724527J1 (THYRDIE01)	3967	4785
				5626906F6 (PLACFER01)	1432	2034
				7714385J1 (SINTFEE02)	2383	3015
				8080767J1 (BONMTUE02)	3370	4094
				7722890J2 (THYRDIE01)	1095	1816

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
15	1806450CB1	KIDCTMT01
16	959690CB1	THYMNOE02
17	7091536CB1	BRAIFER05
18	7472724CB1	PROSNON01
19	5844189CB1	BRAYDIN03
20	7472720CB1	UTRENOT10
21	7583990CB1	BRSTNOT01
22	2058182CB1	CORPNOT02
23	3564377CB1	EOSINOT01
24	1568689CB1	BRAITUT21
25	1393767CB1	LUNGGIN02
26	3029343CB1	LIVRNOC07
27	5507629CB1	BRADDIR01
28	5607780CB1	THYRDIE01

Table 6

Library	Vector	Library Description
BRADDIR01	pINCY	Library was constructed using RNA isolated from diseased choroid plexus tissue of the lateral ventricle, removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident.
BRAIFER05	pINCY	Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation.
BRAITUT21	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the midline frontal lobe of a 61-year-old Caucasian female during excision of a cerebral meningeal lesion. Pathology indicated subfrontal meningotheelial meningioma with no atypia. One ethmoid and mucosal tissue sample indicated meningioma. Family history included cerebrovascular disease, senile dementia, hyperlipidemia, benign hypertension, atherosclerotic coronary artery disease, congestive heart failure, and breast cancer.
BRAYDIN03	pINCY	This normalized library was constructed from 6.7 million independent clones from a brain tissue library. Starting RNA was made from RNA isolated from diseased hypothalamus tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular accident. Patient history included Huntington's disease and emphysema. 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. The library was linearized and recircularized to select for insert containing clones.
BRSTNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the breast tissue of a 56-year-old Caucasian female who died in a motor vehicle accident.
CORPNOT02	pINCY	Library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
EOSINOT01	pINCY	Library was constructed using RNA isolated from microscopically normal eosinophils from 31 non-allergic donors. Donors abstained from prescription and over-the-counter drug use for at least one week prior to donating 200 ml of peripheral venous blood.
KIDCTMT01	pINCY	Library was constructed using RNA isolated from kidney cortex tissue removed from a 65-year-old male during nephroureterectomy. Pathology for the associated tumor tissue indicated grade 3 renal cell carcinoma within the mid-portion of the kidney and the renal capsule.
LIVRNOC07	pINCY	Library was constructed using pooled cDNA from two different donors. cDNA was generated using RNA isolated from liver tissue removed from a 20-week-old Caucasian male fetus who died from Patau's Syndrome (donor A) and a 16-week-old Caucasian female fetus who died from anencephaly (donor B). Family history included mitral valve prolapse in donor B.
LUNGDIN02	pINCY	This normalized lung tissue library was constructed from 7.6x10 ⁵ independent clones from a diseased lung tissue library. Starting RNA was made from RNA isolated from diseased lung tissue. Pathology indicated idiopathic pulmonary disease. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994)

Table 6 (cont.)

Library	Vector	Library Description
PROSNON01	PSPORT1	91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used. This normalized prostate library was constructed from 4.4 M independent clones from a prostate library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.
THYMNOE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from thymus tissue removed from a 3-year-old Hispanic male during a thymectomy and closure of a patent ductus arteriosus. The patient presented with severe pulmonary stenosis and cyanosis. Patient history included a cardiac catheterization and echocardiogram. Previous surgeries included Blalock-Taussig shunt and pulmonary valvotomy. The patient was not taking any medications. Family history included benign hypertension, osteoarthritis, depressive disorder, and extrinsic asthma in the grandparent(s).
THYRDIE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from diseased thyroid tissue removed from a 22-year-old Caucasian female during closed thyroid biopsy, partial thyroidectomy, and regional lymph node excision. Pathology indicated adenomatous hyperplasia. The patient presented with malignant neoplasm of the thyroid. Patient history included normal delivery, alcohol abuse, and tobacco abuse. Previous surgeries included myringotomy. Patient medications included an unspecified type of birth control pills. Family history included hyperlipidemia and depressive disorder in the mother; and benign hypertension, congestive heart failure, and chronic leukemia in the grandparent(s).
UTRENOT10	pINCY	Library was constructed using RNA isolated from pooled uterine endometrial tissue removed from three adult females during endometrial biopsy. Pathology indicated normal endometrium. All three patients were positive for Beta-3 integrin.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 Assembled ESTs: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	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 hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
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 > 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.	
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 Press, Menlo Park, CA, 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.	

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:1-14,
 - 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:1-14,
 - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, and
 - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.
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:15-28.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
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
- b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected
5 from the group consisting of SEQ ID NO:1-14.

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

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

- 10 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:15-28,
- 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:15-28,
- 15 c) a polynucleotide complementary to a polynucleotide of a),
- d) a polynucleotide complementary to a polynucleotide of b), and
- e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a
20 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:

- 25 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
30 present, the amount thereof.

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

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

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

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

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

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

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

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

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

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

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

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

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

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

a pharmaceutically acceptable excipient.

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

5

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
- 10 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:

- 15 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
20 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.

25 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,
- 30 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,
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.

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

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

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

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

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

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

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

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

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

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

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

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-14, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14.

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

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

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

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

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14 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:1-14 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-14 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:1-14.

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.

5 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 completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

10

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

15

 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.

 54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to
20 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 location have the same sequence, and each distinct physical location on the substrate contains
25 nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

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

30 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.

 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.

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

5 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.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

10

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.

15

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.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

20

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

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

25

72. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:17.

73. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:18.

74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:19.

30

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

76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:21.

77. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:22.

78. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:23.

5 79. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:24.

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

81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:26.

10

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

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

<110> INCYTE GENOMICS, INC.
 BAUGHN, Mariah R.
 YAO, Monique G.
 WALIA, Narinder K.
 GIETZEN, Kimberly J.
 THANGAVELU, Kavitha
 LU, Yan
 DING, Li
 YUE, Henry
 TANG, Y. Tom
 LAL, Preeti G.
 BATRA, Sajeev
 LU, Dyung Aina M.
 SANJANWALA, Madhu S.
 ARVIZU, Chandra
 RAMKUMAR, Jayalaxmi
 GRIFFIN, Jennifer A.
 GURURAJAN, Rajagopal
 AZIMZAI, Yalda
 XU, Yuming
 BURFORD, Neil

<120> CYTOSKELETON-ASSOCIATED PROTEINS

<130> PF-0828 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/244,022; 60/247,370; 60/251,831

<151> 2000-10-27; 2000-11-08; 2000-12-07

<160> 28

<170> PERL Program

<210> 1

<211> 580

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1806450CD1

<400> 1

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

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

	515	520	525
Asn His Ala Asp	Pro Asn Ser Arg Arg	Ser Arg His Arg Ser	Arg
	530	535	540
Ser Arg Ser Pro	Asp Ile Gln Ala Lys	Glu Glu Leu Trp Lys	His
	545	550	555
Ile Gln Lys Glu	Leu Val Asp Pro Ser	Gly Leu Ser Glu Glu	Gln
	560	565	570
Leu Lys Glu Ile	Pro Tyr Thr Lys Ile	Glu	
	575	580	

<210> 2

<211> 541

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 959690CD1

<400> 2

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

				275					280				285	
Thr	Val	Thr	Val	Arg	Val	Lys	Lys	Asn	Ser	Thr	Thr	Asp	Gln	Val
				290					295					300
Tyr	Gln	Ala	Ile	Ala	Ala	Lys	Val	Gly	Met	Asp	Ser	Thr	Thr	Val
				305					310					315
Asn	Tyr	Phe	Ala	Leu	Phe	Glu	Val	Ile	Ser	His	Ser	Phe	Val	Arg
				320					325					330
Lys	Leu	Ala	Pro	Asn	Glu	Phe	Pro	His	Lys	Leu	Tyr	Ile	Gln	Asn
				335					340					345
Tyr	Thr	Ser	Ala	Val	Pro	Gly	Thr	Cys	Leu	Thr	Ile	Arg	Lys	Trp
				350					355					360
Leu	Phe	Thr	Thr	Glu	Glu	Glu	Ile	Leu	Leu	Asn	Asp	Asn	Asp	Leu
				365					370					375
Ala	Val	Thr	Tyr	Phe	Phe	His	Gln	Ala	Val	Asp	Asp	Val	Lys	Lys
				380					385					390
Gly	Tyr	Ile	Lys	Ala	Glu	Glu	Lys	Ser	Tyr	Gln	Leu	Gln	Lys	Leu
				395					400					405
Tyr	Glu	Gln	Arg	Lys	Met	Val	Met	Tyr	Leu	Asn	Met	Leu	Arg	Thr
				410					415					420
Cys	Glu	Gly	Tyr	Asn	Glu	Ile	Ile	Phe	Pro	His	Cys	Ala	Cys	Asp
				425					430					435
Ser	Arg	Arg	Lys	Gly	His	Val	Ile	Thr	Ala	Ile	Ser	Ile	Thr	His
				440					445					450
Phe	Lys	Leu	His	Ala	Cys	Thr	Glu	Glu	Gly	Gln	Leu	Glu	Asn	Gln
				455					460					465
Val	Ile	Ala	Phe	Glu	Trp	Asp	Glu	Met	Gln	Arg	Trp	Asp	Thr	Asp
				470					475					480
Glu	Glu	Gly	Met	Ala	Phe	Cys	Phe	Glu	Tyr	Ala	Arg	Gly	Glu	Lys
				485					490					495
Lys	Pro	Arg	Trp	Val	Lys	Ile	Phe	Thr	Pro	Tyr	Phe	Asn	Tyr	Met
				500					505					510
His	Glu	Cys	Phe	Glu	Arg	Val	Phe	Cys	Glu	Leu	Lys	Trp	Arg	Lys
				515					520					525
Glu	Asn	Ile	Phe	Gln	Met	Ala	Arg	Ser	Gln	Gln	Arg	Asp	Val	Ala
				530					535					540

Thr

<210> 3

<211> 570

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7091536CD1

<400> 3

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

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

```

Lys Phe Val Leu Ser Val Leu Arg Leu Leu Leu Val Thr Met Gly
      500                      505                      510
Leu Leu Phe Val Leu Leu Leu Leu Leu Ile Ile Leu Thr Glu Ser
      515                      520                      525
Asp Leu Asp Ile Ala Phe Phe Arg Asp Ile Arg Gln Thr Pro Glu
      530                      535                      540
Phe Glu Gln Phe His Tyr Gln Tyr Phe Cys Pro Leu Arg Arg Trp
      545                      550                      555
Phe Ala Cys Lys Ile Arg Ser Val Val Ser Leu Leu Ile Asp Thr
      560                      565                      570

```

<210> 4
 <211> 163
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7472724CD1

```

<400> 4
Met Glu Asp Gly Lys Arg Glu Arg Trp Pro Thr Leu Met Glu Arg
  1                      5                      10                      15
Leu Cys Ser Asp Gly Phe Ala Phe Pro Gln Tyr Pro Ile Lys Pro
      20                      25                      30
Tyr His Leu Lys Arg Ile His Arg Ala Val Leu His Gly Asn Leu
      35                      40                      45
Glu Lys Leu Lys Tyr Leu Leu Leu Thr Tyr Tyr Asp Ala Asn Lys
      50                      55                      60
Arg Asp Arg Lys Glu Arg Thr Ala Leu His Leu Ala Cys Ala Thr
      65                      70                      75
Gly Gln Pro Glu Met Val His Leu Leu Val Ser Arg Arg Cys Glu
      80                      85                      90
Leu Asn Leu Cys Asp Arg Glu Asp Arg Thr Pro Leu Ile Lys Ala
      95                      100                     105
Val Gln Leu Arg Gln Glu Ala Cys Ala Thr Leu Leu Leu Gln Asn
      110                     115                     120
Gly Ala Asn Pro Asn Ile Thr Asp Phe Phe Gly Arg Thr Ala Leu
      125                     130                     135
His Tyr Ala Val Tyr Asn Glu Asp Thr Ser Met Ile Glu Lys Leu
      140                     145                     150
Leu Ser His Gly Thr Asn Ile Glu Glu Cys Ser Lys Val
      155                     160

```

<210> 5
 <211> 2803
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 5844189CD1

```

<400> 5
Met Asp Gly Val Ala Glu Phe Ser Glu Tyr Val Ser Glu Thr Val
  1                      5                      10                      15

```

Asp	Val	Pro	Ser	Pro	Phe	Asp	Leu	Leu	Glu	Pro	Pro	Thr	Ser	Gly	20	25	30
Gly	Phe	Leu	Lys	Leu	Ser	Lys	Pro	Cys	Cys	Tyr	Ile	Phe	Pro	Gly	35	40	45
Gly	Arg	Gly	Asp	Ser	Ala	Leu	Phe	Ala	Val	Asn	Gly	Phe	Asn	Ile	50	55	60
Leu	Val	Asp	Gly	Gly	Ser	Asp	Arg	Lys	Ser	Cys	Phe	Trp	Lys	Leu	65	70	75
Val	Arg	His	Leu	Asp	Arg	Ile	Asp	Ser	Val	Leu	Leu	Thr	His	Ile	80	85	90
Gly	Ala	Asp	Asn	Leu	Pro	Gly	Ile	Asn	Gly	Leu	Leu	Gln	Arg	Lys	95	100	105
Val	Ala	Glu	Leu	Glu	Glu	Gln	Ser	Gln	Gly	Ser	Ser	Ser	Ser	Tyr	110	115	120
Ser	Asp	Trp	Val	Lys	Asn	Leu	Ile	Ser	Pro	Glu	Leu	Gly	Val	Val	125	130	135
Phe	Phe	Asn	Val	Pro	Glu	Lys	Leu	Arg	Leu	Pro	Asp	Ala	Ser	Arg	140	145	150
Lys	Ala	Lys	Arg	Ser	Ile	Glu	Glu	Ala	Cys	Leu	Thr	Leu	Gln	His	155	160	165
Leu	Asn	Arg	Leu	Gly	Ile	Gln	Ala	Glu	Pro	Leu	Tyr	Arg	Val	Val	170	175	180
Ser	Asn	Thr	Ile	Glu	Pro	Leu	Thr	Leu	Phe	His	Lys	Met	Gly	Val	185	190	195
Gly	Arg	Leu	Asp	Met	Tyr	Val	Leu	Asn	Pro	Val	Lys	Asp	Ser	Lys	200	205	210
Glu	Met	Gln	Phe	Leu	Met	Gln	Lys	Trp	Ala	Gly	Asn	Ser	Lys	Ala	215	220	225
Lys	Thr	Gly	Ile	Val	Leu	Pro	Asn	Gly	Lys	Glu	Ala	Glu	Ile	Ser	230	235	240
Val	Pro	Tyr	Leu	Thr	Ser	Ile	Thr	Ala	Leu	Val	Val	Trp	Leu	Pro	245	250	255
Ala	Asn	Pro	Thr	Glu	Lys	Ile	Val	Arg	Val	Leu	Phe	Pro	Gly	Asn	260	265	270
Ala	Pro	Gln	Asn	Lys	Ile	Leu	Glu	Gly	Leu	Glu	Lys	Leu	Arg	His	275	280	285
Leu	Asp	Phe	Leu	Arg	Tyr	Pro	Val	Ala	Thr	Gln	Lys	Asp	Leu	Ala	290	295	300
Ser	Gly	Ala	Val	Pro	Thr	Asn	Leu	Lys	Pro	Ser	Lys	Ile	Lys	Gln	305	310	315
Arg	Ala	Asp	Ser	Lys	Glu	Ser	Leu	Lys	Ala	Thr	Thr	Lys	Thr	Ala	320	325	330
Val	Ser	Lys	Leu	Ala	Lys	Arg	Glu	Glu	Val	Val	Glu	Glu	Gly	Ala	335	340	345
Lys	Glu	Ala	Arg	Ser	Glu	Leu	Ala	Lys	Glu	Leu	Ala	Lys	Thr	Glu	350	355	360
Lys	Lys	Ala	Lys	Glu	Ser	Ser	Glu	Lys	Pro	Pro	Glu	Lys	Pro	Ala	365	370	375
Lys	Pro	Glu	Arg	Val	Lys	Thr	Glu	Ser	Ser	Glu	Ala	Leu	Lys	Ala	380	385	390
Glu	Lys	Arg	Lys	Leu	Ile	Lys	Asp	Lys	Val	Gly	Lys	Lys	His	Leu	395	400	405
Lys	Glu	Lys	Ile	Ser	Lys	Leu	Glu	Glu	Lys	Lys	Asp	Lys	Glu	Lys	410	415	420
Lys	Glu	Ile	Lys	Lys	Glu	Arg	Lys	Glu	Leu	Lys	Lys	Asp	Glu	Gly	425	430	435
Arg	Lys	Glu	Glu	Lys	Lys	Asp	Ala	Lys	Lys	Glu	Glu	Lys	Arg	Lys			

	440		445		450
Asp Thr Lys Pro	Glu Leu Lys Lys Ile	Ser Lys Pro Asp Leu Lys			
	455		460		465
Pro Phe Thr Pro	Glu Val Arg Lys Thr	Leu Tyr Lys Ala Lys Val			
	470		475		480
Pro Gly Arg Val	Lys Ile Asp Arg Ser	Arg Ala Ile Arg Gly Glu			
	485		490		495
Lys Glu Leu Ser	Ser Glu Pro Gln Thr	Pro Pro Ala Gln Lys Gly			
	500		505		510
Thr Val Pro Leu	Pro Thr Ile Ser Gly	His Arg Glu Leu Val Leu			
	515		520		525
Ser Ser Pro Glu	Asp Leu Thr Gln Asp	Phe Glu Glu Met Lys Arg			
	530		535		540
Glu Glu Arg Ala	Leu Leu Ala Glu Gln	Arg Asp Thr Gly Leu Gly			
	545		550		555
Asp Lys Pro Phe	Pro Leu Asp Thr Ala	Glu Glu Gly Pro Pro Ser			
	560		565		570
Thr Ala Ile Gln	Gly Thr Pro Pro Ser	Val Pro Gly Leu Gly Gln			
	575		580		585
Glu Glu His Val	Met Lys Glu Lys Glu	Leu Val Pro Glu Val Pro			
	590		595		600
Glu Glu Gln Gly	Ser Lys Asp Arg Gly	Leu Asp Ser Gly Ala Glu			
	605		610		615
Thr Glu Glu Glu	Lys Asp Thr Trp Glu	Glu Lys Lys Gln Arg Glu			
	620		625		630
Ala Glu Arg Leu	Pro Asp Arg Thr Glu	Ala Arg Glu Glu Ser Glu			
	635		640		645
Pro Glu Val Lys	Glu Asp Val Ile Glu	Lys Ala Glu Leu Glu Glu			
	650		655		660
Met Glu Glu Val	His Pro Ser Asp Glu	Glu Glu Glu Asp Ala Thr			
	665		670		675
Lys Ala Glu Gly	Phe Tyr Gln Lys His	Met Gln Glu Pro Leu Lys			
	680		685		690
Val Thr Pro Arg	Ser Arg Glu Ala Phe	Gly Gly Arg Glu Leu Gly			
	695		700		705
Leu Gln Gly Lys	Ala Pro Glu Lys Glu	Thr Ser Leu Phe Leu Ser			
	710		715		720
Ser Leu Thr Thr	Pro Ala Gly Ala Thr	Glu His Val Ser Tyr Ile			
	725		730		735
Gln Asp Glu Thr	Ile Pro Gly Tyr Ser	Glu Thr Glu Gln Thr Ile			
	740		745		750
Ser Asp Glu Glu	Ile His Asp Glu Pro	Glu Glu Arg Pro Ala Pro			
	755		760		765
Pro Arg Phe His	Thr Ser Thr Tyr Asp	Leu Pro Gly Pro Glu Gly			
	770		775		780
Ala Gly Pro Phe	Glu Ala Ser Gln Pro	Ala Asp Ser Ala Val Pro			
	785		790		795
Ala Thr Ser Gly	Lys Val Tyr Gly Thr	Pro Glu Thr Glu Leu Thr			
	800		805		810
Tyr Pro Thr Asn	Ile Val Ala Ala Pro	Leu Ala Glu Glu Glu His			
	815		820		825
Val Ser Ser Ala	Thr Ser Ile Thr Glu	Cys Asp Lys Leu Ser Ser			
	830		835		840
Phe Ala Thr Ser	Val Ala Glu Asp Gln	Ser Val Ala Ser Leu Thr			
	845		850		855
Ala Pro Gln Thr	Glu Glu Thr Gly Lys	Ser Ser Leu Leu Leu Asp			
	860		865		870

Thr	Val	Thr	Ser	Ile	Pro	Ser	Ser	Arg	Thr	Glu	Ala	Thr	Gln	Gly	875	880	885
Leu	Asp	Tyr	Val	Pro	Ser	Ala	Gly	Thr	Ile	Ser	Pro	Thr	Ser	Ser	890	895	900
Leu	Glu	Glu	Asp	Lys	Gly	Phe	Lys	Ser	Pro	Pro	Cys	Glu	Asp	Phe	905	910	915
Ser	Val	Thr	Gly	Glu	Ser	Glu	Lys	Arg	Gly	Glu	Ile	Ile	Gly	Lys	920	925	930
Gly	Leu	Ser	Gly	Glu	Arg	Ala	Val	Glu	Glu	Glu	Glu	Glu	Glu	Thr	935	940	945
Ala	Asn	Val	Glu	Met	Ser	Glu	Lys	Leu	Cys	Ser	Gln	Tyr	Gly	Thr	950	955	960
Pro	Val	Phe	Ser	Ala	Pro	Gly	His	Ala	Leu	His	Pro	Gly	Glu	Pro	965	970	975
Ala	Leu	Gly	Glu	Ala	Glu	Glu	Arg	Cys	Leu	Ser	Pro	Asp	Asp	Ser	980	985	990
Thr	Val	Lys	Met	Ala	Ser	Pro	Pro	Pro	Ser	Gly	Pro	Pro	Ser	Ala	995	1000	1005
Thr	His	Thr	Pro	Phe	His	Gln	Ser	Pro	Val	Glu	Glu	Lys	Ser	Glu	1010	1015	1020
Pro	Gln	Asp	Phe	Gln	Glu	Ala	Asp	Ser	Trp	Gly	Asp	Thr	Lys	Arg	1025	1030	1035
Thr	Pro	Gly	Val	Gly	Lys	Glu	Asp	Ala	Ala	Glu	Glu	Thr	Val	Lys	1040	1045	1050
Pro	Gly	Pro	Glu	Glu	Gly	Thr	Leu	Glu	Lys	Glu	Glu	Lys	Val	Pro	1055	1060	1065
Pro	Pro	Arg	Ser	Pro	Gln	Ala	Gln	Glu	Ala	Pro	Val	Asn	Ile	Asp	1070	1075	1080
Glu	Gly	Leu	Thr	Gly	Cys	Thr	Ile	Gln	Leu	Leu	Pro	Ala	Gln	Asp	1085	1090	1095
Lys	Ala	Ile	Val	Phe	Glu	Ile	Met	Glu	Ala	Gly	Glu	Pro	Thr	Gly	1100	1105	1110
Pro	Ile	Leu	Gly	Ala	Glu	Ala	Leu	Pro	Gly	Gly	Leu	Arg	Thr	Leu	1115	1120	1125
Pro	Gln	Glu	Pro	Gly	Lys	Pro	Gln	Lys	Asp	Glu	Val	Leu	Arg	Tyr	1130	1135	1140
Pro	Asp	Arg	Ser	Leu	Ser	Pro	Glu	Asp	Ala	Glu	Ser	Leu	Ser	Val	1145	1150	1155
Leu	Ser	Val	Pro	Ser	Pro	Asp	Thr	Ala	Asn	Gln	Glu	Pro	Thr	Pro	1160	1165	1170
Lys	Ser	Pro	Cys	Gly	Leu	Thr	Glu	Gln	Tyr	Leu	His	Lys	Asp	Arg	1175	1180	1185
Trp	Pro	Glu	Val	Ser	Pro	Glu	Asp	Thr	Gln	Ser	Leu	Ser	Leu	Ser	1190	1195	1200
Glu	Glu	Ser	Pro	Ser	Lys	Glu	Thr	Ser	Leu	Asp	Val	Ser	Ser	Lys	1205	1210	1215
Gln	Leu	Ser	Pro	Glu	Ser	Leu	Gly	Thr	Leu	Gln	Phe	Gly	Glu	Leu	1220	1225	1230
Asn	Leu	Gly	Lys	Glu	Glu	Met	Gly	His	Leu	Met	Gln	Ala	Glu	Asn	1235	1240	1245
Thr	Ser	His	His	Thr	Ala	Pro	Met	Ser	Val	Pro	Glu	Pro	His	Ala	1250	1255	1260
Ala	Thr	Ala	Ser	Pro	Pro	Thr	Asp	Gly	Thr	Thr	Arg	Tyr	Ser	Ala	1265	1270	1275
Gln	Thr	Asp	Ile	Thr	Asp	Asp	Ser	Leu	Asp	Arg	Lys	Ser	Pro	Ala	1280	1285	1290
Ser	Ser	Phe	Ser	His	Ser	Thr	Pro	Ser	Gly	Asn	Gly	Lys	Tyr	Leu			

1295	1300	1305
Pro Gly Ala Ile Thr Ser Pro Asp Glu His Ile Leu Thr Pro Asp		
1310	1315	1320
Ser Ser Phe Ser Lys Ser Pro Glu Ser Leu Pro Gly Pro Ala Leu		
1325	1330	1335
Glu Asp Ile Ala Ile Lys Trp Glu Asp Lys Val Pro Gly Leu Lys		
1340	1345	1350
Asp Arg Thr Ser Glu Gln Lys Lys Glu Pro Glu Pro Lys Asp Glu		
1355	1360	1365
Val Leu Gln Gln Lys Asp Lys Thr Leu Glu His Lys Glu Val Val		
1370	1375	1380
Glu Pro Lys Asp Thr Ala Ile Tyr Gln Lys Asp Glu Ala Leu His		
1385	1390	1395
Val Lys Asn Glu Ala Val Lys Gln Gln Asp Lys Ala Leu Glu Gln		
1400	1405	1410
Lys Gly Arg Asp Leu Glu Gln Lys Asp Thr Ala Leu Glu Gln Lys		
1415	1420	1425
Asp Lys Ala Leu Glu Pro Lys Asp Lys Asp Leu Glu Glu Lys Asp		
1430	1435	1440
Lys Ala Leu Glu Gln Lys Asp Lys Ile Pro Glu Glu Lys Asp Lys		
1445	1450	1455
Ala Leu Glu Gln Lys Asp Thr Ala Leu Glu Gln Lys Asp Lys Ala		
1460	1465	1470
Leu Glu Pro Lys Asp Lys Asp Leu Glu Gln Lys Asp Arg Val Leu		
1475	1480	1485
Glu Gln Lys Glu Lys Ile Pro Glu Glu Lys Asp Lys Ala Leu Asp		
1490	1495	1500
Gln Lys Val Arg Ser Val Glu His Lys Ala Pro Glu Asp Thr Val		
1505	1510	1515
Ala Glu Met Lys Asp Arg Asp Leu Glu Gln Thr Asp Lys Ala Pro		
1520	1525	1530
Glu Gln Lys His Gln Ala Gln Glu Gln Lys Asp Lys Val Ser Glu		
1535	1540	1545
Lys Lys Asp Gln Ala Leu Glu Gln Lys Tyr Trp Ala Leu Gly Gln		
1550	1555	1560
Lys Asp Glu Ala Leu Glu Gln Asn Ile Gln Ala Leu Glu Glu Asn		
1565	1570	1575
His Gln Thr Gln Glu Gln Glu Ser Leu Val Gln Glu Asp Lys Thr		
1580	1585	1590
Arg Lys Pro Lys Met Leu Glu Glu Lys Ser Pro Glu Lys Val Lys		
1595	1600	1605
Ala Met Glu Glu Lys Leu Glu Ala Leu Leu Glu Lys Thr Lys Ala		
1610	1615	1620
Leu Gly Leu Glu Glu Ser Leu Val Gln Glu Gly Arg Ala Arg Glu		
1625	1630	1635
Gln Glu Glu Lys Tyr Trp Arg Gly Gln Asp Val Val Gln Glu Trp		
1640	1645	1650
Gln Glu Thr Ser Pro Thr Arg Glu Glu Pro Ala Gly Glu Gln Lys		
1655	1660	1665
Glu Leu Ala Pro Ala Trp Glu Asp Thr Ser Pro Glu Gln Asp Asn		
1670	1675	1680
Arg Tyr Trp Arg Gly Arg Glu Asp Val Ala Leu Glu Gln Asp Thr		
1685	1690	1695
Tyr Trp Arg Glu Leu Ser Cys Glu Arg Lys Val Trp Phe Pro His		
1700	1705	1710
Glu Leu Asp Gly Gln Gly Ala Arg Pro His Tyr Thr Glu Glu Arg		
1715	1720	1725

Glu Ser Thr Phe Leu Asp Glu Gly Pro Asp Asp Glu Gln Glu Val	1730	1735	1740
Pro Leu Arg Glu His Ala Thr Arg Ser Pro Trp Ala Ser Asp Phe	1745	1750	1755
Lys Asp Phe Gln Glu Ser Ser Pro Gln Lys Gly Leu Glu Val Glu	1760	1765	1770
Arg Trp Leu Ala Glu Ser Pro Val Gly Leu Pro Pro Glu Glu Glu	1775	1780	1785
Asp Lys Leu Thr Arg Ser Pro Phe Glu Ile Ile Ser Pro Pro Ala	1790	1795	1800
Ser Pro Pro Glu Met Val Gly Gln Arg Val Pro Ser Ala Pro Gly	1805	1810	1815
Gln Glu Ser Pro Ile Pro Asp Pro Lys Leu Met Pro His Met Lys	1820	1825	1830
Asn Glu Pro Thr Thr Pro Ser Trp Leu Ala Asp Ile Pro Pro Trp	1835	1840	1845
Val Pro Lys Asp Arg Pro Leu Pro Pro Ala Pro Leu Ser Pro Ala	1850	1855	1860
Pro Gly Pro Pro Thr Pro Ala Pro Glu Ser His Thr Pro Ala Pro	1865	1870	1875
Phe Ser Trp Gly Thr Ala Glu Tyr Asp Ser Val Val Ala Ala Val	1880	1885	1890
Gln Glu Gly Ala Ala Glu Leu Glu Gly Gly Pro Tyr Ser Pro Leu	1895	1900	1905
Gly Lys Asp Tyr Arg Lys Ala Glu Gly Glu Arg Glu Glu Glu Gly	1910	1915	1920
Arg Ala Glu Ala Pro Asp Lys Ser Ser His Ser Ser Lys Val Pro	1925	1930	1935
Glu Ala Ser Lys Ser His Ala Thr Thr Glu Pro Glu Gln Thr Glu	1940	1945	1950
Pro Glu Gln Arg Glu Pro Thr Pro Tyr Pro Asp Glu Arg Ser Phe	1955	1960	1965
Gln Tyr Ala Asp Ile Tyr Glu Gln Met Met Leu Thr Gly Leu Gly	1970	1975	1980
Pro Ala Cys Pro Thr Arg Glu Pro Pro Leu Gly Ala Ala Gly Asp	1985	1990	1995
Trp Pro Pro Cys Leu Ser Thr Lys Glu Ala Ala Ala Gly Arg Asn	2000	2005	2010
Thr Ser Ala Glu Lys Glu Leu Ser Ser Pro Ile Ser Pro Lys Ser	2015	2020	2025
Leu Gln Ser Asp Thr Pro Thr Phe Ser Tyr Ala Ala Leu Ala Gly	2030	2035	2040
Pro Thr Val Pro Pro Arg Pro Glu Pro Gly Pro Ser Met Glu Pro	2045	2050	2055
Ser Leu Thr Pro Pro Ala Val Pro Pro Arg Ala Pro Ile Leu Ser	2060	2065	2070
Lys Gly Pro Ser Pro Pro Leu Asn Gly Asn Ile Leu Ser Cys Ser	2075	2080	2085
Pro Asp Arg Arg Ser Pro Ser Pro Lys Glu Ser Gly Arg Ser His	2090	2095	2100
Trp Asp Asp Ser Thr Ser Asp Ser Glu Leu Glu Lys Gly Ala Arg	2105	2110	2115
Glu Gln Pro Glu Lys Glu Ala Gln Ser Pro Ser Pro Pro His Pro	2120	2125	2130
Ile Pro Met Gly Ser Pro Thr Leu Trp Pro Glu Thr Glu Ala His	2135	2140	2145
Val Ser Pro Pro Leu Asp Ser His Leu Gly Pro Ala Arg Pro Ser			

2150	2155	2160
Leu Asp Phe Pro Ala Ser Ala Phe Gly Phe Ser Ser Leu Gln Pro		
2165	2170	2175
Ala Pro Pro Gln Leu Pro Ser Pro Ala Glu Pro Arg Ser Ala Pro		
2180	2185	2190
Cys Gly Ser Leu Ala Phe Ser Gly Asp Arg Ala Leu Ala Leu Ala		
2195	2200	2205
Pro Gly Pro Pro Thr Arg Thr Arg His Asp Glu Tyr Leu Glu Val		
2210	2215	2220
Thr Lys Ala Pro Ser Leu Asp Ser Ser Leu Pro Gln Leu Pro Ser		
2225	2230	2235
Pro Ser Ser Pro Gly Ala Pro Leu Leu Ser Asn Leu Pro Arg Pro		
2240	2245	2250
Ala Ser Pro Ala Leu Ser Glu Gly Ser Ser Ser Glu Ala Thr Thr		
2255	2260	2265
Pro Val Ile Ser Ser Val Ala Glu Arg Phe Ser Pro Ser Leu Glu		
2270	2275	2280
Ala Ala Glu Gln Glu Ser Gly Glu Leu Asp Pro Gly Met Glu Pro		
2285	2290	2295
Ala Ala His Ser Leu Trp Asp Leu Thr Pro Leu Ser Pro Ala Pro		
2300	2305	2310
Pro Ala Ser Leu Asp Leu Ala Leu Ala Pro Ala Pro Ser Leu Pro		
2315	2320	2325
Gly Asp Met Gly Asp Gly Ile Leu Pro Cys His Leu Glu Cys Ser		
2330	2335	2340
Glu Ala Ala Thr Glu Lys Pro Ser Pro Phe Gln Val Pro Ser Glu		
2345	2350	2355
Asp Cys Ala Ala Asn Gly Pro Thr Glu Thr Ser Pro Asn Pro Pro		
2360	2365	2370
Gly Pro Ala Pro Ala Lys Ala Glu Asn Glu Glu Ala Ala Ala Cys		
2375	2380	2385
Pro Ala Trp Glu Arg Gly Ala Trp Pro Glu Gly Ala Glu Arg Ser		
2390	2395	2400
Ser Arg Pro Asp Thr Leu Leu Ser Pro Glu Gln Pro Val Cys Pro		
2405	2410	2415
Ala Gly Gly Ser Gly Gly Pro Pro Ser Ser Ala Ser Pro Glu Val		
2420	2425	2430
Glu Ala Gly Pro Gln Gly Cys Ala Thr Glu Pro Arg Pro His Arg		
2435	2440	2445
Gly Glu Leu Ser Pro Ser Phe Leu Asn Pro Pro Leu Pro Pro Ser		
2450	2455	2460
Ile Asp Asp Arg Asp Leu Ser Thr Glu Glu Val Arg Leu Val Gly		
2465	2470	2475
Arg Gly Gly Arg Arg Arg Val Gly Gly Pro Gly Thr Thr Gly Gly		
2480	2485	2490
Pro Cys Pro Val Thr Asp Glu Thr Pro Pro Thr Ser Ala Ser Asp		
2495	2500	2505
Ser Gly Ser Ser Gln Ser Asp Ser Asp Val Pro Pro Glu Thr Glu		
2510	2515	2520
Glu Cys Pro Ser Ile Thr Ala Glu Ala Ala Leu Asp Ser Asp Glu		
2525	2530	2535
Asp Gly Asp Phe Leu Pro Val Asp Lys Ala Gly Gly Val Ser Gly		
2540	2545	2550
Thr His His Pro Arg Pro Gly His Asp Pro Pro Pro Leu Pro Gln		
2555	2560	2565
Pro Asp Pro Arg Pro Ser Pro Pro Arg Pro Asp Val Cys Met Ala		
2570	2575	2580

```

Asp Pro Glu Gly Leu Ser Ser Glu Ser Gly Arg Val Glu Arg Leu
      2585                      2590                      2595
Arg Glu Lys Glu Lys Val Gln Gly Arg Val Gly Arg Arg Ala Pro
      2600                      2605                      2610
Gly Lys Ala Lys Pro Ala Ser Pro Ala Arg Arg Leu Asp Leu Arg
      2615                      2620                      2625
Gly Lys Arg Ser Pro Thr Pro Gly Lys Gly Pro Ala Asp Arg Ala
      2630                      2635                      2640
Ser Arg Ala Pro Pro Arg Pro Arg Ser Thr Thr Ser Gln Val Thr
      2645                      2650                      2655
Pro Ala Glu Glu Lys Asp Gly His Ser Pro Met Ser Lys Gly Leu
      2660                      2665                      2670
Val Asn Gly Leu Lys Ala Gly Pro Met Ala Leu Ser Ser Lys Gly
      2675                      2680                      2685
Ser Ser Gly Ala Pro Val Tyr Val Asp Leu Ala Tyr Ile Pro Asn
      2690                      2695                      2700
His Cys Ser Gly Lys Thr Ala Asp Leu Asp Phe Phe Arg Arg Val
      2705                      2710                      2715
Arg Ala Ser Tyr Tyr Val Val Ser Gly Asn Asp Pro Ala Asn Gly
      2720                      2725                      2730
Glu Pro Ser Arg Ala Val Leu Asp Ala Leu Leu Glu Gly Lys Ala
      2735                      2740                      2745
Gln Trp Gly Glu Asn Leu Gln Val Thr Leu Ile Pro Thr His Asp
      2750                      2755                      2760
Thr Glu Val Thr Arg Glu Trp Tyr Gln Gln Thr His Glu Gln Gln
      2765                      2770                      2775
Gln Gln Leu Asn Val Leu Val Leu Ala Ser Ser Ser Thr Val Val
      2780                      2785                      2790
Met Gln Asp Glu Ser Phe Pro Ala Cys Lys Ile Glu Phe
      2795                      2800

```

<210> 6

<211> 1029

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472720CD1

<400> 6

```

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

```

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

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

Gln Met Lys Tyr Phe Leu Ser Ala Leu Pro Thr Arg Arg Asp Pro
 980 985 990
 Glu Ser Pro Cys Val Glu Asn Leu Thr Ser Ile Gly Leu Asn Arg
 995 1000 1005
 Lys Tyr Ile Pro Gln Thr Pro Ile Arg Ile Pro Ile Ser Ser Pro
 1010 1015 1020
 Gln Thr Ser Asn Asn Cys Lys Asn Ser
 1025

<210> 7

<211> 696

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7583990CD1

<400> 7

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

Asn Trp Leu Glu Gly Pro Gly Ser Glu Gln Leu Arg Ala Gln Trp	290	295	300
Gly Ile Gly Asp Ser Ile Arg Ala Ser Gln Ala Leu Gln Gln Lys	305	310	315
His Glu Glu Ile Glu Ser Gln His Ser Glu Trp Phe Ala Val Tyr	320	325	330
Val Glu Leu Asn Gln Gln Ile Ala Ala Leu Leu Asn Ala Gly Asp	335	340	345
Glu Glu Asp Leu Val Glu Leu Lys Ser Leu Gln Gln Gln Leu Ser	350	355	360
Asp Val Cys Tyr Arg Gln Ala Ser Gln Leu Glu Phe Arg Gln Asn	365	370	375
Leu Leu Gln Ala Ala Leu Glu Phe His Gly Val Ala Gln Asp Leu	380	385	390
Ser Gln Gln Leu Asp Gly Leu Leu Gly Met Leu Cys Val Asp Val	395	400	405
Ala Pro Ala Asp Gly Ala Ser Ile Gln Gln Thr Leu Lys Leu Leu	410	415	420
Glu Glu Lys Leu Lys Ser Val Asp Val Gly Leu Gln Gly Leu Arg	425	430	435
Glu Lys Gly Gln Gly Leu Leu Asp Gln Ile Ser Asn Gln Ala Ser	440	445	450
Trp Ala Tyr Gly Lys Asp Val Thr Ile Glu Asn Lys Glu Asn Val	455	460	465
Asp His Ile Gln Gly Val Met Glu Asp Met Gln Leu Arg Lys Gln	470	475	480
Arg Cys Glu Asp Met Val Asp Val Arg Arg Leu Lys Met Leu Gln	485	490	495
Met Val Gln Leu Phe Lys Cys Glu Glu Asp Ala Ala Gln Ala Val	500	505	510
Glu Trp Leu Ser Glu Leu Leu Asp Ala Leu Leu Lys Thr His Ile	515	520	525
Arg Leu Gly Asp Asp Ala Gln Glu Thr Lys Val Leu Leu Glu Lys	530	535	540
His Arg Lys Phe Val Asp Val Ala Gln Ser Thr Tyr Asp Tyr Gly	545	550	555
Arg Gln Leu Leu Gln Ala Thr Val Val Leu Cys Gln Ser Leu Arg	560	565	570
Cys Thr Ser Arg Ser Ser Gly Asp Thr Leu Pro Arg Leu Asn Arg	575	580	585
Val Trp Lys Gln Phe Thr Ile Ala Ser Glu Glu Arg Val His Arg	590	595	600
Leu Glu Met Ala Ile Ala Phe His Ser Asn Ala Glu Lys Ile Leu	605	610	615
Gln Asp Cys Pro Glu Glu Pro Glu Ala Ile Asn Asp Glu Glu Gln	620	625	630
Phe Asp Glu Ile Glu Ala Val Gly Lys Ser Leu Leu Asp Arg Leu	635	640	645
Thr Val Pro Val Val Tyr Pro Asp Gly Thr Glu Gln Tyr Phe Gly	650	655	660
Ser Pro Ser Asp Met Ala Ser Thr Ala Glu Asn Ile Arg Asp Arg	665	670	675
Met Lys Leu Val Asn Leu Lys Arg Gln Gln Leu Arg His Pro Glu	680	685	690
Met Val Thr Thr Glu Ser	695		

<210> 8
 <211> 803
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 2058182CD1

<400> 8

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

				350					355				360	
Trp	Arg	Thr	Cys	Gln	Lys	Leu	Pro	Pro	Gln	Asn	Phe	Val	Asn	Phe
				365					370					375
Arg	Tyr	Leu	Ile	Lys	Phe	Leu	Ala	Lys	Leu	Ala	Gln	Thr	Ser	Asp
				380					385					390
Val	Asn	Lys	Met	Thr	Pro	Ser	Asn	Ile	Ala	Ile	Val	Leu	Gly	Pro
				395					400					405
Asn	Leu	Leu	Trp	Ala	Arg	Asn	Glu	Gly	Thr	Leu	Ala	Glu	Met	Ala
				410					415					420
Ala	Ala	Thr	Ser	Val	His	Val	Val	Ala	Val	Ile	Glu	Pro	Ile	Ile
				425					430					435
Gln	His	Ala	Asp	Trp	Phe	Phe	Pro	Glu	Glu	Val	Glu	Phe	Asn	Val
				440					445					450
Ser	Glu	Ala	Phe	Val	Pro	Leu	Thr	Thr	Pro	Ser	Ser	Asn	His	Ser
				455					460					465
Phe	His	Thr	Gly	Asn	Asp	Ser	Asp	Ser	Gly	Thr	Leu	Glu	Arg	Lys
				470					475					480
Arg	Pro	Ala	Ser	Met	Ala	Val	Met	Glu	Gly	Asp	Leu	Val	Lys	Lys
				485					490					495
Glu	Ser	Pro	Pro	Lys	Pro	Lys	Asp	Pro	Val	Ser	Ala	Ala	Val	Pro
				500					505					510
Ala	Pro	Gly	Arg	Asn	Asn	Ser	Gln	Ile	Ala	Ser	Gly	Gln	Asn	Gln
				515					520					525
Pro	Gln	Ala	Ala	Ala	Gly	Ser	His	Gln	Leu	Ser	Met	Gly	Gln	Pro
				530					535					540
His	Asn	Ala	Ala	Gly	Pro	Ser	Pro	His	Thr	Leu	Arg	Arg	Ala	Val
				545					550					555
Lys	Lys	Pro	Ala	Pro	Ala	Pro	Pro	Lys	Pro	Gly	Asn	Pro	Pro	Pro
				560					565					570
Gly	His	Pro	Gly	Gly	Gln	Ser	Ser	Ser	Gly	Thr	Ser	Gln	His	Pro
				575					580					585
Pro	Ser	Leu	Ser	Pro	Lys	Pro	Pro	Thr	Arg	Ser	Pro	Ser	Pro	Pro
				590					595					600
Thr	Gln	His	Thr	Gly	Gln	Pro	Pro	Gly	Gln	Pro	Ser	Ala	Pro	Ser
				605					610					615
Gln	Leu	Ser	Ala	Pro	Arg	Arg	Tyr	Ser	Ser	Ser	Leu	Ser	Pro	Ile
				620					625					630
Gln	Ala	Pro	Asn	His	Pro	Pro	Pro	Gln	Pro	Pro	Thr	Gln	Ala	Thr
				635					640					645
Pro	Leu	Met	His	Thr	Lys	Pro	Asn	Ser	Gln	Gly	Pro	Pro	Asn	Pro
				650					655					660
Met	Ala	Leu	Pro	Ser	Glu	His	Gly	Leu	Glu	Gln	Pro	Ser	His	Thr
				665					670					675
Pro	Pro	Gln	Thr	Pro	Thr	Pro	Pro	Ser	Thr	Pro	Pro	Leu	Gly	Lys
				680					685					690
Gln	Asn	Pro	Ser	Leu	Pro	Ala	Pro	Gln	Thr	Leu	Ala	Gly	Gly	Asn
				695					700					705
Pro	Glu	Thr	Ala	Gln	Pro	His	Ala	Gly	Thr	Leu	Pro	Arg	Pro	Arg
				710					715					720
Pro	Val	Pro	Lys	Pro	Arg	Asn	Arg	Pro	Ser	Val	Pro	Pro	Pro	Pro
				725					730					735
Gln	Pro	Pro	Gly	Val	His	Ser	Ala	Gly	Asp	Ser	Ser	Leu	Thr	Asn
				740					745					750
Thr	Ala	Pro	Thr	Ala	Ser	Lys	Ile	Val	Thr	Asp	Ser	Asn	Ser	Arg
				755					760					765
Val	Ser	Glu	Pro	His	Arg	Ser	Ile	Phe	Pro	Glu	Met	His	Ser	Asp
				770					775					780

Ser Ala Ser Lys Asp Val Pro Gly Arg Ile Leu Leu Asp Ile Asp
785 790 795
Asn Asp Thr Glu Ser Thr Ala Leu
800

<210> 9

<211> 701

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3564377CD1

<400> 9

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

Gly	Ala	Ser	Val	Leu	Lys	Arg	Leu	Lys	Gln	Thr	Met	Ala	Ser	Asp			
				320					325					330			
Pro	His	Ser	Leu	Glu	Glu	Phe	Cys	Ser	Asp	Pro	His	Ala	Val	Ala			
				335					340					345			
Gly	Ala	Leu	Lys	Ser	Tyr	Leu	Arg	Glu	Leu	Pro	Glu	Pro	Leu	Met			
				350					355					360			
Thr	Phe	Asp	Leu	Tyr	Asp	Asp	Trp	Met	Arg	Ala	Ala	Ser	Leu	Lys			
				365					370					375			
Glu	Pro	Gly	Ala	Arg	Leu	Gln	Ala	Leu	Gln	Glu	Val	Cys	Ser	Arg			
				380					385					390			
Leu	Pro	Pro	Glu	Asn	Leu	Ser	Asn	Leu	Arg	Tyr	Leu	Met	Lys	Phe			
				395					400					405			
Leu	Ala	Arg	Leu	Ala	Glu	Glu	Gln	Glu	Val	Asn	Lys	Met	Thr	Pro			
				410					415					420			
Ser	Asn	Ile	Ala	Ile	Val	Leu	Gly	Pro	Asn	Leu	Leu	Trp	Pro	Pro			
				425					430					435			
Glu	Lys	Glu	Gly	Asp	Gln	Ala	Gln	Leu	Asp	Ala	Ala	Ser	Val	Ser			
				440					445					450			
Ser	Ile	Gln	Val	Val	Gly	Val	Val	Glu	Ala	Leu	Ile	Gln	Ser	Ala			
				455					460					465			
Asp	Thr	Leu	Phe	Pro	Gly	Asp	Ile	Asn	Phe	Asn	Val	Ser	Gly	Leu			
				470					475					480			
Phe	Ser	Ala	Val	Thr	Leu	Gln	Asp	Thr	Val	Ser	Asp	Arg	Leu	Ala			
				485					490					495			
Ser	Glu	Glu	Leu	Pro	Ser	Thr	Ala	Val	Pro	Thr	Pro	Ala	Thr	Thr			
				500					505					510			
Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Pro	Ala	Leu			
				515					520					525			
Ala	Ser	Ala	Ala	Thr	Lys	Glu	Arg	Thr	Glu	Ser	Glu	Val	Pro	Pro			
				530					535					540			
Arg	Pro	Ala	Ser	Pro	Lys	Val	Thr	Arg	Ser	Pro	Pro	Glu	Thr	Ala			
				545					550					555			
Ala	Pro	Val	Glu	Asp	Met	Ala	Arg	Arg	Thr	Lys	Arg	Pro	Ala	Pro			
				560					565					570			
Ala	Arg	Pro	Thr	Met	Pro	Pro	Pro	Gln	Val	Ser	Gly	Ser	Arg	Ser			
				575					580					585			
Ser	Pro	Pro	Ala	Pro	Pro	Leu	Pro	Pro	Gly	Ser	Gly	Ser	Pro	Gly			
				590					595					600			
Thr	Pro	Gln	Ala	Leu	Pro	Arg	Arg	Leu	Val	Gly	Ser	Ser	Leu	Arg			
				605					610					615			
Ala	Pro	Thr	Val	Pro	Pro	Pro	Leu	Pro	Pro	Thr	Pro	Pro	Gln	Pro			
				620					625					630			
Ala	Arg	Arg	Gln	Ser	Arg	Arg	Ser	Pro	Ala	Ser	Pro	Ser	Pro	Ala			
				635					640					645			
Ser	Pro	Gly	Pro	Ala	Ser	Pro	Ser	Pro	Val	Ser	Leu	Ser	Asn	Pro			
				650					655					660			
Ala	Gln	Val	Asp	Leu	Gly	Ala	Ala	Thr	Ala	Glu	Gly	Gly	Ala	Pro			
				665					670					675			
Glu	Ala	Ile	Ser	Gly	Val	Pro	Thr	Pro	Pro	Ala	Ile	Pro	Pro	Gln			
				680					685					690			
Pro	Arg	Pro	Arg	Ser	Leu	Ala	Ser	Glu	Thr	Asn							
				695					700								

<210> 10

<211> 354

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1568689CD1

<400> 10

```

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

```

<210> 11

<211> 605

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1393767CD1

<400> 11

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

```

Tyr Pro Lys Asn Phe Leu Thr Val Gly Asp Trp Thr Ala Arg Ile
      380      385      390
Trp Ser Glu Asp Ser Arg Glu Ser Ser Ile Met Trp Thr Lys Tyr
      395      400      405
His Met Ala Tyr Leu Thr Asp Ala Ala Trp Ser Pro Val Arg Pro
      410      415      420
Thr Val Phe Phe Thr Thr Arg Met Asp Gly Thr Leu Asp Ile Trp
      425      430      435
Asp Phe Met Phe Glu Gln Cys Asp Pro Thr Leu Ser Leu Lys Val
      440      445      450
Cys Asp Glu Ala Leu Phe Cys Leu Arg Val Gln Asp Asn Gly Cys
      455      460      465
Leu Ile Ala Cys Gly Ser Gln Leu Gly Thr Thr Thr Leu Leu Glu
      470      475      480
Val Ser Pro Gly Leu Ser Thr Leu Gln Arg Asn Glu Lys Asn Val
      485      490      495
Ala Ser Ser Met Phe Glu Arg Glu Thr Arg Arg Glu Lys Ile Leu
      500      505      510
Glu Ala Arg His Arg Glu Met Arg Leu Lys Glu Lys Gly Lys Ala
      515      520      525
Glu Gly Arg Asp Glu Glu Gln Thr Asp Glu Glu Leu Ala Val Asp
      530      535      540
Leu Glu Ala Leu Val Ser Lys Ala Glu Glu Glu Phe Phe Asp Ile
      545      550      555
Ile Phe Thr Glu Leu Lys Lys Lys Glu Ala Asp Ala Ile Lys Leu
      560      565      570
Thr Pro Val Pro Gln Gln Pro Ser Pro Glu Glu Asp Gln Val Val
      575      580      585
Glu Glu Gly Glu Glu Ala Ala Gly Glu Glu Gly Asp Glu Glu Val
      590      595      600
Glu Glu Asp Leu Ala
      605

```

<210> 12

<211> 1179

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3029343CD1

<400> 12

```

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

```

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

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

```

Met Glu Val Glu Glu Pro Val Ser Ser Ser Leu Ser Leu His Ile
      965                      970                      975
Ser Ser Tyr Gly Val Gln Ala Gly Thr Ser Ser Pro Gly Asp Ala
      980                      985                      990
Gln Glu Ser Val Leu Pro Pro Pro Val Ala Phe Ala Asp Ala Cys
      995                      1000                     1005
Asn Asp Lys Arg Glu Thr Phe Ser Ser His Phe Glu Arg Pro Tyr
      1010                     1015                     1020
Leu Gln Phe Lys Ala Glu Pro Pro Leu Thr Ser Arg Gly Phe Val
      1025                     1030                     1035
Glu Lys Ser Thr Ala Leu His Arg Ile Ser Ala Glu His Pro Glu
      1040                     1045                     1050
Ser Met Met Ser Glu Val His Glu Arg Ala Leu Gln Gln His Pro
      1055                     1060                     1065
Gln Ala Gln Gly Gly Leu Leu Glu Thr Arg Glu Lys Met His Ala
      1070                     1075                     1080
Asp Asn Asn Phe Thr Lys Thr Gln Asp Arg Leu His Ala Ser Ser
      1085                     1090                     1095
Asp Ala Phe Ser Gly Leu Arg Phe Gln Ser Gly Thr Ser Arg Gly
      1100                     1105                     1110
Tyr Gln Arg Gln Met Val Pro Arg Glu Glu Ile Lys Ser Thr Ser
      1115                     1120                     1125
Ala Lys Ser Ser Val Val Ser Leu Ala Asp Gln Ala Pro Asn Phe
      1130                     1135                     1140
Ser Arg Leu Leu Ser Asn Val Thr Val Met Glu Gly Ser Pro Val
      1145                     1150                     1155
Thr Leu Glu Val Glu Val Thr Gly Phe Pro Glu Pro Thr Leu Thr
      1160                     1165                     1170
Trp Trp Val Ala Tyr Asn Asp Lys Pro
      1175

```

<210> 13

<211> 372

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5507629CD1

<400> 13

```

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

```

Met	Phe	Phe	Glu	His	Leu	Gly	Val	Pro	Ala	Phe	Tyr	Met	Ser	Ile		
				125					130					135		
Gln	Ala	Val	Leu	Ala	Leu	Phe	Ala	Ala	Gly	Phe	Thr	Thr	Gly	Leu		
				140					145					150		
Val	Leu	Asn	Ser	Gly	Ala	Gly	Val	Thr	Gln	Ser	Val	Pro	Ile	Phe		
				155					160					165		
Glu	Gly	Tyr	Cys	Leu	Pro	His	Gly	Val	Gln	Gln	Leu	Asp	Leu	Ala		
				170					175					180		
Gly	Leu	Asp	Leu	Thr	Asn	Tyr	Leu	Met	Val	Leu	Met	Lys	Asn	His		
				185					190					195		
Gly	Ile	Met	Leu	Leu	Ser	Ala	Ser	Asp	Arg	Lys	Ile	Val	Glu	Asp		
				200					205					210		
Ile	Lys	Glu	Ser	Phe	Cys	Tyr	Val	Ala	Met	Asn	Tyr	Glu	Glu	Glu		
				215					220					225		
Met	Ala	Lys	Lys	Pro	Asp	Cys	Leu	Glu	Lys	Val	Tyr	Gln	Leu	Pro		
				230					235					240		
Asp	Gly	Lys	Val	Ile	Gln	Leu	His	Asp	Gln	Leu	Phe	Ser	Cys	Pro		
				245					250					255		
Glu	Ala	Leu	Phe	Ser	Pro	Cys	His	Met	Asn	Leu	Glu	Ala	Pro	Gly		
				260					265					270		
Ile	Asp	Lys	Ile	Cys	Phe	Ser	Ser	Ile	Met	Lys	Cys	Asp	Thr	Gly		
				275					280					285		
Leu	Arg	Asn	Ser	Phe	Phe	Ser	Asn	Ile	Ile	Leu	Ala	Gly	Gly	Ser		
				290					295					300		
Thr	Ser	Phe	Pro	Gly	Leu	Asp	Lys	Arg	Leu	Val	Lys	Asp	Ile	Ala		
				305					310					315		
Lys	Val	Ala	Pro	Ala	Asn	Thr	Ala	Val	Gln	Val	Ile	Ala	Pro	Pro		
				320					325					330		
Glu	Arg	Lys	Ile	Ser	Val	Trp	Met	Gly	Gly	Ser	Ile	Leu	Ala	Ser		
				335					340					345		
Leu	Ser	Ala	Phe	Gln	Asp	Met	Trp	Ile	Thr	Ala	Ala	Glu	Phe	Lys		
				350					355					360		
Glu	Val	Gly	Pro	Asn	Ile	Val	His	Gln	Arg	Cys	Phe					
				365					370							

<210> 14

<211> 1561

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5607780CD1

<400> 14

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

Gln	Ala	Pro	Gly	Ser	Ser	Thr	Leu	Ser	Ser	Ser	Leu	Cys	Arg	Asp
				95						100				105
Lys	Pro	Leu	Arg	Pro	Val	Ile	Leu	Ser	Pro	Thr	Met	Glu	Glu	Gly
				110						115				120
His	Gly	Leu	Asp	Leu	Thr	Tyr	Ile	Thr	Glu	Arg	Ile	Ile	Ala	Val
				125						130				135
Ser	Phe	Pro	Ala	Gly	Cys	Ser	Glu	Glu	Ser	Tyr	Leu	His	Asn	Leu
				140						145				150
Gln	Glu	Val	Thr	Arg	Met	Leu	Lys	Ser	Lys	His	Gly	Asp	Asn	Tyr
				155						160				165
Leu	Val	Leu	Asn	Leu	Ser	Glu	Lys	Arg	Tyr	Asp	Leu	Thr	Lys	Leu
				170						175				180
Asn	Pro	Lys	Ile	Met	Asp	Val	Gly	Trp	Pro	Glu	Leu	His	Ala	Pro
				185						190				195
Pro	Leu	Asp	Lys	Met	Cys	Thr	Ile	Cys	Lys	Ala	Gln	Glu	Ser	Trp
				200						205				210
Leu	Asn	Ser	Asn	Leu	Gln	His	Val	Val	Val	Ile	His	Cys	Arg	Gly
				215						220				225
Gly	Lys	Gly	Arg	Ile	Gly	Val	Val	Ile	Ser	Ser	Tyr	Met	His	Phe
				230						235				240
Thr	Asn	Val	Ser	Ala	Ser	Ala	Asp	Gln	Ala	Leu	Asp	Arg	Phe	Ala
				245						250				255
Met	Lys	Lys	Phe	Tyr	Asp	Asp	Lys	Val	Ser	Ala	Leu	Met	Gln	Pro
				260						265				270
Ser	Gln	Lys	Arg	Tyr	Val	Gln	Phe	Leu	Ser	Gly	Leu	Leu	Ser	Gly
				275						280				285
Ser	Val	Lys	Met	Asn	Ala	Ser	Pro	Leu	Phe	Leu	His	Phe	Val	Ile
				290						295				300
Leu	His	Gly	Thr	Pro	Asn	Phe	Asp	Thr	Gly	Gly	Val	Cys	Arg	Pro
				305						310				315
Phe	Leu	Lys	Leu	Tyr	Gln	Ala	Met	Gln	Pro	Val	Tyr	Thr	Ser	Gly
				320						325				330
Ile	Tyr	Asn	Val	Gly	Pro	Glu	Asn	Pro	Ser	Arg	Ile	Cys	Ile	Val
				335						340				345
Ile	Glu	Pro	Ala	Gln	Leu	Leu	Lys	Gly	Asp	Val	Met	Val	Lys	Cys
				350						355				360
Tyr	His	Lys	Lys	Tyr	Arg	Ser	Ala	Thr	Arg	Asp	Val	Ile	Phe	Arg
				365						370				375
Leu	Gln	Phe	His	Thr	Gly	Ala	Val	Gln	Gly	Tyr	Gly	Leu	Val	Phe
				380						385				390
Gly	Lys	Glu	Asp	Leu	Asp	Asn	Ala	Ser	Lys	Asp	Asp	Arg	Phe	Pro
				395						400				405
Asp	Tyr	Gly	Lys	Val	Glu	Leu	Val	Phe	Ser	Ala	Thr	Pro	Glu	Lys
				410						415				420
Ile	Gln	Gly	Ser	Glu	His	Leu	Tyr	Asn	Asp	His	Gly	Val	Ile	Val
				425						430				435
Asp	Tyr	Asn	Thr	Thr	Asp	Pro	Leu	Ile	Arg	Trp	Asp	Ser	Tyr	Glu

	515		520		525
Ala Pro Gly Thr	Arg Arg Gly Leu Ser	Ala Gln Glu Lys Ala Glu			
	530		535		540
Leu Asp Gln Leu	Leu Ser Gly Phe Gly	Leu Glu Asp Pro Gly Ser			
	545		550		555
Ser Leu Lys Glu	Met Thr Asp Ala Arg	Ser Lys Tyr Ser Gly Thr			
	560		565		570
Arg His Val Val	Pro Ala Gln Val His	Val Asn Gly Asp Ala Ala			
	575		580		585
Leu Lys Asp Arg	Glu Thr Asp Ile Leu	Asp Asp Glu Met Pro His			
	590		595		600
His Asp Leu His	Ser Val Asp Ser Leu	Gly Thr Leu Ser Ser Ser			
	605		610		615
Glu Gly Pro Gln	Ser Ala His Leu Gly	Pro Phe Thr Cys His Lys			
	620		625		630
Ser Ser Gln Asn	Ser Leu Leu Ser Asp	Gly Phe Gly Ser Asn Val			
	635		640		645
Gly Glu Asp Pro	Gln Gly Thr Leu Val	Pro Asp Leu Gly Leu Gly			
	650		655		660
Met Asp Gly Pro	Tyr Glu Arg Glu Arg	Thr Phe Gly Ser Arg Glu			
	665		670		675
Pro Lys Gln Pro	Gln Pro Leu Leu Arg	Lys Pro Ser Val Ser Ala			
	680		685		690
Gln Met Gln Ala	Tyr Gly Gln Ser Ser	Tyr Ser Thr Gln Thr Trp			
	695		700		705
Val Arg Gln Gln	Gln Met Val Val Ala	His Gln Tyr Ser Phe Ala			
	710		715		720
Pro Asp Gly Glu	Ala Arg Leu Val Ser	Arg Cys Pro Ala Asp Asn			
	725		730		735
Pro Gly Leu Val	Gln Ala Gln Pro Arg	Val Pro Leu Thr Pro Thr			
	740		745		750
Arg Gly Thr Ser	Ser Arg Val Ala Val	Gln Arg Gly Val Gly Ser			
	755		760		765
Gly Pro His Pro	Pro Asp Thr Gln Gln	Pro Ser Pro Ser Lys Ala			
	770		775		780
Phe Lys Pro Arg	Phe Pro Gly Asp Gln	Val Val Asn Gly Ala Gly			
	785		790		795
Pro Glu Leu Ser	Thr Gly Pro Ser Pro	Gly Ser Pro Thr Leu Asp			
	800		805		810
Ile Asp Gln Ser	Ile Glu Gln Leu Asn	Arg Leu Ile Leu Glu Leu			
	815		820		825
Asp Pro Thr Phe	Glu Pro Ile Pro Thr	His Met Asn Ala Leu Gly			
	830		835		840
Ser Gln Ala Asn	Gly Ser Val Ser Pro	Asp Ser Val Gly Gly Gly			
	845		850		855
Leu Arg Ala Ser	Ser Arg Leu Pro Asp	Thr Gly Glu Gly Pro Ser			
	860		865		870
Arg Ala Thr Gly	Arg Gln Gly Ser Ser	Ala Glu Gln Pro Leu Gly			
	875		880		885
Gly Arg Leu Arg	Lys Leu Ser Leu Gly	Gln Tyr Asp Asn Asp Ala			
	890		895		900
Gly Gly Gln Leu	Pro Phe Ser Lys Cys	Ala Trp Gly Lys Ala Gly			
	905		910		915
Val Asp Tyr Ala	Pro Asn Leu Pro Pro	Phe Pro Ser Pro Ala Asp			
	920		925		930
Val Lys Glu Thr	Met Thr Pro Gly Tyr	Pro Gln Asp Leu Asp Ile			
	935		940		945

Ile Asp Gly Arg Ile	Leu Ser Ser Lys Glu Ser Met Cys Ser Thr	950	955	960
Pro Ala Phe Pro Val	Ser Pro Glu Thr Pro Tyr Val Lys Thr Ala	965	970	975
Leu Arg His Pro Pro	Phe Ser Pro Pro Glu Pro Pro Leu Ser Ser	980	985	990
Pro Ala Ser Gln His	Lys Gly Gly Arg Glu Pro Arg Ser Cys Pro	995	1000	1005
Glu Thr Leu Thr His	Ala Val Gly Met Ser Glu Ser Pro Ile Gly	1010	1015	1020
Pro Lys Ser Thr Met	Leu Arg Ala Asp Ala Ser Ser Thr Pro Ser	1025	1030	1035
Phe Gln Gln Ala Phe	Ala Ser Ser Cys Thr Ile Ser Ser Asn Gly	1040	1045	1050
Pro Gly Gln Arg Arg	Glu Ser Ser Ser Ser Ala Glu Arg Gln Trp	1055	1060	1065
Val Glu Ser Ser Pro	Lys Pro Met Val Ser Leu Leu Gly Ser Gly	1070	1075	1080
Arg Pro Thr Gly Ser	Pro Leu Ser Ala Glu Phe Ser Gly Thr Arg	1085	1090	1095
Lys Asp Ser Pro Val	Leu Ser Cys Phe Pro Pro Ser Glu Leu Gln	1100	1105	1110
Ala Pro Phe His Ser	His Glu Leu Ser Leu Ala Glu Pro Pro Asp	1115	1120	1125
Ser Leu Ala Pro Pro	Ser Ser Gln Ala Phe Leu Gly Phe Gly Thr	1130	1135	1140
Ala Pro Val Gly Ser	Gly Leu Pro Pro Glu Glu Asp Leu Gly Ala	1145	1150	1155
Leu Leu Ala Asn Ser	His Gly Ala Ser Pro Thr Pro Ser Ile Pro	1160	1165	1170
Leu Thr Ala Thr Gly	Ala Ala Asp Asn Gly Phe Leu Ser His Asn	1175	1180	1185
Phe Leu Thr Val Ala	Pro Gly His Ser Ser His His Ser Pro Gly	1190	1195	1200
Leu Gln Gly Gln Gly	Val Thr Leu Pro Gly Gln Pro Pro Leu Pro	1205	1210	1215
Glu Lys Lys Arg Ala	Ser Glu Gly Asp Arg Ser Leu Gly Ser Val	1220	1225	1230
Ser Pro Ser Ser Ser	Gly Phe Ser Ser Pro His Ser Gly Ser Thr	1235	1240	1245
Ile Ser Ile Pro Phe	Pro Asn Val Leu Pro Asp Phe Ser Lys Ala	1250	1255	1260
Ser Glu Ala Ala Ser	Pro Leu Pro Asp Ser Pro Gly Asp Lys Leu	1265	1270	1275
Val Ile Val Lys Phe	Val Gln Asp Thr Ser Lys Phe Trp Tyr Lys	1280	1285	1290
Ala Asp Ile Ser Arg	Glu Gln Ala Ile Ala Met Leu Lys Asp Lys	1295	1300	1305
Glu Pro Gly Ser Phe	Ile Val Arg Asp Ser His Ser Phe Arg Gly	1310	1315	1320
Ala Tyr Gly Leu Ala	Met Lys Val Ala Thr Pro Pro Pro Ser Val	1325	1330	1335
Leu Gln Leu Asn Lys	Lys Ala Gly Asp Leu Ala Asn Glu Leu Val	1340	1345	1350
Arg His Phe Leu Ile	Glu Cys Thr Pro Lys Gly Val Arg Leu Lys	1355	1360	1365
Gly Cys Ser Asn Glu	Pro Tyr Phe Gly Ser Leu Thr Ala Leu Val			

1370	1375	1380
Cys Gln His Ser Ile Thr Pro Leu Ala Leu Pro Cys Lys Leu Leu		
1385	1390	1395
Ile Pro Glu Arg Asp Pro Leu Glu Glu Ile Ala Glu Ser Ser Pro		
1400	1405	1410
Gln Thr Ala Ala Asn Ser Ala Ala Glu Leu Leu Lys Gln Gly Ala		
1415	1420	1425
Ala Cys Asn Val Trp Tyr Leu Asn Ser Val Glu Met Glu Ser Leu		
1430	1435	1440
Thr Gly His Gln Ala Ile Gln Lys Ala Leu Ser Ile Thr Leu Val		
1445	1450	1455
Gln Glu Pro Pro Pro Val Ser Thr Val Val His Phe Lys Val Ser		
1460	1465	1470
Ala Gln Gly Ile Thr Leu Thr Asp Asn Gln Arg Lys Leu Phe Phe		
1475	1480	1485
Arg Arg His Tyr Pro Val Asn Ser Val Ile Phe Cys Ala Leu Asp		
1490	1495	1500
Pro Gln Asp Arg Lys Trp Ile Lys Asp Gly Pro Ser Ser Lys Val		
1505	1510	1515
Phe Gly Phe Val Ala Arg Lys Gln Gly Ser Ala Thr Asp Asn Val		
1520	1525	1530
Cys His Leu Phe Ala Glu His Asp Pro Glu Gln Pro Ala Ser Ala		
1535	1540	1545
Ile Val Asn Phe Val Ser Lys Val Met Ile Gly Ser Pro Lys Lys		
1550	1555	1560
Val		

<210> 15

<211> 2066

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1806450CB1

<400> 15

```

ggcggctggg cgcctcgggtg gtagctttct ctcctggctg gagacgacca caaccgacat 60
gggctgtttc tgcgctgttc cggaagaatt ttactgcgaa gttttgctcc tggatgaatc 120
caagttaacc cttaccaccc agcagcaggg catcaagaag tcaacgaaag gttccgttgt 180
ccttgaccac gtattccatc acgtaaacct tgtggagata gattattttg ggctacgtta 240
ctgtgacaga agccatcaga cgtattggct ggatcctgca aaaacccttg ctgaacacaa 300
agaactgatc aacactggac ctccatatac tttgtatttt ggtattaaat tctatgctga 360
agatccatgt aaacttaaag aagaaataac cagatatcag tttttcttgc aggtgaagca 420
agatgtcctt cagggccgctc tgccctgtcc cgtcaacact gctgctcagc tgggagcgta 480
tgccatccag tcggagcttg gagattatga cccatataaa catactgcag gatatgtatc 540
tgagtaccgg tttgttcctg atcagaagga agaacttgaa gaagccatag aaaggattca 600
taaaactcta atgggtcaga ttccttctga ggctgagctg aattacttga ggactgccaa 660
atccctggag atgtatggcg ttgacctcca tcccgtctat ggagaaaaca agtctgagta 720
tttcttagga ttaactccgg ttggtgttgt tgtgtacaag aataaaaagc aagtggggaa 780
gtattttctgg cctcggatta caaaggttca cttcaaggag actcaatttg aactcagagt 840
actgggaaaa gattgtaacg aaacctcatt cttttttgaa gctcggagta aaactgcttg 900
caagcacctc tggaagtgcg gtgtggaaca tcatacattt tttagaatgc cagaaaatga 960
atccaattca ctgtcaagaa aactcagcaa gtttgatcc atacgttata agcaccgcta 1020
cagtggcagg acagctttgc aaatgagccg agatctttct attcagcttc cccggcctga 1080
tcagaatgtg acaagaagtc gaagcaagac ttaccctaag cgaatagcac aaacacagcc 1140

```

```

agctgaatca aacacccatca gtaggataac tgcaaacaatg gaaaatggag aaaatgaagg 1200
aacaattaaa attattgcac cttcaccagt aaaaagcttt aagaaagcaa agaataaaaa 1260
tagccctgat acccaaagaa gcaaattctca tgcaccgtgg gaagaaaatg gccccagag 1320
tggaactctac aattctccca gtgatcgac taagtcgcca aagttccctt acacgcgtcg 1380
ccgaaacccc tcctgtggaa gtgacaatga ttctgtacag cctgtgagga ggaggaaagc 1440
ccataacagt ggtgaagatt cagatcttaa gcaaaggagg aggtcacgtt cagctgttaa 1500
caccagcagt ggtagtgaat cagaaaattc taatagagaa caccggaaaa agagaaacag 1560
aatacggcag gagaatgata tggttgattc agcgctcag tgggaagctg tattaaggag 1620
acaaaaggaa aaaaaccacg ccgaccccaa cagcaggcga tccagacaca gatctcgttc 1680
gagaagcccc gatatccaag caaaagaaga gttatggaag cacattcaaa aagaacttgt 1740
ggatccatcc ggattgtccg aagaacaatt aaaagagatt ccatacacta aaatagagtg 1800
agtgcctttc agaactcttc caccaaagct ttattagtgc ttgtgagtaa tccattctaa 1860
ttcttcaatt gtgttccaga cagtgcctta atttgtctt acattttaac caaaactagg 1920
tgacagtagc gaaaaggaga gaaaagtgtg cattaagct acttattcta cactataatc 1980
actatcatct cttattagcc acctcttgt acttggtagg tacaaggggg cttttcctga 2040
ttaatgtcag ttttaaaata gagtat 2066

```

<210> 16

<211> 1912

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 959690CB1

<400> 16

```

atggcgagc aggacgggga agggattcat ccctcagccc ctcacaggaa cggaggtggc 60
ggcgggcgcg gggggtctgg gctccactgc gccgggaacg gcggcggggg aggcggcggc 120
ccgcggtcg tgcgcacgt caagtccgag tccggctacg gcttcaacgt gcggggccaa 180
gtgagcgagg gcgggcaact gcggagcatc aacggggagc tgtacgcgcc gctgcagcat 240
gtgagcgccg tgctgcccgg gggggcgccc gatcgggccg ggggtgcgca gggggaccgc 300
atcctggagg tgaaccacgt gaatgttgag ggggcgacac acaagcaggt ggtggacctg 360
attcgagcag gcgagaagga attgatcttg acagtgttat ctgtacctcc tcatgaggca 420
gataacctag atcccagtg cgaactcgtt ggacaatcat tttatgatta cacagaaaag 480
caagcagtc ccataatcgg cccagatac aaacatgtgg agcagaatgg tgagaagttt 540
gtggtatata atgtttacat ggcagggagg cagctgtgtt ctaagcggta ccgggagttt 600
gctatcctac accagaacct gaagagagag tttgccaact ttacatttcc tcgactccca 660
gggaagtggc cattttcat atcagaacaa caattagatg cccgacgtcg gggattggaa 720
gaatatctag aaaaagtgtg ttcaatacga gtaattggtg agagtacat catgcaggaa 780
ttcctatcag aatccgatga gaactacaat ggtgtgtccg acgtagagct gagagtagca 840
ttaccagatg gaacaacggt tacagtcagg gttaaaaaga acagtactac agaccaagta 900
tatcaggcta tcgcagcaaa ggttggcatg gacagtacga cagtgaatta ctttgcctta 960
tttgaagtga tcagtcactc ctttgtacgt aaattggcac ctaatgagtt tcctcacaaa 1020
ctctacattc agaattatac atcagctgtg ccaggcacct gcttgacct tcgaaagtgg 1080
ctttttacaa cagaagaaga aattctctta aatgacaatg accttgctgt tacctacttc 1140
tttcatcagg cagtcgatga tgtgaagaaa ggttacatca aagcagaaga aaagtcctat 1200
caattacaga agctatacga acaaagaaaa atggtcatgt acctcaacat gctaaggact 1260
tgtgagggct acaatgaaat catctttccc cactgtgcct gtgactccag gaggaagggg 1320
cacgttatca cagccatcag catcacgcac tttaaactgc atgcctgcac tgaagaagga 1380
cagctggaga accaggtaat tgcatttgaa tgggatgaga tgcagcgatg ggacacagat 1440
gaagaaggga tggccttctg tttcgaatat gcacgaggag agaagaagcc ccgatgggtt 1500
aaaatcttca cgccatattt caattacatg catgagtgtc tcgagagggg gttctgcgag 1560
ctcaagtggg gaaaagagaa cttttccag atggcgagggt cacagcagag agatgtggcc 1620
acctagcctt tccttatccc cttcccttcc cttcaccccc atcctcttac tcctttcatg 1680
tccatttcca gacagagtaa ccattaacaa aaaagaagag aaaaagttaa agtcgttata 1740
ttcaaaagcc ctaaactaaa tattattaat aacccccctc gaatttcatg tctctggaat 1800

```

tgagggtgta gtgaacagca gatcgggtcag caccagaagt caactgagtt aaggcaggaa 1860
aagaaataag ccctttccag cacactgcgc cgtaactagt gtgccggctc ga 1912

<210> 17

<211> 2846

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7091536CB1

<400> 17

cggctcgaga tctgggctgg ggggaggcgg tggcggctga gggaaggagg aggataagga 60
ggaggaacga ggccagcagg aggcaacggc agcgacgggg ccgggggtgat ggtgcagggtg 120
cctggggctcg gtgctggagct gccgggctga gggacgcctg gtccagggtc cgcagcgccg 180
ccgcgtcgct cccggggcggg cgggaggaa gatgctgagc aggttgatga gcggcagcag 240
caggagcctg gagcgcgagt acagctgcac cgtgaggctg ctggacgaca gcgagtacac 300
ctgcaccatc cagagagatg ccaaaggcca gtacctgttt gaccttcttt gccaccatct 360
gaacctactt gagaaagact attttggtat ccgctttgta gaccagata agcagcggca 420
ttggctggaa tttacaaagt ctgtggtgaa acaattgaga tcccagcctc cattcaccat 480
gtgcttcctg gtgaagtgtt atcctgcaga cctgctgct ctgaaagaag aaataaccag 540
gtatttagtc ttcctgcaga tcaaaaggga tctctaccat ggccgactcc tctgtaaaac 600
atcggtgct ccttggttag cagcttacat cctcaagcg gagattgggg attatgactc 660
agtgaacac cctgaaggct acagctccaa gttccagttt ttccctaaac attcagagaa 720
gctggaaagg aaaattgctg agattcacaa gacggaactg agtgggtcaaa caccagcaac 780
atcagagctg aacttcttaa gaaaagcaca gacattggaa acatatggag tggatcctca 840
cccatgtaag gacgtgtcag gaaatgctgc atttctggcc ttcactcctt ttgggtttgt 900
tgttcttcaa ggaaacaaga ggggtccactt cattaaatgg aatgaggtga ccaagctgaa 960
atttgaagga aagactttct atttatacgt aagtcagaaa gaggaaaaga aaattattct 1020
tacatatctt gctccaactc ctgaagcgtg taagcacctc tggaaatgtg gaatcgagaa 1080
ccaagccttc tacaagctgg agaagtcaag ccaagtccgc acagtgtcca gcagcaattt 1140
attctttaaa gggagccggt tccgatacag tggccgagtt gcaaagggaag tcatggaatc 1200
aagtgtcaag atcaaacggg agccaccgga aatacacaga gcagggtatg ttcccagccg 1260
gagctgtccc tccataacct atggcccaag gctgagcagc gtcccagga cccgcagaag 1320
agctgttcac atctccatca tgggaaggcct agagtcctta cgggacagtg cccattccac 1380
accagtgcgt tccacttccc atggggacac ctctctgcct cacgtgagaa gcagccggac 1440
agatagcaat gagcgagtag ctgtgattgc agacgaggcc tacagccctg cagacagcgt 1500
gctgcccacc cctgtggctg agcacagcct ggagctgatg ttgctttccc ggcagatcaa 1560
tggagccacc tgcagcattg aggaggagaa ggaatctgaa gccagcacc caactgctac 1620
agaggtggag gcccttgggg gagagctgag ggccctgtgt caggggcaca gcgggcccga 1680
ggaggaacag gtgaataagt ttgttctaag tgtcctccgt ttgtccttg tgaccatggg 1740
actcctcttt gttttgctcc tctcctgat catccttacc gagtctgacc ttgacattgc 1800
ctttttccgt gatatccgcc agacccccga gtttgaacaa ttccactatc aatacttttg 1860
tcccctcagg cgatgggttg cctgcaaaat ccgctcagt gtgagcctgc tcattgacac 1920
ctgagaaggc atgactcctc ccaaaaacta gccagggtgga ccaaggaacc cggctaccca 1980
ttcccagcaa tgggacccat cgcggaacca tcggcacata taccaagtcc tcctctcatg 2040
actcaaagtc cactgcagcc taggaggtg tttcccagaa gaagaaagg atagggtcat 2100
gccctgtcta acaaaactgg gaaaactcat tttcttcaga agttatttca agaaaggctc 2160
agcgactctg tttctcatct ttccaatttg caggataatt tttgggtttt gaattttgat 2220
ttttcataga tgtatattat tttgaagtat caaataaaaa taattttatt tactattact 2280
gattattgca gctagtactc acctagcaga ggggacacta gttgaaaact agagagctgc 2340
tgtcctctgt attctgcagg agcttttct gctgggtgcca ctgggttcca gtagactcat 2400
cactgcagcc tcagcagggc aggccaggat ctggacaatg gggactgttt agttttttgt 2460
ttgttttttt tgccagccag aactttttaa aaagtaaaca tccatgtaga atgattaaat 2520
ggaaagtgtc ttcttatgat ggtctgagtt ggattttctt ttccttttgt tttttcaaat 2580
ctgagcagag tgggcatctg aagggaaccac cgctacagca acagcagcat caggggtggg 2640

```

gtaagtctgt ccccttagtc tagagcctag tgggcatcac catagttttg tataatgaaa 2700
ctagacttaa cgtgattttt tttttccgaa gaacctcaag acttttatag tgctccaggg 2760
gcgtaaacg aattcagggg taccagagat agatagttct gtcagaattg tgggacaaag 2820
tgtagttaag agaaagcaga gttaag 2846

```

<210> 18

<211> 1200

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472724CB1

<400> 18

```

gggagacagt taaacaggcc ctggggcagg gcgcgcctcg cgctccaggg agccccgccc 60
tcccgcggca cctccgcagc aaccgcccgc tgccactgggc gcgcgagagc tgctagggcg 120
gtttctctgc ctgggccttg ttgggcaggg ccggctaagg tgcgcgtgct cgctggttct 180
aacccttctg ttgggcgttt ctgctgagag gcgggaggcg ctgagagctt gtgcggaggt 240
ccgtggacag actgctttgc tcgttggtgc tcttcggagg cggcgatccc cgaaggcgag 300
ctgaaatacg gctgcaggct acaatttgca gccgacgatt atggaagacg gcaagcgga 360
gaggtggccc accctcatgg agcgcttggt ctcgatggc ttcgcatctc cccaatacc 420
cattaaaccg tatcatctga agaggatcca cagagctgtc ttacatggta atctagagaa 480
actgaagtac cttctgctca cgtattatga cgccaataag agagacagga aggaaaggac 540
cgccctacat ttggcctgtg ccaactggcca accggaaatg gtacatctcc tgggtgtccag 600
aagatgtgag cttaacctct gcgaccgtga agacaggaca cctctgatca aggtgttaca 660
actgaggcag gaggtctgtg caactcttct gctgcaaaat ggcgccaatc caaatattac 720
ggatttcttt ggaaggactg ctctgcacta cgctgtgtat aatgaagata catccatgat 780
agaaaaactt ctttcacatg gtacaaatat tgaagaatgc agcaaggat aggtcaacca 840
atgttatttt caaactatct gaaatgaatt tattttaaca ttgacacatg taagggtcaa 900
tttttcatat ttggaagctc aaacattcct tgaatgaaaa tattttgaaa tgccttaact 960
gtctaagatt ttacttttaa tattggaact tttaaagaag cattataggg aacagccttt 1020
tttcatgcac ttatggtaaa taactataaa aacaaatgaa ttacaataaa tttataattc 1080
atgacaactg aatttgggaa aggtaatagt taagtgtttt tccactaaat tacttttttt 1140
ctaatacgtg tgaagtgaca caggaaagta aaattgtccc ttataaatag gctttatttt 1200

```

<210> 19

<211> 10253

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5844189CB1

<400> 19

```

ttccctgaag ctgccggctg aggcgggagc tgccgcctcc atgagaggct tcctcctaca 60
ccccagggcc agaggaccct ttgccaccag agtgagatcc tagagaccat catcctggta 120
aatcccagtg cagacagcat cagctctgag gttcatcatc ttcttagcag ctcatcagct 180
tataaactac taatcttgag tgggcaaagt ttagagcctg ggggagacct catcctacag 240
agtggcacct actcatatga aaactttgcc caggctcctc acaaccccga gatttcccaa 300
ttgctcagca atagagacc cgggatacgg gccttcctta ccgtgtcctg cttaggggaa 360
ggtgattgga gccacctggg attatccagt tccaagaga ccctgcacct ccggctaaac 420
cctgagccca ctctgcccac catggacggc gtggctgagt tctccgagta tgtctctgag 480
actgtggaag tgccatcccc atttgaccta ctgagcccc ccacctcagg gggcttcctc 540
aagctctcca agccttggtg ctacatcttc ccagggtggc gtggggactc tgccctcttt 600
gctgtcaatg gtttcaacat cctggtggat ggtggctctg atcgcaagtc ctgttttttg 660

```

aagctggtac	ggcacttggga	ccgcattgac	tccgtgctac	tcacacacat	tggggcagac	720
aacctgccag	gcatcaatgg	actactgcag	cgcaaagtgg	cagagctaga	ggaggagcag	780
tcccagggct	ctagcagtta	cagcgactgg	gtgaagaacc	ttatctctcc	tgagcttggga	840
gttgtctttt	tcaacgtgcc	tgagaagctg	cggcttcctg	atgcctcccg	gaaagccaag	900
cgtagcattg	aggaggcctg	cctcactctg	cagcacttaa	accgcctggg	catccaggct	960
gagcctctat	atcgtgtggt	cagcaatacc	attgagccac	tgaccctctt	ccacaaaatg	1020
ggtgtggggc	ggctggacat	gtatgtcctc	aaccctgtca	aggacagcaa	ggagatgcag	1080
ttcctcatgc	aaaagtgggc	aggcaatagt	aaagccaaga	caggcatcgt	gctgcccatt	1140
gggaaggagg	ctgagatctc	cgtgccctac	cttacctcta	tcactgctct	ggtggtctgg	1200
ctaccagcca	atcccactga	gaagattgtg	cgtgtgcttt	ttccaggaaa	tgctccccc	1260
aacaagatct	tggagggcct	agaaaagctt	cggcatctgg	acttctctcg	ttaccctgtg	1320
gccacgcaga	aggacctggc	ttctggggct	gtgcctacca	acctcaagcc	cagcaaaatg	1380
aaacagcggg	ctgatagcaa	ggagagcctc	aaagccacta	ccaagacggc	cgtgagcaag	1440
ttggccaaac	gggaggaggt	ggtagaagag	ggagccaagg	aggcacgttc	agagctggcc	1500
aaggagttag	ccaagacaga	gaagaaggca	aaagagtcac	ctgagaagcc	cccagagaag	1560
cctgccaaag	ctgagagggt	gaagacagag	tcaagtggag	cactgaaggc	agagaagcga	1620
aagctgatca	aagacaaggt	agggaaaaag	caccttaaag	aaaagatatc	aaagctggaa	1680
gaaaaaaaaa	acaaggagaa	aaaagagatc	aaaaaggaga	ggaaagagct	caagaaggat	1740
gaaggaagga	aggaggagaa	gaaggatgcc	aagaaggagg	agaagaggaa	agataccaaa	1800
cctgagctca	agaagatttc	caagccagac	ctaaagccct	ttactcctga	ggtacgtaag	1860
acctcttata	aagccaaggt	ccctggaaga	gtcaaaatag	acaggagccg	tgctatccgt	1920
ggggagaagg	agctgtcttc	tgagccccag	acacccccag	cccagaaggg	aactgtacca	1980
ctccaacca	tcagtgggca	cagggagctg	gtcctatcct	caccagagga	cctcacacag	2040
gactttgagg	agatgaagcg	tgaggagagg	gctttgctgg	ctgaacaaa	ggacacagga	2100
ctaggagata	agccattccc	tctagacact	gcagaggagg	gacccccaa	tacagctatc	2160
cagggaaac	caccctctgt	tccagggctg	ggacaagaag	aacatgtgat	gaaggagaaa	2220
gagcttgtcc	cagaggctcc	tgaggaacaa	ggcagcaagg	acagaggcct	agactctggg	2280
gctgaaacag	aggaagagaa	agatacctgg	gaggaaaaga	agcagagggg	agcagagagg	2340
ctcccagaca	gaacagaagc	cagagaggaa	agtgaacctg	aagtaaagga	ggatgtgata	2400
gaaaaggctg	agttagaaga	aatggaggag	gtacaccctt	cagatgagga	ggaagaggac	2460
gcgacaaaag	ctgagggttt	ttacccaaaa	catatgcagg	aacccttgaa	ggtaactcca	2520
aggagccggg	aggcttttgg	gggtcgggaa	ttgggactcc	agggcaaggc	ccctgagaag	2580
gagacctcgt	tattcctaag	cagcctgacc	acacctgcag	gagccactga	gcatgtctct	2640
tacatccagg	atgagacaat	ccctggctac	tcagagactg	agcagaccat	ctcagatgag	2700
gagatccatg	atgagccgga	ggagcgccca	gctccaccca	gatttcatac	aagtacatat	2760
gacctgcccg	ggcctgaagg	tgctggccca	ttcgaaagcca	gccaacctgc	cgatagtgtc	2820
gttcctgcta	cctctggcaa	agtctatgga	acgccagaga	ctgaactcac	ctacccact	2880
aacatagtgg	ctgccccctt	ggctgaagag	gaacatgtgt	cctcggccac	ttcaatcact	2940
gagtgtgaca	aactttcttc	ctttgccaca	tcagtggctg	aggaccaatc	tgtggcctca	3000
cttacagctc	cccagacaga	ggagacaggc	aagagctccc	tgctgcttga	cacagtcaca	3060
agcatccctt	cctcccgtac	tgaagctacg	cagggcttgg	actatgtgcc	atcagctggg	3120
accatctcac	ccacctcctc	actggaagaa	gacaagggct	tcaaatcacc	acctgtgag	3180
gacttctctg	tgactgggga	gtcagagaag	agaggagaga	tcatagggaa	aggcttgtct	3240
ggagagagag	ctgtggaaga	ggaagaggag	gagacagcaa	acgtagagat	gtctgagaaa	3300
ctttgcagtc	aatatggaac	tccagtgttt	agtggccctg	ggcatgccct	acatccagga	3360
gaaccagccc	ttggagaagc	agaggagcgg	tgcccttagcc	cagatgacag	cacagtgaag	3420
atggcttctc	ctccaccatc	tggcccaccc	agtgccaccc	acacaccctt	tcacagtcc	3480
ccagtggaa	aaaagtctga	gccccaaagc	tttcaggagg	cagactcctg	gggagacact	3540
aagcgcacac	caggtgtggg	caaagaagat	gctgctgagg	agacagtcaa	gccagggcct	3600
gaagagggca	cactagagaa	ggaagagaaa	gttcctcctc	ccaggagccc	ccaggcccag	3660
gaagcacctg	tcaacattga	tgaggggctt	acaggctgta	ccattcaact	gttgccagca	3720
caggataaag	caatagtctt	tgagattatg	gaggcaggag	agcccacagg	cccaattctg	3780
ggagcagaag	cccttcccg	aggtttgagg	actttacccc	aagaacctgg	caaacctcag	3840
aaagatgagg	tgctcagata	tcctgaccga	agcctctctc	ctgaagatgc	agaatccctc	3900
tctgtctctc	gcgtgccctc	cccagacact	gccaaacca	agcctacccc	caagtctccc	3960
tgtggcctga	cagaacagta	cctacacaaa	gaccgttggc	cagaggtatc	tccagaagac	4020
accagtcac	tttctctgtc	agaagagagt	cccagcaagg	agacctccct	ggatgtctct	4080

tctaagcagc	tctctccaga	aagccttggc	accctccagt	ttggggaact	aaaccttggg	4140
aaggaagaaa	tgggycatct	gatgcaggcc	gagaacacct	ctcaccacac	agctcccatg	4200
tctgttccag	agccccatgc	agccacagcg	tcacctccca	cagatgggac	aactcgatac	4260
tctgcacaga	cagacatcac	agatgacagc	cttgacagga	agtcacctgc	cagctcattc	4320
tctcactcta	caccttcagg	aaatgggaag	tacttacctg	gggcgatcac	aagccctgat	4380
gaacacattc	tgacacctga	tagctccttc	tccaagagtc	ctgagtcctt	gccaggccct	4440
gccttggagg	acattgccat	aaagtgggaa	gataaagtcc	caggggttgaa	agacagaacc	4500
tcagaacaga	agaaggaacc	tgagccaaag	gatgaagttt	tacagcagaa	agacaaaact	4560
ctggagcaca	aggaggtggt	agagccgaag	gatacagcca	tctatcagaa	agatgaggct	4620
ctgcatgtaa	agaatgaggc	tgtgaaacag	caggataagg	ctttagaaca	aaagggcaga	4680
gacttagagc	aaaaagacac	agccctagaa	cagaaggaca	aggccctgga	accaaaagac	4740
aaagacttag	aagaaaaaga	caaggccctg	gaacagaagg	ataagattcc	agaagagaaa	4800
gacaaagcct	tagaacaaaa	ggatacagcc	ctggaaacaga	aggacaaggc	cctggaacca	4860
aaagataaag	acttggaaca	aaaggacagg	gtcctagAAC	agaaggagaa	gatcccagaa	4920
gagaaagaca	aagccttaga	tcaaaaagtc	agaagtgttg	aacataaggc	tccggaggac	4980
acggtcgcgtg	aaatgaagga	cagagaccta	gaacagacag	acaaagcccc	tgaacagaaa	5040
caccaggccc	aggaacaaaa	ggataaagtc	tcagaaaaga	aggatcaggc	cttagaacca	5100
aaatactggg	ctttgggaca	gaaggatgaa	gccctggaac	aaaacattca	ggctctggaa	5160
gagaaccacc	aaactcagga	gcaggagagc	ctagtgcagg	aggataaaac	caggaaacca	5220
aagatgctag	aggaaaaatc	cccagaaaag	gtcaaggcca	tggaagagaa	gttagaagct	5280
cttctggaga	agaccaaagc	tctgggcctg	gaagagagcc	tagtgcagga	gggcaggggc	5340
agagagcagg	aagaaaagta	ctggaggggg	caggatgtgg	tccaggagtg	gcaagaaaca	5400
tctcctacca	gagaggagcc	ggctggagaa	cagaaagagc	ttgccccggc	atgggaggac	5460
acatctcctg	agcaggacaa	taggtattgg	aggggcagag	aggatgtggc	cttggaaacag	5520
gacacatact	ggagggagct	aagctgtgag	cggaaggctc	ggttccctca	cgagctggat	5580
ggccaggggg	cccgccacac	ctacactgag	gaacgggaaa	gcactttcct	agatgagggc	5640
ccagatgatg	agcaagaagt	accctgcgg	gaacacgcaa	cccggagccc	ctgggcctca	5700
gacttcaagg	atttccagga	atcctcacca	cagaaggggc	tagaggtgga	gcgctggcct	5760
gctgaatcac	cagttgggtt	gccaccagag	gaagaggaca	aactgaccgc	ctctcccttt	5820
gagatcatct	cccctccagc	ttccccacct	gagatgggtg	gacaaaggg	tccttcagcc	5880
ccaggacaag	agagtccctat	cccagaccct	aagctcatgc	cacacatgaa	gaatgaacct	5940
actactccct	catggctggc	tgacatccca	ccctgggtgc	ccaaggacag	accctccccc	6000
cctgcacccc	tctccccagc	tcctgggtccc	cccacacctg	ccccggaatc	ccatactcct	6060
gcacccttct	cttggggcac	agccgagtat	gacagtgtgg	tggctgcagt	gcaggagggg	6120
gcagctgagt	tgggaaggtg	gccatactcc	cccctgggga	aggactaccg	caaggctgaa	6180
ggggaaagg	aagaagaagg	tagggctgag	gctcctgaca	aaagctcaca	cagctcaaag	6240
gtaccagagg	ccagcaaaaag	ccatgccacc	acggagcctg	agcagactga	gccggagcag	6300
agagagccca	cacctatctc	tgatgagaga	agctttcagt	atgcagacat	ctatgagcag	6360
atgatgctta	ctgggcttgg	ccctgcatgc	cccactagag	agcctccact	tggagcagct	6420
ggggattggc	ccccatgcct	ctcaaccaag	gaggcagctg	ccggccgaaa	cacatctgca	6480
gagaaggagc	tttcatctcc	tatctcacc	aagagcctcc	agtctgacac	tccaaccttc	6540
agctatgcag	ccctggcagg	accactgtta	cccccaaggc	cagagccagg	gccaagtatg	6600
gagcccagcc	tcaccccacc	tgcagttccc	ccccgtgctc	ctatcctgag	caaaggccca	6660
agccccctc	ttaatggtaa	catcctgagc	tgcagcccag	ataggaggct	cccatcccc	6720
aaggaatcag	gccggagtc	ctgggatgac	agcactagt	actcagaact	ggagaagggg	6780
gctcgggaac	agccagaaaa	agaggcccaa	tccccaaagt	ctcctcacc	cattcctatg	6840
gggtccccc	cattatggcc	agaaactgag	gcacatgtta	gccctccctt	ggactcacac	6900
ctggggcctg	cccgaaccag	tctggacttc	cctgcttcag	cctttggctt	ctcctcattg	6960
cagccagctc	ccccacagct	gccctctcca	gctgaacccc	gctcggcacc	ctgtggtctc	7020
cttgccctct	ctggggatcg	agctctggct	ctggctccag	gacccccac	cagaaccccg	7080
catgatgaat	acctggaagt	gaccaaggcc	cccagcctgg	attcctcact	gccccagctc	7140
ccatcaccca	gttctcctgg	ggccctctct	ctctccaatc	tgccacgacc	tgctcacca	7200
gccctgtctg	agggctcctc	ctctgaggct	accacgcctg	tgatttcaag	tgtggcggag	7260
cgcttctctc	caagccttga	ggctgcagaa	caggagtctg	gagagctgga	cccaggaatg	7320
gaaccagctg	cccacagcct	ctgggacctc	actcctctga	gccagcacc	cccagcttca	7380
ctggacttgg	ccctagctcc	agctccaagc	ctgcctggag	acatgggtga	tggcatcctg	7440
ccgtgccacc	tggagtgtct	agaggcagcc	acggagaagc	caagccctt	ccaggttccc	7500

```

tctgaggatt gtgcagccaa tggcccaact gaaaccagcc ctaaccccc aggcctgcc 7560
ccagccaagg ctgaaaatga agaggctgcg gcttgccctg cctgggaacg tggggcctgg 7620
cctgaaggag ctgagaggag ctcccggcct gacacattgc tctccccctga gcagccagtg 7680
tgtcctgcag ggggctccgg gggcccaccc agcagtgcct ctccctgaggt cgaagctggg 7740
ccccagggat gtgccactga gcctcggccc catcgtgggg agctctcccc atccttcctg 7800
aaccacctc tgcccccatc catagatgat agggacctct caactgagga agttcggtta 7860
gtaggaagag gggggcgggcg ccgggtaggg gggccaggga ccaactgggg cccatgccct 7920
gtgactgatg agacaccccc tacatcagcc agtgactcag gctcctcaca gtcagattct 7980
gatgtcccg cagaaactga ggagtgtccg tccatcacag ctgaggcagc cctcgactca 8040
gatgaagatg gagacttcct acctgtggac aaagctgggg gtgtcagtgg tactcaccac 8100
cccaggcctg gccatgacct acctcctctc ccacagccag acccccgcct atccctccc 8160
cgccctgatg tgtgcatggc tgaccccgag gggctcagct cagagtctgg gagagttag 8220
aggctacggg agaaggaaaa ggttcagggg cgagtgggg gcaggggccc aggcaggcc 8280
aagccagcgt cccctgcacg gcgtctggat ctctggggaa aacgctcacc caccctggt 8340
aaagggcctg cagatcgagc atcccgggcc ccacctcgac cacgcagcac cacaagccag 8400
gtcacccag cagaggaaaa ggatggacac agcccatgt ccaaaggcct agtcaatgga 8460
ctcaaggcag gaccaatggc cttgagttcc aagggcagct ctgggtgccc tgtatatgtg 8520
gatctgcct acatcccgaa tcattgcagt ggcaagactg ctgacctga cttcttccgt 8580
cgagtgcgtg catcctacta tgtgttcagt gggaatgacc ctgccaatgg cgagccaagc 8640
cgggctgtgc tggatgccct gctggagggc aaggcccagt ggggggagaa tcttcaggtg 8700
actctgatcc ctactcatga cacggaggtg actcgtgagt ggtaccaaca aactcatgag 8760
cagcagcaac aactgaatgt cctggtcctg gctagcagca gcaccgtgg gatgcaggat 8820
gagtccttcc ctgcctgcaa gattgagttc tgaaagagcc gccctccctt cccaaggat 8880
ccactcccc agctccttta gagaatggct actgctgagt cctttggggg tgagggagat 8940
gggagctagg gggaggggag ggagatgtct tgttgtgggg acttgggctg ggctaaatgg 9000
gaggggttgt ccctccccat catccattcc tgtgaggtgt ctcaaacc aaagtaacagg 9060
gagaggatgg gggagggggac aaattagaat aggatagcat ctgatgcctg agaaccctct 9120
cctagcactg tcaaagtctg gtattgaatg gggactgagg atgggtctca gagagcaacc 9180
tcctccctcg tagagggaga ttatatcccc aactccaggg acctctttat ctcaatctat 9240
ttatttggca tcctgggaag gatttccaat agtaatttat gtgacctggg gcaggatacc 9300
gtcagtgagg tgcccagagc tgcacccttt cctccatttc ccacccccca tctcctcaac 9360
caccagggtc tgagttctag cagggtcctg ggggtatccc actgctatac tgttctactg 9420
cttccctcag tatctgaatg tctcaattta aaacttgaag ctcttttagac caatagactg 9480
gtgagaggag aaaggagctt atcccccaga cctgcttta taccattcac atcccagggc 9540
tgtgtccaga cagcacaaaa cggcaaggag agcccaagcc ccaatgccag aattcttcca 9600
aactccctga ctctttgaag tttttactca cccatttca attatcctga tcccttctca 9660
tcccctgctt ggcttctctg catgtggtca tctgctgtgg cttggtgttt aatgggttaa 9720
aaataagcca ctgcctgaca tccaacatt tgacacccca gcaatgtgtg actccccc aa 9780
cattccacta tgccatcctg cagctgaaat gggaacactg gctgcctctc caaacccgct 9840
cttgacaga ggatctggga ggtggaagcc aggccagagg acttggggaa aatgagatgg 9900
aggaaggaaa aaggggagaag ctgagccaca gcttaactcc tacagagtga aatgaaaacg 9960
ggctgaaaat accaccccag gagaggacct cgccccaaagc aagccagtga gcagccctgc 10020
cagactactg ccagactgag aaaccagaa gctggtagtc atgtgggctt gccttctctg 10080
ccaaacgact gggaaaccaa aatgagccca ccttgtgttc ttcctagctc caccctcccc 10140
gtgctgctgt gttctgctcc tcccacgct tccctgctat agttcccagc tgcgtgtaacg 10200
gagccacctc caactctaac aataaaccaa gttcattgca gaaaaaaaaa aaa 10253

```

<210> 20

<211> 3851

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472720CB1

<400> 20

```

cggcagcaac  tgcctgtcag  gcgcgcgcgc  aacggcctttg  cgaggctcac  tcggctctgag  60
aggtcggagg  ctgcgagtg  cgctgctgaa  ggcctgtggtg  gaccgggctg  gatcgcggat  120
tgtggagtag  attatagatt  tgaaatagcg  gaggctggggg  tggatcgggg  cttgggggtg  180
gataggggat  ttggggctgg  gtcggccggg  gtcggggagg  ggggtggtga  aaaggtgaca  240
gggagctgcc  ctgcgtcaag  agccgggtgt  tgggggtctg  agaagaagtc  accaatatga  300
agttattcgg  cttcgggagc  cgcagggggc  agacggccca  gggctccata  gaccacgtct  360
acacgggttc  cggataccga  atccgggact  ccgaactgca  gaagatccac  agggcagctg  420
tcaaaggcga  tgcgcgggag  gtggagcgct  gcttggcgcg  caggagcgga  gacctggacg  480
ccctggacaa  gcagcacaga  actgctctac  acttggcctg  tgccagtggc  catgtgcaag  540
tggtcactct  cctggttaac  agaaaatgcc  agattgatgt  ctgtgacaaa  gaaaacagaa  600
cgcctttgat  acaggctgtc  cattgccagg  aagaggcttg  tgccgttatt  ctgctggaac  660
atggcgccaa  tccaaacctt  aaggatatct  acggcaacac  tgctctccat  tatgccgtgt  720
atagtggag  cacctcactg  gcagaaaaac  tgctttccca  tgggtgcacat  attgaagcac  780
tggacaagga  caataatacc  ccacttttat  tcgctataat  ttgcaagaaa  gagaaaatgg  840
tgggaatttt  attgaaaaag  aaagcagtg  acaatgccgt  tgataggctg  agacggtcag  900
ctctcatact  tgctgtatac  tatgactcac  caggatttgt  caatatcctt  cttaaagcaa  960
atattgatgt  cttcgtctca  gacatgtgtg  gacgagatgc  agaagattat  gctatttctc  1020
atcatttgac  aaaaattcaa  caacaaatct  tggaacataa  aaagaagata  cttaaaaagg  1080
agaaatcaga  tgttggaaag  tctgatgaat  ctgcagtcag  cattttccat  gaactgcgtg  1140
tggattcatt  gcctgcctcg  gatgacaaag  acttgaatgt  tgctactaag  tgtgtccccg  1200
agaaagtgtc  agagccttta  cctggatcct  cgcataaaaa  aggaacaga  atagtcaatg  1260
gacaaggaga  agggcctcct  gcaaaacatc  ctctcttgaa  gcctagcact  gaagtggag  1320
atcctgctgt  gaaaggagca  gtacaaagaa  agaattgtaca  gacattgaga  gcagaacaa  1380
ccttaccagt  ggcttcagag  gaagagcaac  aaaggcatga  aagaagtga  aagaagcaac  1440
cacagggtcaa  agaaggaaat  aatacaaaac  aaagtgaaaa  aatacaactt  tcagaaaata  1500
tatgtgatag  tacatcttct  gctgctgctg  gcagattaac  ccaacaaaga  aagattggga  1560
aaacgtatcc  tcagcaatct  cccaagaagc  tgaagggaag  gcatgataga  tgcacttaaa  1620
aacaagaaaa  tgaagaaaaa  acaaatgtta  atatgctgta  caaaaaaat  agagaagaat  1680
tagaaaggaa  agagaaacaa  tataagaaag  aagttgaagc  aaaacaactt  gaaccaactg  1740
ttcagtcact  agagatgaaa  tcaagactg  caagaaatac  tccaaatcgg  gattttcata  1800
atcatgaaga  aatgaaaggt  ctgatggatg  aaaattgcat  tttgaaggca  gatattgcta  1860
tactcagaca  ggaaatatgt  acaatgaaaa  atgacaactt  ggaaaaagaa  aataaatatc  1920
ttaaggacat  taaaattgtt  aaagaaacaa  atgctgccct  tgaaaagtat  ataaaactca  1980
atgaggaaat  gataacagaa  acagcattcc  ggtatcaaca  agagcttaat  gatctcaagg  2040
ctgagaatac  aaggctcaat  gccgaactgt  tgaaggaaaa  agaaagcaag  aaaagactgg  2100
aagctgacat  tgaatcttat  cagtctagac  tggctgctgc  tataagcaaa  cacagtgaag  2160
gtgtgaaaa  agaaagaaac  ctaaaacttg  ctttagagag  aacacaagat  gtttctgtac  2220
aagtagaaat  gagttctgct  atttccaaag  taaaagctga  gaattgagttt  cttactgaac  2280
aactttctga  aacacaaatt  aaattcaata  ctttaaaaga  taagttccgt  aagacaagag  2340
atagtctcag  aaaaaagtca  ttggcttttag  aaactgtaca  aaacgacct  agccaaacac  2400
agcagcaaac  acaggaaatg  aaagagatgt  atcaaaatgc  agaagctaaa  gtgaataatt  2460
ccactggaaa  gtggaactgt  gtagaagaga  ggatatgtca  cctccaacgt  gaaaatgcgt  2520
ggcttgtaga  gcaactagat  gacgttcatc  agaaagagga  tcataaagag  acagtaacta  2580
atatccaaag  aggctttatt  gagagtggaa  agaaagacct  cgtgctagaa  gagaaaagta  2640
agaagctaatt  gaatgaatgt  gatcatttaa  aagaaagtct  ctttcagtat  gagagagaga  2700
aagcagaagg  agtacctaaa  aaagaaaatg  aagaattaag  aaaacttttt  gagttaatat  2760
catcactgaa  atataatgtg  aatcgaataa  gaaagaaaaa  tgatgaatta  gaagaaggag  2820
caactggata  taagaaactc  ctggaaatga  caataaatat  gttaaatgta  tttggaaatg  2880
aagactttga  ttgcatgga  gacttaaaaa  cagatcaact  gaaaatggat  attctgatta  2940
agaagctaaa  acagaaggaa  caagcacaat  atgaaaaaca  attagagcag  ttaaacagg  3000
ataatatggc  ttcactaaat  aaaaagggaac  tcacacttaa  agatgtggaa  tgtaaatct  3060
cagaaatgaa  aactgcttat  gaagaggtta  caaccgaatt  agaagaatat  aaggagcct  3120
ttgcagcagc  attgaaagct  aacaattcca  tgtcaaaaaa  gttacaaaaa  tcaaataaga  3180
aaatagcagt  gattagcatg  aagctcctca  tggagaaaga  gcagatgaaa  tattttctca  3240
gtgctcttcc  tacaaggcga  gaccagagat  caccttgtgt  tgaaaatctt  actagtatag  3300
gactcaacag  aaaatatatt  ccccaaacac  ccataagaat  tcctatttca  agcccacaga  3360
cttcaaataa  ctgcaagaac  tcctagactg  tgatggagct  ggactgtgta  gaacaaataa  3420

```

```

ctagagaaac aaagagaatt gttgctgtgt tgaacacttg ctcccgctcta cttactttctc 3480
tataatccac tgccatggaa tgagtgtatt ttcttagaag cagaggtgga gccactgagg 3540
aagcacaggc gagccctccc cagcacgtgc tcaactggtcc ccaacagAAC aaccgctgcc 3600
gcatccatga ggctcccatt gtggtgggtt gtgtcacccc acaatgtcac tgttgctgag 3660
ccccatcgc ctctgtgttg tggagcagtt agagacacac tgtggtgtct gagggtctct 3720
gtgtgaagga ccgttttcta ggtgagaggg acatctcaac acagctgact gatcagactc 3780
agccgttttg cacaccctgg tcagaatgaa acattccttg ggggaactcgg gccgtgagaa 3840
gcatcctccc g                                     3851

```

<210> 21

<211> 3100

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7583990CB1

<400> 21

```

ccgccccag cccgccttcc tcccgccgcg ccctccgcct ccgcccgcac ctcgctagcg 60
ttcccctgtc ttccccaacg ccccgagacc gccggccgct agcgtcagcg ccagccagaa 120
ttaaggaagt tcaactggagt aaaatggagg cctcagtaat attaccatt ctgaagaaaa 180
aactagcctt cctttcagga ggaaaggaca gacggagtgg cctcattttg acaattccat 240
tatgcctcga acagacaaat atggatgagc tgagtgtcac cttagactac ctactcagca 300
ttccaagtga gaagtgtgaa gctagaggat ttaccgtgat tgtggatggc agaaaatcac 360
agtggaaatgt ggtgaaaaca gtagtcgtaa tgctacagaa tgttggtcca gctgaggtgt 420
cccttggttg tgtggttaaag ccagatgaat tctgggataa gaaagtaacg cttttttgtt 480
tttggaagga gaaggataga cttggccttg aggttatttt agtgtccgcc acaaaattga 540
ctcggttatat agaaccatgc caattaacag aagatttttg tgggagtctc acctatgatc 600
acatggactg gttaaataag aggctggttt ttgagaagtt taaaaggaa tctacatcat 660
tattagatga acttgctttg attacaatg gaagtataa aggaatcag caagagaaag 720
aaaggtctgt ggatttaaac tttcttccat cggttgatcc tgaaacagtt cttcagacag 780
ggcatgaatt gttgtccgaa ttacagcagc gtcgatttaa tggctcagac ggaggggttt 840
catggtctcc tatggatgat gaacttcttg cacagccaca ggttatgaaa ttattagatt 900
cactccgaga gcaatatacc cgctaccagg aagttttagt gcaacgtagc aagcgcacac 960
agttagaaga gattcaacag aaggtaatgc aggtggtgaa ctggctagaa gggcctggat 1020
cagaacaact aagagcccag tggggcattg gagactccat tagggcctcc caggccctac 1080
agcagaaaca cgaagagatt gagagccagc acagtgaatg gtttgcaagt tatgtggaac 1140
ttaatcagca aattgcagca ctcttgaaat ctggcgatga ggaagatctt gtggaactaa 1200
agtactgca gcaacaactt agtgatgttt gttatcgaca ggccagtcag ctggaattta 1260
ggcaaaatct cttacaagca gctcttgaat ttcatggtgt tgcccaagat ttgtctcagc 1320
agttggatgg cttattaggg atgttgtgcy tagatgtagc accagctgat ggagcatcga 1380
ttcagcaaac tttaaaactg cttgaagaga agctgaaaag tgttgatgtg ggattgcaag 1440
gtttgcgtga aaaagggtcaa ggtctccttg atcagatctc caatcaggca tcctgggcct 1500
atggaaagga tgtaaccatt gaaaataaag aaaatgtgga ccacatacaa ggagtgtggt 1560
aagatatgca gcttagaaaa caaagatgtg aagacatggt agatgtgcga aggttaaaga 1620
tgcttcagat ggtgcagttg tttaaatgtg aagaagatgc tgcccaggca gtagaatggc 1680
taagtgaact tctggatgct ctgcttaaga ctcacatcag attgggcgat gatgctcaag 1740
aaacgaaagt tttgctggaa aagcatagaa aatttggttg tgttgacag agcacttatg 1800
actatggcag gcagttgcta caggccacag ttgtgttatg ccaatctttg cgctgcactt 1860
ctcggtcatc tggggatata cttcctcgac tgaacagagt atggaaacaa tttacaatag 1920
catctgaaga gagagtacat agattggaaa tggctattgc atttactca aatgctgaaa 1980
agattttgca ggactgtcca gaagagcctg aagctattaa tgatgaggag caatttgatg 2040
aaattgaagc agttgggaaa tcacttttgg atagattaac tgttccagta gtttatcctg 2100
atggaaccga acaatatttt gggagtccaa gtgacatggc ttctactgca gaaaacatca 2160
gagacaggat gaaactagtt aatctcaaaa ggcagcagct gagacatcct gaaatggtga 2220
ccacagagag ctaatagcta ccagctacct acagatttgc agttcataat cccgcatggt 2280

```

```

gtcaacatac tacagcatta gccaccacac cttaagatgc atttcacagc caaaataagt 2340
ctcatttctt ttcattgacac atttctcttt acatgttaac accttgctac taccaaggca 2400
taattactta acatgcttcg aggctgtaga ttccaagtat cttaaaagaa ggaactataa 2460
acattgcact gaaaacttgc tttaaagctt tacctgacct gtcagtttgt agacaaacaa 2520
ctgataataa gctttgaatg gtgctaataa gagtaggaat tctctctatt aaaaagaaaa 2580
aaaaaagttg cccttcctcc acaggtgatt tagtaaattt agacagtagt taaactcttg 2640
ttagtagaca gtggtgtcct caaaatttta ctttgtaatt cttcagaatt gattatTTTT 2700
attgtgtcaa tacagagaaa gcctttcaga tctttgatat atcatagtca ttaaagacc 2760
ttttcctatt tgtattgata atgtattaaa agttgtttgt gcttaataaa agacttcttt 2820
aaacatctta ttttaatttag tagttacatc ctatttccaa acatgagtg cttattttaa 2880
agggcattct taggactgtg aggatgggtt aatatttgtt ttttcattgt ggttgcattg 2940
attttagaca ggaaatacat atgtaagcat gtgtatataa taaataagca tgttttatca 3000
tgaaaaatta ttgtgaacaa tttagatctt taagaactta ttaataatgg aatactattt 3060
ctaatttttc tctttttcaa cttgaaaaat attctcaaaa 3100

```

<210> 22

<211> 3248

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2058182CB1

<400> 22

```

gcggaagcg atgtagtagc tgccaggctg tccccgccc tgcccgccc gagccccg 60
ggcgccgccc gccaccgccc ccatgaagaa gcagttcaac cgcattgaagc agctggctaa 120
ccagaccgtg ggcagagctg agaaaacaga agtcttagt gaagatctat tacagattga 180
gagacgcctg gacacgggtg ggtcaatatg ccaccattcc cataagcgct tgggtggcatg 240
ttccagggc cagcatggca ccgatgccga gaggagacac aaaaaactgc ctctgacagc 300
tcttgctcaa aatatgcaag aagcatcgac tcagctggaa gactctctcc tggggaagat 360
gctggagacg tgtggagatg ctgagaatca gctggctctc gagctctccc agcacgaagt 420
ctttgttgag aaggagatcg tggaccctct gtacggcata gctgaggtgg agattcccaa 480
catccagaag cagaggaagc agcttgcaag attggtgtta gactgggatt cagtcagagc 540
caggtggaac caagctcaca aatcctcagg aaccaacttt caggggcttc catcaaaaat 600
agatactcta aaggaagaga tggatgaagc tggaaataaa gtagaacagt gcaaggatca 660
acttgacgca gacatgtaca actttatggc caaagaaggg gagtatggca aattctttgt 720
tacgttatta gaagcccaag cagattacca tagaaaagca ttagcagtct tagaaaagac 780
cctccccgaa atgcgagccc atcaagataa gtggcgga aaaccagcct ttgggactcc 840
cctagaagaa cacctgaaga ggagcggcg cgagattgag ctgcccattg aagcctgtgt 900
catgtgctt ctggagacag gcatgaagga ggaggcctt ttccgaattg gggctggggc 960
ctccaagtta aagaagctga aagctgctt ggactgttct acttctcacc tggatgagtt 1020
ctattcagac ccccatgctg tagcaggtgc tttaaaatcc tatttacggg aattgcctga 1080
acctttgatg acttttaatc tgtatgaaga atggacacaa gttgcaagtg tgcaggatca 1140
agacaaaaaa cttcaagact tgtggagaac atgtcagaag ttgccaccac aaaatTTTT 1200
taactttaga tatttgatca agttccttgc aaagcttgct cagaccagcg atgtgaataa 1260
aatgactccc agcaacattg cgattgtgtt aggcctaac ttgttatggg ccagaaatga 1320
aggaacactt gctgaaatgg cagcagccac atccgtccat gtggttgagc tgattgaacc 1380
catcattcag catgccgact ggttcttccc tgaagagggtg gaatttaatg tatcagaagc 1440
atttgtacct ctcaccacc cgagttctaa tcaactcatt cactggaa acgactctga 1500
ctcggggacc ctggagagga agcggcctgc tagcatggcg gtgatggaag gagacttggg 1560
gaagaaggaa agtctctcca aaccgaagga cctgtatct gcagctgtgc cagcaccagg 1620
gagaaacaac agtcagatag catctggcca aaatcagccc caggcagctg ctggctccca 1680
ccagctctcc atgggccaac ctcacaatgc tgcagggccc agcccgcata cactgcgccg 1740
agctgttaaa aaaccgctc cagcaccccc gaaaccgggc aaccacctc ctggccacc 1800
cgggggcccag agttcttcag gaacatctca gcatccacc agtctgtcac caaagccacc 1860
caccgaagc cctctctctc ccaccagca caggggccag cctccaggcc agccctccgc 1920

```

```

cccccccag ctctcagcac cccggaggta ctccagcagc ttgtctccaa tccaagctcc 1980
caatcaccca ccgcgcgcagc cccctacgca ggccacgcca ctgatgcaca ccaaacccaa 2040
tagccagggc cctcccaccc ccatggcatt gcccagtgag catggacttg agcagccatc 2100
tcacacccct ccccagactc caacgcccc cagtactccg cccctaggaa aacagaaccc 2160
cagtctgcca gctcctcaga ccctggcagg gggtaacctt gaaactgcac agccacatgc 2220
tggaacctta ccgagaccga gaccagtacc aaagccaagg aaccggccca gcgtgcccc 2280
accccccaa cctcctgggtg tccactcagc tggggacagc agcctcacca acacagcacc 2340
aacagcttcc aagatagtaa cagactccaa ttccagggtt tcagaaccgc atcgagcat 2400
ctttcctgaa atgcactcag actcagccag caaagacgtg cctggccgca tcctgctgga 2460
tatagacaaat gataccgaga gcactgccct gtgaagaaag ccctttccca gccctccacc 2520
acttcacccc tggcgagtggt agcaggggca ggcgaacctc tttctttgca gaccgaacag 2580
tgaaaagctt tcagtggagg acaaaggagg gcctcactgt gcgggacctg gccttctgca 2640
cgccccaaag agaaccctga ggccaccact aaagctgaat gacctgtgtc ttgaagaagt 2700
tggtctttctt tacatgggaa ggaaatcatg ccaaaaaaat ccaaaacaaa gaagtacctg 2760
gagtggagag agtattcctg ctgaaacgcg cataggaagc ttttgtccct gctgttaatg 2820
cgggcagcac ctacagcaac ttggaatgag taagaagcag tgcgttaact atctatttaa 2880
taaaatgctc tcattatgca agtcgcctac tctctgctac ctggacgttc attcttatgt 2940
attaggaggg aggtgcgct ccttcagact tgctgcagaa tcattttgta tcatgtatgg 3000
tctgtgtctc ccagtcctcc tcagaacccat gcccatggat ggtgactgct ggctctgtca 3060
cctcatcaaa ctggatgtga cccatgccgc ctcggtggat tgtcggaatg tagacagaaa 3120
tgtactgttc tttttttttt ttttaacaa tgtaattgct acttgataag gaccgaacat 3180
tattctagtt tcatgtttaa tttgaattaa atatattctg tggtttataat gaaaaaaaaa 3240
aaaaaaaaa 3248

```

<210> 23

<211> 2592

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3564377CB1

<400> 23

```

gcgagagcgc cgcccaccca tccgggggcaa gagccgcgcc gcaggagagg caggctggac 60
cgggggctcc ccggggccgc gaccccgcc gtgacccgc agccccagc tcgccccaa 120
gatgatgaag aggcagctgc accgcatgct gcagctggcc cagacgggca gcttgggacg 180
caccctggag accgctgagt tcctgggtga ggacctgctg caggtagaac agcggctgga 240
gccggccaag cgggcagccc acaacatcca caagcggctg caggcctgtc tgcagggcca 300
gagcggggca gacatggaca agcgggtgaa gaagcttccc ctcatggctc tgtccaccac 360
gatggctgag agcttcaagg agctggaccc tgattccagc atggggaagg ccttggagat 420
gagctgtgcc atccagaatc agctggccc ccatcctggc gagtttgaga tgaccctgga 480
gagggacgtc ctgcagccac tcagcaggct gagtgaggag gagctgccag ccacctcaa 540
acacaagaaa agcctccaga agctcgtgtc cgactggaac aactcaaga gcaggctcag 600
tcaggcaacc aagaattcag gcagcagtcaggcctagga ggcagcccgg gtagtcacag 660
ccatacgacc atggccaaca aggtggagac gctgaaggag gaggaggagg agctgaagag 720
gaaagtggag caatgcaggg acgagtactt ggctgacctg taccactttg ttaccaagga 780
ggactcctat gccaaactact tcattcgtct cctggagatt caggccgatt accatcgag 840
gtcactgagc tcgctggaca cagccctggc tgagctgagg gagaaccacg gccaagcaga 900
ccactccctc tcgatgacag ccacccactt cccagggtg tatggggtgt cgctggcaac 960
ccacctgcaa gagctgggccc gggagattgc cctgcccac gaggcctgcg tcgatgatgt 1020
gctttctgag ggcaggaagg aagagggtct cttccgtctg gctgctgggg cctcgggtgt 1080
gaagcgtctc aagcagacaa tggcctcgga ccccccacagc ctggaggagt tctgctccga 1140
cccgcacgct gtggcagggt ccctcaagtc ctatctgcgg gagctgccag agcctctgat 1200
gaccttcgac ctctatgatg actggatgag ggcagccagc ctgaaggagc cagggggccc 1260
gctgcaggcc ctccaagagg tgtgcagccg cctaccccc gagaacctca gcaacctcag 1320
gtacctgatg aagttcctgg cacggctggc cgaggagcag gaggtgaaca agatgacacc 1380

```

```

cagcaacatc gccatagtc tgggacccaa cttgtctgtg ccacctgaga aagaagggga 1440
ccaggcccag ctggatgcag cctccgtgtc ttccatccag gtggtgggcg tcgtcgaggc 1500
gctgatccag agcgcagaca ccctcttccc tggagacatc aacttcaacg tgtcaggcct 1560
cttctcagct gttaccctcc aggacacagt cagtgcaggg ctggcctctg aggaacttcc 1620
gtccactgcc gtgcccaccc cagccaccac cccggctccg gctccggctc cagctccagc 1680
tccggcccca gccttggctt cagcggctac caaggaaagg acagagtctg aggtgcctcc 1740
cagaccagcc tcccccaagg tcaccaggag tcccccgagg acagctgcc cagtggagga 1800
catggctcgg aggaccaagc gcccgggccc agcccgggcc accatgccgc cccccagggt 1860
ctccggctcc cgctcctccc ctccagcccc gcccttgccc cctggctctg gcagccctgg 1920
gaccccccaa gccctgcccc gacgtctggt tggcagcagc ctccgagccc ccacagtgcc 1980
acccccgtta cccccacac cccctcagcc tgcccgggcg caaagccggc gttcaccagc 2040
ctcccccagc ccggcctccc caggtccagc ctccccccag ccagtctctt tgagtaacct 2100
tgacacagtg gacctggggg ctgccacagc agagggagga gccctgagg ctatcagtgg 2160
ggtcccccact ccccagcta cccccctca gcccgcccc aggagccttg cctcagagac 2220
caactgagtg gctggtttct ccctaagcag ccctcagcac cccctccctc ccacctggc 2280
cctcccagga cagctctcgc ccccacaaa ggggcatggg cctccagcct ttgcccacaa 2340
gtgcctcagt gccactggg tcggccccc tggccaggag ggctcaggac aatcctctat 2400
ttcctgacct tttcctcgtc caccctgggc ttggggagccc cccaccgga ctctccactc 2460
tccggcaggt cctaggggag ccaccggaag gaaggagagg tttgcctgct cctacgggac 2520
tgattcttct cttgccgaca tgttttttgt aaggctggta aataaattat tttggacaaa 2580
aaaaaaaaaa aa 2592

```

<210> 24

<211> 2004

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1568689CB1

<400> 24

```

ggccccgcgc cggagcagtg ccggagcccc gccagagccc gacttcagcc ccagccagat 60
cccgctcaa cggaggcgga acggcggacc ccgtaccctg gcagcatcgg agcaccggcg 120
ggtgaaggca aggtcccttg actggtcata tacctcttgt ggccctggca gaatcaagat 180
gaggccctgt catgcctccc cagtgcaggg tacagtctga gcagacagca tggcctgcca 240
ctggcagtg acaccatgtc tgcaggaggt ggccgggccc ttgcttggca agtgttcccc 300
cccatgcca cttgccgggt ctatggcaca gtggcacacc aagatgggca cctgctggtg 360
ttgggggggt gtggccgggc tggactgccc ctggacactg ctgagacact ggacatggcc 420
tcgcacacat ggctggcact ggcacccctg cccactgccc gggctggtgc agctgcggta 480
gttctgggca agcagggtgct agtgggtgggt ggtgtggatg aggtccagag cccggtagct 540
gctgtagagg ccttcctgat ggatgagggc cgctgggagc gtcgggccac cctccctcaa 600
gcagccatgg gggttgcaac tgtggagaga gatggtatgg tgtatgctct ggggggaatg 660
ggccctgaca cgcccccca ggcccaggta cgtgtgtatg agccccgtcg ggactgctgg 720
ctttcgctac cctccatgcc cacaccctgc tatggggcct ccaccttctt gcacgggaac 780
aagatctatg tcctgggggg ccgccagggc aagctcccgg tgactgcttt tgaagccttt 840
gatctggagg ccgctacatg gacccggcat ccaagcctac ccagccgtcg ggcctttgct 900
ggctgcgcca tggctgaagg cagcgtcttt agcctgggtg gcctgcagca gcctggggcc 960
cacaacttct actctcgccc acactttgtc aacactgtgg agatgtttga cctggagcat 1020
gggtcctgga ccaaattgcc ccgcagcctg cgcagagggg ataagagggc agactttgtg 1080
gttgggtccc ttggggggcca cattgtggcc attggggggc ttggaaacca gccatgtcct 1140
ttgggtctct tggagagctt tagccttgca cggcggcgct gggaggcatt gcctgccatg 1200
ccactgccc gctgctcctg ctctagtctg caggctgggc cccggctgtt tgttattggg 1260
ggtgtggccc agggccccag tcaagccgtg gaggcactgt gtctgcgtga tggggtctga 1320
aggcttgggt ggagctgtcc actggagcag ctcatgcca gaggcagcta tttctatggc 1380
tccttttgct gctgaggaca ctactgtgg ctctgtggga tgagagaggc atgggggtga 1440
gcacttgaaa cactgccttg gggccttggg ttaggggagc ctttgtcttt agtgcaggac 1500

```

```

acacatatgc ttacacctac ctttatcacc attcgttcat gaatcatgcc tagctccatc 1560
cttgccctgg gacctactag gccttccatc caactgggaa atggggagaa gcaaagctgg 1620
cctcatgctc ttcagggtca gttcctatct ggagttgacc aggcctaccc cagttgccat 1680
tcctgaaaaa tctcagctgc caggctgcct ttaggggtccc tgtagacca ggagagttga 1740
gaggggtggg gacacagaga gaatagagag gatgtgggaa ctgccagagg gccggagcgc 1800
aggagttcaa gtggaggaat gctggccttg agccctctac actgctggtt gtatgacctt 1860
ggacaagtca cttcacctct ctgtgcctca gcacccatcat ctataaatgg ggatctctga 1920
aaccttcta ccctacctac ctcacagggc tgtgtgagg acccagggag tttggatgtg 1980
gaagtaaaag tgctgctaaa aaaa 2004

```

<210> 25

<211> 2250

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1393767CB1

<400> 25

```

ggagcaccgg gtggattgga cgcttcccca gagaccaga agcagaagga gcggaccag 60
gcagccggca ccatggagat tgtgtacgtg tacgtcaaga agcgagcga gttcggaag 120
cagtgaatt tctcggaccg ccaggccgag ctgaacatcg acatcatgcc caaccctgag 180
ctggccgagc agttcgtgga gcggaaccca gtggacacgg gcatccagtg ctcgatcagc 240
atgtcggaac acgaggccaa ctcagagcgg tttgagatgg agaccgggg agttaaccat 300
gtcagggggg gctggcccaa ggacgtgaac cccttgagc tggagcagac catccgtttc 360
cggaagaaag tggagaaaga tgagaactac gttaacgcca tcatgcagct cggctctatc 420
atggagcact gcatcaagca gaacaatgcc attgacatct atgaagagta tttcaatgac 480
gaggaggcca tggaaagtga ggaggaggac ccttcagcta aaaccatcaa tgtgttcagg 540
gacccccagg aaatcaagag ggctgccaca cactctcct ggcaccccg tggcaacagg 600
aagttggcag tggcatactc ctgcttggat tttcagcggg cactgtggg catgagcagc 660
gattcataca tctgggacct ggaaaacccc aacaagcctg aacttgctct gaagccatcg 720
tctccactcg tgacgttggg gttcaacccc aaagattccc acgtactcct ggggtggctgc 780
tacaatggac agatagcctg ctgggacacc cgaaagggca gcctggtggc ggagctatcc 840
accattgagt ccagccaccg agaccctgtg tatggcacca tctggctgca gtcgaagacg 900
ggcaccgagt gcttctcagc ttccacggat gggcaggtca tgtggtggga catccgaaag 960
atgagcgagc cactgaagt tgtgatcttg gacatcacca agaaggaaca gttggaaaat 1020
gccttggggg ccatctccct ggagttcgaa tctactttgc ccaccaagtt catggtgggg 1080
accgagcagg gcacgtcat ctctgcaac cgcaaggcca agacgtcagc tgaaaagatt 1140
gtgtgcacct tcccgggcca tcatggcccc atctacgccc tccagagaaa ccccttctac 1200
ccgaagaact tctgacgggt tggcgactgg acagcccgca tttggtctga agacagccgg 1260
gaatcgcca tcatgtggac caagtaccac atggcttacc tcatgatgc tgcctggagc 1320
cccgtgaggc cgaccgtttt ctttaccacc aggatggagc gaaccctgga tatctgggac 1380
ttcatgttcg agcagtgcga tcccaccctc agcttgaagg tgtgtgacga ggccctcttc 1440
tgcctccggg tgcaggacaa tgggtgtctc atcgctgag gctcccagct ggggacaacc 1500
accctgctgg aggtctcgcc tgggtctctt accctccaga ggaatgagaa gaacgtagcc 1560
tcttccatgt ttgagcgtga gaccggcgga gagaagatcc tggaggccag gcaccgggag 1620
atgcccgtga aggagaaggg taaggcggag ggcagggatg aggagcagac cgatgaggag 1680
ctggccgtag acctggaggc gctggtcagc aaggccgagg aggagtctt cgacatcatc 1740
ttcacagagc tgaagaagaa ggaggcagac gccataaagc tgacgccagt gcctcagcaa 1800
ccaagtccag aagaagacca ggtggtggag gagggagagg aagcagcggg ggaagaaggg 1860
gatgaagaag tggagaaga cttagcctag aatcagcct tcgactgagg cgctatccct 1920
gtgtgccttc ctttccacc tcttgacct caaccagact tgcattggca tggcagggcc 1980
tcgggaagac cttcaggagt ggggaagggt ttctcctcca tgatcgacct tctcgtcca 2040
cctacaaatc aggaacagaa agtctgtcca ctttgaaaat acctttccag gcagctccct 2100
gaccatttgg acacattgcc acgacaggag cctccaagta tgtgggaggg gacggggcggg 2160
acgagcttgg ctgttctgct gcacctgaat gctttctggt atcctaattc ttgtaaaatt 2220

```

aatgaatcg taacaataaa aaaaaaaaaa

2250

<210> 26

<211> 3728

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 3029343CB1

<400> 26

```

atggattatg aacaccatga aagatggccc aggtttaaca ggatgttcct ggacaagtca 60
ggagcacagt ctaaggcatt tgatgtactt ggaagagttg aagcttacct taagctcctt 120
aaatcagagg gtttaagtct ggctgttttg gcagtgaggc atgaggaatt acacagaaaa 180
attaaagact gcacaactga tgctttgcaa aagggacaaa ccttaatcag ccaagtagac 240
tcctgcagca ccaggcctca gggacaatca aagccatata aaactgacct caaatcccca 300
gaacctgtcc cgcgtccagt cagggagctg cacatcaagg aagtgtgctc caggcacgag 360
ggggcccatga gtacagtgga tgttgcggtc acttcttcag agaagggaga cacaatccga 420
aagtctgaga tcaagacagg ccaaatgaaa ggctctcagg tgtccggcat ccatgagatg 480
atggggtgca ttaagagacg agtggatcat ctgaccgaac agtgttcagc gcacaaggaa 540
tatgtcttta agaaacaaca actaacagcc tcagtggagg gttacctacg gaagggtgaa 600
atgtcaattc agaaaatcag tccagtactt tctaattgaa tggatgttgg ttctaccctg 660
tctgaatcag agaagatttt gaataaatat ctggaactag atatccaagc taaggagaca 720
tcacatgaat tagaagcagc tgcaaaaacc atgatggaga aaaatgaatt tgtatctgat 780
gaaatggtat cactttcctc taaagctaga tggctagcag aagaattaaa cctatttggc 840
caaagcattg actatagatc gcaagtcctg caaacttacg tggcatttct gaagtcacat 900
gaggaggtag agatgcagtt tcagagctta aaagaatttt atgaaaccga aatccctcag 960
aaggagcagg atgatgctaa agccaagcat tgttctgact cggtgagaa gcagtggcag 1020
ctatttttaa agaagagttt tataacacaa gatctagggc ttgagttcct taattttaata 1080
aatatggcaa aagagaacga gatattagat gtgaaaaatg aagtgtacct catgaagaac 1140
accatggaaa accagaaagc agaacgggaa gaacttagcc tccttcggct ggcattggcag 1200
cttaaagcca cggaaagcaa gcctggaaaa cagcagtggg cggcattcaa agagcaactt 1260
aaaaagactt ctcaacaact aaaacttctt caggaagcac ttatgcctgt gtctgcactt 1320
gacctcggag ggagcctcca gtctatttta gatctacgac aaaaatggaa tgacatgaag 1380
cctcagttcc agcaattgaa tgatgaggtt cagtacatta tgaaagaatc agaggagtta 1440
actggcagag gagccctgtt aaaagaaaag tctcaacaac tgaaggacct tattcacttc 1500
catcaaaaac agaaagagag aatccaggat tacgaggata tcctgtacaa ggtggtccag 1560
ttccatcaag tcaaggaaga gctgggacgt ctcatcaaat caagagagct ggagtttgta 1620
gagcagccga aggaactggg tgatgcccac gatgtgcaga ttcacctccg gtgctctcag 1680
gaaaagcaag cccgtgtaga ccatctccac agactggccc tttccttagg agtcgacatc 1740
atctcatcag tgcagcggcc tcaactgctt aatgtttctg caaagaacct acagcagcag 1800
ctggagctcc ttgaggagga cagcatgaag tggcgtgcca aagctgagga gtatggacgg 1860
accctgtccc gtatgttgga gtactgcgcc atgagagacg agataaatga gctcaaagac 1920
tcattcaaag atatcaaaaa gaaattcaat aatttgaagt ttaattacac taagaaaaat 1980
gaaaaatctc ggaatctgaa ggcgttataa tatcaaatc agcaagttga tatgtatgct 2040
gaaaaaatgc aggtcttgaa aaggaaaatg gaaaaagtta gtaataaaac ctctgattct 2100
ttcttaaat atccaagtga taaagttaat gtccttttgg aagtcatgaa ggatttgcaa 2160
aaacatgtgg atgactttga caaagtgtgt acagattaca agaagaattt ggacctgact 2220
gagcatttcc aggaggtgat agaagagtgt catttttggg acgaagatgc aagtgcacac 2280
gttgtaagag ttggaaaata ttccacagag tgcaagacaa aggaagctgt gaaaattctc 2340
caccagcagt ttaataagtt tattgcacct tcagtgccgc agcaagaaga aaggattcag 2400
gaggccactg accttgctca gcacttatat ggtttggaag aaggacagaa atatattgag 2460
aaaatagtga caaaacacaa agaggttctt gaatctgtga ctgaattatg tgagtcccg 2520
acagagctcg aagaaaaact gaagcaggga gatgttttaa agatgaatcc gaatttgaa 2580
gacttccatt atgattacat tgacttgcta aaggaaccag caaaaaataa gcagacaata 2640
ttcaatgaag aaaggaataa ggggcaggtg caggtggcag atcttttggg catcaatgga 2700

```

```

acaggggaag agcgactacc acaagacctg aaggtgtcca ctgacaagga ggggtggcgtc 2760
caggacctgc tcctgcctga agacatgctc tcaggggaag aatatgagtg tgtctcacct 2820
gatgacatct ccttgccctc tctcccagga agccctgagt ccccccctgc accatctgac 2880
atggaggtgg aagagcctgt cagctcctcc ctcagccttc acataagcag ctatgggggtg 2940
caggctggga ccagcagccc aggggatgcc caggaatctg ttcttccacc acctgttgcc 3000
tttgcggtatg catgcaatga taagagagaa acattttcaa gtcattttga gaggccttac 3060
ctccagttca aagctgagcc cccactaacc tccagaggat tcgtggaaaa gagtactgcc 3120
ttacacagaa tcagtgtctga acatccagag agcatgatga gtgaagtgc tggagagagct 3180
ttacagcagc accctcaggc tcagggtggt ttgctagaaa cacgggagaa aatgcatgct 3240
gataataact tcactaaaac ccaagatagg ctgcatgctt cctctgatgc attctcgggc 3300
ctcaggtttc aatcaggcac cagcaggggc tatcagaggc aaatggttcc tcgagaagag 3360
attaaaagca catcagcaaa gagcagcgtg gtcagcctag ctgaccaggc acctaatctc 3420
tccaggctcc tgtctaattg aactgtcatg gaaggttctc cagtgaactt ggaagttgaa 3480
gtaacaggat ttccagagcc tacactgaca tgggtgggtag cctataatga caagccataa 3540
atggaaacaa attcaatcac agagaaaaat cattctgtga gcactaactg aaaggtttag 3600
ggctagctga ttaatatct atgacactgc aactctgcat gattcaatct cacatcagac 3660
ccctctcatt ttagtagcag catagttaat acctttaaga aaaataaaaag gtaaccatat 3720
aaagtact                                     3728

```

<210> 27

<211> 2241

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5507629CB1

<400> 27

```

taattgcgta cattttgctt ggaaatcaca ggaagaaaca caacaaagca ttgcccagg 60
tacagaaaca ataataaacg tacttaaaac tcatttaaat tccagcattt ttctccttat 120
tttagaagtt taacatttca gaagcagaca ctgcctccct tcctgcaaca actttctcct 180
cttaagtctc atttttcccc agtttcaagg gcgaattcta gaactccccg gaaccgccac 240
cagttaacca gaattccgtg gctttggaaa taaaactgct gttatagctc ttctgggtat 300
ttgagaaatg cacttgtgaa ggggttagagt tgaatctttt gatgcgaaag tcgggttttc 360
ctgatactgg gattccggga ttccagggtg tgggggtggc caattcctgc gagaagcaat 420
agcggggcgt aacatgagga gcacgggtgcg tccagcagat ccttccgcct ggggccctgc 480
cgacccccct cctgtgcccc caggactctg gcctcacccg gccgtgccgg ggcctctgtg 540
acgcgggcgt ccaggcactc ggccccggcc gagcccgtag ctagagcggc tcagagacag 600
gagggcgccg cagcagcggc ggcatgaacc actgccagct accgggtggg atcgacaacg 660
gctcggggaat gatcaaggcg ggcgtggctg ggtgccggga gccccagttt atctaccgga 720
acattatcgg ccgcgccaag ggccagagcc gcgcggccca gggcgggcta gaactctgcg 780
tgggcgacca agctcaggac tggaggagct cgctgttcat cagttacca gtggagcgtg 840
gtctcattac ttcatgggag gacatggaga tcatgtggaa gcatactat gactataacc 900
taaagctgaa gccgtgtgat ggcccagttc tgattactga gccagcgtg aaccactgg 960
ccaaccggca acagatcacg gaaatgtttt ttgagcatct ggggtgttct gccttctata 1020
tgtccatcca ggctgtgctg gctctctttg ctgctggctt cactactggc cttgtgctga 1080
attcagggtg tgggggttacc cagagtgtgc ccattcttga ggggttactgt ctgcctcatg 1140
gtgtgcagca actggatctg gcaggccttg acctcaccaa ctacctcatg gtgctaata 1200
agaaccatgg tatcatgttg ctcatgtctt cagacagaaa gattgttgaa gacatcaagg 1260
agagcttttg ttatgtggca atgaactacg aagaggaaat ggccaagaaa cccgattgtc 1320
tagagaaagt ttaccaacta cctgatggga aggtcatcca gctccatgac cagctctttt 1380
cttgtccaga ggccctcttc tctccgtgtc atatgaacct tgaggccctt ggcattgata 1440
agatatgctt cagcagcata atgaaatgtg atacaggcct gaggaattcc ttcttttcca 1500
atattatcct tgccggggga tcaacctctt tccctggttt agacaagcgg ttagttaagg 1560
atatagcaaa ggtggctcct gccaacaccg ctgtgcaagt tatagctcct ccagaaagga 1620
aaatatcagt gtggatggga ggttctatct ttgcatcctt gtctgccttc caggacatgt 1680

```

```

ggatcactgc tgcagaatth aaagaagttg gacccaacat agtacaccaa agatgcttct 1740
gaaatacaga taaaatgggt ggaagaaaat gttttgagta tatgtgacag aaaacttttg 1800
atattatatg tttctgggag aagagaaaat acttcaccta ttgggatgcc aatattttctg 1860
ttgtatttct ataatgggtt tgggggataa taatgggtgaa gctcaagaac agatgtctat 1920
tgagtagaac caagttaaaa taatgtttcc catagtgttt cttctataac ttgacgtttg 1980
tgagcttata tttcccttgg aagagagcat ttgtggtaca atatgctatg tgccaaatga 2040
gtgataagat ttaagcttat tgaagtttag ggaaagaagg ttgctgtggt gaggaacgag 2100
actccatagc agaggtatgc catcatggaa ggggtggcat tgggatggag cgcagatatc 2160
caggcaagca tactaaaatg aacaagttgc taaagatgag aatgaacaaa gcatattcag 2220
ggcatgttta atagactgat t

```

<210> 28

<211> 5203

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5607780CB1

<400> 28

```

ggtttgatgg tcctgggtgga agtagccttg gcggctgctg gttttcaaag cctctgacaa 60
agctgtgcat atcacccgtg actgcaccct gaggaagac gaaaagtcag cctctccctt 120
acgtaggacg cttgtaaaact ttcccagacg cactgttggt cttagaattt gttctgacct 180
tgacattttg caggacttgt ccaaggtgga tctgagcctc ctcatgtgct ccaggcagag 240
atcaggattc ggatgcatca ccaactgggt gaagatgggg actaggcacc ctgcacaccc 300
tgcacagccg gaagaattaa cctcgagtct gcacgctttt aagaacaagg cctttaaaaa 360
atccaaagtg tgtggagttt gcaaacaaat tattgacggt caaggtattt catgccgagc 420
ctgcaagtat tcctgccaca agaaatgtga agccaagggt gtgattccct gccgtgtgca 480
agtccgactg gaacaggctc caggagttc cacgctgtcc agttctctct gccgtgataa 540
acctctgcgg ccctcatcc tgagtcccac catggaggag ggccatgggc tggacctcac 600
ttacatcacg gagcgcatca tcgtgtgtc cttccctgcc ggctgctctg aggagtccta 660
cctgcacaac ctacaggagg tcacgcgat gctcaagtcc aagcacgggg acaactacct 720
ggtattaaac ctttcagaaa agagatatga ccttacgaag cttaaccaa agatcatgga 780
tgtgggctgg ccagagctcc acgcaccgcc cctggataag atgtgtacca tatgcaaggc 840
gcaggagtcc tggctgaaca gcaacctcca gcatgtggtc gtcattcact gcaggggagg 900
gaaaggacgc ataggagtgg tcatatcatc ctacatgcat ttcaccaacg tctcagccag 960
cgccgaccag gcccttgaca ggtttgcaat gaagaagttt tatgatgaca aagtttcagc 1020
tttaatgcag ccttcccaaa aacggtatgt tcagttcctc agtgggctcc tgtccggatc 1080
ggtgaaaatg aatgcctctc ccctgttcct gcattttgtc atcctccacg gcacccccaa 1140
cttcgacaca ggtggagtgt gccggccctt tctgaagctc taccaagcca tgcagcctgt 1200
gtacacctcc gggatctaca acgttgcccc agaaaacccc agcaggatct gcacgtcat 1260
cgagccggcc cagcttctga agggagatgt catggtgaaa tgctaccaca agaaataaccg 1320
ctcggccacc cgtgacgtca tttccgcct gcagtttcac actggggctg tgcagggcta 1380
cgggctgggtg tttgggaagg aggatctgga caatgccagc aaagatgacc gttttcctga 1440
ctatgggaag gttgaattag tcttctctgc cacgcctgag aagattcaag ggtccgaaca 1500
cttgataaac gaccacggtg tgattgtgga ctacaacaca acagaccac tgatacgtg 1560
ggactcgtac gagaacctca gtgcagatgg agaagtgcta cacacgcagg gccctgtcga 1620
tggcagcctt tacgcgaagg tgaggaagaa aagctcctcg gatcctggca tcccagggtg 1680
ccccaggca atcccggcca ccaacagccc agaccacagt gaccacacct tgtctgtcag 1740
cagtgactcc ggccactcta cagcctctgc caggacggat aagacggaag agcgcctggc 1800
cccaggaacc aggaggggcc tgagtgtcca ggagaaggct gagttggacc agctgctcag 1860
tggctttggc ctggaagatc ctggaagctc cctcaaggaa atgactgatg ctggaagcaa 1920
gtacagtggg acccgccacg tgggtgccag ccaggttcac gtgaatggag acgctgctct 1980
gaaggatcgg gagacagaca ttctggatga cgagatgccc caccacgacc tgcacagtgt 2040
ggacagcctt gggacctgt cctcctcgga agggcctcag tcggcccacc tgggtccctt 2100
cacctgccac aagagcagcc agaactcact cctatctgac ggttttgga gcaacgttgg 2160

```

tgaagatccg	cagggcaccc	tcgttccgga	cctgggcctt	ggcatggacg	gccctatga	2220
gcgggagcgg	acttttggga	gtcgagagcc	caagcagccc	cagccccctgc	tgagaaagcc	2280
ctcagtgtcc	gcccagatgc	aggcctatgg	gcagagcagc	tactccacac	agacctgggt	2340
gcgccagcag	cagatggttg	tagctcacca	gtatagcttc	gccccagatg	gggaggcccg	2400
gctggtgagc	cgctgccctg	cagacaatcc	tggcctcgtc	cagggcccagc	ccagagtgcc	2460
actcaccccc	acccgagggga	ccagcagtag	ggtggctgtc	cagaggggtg	taggcagtgg	2520
gccacatccc	cctgacacac	agcagccctc	tcccagcaaa	gcgttcaaac	ccaggtttcc	2580
aggagaccag	gttgtgaatg	gagccggccc	agagctgagc	acaggccctc	ccccaggctc	2640
gccaccctcg	gacatcgacc	agtccatcga	gcagctcaac	aggctgatcc	tggagctgga	2700
tcccaccttc	gagcccatcc	ctaccacatc	gaacgccctc	ggtagccagg	ccaatggctc	2760
tgtgtctcca	gacagcgtgg	gaggcgggct	ccgggcaagc	agcaggctgc	ctgacacagg	2820
agagggcccc	agcagggcca	ccgggcggca	aggctcctct	gctgaacagc	ccctgggcgg	2880
gagactcagg	aagctgagcc	tggggcagta	cgacaacgat	gctggggggc	agctgccctt	2940
ctccaaatgt	gcatggggaa	aggctggtgt	ggactatgcc	ccaaacctgc	cgccattccc	3000
ctcaccagcg	gacgtcaaag	agacgatgac	ccctggctat	ccccaggacc	tcgatattat	3060
cgatggcaga	attttaagta	gcaaggagtc	catgtgttca	actccagcat	ttcctgtgtc	3120
tccagagaca	ccttatgtga	aaacagcgct	gcgccatcct	ccgttcagcc	cacctgagcc	3180
cccgtgagc	agcccagcca	gtcagcacia	aggaggacgt	gaaccacgaa	gctgccctga	3240
gacgctcact	cacgctgtgg	ggatgtcaga	gagccccatc	ggacccaaat	ccacgatgct	3300
ccgggctgat	gcgtcctcga	cgccctcctt	tcagcaggct	tttgcttctt	cctgcaccat	3360
ttccagcaac	ggccctgggc	agaggagaga	gagctcctct	tctgcagaac	gccagtgggt	3420
ggagagcagc	cccaagccca	tggtttccct	gctggggagc	ggccggccca	ccggaagtcc	3480
cctcagcgct	gagttctccg	gtaccaggaa	ggactcccca	gtgctgtcct	gcttcccggc	3540
gtcagagctc	caggctcctt	tccacagcca	tgagctgtcc	ctagcagagc	caccggactc	3600
cctggcgccct	cccagcagcc	aggccttccct	gggcttcggc	accgccccag	tgggaagtgg	3660
ccttccgccc	gaggaggacc	tgggggcctt	gctggccaat	tctcatggag	cgtcaccgac	3720
cccagcatc	ccgctgacag	cgacaggggc	tgccgacaat	ggcttccctgt	cccacaactt	3780
tctcacggtg	gcgcctggac	acagcagcca	ccacagtcca	ggcctgcagg	gccagggtgt	3840
gaccctgccc	gggcagccac	ccctccctga	gaagaagcgg	gcctcggagg	gggatcggtc	3900
tttgggctca	gtctctccct	cctccagtgg	cttctccagc	ccgcacagcg	ggagcaccat	3960
cagtatcccc	ttcccaaagt	tccttcccga	cttttccaag	gcttcagaag	cggcctcacc	4020
tctgccagat	agtccagggt	ataaacttgt	gatcgtgaaa	tttgttcaag	acacttccaa	4080
gttctggtac	aaggcgata	tttcaagaga	acaagccatc	gccatgttga	aggacaagga	4140
gccgggctca	ttcattgttc	gagacagcca	ttccttccga	ggggcctatg	gcctggccat	4200
gaagggtggc	acgcccccac	cttcagtcct	gcagctgaac	aagaaagctg	gagatttggc	4260
caatgaactc	gtccggcact	ttttgatcga	gtgtaccccg	aaggagtgct	ggttgaaagg	4320
gtgctcgaat	gaaccatatt	tcgggagcct	gacggccttg	gtgtgccagc	attccatcac	4380
gcccttggcc	ttgccttgca	agctgccttat	cccagagaga	gatccattgg	aggaaatagc	4440
agaaagtctt	ccccagacgg	cagccaattc	agcagctgag	ctgttgaagc	agggggcagc	4500
ctgcaacgtg	tgggtacttga	actctgtgga	gatggagtcc	ctcaccggcc	accaggcgat	4560
ccagaaggcc	ctgagcatca	ccctgggtcca	ggagcctcca	cctgtgtcca	cagttgtgca	4620
cttcaagggt	tcagcccagg	gcatcaccct	gacagacaat	cagaggaagc	tcttcttccg	4680
gaggcattac	cccgtgaaca	gtgtgatttt	ctgtgccttg	gaccacaag	acaggaagtg	4740
gatcaaagat	ggcccttcct	caaaagtctt	tggatttgtg	gcccggagc	agggcagtg	4800
cacggataat	gtgtgccacc	tgtttgcaga	gcatgaccct	gagcagcctg	ccagtgccat	4860
tgtcaacttc	gtatcaaagg	tcatgattgg	ttccccaaag	aaggctctgag	aactccccct	4920
cctccctgga	cccaccgatg	cctctcgaag	ccctggagac	agccgttggg	tgaggggtgg	4980
gccccactt	tttaccaaac	tagtaaacct	gacattccag	gcccagtagg	ggaaagagga	5040
tcttccagct	ctgcaaaaac	agaacaaac	aacatcaccg	tgaattggcc	tttccctgaa	5100
gtgacttatc	tgacacatct	ctgtagccac	atgctttttg	ggtagaagaa	gctggggcatg	5160
ggtgcacccc	accccttagg	gtcccatggt	gaaagggaca	tgc		5203