

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
13 February 2003 (13.02.2003)

PCT

(10) International Publication Number
WO 03/012065 A2(51) International Patent Classification⁷: **C12N**

(21) International Application Number: PCT/US02/24521

(22) International Filing Date: 1 August 2002 (01.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/309,627	2 August 2001 (02.08.2001)	US
60/310,933	7 August 2001 (07.08.2001)	US
60/311,323	9 August 2001 (09.08.2001)	US
60/317,820	7 September 2001 (07.09.2001)	US
60/322,264	14 September 2001 (14.09.2001)	US
60/326,098	28 September 2001 (28.09.2001)	US
60/343,007	19 December 2001 (19.12.2001)	US
60/364,494	15 March 2002 (15.03.2002)	US
60/375,539	24 April 2002 (24.04.2002)	US

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street A # 205, San Leandro, CA 94577 (US). **HE, Ann** [CN/US]; 4601 Catalina Drive, San Jose, CA 95129 (US). **AU-YOUNG, Janice, K.** [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). **LEE, Soo Yeun** [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **LAL, Preeti, G.** [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **ISON, Craig, H.** [US/US]; 1242 Weathersfield Way, San Jose, CA 95118 (US).

YANG, Junming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **LI, Joana, X.** [US/US]; 1264 Geneva Avenue, San Francisco, CA 94112 (US). **EMERLING, Brooke, M.** [US/US]; 1735 Woodland Avenue # 71, Palo Alto, CA 94303 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **WARREN, Bridget, A.** [US/US]; 1810 S. El Camino Real #B103, Encinitas, CA 94024 (US). **HAFALIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **MARQUIS, Joseph, P.** [US/US]; 4428 Lazy Lane, San Jose, CA 95135 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

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

(54) Title: KINASES AND PHOSPHATASES

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



WO 03/012065 A2

KINASES AND PHOSPHATASES

TECHNICAL FIELD

The invention relates to novel nucleic acids, kinases and phosphatases encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and kinases and phosphatases,

BACKGROUND OF THE INVENTION

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups. Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors can activate kinases, which can occur as cell surface receptors or as the activator of the final effector protein, as well as other locations along the signal transduction pathway. Cascades of kinases occur, as well as kinases sensitive to second messenger molecules. This system allows for the amplification of weak signals (low abundance growth factor molecules, for example), as well as the synthesis of many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and cell cycle progression.

KINASES

Kinases comprise the largest known enzyme superfamily and vary widely in their target molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the acceptor molecule, causing internal conformational changes and potentially influencing intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response

to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

5 Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked
10 to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine
15 residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to
20 the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence
25 motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Book, Vol I, pp. 17-20 Academic Press,
30 San Diego CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

35 **Protein Tyrosine Kinases**

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) *Annu. Rev. Cell Biol.* 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Substrates for tyrosine kinases can be identified using anti-phosphotyrosine antibodies to screen tyrosine-phosphorylated cDNA expression libraries. Fish, so named for tyrosine-phosphorylated in Src-transformed fibroblast, is a tyrosine kinase substrate which has been identified by such a technique. Fish has five SH3 domains and a phospho homology (PX) domain. Fish is suggested to be involved in signalling by tyrosine kinases and have a role in the actin cytoskeleton (Lock, P. et al (1998) *EMBO J.* 17:4346-4357).

SHP-2, an SH2-domain-containing phosphotyrosine phosphatase, is a positive signal transducer for several receptor tyrosine kinases (RTKs) and cytokine receptors. Phosphotyrosine phosphatases are critical positive and negative regulators in the intracellular signalling pathways that result in growth-factor-specific cell responses such as mitosis, migration, differentiation, transformation, survival or death. Signal-regulatory proteins (SIRPs) comprise a new gene family of at least 15 members, consisting of two subtypes distinguished by the presence or absence of a cytoplasmic SHP-2-binding domain. The SIRP-alpha subfamily members have a cytoplasmic SHP2-binding domain and includes SIRP-alpha-1, a transmembrane protein, a substrate of activated RTKs and which binds to SH2 domains. SIRPs have a high degree of homology with immune antigen recognition molecules. The SIRP-beta subfamily lacks the cytoplasmic tail. The SIRP-beta-1 gene encodes a polypeptide of 398 amino acids. SIRP family members are generally involved in regulation of signals which define different physiological and pathological processes (Kharitonov, A. et al

(1997) Nature 386:181-186). Two possible areas of regulation include determination of brain diversity and genetic individuality (Sano, S et al (1999) Biochem. J. 344 Pt 3:667-675) and recognition of self which fails in diseases such as hemolytic anemia (Oldenborg, P.-A et al (2000) Science 288:2051-2054).

5 **Protein Serine/Threonine Kinases**

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

One member of the ERK family of MAP kinases, ERK 7, is a novel 61-kDa protein that has motif similarities to ERK1 and ERK2, but is not activated by extracellular stimuli as are ERK1 and ERK2 nor by the common activators, c-Jun N-terminal kinase (JNK) and p38 kinase. ERK7 regulates its nuclear localization and inhibition of growth through its C-terminal tail, not through the kinase domain as is typical with other MAP kinases (Abe, M.K. (1999) Mol. Cell. Biol. 19:1301-1312).

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's

Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) J. Biol. Chem. 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) J. Biol. Chem. 273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms, α , β , γ , δ , and ϵ . Fish et al. identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al., *supra*).

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state. Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al. have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) Science 288:483-491).

Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al. (2000) Biochimie 82:1123-1127). HIPKs contain a conserved protein kinase domain separated from a domain that interacts with homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim, Y.H. et al. (1998) J. Biol. Chem. 273:25875-25879; Wang, Y. et al. (2001) Biochim. Biophys. Acta 1518:168-172). HIPKs act as corepressors for homeodomain transcription factors. This corepressor

activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each of which are important in the regulation of cellular protein function (Kim, Y.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:12350-12355).

The human h-warts protein, a homolog of *Drosophila* warts tumor suppressor gene, maps to
5 chromosome 6q24-25.1. It has a serine/threonine kinase domain and is localized to centrosomes in interphase cells. It is involved in mitosis and functions as a component of the mitotic apparatus (Nishiyama, Y. et al. (1999) FEBS Lett. 459:159-165).

The UNC-51 serine/threonine kinase of Caenorhabditis elegans is required for axon formation. Its murine homolog is expressed in granule cells of the cerebellar cortex (Tomoda, T. et al. (1999) Neuron 24:833-846). The human homolog of UNC-51, ULK1 (UNC-51 (*C. elegans*)-like
10 kinase 1), is highly conserved among vertebrates. It is composed of 1050 amino acids, has a calculated MW of 112.6 kDa and a pI of 8.80. ULK1 is ubiquitously expressed in adult tissues while UNC-51 has been specifically located in the nervous system of C. elegans. ULK1 has been mapped to human chromosome 12q24.3 (Kuroyanagi, H. et al. (1998) Genomics 51:76-85).

15 **Calcium-Calmodulin Dependent Protein Kinases**

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also
20 activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) EMBO J. 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of
25 catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be enriched in mammalian forebrain. This protein is associated with vesicles in both axons and
30 dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

Mitogen-Activated Protein Kinases

The mitogen-activated protein kinases (MAP), which mediate signal transduction from the
35 cell surface to the nucleus via phosphorylation cascades, are another STK family that regulates

intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) *Nature* 365:781-783). There are three kinase modules comprising the MAP kinase cascade: MAPK (MAP), MAPK kinase (MAP2K, MAPKK, or MKK), and MKK kinase (MAP3K, MAPKKK, OR MEKK) (Wang, X.S. et al (1998) *Biochem. Biophys. Res. Commun.* 253:33-37). The extracellular-regulated kinase (ERK) pathway is activated by growth factors and mitogens, for example, epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, or endotoxic lipopolysaccharide (LPS). The closely related though distinct parallel pathways, the c-Jun N-terminal kinase (JNK), or stress-activated kinase (SAPK) pathway, and the p38 kinase pathway are activated by stress stimuli and proinflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.. MAP kinase signaling pathways are present in mammalian cells as well as in yeast.

MAPKKK6 (MAP3K6) is one of numerous MAP3Ks identified. Isolated from skeletal muscle, MAP3K6 is 1,280 amino acids in length with 11 kinase subdomains and is detected in several tissues. The highest expression has been found in heart and skeletal muscle. MAP3K6 has 45% amino acid sequence identity with MAP3K5, while their catalytic domains share 82% identity. MAP3K6 interaction with MAP3K5 *in vivo* was confirmed by coimmunoprecipitation. Recombinant MAP3K6 has been shown to weakly activate the JNK but not the p38 kinase or ERK pathways (Wang, X.S. et al. *supra*)

Cyclin-Dependent Protein Kinases

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) *EMBO J.* 17:470-481).

Checkpoint and Cell Cycle Kinases

In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome

segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) Science 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) Science 277:1501-1505). Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-19408). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) J. Biol Chem. 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune diseases, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a

C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors

5 (Sanjo, H. et al. (1998) J. Biol. Chem. 273:29066-29071). DRAK1 and DRAK2 are STKs that share homology with the death-associated protein kinases (DAP kinases), known to function in interferon- γ induced apoptosis (Sanjo et al., *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with

10 apoptosis when transfected into NIH3T3 cells (Sanjo et al., *supra*). However, deletion of either the N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving

15 the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8

20 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the

25 CD95 apoptosis pathway (Inohara et al., *supra*).

Mitochondrial Protein Kinases

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the

30 mitochondrial matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes,

35 regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory

enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) Adv. Enzyme Regul. 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) *supra*).

KINASES WITH NON-PROTEIN SUBSTRATES

10 Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) Curr. Opin. Cell Biol. 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP₂). PIP₂ is then cleaved into inositol triphosphate (IP₃) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) biphosphate (PIP₂) to PI (3,4,5) P₃ (PIP₃). PIP₃ then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) Biochem. J. 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, *supra*).

An example of lipid kinase phosphorylation activity is the phosphorylation of D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) J. Biol. Chem. 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., *supra*).

Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) Cancer Res. 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zelevnikar, R.J. et al. (1995) J. Biol. Chem. 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity in order to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenylyl cyclase, and production

of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and GTP levels also control the activity of certain oncogenic proteins such as p21^{ras} known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) Cancer Res. 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of p21^{ras} and promote
5 oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and bucciclovir (Miller, W.H. and R.L. Miller (1980) J. Biol. Chem. 255:7204-7207;
10 Stenberg, K. et al. (1986) J. Biol. Chem. 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2.

15 Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and S.
20 Eriksson (1995) Pharmacol. Ther. 67:155-186).

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues
25 serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups. Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors can activate kinases, which can occur as cell surface receptors or as the activator of the final effector protein, but can also occur along the signal transduction pathway. Cascades of kinases occur, as well as kinases sensitive to second messenger molecules. This system allows for the amplification of
30 weak signals (low abundance growth factor molecules, for example), as well as the synthesis of many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and cell cycle progression.

PHOSPHATASES

Protein phosphatases are generally characterized as either serine/threonine- or tyrosine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs, for dual specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues.

- 5 The protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle and cell signaling processes. Another family of phosphatases is the acid phosphatase or histidine acid phosphatase (HAP) family whose members hydrolyze phosphate esters at acidic pH conditions.

- 10 PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal and membranous structures in most tissues, especially the brain. Some PSPs require divalent cations, such as Ca^{2+} or Mn^{2+} , for activity. PSPs play important roles in glycogen metabolism, muscle contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic metabolism (reviewed in Cohen, P. (1989) *Annu. Rev. Biochem.* 58:453-508). PSPs can be separated into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7.
- 15 Members of this class are composed of a homologous catalytic subunit bearing a very highly conserved signature sequence, coupled with one or more regulatory subunits (PROSITE PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular localization of PSPs and substrate specificity. The PPM class consists of several closely related isoforms of PP2C and is evolutionarily unrelated to the PPP class.

- 20 PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A number of isoforms have been identified, with the alpha and beta forms being produced by alternative splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-
- 25 monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) *Curr. Opin. Neurobiol.* 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) *Otolaryngol. Head Neck Surg.* 121:463-468).

- 30 PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A et al. (1999) *Trends Biosci.* 24:186-191). These "B-type" subunits are cell type- and tissue-
- 35 specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the

holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII
5 kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen. Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through the cell cycle (Pallas, D.C. et al. (1992) J. Virol. 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development, cardiovascular diseases, and lipid disorders, and in the
10 assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of kinases. PP2A, in fact, can dephosphorylate and modulate the activities of more than 30 protein kinases *in vitro*, and other evidence suggests that the same is true *in vivo* for such kinases as PKB, PKC, the calmodulin-dependent kinases, ERK family MAP kinases, cyclin-dependent kinases, and the I κ B kinases (reviewed in Millward et al., *supra*). PP2A is itself a substrate for CKI and CKII
15 kinases, and can be stimulated by polycationic macromolecules. A PP2A-like phosphatase is necessary to maintain the G1 phase destruction of mammalian cyclins A and B (Bastians, H. et al. (1999) Mol. Biol. Cell 10:3927-3941). PP2A is a major activity in the brain and is implicated in regulating neurofilament stability and normal neural function, particularly the phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been proposed to lead to the
20 neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, *supra*).

PP2B, or calcineurin, is a Ca²⁺-activated dimeric phosphatase and is particularly abundant in the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the calcium/calmodulin complex. Calcineurin is the target of the immunosuppressant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit
25 phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and
30 Mumby, *supra*).

Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci. 22:245-251). One of them, PP5, contains regulatory domains with tetratricopeptide repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids *in vitro* and appears to be involved in a number of signaling pathways, including those controlled by atrial
35 natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell

Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent cations (mainly Mn^{2+} or Mg^{2+}) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M. et al. (1998) EMBO J. 17:4744-4752).

In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site consensus sequence includes a cysteine residue which executes a nucleophilic attack on the phosphate moiety during catalysis (Neel, B.G. and N.K. Tonks (1997) Curr. Opin. Cell Biol. 9:193-204).

Receptor PTPs are made up of an N-terminal extracellular domain of variable length, a transmembrane region, and a cytoplasmic region that generally contains two copies of the catalytic domain. Although only the first copy seems to have enzymatic activity, the second copy apparently affects the substrate specificity of the first. The extracellular domains of some receptor PTPs contain fibronectin-like repeats, immunoglobulin-like domains, MAM domains (an extracellular motif likely to have an adhesive function), or carbonic anhydrase-like domains (PROSITE PDOC 00323). This wide variety of structural motifs accounts for the diversity in size and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte activation, and cell proliferation. PTPs μ and κ are involved in cell-cell contacts, perhaps regulating cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility, most of them via the integrin/tyrosine kinase signaling pathway (reviewed in Neel and Tonks, *supra*). CD45 phosphatases regulate signal transduction and lymphocyte activation (Ledbetter, J.A. et al. (1988) Proc. Natl. Acad. Sci. USA 85:8628-8632). Soluble PTPs containing Src-homology-2 domains have been identified (SHPs), suggesting that these molecules might interact with receptor tyrosine kinases. SHP-1 regulates cytokine receptor signaling by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell receptor and c-Kit (reviewed in Neel and Tonks, *supra*). M-phase inducer phosphatase plays a key role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell division (Sadhu, K. et al. (1990) Proc. Natl. Acad. Sci. USA 87:5139-5143). In addition, the genes encoding at least eight PTPs have been mapped to chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma, leukemia, adenocarcinoma, and

neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992) *Annu. Rev. Cell Biol.* 8:463-493). The PTP enzyme active site comprises the consensus sequence of the MTM1 gene family. The MTM1 gene is responsible for X-linked recessive myotubular myopathy, a congenital muscle disorder that has been linked to Xq28 (Kioschis, P. et al., (1998) *Genomics* 54:256-266). Many PTKs are encoded by oncogenes, and it is well known that oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may serve to prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine phosphorylation in cells. This is supported by studies showing that overexpression of PTP can suppress transformation in cells and that specific inhibition of PTP can enhance cell transformation (Charbonneau and Tonks, *supra*).

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs. DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C. The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular signal-regulated kinase), JNK (c-Jun N-terminal kinase), and p38 on both tyrosine and threonine residues (PROSITE PDOC 00323, *supra*). In the activated state, these kinases have been implicated in neuronal differentiation, proliferation, oncogenic transformation, platelet aggregation, and apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996) *J. Biol. Chem.* 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri, D. and M. Ittmann (1999) *Hum. Pathol.* 30:419-424) and abnormalities in its expression are associated with numerous cancers (reviewed in Tamura, M. et al. (1999) *J. Natl. Cancer Inst.* 91:1820-1828).

Histidine acid phosphatase (HAP; EXPASY EC 3.1.3.2), also known as acid phosphatase, hydrolyzes a wide spectrum of substrates including alkyl, aryl, and acyl orthophosphate monoesters and phosphorylated proteins at low pH. HAPs share two regions of conserved sequences, each centered around a histidine residue which is involved in catalytic activity. Members of the HAP family include lysosomal acid phosphatase (LAP) and prostatic acid phosphatase (PAP), both sensitive to inhibition by L-tartrate (PROSITE PDOC00538).

LAP, an orthophosphoric monoester of the endosomal/lysosomal compartment is a housekeeping gene whose enzymatic activity has been detected in all tissues examined (Geier, C. et al. (1989) *Eur. J. Biochem.* 183:611-616). LAP-deficient mice have progressive skeletal disorder and an increased disposition toward generalized seizures (Saftig, P. et al. (1997) *J. Biol. Chem.* 272:18628-18635). LAP-deficient patients were found to have the following clinical features:

intermittent vomiting, hypotonia, lethargy, opisthotonos, terminal bleeding, seizures, and death in early infancy (Online Mendelian Inheritance in Man (OMIM) *200950).

PAP, a prostate epithelium-specific differentiation antigen produced by the prostate gland, has been used to diagnose and stage prostate cancer. In prostate carcinomas, the enzymatic activity of PAP was shown to be decreased compared with normal or benign prostate hypertrophy cells (Foti, A.G. et al. (1977) *Cancer Res.* 37:4120-4124). Two forms of PAP have been identified, secreted and intracellular. Mature secreted PAP is detected in the seminal fluid and is active as a glycosylated homodimer with a molecular weight of approximately 100-kilodalton. Intracellular PAP is found to exhibit endogenous phosphotyrosyl protein phosphatase activity and is involved in regulating prostate cell growth (Meng, T.C. and M.F. Lin (1998) *J. Biol. Chem.* 34:22096-22104).

Synaptojanin, a polyphosphoinositide phosphatase, dephosphorylates phosphoinositides at positions 3, 4 and 5 of the inositol ring. Synaptojanin is a major presynaptic protein found at clathrin-coated endocytic intermediates in nerve terminals, and binds the clathrin coat-associated protein, EPS15. This binding is mediated by the C-terminal region of synaptojanin-170, which has 3 Asp-Pro-Phe amino acid repeats. Further, this 3 residue repeat had been found to be the binding site for the EH domains of EPS15 (Haffner, C. et al. (1997) *FEBS Lett.* 419:175-180). Additionally, synaptojanin may potentially regulate interactions of endocytic proteins with the plasma membrane, and be involved in synaptic vesicle recycling (Brodin, L. et al. (2000) *Curr. Opin. Neurobiol.* 10:312-320). Studies in mice with a targeted disruption in the synaptojanin 1 gene (*Synj1*) were shown to support coat formation of endocytic vesicles more effectively than was seen in wild-type mice, suggesting that *Synj1* can act as a negative regulator of membrane-coat protein interactions. These findings provide genetic evidence for a crucial role of phosphoinositide metabolism in synaptic vesicle recycling (Cremona, O. et al. (1999) *Cell* 99:179-188).

Expression profiling

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

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling

cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder. For example, both the levels and sequences expressed in T cells from subjects with autoimmune disease may be compared with the levels and sequences expressed in T cells from normal subjects.

- 5 The potential application of gene expression profiling is relevant to improving the diagnosis, prognosis, and treatment of cancers, such as breast cancer and lung cancer.

Breast cancer

- More than 180,000 new cases of breast cancer are diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (Gish, K. (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).
- 10
- 15

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

- 20 The relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie, K. et al. (1993) Cancer and Metastasis Rev. 12:255-274, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S.S. et al. (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene
- 25
- 30
- 35

downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down-regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou, Z. et al. (1998) Int. J. Cancer 78:95-99; Chen, L. et al. (1990) Oncogene 5:1391-1395; Ulrix, W. et al (1999) FEBS Lett 455:23-26; Sager, R. et al. (1996) Curr. Top. Microbiol. Immunol. 213:51-64; and Lee, S.W. et al. (1992) Proc. Natl. Acad. Sci. USA 89:2504-2508).

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

Epidermal growth factor (EGF) was originally discovered in crude preparations of nerve growth factor prepared from mouse submaxillary glands as an activity that induced early eyelid opening, incisor eruption, hair growth inhibition, and stunting of growth when injected into newborn mice. Human EGF was isolated from urine based on its inhibitory effect on gastric secretion and named urogastrone, accordingly. EGF is prototypic of a family of growth factors that are derived from membrane-anchored precursors. All members of this family are characterized by the presence of at least one EGF structural unit (defined by the presence of a conserved 6-cysteine motif that forms three disulfide bonds) in their extracellular domain. A wide variety of *in vitro* and *in vivo* biological effects have been ascribed to EGF and other members of the EGF family. The best documented activity of EGF is its ability to promote proliferation and differentiation of mesenchymal and epithelial cells. *In vitro*, EGF is a mitogen for fibroblasts, epithelial and endothelial cells, and promotes colony formation of epidermal cells in culture. *In vivo*, EGF induces epithelial development, promotes angiogenesis, and inhibits gastric acid secretion.

EGF is a potent mitogen for normal as well as malignant mammary epithelial cells. In addition, EGF was found to induce tumor progression, and is highly expressed in breast carcinoma cells. EGF produces its effect by binding to its receptor, the protein tyrosine kinase EGF receptor (EGFR) also known as erbB2. Ligation of EGF to the EGFR activates the tyrosine kinase domain of EGFR and the phosphorylation of multiple substrates, thereby regulating gene expression. Since EGF plays an important role in growth as well as malignant progression, we investigated the effect of EGF on gene expression in multiple breast carcinoma cell lines. However, more recent studies have indicated that the EGFR may participate in signaling cascades unrelated to the EGF pathway.

Lung cancer

Lung cancer is the leading cause of cancer death in the United States, affecting more than 100,000 men and 50,000 women each year. Nearly 90% of the patients diagnosed with lung cancer

are cigarette smokers. Tobacco smoke contains thousands of noxious substances that induce carcinogen metabolizing enzymes and covalent DNA adduct formation in the exposed bronchial epithelium. In nearly 80% of patients diagnosed with lung cancer, metastasis has already occurred. Most commonly lung cancers metastasize to pleura, brain, bone, pericardium, and liver. The decision to treat with surgery, radiation therapy, or chemotherapy is made on the basis of tumor histology, response to growth factors or hormones, and sensitivity to inhibitors or drugs. With current treatments, most patients die within one year of diagnosis. Earlier diagnosis and a systematic approach to identification, staging, and treatment of lung cancer could positively affect patient outcome.

Lung cancers progress through a series of morphologically distinct stages from hyperplasia to invasive carcinoma. Malignant lung cancers are divided into two groups comprising four histopathological classes. The Non Small Cell Lung Carcinoma (NSCLC) group includes squamous cell carcinomas, adenocarcinomas, and large cell carcinomas and accounts for about 70% of all lung cancer cases. Adenocarcinomas typically arise in the peripheral airways and often form mucin secreting glands. Squamous cell carcinomas typically arise in proximal airways. The histogenesis of squamous cell carcinomas may be related to chronic inflammation and injury to the bronchial epithelium, leading to squamous metaplasia. The Small Cell Lung Carcinoma (SCLC) group accounts for about 20% of lung cancer cases. SCLCs typically arise in proximal airways and exhibit a number of paraneoplastic syndromes including inappropriate production of adrenocorticotropin and anti-diuretic hormone.

Lung cancer cells accumulate numerous genetic lesions, many of which are associated with cytologically visible chromosomal aberrations. The high frequency of chromosomal deletions associated with lung cancer may reflect the role of multiple tumor suppressor loci in the etiology of this disease. Deletion of the short arm of chromosome 3 is found in over 90% of cases and represents one of the earliest genetic lesions leading to lung cancer. Deletions at chromosome arms 9p and 17p are also common. Other frequently observed genetic lesions include overexpression of telomerase, activation of oncogenes such as K-ras and c-myc, and inactivation of tumor suppressor genes such as RB, p53 and CDKN2.

Genes differentially regulated in lung cancer have been identified by a variety of methods. Using mRNA differential display technology, Manda et al. (1999; Genomics 51:5-14) identified five genes differentially expressed in lung cancer cell lines compared to normal bronchial epithelial cells. Among the known genes, pulmonary surfactant apoprotein A and alpha 2 macroglobulin were down regulated whereas nm23H1 was upregulated. Petersen et al. (2000; Int J. Cancer, 86:512-517) used suppression subtractive hybridization to identify 552 clones differentially expressed in lung tumor derived cell lines, 205 of which represented known genes. Among the known genes,

thrombospondin-1, fibronectin, intercellular adhesion molecule 1, and cytokeratins 6 and 18 were previously observed to be differentially expressed in lung cancers. Wang et al. (2000; Oncogene 19:1519-1528) used a combination of microarray analysis and subtractive hybridization to identify 17 genes differentially overexpressed in squamous cell carcinoma compared with normal lung epithelium. Among the known genes they identified were keratin isoform 6, KOC, SPRC, IGFb2, connexin 26, plakofillin 1 and cytokeratin 13.

The potential application of gene expression profiling is also relevant to improving the diagnosis, prognosis, and treatment of inflammation and immune system disorders.

Immune response

The immune system provides defense against a variety of potentially pathogenic organisms, tumors, autoimmunity, and allergy. A wide range of cell types are involved in an immune response, each with different roles. Encounter with antigen triggers the innate immune response by natural killer cells, eosinophils, macrophages, and dendritic cells to produce soluble molecules such as cytokines and chemokines. These factors attract other immune cells such as T and B lymphocytes and antigen presenting cells (APC) to the site of inflammation. T and B lymphocytes mediate adaptive immune responses to pathogens such as bacteria, parasites, and viruses. Both T and B cells are able to recognize antigen through antigen-specific receptors by somatically rearranging T cell receptor (TCR) or antibody gene segments, respectively. While antibodies bind directly to antigen, TCR bind to antigen that has been proteolytically processed into 8 to 20 amino acid peptides and bound to major histocompatibility complex (MHC) molecules on the surface of B cells, macrophages, dendritic cells, and other APC.

T cells are divided into two major groups, CD4+ T helper (Th) cells, and CD8+ cytotoxic T lymphocytes (CTL). Immune responses are primarily regulated by CD4+ Th cells, which fall into two subclasses based on the kinds of cytokines they secrete. Th1 cells secrete mainly interferon-gamma and interleukin (IL)-2, regulate the responses of CTLs, B cells, and macrophages, and orchestrate the removal of intracellular pathogens. In contrast, Th2 cells secrete primarily IL-4 and IL-10 and promote the development of certain antibody responses such as IgG1, IgA, and IgE, an excess of the latter triggering allergic responses. By enhancing the antibody response Th2 cells regulate removal of extracellular pathogens, including various bacteria and parasites.

The first critical step in an adaptive immune response is the activation of effector T cells resulting in the differentiation and proliferation of T cells. The interaction of the TCR and its associated CD3 complex with peptide bound to MHC proteins presented on the surface of APC provides a necessary signal to activate T cells. This signal alone, however, does not completely activate T cells, and in some cases, suppresses or anergizes T cells. Optimal activation is achieved through a second costimulatory signal derived from the interaction of CD28 on the T cells with its

ligands, CD80 and CD86, on the APC. The aggregation of these molecules on the surface of the cell transduces signals through the cytosol, ultimately leading to the transcription of genes that induce proliferation and differentiation into effector CTL or Th cells. Antibodies to CD3 and CD28 can directly activate T cells, serving as a model for T cell activation induced by the peptide/MHC

5 complex and CD80/86.

Ligation of the TCR initiates not only stimulatory signal transduction pathways, such as the MAP kinase (mitogen-activated protein kinase)-dependent signaling pathways, and the NFAT (nuclear factor of activated T cells)-dependent signaling pathway, but also inhibitory signal transduction pathways such as the PKA (the cAMP-dependent protein kinase)-dependent signaling

10 pathway (Li, L.; Yee, C.; and Beavo, J.A. (1999) Science 283:848-851). The expression of IL-2, which is required for proliferation and differentiation of T cells, depends on the NFAT and MAP kinase pathways. The PKA pathway can inhibit these pathways, however, such that a signal through the TCR alone may induce a non-responsive state, or anergy. Costimulation through CD28 may invoke TCR-independent pathways that can overcome PKA inhibition of IL-2 expression. Signaling

15 through the TCR without the costimulatory signals renders T cells anergic, preventing them proliferating and differentiating into effector cells.

Abnormal regulation of T cell activation causes a number of diseases, disorders, and conditions of the immune system. Organ-specific autoimmune disorders, such as insulin-dependent diabetes mellitus, multiple sclerosis, rheumatoid arthritis, systemic lupus erythematosus, chronic graft

20 versus host disease, and Crohn's disease, result from an abnormal T cell response to self tissues. T cells also mediate allergic and other atopic immune responses. The ability to control T cell activation through either the TCR/CD3 complex or the CD28 costimulatory pathway could help to suppress organ transplant rejection, enhance vaccine efficacy, and eradicate tumors and pathogens such as bacteria and viruses that cause serious infectious disease.

25 The identification of molecules upregulated during T cell activation may reveal new lymphocyte-binding and antigen-processing molecules, transmembrane and secretory products, and transcription factors that may help to explain the specialized features of T cells and allow manipulation of the immune system.

Peripheral blood mononuclear cells (PBMCs) can be classified into discrete cellular

30 populations representing the major cellular components of the immune system. PBMCs contain about 52% lymphocytes (12% B lymphocytes, 40% T lymphocytes {25% CD4+ and 15% CD8+}), 20% NK cells, 25% monocytes, and 3% various cells the include dendritic cells and progenitor cells. The proportions, as well as the biology of these cellular components tend to vary slightly between healthy individuals, depending on factors such as age, gender, past medical history, and genetic

35 background.

IL-4 is a pleiotropic cytokine produced by activated T cells, mast cells, and basophils. It was initially identified as a B cell differentiation factor (BCDF) and a B cell stimulatory factor (BSF1). Subsequent to the molecular cloning and expression of both human and mouse IL-4, numerous other functions have been ascribed to B cells and other hematopoietic and non-hematopoietic cells including T lymphocytes, monocytes, macrophages, mast cells, myeloid and erythroid progenitors, fibroblasts, endothelial cells, etc. IL-4 exhibits anti-tumor effects both *in vivo* and *in vitro*. Recently, IL-4 was identified as an important regulator for the CD4+ subset (Th1-like vs. Th2-like) of development. The biological effects of IL-4 are mediated by the binding of IL-4 to specific cell surface receptors. The functional high-affinity receptor for IL-4 consists of a ligand-binding subunit (IL-4 R) and a second subunit (b chain) that can modulate the ligand binding affinity of the receptor complex. In certain cell types, the gamma chain of the IL-2 receptor complex is a functional b chain of the IL-4 receptor complex.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

SUMMARY OF THE INVENTION

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

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. Another

embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-18.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-18. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:19-36.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid

sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method

comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

5 Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active
10 fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. Other embodiments provide a method of treating a disease or
15 condition associated with decreased or abnormal expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a
20 naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample
25 comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

30 Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a

biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in
5 the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds
10 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:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ
15 ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the
20 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:1-18, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, c) a biologically active
25 fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18. 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)
30 comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness
35 in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a

polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the
5 compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide
10 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36, 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
15 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:19-36, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the
20 group consisting of SEQ ID NO:19-36, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of
25 hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

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

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

35 Table 3 shows structural features of polypeptide embodiments, including predicted motifs

and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

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

5 Table 5 shows representative cDNA libraries for polynucleotide embodiments.

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

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

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

DESCRIPTION OF THE INVENTION

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

As used herein and in the appended claims, the singular forms "a," "an," and "the" include
20 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.
25 Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed
30 as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"KPP" refers to the amino acid sequences of substantially purified KPP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and
35 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 KPP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

- 5 An “allelic variant” is an alternative form of the gene encoding KPP. 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.
- 10 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 KPP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as KPP or a polypeptide with at least one functional characteristic of KPP. Included within this definition are
- 15 polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding KPP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding KPP. 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 KPP.
- 20 Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of KPP 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
- 25 hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

- The terms “amino acid” and “amino acid sequence” can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or
- 30 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.

- 35 Amplification may be carried out using polymerase chain reaction (PCR) technologies or other

nucleic acid amplification technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of KPP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP 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 KPP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

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

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

The term “intramer” refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in

the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

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

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides
10 having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either
15 transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

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

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

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

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied
35 Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been

assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

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

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

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

35 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
 40 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 KPP or a polynucleotide encoding KPP which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:19-36 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:19-36, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:19-36 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:19-36 from related polynucleotides. The precise length of a fragment of SEQ ID NO:19-36 and the region of SEQ ID NO:19-36 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:1-18 is encoded by a fragment of SEQ ID NO:19-36. A fragment of SEQ ID NO:1-18 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-18. For example, a fragment of SEQ ID NO:1-18 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-18. The precise

length of a fragment of SEQ ID NO:1-18 and the region of SEQ ID NO:1-18 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

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

“Homology” refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of 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 similarity” between aligned polynucleotide sequences.

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

compare two nucleotide sequences, one may use blastn with the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

5 *Penalty for mismatch: -2*

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

10 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous
15 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
20 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
25 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
30 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.

35 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

5 *Open Gap: 11 and Extension Gap: 1 penalties*

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

10 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
15 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.

20 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
25 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
30 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
35 sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

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

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

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of KPP 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 KPP which is useful in any of the antibody production methods disclosed herein or known in the

art.

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

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

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

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

“Operably linked” refers to the situation in which a first nucleic acid sequence is placed in a
15 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
20 comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

“Post-translational modification” of an KPP may involve lipidation, glycosylation,
25 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 KPP.

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

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

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989; Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular
10 Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such
15 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
20 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
25 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
30 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
35 oligonucleotide selection are not limited to those described above.

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

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

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

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

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

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

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

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

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

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

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

"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
15 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
20 of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

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

provided in references such as Sambrook et al. (1989), *supra*.

A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-5 1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have 10 significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A 15 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.

20 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 85%, at least 90%, at least 91%, at least 92%, at least 25 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

Various embodiments of the invention include new human kinases and phosphatases (KPP), 30 the polynucleotides encoding KPP, and the use of these compositions for the diagnosis, treatment, or prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated 35 to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is

denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are kinases and phosphatases. For example, SEQ ID NO:1 is 67% identical, from residue E149 to residue T1181, to human liprin-alpha2 (GenBank ID g3309533) 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:1 also contains SAM (for sterile alpha motif) domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.)

Data from additional BLAST analyses provide further corroborative evidence that SEQ ID NO:1 is a protein which interacts with protein tyrosine phosphatases (PTPs), in particular LAR type PTPs. In an alternative example, SEQ ID NO:2 is 85% identical, from residue M1 to residue Q170, to human protein tyrosine phosphatase PTPCAAX1 (GenBank ID g1777755) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $2.5e-76$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from additional BLAST analyses provide corroborative evidence that SEQ ID NO:2 is a protein tyrosine phosphatase. In an alternative example, SEQ ID NO:3 is 99% identical, from residue S195 to residue F615, to human RAF1, an oncogene product having kinase activity (GenBank ID g496091) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $4.7e-226$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains a phorbol ester/diacylglycerol binding domain, a Raf-like Ras-binding domain, and a protein kinase 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:3 is a GTPase-associated protein kinase. In an alternative example, SEQ ID NO:4 is 100% identical, from residue F13 to residue K642, to human protein-tyrosine phosphatase (GenBank ID g35792) 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:4 also contains a protein-tyrosine phosphatase 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:4 is a protein-tyrosine phosphatase. In an alternative example, SEQ ID NO:7 is 100% identical, from residue M1 to residue L123 and from residue S124 to residue I180, to a human protein tyrosine phosphatase homolog (GenBank ID g4689110) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is $6.7e-93$, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. Data from BLIMPS, BLAST-PRODOM, BLAST-DOMO, and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:7 is a tyrosine phosphatase. In an alternative example, SEQ ID NO:12 is 92% identical, from residue M1 to residue S1268, to human synaptojanin (GenBank ID g2702323) 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:12 also contains a inositol polyphosphate phosphatase family catalytic domain and a

SAC I homology 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 and other BLAST analyses provide further corroborative evidence that SEQ ID NO:12 is a synaptojanin. In an alternative example, SEQ ID NO:14 is 85% identical, from residue M1 to residue L778, to mouse serine/threonine- and tyrosine-specific protein kinase, Nek1 (GenBank ID g256855) 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:14 also contains a protein kinase 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:14 is a serine/threonine- and tyrosine-specific protein kinase. SEQ ID NO:5-6, SEQ ID NO:8-11, SEQ ID NO:13-13, and SEQ ID NO:15-18 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-18 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:19-36 or that distinguish between SEQ ID NO:19-36 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by

an “exon stitching” algorithm. For example, a polynucleotide sequence identified as FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄ represents a “stitched” sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N_{1,2,3...}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an “exon-stretching” algorithm. For example, a polynucleotide sequence identified as FLXXXXXX_gAAAAA_gBBBBB_1_N is a “stretched” sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the “exon-stretching” algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the “exon-stretching” algorithm, a RefSeq identifier (denoted by “NM,” “NP,” or “NT”) may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

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

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

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

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to

assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

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

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

Various embodiments also encompass polynucleotides which encode KPP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:19-36, which encodes KPP. The polynucleotide sequences of SEQ ID NO:19-36, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding KPP. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding KPP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:19-36 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:19-36. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of KPP.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant

of a polynucleotide encoding KPP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding KPP, 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
5 alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding KPP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding KPP. For example, a polynucleotide comprising a sequence of SEQ ID NO:24 and a
10 polynucleotide comprising a sequence of SEQ ID NO:34 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:27 and a polynucleotide comprising a sequence of SEQ ID NO:35 are splice variants of each other; and a polynucleotide comprising a sequence of SEQ ID NO:29 and a polynucleotide comprising a sequence of SEQ ID NO:36 are splice variants of each other. Any one of the splice variants described above can encode a polypeptide which contains
15 at least one functional or structural characteristic of KPP.

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

Although polynucleotides which encode KPP and its variants are generally capable of
25 hybridizing to polynucleotides encoding naturally occurring KPP under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding KPP 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
30 codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding KPP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode KPP and KPP
35 derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic

polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding KPP or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:19-36 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding KPP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are

known in the art (Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

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

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

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

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter KPP-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 Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of KPP, 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
5 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
10 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.

15 In another embodiment, polynucleotides encoding KPP may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, KPP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase
20 techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of KPP, 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
25 polypeptide having a sequence of a naturally occurring polypeptide.

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

30 In order to express a biologically active KPP, the polynucleotides encoding KPP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in
35 polynucleotides encoding KPP. Such elements may vary in their strength and specificity. Specific

initiation signals may also be used to achieve more efficient translation of polynucleotides encoding KPP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding KPP 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 natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding KPP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding KPP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5:350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:6340-6344; Buller, R.M. et al. (1985) *Nature* 317:813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31:219-226; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending

upon the use intended for polynucleotides encoding KPP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding KPP can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPO1 plasmid (Invitrogen). Ligation of polynucleotides encoding KPP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria
5 containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of KPP are needed, e.g. for the production of antibodies, vectors which direct high
10 level expression of KPP 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 KPP. 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
15 vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

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

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

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-5 355).

For long term production of recombinant proteins in mammalian systems, stable expression of KPP in cell lines is preferred. For example, polynucleotides encoding KPP 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.

10 Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

15 Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *ap^r* cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to

20 methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA

25 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 (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

30 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 KPP is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding KPP can be identified by the absence of marker gene function.

Alternatively, a marker gene can be placed in tandem with a sequence encoding KPP under the

35 control of a single promoter. Expression of the marker gene in response to induction or selection

usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding KPP and that express KPP 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
5 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 KPP 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
10 fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on KPP is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience,
15 New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

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

Host cells transformed with polynucleotides encoding KPP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a
30 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 KPP may be designed to contain signal sequences which direct secretion of KPP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the
35 inserted polynucleotides or to process the expressed protein in the desired fashion. Such

modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a “prepro” or “pro” form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms
5 for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding KPP may be ligated to a heterologous sequence resulting in translation of a fusion protein
10 in any of the aforementioned host systems. For example, a chimeric KPP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of KPP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose
15 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
20 that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the KPP encoding sequence and the heterologous protein sequence, so that KPP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and
25 purification of fusion proteins.

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

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

In related embodiments, variants of KPP can be used to screen for binding of test compounds, such as antibodies, to KPP, a variant of KPP, or a combination of KPP and/or one or more variants KPP. In an embodiment, a variant of KPP can be used to screen for compounds that bind to a variant of KPP, but not to KPP having the exact sequence of a sequence of SEQ ID

- 5 NO:1-18. KPP variants used to perform such screening can have a range of about 50% to about 99% sequence identity to KPP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

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

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

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

In an embodiment, anticalins can be screened for specific binding to KPP, fragments of KPP, or variants of KPP. Anticalins are ligand-binding proteins that have been constructed based on a

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

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

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with KPP, either in solution or affixed to a solid support, and detecting the binding of KPP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

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

KPP, fragments of KPP, or variants of KPP may be used to screen for compounds that modulate the activity of KPP. Such compounds may include agonists, antagonists, or partial or

inverse agonists. In one embodiment, an assay is performed under conditions permissive for KPP activity, wherein KPP is combined with at least one test compound, and the activity of KPP in the presence of a test compound is compared with the activity of KPP in the absence of the test compound. A change in the activity of KPP in the presence of the test compound is indicative of a compound that modulates the activity of KPP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising KPP under conditions suitable for KPP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of KPP 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 KPP 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 KPP 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 KPP 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 KPP 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 KPP, e.g., by secreting KPP 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
 5 between regions of KPP and kinases and phosphatases. The expression of KPP is closely associated with allocortex and cingulate brain tissue, unstimulated astrocyte cells, ovarian tissue, and spleen tissue. In addition, examples of tissues expressing KPP can be found in Table 6 and can also be found in Example XI. Therefore, KPP appears to play a role in cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell
 10 proliferative disorders, and cancers. In the treatment of disorders associated with increased KPP expression or activity, it is desirable to decrease the expression or activity of KPP. In the treatment of disorders associated with decreased KPP expression or activity, it is desirable to increase the expression or activity of KPP.

Therefore, in one embodiment, KPP or a fragment or derivative thereof may be administered
 15 to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers 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,
 20 salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy
 25 (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation,
 30 osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder,
 35 such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed

connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia,

5 gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial

10 dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins,

15 thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease,

20 infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive

25 pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia,

30 diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS),

35 Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis,

anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic

5 gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative

10 colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a 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

15 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, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the

20 nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, 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,

25 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, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development

30 such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental

35 retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial

dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM₂ gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers 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, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease.

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

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

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

In a further embodiment, an antagonist of KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP. Examples of such disorders include, but are not limited to, those cardiovascular diseases, immune system disorders,

neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers described above. In one aspect, an antibody which specifically binds KPP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express KPP.

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

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic
10 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.

15 An antagonist of KPP may be produced using methods which are generally known in the art. In particular, purified KPP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind KPP. Antibodies to KPP 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
20 produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have advantages in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

25 For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with KPP 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
30 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 KPP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of
35 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 KPP 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 KPP may be prepared using any technique which provides for the
5 production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

10 In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain
15 antibodies may be adapted, using methods known in the art, to produce KPP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

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

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

Various immunoassays may be used for screening to identify antibodies having the desired
30 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 KPP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering KPP epitopes is generally used, but a competitive binding assay may also be
35 employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for KPP. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of KPP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions.

- 5 The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple KPP epitopes, represents the average affinity, or avidity, of the antibodies for KPP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular KPP 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
- 10 KPP-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 KPP, 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
- 15 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

20 of KPP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding KPP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene

25 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 KPP. 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 KPP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

- 30 In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102:469-475; Scanlon, K.J. et al. (1995) 9:1288-1296).
- 35 Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as

retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding KPP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in KPP expression or regulation causes disease, the expression of KPP 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 KPP are treated by constructing mammalian expression vectors encoding KPP and introducing these vectors by mechanical means into KPP-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 KPP 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). KPP 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 KPP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

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

7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver
5 polynucleotides encoding KPP to cells which have one or more genetic abnormalities with respect to the expression of KPP. 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
10 described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; *Annu. Rev. Nutr.* 19:511-544) and Verma, I.M. and N. Somia (1997; *Nature* 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding KPP to target cells which have one or more genetic abnormalities with
15 respect to the expression of KPP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing KPP 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.*
20 169:385-395). 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
25 patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; *J. Virol.* 73:519-532) and Xu, H. et al. (1994; *Dev. Biol.* 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells
30 with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding KPP 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
35 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 KPP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of KPP-coding RNAs and the synthesis of high levels of KPP 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 KPP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

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

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

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

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

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3'
10 ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous
15 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 KPP. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides,
20 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 KPP expression or activity, a compound which specifically inhibits expression of the polynucleotide
25 encoding KPP may be therapeutically useful, and in the treatment of disorders associated with decreased KPP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding KPP may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method
30 commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a
35 polynucleotide encoding KPP 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 KPP 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 KPP. 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 (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 KPP, antibodies to KPP, and mimetics, agonists, antagonists, or inhibitors of KPP.

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 KPP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, KPP 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 KPP or fragments thereof, antibodies of KPP, and agonists, antagonists or inhibitors of KPP, 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 KPP may be used for the diagnosis of disorders characterized by expression of KPP, or in assays to monitor patients being treated with KPP or agonists, antagonists, or inhibitors of KPP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for KPP include methods which utilize the antibody and a label to detect KPP 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 KPP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of KPP expression. Normal or standard values for KPP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to KPP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of KPP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding KPP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect

and quantify gene expression in biopsied tissues in which expression of KPP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of KPP, and to monitor regulation of KPP levels during therapeutic intervention.

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

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the KPP 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:19-36 or from genomic sequences including promoters, enhancers, and introns of the KPP gene.

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

Polynucleotides encoding KPP may be used for the diagnosis of disorders associated with expression of KPP. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers 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, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy

(APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel

5 syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal,

10 parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of

15 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, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and

20 mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease such as arteriovenous fistula, atherosclerosis,

25 hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve,

30 mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary

35 hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive

pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a 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, prion diseases 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 including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other

5 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, Tourette's disorder, progressive supranuclear palsy, corticobasal

10 degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal

15 dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and

20 sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such as Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM₂ gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus,

25 lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease,

30 hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland,

35 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, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease. Polynucleotides encoding KPP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat
5 ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered KPP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, polynucleotides encoding KPP may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding KPP may be labeled by standard methods and added to a fluid
10 or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding KPP in the sample indicates the presence of the associated disorder. Such assays may also be used to
15 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 KPP, 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
20 fragment thereof, encoding KPP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to
25 establish the presence of a disorder.

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

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals
35 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 KPP 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
5 polynucleotide encoding KPP, or a fragment of a polynucleotide complementary to the polynucleotide encoding KPP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

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

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

diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) *Trends Mol. Med.* 7:507-512; Kwok, P.-Y. and Z. Gu
5 (1999) *Mol. Med. Today* 5:538-543; Nowotny, P. et al. (2001) *Curr. Opin. Neurobiol.* 11:637-641).

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

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

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

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

hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

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

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

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

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected

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

A proteomic profile may also be generated using antibodies specific for KPP to quantify the
20 levels of KPP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; 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-
25 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,
30 N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

35 In another embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

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

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

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

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data 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 KPP on a physical map
5 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,
10 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
15 further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, KPP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug
20 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 KPP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application
25 WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with KPP, or fragments thereof, and washed. Bound KPP is then detected by methods well known in the art. Purified KPP 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.

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

In additional embodiments, the nucleotide sequences which encode KPP may be used in any
35 molecular biology techniques that have yet to be developed, provided the new techniques rely on

properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

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

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/309,627; U.S. Ser. No. 60/310,933; U.S. Ser. No. 60/311,323; U.S. Ser. No. 60/317,820; U.S. Ser. No. 60/322,264; U.S. Ser. No. 60/326,098; U.S. Ser. No. 60/375,539; U.S.
10 Ser. No. 60/343,007; and U.S. Ser. No. 60/364,494 are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database
15 (Incyte Genomics, Palo Alto CA) and shown in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and
20 ethanol, or by other routine methods.

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

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Invitrogen), using the recommended
30 procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or
35 preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites

of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPO1 plasmid (Invitrogen), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof.

5 Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using
10 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.

15 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
20 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
25 cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides
30 were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences were selected for extension using the techniques disclosed in
35 Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public

5 databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft,

10 D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and

15 HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs

20 based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS,

25 DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as

30 incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used,

35 the second column provides brief descriptions thereof, the third column presents appropriate

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

5 The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:19-36. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

10 Putative kinases and phosphatases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpr and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an
15 assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode kinases and phosphatases, the encoded polypeptides were analyzed by querying against PFAM models for kinases and phosphatases. Potential kinases and phosphatases were also
20 identified by homology to Incyte cDNA sequences that had been annotated as kinases and phosphatases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpr public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any
25 Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length
30 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
35 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
5 extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by
10 cDNA sequence. Intervals thus identified were then “stitched” together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences
15 were translated and compared by BLAST analysis to the genpept and gbpri 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

20 Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in
25 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
30 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 KPP Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:19-36 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other
35 implementations of the Smith-Waterman algorithm. Sequences from these databases that matched

SEQ ID NO:19-36 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

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

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:

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

30

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

35

(HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

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

VIII. Extension of KPP Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

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

High fidelity amplification was obtained by PCR using methods well known in the art. PCR

was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair

5 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.

10 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
15 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
20 sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). 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 Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and
25 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 Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step
30 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
35 and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator

cycle sequencing ready reaction kit (Applied Biosystems).

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

5 IX. Identification of Single Nucleotide Polymorphisms in KPP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:19-36 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was
10 used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated
15 algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

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

30 X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:19-36 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06

software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ - 32 P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 5 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 10 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

15 The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested 20 substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) *Science* 270:467-470; 25 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 30 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 35 complementarity and the relative abundance of each polynucleotide which hybridizes to an element

on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (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 μ l 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 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences). Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic 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
5 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
10 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
15 SSC), and dried.

Detection

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

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially.
25 Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the
30 laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that

location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the
5 two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high
10 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
15 integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibited at least about a two-fold change in expression, a signal-to-background ratio of at least 2.5, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

20 Expression

SEQ ID NO:24 and SEQ ID NO:34 showed differential expression in association with lung cancer, as determined by microarray analysis. Gene expression profiles were obtained by comparing the results of competitive hybridization experiments. Messenger RNA isolated from grossly uninvolved lung tissue with no visible abnormalities was compared to cancerous lung tissue from
25 matched donors (Roy Castle International Centre for Lung Cancer Research, Liverpool, UK). In ten matched tissue experiments, the expression of SEQ ID NO:24 was decreased by at least two-fold in lung tumor tissue as compared to as compared to normal lung tissue from matched donors. Testing involved two experiments using lung adenocarcinoma, and eight experiments using lung squamous cell carcinoma. In a separate matched tissue experiment, the expression of SEQ ID NO:34 was
30 decreased by at least two-fold in lung tumor tissue as compared to as compared to normal lung tissue from matched donors. Thus, in various embodiments, SEQ ID NO:24 and SEQ ID NO:34 can each be used for one or more of the following: i) monitoring treatment of lung cancer, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

SEQ ID NO:26 showed differential expression in association with inflammation and immune

responses, as determined by microarray analysis. PBMCs were collected from the blood of 6 healthy volunteer donors using standard gradient separation. The PBMCs from each donor were placed in culture for 2 hours in the presence or absence of recombinant IL-4. PBMCs were treated with interleukin 4 (IL-4) and compared to untreated cells. Treated PBMCs and untreated control PBMCs from the individual donors were pooled according to their respective treatment. The expression of SEQ ID NO:26 was decreased by at least two-fold in cells treated with IL-4. Thus, in various embodiments, SEQ ID NO:26 can be used for one or more of the following: i) monitoring treatment of immune system and/or inflammation disorders, ii) diagnostic assays for immune system and/or inflammation disorders, and iii) developing therapeutics and/or other treatments for immune system and/or inflammation disorders.

In another example, SEQ ID NO:26 showed differential expression in association with breast cancer, as determined by microarray analysis. The gene expression profiles of a nonmalignant mammary epithelial (HMEC) cell line, and a nonmalignant mammary gland cell line isolated from a 36-year-old woman with fibrocystic breast disease (MCF-10A), were compared to the gene expression profiles of breast carcinoma lines at different stages of tumor progression. Cell lines compared included: a) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female; b) T-47D, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with an infiltrating ductal carcinoma of the breast; c) Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female; d) BT-20, a breast carcinoma cell line derived *in vitro* from tumor mass isolated from a 74-year-old female; e) MDA-mb-231, a breast tumor cell line isolated from the pleural effusion of a 51-year old female; f) MDA-mb-435S, a spindle shaped strain that evolved from the parent line (435) isolated from the pleural effusion of a 31-year-old female with metastatic, ductal adenocarcinoma of the breast; and g) MDA-MB-468, a line isolated from the pleural effusion of a 51-year old female with metastatic adenocarcinoma of the breast. The expression of SEQ ID NO:26 was decreased by at least two-fold in the MDA-MB-468 breast adenocarcinoma cells as compared to HMEC cells. Therefore, in various embodiments, SEQ ID NO:26 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

The significant differential expression profiles of SEQ ID NO:24, SEQ ID NO:26, and SEQ ID NO:34 under various conditions, further establish their utility as diagnostic markers, tools for disease staging, or as therapeutic targets for diseases in which kinases and phosphatases are involved.

XII. Complementary Polynucleotides

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

XIII. Expression of KPP

Expression and purification of KPP is achieved using bacterial or virus-based expression systems. For expression of KPP 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 KPP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of KPP 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 KPP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

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

Examples XVII, XVIII, XIX, XX, and XXI, where applicable.

XIV. Functional Assays

KPP function is assessed by expressing the sequences encoding KPP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression
5 vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences
10 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
15 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
20 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 KPP on gene expression can be assessed using highly purified populations
25 of cells transfected with sequences encoding KPP 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.
30 Expression of mRNA encoding KPP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of KPP Specific Antibodies

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

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

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

XVI. Purification of Naturally Occurring KPP Using Specific Antibodies

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

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

XVII. Identification of Molecules Which Interact with KPP

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

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

KPP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S.

Patent No. 6,057,101).

XVIII. Demonstration of KPP Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein substrate by KPP in the presence of [γ - ^{32}P]ATP. KPP is incubated with the protein substrate,
5 ^{32}P -ATP, and an appropriate kinase buffer. The ^{32}P incorporated into the substrate is separated from free ^{32}P -ATP by electrophoresis and the incorporated ^{32}P is counted using a radioisotope counter. The amount of incorporated ^{32}P is proportional to the activity of KPP. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

10 In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma ^{32}P -ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated ^{32}P -peptide product. The binding sample then undergoes a centrifugal ultrafiltration
15 process with a membrane which will retain the product-avidin complex and allow passage of free gamma ^{32}P -ATP. The reservoir of the centrifuged unit containing the ^{32}P -peptide product as retentate is then counted in a scintillation counter. This procedure allows the assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation,
20 Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes include but are not limited to: Histone H1 (Sigma) and p34^{cdc2}kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) Methods Enzymol. 200:62-81).

In another alternative, protein kinase activity of KPP is demonstrated in an assay containing
25 KPP, 50 μl of kinase buffer, 1 μg substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10 μg ATP, and 0.5 μCi [γ - ^{32}P]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [γ - ^{32}P]ATP is removed by washing and the incorporated radioactivity is measured using a scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and resolved on a
30 12% SDS polyacrylamide gel followed by autoradiography. The amount of incorporated ^{32}P is proportional to the activity of KPP.

In yet another alternative, adenylate kinase or guanylate kinase activity of KPP may be measured by the incorporation of ^{32}P from [γ - ^{32}P]ATP into ADP or GDP using a gamma radioisotope counter. KPP, in a kinase buffer, is incubated together with the appropriate nucleotide

mono-phosphate substrate (AMP or GMP) and ^{32}P -labeled ATP as the phosphate donor. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is excised and counted. The radioactivity recovered is proportional to the activity of KPP.

In yet another alternative, other assays for KPP include scintillation proximity assays (SPA), scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of KPP activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

In another alternative, phosphatase activity of KPP is measured by the hydrolysis of para-nitrophenyl phosphate (PNPP). KPP is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β -mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) *Mol. Cell. Biol.* 14:3752-62). Alternatively, acid phosphatase activity of KPP is demonstrated by incubating KPP-containing extract with 100 μl of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 μl of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) *J. Biol. Chem.* 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of KPP in the assay.

In the alternative, KPP activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM KPP in a final volume of 30 μl containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% β -mercaptoethanol and 10 μM substrate, ^{32}P -labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30°C for 10-15 min. Reactions are quenched with 450 μl of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $\text{Na}_4\text{P}_2\text{O}_7$, and 2 mM NaH_2PO_4 , then centrifuged at $12,000 \times g$ for 5 min. Acid-soluble ^{32}P i is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) *J. Biol. Chem.* 274:23666-23672).

XIX. Kinase Binding Assay

Binding of KPP to a FLAG-CD44 cyt fusion protein can be determined by incubating KPP with anti-KPP-conjugated immunoaffinity beads followed by incubating portions of the beads (having 10-20 ng of protein) with 0.5 ml of a binding buffer (20 mM Tris-HCL (pH 7.4), 150 mM NaCl, 0.1% bovine serum albumin, and 0.05% Triton X-100) in the presence of ^{125}I -labeled FLAG-CD44cyt fusion protein (5,000 cpm/ng protein) at 4 °C for 5 hours. Following binding, beads were washed thoroughly in the binding buffer and the bead-bound radioactivity measured in a scintillation

counter (Bourguignon, L.Y.W. et al. (2001) J. Biol. Chem. 276:7327-7336). The amount of incorporated ^{32}P is proportional to the amount of bound KPP.

XX. Identification of KPP Inhibitors

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVII. KPP activity is measured for each well and the ability of each compound to inhibit KPP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance KPP activity.

XXI. Identification of KPP Substrates

A KPP "substrate-trapping" assay takes advantage of the increased substrate affinity that may be conferred by certain mutations in the PTP signature sequence of protein tyrosine phosphatases. KPP bearing these mutations form a stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a clone encoding the catalytic domain of KPP is performed using a method standard in the art or a commercial kit, such as the MUTA-GENE kit from BIO-RAD. For expression of KPP mutants in *Escherichia coli*, DNA fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding KPP or a glutathione S-transferase (GST)-KPP fusion protein. KPP mutants are expressed in *E. coli* and purified by chromatography.

The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated transfection with 20 μg of CsCl-purified DNA per 10-cm dish of cells or 8 μg per 6-cm dish. Forty-eight hours after transfection, cells are stimulated with 100 ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are lysed in 50 mM Tris-HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 $\mu\text{g}/\text{ml}$ leupeptin/5 $\mu\text{g}/\text{ml}$ aprotinin/1 mM benzamidine (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). KPP is immunoprecipitated from lysates with an appropriate antibody. GST-KPP fusion proteins are precipitated with glutathione-Sepharose, 4 μg of mAb or 10 μl of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or further purified to identify substrate molecules (Flint, A.J. et al. (1997) Proc. Natl. Acad. Sci. USA 94:1680-1685).

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for

using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to

5 the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
5532044	1	5532044CD1	19	5532044CB1	
7497825	2	7497825CD1	20	7497825CB1	90160686CA2, 90160786CA2, 90188642CA2, 90188726CA2
6464221	3	6464221CD1	21	6464221CB1	90106952CA2
90044205	4	90044205CD1	22	90044205CB1	90044484CA2
603340	5	603340CD1	23	603340CB1	90108796CA2, 90108856CA2, 90108864CA2, 90108880CA2, 90108888CA2
71059650	6	71059650CD1	24	71059650CB1	
7499961	7	7499961CD1	25	7499961CB1	90042613CA2, 90042621CA2, 90042645CA2, 90042705CA2, 90042721CA2, 90042737CA2
7500309	8	7500309CD1	26	7500309CB1	6174347CA2
2446376	9	2446376CD1	27	2446376CB1	90206557CA2
5562847	10	5562847CD1	28	5562847CB1	90189221CA2
7493671	11	7493671CD1	29	7493671CB1	90048842CA2, 90049235CA2, 90049402CA2
7499991	12	7499991CD1	30	7499991CB1	

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7499977	13	7499977CD1	31	7499977CB1	90064938CA2, 90064946CA2, 90065022CA2, 90065030CA2, 90065046CA2, 90133310CA2, 90133402CA2
8043337	14	8043337CD1	32	8043337CB1	
5947007	15	5947007CD1	33	5947007CB1	
7510576	16	7510576CD1	34	7510576CB1	9006508CA2
90111008	17	90111008CD1	35	90111008CB1	90110916CA2, 90110924CA2, 90111024CA2, 90206589CA2
7503305	18	7503305CD1	36	7503305CB1	90049151CA2, 90049159CA2, 90049175CA2, 90049191CA2, 90049275CA2, 90049283CA2, 90049334CA2

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	5532044CD1	g3309533	0.0	[Homo sapiens] liprin-alpha2 Serra-Pages, C. et al. (1998) Liprins, a family of LAR transmembrane protein-tyrosine phosphatase-interacting proteins. J. Biol. Chem. 273:15611-15620.
2	7497825CD1	g1777755	2.5E-76	[Homo sapiens] protein tyrosine phosphatase PTPCAAX1 Cates, C. A. et al. (1996) Prenylation of oncogenic human PTP(CAAX) protein tyrosine phosphatases. Cancer Lett. 110:49-55.
3	6464221CD1	g496091	4.7E-226	[Homo sapiens] RAF1 (protein kinase) Bonner, T.I. et al. (1985) Mol. Cell Biol. 5:1400-1407
4	90044205CD1	g35792	0.0	[Homo sapiens] protein-tyrosine phosphatase Krueger, N.X. et al. (1990) EMBO J. 9 (10), 3241-3252
5	603340CD1	g20196841 g3218467	3.0E-58 2.8E-54	[Mus musculus] phosphatase, orphan 1 [Gallus gallus] putative phosphatase
6	71059650CD1	g12853066 g8809636	0.0 4.3E-30	[Mus musculus] Protein kinase domain containing protein, ISS putative [Arabidopsis thaliana] receptor-like protein kinase Sato, S. et al. (2000) DNA Res. 7:31-63
7	7499961CD1	g4689110	6.7E-93	[Homo sapiens] protein tyrosine phosphatase homolog
8	7500309CD1	g36623	7.6E-17	[Homo sapiens] serine/threonine protein kinase Meyerson, M. et al. (1992) EMBO J. 11:2909-2917
9	2446376CD1	g8979826	2.8E-127	[Homo sapiens] d1776F14.2 (a novel protein member of the PTPNS (protein tyrosine phosphatase, non-receptor type substrate) family
10	5562847CD1	g4220888	3.2E-190	[Rattus norvegicus] extracellular signal-regulated kinase 7; ERK7

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Abe, M. K. et al. (1999) Extracellular signal-regulated kinase 7 (ERK7), a novel ERK with a C-terminal domain that regulates its activity, its cellular localization, and cell growth. Mol. Cell. Biol. 19:1301-1312
11	7493671CD1	g5738136	1.6E-83	[Homo sapiens] WARTS protein kinase
				Nishiyama, Y. et al. (1999) A human homolog of Drosophila warts tumor suppressor, h-warts, localized to mitotic apparatus and specifically phosphorylated during mitosis. FEBS Lett. 459:159-165
12	7499991CD1	g2702323	0.0	[Homo sapiens] synaptotagmin
				Haffner, C. et al. (1997) Synaptotagmin 1: localization on coated endocytic intermediates in nerve terminals and interaction of its 170 kDa isoform with Eps15. FEBS Lett. 419:175-180
13	7499977CD1	g14250738	3.1E-105	[Homo sapiens] (BC008838) mitogen-activated protein kinase kinase 5
14	8043337CD1	g256855	0.0	[Mus sp.] serine/threonine- and tyrosine-specific protein kinase, Nek1
				Letwin, K. et al. (1992) EMBO J. 11 (10), 3521-3531
15	5947007CD1	g12666523	1.8E-212	[Homo sapiens] protein phosphatase 1B1 43 kDa isoform
16	7510576CD1	g7573352	4.8E-23	[Arabidopsis thaliana] MAP3K delta-1 protein kinase
		727676 T27C10.5	1.1E-24	[Caenorhabditis elegans][Protein kinase] Probable tyrosine kinase with similarity to human and D. melanogaster tyrosine protein kinases
				Bateman, A. et al. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. Nucleic Acids Res 27, 260-2 (1999).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Hill, A. A. et al. Genomic analysis of gene expression in <i>C. elegans</i> . Science 290,809-812 (2000).
		624552 Matk	2.0E-23	[<i>Rattus norvegicus</i>] Brain-associated tyrosine kinase, a member of Csk subfamily of non-receptor protein tyrosine kinases, may function as a brain-specific regulator of kinases involved in the development and maintenance of the nervous system.
				Kaneko, Y. et al. Presence of alternative 5' untranslated sequences and identification of cells expressing ctk transcripts in the brain and testis. Oncogene 10, 945-52. (1995).
				Yamashita, H. et al. The Csk homologous kinase associates with TrkA receptors and is involved in neurite outgrowth of PC12 cells. J Biol Chem 274, 15059-65. (1999).
17	90111008CD1	g8979826	1.0E-63	[3' and 5' incom][<i>Homo sapiens</i>] dJ776F14.2 (a novel protein member of the PTPNS (protein tyrosine phosphatase, non-receptor type substrate) family)
		g2052056	1.0E-21	[<i>Homo sapiens</i>] SIRP-alpha1 Kharitonov, A. et al. (1997) Nature 386:181-186.
		341116 PTPNS1	9.1E-23	[<i>Homo sapiens</i>][Plasma membrane] Protein tyrosine phosphatase non-receptor type substrate 1, a receptor-like protein that is a substrate for phosphotyrosine phosphatases and a regulatory protein in signal transduction pathways. Muller, H. et al. (2001) Genes And Development 15:267-285.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		694548 Ptpns1	5.5E-22	[Rattus norvegicus][Receptor (signalling)][Plasma membrane] Protein tyrosine phosphatase non-receptor type substrate 1, a receptor-like protein that is a substrate for phosphotyrosine phosphatases and a regulatory protein in signal transduction pathways, has a role in cell adhesion signaling. Adams, S. et al. (1998) J. Immunol. 161:1853-1859.
18	7503305CD1	g5738136	2.0E-61	[Homo sapiens] WARTS protein kinase
		g4324434	1.9E-57	[Homo sapiens] large tumor suppressor 1 Tao, W. et al. (1999) Human homologue of the Drosophila melanogaster lats tumour suppressor modulates CDC2 activity. Nat. Genet. 21, 177-181 St John, M. A. et al. (1999) Mice deficient of Lats1 develop soft-tissue sarcomas, ovarian tumours and pituitary dysfunction. Nat. Genet. 21, 182-186
		340824 LATS1	1.5E-58	[Homo sapiens][Protein kinase; Transferase] Large Tumor Suppressor 1, a serine/threonine kinase that interacts with Zyxin (ZXX) and CDC2, negatively regulates CDC2, functions in cell cycle control and control of mitosis Tao, W. et al. (supra)
		587519 Lats2	3.9E-18	[Mus musculus][Nuclear] Large tumor suppressor 2, a putative protein kinase, member of a subfamily of protein kinases that includes Lats1 Yabuta, N. et al. (2000) Structure, expression, and chromosome mapping of LATS2, a mammalian homologue of the Drosophila tumor suppressor gene lats/warts. Genomics 63, 263-270

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	5532044CD1	1363	S43 S123 S128 S146 S156 S277 S314 S423 S461 S508 S531 S537 S560 S565 S647 S665 S705 S725 S740 S746 S760 S818 S865 S895 S953 S1051 S1149 S1249 S1291 S1306 S1338 T7 T180 T181 T225 T255 T257 T265 T299 T357 T363 T669 T673 T715 T728 T813 T830 T932 T943 T954 T956 T1008 T1132 T1202 T1212 T1243	N109 N759	SAM domain (Sterile alpha motif): D971-K1035, A856-V922	HMMER_PFAM
					PROTEIN LAR-INTERACTING LIPRIN-ALPHA2 PD025515: Q410-S936 PD023244: Q28-L270	BLAST_PRODROM
					PROTEIN LAR-INTERACTING COILED COIL LIKE LIPRIN-BETA1 LIPRIN-ALPHA2 PD014510: Y969-D1140	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN COILED COIL CHAIN MYOSIN REPEAT HEAVY ATPBINDING FILAMENT HEPTAD PD000002: R227-K472	BLAST_PRODOM
					TRICHOHYALIN DM03839 P37709 632-1103: E36-K490, Q598-Q751, E833-D847, L1034-D1057	BLAST_DOMO
					TRICHOHYALIN DM03839 P22793 921-1475: E27-K490, Q598-E624	BLAST_DOMO
					CYTOSKELETAL KERATIN DM01288 P14731 1-584: P141-Q488	BLAST_DOMO
					INTERMEDIATE FILAMENTS DM00061 Q01240 51-476: H160-L489	BLAST_DOMO
2	7497825CD1	170	T10 T53 Y37 Y50 Y96	N24	PHOSPHATASE PROTEIN TYROSINE PTPCAAX1 NUCLEAR PROTEIN TYROSINE PD007217: M1-E33	BLAST_PRODOM
3	6464221CD1	615	S43 S77 S177 S219 S250 S319 S571 S586 S609 T57 T448 T458 Y307	N208 N584	Phorbol esters/diacylglycerol binding domain: H139-C184	HMMER-PFAM
					Raf-like Ras-binding domain: N56-L131	HMMER-PFAM
					Protein kinase domain: V316-L575	HMMER-PFAM
					Phorbol esters / diacylglycerol binding domain: L149-A204	PROFILESCAN
					Protein kinases signatures and profile: Q411-R462	PROFILESCAN

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
					Diacylglycerol/phorbol-ester binding signature PR00008: H173-V185, L136-A150, C152-N161, N161-F172	BLIMPS-PRINTS
					Tyrosine kinase catalytic domain signature PR00109: T388-H401, Y425-L443, G474-I484, S496-I518, C544-F566	BLIMPS-PRINTS
					Serine/threonine protein kinase, phorbol ester binding, protooncogene, RAF, ATP-binding, transferase: PD006130: Y38-T138 PD009334: S195-H251 PD015890: M1-A42 PD007121: S579-F615	BLAST-PRODOM
					Protein kinase domain: DM00004 B26126 305-552: L318-F566 DM00004 P05625 351-598: L318-F566 DM00004 P00531 84-331: L318-F566 DM00004 P27966 85-332: R321-F566	BLAST-DOMO
					Phorbol esters / diacylglycerol binding domain: H139-C184	MOTIFS
					Protein kinases ATP-binding region signature: I322-K342	MOTIFS
					Serine/Threonine protein kinases active-site signature: I431-L443	MOTIFS

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
4	90044205CD1	642	S2 S22 S24 S70 S121 S135 S242 S347 S413 S427 S507 S606 T156 T215 T342 T361 T376 T448 T489 T501 Y423 Y479	N3 N133 N425 N499	Protein-tyrosine phosphatase: N101-E335, N393-D630	HMMER_PFAM
					Tyrosine specific protein phosphatases proteins BL00383: K104-V118, S129-I137, D159-S169, Q237-P249, V275-G285, R608-F623	BLIMPS_BLOCKS
					Tyrosine specific protein phosphatases signature and profiles: I253-R313, M549-L607	PROFILESCAN
					Protein tyrosine phosphatase signature PR00700: D130-I137, F438-E458, L598-H613, M614-C624	BLIMPS_PRINTS
					HYDROLASE PHOSPHATASE PROTEIN PROTEINTYROSINE PRECURSOR SIGNAL TYROSINE TRANSMEMBRANE GLYCOPROTEIN RECEPTOR PD000167: N393-Y625, N101-E335	BLAST_PRODOM
					HYDROLASE PHOSPHATASE PROTEIN PROTEINTYROSINE TYROSINE PRECURSOR SIGNAL TRANSMEMBRANE GLYCOPROTEIN RECEPTOR PD000155: R233-Y337, S507-F631	BLAST_PRODOM
					PHOSPHATASE HYDROLASE PRECURSOR SIGNAL GLYCOPROTEIN TRANSMEMBRANE PHOSPHORYLATION TYROSINE PROTEINTYROSINE RPTP PD006602: F13-E100	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN-TYROSINE-PHOSPHATASE DM00089 P23469 121-403: E63-V346, E370-D637 DM00089 P23469 405-692: S347-F635, E79-Y336 DM00089 P18433 226-510: L62-V346, E370-V628 DM00089 P18433 512-794: S347-F635, E79-Y336 Tyrosine specific protein phosphatases active site: V275-F287, V570-F582	BLAST_DOMO
5	603340CD1	241	S38 S113 S227 T12	N170	PUTATIVE PHOSPHATASE PD123939: K2-K240	BLAST_PRODOM
6	71059650CD1	2527	S145 S263 S383 S410 S520 S754 S850 S858 S870 S881 S908 S912 S954 S958 S961 S979 S990 S1001 S1044 S1066 S1143 S1159 S1204 S1233 S1238 S1245 S1345 S1536	N126 N130 N474 N720 N800 N863 N1089 N1101 N1135 N1269 N1489 N1506 N1577 N2164 N2295 N2308	Leucine Rich Repeat: R1246-E1268, S1036-S1058, E1130-P1153, S1174-P1196, N1269-S1292, H1012- K1035, H1197-K1218, T1084-T1103, K1108-K1129, Y983-C1005, N1221-Y1242	HMMER_PFAM
			S1627 S1636 S1672 S1721 S1769 S1786 S1815 S1913 S2180 S2411 S2440 S2454 S2465 S2525 T15 T46 T424		Protein kinase domain: L1885-V2128	HMMER_PFAM

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
			T475 T615 T833 T985 T1033 T1038 T1079 T1084 T1271 T1357 T1410 T1452 T1612 T2141 T2174 T2190 T2237 T2279 T2310 T2380 T2397 T2441 T2480 T2483 T2494 T2524 Y1699 Y1718		Protein kinases signatures and profile: L1970-Q2022	PROFILES CAN
					Tyrosine kinase catalytic domain signature PR00109: M1947-Q1960, Y1984-L2002, A2054-G2076, P2103-S2125	BLIMPS_PRINTS
					Transforming protein P21 RAS signature PR00449: M1335-K1356, R1381-S1403, S1443-V1456	BLIMPS_PRINTS
					PROTEIN KINASE DOMAIN DM00004 [P42679 236-470: L1885-P2123 [I49621 195-428: L1885-P2123 [I38044 100-349: E1882-P2123 [Q05609 553-797: L1885-T2068	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): G1341-T1348	MOTIFS
					Protein kinases ATP-binding region signature: L1885-K1906	MOTIFS

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
					Serine/Threonine protein kinases active-site signature: I1990-L2002	MOTIFS
7	7499961CD1	180	S16 S35 S69 S78 S124 T20 T47 T77 T109 Y44 Y66	N151 N152	Tyrosine specific protein phosphatases proteins BL00383: K57-V71, S82-I90, D112-V122 Protein tyrosine phosphatase signature PR00700: S83-I90, Y99-E119 HEMATOPOIETIC CELL PROTEIN TYROSINE PHOSPHATASE 70ZPEP HYDROLASE PD143889: S124-I180 PD16699: M1-K53 HYDROLASE PHOSPHATASE PROTEIN TYROSINE PRECURSOR SIGNAL TYROSINE TRANSMEMBRANE GLYCOPROTEIN RECEPTOR PD000167: N54-S150 PROTEIN-TYROSINE-PHOSPHATASE DM00089 [P29352 22-291: E22-L123 [Q05209 14-295: K21-L123 [S48748 14-295: K21-L123 [JH0609 14-296: K21-L125	BLIMPS_BLOCKS BLIMPS_PRINTS BLAST_PRODOM BLAST_PRODOM BLAST_DOMO
8	7500309CD1	848	S35 S77 S175 S210 S259 S293 S321 S374 S416 S520 S538 S760 S764 S784 S789 T245 T263 T300 T501 T704 T709 Y326 Y437	N651	TBC domain: A463-N673	HMMER_PFAM

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
					Protein kinase domain: Q43-F273	HMMER_PFAM
					Protein kinases signatures and profile: T92-D145	PROFILESCAN
					PROTEIN CHROMOSOME TRANSMEMBRANE CELL DIVISION I OF ONCOGENE COSMID SIMILAR PD001799: V531-I679	BLAST_PRODOR
					PROTEIN KINASE DOMAIN DM00004	BLAST_DOMO
					[P32562 84-327: K37-K214 Q00534 15-290: M1-R261 S59359 20-279: I44-D208 P50528 43-286: V23-I209	
9	2446376CD1	342	S34 S35 S65 S77 S91 S221 S233 S260 S269 S314 T171 T313	N116 N179 N231	signal_cleavage: M1-A30	SPSCAN
					Signal Peptide: L11-A30	HMMER
					Signal Peptide: M5-A30	HMMER
					Signal Peptide: L14-G32	HMMER
					Signal Peptide: M5-G32	HMMER
					Immunoglobulin domain: G173-V243, G53-R129	HMMER_PFAM
					Cytosolic domain: A312-E342	TMHMMER
					Transmembrane domain: L289-L311	
					Non-cytosolic domain: M1-L288	
					Protein kinase domain: Y13-V304	HMMER_PFAM
10	5562847CD1	544	S75 S161 S192 S238 S294 S331 S403 S447 T3 T57 T150 T273 T352 Y89			

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
					Tyrosine kinase catalytic domain signature PR00109: F127-L145, V199-T221, T273-A295	BLIMPS_PRINTS
					KINASE TRANSFERASE PROTEIN SERINE/THREONINE PROTEIN ATP BINDING II PHOSPHORYLATION CASEIN ALPHA CHAIN PD002608: V178-Y303	BLAST_PRODOM
					KINASE PROTEIN TRANSFERASE ATPBINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: Y183-Y303, L14-D98, Y119-L162	BLAST_PRODOM
					PROTEIN KINASE DOMAIN DM00004 [Q11179]15-300: L15-S294 [A56492]16-294: Q18-A295 [S40469]40-339: L19-E313 [S40471]45-344: R17-E313	BLAST_DOMO
					Cell attachment sequence: R470-D472	MOTIFS
					Protein kinases ATP-binding region signature: L19-K42	MOTIFS
11	7493671CD1	185	S4 S26 S37 S58 S94 S97 S142 S162	N22 N40 N53 N90 N93 N128 N160	UBA (ubiquitin associated) domain: V101-M141	HMMER_PFAM

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
12	7499991CD1	1269	S42 S105 S110 S111 S148 S266 S470 S471 S501 S513 S519 S712 S796 S820 S830 S899 S903 S924 S1045 S1084 S1130 T99 T221	N161 N367 N555 N602 N828 N961	Inositol polyphosphate phosphatase family, catalytic domain: K531-E874	HMMER_PFAM
			T288 T378 T557 T657 T783 T805 T941 T1055 T1151 T1188 Y525		SacI homology domain: A59-L347	HMMER_PFAM
					Inositol polyphosphate phosphatase family, catalytic domain PF00783: F726-I735, R800-L809	BLIMPS_PFAM
					PROTEIN INOSITOL HYDROLASE 5 PHOSPHATASE SYNAPTOJANIN POLYPHOSPHATE PHOSPHATASE TYPE I POLYPHOSPHATE 5 PHOSPHATASE PD002029: W540-E874	BLAST_PRODROM
					SYNAPTOJANIN ENDOCYTOSIS KIAA0348 II ISOFORM ALPHA DELTA SAC SYNAPTOJANIN 1 PD011649: A875-E1104	BLAST_PRODROM
					PROTEIN SYNAPTOJANIN TRANSMEMBRANE INTERGENIC REGION KIAA0274 ENDOCYTOSIS HYDROLASE T22J18.20 C34B7.2 PD004175: G83-H336; V243-E400, M85-D176	BLAST_PRODROM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN SYNAPTOJANIN' TRANSMEMBRANE ENDOCYTOSIS KIAA0348 II ISOFORM ALPHA DELTA SACSYNAPTOJANIN1 YOR3231W PD004400: M401-T582	BLAST_PRODOM
					SYNDROME; YOR109W; OCULOCEREBRORENAL; MEMBRANE; DM02714 P50942 599-979: K580-R811, L847-P1154 S61667 574-958: K577-R811, K836-N908, V1005-P1072 Q01968 323-658: A569-R811, K836-E906 YOR109W; MEMBRANE; DM02715 P50942 65-597: E118-L480	BLAST_DOMO
						BLAST_DOMO
13	7499977CD1	199	S74 T58 T71 T88		signal_cleavage: M1-N14	SPSCAN
					Octicosapeptide repeat: T57-Y86	HMMER_Pfam
					Protein kinase domain: P111-I170	HMMER_Pfam
					KINASE MAP DUAL SPECIFICITY MITOGEN ACTIVATED PROTEIN MAPKK MAPK/ERK TRANSFERASE SERINE/THREONINE PROTEIN PD016599: M1-L122	BLAST_PRODOM
					KINASE MAP DUAL SPECIFICITY MITOGEN ACTIVATED PROTEIN MAPKK MAPK/ERK TRANSFERASE SERINE/THREONINE PROTEIN PD014653: I170-P199	BLAST_PRODOM
					PROTEIN KINASE DOMAIN DM00004 A56708 168-399: L120-P161	BLAST_DOMO

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
14	8043337CD1	I214	S23 S41 S48 S69 S155 S207 S251 S359 S438 S563 S644 S663 S664 S711 S736 S754 S762 S824 S841 S909 S929 S964 S1008 S1048 S1082 S1092 S1125 S1131 T198 T351 T651 T738 T742 T787 T865 T874 T896 T1018 T1024 T1058 Y30	N36 N73 N154 N182 N543 N687 N785 N924 N1099	Protein kinase domain: Y4-I258	HMMER_PFAM
					Protein kinases signatures and profile: Q104-T156	PROFILESCAN
					Tyrosine kinase catalytic domain signature PR00109: M80-N93, H118-L136, S184-G206, Y227-V249	BLIMPS_PRINTS
					SERINE/THREONINE PROTEIN KINASE NIMA-RELATED PROTEIN 1 TRANSFERASE ATP-BINDING MITOSIS NUCLEAR PHOSPHORYLATION CELL CYCLE DIVISION TYROSINE PROTEIN PD144030: K255-L778	BLAST_PRODOR

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
					PROTEIN KINASE DOMAIN DM00004 P51954 6-248: R6-V249 DM00004 P51957 8-251: L7-V249 DM00004 Q08942 22-269: I10-V249, N1091-S1111 DM00004 P51955 10-261: L7-V249	BLAST_DOMO
					Protein kinases ATP-binding region signature: I10-K33	MOTIFS
					Serine/Threonine protein kinases active-site signature: I124-L136	MOTIFS
15	5947007CD1	411	S117 S164 S189 S274 S284 S334 S342 S351 T34 T104 T105 Y231 Y391	N143 N217 N311	Protein phosphatase 2C: L46-M312	HMMER_PFAM
					Protein phosphatase 2C proteins BL01032: D259-ID271, S308-V317, M25-N44, Q54-H64, F79-G88, N125-R142, S154-I163, N172-V211, R215-D228	BLIMPS_BLOCKS
					PROTEIN PHOSPHATASE 2C MAGNESIUM HYDROLASE MANGANESE MULTIGENE FAMILY PP2C ISOFORM PD001101: L46-G236, K136-E335, D310-E354	BLAST_PRODOM
					PROTEIN PHOSPHATASE 2C HYDROLASE ISOFORM 1A MAGNESIUM MANGANESE MULTIGENE FAMILY PD007496: E346-W411	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PROTEIN PHOSPHATASE 2C DM00377 P35815 1-304: M25-A329 DM00377 JC2524 1-304: M25-A329 DM00377 P36993 1-304: M25-A329 DM00377 S39781 1-304: M25-A329	BLAST_DOMO
					Protein phosphatase 2C signature: F79-A87	MOTIFS
16	7510576CD1	224	S42		Protein kinase domain: L6-L224	HMIMER_PFAM
					Serine/Threonine protein kinases, catalytic domain: K8-L224	HMIMER_SMART
					Tyrosine kinase, catalytic domain: L6-L221	HMIMER_SMART
					Protein kinases ATP-binding region proteins BL00107: Y113-K143, Q182-T197	BLIMPS_BLOCKS
					Protein kinases signatures and profile: L99-Q151	PROFILESCAN
					Tyrosine kinase catalytic domain signature PR00109: M76-Q89, Y113-L131, A183-G205	BLIMPS_PRINTS
					PROTEIN KINASE DOMAIN DM00004 I49621 195-428: L14-G199 Q05609 553-797: L14-T197 P42679 236-470: L14-G199 P08414 44-285: L14-L196	BLAST_DOMO
					Protein kinases ATP-binding region signature: L14-K35	MOTIFS
					Serine/Threonine protein kinases active-site signature: I119-L131	MOTIFS
17	90111008CD1	244	S34 S35 S123 S135 S162 S171 S216 T73 T215	N81 N133	signal_cleavage: M1-A30 Signal Peptide: M5-A30, M5-G32 Immunoglobulin domain: G75-V145	SFSCAN HMIMER HMIMER_PFAM

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
					Cytosolic domain: A214-E244 Transmembrane domain: L191-L213 Non-cytosolic domain: M1-L190	TMHMMER
18	7503305CD1	142	S4 S26 S37 S58 S94 S97	N22 N40 N53 N90 N93		

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
19/5532044CB1/ 4723	1-627, 1-2523, 1-4617, 55-570, 55-624, 59-717, 226-836, 226-845, 229-565, 299-811, 304-867, 555-1203, 710-1203, 751-1206, 829-1289, 829-1422, 853-1534, 894-1465, 1018-1673, 1126-1785, 1193-1598, 1225-1534, 1231-1785, 1374-1503, 1374-1505, 1391-1509, 1496-1673, 1512-2006, 1542-1673, 1676-2397, 1684-2315, 1733-2165, 1965-2523, 2014-2486, 2088-2531, 2142-2490, 2159-2423, 2505-2931, 3330-3653, 4420-4723
20/7497825CB1/ 513	1-513, 40-513
21/6464221CB1/ 3087	1-797, 5-837, 25-870, 322-886, 341-1022, 721-1159, 756-1390, 917-1182, 917-1500, 936-1219, 951-1372, 966-1372, 985-1551, 990-1636, 1043-1318, 1051-1574, 1073-1317, 1368-1914, 1369-1634, 1378-1620, 1380-1620, 1408-1868, 1415-1665, 1430-2072, 1441-1987, 1460-1980, 1621-2299, 1721-2271, 1791-2387, 1862-2251, 1897-2422, 2000-2457, 2118-2815, 2127-2583, 2277-2802, 2289-2640, 2310-2727, 2364-2989, 2378-2716, 2473-3060, 2483-3015, 2501-3073, 2534-3087, 2537-3071, 2558-3075, 2558-3087
22/90044205CB1/ 1997	1-217, 1-245, 1-263, 1-411, 1-532, 1-540, 1-797, 1-824, 22-273, 24-253, 27-540, 27-765, 28-284, 28-587, 28-637, 28-663, 36-497, 36-622, 36-826, 42-822, 79-317, 79-358, 79-365, 79-538, 79-579, 79-583, 79-584, 79-587, 79-597, 79-641, 79-642, 79-644, 79-653, 79-657, 79-659, 79-675, 79-694, 79-724, 79-733, 79-738, 79-739, 79-793, 79-800, 81-339, 117-472, 170-428, 203-448, 223-680, 357-1222, 361-1222, 384-1194, 387-1304, 390-1222, 429-1201, 429-1257, 430-464, 430-541, 447-1206, 532-1219, 556-1219, 560-1255, 603-1222, 650-1222, 684-1302, 738-1535, 853-1507, 893-1336, 893-1406, 908-1561, 1014-1232, 1029-1278, 1057-1877, 1117-1445, 1121-1257, 1130-1989, 1131-1997, 1146-1996, 1162-1772, 1258-1321, 1306-1340, 1306-1349, 1306-1351, 1306-1417, 1384-1417, 1386-1417, 1386-1418
23/603340CB1/ 1183	1-405, 2-577, 60-744, 63-588, 310-1072, 364-633, 398-1035, 415-643, 421-661, 586-1061, 616-1183, 628-1116, 663-1102, 664-1112, 676-1102, 698-1116, 701-1109, 703-1111, 719-1023, 725-1116, 742-1111, 750-1114, 791-1155, 841-1116, 1040-1114

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
24/71059650CB1/ 8752	1-434, 2-2294, 137-789, 144-789, 154-177, 346-1028, 554-1004, 554-1014, 837-1093, 837-1122, 837-1371, 871-1152, 971-1476, 1087-1476, 1277-1739, 1321-1737, 1345-1476, 1374-1476, 1418-1552, 1420-1552, 1431-1821, 1431-1856, 1431-1857, 1431-1860, 1431-1861, 1485-2081, 1555-1790, 1555-1979, 1555-2028, 1555-2041, 1555-2056, 1555-2111, 1555-2170, 1580-2161, 1584-2206, 1709-2204, 1744-2213, 1849-2204, 2003-2237, 2011-2499, 2151-2775, 2165-2775, 2178-2775, 2195-2775, 2203-2775, 2215-2775, 2217-2775, 2232-2775, 2233-2775, 2239-2775, 2241-2775, 2247-2775, 2271-2775, 2304-2775, 2325-2775, 2343-2775, 2396-3008, 2396-3030, 2396-3034, 2396-3174, 2396-3175, 2396-3183, 2400-3181, 2456-2655, 2563-2775, 2599-3486, 2608-3486, 2611-3453, 2682-3486, 2684-3486, 2713-3453, 2714-3298, 2750-3486, 2793-3068, 2793-3176, 2793-3177, 2793-3296, 2796-3203, 2811-3296, 2877-3486, 2878-7019, 2971-3296, 3096-3383, 3096-3565, 3110-3492, 3150-3734, 3274-3479, 3453-3803, 3453-3843, 3453-3920, 3453-4006, 3453-4031, 3453-4032, 3453-4038, 3453-4041, 3453-4043, 3453-4147, 3453-4158, 3455-3960, 3475-3734, 3499-4059, 3632-4211, 3701-4170, 3703-4357, 3716-4209, 3742-4045, 3743-4143, 3773-4386, 3805-4508, 3811-4454, 3818-4461, 3825-4524, 3835-4411, 3860-4457, 3863-4545, 3863-4561, 3880-4558, 3884-4454, 3891-4485, 3914-4545, 3931-4520, 3935-4425, 3940-4545, 3950-4444, 4003-4631, 4123-4791, 4158-4845, 4182-4829, 4190-4776, 4190-4871, 4255-4941, 4265-4941, 4291-4934, 4305-4963, 4310-5004, 4311-4842, 4469-5014, 4476-5046, 4511-5201, 4599-5362, 4617-5362, 4661-5252, 4687-5362, 4695-5356, 4698-5302, 4730-5362, 4753-5362, 4755-5362, 4759-5362, 4768-5362, 4791-5215, 4797-5362, 4860-5362, 4997-5303, 5018-5362, 5050-5596, 5201-5362, 5204-5596, 5466-5653, 5565-6189, 5671-6271, 5671-6313, 5730-6382, 5809-5956, 5810-6402, 5899-6203, 5899-6379, 5899-6667, 6012-6547, 6096-6559, 6108-6816, 6146-6294, 6160-6236, 6182-6779, 6236-6773, 6237-6863, 6251-6865, 6253-6690, 6253-6701, 6253-6727, 6253-6817, 6253-6864, 6253-6918, 6253-6946, 6253-6961, 6253-6973, 6253-7008, 6253-7014, 6253-7016, 6253-7017, 6253-7019, 6253-7020, 6253-7023, 6254-6803, 6257-6881, 6277-6630, 6298-7020, 6320-6858, 6320-6863, 6340-6980, 6350-6945, 6390-6863, 6390-6945, 6396-6696, 6396-6863, 6396-6871, 6397-6863, 6399-6863,

Table 4

Polynucleotide SEQ ID NO:/ Incye ID/ Sequence Length	Sequence Fragments
	6409-6952, 6415-6861, 6415-6863, 6421-6863, 6468-6856, 6475-7236, 6502-7117, 6535-6918, 6542-6862, 6582-7366, 6606-7262, 6744-7202, 6756-7105, 6756-7148, 6758-7102, 6858-7288, 6874-7713, 7010-7775, 7220-7886, 7380-8073, 7506-8222, 7531-8182, 7547-8222, 7621-8188, 7630-8571, 7642-8222, 7654-8190, 7661-8190, 7681-8190, 7682-8122, 7682-8190, 7688-8189, 7693-8190, 7722-8221, 7736-8328, 7748-8190, 7749-8169, 7749-8190, 7762-8457, 7776-8450, 7783-8328, 7797-8328, 7834-8100, 7835-8119, 7878-8116, 7934-8349, 7955-8252, 7969-8412, 7977-8328, 7987-8752, 7994-8328, 8009-8523, 8029-8100, 8052-8748, 8054-8100, 8060-8393, 8060-8524, 8067-8310, 8098-8746, 8101-8188, 8101-8210, 8101-8212, 8101-8238, 8101-8370, 8101-8491, 8129-8453
25/7499961CB1/ 868	1-181, 1-292, 1-322, 1-766, 1-775, 1-809, 1-868, 2-767, 6-797, 50-803, 170-809, 269-797, 457-811, 534-809, 604-811, 605-868
26/7500309CB1/ 3165	1-541, 1-542, 1-603, 1-668, 1-691, 1-767, 1-782, 1-905, 2-376, 2-757, 11-682, 11-688, 13-342, 13-425, 13-595, 13-600, 13-620, 13-641, 13-677, 13-685, 13-686, 13-781, 14-524, 28-456, 28-457, 31-538, 55-677, 79-910, 83-316, 102-680, 113-1015, 130-354, 131-985, 154-437, 388-1152, 927-1658, 1320-2003, 1385-1758, 1404-2193, 1429-1503, 1526-2156, 1553-2088, 1605-2394, 1713-2191, 1770-2294, 1777-2302, 1785-2346, 1787-2346, 1824-2680, 1849-2120, 1853-2133, 1902-2398, 1902-2437, 1913-2475, 1942-2124, 1942-2407, 1942-2418, 1942-2510, 1942-2542, 1957-2238, 1972-2262, 2076-2302, 2129-2868, 2163-2868, 2245-2960, 2262-2928, 2263-2776, 2303-2686, 2311-2702, 2390-3104, 2405-2702, 2410-3165, 2415-2851, 2417-3116, 2498-2849
27/2446376CB1/ 1630	1-259, 1-349, 111-349, 218-483, 221-769, 221-971, 221-1626, 229-349, 229-393, 350-715, 510-1076, 510-1094, 510-1123, 603-1123, 716-1123, 726-957, 726-1055, 726-1056, 726-1057, 727-1057, 734-1055, 766-1057, 818-1057, 822-1630, 835-1057, 837-1057, 838-1057, 839-1057, 970-1057, 997-1057
28/5562847CB1/ 1802	1-422, 337-895, 660-1218, 660-1220, 794-1280, 960-1212, 984-1242, 996-1275, 1011-1244, 1011-1523, 1011-1539, 1030-1531, 1048-1660, 1089-1585, 1099-1577, 1101-1788, 1157-1802, 1214-1649, 1244-1544, 1244-1632, 1316-1444, 1468-1767
29/7493671CB1/ 835	1-708, 1-724, 1-762, 1-797, 1-809, 1-810, 1-811, 1-814, 1-835, 367-534, 553-673, 554-835

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
30/7499991CB1/ 4814	1-303, 1-740, 15-4814, 468-1129, 527-920, 528-1127, 801-1058, 825-1129, 975-1568, 1648-2017, 1828-2101, 2095-2353, 2238-2865, 2346-2635, 2346-2810, 2346-2831, 2346-2836, 2346-2862, 2346-2867, 2346-2869, 2346-2877, 2346-2892, 2346-2926, 2346-2932, 2346-2937, 2349-2822, 2601-2858, 2607-3237, 2633-3289, 2700-3212, 2704-3418, 2708-2944, 2754-3374, 2805-3418, 2821-3418, 2832-3418, 2854-3418, 2857-3418, 2860-3418, 2868-3418, 2870-3418, 2978-3389, 2993-3418, 3007-3389, 3018-3219, 3018-3224, 3018-3418, 3018-3419, 3018-3421, 3018-3423, 3026-3244, 3051-3427, 3061-3421, 3066-3418, 3226-3421, 3252-3418, 3260-3418, 3277-3853, 3342-3421, 3420-3779, 3420-3785, 3438-3785, 3492-4049, 3536-4046, 3536-4084, 3650-3862, 3919-4814, 3969-4569, 4002-4265, 4096-4228, 4368-4590, 4440-4728
31/7499977CB1/ 1076	1-289, 3-418, 7-178, 72-780, 72-809, 72-814, 72-945, 74-999, 121-999, 146-999, 187-999, 227-999, 318-998, 461-1045, 465-1050, 472-1015, 482-1076, 484-1053, 486-1065, 505-1043, 524-955, 528-1032, 544-1051, 549-849, 563-827, 565-1008, 571-877, 572-790, 572-1046, 572-1053, 586-1016, 586-1059, 592-1059, 593-1056, 595-1040, 597-1074, 598-1053, 598-1055, 603-1055, 606-1053, 606-1056, 606-1059, 608-1054, 611-1057, 621-917, 625-1055, 629-1076, 632-1042, 637-1055, 641-1053, 641-1059, 644-1062, 653-1060, 656-964, 666-946, 674-1053, 679-1055, 684-1051, 706-1052, 727-1030, 730-1051, 737-1000, 744-1060, 755-1037, 761-1049, 774-1062, 797-1058, 803-1055, 817-1044, 821-1057, 857-970, 863-1074, 881-1040, 883-1073, 909-1030, 929-1055, 948-1049, 958-1076
32/8043337CB1/ 4683	1-849, 7-774, 13-807, 13-822, 63-839, 93-746, 96-736, 107-822, 121-987, 126-708, 205-699, 220-842, 224-896, 231-859, 256-935, 277-949, 285-1158, 291-1038, 295-1039, 316-1061, 334-983, 388-1154, 473-975, 637-1374, 662-1203, 671-937, 684-968, 684-1464, 703-1373, 723-1154, 906-1586, 920-1191, 920-1421, 920-1525, 920-1629, 921-1590, 930-1303, 960-1272, 963-1503, 964-1544, 979-1574, 984-1716, 1026-1567, 1066-1678, 1091-1726, 1095-1793, 1437-1683, 1437-1975, 1443-1664, 1454-1877, 1577-2454, 1631-2455, 1651-2455, 1804-2455, 1856-2453, 1901-2462, 2040-2741, 2271-2610, 2311-2669, 2311-2713, 2311-2739, 2311-2797, 2311-2799, 2311-2850, 2311-2872, 2311-2961, 2311-2967, 2311-2978, 2311-3002, 2311-3003, 2311-3041, 2311-3055, 2311-3093, 2311-3100, 2311-3106, 2311-3108, 2311-3132, 2311-3141, 2311-3147, 2311-3153, 2312-2592, 2321-2889, 2387-3192, 2405-3038, 2405-3126, 2406-2669, 2416-3025, 2836-3126, 2836-3475, 2836-3580, 2876-3463, 3105-3728, 3167-3868, 3188-3735, 3202-3929, 3276-4001, 3289-3950, 3356-4019, 3382-3625, 3382-4142, 3382-4149, 3382-4188, 3382-4253, 3391-4128, 3533-4111, 3533-4122, 3628-4169, 3641-4230, 3792-4058, 3792-4342, 3807-4298, 3809-4105, 3945-4553, 4015-4646, 4248-4683

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
33/5947007CB1/ 3129	1-575, 208-575, 217-1196, 339-655, 339-973, 720-1195, 867-1355, 1080-1521, 1080-1539, 1080-1641, 1080-3123, 1081-1655, 1111-1436, 1139-1856, 1166-1391, 1166-1500, 1174-1439, 1213-1368, 1273-1561, 1274-1389, 1400-1656, 1432-1722, 1447-1670, 1447-1846, 1478-1738, 1503-1763, 1503-2089, 1530-1796, 1546-1952, 1573-1876, 1629-1784, 1654-2137, 1665-1932, 1712-2247, 1766-2010, 1783-2035, 1785-2061, 1796-2065, 1865-2495, 1902-2284, 1916-2091, 1916-2130, 1916-2325, 1916-2349, 1942-2193, 1946-2108, 1958-2326, 1959-2190, 1960-2208, 1960-2214, 1960-2228, 1960-2247, 1979-2214, 1980-2627, 2034-2320, 2067-2360, 2119-2365, 2143-2413, 2143-2448, 2143-2555, 2145-2436, 2186-2411, 2187-2741, 2201-2824, 2211-2493, 2212-2886, 2262-2511, 2282-2551, 2300-2946, 2302-2523, 2303-2510, 2303-2571, 2303-2752, 2318-2566, 2339-2780, 2341-2582, 2349-2638, 2350-2619, 2350-2630, 2350-2657, 2360-2582, 2360-2619, 2360-2672, 2360-2951, 2375-2590, 2386-3097, 2413-2703, 2419-3035, 2449-2715, 2453-3112, 2463-2975, 2465-2676, 2475-2676, 2488-2802, 2504-2759, 2509-3109, 2512-3011, 2526-2676, 2531-2819, 2531-3062, 2534-3088, 2537-2740, 2537-2797, 2537-3066, 2548-2812, 2550-2797, 2552-2828, 2563-2784, 2567-3102, 2579-2908, 2580-2898, 2592-3112, 2610-2850, 2610-3095, 2612-2951, 2627-3106, 2631-3105, 2642-2806, 2650-2894, 2657-3115, 2658-3122, 2664-3050, 2664-3115, 2665-2928, 2668-2957, 2669-2939, 2669-3113, 2676-3129, 2679-3106, 2679-3128, 2687-3108, 2692-3113, 2696-2990, 2699-3105, 2701-3129, 2703-3105, 2706-3112, 2711-3105, 2712-3105, 2712-3129, 2713-3122, 2716-3105, 2720-3129, 2723-3129, 2734-3113, 2736-3129, 2737-3127, 2749-3117, 2754-2985, 2778-3115, 2788-3103, 2788-3105, 2796-3110, 2818-3074, 2823-3105, 2825-3125, 2850-3124, 2855-3103, 2866-3111, 2905-3114, 2917-3118, 2917-3129, 2918-3127, 2918-3129, 2923-3120, 2927-3111, 2928-3114, 2933-3109, 2967-3109, 2993-3101
34/7510576CB1/ 2580	1-665, 1-2580, 3-640, 3-661, 3-665, 6-216, 38-665, 54-666, 54-679, 54-697, 54-717, 54-937, 55-605, 64-465, 64-557, 64-608, 153-665, 177-665, 184-665, 192-665, 199-665, 211-715, 217-665, 312-665, 337-1259, 365-665, 720-1476, 756-1497, 759-1044, 761-1195, 761-1198, 761-1214, 761-1280, 761-1285, 761-1361, 765-1025, 765-1110, 782-1760, 785-1381, 790-1066, 802-1465, 815-1014, 837-1110, 844-1277, 849-1444, 850-1381, 851-1372, 874-1182, 874-1242, 874-1328, 874-1355, 874-1361, 877-1381, 887-1434, 907-1361, 921-1587, 949-1381, 957-1038, 962-1353, 978-1348, 978-1447, 979-1575, 980-1376, 1006-1922, 1049-1634, 1059-1922, 1078-1923, 1081-1774, 1108-1919, 1124-1634, 1154-1923, 1178-1923, 1183-1654, 1183-1725, 1184-1442, 1185-1923, 1193-1361, 1194-1725, 1198-1634, 1207-1922, 1214-1634, 1228-1468, 1252-1691, 1260-1922, 1272-1367, 1293-1571, 1329-1625, 1343-1923, 1352-1643, 1355-1891, 1358-1604, 1361-1526, 1362-1448, 1362-1609, 1362-1891, 1364-1725, 1368-1923, 1382-1693, 1382-1883, 1382-1891, 1383-1823, 1383-1891, 1389-1890, 1390-1891, 1392-1922, 1423-1922,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
	1437-2029, 1449-1891, 1450-1870, 1450-1891, 1463-2158, 1477-2151, 1484-2029, 1498-2029, 1535-1755, 1536-1820, 1579-1817, 1635-2050, 1656-1953, 1670-2113, 1678-2029, 1695-2029, 1710-2224, 1753-2506, 1761-2094, 1761-2225, 1768-2011, 1830-2154, 1881-2455, 1882-2431, 1892-2181, 1892-2326, 1892-2422, 1892-2453, 1909-2414, 1910-2421, 1918-2416, 1926-2203, 1940-2193, 1968-2414, 1978-2269, 1984-2408, 2001-2278, 2030-2225, 2030-2417, 2030-2449, 2030-2455, 2048-2580, 2052-2414, 2069-2461, 2070-2454, 2078-2293, 2117-2455, 2157-2445, 2168-2459, 2219-2459, 2239-2414, 2240-2445, 2246-2455, 2255-2455, 2262-2389, 2262-2409, 2262-2454, 2262-2459, 2281-2461, 2300-2554, 2336-2580, 2391-2461
35/90111008CB1/ 1301	1-803, 1-885, 3-227, 3-522, 3-748, 3-803, 3-805, 3-827, 3-869, 4-802, 197-361, 400-631, 446-1301, 496-1299, 524-1300, 524-1301, 545-1298, 569-1299, 571-1299, 642-1299, 655-1299, 673-1299, 795-1212, 1036-1236
36/7503305CB1/ 1675	1-669, 17-706, 18-235, 35-314, 35-624, 35-1675, 64-377, 319-721, 319-808, 319-968, 319-1110, 319-1159, 319-1190, 319-1209, 319-1222, 319-1272, 343-721, 407-696, 725-948, 725-1206, 736-1582, 753-1580, 769-1582, 771-1580, 774-1582, 836-1580, 849-1580, 851-1582, 852-1573, 852-1582, 858-1582, 864-1580, 865-1582, 873-1109, 874-1580, 874-1582, 875-1109, 896-1188, 905-1580, 907-1580, 909-1580, 919-1545, 965-1580, 978-1580, 982-1582, 989-1582, 1014-1580, 1016-1537, 1016-1580, 1021-1582, 1037-1582, 1039-1552, 1040-1580, 1041-1622, 1089-1559, 1112-1576, 1150-1563, 1217-1560

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
19	5532044CB1	BRABDIR03
21	6464221CB1	LSUBNOT03
22	90044205CB1	THP1TX04
23	603340CB1	MCLRUNF01
24	71059650CB1	OVARNOT12
25	7499961CB1	BRAVUNT02
26	7500309CB1	BRAITDR03
27	2446376CB1	SPLNDIC01
28	5562847CB1	OVARTUE01
29	7493671CB1	BRSTTUT17
30	7499991CB1	BRAUNOR01
31	7499977CB1	STOMFET01
32	8043337CB1	BRONDIT01
33	5947007CB1	PROSNOT16
34	7510576CB1	LUNGNOT02
35	90111008CB1	LUNGNOE02
36	7503305CB1	OVARTUT04

Table 6

Library	Vector	Library Description
BRABDIR03	pINCY	This random primed library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male who died from a cerebrovascular accident. Serologies were negative. Patient history included Huntington's disease, emphysema, and tobacco abuse (3-4 packs per day for 40 years).
BRAITDR03	PCDNA2.1	This random primed library was constructed using RNA isolated from allocortex, cingulate posterior tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior
		hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloidal goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.
BRAVUNT02	PSORT1	Library was constructed using pooled RNA isolated from separate populations of unstimulated astrocytes.
BRONDI01	pINCY	Library was constructed using RNA isolated from right lower lobe bronchial tissue removed from a pool of 3 asthmatic Caucasian male and female donors, 22- to 51-years-old during bronchial pinch biopsies. Patient history included atopy as determined by positive skin tests to common aero-allergens.

Table 6

Library	Vector	Library Description
BRSTTUT17	pINCY	Library was constructed using RNA isolated from left breast tumor tissue removed from a 65-year-old Caucasian female during a unilateral radical mastectomy. Pathology indicated invasive and in situ grade 3, nuclear grade 2 ductal carcinoma. Patient history included hyperlipidemia and uterine leiomyoma. Family history included stomach cancer, myocardial infarction, atherosclerotic coronary artery disease, prostate cancer, benign hypertension, breast cancer, and hyperlipidemia.
LSUBNOT03	pINCY	Library was constructed using RNA isolated from submandibular gland tissue obtained from a 68-year-old Caucasian male during a sialoadenectomy. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
LUNGNOE02	PSPORT	This 5' biased random primed library was constructed using RNA isolated from lung tissue removed from a 35-year-old Caucasian female during who died from a cerebrovascular accident. Serologies were negative. Patient history included mononucleosis, high blood pressure during pregnancies and alcohol use.
LUNGNOT02	PBLUESCRIPT	Library was constructed using RNA isolated from the lung tissue of a 47-year-old Caucasian male, who died of a subarachnoid hemorrhage.
MCLRUNF01	PCMV-ICIS	This full-length enriched library was constructed using RNA isolated from pooled, untreated peripheral blood mononuclear cell tissue removed from six donors. The donor pool consisted of a 35-year-old Caucasian male, a 54-year-old Caucasian male, a 57-year-old Caucasian male, a 28-year-old mixed race female, a 29-year-old Caucasian female, and a 44-year-old Caucasian female. The cells were pooled and untreated for 24 hours.
OVARNOT12	pINCY	Library was constructed using RNA isolated from ovarian tissue removed from a 36-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, soft tissue excision, and an incidental appendectomy. Pathology for the associated tumor tissue indicated one intramural and one subserosal leiomyoma of the myometrium. The cervix showed mild chronic cervicitis. A left inner thigh excision showed a lipoma. Family history included hyperlipidemia, acute myocardial infarction, atherosclerotic coronary artery disease, type II diabetes, and chronic liver disease.
OVARTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from left ovary tumor tissue removed from a 44-year-old female. Pathology indicated grade 4 (of 4) serous carcinoma replacing both the right and left ovaries forming solid mass cystic masses. Neoplastic deposits were identified in para-ovarian soft tissue.

Table 6

Library	Vector	Library Description
OVRTUT04	pINCY	Library was constructed using RNA isolated from ovarian tumor tissue removed from a 53-year-old Caucasian female during a total abdominal hysterectomy, removal of the fallopian tubes and ovaries, regional lymph node excision, peritoneal tissue destruction, and incidental appendectomy. Pathology indicated grade 1 transitional cell carcinoma of the right ovary. The left ovary had a hemorrhagic corpus luteum. The uterus had multiple leiomyomas (1 submucosal, 11 intramural), and the endometrium was inactive. The cul-de-sac contained abundant histiocytes and rare clusters of mesothelial cells. Patient history included breast fibrosclerosis and chronic stomach ulcer. Family history included acute stomach ulcer with perforation, breast cancer, bladder cancer, rectal/anal cancer, benign hypertension, coronary angioplasty, and hyperlipidemia.
PROSNOT16	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis, and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
SPLNDIC01	pINCY	This large size-fractionated library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from spleen tissue removed from an 8-year-old Black male (donor A) who died from anoxia and from diseased spleen tissue removed from a 14-year-old Asian male (donor B) during a total splenectomy. Pathology for donor B indicated changes consistent with idiopathic thrombocytopenic purpura. Serologies were negative for donor A. Donor B presented with bruising. Patient medications included DDAVP, Versed, labetalol (donor A), and Vincristine (donor B).
STOMFET01	pINCY	Library was constructed using RNA isolated from the stomach tissue of a Caucasian female fetus, who died at 20 weeks' gestation.
THP1TXT04	pINCY	Library was constructed using RNA isolated from stimulated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old male (Abbott Sample) with acute monocytic leukemia (Int. J. Cancer 26 (1980):171).

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

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
34	7510576	1606422H1	SNP00010291	26	2287	A	C	A	noncoding	0.63	n/a	n/a	n/a
34	7510576	1606422H1	SNP00010292	46	2307	C	T	C	noncoding	0.55	n/a	n/a	n/a
34	7510576	3046547H1	SNP00010291	128	2284	A	C	A	noncoding	0.63	n/a	n/a	n/a
34	7510576	3046547H1	SNP00010292	148	2304	C	T	C	noncoding	0.55	n/a	n/a	n/a
34	7510576	414383H1	SNP00010291	168	2286	A	C	A	noncoding	0.63	n/a	n/a	n/a
34	7510576	414383H1	SNP00010292	148	2306	C	T	C	noncoding	0.55	n/a	n/a	n/a
36	7503305	5900025H1	SNP00065406	275	1175	T	T	C	noncoding	0.38	n/a	n/a	n/a
36	7503305	6051950J1	SNP00065406	383	1177	T	T	C	noncoding	0.38	n/a	n/a	n/a

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting
5 of SEQ ID NO:1-18,
 - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-2, SEQ ID NO:5-6, SEQ ID NO:8-11, SEQ ID NO:13-14, and SEQ ID NO:16-18,
 - 10 c) a polypeptide comprising a naturally occurring amino acid sequence at least 92% identical to the amino acid sequence of SEQ ID NO:12,
 - d) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO:15,
 - e) a polypeptide comprising a naturally occurring amino acid sequence at least 98%
15 identical to the amino acid sequence of SEQ ID NO:4,
 - f) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to the amino acid sequence of SEQ ID NO:3,
 - g) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18, and
 - 20 h) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.
2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.
25
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
- 30 5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
35

7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
- 5 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
- 10 b) recovering the polypeptide so expressed.
10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.
- 15 11. An isolated antibody which specifically binds to a polypeptide of claim 1.
12. An isolated polynucleotide selected from the group consisting of:
- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:19-36,
- 20 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:19-21, SEQ ID NO:23-24, SEQ ID NO:26-30, SEQ ID NO:32-35,
- c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 92% identical to the polynucleotide sequence of SEQ ID NO:31,
- 25 d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 96% identical to the polynucleotide sequence of SEQ ID NO:22,
- e) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to the polynucleotide sequence of SEQ ID NO:36,
- f) a polynucleotide complementary to a polynucleotide of a),
- 30 g) a polynucleotide complementary to a polynucleotide of b),
- h) a polynucleotide complementary to a polynucleotide of c),
- i) a polynucleotide complementary to a polynucleotide of d),
- j) a polynucleotide complementary to a polynucleotide of e), and
- k) an RNA equivalent of a)-j).
- 35

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

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide
5 having a sequence of a polynucleotide of claim 12, the method comprising:
a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said
10 target polynucleotide or fragments thereof, and
b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

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

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
20 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.
25

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

19. A method for treating a disease or condition associated with decreased expression of
30 functional KPP, 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:
35 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.

5

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

10 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.

15 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 KPP, comprising administering to a patient in need of such treatment a composition of claim 24.

20

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
25 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:

- 30 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
35 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.

5 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,
- 10 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:

- 15 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
- 20 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
- 25 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 KPP in a biological sample, the method comprising:

- 30 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,
 - 5 d) a F(ab')₂ fragment, or
 - e) a humanized antibody.
32. A composition comprising an antibody of claim 11 and an acceptable excipient.
- 10 33. A method of diagnosing a condition or disease associated with the expression of KPP 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.
- 15 35. A method of diagnosing a condition or disease associated with the expression of KPP in a subject, comprising administering to said subject an effective amount of the composition of claim 34.
- 20 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:1-18, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
 - 25 b) isolating antibodies from the animal, and
 - c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.
- 30 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:1-18, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- 5 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
- 10 e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18.

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

15 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.

20 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:1-18 in a sample, the method comprising:

- 25 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:1-18 in the sample.

30

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-18 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:1-18.

5 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:

- 10 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.

15

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
20 claim 12.

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

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

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

30

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 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 nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

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

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

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

15

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.

20

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

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

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

25

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

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

30

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

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
- 5 71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
- 10 74. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:19.
75. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
15 NO:20.
76. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:21.
- 20 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.
- 25 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
30 NO:25.
81. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:26.

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

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

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

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

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

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

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

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

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

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

<110> INCYTE GENOMICS, INC.
BAUGHN, Mariah R.
YUE, Henry
WALIA, Narinder K.
HE, Ann
AU-YOUNG, Janice K.
LEE, Soo Yeun
GIETZEN, Kimberly J.
LAL, Preeti G.
ELLIOTT, Vicki S.
ISON, Craig H.
YANG, Jumming
LEE, Ernestine A.
LI, Joana X.
EMERLING, Brooke M.
RICHARDSON, Thomas W.
WARREN, Bridget A.
HAFALIA, April J.A.
MARQUIS, Joseph P.

<120> KINASES AND PHOSPHATASES

<130> PF-1110 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/309,627

<151> 2001-08-02

<150> 60/310,933

<151> 2001-08-07

<150> 60/311,323

<151> 2001-08-09

<150> 60/317,820

<151> 2001-09-07

<150> 60/322,264

<151> 2001-09-14

<150> 60/326,098

<151> 2001-09-28

<150> 60/343,007

<151> 2001-12-19

<150> 60/364,494

<151> 2002-03-15

<150> 60/375,539

<151> 2002-04-24

<160> 36

<170> PERL Program

<210> 1

<211> 1363

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5532044CD1

<400> 1

Met	Cys	Glu	Val	Met	Pro	Thr	Ile	Asn	Glu	Gly	Asp	Pro	Leu	Gly
1				5					10					15
Pro	Pro	His	Gly	Ala	Asp	Ala	Asp	Ala	Asn	Phe	Glu	Gln	Leu	Met
				20					25					30
Val	Asn	Met	Leu	Asp	Glu	Arg	Glu	Lys	Leu	Leu	Glu	Ser	Leu	Arg
				35					40					45
Glu	Ser	Gln	Glu	Thr	Leu	Ala	Ala	Thr	Gln	Ser	Arg	Leu	Gln	Asp
				50					55					60
Ala	Ile	His	Glu	Arg	Asp	Gln	Leu	Gln	Arg	His	Leu	Asn	Ser	Ala
				65					70					75
Leu	Pro	Gln	Glu	Phe	Ala	Thr	Leu	Thr	Arg	Glu	Leu	Ser	Met	Cys
				80					85					90
Arg	Glu	Gln	Leu	Leu	Glu	Arg	Glu	Glu	Glu	Ile	Ser	Glu	Leu	Lys
				95					100					105
Ala	Glu	Arg	Asn	Asn	Thr	Arg	Leu	Leu	Leu	Glu	His	Leu	Glu	Cys
				110					115					120
Leu	Val	Ser	Arg	His	Glu	Arg	Ser	Leu	Arg	Met	Thr	Val	Val	Lys
				125					130					135
Arg	Gln	Ala	Gln	Ser	Pro	Ser	Gly	Val	Ser	Ser	Glu	Val	Glu	Val
				140					145					150
Leu	Lys	Ala	Leu	Lys	Ser	Leu	Phe	Glu	His	His	Lys	Ala	Leu	Asp
				155					160					165
Glu	Lys	Val	Arg	Glu	Arg	Leu	Arg	Ala	Ala	Leu	Glu	Arg	Val	Thr
				170					175					180
Thr	Leu	Glu	Glu	Gln	Leu	Ala	Gly	Ala	His	Gln	Gln	Val	Ser	Ala
				185					190					195
Leu	Gln	Gln	Gly	Ala	Gly	Val	Arg	Asp	Gly	Ala	Ala	Glu	Glu	Glu
				200					205					210
Gly	Thr	Val	Glu	Leu	Gly	Pro	Lys	Arg	Leu	Trp	Lys	Glu	Asp	Thr
				215					220					225
Gly	Arg	Val	Glu	Glu	Leu	Gln	Glu	Leu	Leu	Glu	Lys	Gln	Asn	Phe
				230					235					240
Glu	Leu	Ser	Gln	Ala	Arg	Glu	Arg	Leu	Val	Thr	Leu	Thr	Thr	Thr
				245					250					255
Val	Thr	Glu	Leu	Glu	Glu	Asp	Leu	Gly	Thr	Ala	Arg	Arg	Asp	Leu
				260					265					270
Ile	Lys	Ser	Glu	Glu	Leu	Ser	Ser	Lys	His	Gln	Arg	Asp	Leu	Arg
				275					280					285
Glu	Ala	Leu	Ala	Gln	Lys	Glu	Asp	Met	Glu	Glu	Arg	Ile	Thr	Thr
				290					295					300
Leu	Glu	Lys	Arg	Tyr	Leu	Ala	Ala	Gln	Arg	Glu	Ala	Thr	Ser	Ile
				305					310					315
His	Asp	Leu	Asn	Asp	Lys	Leu	Glu	Asn	Glu	Leu	Ala	Asn	Lys	Glu
				320					325					330
Ser	Leu	His	Arg	Gln	Cys	Glu	Glu	Lys	Ala	Arg	His	Leu	Gln	Glu
				335					340					345
Leu	Leu	Glu	Val	Ala	Glu	Gln	Lys	Leu	Gln	Gln	Thr	Met	Arg	Lys
				350					355					360
Ala	Glu	Thr	Leu	Pro	Glu	Val	Glu	Ala	Glu	Leu	Ala	Gln	Arg	Ile
				365					370					375
Ala	Ala	Leu	Thr	Lys	Ala	Lys	Arg	His	Gly	Asn	Ile	Glu	Glu	His
				380					385					390
Leu	Arg	Gln	Leu	Glu	Gly	Gln	Leu	Glu	Glu	Lys	Asn	Gln	Glu	Leu
				395					400					405
Ala	Arg	Val	Arg	Gln	Arg	Glu	Lys	Met	Asn	Glu	Asp	His	Asn	Lys
				410					415					420
Arg	Leu	Ser	Asp	Thr	Val	Asp	Arg	Leu	Leu	Ser	Glu	Ser	Asn	Glu
				425					430					435

Arg	Leu	Gln	Leu	His	Leu	Lys	Glu	Arg	Met	Ala	Ala	Leu	Glu	Glu
				440					445					450
Lys	Asn	Thr	Leu	Ile	Gln	Glu	Leu	Glu	Ser	Ser	Gln	Arg	Gln	Ile
				455					460					465
Glu	Glu	Gln	His	His	His	Lys	Gly	Arg	Leu	Ser	Glu	Glu	Ile	Glu
				470					475					480
Lys	Leu	Arg	Gln	Glu	Val	Asp	Gln	Leu	Lys	Gly	Arg	Gly	Gly	Pro
				485					490					495
Phe	Val	Asp	Gly	Val	His	Ser	Arg	Ser	His	Met	Gly	Ser	Ala	Ala
				500					505					510
Asp	Val	Arg	Phe	Ser	Leu	Gly	Thr	Thr	Thr	His	Ala	Pro	Pro	Gly
				515					520					525
Val	His	Arg	Arg	Tyr	Ser	Ala	Leu	Arg	Glu	Glu	Ser	Ala	Lys	Asp
				530					535					540
Trp	Glu	Thr	Ser	Pro	Leu	Pro	Gly	Met	Leu	Ala	Pro	Ala	Ala	Gly
				545					550					555
Pro	Ala	Phe	Asp	Ser	Asp	Pro	Glu	Ile	Ser	Asp	Val	Asp	Glu	Asp
				560					565					570
Glu	Pro	Gly	Gly	Leu	Val	Gly	Ser	Ala	Asp	Val	Val	Ser	Pro	Ser
				575					580					585
Gly	His	Ser	Asp	Ala	Gln	Thr	Leu	Ala	Met	Met	Leu	Gln	Glu	Gln
				590					595					600
Leu	Asp	Ala	Ile	Asn	Glu	Glu	Ile	Arg	Met	Ile	Gln	Glu	Glu	Lys
				605					610					615
Glu	Ser	Thr	Glu	Leu	Arg	Ala	Glu	Glu	Ile	Glu	Thr	Arg	Val	Thr
				620					625					630
Ser	Gly	Ser	Met	Glu	Ala	Leu	Asn	Leu	Lys	Gln	Leu	Arg	Lys	Arg
				635					640					645
Gly	Ser	Ile	Pro	Thr	Ser	Leu	Thr	Ala	Leu	Ser	Leu	Ala	Ser	Ala
				650					655					660
Ser	Pro	Pro	Leu	Ser	Gly	Arg	Ser	Thr	Pro	Lys	Leu	Thr	Ser	Arg
				665					670					675
Ser	Ala	Ala	Gln	Asp	Leu	Asp	Arg	Met	Gly	Val	Met	Thr	Leu	Pro
				680					685					690
Ser	Asp	Leu	Arg	Lys	His	Arg	Arg	Lys	Leu	Leu	Ser	Pro	Val	Ser
				695					700					705
Arg	Glu	Glu	Asn	Arg	Glu	Asp	Lys	Ala	Thr	Ile	Lys	Cys	Glu	Thr
				710					715					720
Ser	Pro	Pro	Ser	Ser	Pro	Arg	Thr	Leu	Arg	Leu	Glu	Lys	Leu	Gly
				725					730					735
His	Pro	Ala	Leu	Ser	Gln	Glu	Glu	Gly	Lys	Ser	Ala	Leu	Glu	Asp
				740					745					750
Gln	Gly	Ser	Asn	Pro	Ser	Ser	Ser	Asn	Ser	Ser	Gln	Asp	Ser	Leu
				755					760					765
His	Lys	Gly	Ala	Lys	Arg	Lys	Gly	Ile	Lys	Ser	Ser	Ile	Gly	Arg
				770					775					780
Leu	Phe	Gly	Lys	Lys	Glu	Lys	Gly	Arg	Leu	Ile	Gln	Leu	Ser	Arg
				785					790					795
Asp	Gly	Ala	Thr	Gly	His	Gly	Leu	Cys	Pro	Leu	Pro	Asn	Gly	Val
				800					805					810
Leu	Leu	Thr	Asp	Ser	Glu	Phe	Ser	Met	Gln	Glu	Pro	Met	Val	Pro
				815					820					825
Ala	Lys	Leu	Gly	Thr	Gln	Ala	Glu	Lys	Asp	Arg	Arg	Leu	Lys	Lys
				830					835					840
Lys	His	Gln	Leu	Leu	Glu	Asp	Ala	Arg	Arg	Lys	Gly	Met	Pro	Phe
				845					850					855
Ala	Gln	Trp	Asp	Gly	Pro	Thr	Val	Val	Ser	Trp	Leu	Glu	Leu	Trp
				860					865					870
Val	Gly	Met	Pro	Ala	Trp	Tyr	Val	Ala	Ala	Cys	Arg	Ala	Asn	Val
				875					880					885
Lys	Ser	Gly	Ala	Ile	Met	Ser	Ala	Leu	Ser	Asp	Thr	Glu	Ile	Gln
				890					895					900
Arg	Glu	Ile	Gly	Ile	Ser	Asn	Ala	Leu	His	Arg	Leu	Lys	Leu	Arg

	905		910		915
Leu Ala Ile Gln	Glu Met Val Ser Leu	Thr Ser Pro Ser Ala	Pro		
	920		925		930
Pro Thr Ser Arg	Thr Ser Ser Gly Asn	Val Trp Val Thr His	Glu		
	935		940		945
Glu Met Glu Thr	Leu Glu Thr Ser Thr	Lys Thr Asp Ser Glu	Glu		
	950		955		960
Gly Ser Trp Ala	Gln Thr Leu Ala Tyr	Gly Asp Met Asn His	Glu		
	965		970		975
Trp Ile Gly Asn	Glu Trp Leu Pro Ser	Leu Gly Leu Pro Gln	Tyr		
	980		985		990
Arg Ser Tyr Phe	Met Glu Cys Leu Val	Asp Ala Arg Met Leu	Asp		
	995		1000		1005
His Leu Thr Lys	Lys Asp Leu Arg Val	His Leu Lys Met Val	Asp		
	1010		1015		1020
Ser Phe His Arg	Thr Ser Leu Gln Tyr	Gly Ile Met Cys Leu	Lys		
	1025		1030		1035
Arg Leu Asn Tyr	Asp Arg Lys Glu Leu	Glu Lys Arg Arg Glu	Glu		
	1040		1045		1050
Ser Gln His Glu	Ile Lys Asp Val Leu	Val Trp Thr Asn Asp	Gln		
	1055		1060		1065
Val Val His Trp	Val Gln Ser Ile Gly	Leu Arg Asp Tyr Ala	Gly		
	1070		1075		1080
Asn Leu His Glu	Ser Gly Val His Gly	Ala Leu Leu Ala Leu	Asp		
	1085		1090		1095
Glu Asn Phe Asp	His Asn Thr Leu Ala	Leu Ile Leu Gln Ile	Pro		
	1100		1105		1110
Thr Gln Asn Thr	Gln Ala Arg Gln Val	Met Glu Arg Glu Phe	Asn		
	1115		1120		1125
Asn Leu Leu Ala	Leu Gly Thr Asp Arg	Lys Leu Asp Asp Gly	Asp		
	1130		1135		1140
Asp Lys Val Phe	Arg Arg Ala Pro Ser	Trp Arg Lys Arg Phe	Arg		
	1145		1150		1155
Pro Arg Glu His	His Gly Arg Gly Gly	Met Leu Ser Ala Ser	Ala		
	1160		1165		1170
Glu Thr Leu Pro	Ala Gly Phe Arg Val	Ser Thr Leu Gly Thr	Leu		
	1175		1180		1185
Gln Pro Pro Pro	Ala Pro Pro Lys Lys	Ile Met Pro Glu Ala	Ala		
	1190		1195		1200
Leu Thr Leu Arg	Gly Glu Arg Val Cys	Ala Leu Thr Arg Arg	Ser		
	1205		1210		1215
Val Cys Ser Pro	Ala Ala Ala Asp Phe	Ser Gln Leu Gly Arg	Arg		
	1220		1225		1230
Gly Arg Arg Asp	Trp Arg Pro Pro Arg	Phe Gly Pro Thr Pro	Ala		
	1235		1240		1245
Asp Pro Pro Ser	Ser Arg Pro Gly Ser	Asp Gly Gly Val Pro	Val		
	1250		1255		1260
Ala Arg Ala Pro	Ser Leu Val Arg Asp	Phe Pro Val Ala Leu	Asp		
	1265		1270		1275
Leu Arg Ile Tyr	Ser Ser Thr Pro Ser	Ala Pro His Tyr Pro	Glu		
	1280		1285		1290
Ser His Arg Val	Ser Val Val Pro Glu	Asn Val Gln Asp Leu	Cys		
	1295		1300		1305
Ser Gly Trp Asp	Asp Gly Thr Glu Gly	Phe Ile Lys Val Arg	Gly		
	1310		1315		1320
Lys Glu Lys His	Leu Ala Cys Val Ser	Trp Ile Gly Ala Ser	Ala		
	1325		1330		1335
Val Pro Ser Tyr	Arg Trp Leu Glu Gln	Leu Leu Gly Asn Leu	Ile		
	1340		1345		1350
Thr Leu Ser Ser	Glu Gln Gly Leu Gly	Leu Gln Val Arg			
	1355		1360		

<210> 2

<211> 170
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7497825CD1

<400> 2
 Met Asn His Pro Ala Pro Val Lys Val Thr Tyr Lys Asn Met Arg
 1 5 10 15
 Phe Pro Ile Thr His Asn Pro Thr Asn Val Thr Leu Asn Lys Phe
 20 25 30
 Ile Glu Glu Leu Lys Lys Tyr Gly Ala Thr Thr Ile Val Arg Val
 35 40 45
 Cys Glu Ala Thr Tyr Asp Thr Thr Leu Val Glu Lys Glu Gly Ile
 50 55 60
 His Val Leu Asn Trp Pro Phe Gly Asp Gly Ala Pro Pro Ser Asn
 65 70 75
 Gln Ile Val Ala Asp Trp Leu His Phe Val Lys Ile Lys Phe Cys
 80 85 90
 Glu Glu Pro Gly Cys Tyr Ile Ala Val Asn Cys Ile Val Gly Leu
 95 100 105
 Gly Lys Ala Pro Val Leu Val Ala Leu Ala Ser Val Glu Gly Gly
 110 115 120
 Met Lys His Glu Asp Ala Val Gln Phe Ile Gly Gln Lys Arg Ser
 125 130 135
 Gly Ala Phe Lys Ser Lys Gln Leu Leu Tyr Leu Glu Lys Tyr His
 140 145 150
 Pro Lys Met Arg Leu Arg Phe Lys Asp Ser Asn Ser His Ile Asn
 155 160 165
 Asn Cys Cys Ile Gln
 170

<210> 3
 <211> 615
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 6464221CD1

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

Phe Cys Asp Ile	140		145		150
Cys Gln Lys Phe Leu	155	Leu Asn Gly Phe Arg	Cys		
Gln Thr Cys Gly Tyr	160	His Cys Ser Thr Lys	Val		
Pro Thr Met Cys Val	170	Asp Trp Ser Asn Ile	Arg Gln Leu Phe	Ser	
Gln His Arg Tyr Ser	185	Thr Pro His Ala Phe	Thr Phe Asn Thr	Ser	
Ser Pro Ser Ser Glu	200	Gly Ser Leu Ser Gln	Arg Gln Arg Ser	Thr	
Ser Thr Pro Asn Val	215	His Met Val Ser Thr	Thr Leu Pro Val	Asp	
Ser Arg Met Ile Glu	230	Asp Ala Ile Arg Ser	His Ser Glu Ser	Ala	
Ser Pro Ser Ala Leu	245	Ser Ser Ser Pro Asn	Asn Leu Ser Pro	Thr	
Gly Trp Ser Gln Pro	260	Lys Thr Pro Val Pro	Ala Gln Arg Glu	Arg	
Ala Pro Val Ser Gly	275	Thr Gln Glu Lys Asn	Lys Ile Arg Pro	Arg	
Gly Gln Arg Asp Ser	290	Ser Ser Tyr Tyr Trp	Glu Ile Glu Ala Ser	Glu	
Val Met Leu Ser Thr	305	Arg Ile Gly Ser Gly	Ser Phe Gly Thr	Val	
Tyr Lys Gly Lys Trp	320	His Gly Asp Val Ala	Val Lys Ile Leu	Lys	
Val Val Asp Pro Thr	335	Pro Glu Gln Phe Gln	Ala Phe Arg Asn	Glu	
Val Ala Val Leu Arg	350	Lys Thr Arg His Val	Asn Ile Leu Leu	Phe	
Met Gly Tyr Met Thr	365	Lys Asp Asn Leu Ala	Ile Val Thr Gln	Trp	
Cys Glu Gly Ser Ser	380	Leu Tyr Lys His Leu	His Val Gln Glu	Thr	
Lys Phe Gln Met Phe	395	Gln Leu Ile Asp Ile	Ala Arg Gln Thr	Ala	
Gln Gly Met Asp Tyr	410	Leu His Ala Lys Asn	Ile Ile His Arg	Asp	
Met Lys Ser Asn Asn	425	Ile Phe Leu His Glu	Gly Leu Thr Val	Lys	
Ile Gly Asp Phe Gly	440	Leu Ala Thr Val Lys	Ser Arg Trp Ser	Gly	
Ser Gln Gln Val Glu	455	Gln Pro Thr Gly Ser	Val Leu Trp Met	Ala	
Pro Glu Val Ile Arg	470	Met Gln Asp Asn Asn	Pro Phe Ser Phe	Gln	
Ser Asp Val Tyr Ser	485	Tyr Gly Ile Val Leu	Tyr Glu Leu Met	Thr	
Gly Glu Leu Pro Tyr	500	Ser His Ile Asn Asn	Arg Asp Gln Ile	Ile	
Phe Met Val Gly Arg	515	Gly Tyr Ala Ser Pro	Asp Leu Ser Lys	Leu	
Tyr Lys Asn Cys Pro	530	Lys Ala Met Lys Arg	Leu Val Ala Asp	Cys	
Val Lys Lys Val Lys	545	Glu Glu Arg Pro Leu	Phe Pro Gln Ile	Leu	
Ser Ser Ile Glu Leu	560	Leu Gln His Ser Leu	Pro Lys Ile Asn	Arg	
Ser Ala Ser Glu Pro	575	Ser Leu His Arg Ala	Ala His Thr Glu	Asp	
Ile Asn Ala Cys Thr	590	Leu Thr Thr Ser Pro	Arg Leu Pro Val	Phe	
	605			615	

<210> 4
 <211> 642
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 90044205CD1

<400> 4

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

```

Pro Ala Asn Met Lys Lys Ala Arg Val Ile Gln Ile Ile Pro Tyr
395 400
Asp Phe Asn Arg Val Ile Leu Ser Met Lys Arg Gly Gln Glu Tyr
410 415 420
Thr Asp Tyr Ile Asn Ala Ser Phe Ile Asp Gly Tyr Arg Gln Lys
425 430 435
Asp Tyr Phe Ile Ala Thr Gln Gly Pro Leu Ala His Thr Val Glu
440 445 450
Asp Phe Trp Arg Met Ile Trp Glu Trp Lys Ser His Thr Ile Val
455 460 465
Met Leu Thr Glu Val Gln Glu Arg Glu Gln Asp Lys Cys Tyr Gln
470 475 480
Tyr Trp Pro Thr Glu Gly Ser Val Thr His Gly Glu Ile Thr Ile
485 490 495
Glu Ile Lys Asn Asp Thr Leu Ser Glu Ala Ile Ser Ile Arg Asp
500 505 510
Phe Leu Val Thr Leu Asn Gln Pro Gln Ala Arg Gln Glu Glu Gln
515 520 525
Val Arg Val Val Arg Gln Phe His Phe His Gly Trp Pro Glu Ile
530 535 540
Gly Ile Pro Ala Glu Gly Lys Gly Met Ile Asp Leu Ile Ala Ala
545 550 555
Val Gln Lys Gln Gln Gln Gln Thr Gly Asn His Pro Ile Thr Val
560 565 570
His Cys Ser Ala Gly Ala Gly Arg Thr Gly Thr Phe Ile Ala Leu
575 580 585
Ser Asn Ile Leu Glu Arg Val Lys Ala Glu Gly Leu Leu Asp Val
590 595 600
Phe Gln Ala Val Lys Ser Leu Arg Leu Gln Arg Pro His Met Val
605 610 615
Gln Thr Leu Glu Gln Tyr Glu Phe Cys Tyr Lys Val Val Gln Asp
620 625 630
Phe Ile Asp Ile Phe Ser Asp Tyr Ala Asn Phe Lys
635 640

```

<210> 5

<211> 241

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 603340CD1

<400> 5

```

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

```

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

Asp

<210> 6

<211> 2527

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 71059650CD1

<400> 6

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

Cys	Cys	Leu	Leu	His	Arg	Leu	Thr	Leu	Gly	Asn	Phe	Phe	Asn	Ile
				275					280					285
Leu	Val	Leu	Asn	Glu	Val	His	Glu	Phe	Val	Val	Lys	Ala	Val	Gln
				290					295					300
Gln	Tyr	Pro	Glu	Asn	Ala	Ala	Leu	Gln	Ile	Ser	Ala	Leu	Ser	Cys
				305					310					315
Leu	Ala	Leu	Leu	Thr	Glu	Thr	Ile	Phe	Leu	Asn	Gln	Asp	Leu	Glu
				320					325					330
Glu	Lys	Asn	Glu	Asn	Gln	Glu	Asn	Asp	Asp	Glu	Gly	Glu	Glu	Asp
				335					340					345
Lys	Leu	Phe	Trp	Leu	Glu	Ala	Cys	Tyr	Lys	Ala	Leu	Thr	Trp	His
				350					355					360
Arg	Lys	Asn	Lys	His	Val	Gln	Glu	Ala	Ala	Cys	Trp	Ala	Leu	Asn
				365					370					375
Asn	Leu	Leu	Met	Tyr	Gln	Asn	Ser	Leu	His	Glu	Lys	Ile	Gly	Asp
				380					385					390
Glu	Asp	Gly	His	Phe	Pro	Ala	His	Arg	Glu	Val	Met	Leu	Ser	Met
				395					400					405
Leu	Met	His	Ser	Ser	Ser	Lys	Glu	Val	Phe	Gln	Ala	Ser	Ala	Asn
				410					415					420
Ala	Leu	Ser	Thr	Leu	Leu	Glu	Gln	Asn	Val	Asn	Phe	Arg	Lys	Ile
				425					430					435
Leu	Leu	Ser	Lys	Gly	Ile	His	Leu	Asn	Val	Leu	Glu	Leu	Met	Gln
				440					445					450
Lys	His	Ile	His	Ser	Pro	Glu	Val	Ala	Glu	Ser	Gly	Cys	Lys	Met
				455					460					465
Leu	Asn	His	Leu	Phe	Glu	Gly	Ser	Asn	Thr	Ser	Leu	Asp	Ile	Met
				470					475					480
Ala	Ala	Val	Val	Pro	Lys	Ile	Leu	Thr	Val	Met	Lys	Arg	His	Glu
				485					490					495
Thr	Ser	Leu	Pro	Val	Gln	Leu	Glu	Ala	Leu	Arg	Ala	Ile	Leu	His
				500					505					510
Phe	Ile	Val	Pro	Gly	Met	Pro	Glu	Glu	Ser	Arg	Glu	Asp	Thr	Glu
				515					520					525
Phe	His	His	Lys	Leu	Asn	Met	Val	Lys	Lys	Gln	Cys	Phe	Lys	Asn
				530					535					540
Asp	Ile	His	Lys	Leu	Val	Leu	Ala	Ala	Leu	Asn	Arg	Phe	Ile	Gly
				545					550					555
Asn	Pro	Gly	Ile	Gln	Lys	Cys	Gly	Leu	Lys	Val	Ile	Ser	Ser	Ile
				560					565					570
Val	His	Phe	Pro	Asp	Ala	Leu	Glu	Met	Leu	Ser	Leu	Glu	Gly	Ala
				575					580					585
Met	Asp	Ser	Val	Leu	His	Thr	Leu	Gln	Met	Tyr	Pro	Asp	Asp	Gln
				590					595					600
Glu	Ile	Gln	Cys	Leu	Gly	Leu	Ser	Leu	Ile	Gly	Tyr	Leu	Ile	Thr
				605					610					615
Lys	Lys	Asn	Val	Phe	Ile	Gly	Thr	Gly	His	Leu	Leu	Ala	Lys	Ile
				620					625					630
Leu	Val	Ser	Ser	Leu	Tyr	Arg	Phe	Lys	Asp	Val	Ala	Glu	Ile	Gln
				635					640					645
Thr	Lys	Gly	Phe	Gln	Thr	Ile	Leu	Ala	Ile	Leu	Lys	Leu	Ser	Ala
				650					655					660
Ser	Phe	Ser	Lys	Leu	Leu	Val	His	His	Ser	Phe	Asp	Leu	Val	Ile
				665					670					675
Phe	His	Gln	Met	Ser	Ser	Asn	Ile	Met	Glu	Gln	Lys	Asp	Gln	Gln
				680					685					690
Phe	Leu	Asn	Leu	Cys	Cys	Lys	Cys	Phe	Ala	Lys	Val	Ala	Met	Asp
				695					700					705
Asp	Tyr	Leu	Lys	Asn	Val	Met	Leu	Glu	Arg	Ala	Cys	Asp	Gln	Asn
				710					715					720
Asn	Ser	Ile	Met	Val	Glu	Cys	Leu	Leu	Leu	Leu	Gly	Ala	Asp	Ala
				725					730					735
Asn	Gln	Ala	Lys	Glu	Gly	Ser	Ser	Leu	Ile	Cys	Gln	Val	Cys	Glu

Lys Glu Ser Ser	740	Lys Leu Val Glu	745	Leu Leu Asn Ser	750
Ser Arg Glu Gln	755	Val Arg Lys Ala	760	Leu Thr Ile Ser Ile	765
Lys Gly Asp Ser	770	Gln Ile Ile Ser Leu	775	Leu Leu Arg Arg Leu	780
Leu Asp Val Ala	785	Asn Asn Ser Ile Cys	790	Leu Gly Gly Phe Cys	795
Gly Lys Val Glu	800	Pro Ser Trp Leu Gly	805	Pro Leu Phe Pro Asp	810
Thr Ser Asn Leu	815	Arg Lys Gln Thr Asn	820	Ile Ala Ser Thr Leu	825
Arg Met Val Ile	830	Arg Tyr Gln Met Lys	835	Ser Ala Val Glu Glu	840
Thr Ala Ser Gly	845	Ser Asp Gly Asn Phe	850	Ser Glu Asp Val Leu	855
Lys Phe Asp Glu	860	Trp Thr Phe Ile Pro	865	Asp Ser Ser Met Asp	870
Val Phe Ala Gln	875	Ser Asp Asp Leu Asp	880	Ser Ser Ser Met Asp	885
Ser Phe Leu Val	890	Lys Lys Lys Ser Asn	895	Ser Glu Gly Ser Glu	900
Phe Tyr Arg Asp	905	Ala Val Leu Gln Arg	910	Ser Ile Ser Val Gly	915
Arg His Ser Asn	920	Ser Leu Gly Pro Ile	925	Cys Ser Pro Asn Leu	930
Leu Lys Arg Lys	935	Arg Lys Ile Leu Ser	940	Phe Asp His Glu Asp	945
Ser Ser Lys Leu	950	Arg Lys Ile Leu Ser	955	Ser Asp Asp Ser Leu	960
Ser Leu Ala Ser	965	Gln Ser His Met Arg	970	His Ser Asp Ser Ile	975
Ala Asn Glu Leu	980	Arg Glu Tyr Ile Thr	985	Ser Leu Asp Leu Ser	990
Ile Ser Val His	995	Leu Glu His Leu Glu	1000	Ser Gln Lys Cys Cys	1005
Asn Ala Leu Thr	1010	Ser Phe Pro Gln Gln	1015	Leu Glu Leu His Gln	1020
Ser Leu Thr His	1025	Leu Asp Leu His Ser	1030	Leu Cys Glu Thr Leu	1035
Pro Ser Tyr Leu	1040	Leu Lys Met Ser Cys	1045	Leu Phe Thr Ser Phe	1050
Ser Arg Asn Asp	1055	Ile Gly Pro Ser Val	1060	Ala Asn Leu Asp Val	1065
Lys Cys Pro Thr	1070	Leu Lys Gln Phe Asn	1075	Leu Asp Pro Thr Val	1080
Ser Phe Val Pro	1085	Glu Leu Lys Gln Phe	1090	Ser Tyr Asn Gln Leu	1095
Gln Leu Ile Leu	1100	Glu Asn Leu Thr Asp	1105	Val Val Glu Lys Leu	1110
Leu Arg Leu Lys	1115	Gly Asn Lys Ile Ser	1120	Gly Ile Cys Ser Pro	1125
Ile Ser Ser Leu	1130	Ser Glu Asn Phe Leu	1135	Leu Ser Lys Asn His	1140
Glu Ser Phe Ser	1145	Ala Arg Met Asn Phe	1150	Ala Cys Pro Lys Val	1155
Leu Pro Pro Ser	1160	Met Thr Ile Leu Lys	1165	Ala Ala Met Pro Phe	1170
Ser Cys Ile Pro	1175	Glu Met Thr Ile Leu	1180	Ser Gln Asn Lys Phe	1185
Leu Asp Met Ser	1190	Leu Asn Leu Pro His	1195	Leu Arg Ser	1200
	1205	Asn Asp Ile Gln Tyr	1210	Leu Pro Gly Pro Ala	1215

His	Trp	Lys	Ser	Leu	Asn	Leu	Arg	Glu	Leu	Leu	Phe	Ser	His	Asn
				1220					1225					1230
Gln	Ile	Ser	Ile	Leu	Asp	Leu	Ser	Glu	Lys	Ala	Tyr	Leu	Trp	Ser
				1235					1240					1245
Arg	Val	Glu	Lys	Leu	His	Leu	Ser	His	Asn	Lys	Leu	Lys	Glu	Ile
				1250					1255					1260
Pro	Pro	Glu	Ile	Gly	Cys	Leu	Glu	Asn	Leu	Thr	Ser	Leu	Asp	Val
				1265					1270					1275
Ser	Tyr	Asn	Leu	Glu	Leu	Arg	Ser	Phe	Pro	Asn	Glu	Met	Gly	Lys
				1280					1285					1290
Leu	Ser	Lys	Ile	Trp	Asp	Leu	Pro	Leu	Asp	Glu	Leu	His	Leu	Asn
				1295					1300					1305
Phe	Asp	Phe	Lys	His	Ile	Gly	Cys	Lys	Ala	Lys	Asp	Ile	Ile	Arg
				1310					1315					1320
Phe	Leu	Gln	Gln	Arg	Leu	Lys	Lys	Ala	Val	Pro	Tyr	Asn	Arg	Met
				1325					1330					1335
Lys	Leu	Met	Ile	Val	Gly	Asn	Thr	Gly	Ser	Gly	Lys	Thr	Thr	Leu
				1340					1345					1350
Leu	Gln	Gln	Leu	Met	Lys	Thr	Lys	Lys	Ser	Asp	Leu	Gly	Met	Gln
				1355					1360					1365
Ser	Ala	Thr	Val	Gly	Ile	Asp	Val	Lys	Asp	Trp	Pro	Ile	Gln	Ile
				1370					1375					1380
Arg	Asp	Lys	Arg	Lys	Arg	Asp	Leu	Val	Leu	Asn	Val	Trp	Asp	Phe
				1385					1390					1395
Ala	Gly	Arg	Glu	Glu	Phe	Tyr	Ser	Thr	His	Pro	His	Phe	Met	Thr
				1400					1405					1410
Gln	Arg	Ala	Leu	Tyr	Leu	Ala	Val	Tyr	Asp	Leu	Ser	Lys	Gly	Gln
				1415					1420					1425
Ala	Glu	Val	Asp	Ala	Met	Lys	Pro	Trp	Leu	Phe	Asn	Ile	Lys	Ala
				1430					1435					1440
Arg	Ala	Ser	Ser	Ser	Pro	Val	Ile	Leu	Val	Gly	Thr	His	Leu	Asp
				1445					1450					1455
Val	Ser	Asp	Glu	Lys	Gln	Arg	Lys	Ala	Cys	Met	Ser	Lys	Ile	Thr
				1460					1465					1470
Lys	Glu	Leu	Leu	Asn	Lys	Arg	Gly	Phe	Pro	Ala	Ile	Arg	Asp	Tyr
				1475					1480					1485
His	Phe	Val	Asn	Ala	Thr	Glu	Glu	Ser	Asp	Ala	Leu	Ala	Lys	Leu
				1490					1495					1500
Arg	Lys	Thr	Ile	Ile	Asn	Glu	Ser	Leu	Asn	Phe	Lys	Ile	Arg	Asp
				1505					1510					1515
Gln	Leu	Val	Val	Gly	Gln	Leu	Ile	Pro	Asp	Cys	Tyr	Val	Glu	Leu
				1520					1525					1530
Glu	Lys	Ile	Ile	Leu	Ser	Glu	Arg	Lys	Asn	Val	Pro	Ile	Glu	Phe
				1535					1540					1545
Pro	Val	Ile	Asp	Arg	Lys	Arg	Leu	Leu	Gln	Leu	Val	Arg	Glu	Asn
				1550					1555					1560
Gln	Leu	Gln	Leu	Asp	Glu	Asn	Glu	Leu	Pro	His	Ala	Val	His	Phe
				1565					1570					1575
Leu	Asn	Glu	Ser	Gly	Val	Leu	Leu	His	Phe	Gln	Asp	Pro	Ala	Leu
				1580					1585					1590
Gln	Leu	Ser	Asp	Leu	Tyr	Phe	Val	Glu	Pro	Lys	Trp	Leu	Cys	Lys
				1595					1600					1605
Ile	Met	Ala	Gln	Ile	Leu	Thr	Val	Lys	Val	Glu	Gly	Cys	Pro	Lys
				1610					1615					1620
His	Pro	Lys	Gly	Ile	Ile	Ser	Arg	Arg	Asp	Val	Glu	Lys	Phe	Leu
				1625					1630					1635
Ser	Lys	Lys	Arg	Lys	Phe	Pro	Lys	Asn	Tyr	Met	Thr	Gln	Tyr	Phe
				1640					1645					1650
Lys	Pro	Leu	Glu	Lys	Phe	Gln	Ile	Ala	Leu	Pro	Ile	Gly	Glu	Glu
				1655					1660					1665
Tyr	Leu	Leu	Val	Pro	Ser	Ser	Leu	Ser	Asp	His	Arg	Pro	Val	Ile
				1670					1675					1680
Glu	Leu	Pro	His	Cys	Glu	Asn	Ser	Glu	Ile	Ile	Ile	Arg	Leu	Tyr

	1685		1690		1695
Glu Met Pro Tyr Phe		Pro Met Gly Phe Trp		Ser Arg Leu Ile Asn	
	1700		1705		1710
Arg Leu Leu Glu Ile		Ser Pro Tyr Met Leu		Ser Gly Arg Glu Arg	
	1715		1720		1725
Ala Leu Arg Pro Asn		Arg Met Tyr Trp Arg		Gln Gly Ile Tyr Leu	
	1730		1735		1740
Asn Trp Ser Pro Glu		Ala Tyr Cys Leu Val		Gly Ser Glu Val Leu	
	1745		1750		1755
Asp Asn His Pro Glu		Ser Phe Leu Lys Ile		Thr Val Pro Ser Cys	
	1760		1765		1770
Arg Lys Gly Cys Ile		Leu Leu Gly Gln Val		Val Asp His Ile Asp	
	1775		1780		1785
Ser Leu Met Glu Glu		Trp Phe Pro Gly Leu		Leu Glu Ile Asp Ile	
	1790		1795		1800
Cys Gly Glu Gly Glu		Thr Leu Leu Lys Lys		Trp Ala Leu Tyr Ser	
	1805		1810		1815
Phe Asn Asp Gly Glu		Glu His Gln Lys Ile		Leu Leu Asp Asp Leu	
	1820		1825		1830
Met Lys Lys Ala Glu		Glu Gly Asp Leu Leu		Val Asn Pro Asp Gln	
	1835		1840		1845
Pro Arg Leu Thr Ile		Pro Ile Ser Gln Ile		Ala Pro Asp Leu Ile	
	1850		1855		1860
Leu Ala Asp Leu Pro		Arg Asn Ile Met Leu		Asn Asn Asp Glu Leu	
	1865		1870		1875
Glu Phe Glu Gln Ala		Pro Glu Phe Leu Leu		Gly Asp Gly Ser Phe	
	1880		1885		1890
Gly Ser Val Tyr Arg		Ala Ala Tyr Glu Gly		Glu Glu Val Ala Val	
	1895		1900		1905
Lys Ile Phe Asn Lys		His Thr Ser Leu Arg		Leu Leu Arg Gln Glu	
	1910		1915		1920
Leu Val Val Leu Cys		His Leu His His Pro		Ser Leu Ile Ser Leu	
	1925		1930		1935
Leu Ala Ala Gly Ile		Arg Pro Arg Met Leu		Val Met Glu Leu Ala	
	1940		1945		1950
Ser Lys Gly Ser Leu		Asp Arg Leu Leu Gln		Gln Asp Lys Ala Ser	
	1955		1960		1965
Leu Thr Arg Thr Leu		Gln His Arg Ile Ala		Leu His Val Ala Asp	
	1970		1975		1980
Gly Leu Arg Tyr Leu		His Ser Ala Met Ile		Ile Tyr Arg Asp Leu	
	1985		1990		1995
Lys Pro His Asn Val		Leu Leu Phe Thr Leu		Tyr Pro Asn Ala Ala	
	2000		2005		2010
Ile Ile Ala Lys Ile		Ala Asp Tyr Gly Ile		Ala Gln Tyr Cys Cys	
	2015		2020		2025
Arg Met Gly Ile Lys		Thr Ser Glu Gly Thr		Pro Gly Phe Arg Ala	
	2030		2035		2040
Pro Glu Val Ala Arg		Gly Asn Val Ile Tyr		Asn Gln Gln Ala Asp	
	2045		2050		2055
Val Tyr Ser Phe Gly		Leu Leu Leu Tyr Asp		Ile Leu Thr Thr Gly	
	2060		2065		2070
Gly Arg Ile Val Glu		Gly Leu Lys Phe Pro		Asn Glu Phe Asp Glu	
	2075		2080		2085
Leu Glu Ile Gln Gly		Lys Leu Pro Asp Pro		Val Lys Glu Tyr Gly	
	2090		2095		2100
Cys Ala Pro Trp Pro		Met Val Glu Lys Leu		Ile Lys Gln Cys Leu	
	2105		2110		2115
Lys Glu Asn Pro Gln		Glu Arg Pro Thr Ser		Ala Gln Val Phe Asp	
	2120		2125		2130
Ile Leu Asn Ser Ala		Glu Leu Val Cys Leu		Thr Arg Arg Ile Leu	
	2135		2140		2145
Leu Pro Lys Asn Val		Ile Val Glu Cys Met		Val Ala Thr His His	
	2150		2155		2160

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

<210> 7

<211> 180

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7499961CD1

<400> 7

Met Asp Gln Arg Glu Ile Leu Gln Lys Phe Leu Asp Glu Ala Gln
 1 5 10 15

```

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

```

<210> 8

<211> 848

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500309CD1

<400> 8

```

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

```

	215		220		225
Leu Ile Val Leu	Ala Glu Glu His Gly	Cys Leu Asp Ile Ile	Lys		
	230		235		240
Glu Leu Pro Glu	Thr Val Ile Asp Leu	Leu Asn Lys Cys Leu	Thr		
	245		250		255
Phe His Pro Ser	Lys Arg Pro Thr Pro	Asp Glu Leu Met Lys	Asp		
	260		265		270
Lys Val Phe Ser	Glu Val Ser Pro Leu	Tyr Thr Pro Phe Thr	Lys		
	275		280		285
Pro Ala Ser Leu	Phe Ser Ser Ser Leu	Arg Cys Ala Asp Leu	Thr		
	290		295		300
Leu Pro Glu Asp	Ile Ser Gln Leu Cys	Lys Asp Ile Asn Asn	Asp		
	305		310		315
Tyr Leu Ala Glu	Arg Ser Ile Glu Glu	Val Tyr Tyr Leu Trp	Cys		
	320		325		330
Leu Ala Gly Gly	Asp Leu Glu Lys Glu	Leu Val Asn Lys Glu	Ile		
	335		340		345
Ile Arg Ser Lys	Pro Pro Ile Cys Thr	Leu Pro Asn Phe Leu	Phe		
	350		355		360
Glu Asp Gly Glu	Ser Phe Gly Gln Gly	Arg Asp Arg Ser Ser	Leu		
	365		370		375
Leu Asp Asp Thr	Thr Val Thr Leu Ser	Leu Cys Gln Leu Arg	Asn		
	380		385		390
Arg Leu Lys Asp	Val Gly Gly Glu Ala	Phe Tyr Pro Leu Leu	Glu		
	395		400		405
Asp Asp Gln Ser	Asn Leu Pro His Ser	Asn Ser Asn Asn Glu	Leu		
	410		415		420
Ser Ala Ala Ala	Thr Leu Pro Leu Ile	Ile Arg Glu Lys Asp	Thr		
	425		430		435
Glu Tyr Gln Leu	Asn Arg Ile Ile Leu	Phe Asp Arg Leu Leu	Lys		
	440		445		450
Ala Tyr Pro Tyr	Lys Lys Asn Gln Ile	Trp Lys Glu Ala Arg	Val		
	455		460		465
Asp Ile Pro Pro	Leu Met Arg Gly Leu	Thr Trp Ala Ala Leu	Leu		
	470		475		480
Gly Val Glu Gly	Ala Ile His Ala Lys	Tyr Asp Ala Ile Asp	Lys		
	485		490		495
Asp Thr Pro Ile	Pro Thr Asp Arg Gln	Ile Glu Val Asp Ile	Pro		
	500		505		510
Arg Cys His Gln	Tyr Asp Glu Leu Leu	Ser Ser Pro Glu Gly	His		
	515		520		525
Ala Lys Phe Arg	Arg Val Leu Lys Ala	Trp Val Val Ser His	Pro		
	530		535		540
Asp Leu Val Tyr	Trp Gln Gly Leu Asp	Ser Leu Cys Ala Pro	Phe		
	545		550		555
Leu Tyr Leu Asn	Phe Asn Asn Glu Ala	Leu Ala Tyr Ala Cys	Met		
	560		565		570
Ser Ala Phe Ile	Pro Lys Tyr Leu Tyr	Asn Phe Phe Leu Lys	Asp		
	575		580		585
Asn Ser His Val	Ile Gln Glu Tyr Leu	Thr Val Phe Ser Gln	Met		
	590		595		600
Ile Ala Phe His	Asp Pro Glu Leu Ser	Asn His Leu Asn Glu	Ile		
	605		610		615
Gly Phe Ile Pro	Asp Leu Tyr Ala Ile	Pro Trp Phe Leu Thr	Met		
	620		625		630
Phe Thr His Val	Phe Pro Leu His Lys	Ile Phe His Leu Trp	Asp		
	635		640		645
Thr Leu Leu Leu	Gly Asn Ser Ser Phe	Pro Phe Cys Ile Gly	Val		
	650		655		660
Ala Ile Leu Gln	Gln Leu Arg Asp Arg	Leu Leu Ala Asn Gly	Phe		
	665		670		675
Asn Glu Cys Ile	Leu Leu Phe Ser Asp	Leu Pro Glu Ile Asp	Ile		
	680		685		690

Glu	Arg	Cys	Val	Arg	Glu	Ser	Ile	Asn	Leu	Phe	Cys	Trp	Thr	Pro
				695					700					705
Lys	Ser	Ala	Thr	Tyr	Arg	Gln	His	Ala	Gln	Pro	Pro	Lys	Pro	Ser
				710					715					720
Ser	Asp	Ser	Ser	Gly	Gly	Arg	Ser	Ser	Ala	Pro	Tyr	Phe	Ser	Ala
				725					730					735
Glu	Cys	Pro	Asp	Pro	Pro	Lys	Thr	Asp	Leu	Ser	Arg	Glu	Ser	Ile
				740					745					750
Pro	Leu	Asn	Asp	Leu	Lys	Ser	Glu	Val	Ser	Pro	Arg	Ile	Ser	Ala
				755					760					765
Glu	Asp	Leu	Ile	Asp	Leu	Cys	Glu	Leu	Thr	Val	Thr	Gly	His	Phe
				770					775					780
Lys	Thr	Pro	Ser	Lys	Lys	Thr	Lys	Ser	Ser	Lys	Pro	Lys	Leu	Leu
				785					790					795
Val	Val	Asp	Ile	Arg	Asn	Ser	Glu	Glu	Lys	His	Gln	His	Ser	Ile
				800					805					810
Gln	Cys	Cys	Leu	His	Cys	Arg	Arg	Gly	Ala	Tyr	Pro	Gly	Pro	Leu
				815					820					825
His	Cys	Tyr	Ala	Pro	Glu	Leu	Gln	Arg	Glu	Gly	His	Cys	His	Arg
				830					835					840
Gly	Ala	Cys	Gly	Lys	Thr	His	Ser							
				845										

<210> 9

<211> 342

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2446376CD1

<400> 9

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

Ser	Ile	Leu	Leu	Gln	Asn	Val	Ser	Ser	Glu	Asp	Ala	Gly	Thr	Tyr	
				230					235					240	
Tyr	Cys	Val	Lys	Phe	Gln	Arg	Lys	Pro	Asn	Arg	Gln	Tyr	Leu	Ser	
				245					250					255	
Gly	Gln	Gly	Thr	Ser	Leu	Lys	Val	Lys	Ala	Lys	Ser	Thr	Ser	Ser	
				260					265					270	
Lys	Glu	Ala	Glu	Phe	Thr	Ser	Glu	Pro	Ala	Thr	Glu	Met	Ser	Pro	
				275					280					285	
Thr	Gly	Leu	Leu	Val	Val	Phe	Ala	Pro	Val	Val	Leu	Gly	Leu	Lys	
				290					295					300	
Ala	Ile	Thr	Leu	Ala	Ala	Leu	Leu	Leu	Ala	Leu	Ala	Thr	Ser	Arg	
				305					310					315	
Arg	Ser	Pro	Gly	Gln	Glu	Asp	Val	Lys	Thr	Thr	Gly	Pro	Ala	Gly	
				320					325					330	
Ala	Met	Asn	Thr	Leu	Ala	Trp	Ser	Lys	Gly	Gln	Glu				
				335					340						

<210> 10

<211> 544

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5562847CD1

<400> 10

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

```

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

```

<210> 11
 <211> 185
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7493671CD1

```

<400> 11
Met Lys Arg Ser Glu Lys Pro Glu Gly Tyr Arg Gln Met Arg Pro
1 5 10 15
Lys Thr Phe Pro Ala Ser Asn Tyr Thr Val Ser Ser Arg Gln Met
20 25 30
Leu Gln Glu Ile Arg Glu Ser Leu Arg Asn Leu Ser Lys Pro Ser
35 40 45
Asp Ala Ala Lys Ala Glu His Asn Met Ser Lys Met Ser Thr Glu
50 55 60
Asp Pro Arg Gln Val Arg Asn Pro Pro Lys Phe Gly Thr His His
65 70 75
Lys Ala Leu Gln Glu Ile Arg Asn Ser Leu Leu Pro Phe Ala Asn
80 85 90
Glu Thr Asn Ser Ser Arg Ser Thr Ser Glu Val Asn Pro Gln Met
95 100 105

```

```

Leu Gln Asp Leu Gln Ala Ala Gly Phe Asp Glu Asp Met Val Ile
      110      115
Gln Ala Leu Gln Lys Thr Asn Asn Arg Ser Ile Glu Ala Ala Ile
      125      130
Glu Phe Ile Ser Lys Met Ser Tyr Gln Asp Pro Arg Arg Glu Gln
      140      145
Met Ala Ala Ala Ala Ala Arg Pro Ile Asn Ala Ser Met Lys Pro
      155      160
Gly Trp Ile Ile Ser Arg Cys Pro Gly Ser Asn Glu Lys Asp Ala
      170      175
Leu Pro Lys Arg Ile
      185

```

<210> 12

<211> 1269

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7499991CD1

<400> 12

```

Met Ala Phe Ser Lys Gly Phe Arg Ile Tyr His Lys Leu Asp Pro
  1      5      10
Pro Pro Phe Ser Leu Ile Val Glu Thr Arg His Lys Glu Glu Cys
      20      25
Leu Met Phe Glu Ser Gly Ala Val Ala Val Leu Ser Ser Ala Glu
      35      40
Lys Glu Ala Ile Lys Gly Thr Tyr Ser Lys Val Leu Asp Ala Tyr
      50      55
Gly Leu Leu Gly Val Leu Arg Leu Asn Leu Gly Asp Thr Met Leu
      65      70
His Tyr Leu Val Leu Val Thr Gly Cys Met Ser Val Gly Lys Ile
      80      85
Gln Glu Ser Glu Val Phe Arg Val Thr Ser Thr Glu Phe Ile Ser
      95      100
Leu Arg Ile Asp Ser Ser Asp Glu Asp Arg Ile Ser Glu Val Arg
      110      115
Lys Val Leu Asn Ser Gly Asn Phe Tyr Phe Ala Trp Ser Ala Ser
      125      130
Gly Ile Ser Leu Asp Leu Ser Leu Asn Ala His Arg Ser Met Gln
      140      145
Glu Gln Thr Thr Asp Asn Arg Phe Phe Trp Asn Gln Ser Leu His
      155      160
Leu His Leu Lys His Tyr Gly Val Asn Cys Asp Asp Trp Leu Leu
      170      175
Arg Leu Met Cys Gly Gly Val Glu Ile Arg Thr Ile Tyr Ala Ala
      185      190
His Lys Gln Ala Lys Ala Cys Leu Ile Ser Arg Leu Ser Cys Glu
      200      205
Arg Ala Gly Thr Arg Phe Asn Val Arg Gly Thr Asn Asp Asp Gly
      215      220
His Val Ala Asn Phe Val Glu Thr Glu Gln Val Val Tyr Leu Asp
      230      235
Asp Ser Val Ser Ser Phe Ile Gln Ile Arg Gly Ser Val Pro Leu
      245      250
Phe Trp Glu Gln Pro Gly Leu Gln Val Gly Ser His Arg Val Arg
      260      265
Met Ser Arg Gly Phe Glu Ala Asn Ala Pro Ala Phe Asp Arg His
      275      280
Phe Arg Thr Leu Lys Asn Leu Tyr Gly Lys Gln Ile Ile Val Asn
      290      295

```

Leu	Leu	Gly	Ser	Lys	Glu	Gly	Glu	His	Met	Leu	Ser	Lys	Ala	Phe
				305					310					315
Gln	Ser	His	Leu	Lys	Ala	Ser	Glu	His	Ala	Ala	Asp	Ile	Gln	Met
				320					325					330
Val	Asn	Phe	Asp	Tyr	His	Gln	Met	Val	Lys	Gly	Gly	Lys	Ala	Glu
				335					340					345
Lys	Leu	His	Ser	Val	Leu	Lys	Pro	Gln	Val	Gln	Lys	Phe	Leu	Asp
				350					355					360
Tyr	Gly	Phe	Phe	Tyr	Phe	Asn	Gly	Ser	Glu	Val	Gln	Arg	Cys	Gln
				365					370					375
Ser	Gly	Thr	Val	Arg	Thr	Asn	Cys	Leu	Asp	Cys	Leu	Asp	Arg	Thr
				380					385					390
Asn	Ser	Val	Gln	Ala	Phe	Leu	Gly	Leu	Glu	Met	Leu	Ala	Lys	Gln
				395					400					405
Leu	Glu	Ala	Leu	Gly	Leu	Ala	Glu	Lys	Pro	Gln	Leu	Val	Thr	Arg
				410					415					420
Phe	Gln	Glu	Val	Phe	Arg	Ser	Met	Trp	Ser	Val	Asn	Gly	Asp	Ser
				425					430					435
Ile	Ser	Lys	Ile	Tyr	Ala	Gly	Thr	Gly	Ala	Leu	Glu	Gly	Lys	Ala
				440					445					450
Lys	Leu	Lys	Asp	Gly	Ala	Arg	Ser	Val	Thr	Arg	Thr	Ile	Gln	Asn
				455					460					465
Asn	Phe	Phe	Asp	Ser	Ser	Lys	Gln	Glu	Ala	Ile	Asp	Val	Leu	Leu
				470					475					480
Leu	Gly	Asn	Thr	Leu	Asn	Ser	Asp	Leu	Ala	Asp	Lys	Ala	Arg	Ala
				485					490					495
Leu	Leu	Thr	Thr	Gly	Ser	Leu	Arg	Val	Ser	Glu	Gln	Thr	Leu	Gln
				500					505					510
Ser	Ala	Ser	Ser	Lys	Val	Leu	Lys	Ser	Met	Cys	Glu	Asn	Phe	Tyr
				515					520					525
Lys	Tyr	Ser	Lys	Pro	Lys	Lys	Ile	Arg	Val	Cys	Val	Gly	Thr	Trp
				530					535					540
Asn	Val	Asn	Gly	Gly	Lys	Gln	Phe	Arg	Ser	Ile	Ala	Phe	Lys	Asn
				545					550					555
Gln	Thr	Leu	Thr	Asp	Trp	Leu	Leu	Asp	Ala	Pro	Lys	Leu	Ala	Gly
				560					565					570
Ile	Gln	Glu	Phe	Gln	Asp	Lys	Arg	Ser	Lys	Pro	Thr	Asp	Ile	Phe
				575					580					585
Ala	Ile	Gly	Phe	Glu	Glu	Met	Val	Glu	Leu	Asn	Ala	Gly	Asn	Ile
				590					595					600
Val	Asn	Ala	Ser	Thr	Thr	Asn	Gln	Lys	Leu	Trp	Ala	Val	Glu	Leu
				605					610					615
Gln	Lys	Thr	Ile	Ser	Arg	Asp	Asn	Lys	Tyr	Val	Leu	Leu	Ala	Ser
				620					625					630
Glu	Gln	Leu	Val	Gly	Val	Cys	Leu	Phe	Val	Phe	Ile	Arg	Pro	Gln
				635					640					645
His	Ala	Pro	Phe	Ile	Arg	Asp	Val	Ala	Val	Asp	Thr	Val	Lys	Thr
				650					655					660
Gly	Met	Gly	Gly	Ala	Thr	Gly	Asn	Lys	Gly	Ala	Val	Ala	Ile	Arg
				665					670					675
Met	Leu	Phe	His	Thr	Thr	Ser	Leu	Cys	Phe	Val	Cys	Ser	His	Phe
				680					685					690
Ala	Ala	Gly	Gln	Ser	Gln	Val	Lys	Glu	Arg	Asn	Glu	Asp	Phe	Ile
				695					700					705
Glu	Ile	Ala	Arg	Lys	Leu	Ser	Phe	Pro	Met	Gly	Arg	Met	Leu	Phe
				710					715					720
Ser	His	Asp	Tyr	Val	Phe	Trp	Cys	Gly	Asp	Phe	Asn	Tyr	Arg	Ile
				725					730					735
Asp	Leu	Pro	Asn	Glu	Glu	Val	Lys	Glu	Leu	Ile	Arg	Gln	Gln	Asn
				740					745					750
Trp	Asp	Ser	Leu	Ile	Ala	Gly	Asp	Gln	Leu	Ile	Asn	Gln	Lys	Asn
				755					760					765
Ala	Gly	Gln	Val	Phe	Arg	Gly	Phe	Leu	Glu	Gly	Lys	Val	Thr	Phe

	770		775		780
Ala Pro Thr Tyr Lys Tyr Asp Leu Phe Ser Asp Asp Tyr Asp Thr					
	785		790		795
Ser Glu Lys Cys Arg Thr Pro Ala Trp Thr Asp Arg Val Leu Trp					
	800		805		810
Arg Arg Arg Lys Trp Pro Phe Asp Arg Ser Ala Glu Asp Leu Asp					
	815		820		825
Leu Leu Asn Ala Ser Phe Gln Asp Glu Ser Lys Ile Leu Tyr Thr					
	830		835		840
Trp Thr Pro Gly Thr Leu Leu His Tyr Gly Arg Ala Glu Leu Lys					
	845		850		855
Thr Ser Asp His Arg Pro Val Val Ala Leu Ile Asp Ile Asp Ile					
	860		865		870
Phe Glu Val Glu Ala Glu Glu Arg Gln Asn Ile Tyr Lys Glu Val					
	875		880		885
Ile Ala Val Gln Gly Pro Pro Asp Gly Thr Val Leu Val Ser Ile					
	890		895		900
Lys Ser Ser Leu Pro Glu Asn Asn Phe Phe Asp Asp Ala Leu Ile					
	905		910		915
Asp Glu Leu Leu Gln Gln Phe Ala Ser Phe Gly Glu Val Ile Leu					
	920		925		930
Ile Arg Phe Val Glu Asp Lys Met Trp Val Thr Phe Leu Glu Gly					
	935		940		945
Ser Ser Ala Leu Asn Val Leu Ser Leu Asn Gly Lys Glu Leu Leu					
	950		955		960
Asn Arg Thr Ile Thr Ile Ala Leu Lys Ser Pro Asp Trp Ile Lys					
	965		970		975
Asn Leu Glu Glu Glu Met Ser Leu Glu Lys Ile Ser Ile Ala Leu					
	980		985		990
Pro Ser Ser Thr Ser Ser Thr Leu Leu Gly Glu Asp Ala Glu Val					
	995		1000		1005
Ala Ala Asp Phe Asp Met Glu Gly Asp Val Asp Asp Tyr Ser Ala					
	1010		1015		1020
Glu Val Glu Glu Leu Leu Pro Gln His Leu Gln Pro Ser Ser Ser					
	1025		1030		1035
Ser Gly Leu Gly Thr Ser Pro Ser Ser Ser Pro Arg Thr Ser Pro					
	1040		1045		1050
Cys Gln Ser Pro Thr Ile Ser Glu Gly Pro Val Pro Ser Leu Pro					
	1055		1060		1065
Ile Arg Pro Ser Arg Ala Pro Ser Arg Thr Pro Gly Pro Pro Ser					
	1070		1075		1080
Ala Gln Ser Ser Pro Ile Asp Ala Gln Pro Ala Thr Pro Leu Pro					
	1085		1090		1095
Gln Lys Asp Pro Ala Gln Pro Leu Glu Pro Lys Arg Pro Pro Pro					
	1100		1105		1110
Pro Arg Pro Val Ala Pro Pro Thr Arg Pro Ala Pro Pro Gln Arg					
	1115		1120		1125
Pro Pro Pro Pro Ser Gly Arg Ser Gln Pro Ser Pro Gln Ala Gly					
	1130		1135		1140
Leu Ala Gly Pro Gly Pro Ala Gly Tyr Ser Thr Ala Arg Pro Thr					
	1145		1150		1155
Ile Pro Pro Arg Ala Gly Val Ile Ser Ala Pro Gln Ser His Ala					
	1160		1165		1170
Arg Ala Ser Ala Gly Arg Leu Thr Pro Glu Ser Gln Ser Lys Thr					
	1175		1180		1185
Ser Glu Thr Ser Lys Gly Ser Thr Phe Leu Pro Glu Pro Leu Lys					
	1190		1195		1200
Pro Gln Ala Ala Phe Pro Pro Gln Ser Ser Leu Pro Pro Pro Ala					
	1205		1210		1215
Gln Arg Leu Gln Glu Pro Leu Val Pro Val Ala Ala Pro Met Pro					
	1220		1225		1230
Gln Ser Gly Pro Gln Pro Asn Leu Glu Thr Pro Pro Gln Pro Pro					
	1235		1240		1245

Pro Arg Ser Arg Ser Ser His Ser Leu Pro Ser Glu Ala Ser Ser
 1250 1255 1260
 Gln Pro Gln Gln Glu Gln Pro Ser Gly
 1265

<210> 13
 <211> 199
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7499977CD1

<400> 13
 Met Leu Trp Leu Ala Leu Gly Pro Phe Pro Ala Met Glu Asn Gln
 1 5 10 15
 Val Leu Val Ile Arg Ile Lys Ile Pro Asn Ser Gly Ala Val Asp
 20 25 30
 Trp Thr Val His Ser Gly Pro Gln Leu Leu Phe Arg Asp Val Leu
 35 40 45
 Asp Val Ile Gly Gln Val Leu Pro Glu Ala Thr Thr Thr Ala Phe
 50 55 60
 Glu Tyr Glu Asp Glu Asp Gly Asp Arg Ile Thr Val Arg Ser Asp
 65 70 75
 Glu Glu Met Lys Ala Met Leu Ser Tyr Tyr Tyr Ser Thr Val Met
 80 85 90
 Glu Gln Gln Val Asn Gly Gln Leu Ile Glu Pro Leu Gln Ile Phe
 95 100 105
 Pro Arg Ala Cys Lys Pro Pro Gly Glu Arg Asn Ile His Gly Leu
 110 115 120
 Lys Leu Leu Gln Cys Ile Val Asp Glu Asp Ser Pro Val Leu Pro
 125 130 135
 Val Gly Glu Phe Ser Glu Pro Phe Val His Phe Ile Thr Gln Cys
 140 145 150
 Met Arg Lys Gln Pro Lys Glu Arg Pro Ala Pro Glu Glu Leu Met
 155 160 165
 Gly His Pro Phe Ile Val Gln Phe Asn Asp Gly Asn Ala Ala Val
 170 175 180
 Val Ser Met Trp Val Cys Arg Ala Leu Glu Glu Arg Arg Ser Gln
 185 190 195
 Gln Gly Pro Pro

<210> 14
 <211> 1214
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 8043337CD1

<400> 14
 Met Glu Lys Tyr Val Arg Leu Gln Lys Ile Gly Glu Gly Ser Phe
 1 5 10 15
 Gly Lys Ala Ile Leu Val Lys Ser Thr Glu Asp Gly Arg Gln Tyr
 20 25 30
 Val Ile Lys Glu Ile Asn Ile Ser Arg Met Ser Ser Lys Glu Arg
 35 40 45
 Glu Glu Ser Arg Arg Glu Val Ala Val Leu Ala Asn Met Lys His
 50 55 60
 Pro Asn Ile Val Gln Tyr Arg Glu Ser Phe Glu Glu Asn Gly Ser

				65					70					75
Leu	Tyr	Ile	Val	Met	Asp	Tyr	Cys	Glu	Gly	Gly	Asp	Leu	Phe	Lys
				80					85					90
Arg	Ile	Asn	Ala	Gln	Lys	Gly	Val	Leu	Phe	Gln	Glu	Asp	Gln	Ile
				95					100					105
Leu	Asp	Trp	Phe	Val	Gln	Ile	Cys	Leu	Ala	Leu	Lys	His	Val	His
				110					115					120
Asp	Arg	Lys	Ile	Leu	His	Arg	Asp	Ile	Lys	Ser	Gln	Asn	Ile	Phe
				125					130					135
Leu	Thr	Lys	Asp	Gly	Thr	Val	Gln	Leu	Gly	Asp	Phe	Gly	Ile	Ala
				140					145					150
Arg	Val	Leu	Asn	Ser	Thr	Val	Glu	Leu	Ala	Arg	Thr	Cys	Ile	Gly
				155					160					165
Thr	Pro	Tyr	Tyr	Leu	Ser	Pro	Glu	Ile	Cys	Glu	Asn	Lys	Pro	Tyr
				170					175					180
Asn	Asn	Lys	Ser	Asp	Ile	Trp	Ala	Leu	Gly	Cys	Val	Leu	Tyr	Glu
				185					190					195
Leu	Cys	Thr	Leu	Lys	His	Ala	Phe	Glu	Ala	Gly	Ser	Met	Lys	Asn
				200					205					210
Leu	Val	Leu	Lys	Ile	Ile	Ser	Gly	Ser	Phe	Pro	Pro	Val	Ser	Leu
				215					220					225
His	Tyr	Ser	Tyr	Asp	Leu	Arg	Ser	Leu	Val	Ser	Gln	Leu	Phe	Lys
				230					235					240
Arg	Asn	Pro	Arg	Asp	Arg	Pro	Ser	Val	Asn	Ser	Ile	Leu	Glu	Lys
				245					250					255
Gly	Phe	Ile	Ala	Lys	Arg	Ile	Glu	Lys	Phe	Leu	Ser	Pro	Gln	Leu
				260					265					270
Ile	Ala	Glu	Glu	Phe	Cys	Leu	Lys	Thr	Phe	Ser	Lys	Phe	Gly	Ser
				275					280					285
Gln	Pro	Ile	Pro	Ala	Lys	Arg	Pro	Ala	Ser	Gly	Gln	Asn	Ser	Ile
				290					295					300
Ser	Val	Met	Pro	Ala	Gln	Lys	Ile	Thr	Lys	Pro	Ala	Ala	Lys	Tyr
				305					310					315
Gly	Ile	Pro	Leu	Ala	Tyr	Lys	Lys	Tyr	Gly	Asp	Lys	Lys	Leu	His
				320					325					330
Glu	Lys	Lys	Pro	Leu	Gln	Lys	His	Lys	Gln	Ala	His	Gln	Thr	Pro
				335					340					345
Glu	Lys	Arg	Val	Asn	Thr	Gly	Glu	Glu	Arg	Arg	Lys	Ile	Ser	Glu
				350					355					360
Glu	Ala	Ala	Arg	Lys	Arg	Arg	Leu	Glu	Phe	Ile	Glu	Lys	Glu	Lys
				365					370					375
Lys	Gln	Lys	Asp	Gln	Ile	Ile	Ser	Leu	Met	Lys	Ala	Glu	Gln	Met
				380					385					390
Lys	Arg	Gln	Glu	Lys	Glu	Arg	Leu	Glu	Arg	Ile	Asn	Arg	Ala	Arg
				395					400					405
Glu	Gln	Gly	Trp	Arg	Asn	Val	Leu	Ser	Ala	Gly	Gly	Ser	Gly	Glu
				410					415					420
Val	Lys	Ala	Pro	Phe	Leu	Gly	Ser	Gly	Trp	Thr	Ile	Ala	Pro	Ser
				425					430					435
Ser	Phe	Ser	Ser	Arg	Gly	Gln	Tyr	Glu	His	Tyr	His	Ala	Ile	Phe
				440					445					450
Asp	Gln	Met	Gln	Gln	Gln	Arg	Ala	Glu	Asp	Asn	Glu	Ala	Lys	Trp
				455					460					465
Lys	Arg	Glu	Ile	Tyr	Gly	Arg	Gly	Leu	Pro	Glu	Arg	Gln	Lys	Gly
				470					475					480
Gln	Leu	Ala	Val	Glu	Arg	Ala	Lys	Gln	Val	Glu	Glu	Phe	Leu	Gln
				485					490					495
Arg	Lys	Arg	Glu	Ala	Met	Gln	Asn	Lys	Ala	Arg	Ala	Glu	Gly	His
				500					505					510
Met	Val	Tyr	Leu	Ala	Arg	Leu	Arg	Gln	Ile	Arg	Leu	Gln	Asn	Phe
				515					520					525
Asn	Glu	Arg	Gln	Gln	Ile	Lys	Ala	Lys	Leu	Arg	Gly	Glu	Lys	Lys
				530					535					540

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

1010	1015	1020
Asp Val Pro Thr Val Gly Asp Val Arg Gln Asp Asn Leu Glu Ile		
1025	1030	1035
Asp Glu Ile Glu Asp Glu Asn Ile Lys Glu Gly Pro Ser Asp Ser		
1040	1045	1050
Glu Asp Ile Val Phe Glu Glu Thr Asp Thr Asp Leu Gln Glu Leu		
1055	1060	1065
Gln Ala Ser Met Glu Gln Leu Leu Arg Glu Gln Pro Gly Glu Glu		
1070	1075	1080
Tyr Ser Glu Glu Glu Glu Ser Val Leu Lys Asn Ser Asp Val Glu		
1085	1090	1095
Pro Thr Ala Asn Gly Thr Asp Val Ala Asp Glu Asp Asp Asn Pro		
1100	1105	1110
Ser Ser Glu Ser Ala Leu Asn Glu Glu Trp His Ser Asp Asn Ser		
1115	1120	1125
Asp Gly Glu Ile Ala Ser Glu Cys Glu Cys Asp Ser Val Phe Asn		
1130	1135	1140
His Leu Glu Glu Leu Arg Leu His Leu Glu Gln Glu Met Gly Phe		
1145	1150	1155
Glu Lys Phe Phe Glu Val Tyr Glu Lys Ile Lys Ala Ile His Glu		
1160	1165	1170
Asp Glu Asp Glu Asn Ile Glu Ile Cys Ser Lys Ile Val Gln Asn		
1175	1180	1185
Ile Leu Gly Asn Glu His Gln His Leu Tyr Ala Lys Ile Leu His		
1190	1195	1200
Leu Val Met Ala Asp Gly Ala Tyr Gln Glu Asp Asn Asp Glu		
1205	1210	

<210> 15

<211> 411

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5947007CD1

<400> 15

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

	185		190		195
Cys Asn Pro Arg	Glu Lys Glu Arg Ile	Gln Asn Ala Gly Gly	Ser		
	200		205		210
Val Met Ile Gln	Arg Val Asn Gly Ser	Leu Ala Val Ser Arg	Ala		
	215		220		225
Leu Gly Asp Tyr	Asp Tyr Lys Cys Val	Asp Gly Lys Gly Pro	Thr		
	230		235		240
Glu Gln Leu Val	Ser Pro Glu Pro Glu	Val Tyr Glu Ile Leu	Arg		
	245		250		255
Ala Glu Glu Asp	Glu Phe Ile Ile Leu	Ala Cys Asp Gly Ile	Trp		
	260		265		270
Asp Val Met Ser	Asn Glu Glu Leu Cys	Glu Tyr Val Lys Ser	Arg		
	275		280		285
Leu Glu Val Ser	Asp Asp Leu Glu Asn	Val Cys Asn Trp Val	Val		
	290		295		300
Asp Thr Cys Leu	His Lys Gly Ser Arg	Asp Asn Met Ser Ile	Val		
	305		310		315
Leu Val Cys Phe	Ser Asn Ala Pro Lys	Val Ser Asp Glu Ala	Val		
	320		325		330
Lys Lys Asp Ser	Glu Leu Asp Lys His	Leu Glu Ser Arg Val	Glu		
	335		340		345
Glu Ile Met Glu	Lys Ser Gly Glu Glu	Gly Met Pro Asp Leu	Ala		
	350		355		360
His Val Met Arg	Ile Leu Ser Ala Glu	Asn Ile Pro Asn Leu	Pro		
	365		370		375
Pro Gly Gly Gly	Leu Ala Gly Lys Arg	Asn Val Ile Glu Ala	Val		
	380		385		390
Tyr Ser Arg Leu	Asn Pro His Arg Glu	Ser Asp Gly Gly Ala	Gly		
	395		400		405
Asp Leu Glu Asp	Pro Trp				
	410				

<210> 16

<211> 224

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510576CD1

<400> 16

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

				155					160					165
Gly	Phe	Arg	Ala	Pro	Glu	Val	Ala	Arg	Gly	Asn	Val	Ile	Tyr	Asn
				170					175					180
Gln	Gln	Ala	Asp	Val	Tyr	Ser	Phe	Gly	Leu	Leu	Leu	Tyr	Asp	Ile
				185					190					195
Leu	Thr	Thr	Gly	Gly	Arg	Ile	Val	Glu	Gly	Leu	Lys	Phe	Pro	Asn
				200					205					210
Glu	Phe	Asp	Glu	Leu	Glu	Ile	Gln	Gly	Lys	Leu	Pro	Gly	Leu	
				215					220					

<210> 17

<211> 244

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 90111008CD1

<400> 17

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

<210> 18

<211> 142

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503305CD1

<400> 18

```

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

```

<210> 19

<211> 4723

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5532044CB1

<400> 19

```

cgcgcttgagg cggcgaggagg tgcagctacc tcggcgccgg ctcggctcta ggacgtgtcg 60
ggctgcacgg gtcggggggcg gctccctgct gcttccgagc ctagccgcct ggccttcggc 120
atttgcaggc agacacactg ccctccagaa acgatgggtg tcgctagatc cagctgggac 180
tctgccagcc tccggcactc agtctaagag aaggggccatc ttccagccct gctgtccctg 240
cctgtcatgg ccacattcgg ctgctgtctg catgtctggg gacacagggc acccagcccc 300
agccctgaga agttaagcca ggccttctgg ctcatgtcca caggggcact ccagacccca 360
ggcccttttc agagcaaaaag caactgatgc tatgggagaa gccctgccc acctgtccac 420
tcgcctggca ctgcccactc tgagaccat gcactgggtt cccctggagg tgccaaccct 480
gtgagtcctt cctgtcccc tgacgctgag aaggccctgc caacccccac catgtgtgag 540
gtgatgcca caatcaatga gggggacccc ctgggtcccc ctcatggcgc cgatgctgac 600
gccaacttcg agcagctgat ggtgaacatg ctggacgagc gggagaagtt gctggagtct 660
cttcgggaga ctcaggagac gttggcggcc acacagagcc ggctccagga tgccatacac 720
gagcgggacc agctccagcg ccaccttaac tccgccctcc cccaggaatt tgccacctta 780
acccgggagc tgagcatgtg tcgggagcag cttctagagc gggaggaaga gatatacaga 840
ctgaaagcag aacggaataa cacacggctg cttctggaac atctggagtg cctgggtgtc 900
cgccatgaac ggtcactgct gatgactgtg gtgaagcgcc agggccagtc accttcgggg 960
gtctccagtg aggtggagg gctgaaggcc ctcaagtcac tgtttgagca ccacaaggcc 1020
ctggatgaga aggtgcgaga gcggctccgg gcagcgctgg agcgagtcac caccttggag 1080
gagcagctgg caggtgccc ccagcagggt tctgccctgc agcagggggc aggggtgctg 1140
gatggagcgg cagaagagga ggggactgtg gagctggggc cgaaacgcct gtggaaggag 1200
gatacggggc gggtagagga gctgcaggag ctcctggaga agcagaactt tgagttgagc 1260
caggcccggg agcgactggt caccctaaca acaaccgtga ctgaactcga ggaggacctg 1320
ggcacggccc gccgggacct catcaagtgc gaggagctga gcagcaagca tcagcgggac 1380
ctccgggagg ctctggccca gaaggaggac atggaagagc ggattactac actggagaag 1440
cgctacctgg ctgctcagcg tgaggcaaca tccatccatg acctcaatga caagctggag 1500
aatgagctgg ccaacaagga gtccctgcac cgccagtgtg aggagaaggc ccgacacctg 1560
caggagctgc tggaggtggc agagcagaag ctgcagcaga cgatgcgcaa ggcagagacg 1620
ctgccagagg tggaggctga gctggcccag agaattgcag ccctcaccaa ggcaaacg 1680
catggcaaca ttgaggagca cctgcggcag ctggaggggc agctggagga gaagaaccag 1740
gagctggcac ggggtgcgcca gcgggaaaag atgaatgagg accacaacaa gcggctgtcg 1800
gacacagtgg accggctgct cagcgagtcc aacgagcgtc tgcagctcca cctgaaggag 1860

```

```

cgcatggctg ccctggagga gaagaacacg ttgatccagg agttggagag ctcccagcgg 1920
cagattgagg agcagcacca ccacaagggc cgcctgtctg aagagattga gaagctgcgc 1980
caagaggtgg accagctgaa gggccgaggg gggccgtttg tggatggcgt ccactccagg 2040
tcgcacatgg gcaagtgcagc agacgtgcgg ttctccctgg gcacaaccac acacgcaccc 2100
ccaggcgtgc atcgccgcta ctcggcattg agggaagagt ctgccaagga ctgggagact 2160
tctccactgc ctgggatgct gggcccggca gctggccctg cctttgacag tgaccctgag 2220
atctccgacg tggatgagga tgagccaggg ggtctgggtg gctctgcgga tgttgtctcc 2280
cccagcggcc actcagatgc ccagaccctg gccatgatgc tgcaggagca gctggatgcc 2340
atcaatgagg aatcaggat gattcaggaa gagaaggagt ccacggagct ccgcgcggag 2400
gagattgaga cgcgtgtaac cagtggcagc atggaagccc taaacctgaa gcagctgcgc 2460
aagcgtggtt ccatccccac ctctctgacg gccctgtccc tggccagcgc gtccccacca 2520
ctcagcgggc gctccacacc taagctcacc tcccgagtg ctgccagga cctggaccga 2580
atgggggtca tgaccctgcc cagtgactta agaaagcata ggaggaagct gctgtcgcca 2640
gtgtctcggg aagagaaccg agaggataaa gccaccataa aatgtgagac ttctcctcct 2700
tcctcaccca ggacgtgcg gctagagaag cttggccacc cagccctgag ccaggaagaa 2760
ggcaagagtg ccttgaggga tcagggcagc aaccccagca gcagcaacag cagccaggac 2820
tccttgacac agggcgccaa gcgcaagggc atcaagtcgt ccattggccg cctgtttggg 2880
aagaaggaga agggcaggct gatccagctg agtcgggatg gagccacagg ccatggtctc 2940
tgtcccttcc ctaacggagt tctgctaaca gactccgaat tcagtatgca ggagcctatg 3000
gtgcttgcca agctggggac ccaggcagag aaggaccggc ggctaaagaa gaaacaccag 3060
ctgcttgaag atgcccgcag gaaaggaatg ccctttgccc agtgggatgg tcctactgtg 3120
gtctcctggt tggagctctg ggtggggatg cctgcctggt atgtggcagc ctgccgggccc 3180
aacgtcaaga gtggtgccat catgtccgct ctgtcggaca cagagatcca gcgggagatc 3240
ggcatcagca atgcccgtca ccggctcaag ctccgcctgg ccattcagga gatggtgtca 3300
ttgaccagcc cctctgcccc acccacctcc aggacttctt ctgggaatgt ctgggtcacc 3360
catgaagaga tggaaactct ggaaacatct actaaaacag acagtgagga gggcagctgg 3420
gctcagaccc tggcctatgg ggacatgaac catgagtgga ttgggaatga atggctacc 3480
agcctggggc tcccgcagta ccgcagctac ttcatggagt gcctggtgga cgccgcagtg 3540
ctggaccacc tcaccaagaa ggacctgcgg gtccacctga agatggtgga cagcttccat 3600
cgaaccagtc ttcagtatgg catcatgtgt ctgaagaggc tgaattatga ccggaaggag 3660
ctggagaaga ggcgagagga gagccagcat gagatcaagg atgtgttagt ctggaccaac 3720
gaccaggtgg ttcattgggt ccagtctatt gggctccggg actacgcagg aaacctgcat 3780
gagagtgggtg tgcatggagc cttgctggcc ctggacgaga acttcgacca caacacactg 3840
gccctgatcc tcagatccc cacacagaac acccaggcac gccaatgat ggaaagagag 3900
ttcaataacc tgttggcctt gggcacagac cggaaagctgg atgacgggga tgacaagggtg 3960
tttcgccgcg cgccctcctg gaggaagcgc ttccggccgc gggagcacca cggctgcggc 4020
ggcatgctca gcgcttccgc ggagaccctc ccggcggggt tccgtgtgtc caccctgggg 4080
accctgcagc ccccaccggc ccgcgcaaaag aagatcatgc ctgaagctgc actaacgctc 4140
cgcggggagc gtgtgtgcgc actaaccgcg cgctctgtgt gttctccgcg ggctgccgac 4200
ttctcccagc tgggacggcg gggtcgcaga gactggagac ctccacgggt cggacctact 4260
cctgtgacc ccccatcctc ccgccccggg tctgacgggg gagtgcctgt ggctcgggct 4320
cctagtctcg tccgtgactt tccggttgcc ctggatctca gaatatattc gtccaccccc 4380
tcggcacccc attaccccga gtcccaccgt ggtccgttg tgctgaaaa tgttcaagac 4440
ttatgttcag ggtgggatga tggaaaccgag ggcttcatca aagtgaagg aaaggaaaag 4500
catctggcat gtgtttcttg gataggggcc agtgcagtgc catcctacag gtggctggag 4560
cagctgcttg gcaacctgat caccttgagt tctgagcagg gactaggctt gcaggtgaga 4620
taatgggcca gggcacccag tccagaagga gcaatggcac ctgggcagtg ccagggtcta 4680
aagcccgtg ctcttttctg gtagaggaga ggcccatcac tgg 4723

```

<210> 20

<211> 513

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7497825CB1

<400> 20

```

atgaaccacc cagctcctgt gaaagtcaca tacaagaaca tgagatttcc tattacacac 60
aatccaacca atgtgacctt aaataaattt atagaggagc ttaagaagta tggagctacc 120
acaatagtaa gagtatgtga agcaacttat gacactactc ttgtggagaa agaaggtatc 180
catgttctca attggccttt tggatgatgg gcaccaccat ccaaccagat tgttgctgat 240

```

```

tggttacatt ttgtaaaaat taagttttgt gaagaacctg gttgttatat tgetgttaat 300
tgcattgtag gccttgggaa agctccagta cttgttgccc tagcatcagt tgaagggtga 360
atgaaacatg aagatgcagt acaattcata ggacaaaagc ggagtggagc ttttaaaagc 420
aagcaacttt tgtattttgga gaagtatcat cctaaaatgc ggctgcgctt caaagattcc 480
aatagtcata taaacaactg ttgcattcaa taa 513

```

<210> 21

<211> 3087

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6464221CB1

<400> 21

```

gcttggggccg ccatcttaga tggcggggagt aagaggaaaa cgattgtgag gcggggaacgg 60
ctttctgctg ctttttttgg gccccgaaaa gggtcagctg gccggggcttt gggggcgcggtg 120
ccctgaggcg cggagcgcggt ttgctacgat gcgggggctg ctcggggctc cgtcccctgg 180
gctggggacg cgccgaatgt gaccgcctcc cgctccctca cccgccgcgg ggaggaggag 240
cgggcgagaa gctgccgccc aacgacagga cgttggggcg gcctggctcc ctcagggttta 300
agaattgttt aagctgcctc aatggagcac atacaggag cttggaagac gatcagcaat 360
ggttttggat tcaaagatgc cgtgtttgat ggctccagct gcatctctcc tacaatagtt 420
cagcagtttg gctatcagcg ccgggcatca gatgatggca aactcacaga tccttctaag 480
acaagcaaca ctatccgtgt tttcttgccc aacaagcaaa gaacagtggg caatgtgcga 540
aatggaatga ccttgcatga ctgccttatg aaagcactca aggtgagggg cctgcaacca 600
gagtgtgtgt cagtgttcag acttctccac gaacacaaag gtaaaaaagc acgcttagat 660
tggaatactg atgtgcgtc tttgattgga gaagaacttc aagtagattt cctggatcat 720
gttcccctca caacacacaa ctttgcctcg aagacgttcc tgaagcttgc cttctgtgac 780
atctgtcaga aattcctgct caatggattt cgatgtcaga cttgtggcta caaatttcat 840
gagcactgta gcaccaaagt acctactatg tgtgtggact ggagtaacat cagacaactc 900
ttttctcagc acagatattc tacacctcac gccttcacct ttaacacctc cagtccctca 960
tctgaagggt cctctctcca gaggcagagg tcgacatcca cacctaattg ccacatgggtc 1020
agcaccaccc tgctgttgga cagcaggatg attgaggatg caattcgaag tcacagcgaa 1080
tcagcctcac cttcagccct gtccagtagc cccaacaatc tgagcccaac aggcctggta 1140
cagccgaaaa cccccgtgcc agcacaaaga gagcgggcac cagtatctgg gacccaggag 1200
aaaaacaaaa ttaggcctcg tggacagaga gattcaagct attattggga aatagaagcc 1260
agtgaagtga tgctgtccac tcggattggg tcaggctctt ttggaactgt ttataagggg 1320
aaatggcacg gagatgttgc agtaaagatc ctaaagggtg tcgacccaac ccagagcaa 1380
ttccaggcct tcaggaatga ggtggctgtt ctgcgcgaaa cacggcatgt gaacattctg 1440
cttttcatgg ggtacatgac aaaggacaac ctggcaattg tgacccagtg gtgcgagggc 1500
agcagcctct acaaacacct gcatgtccag gagaccaagt ttcagatgtt ccagctaatt 1560
gacattggcc ggcagacggc tcagggaatg gactatttgc atgcaaagaa catcatccat 1620
agagacatga aatccaacaa tatatttctc catgaaggct taacagtga aattggagat 1680
tttggtttgg caacagtaaa gtcacgctgg agtggttctc agcaggttga acaacctact 1740
ggctctgtcc tctggatggc ccagagggtg atccgaatgc aggatacaa ccattcagt 1800
ttccagtcgg atgtctactc ctatggcatc gtatgtatg aactgatgac gggggagctt 1860
ccttattctc acatcaacaa ccgagatcag atcatcttca tgggtgggccc aggatatgcc 1920
tccccagatc ttagtaagct atataagaac tgccccaaa caatgaagag gctggtagct 1980
gactgtgtga agaaagtaaa ggaagagagg cctcttttcc cccagatcct gtcttccatt 2040
gagctgtccc aacactctct accgaagatc aaccggagcg cttccgagcc atccttgc 2100
cgggcagccc aactgagga tatcaatgct tgcacgctga ccacgtcccc gaggtgcct 2160
gtcttctagt tgacttttga cctgtcttca ggctgccagg ggaggaggag aagccagcag 2220
gcaccacttt tctgtctcct ttctccagag gcagaacaca tgttttcaga gaagctgctg 2280
ctaaggacct tctagactgc tcacagggcc ttaacttcat gttgccttct tttctatccc 2340
tttgggcccct gggagaagga agccatttgc agtgcgtgtg tgtcctgctc cctccccaca 2400
ttccccatgc tcaaggccca gccttctgta gatgcgcaag tggatgttga tggtagtaca 2460
aaaagcaggg gccagcccc agctgttggc tacatgagta tttagaggaa gtaaggtagc 2520
aggcagtcca gccctgatgt ggagacacat gggattttgg aaatcagctt ctggagggaat 2580
gcatgtcaca ggccgggactt tcttcagaga gtggtgcagc gccagacatt ttgcacataa 2640
ggcaccaaac agcccaggac tgccgagact ctggccgccc gaaggagcct gctttgttac 2700
tatggaactt ttcttagggg acacgtcctc ctttcacagc ttctaagggtg tccagtgc 2760
tgggatgggt ttccaggcaa ggcactcggc caatccgcat ctcagccctc tcaggagaga 2820

```

```

gtcttccatc atgctgaatt ttgtcttcca ggagctgccc ctatggggcg gggccgcagg 2880
gccagccttg tttctctaac aaacaaacaa acaaacagcc ttgtttctct agtcacatca 2940
tgtgtataca aggaagccag gaatacaggt tttcttgatg atttgggttt taattttgtt 3000
tttattgcac ctgacaaaat acagttatct gatggtcctt caattatgtt attttaataa 3060
aataaattaa atttaggtgt aaaaaaa 3087

```

<210> 22
 <211> 1997
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 90044205CB1

```

<400> 22
ggcctgggag acacacagag gccaggcctt agcgcggtc agccatgagc aacaggagta 60
gcttttcccg gctcacctgg ttcaggaagc agaggaaagc tgtggtcagc accagcgaca 120
agaagatgcc caacggaatc ttggaggagc aagagcagca aagggtgatg ctgctcagca 180
ggtcaccctc agggcccaag aagtattttc ccatccccgt ggagcacctg gaggaggaga 240
tccgtatcag atccgccgac gactgcaagc agtttcggga ggagttcaac tcattgccat 300
ctggacacat acaaggaact tttgaactgg caaataaaga agaaaacaga gaaaaaaaca 360
gatatcccaa catccttccc aatgaccatt ctagggtgat tctgagccaa ctggatggaa 420
ttccctgttc agactacatc aatgcttcc tcatagatgg ttacaaagag aagaataaat 480
tcatagcagc tcaaggtccc aaacaggaaa cggttaacga cttctggaga atggtctggg 540
agcaaaagtc tgcgaccatc gtcatgttaa caaacttgaa agaaaggaaa gaggaaaagt 600
gccatcagta ctggcccagc caaggctgct ggacctatgg aaacatccgg gtgtgcgtgg 660
aggactgcgt ggttttggte gactacacca tccggaagtt ctgcatacag ccacagctcc 720
ccgacggctg caaagccccc aggctggtct cccagctgca cttcaccagc tggcccagct 780
tcggagtgc ttttaccccc attgggatgc tgaagttcct caagaaagta aagacgctca 840
accccgctgc cgctgggccc atcgtggtcc actgtagcgc gggcgtgggc cggacgggca 900
ccttcattgt gatcgatgcc atgatggcca tgatgcacgc ggagcagaag gtggatgtgt 960
ttgaatttgt gtctcgaatc cgtaatcagc gccctcagat ggttcaaacg gatatgcagt 1020
acacgttcat ctaccaagcc ttactcgagt actacctcta cggggacaca gagctggagc 1080
tgtcctccct ggagaagcac ctgcagacca tgcacggcac caccaccac ttcgacaaga 1140
tcgggctgga ggaggagtcc aggaaattga caaatgtccg gatcatgaag gagaacatga 1200
ggacgggcaa cttgccggca aacatgaaga aggcagggt catccagatc atcccgtatg 1260
acttcaaccg agtgatcctt tccatgaaaa ggggtcaaga atacacagac tacatcaacg 1320
catccttcat agacggctac cgacagaagg actatttcat cgccaccag gggccactgg 1380
cacacacggg tgaggacttc tggaggatga tctgggaatg gaaatcccac actatcgtga 1440
tgctgacgga ggtgcaggag agagagcagg ataaatgcta ccagtattgg ccaaccgagg 1500
gctcagttac tcatggagaa ataacgattg agataaagaa tgataccctt tcagaagcca 1560
tcagtatacg agactttctg gtcactctca atcagcccca ggcccggcag gaggagcagg 1620
tccgagtagt gcgccagttt cacttccacg gctggcctga gatcgggatt cccgccgagg 1680
gcaaaggcat gattgacctc atgcagccg tgcagaagca gcagcagcag acaggcaacc 1740
accccatcac cgtgcactgc agtgccggag ctgggcgaac aggtacattc atagccctca 1800
gcaacathtt ggagcgagta aaagccgagg gacttttaga tgtatttcaa gctgtgaaga 1860
gtttacgact tcagagacca catatggtgc aaaccctgga acagtatgaa ttctgctaca 1920
aagtggatca agattttatt gatataatct ctgattatgc taatttcaaa tgaagattcc 1980
tgccttaaaa tatTTTT 1997

```

<210> 23
 <211> 1183
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 603340CB1

```

<400> 23
ggcggttcgg gctagagaag agaggcgctt gcgcttgcca gctgggcttg tgagtggggc 60
tgccgagagg gcaggcgtgg ggcgaggcca aaggactgaa cccgcaggag cgtcacgggc 120

```

```

gccggggcg  ctgccgacgg  cgggactggg  ttttctatca  gatgttccac  gtaataatgc  180
tgaggttaag  aagtttccat  tattttgctc  caaaccagaa  gactctgttc  cctgtatata  240
gaataggagt  aatatttgaa  aacaactggc  tgatgtttaa  aactgaagat  tgtcatgatt  300
gtttatccta  atcccaatgc  tgaagtaaga  ttgtcttgga  aataactaag  tggggtaatc  360
caaactctatt  tctggaacca  tgaaaatttt  gctagttttt  gactttgaca  atacaatcat  420
agatgacaat  agtgacactt  ggattgtaca  atgtgctccc  aacaaaaagc  ttcctattga  480
actacgtgat  tcttatcgaa  aaggattttg  gacagaattt  atgggcagag  tctttaagta  540
tttgggagat  aagggtgtaa  gagaacatga  aatgaaaaga  gcagtgcacat  cattgccttt  600
cactccaggg  atggttggaac  tcttcaactt  tataagaaag  aataaggata  aatttgactg  660
cattattatt  tcagattcaa  attcggctct  catagattgg  gttttagaag  ctgccagttt  720
tcatgacata  tttgataaag  tgtttacaaa  tccagcagct  tttaatagca  atggctcatc  780
cactgttgaa  aattatcata  ctcatctctg  caatagatgc  ccaaagaatc  tttgcaaaaa  840
ggtagttttg  atagaatttg  tagataaaca  gttacaacag  ggagtgaatt  atacacaaat  900
tgtttatatt  ggtgatggtg  gaaatgatgt  ctgtccagtc  acctttttta  agaagatga  960
tgttgccatg  ccacggaaag  gatatacctt  acagaaaact  ctttccagaa  tgtctcaaaa  1020
tcttgagcct  atggaatatt  ctgttgtagt  ttggctctca  ggtgttgata  taatttctca  1080
tttacaattt  ctaataaagg  attaatatgt  cagcaattga  aaagtgtatc  acttaagatt  1140
aggttcagca  gcataaaact  aaaaccccaa  ataccagagg  ctt  1183

```

<210> 24

<211> 8752

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 71059650CB1

<400> 24

```

gagatctgga  gataggcgcg  ctttcatttt  tcttgcctcc  aagtgccag  accttctgt  60
ggggcctgca  ggatccccgg  ctggcgggtc  gcgagggtg  gccggcctg  gctgctgcac  120
tgcgcgctc  cgctgcgggg  ctccgggcct  gtggactcag  cggagtcgc  tgagtcagtt  180
tcttcccgcg  cgactcccgg  ccgcgcgcgc  gctgcgggtg  aatctggtcc  caggaggcgg  240
cgctccgccc  gggctccggtc  taggcgtgcg  tggggggccac  ggtcacggtc  atcccagcca  300
ggccccggtc  cagcagcccc  acggccgcgc  ccagagttct  gcgcggcccc  tcgcctcgcc  360
ggagcctctg  gcagggcccc  gagctcgttt  ttggggcctg  agtgggggag  gaggaagccg  420
agcaggaggg  ctocggagag  ggagggcaac  gcggggcggg  gagctgcctc  ctctctcata  480
aacaggcggg  cgtgggcgcc  gatggggccc  gcggggagcg  ctggctgcgg  gcggtgagct  540
gagctcgccc  ccggggagct  gtggccggcg  cccctgccgg  ttccctgagc  agcggacgtt  600
catgctggga  gggcggcggg  ttggaagcag  gtgccaccat  ggctagtggc  agctgtcagg  660
ggtgcgaaga  ggacgaggaa  actctgaaga  agttgatagt  caggctgaac  aatgtccagg  720
aaggaaaaaca  gatagaaacg  ctggtccaaa  tcttgaggga  tctgctggtg  ttcacgtact  780
ccgagcacgc  ctccaagtta  tttcaaggca  aaaatatcca  tgtgcctctg  ttgatcgtct  840
tggaactccta  tatgagagtc  gcgagtgtgc  agcaggtggg  ttggtcactt  ctgtgcaaat  900
taatagaagt  ctgtccagg  acaatgcaaa  gcttaatggg  accccaggat  gttggaaatg  960
attgggaagt  ccttggtgtt  caccaattga  ttcttaaaat  gctaacagtt  cataatgcca  1020
gtgtaaactt  gtcagtgatt  ggactgaaga  ccttagatct  cctcctaact  tcaggtaaaa  1080
tcaccttgct  gatattggat  gaagaaagtg  atattttcat  gttaattttt  gatgccatgc  1140
actcatttcc  agccaatgat  gaagtccaga  aacttggatg  caaagcttta  catgtgctgt  1200
ttgagagagt  ctgagaggag  caactgactg  aatttggtga  gaacaaagat  tatatgatat  1260
tgtaagtgc  gtttaacaaat  tttaaagatg  aagaggaaat  tgtgcttcat  gtgctgcatt  1320
gtttacattc  cctagcgatt  ccttgcaata  atgtggaagt  cctcatgagt  ggcaatgtca  1380
ggtgttataa  tattgtggtg  gaagctatga  aagcattccc  tatgagtga  agaattcaag  1440
aagtgagttg  ctgtttgctc  cataggctta  cattaggtaa  ttttttcaat  atcctggtat  1500
taaacgaagt  ccatgagttt  gtggtgaaag  ctgtgcagca  gtaccagag  aatgcagcat  1560
tgcagatctc  agcgctcagc  tgtttgccc  tctcactga  gactattttc  ttaaatcaag  1620
atthagagga  aaagaatgag  aatcaagaga  atgatgatga  gggggaagaa  gataaattgt  1680
tttggctgga  agcctgttac  aaagcattaa  cgtggcatag  aaagaacaag  cacgtgcagg  1740
aggccgcag  ctgggcacta  aataatctcc  ttatgtacca  aaacagttta  catgagaaga  1800
ttggagatga  agatggccat  ttcccagctc  atagggaagt  gatgctctcc  atgctgatgc  1860
attcttcac  aaagggaagt  ttccaggcat  ctgcgaatgc  attgtcaact  ctcttagaac  1920
aaaatgttaa  tttcagaaaa  atactgttat  caaaaggaat  acacctgaat  gttttggagt  1980
taatgcagaa  gcataacat  tctcctgaag  tggctgaaag  tggctgtaaa  atgctaaatc  2040

```

atcttttttga	aggaagcaac	acttccttgg	atataatggc	agcagtgggc	cccaaaatac	2100
taacagttat	gaaacgtcat	gagacatcat	taccagtgc	gctggaggcg	cttcgagcta	2160
ttttacattt	tatagtgcct	ggcatgccag	aagaatccag	ggaggataca	gaatttcac	2220
ataagctaaa	tatggttaaa	aaacagtgtt	tcaagaatga	tattcacaaa	ctggctctag	2280
cagcttttga	caggttcatt	ggaaatcctg	ggattcagaa	atgtggatta	aaagtaattt	2340
cttctattgt	acattttcct	gatgcattag	agatgttatc	cctggaaggt	gctatggatt	2400
cagtgtttca	cacactgcag	atgtatccag	atgaccaaga	aattcagtgt	ctgggtttta	2460
gtcttatagg	atacttgatt	acaaagaaga	atgtgttcat	aggaactgga	catctgctgg	2520
caaaaattct	ggtttccagc	ttataccgat	taaaggatgt	tgctgaaata	cagactaaag	2580
gatttcagac	aatcttagca	atcctcaa	tgctcagcatc	tttttctaag	ctgctgggtc	2640
atcattcatt	tgacttagta	atattccatc	aaatgtcttc	caatatcatg	gaacaaaagg	2700
atcaacagtt	tctaaacctc	tggttgcaagt	gttttgcaaa	agtagctatg	gatgattact	2760
taaaaaatgt	gatgctagag	agagcgtgtg	atcagaataa	cagcatcatg	gttgaatgct	2820
tgcttctatt	gggagcagat	gccaatcaag	caaaggaggg	atcttcttta	atttgtcagg	2880
tatgtgagaa	agagagcagt	cccaaattgg	tggaactcct	actgaatagt	ggatctcgtg	2940
aacaagatgt	acgaaaagcg	ttgacgataa	gcattgggaa	aggtgacagc	cagatcatca	3000
gcttgctctt	aaggaggctg	gccctggatg	tggccaacaa	tagcatttgc	cttggaggat	3060
tttgtatagg	aaaagttgaa	ccttcttggc	ttggctcctt	atttccagat	aagacttcta	3120
atttaaggaa	acaaacaaat	atagcatcta	cactagcaag	aatggtgatc	agatatcaga	3180
tgaaaagtgc	tgtggaagaa	ggaacagcct	caggcagcga	tggaaatttt	tctgaagatg	3240
tgctgtctaa	atttgatgaa	tggaccttta	ttcctgactc	ttctatggac	agtgtgtttg	3300
ctcaaagtga	tgacctggat	agtgaaggaa	gtgaaggctc	atttcttgtg	aaaaagaaat	3360
ctaattcaat	tagtgttagga	gaattttacc	gagatgccgt	attacagcgt	tgctcaccaa	3420
atttgcaaag	acattccaat	tccttggggc	ccatttttga	tcatgaagat	ttactgaagc	3480
gaaaaagaaa	aatattatct	tcagatgatt	cactcaggct	atcaaaactt	caatcccata	3540
tgaggcattc	agacagcatt	tcttctctgg	cttctgagag	agaatatatt	acatcactag	3600
acctttcagc	aaatgaacta	agagatatgt	atgcccctaag	ccagaaatgc	tgtataagtg	3660
ttcattttga	gcatcttgaa	aagctggagc	ttcaccagaa	tgactcacg	agctttccac	3720
aacagctatg	tgaaactctg	aagagtttga	cacattttga	cttgccacagt	aataaattta	3780
catcatttcc	ttcttattttg	ttgaaaatga	gttgtattgc	taatcttgat	gtctctcgaa	3840
atgacattgg	accttcagtg	gttttagatc	ctacagtga	atgtccaact	ctgaaacagt	3900
ttaacctgtc	atataaccag	ctgtcttttg	tacctgagaa	cctcactgat	gtggtagaga	3960
aactggagca	gctcatttta	gaaggaaata	aaatatcagg	gatatgctcc	cccttgagac	4020
tgaaggaaact	gaagatttta	aaccttagta	agaaccacat	ttcatcccta	tcagagaact	4080
ttcttgaggc	ttgtctctaa	gtggagagtt	tcagtgccag	aatgaatttt	cttgctgcta	4140
tgcccttctt	gectctctct	atgacaatcc	taaaattatc	tcagaacaaa	ttttcctgta	4200
ttccagaagc	aatttttaaat	cttccacact	tgcggtcttt	agatatgagc	agcaatgata	4260
ttcagtacct	accagggtccc	gcacactgga	aatctttgaa	cttaagggaa	ctcttattta	4320
gccataatca	gatcagcatc	ttggacttga	gtgaaaaagc	atattttatg	tctagagtag	4380
agaaactgca	tctttctcac	aataaactga	aagagattcc	tcttgagatt	ggctgtcttg	4440
aaaatctgac	atctctggat	gtcagttaca	acttggaaact	aagatccttt	cccaatgaaa	4500
tgggggaaatt	aagcaaaata	tgggatcttc	ctttggatga	actgcatctt	aaacttgatt	4560
ttaaacatat	aggatgtaaa	gccaaagaca	tcataagggt	tcttcaacag	cgattaaaaa	4620
aggctgtgcc	ttataaccga	atgaaactta	tgattgtggg	aaatactggg	agtggtaaaa	4680
ccaccttatt	gcagcaatta	atgaaaacca	agaaatcaga	tcttgggaatg	caaagtgcc	4740
cagttggcat	agatgtgaaa	gactggccta	tccaaataag	agacaaaaga	aagagagatc	4800
tcgtcctaaa	tgtgtgggat	tttgcaggct	gtgaggaatt	ctatagtact	catccccatt	4860
ttatgacgca	gcgagcattg	taccttgctg	tctatgacct	cagcaaggga	caggctgaag	4920
ttgatgccat	gaagccttgg	ctcttcaata	taaaggctcg	cgcttcttct	tccccgtgta	4980
ttctcgttgg	cacacatttg	gatgtttctg	atgagaagca	gcgcaaagcc	tgcatgagta	5040
aaatcaccaa	ggaactcctg	aataagcgag	ggttccctgc	catacgagat	taccactttg	5100
tgaatgccac	cgaggaatct	gatgctttgg	caaaactctg	gaaaaccatc	ataaacgaga	5160
gccttaattt	caagatccga	gatcagcttg	ttgttggaca	gctgattcca	gactgctatg	5220
tagaacttga	aaaaatcatt	ttatcggagc	gtaaaaatgt	gccaattgaa	tttcccgtaa	5280
ttgaccggaa	acgattatta	caactagtga	gagaaaatca	gctgcagtta	gatgaaaatg	5340
agcttccctc	cgcagttcac	tttctaaatg	aatcaggagt	ccttcttcat	tttcaagacc	5400
cagcactgca	gttaagtga	ttgtactttg	tggaaaccaa	gtggctttgt	aaaatcatgg	5460
cacagatttt	gacagtga	gtggaagggt	gtccaaaaca	ccctaaggga	attatttcgc	5520
gtagagatgt	ggaaaaattt	ctttcaaaga	aaaggaaatt	tccaaagaac	tacatgacac	5580
agtattttta	gcccttagaa	aaattccaga	ttgcttttgc	aataggagaa	gaatatttgc	5640
tggttccaag	cagtttgtct	gaccacaggc	ctgtgataga	gcttcccat	tgtgagaact	5700
ctgaaattat	catccgacta	tatgaaatgc	cttattttcc	aatgggattt	tgggtcaagat	5760
taatcaatcg	attacttgag	atttcacctt	acatgctttc	agggagagaa	cgagcacttc	5820

```

gccccaaacag aatgtatttgg cgacaaggca tttacttaaa ttggtctcct gaagcttatt 5880
gtcttggttagg atctgaagtc ttagacaatc atccagagag tttcttaaaa attacagttc 5940
cttcttgtagt aaaaggctgt attcttttgg gccaaagtgt ggaccacatt gattctctca 6000
tggaagaatg gtttcctggg ttgctggaga ttgatatttg tggatgaagga gaaactctgt 6060
tgaagaaatg ggcattatat agttttaatg atggcgaga acatcaaaaa atcttacttg 6120
atgacttgat gaagaaagca gaggaaggag atctcttagt aaatccagat caaccaaggc 6180
tcaccattcc aatatctcag attgcccctg acttgatttt ggctgacctg cctagaaata 6240
ttatgttgaa taatgatgag ttggaatttg aacaagctcc agagtctctc ctaggtgatg 6300
gcagtttttg atcagttttac cgagcagcct atgaaggaga agaagtggct gtgaagattt 6360
ttaataaaca tatatcactc aggcgtgtaa gacaagagct tgggtgctt tgccacctcc 6420
accacccag tttgatattc ttgctggcag ctgggattcg tccccggatg ttggtgatgg 6480
agttgacctc caagggttcc ttggatcgcc tgcttcagca ggacaaagcc agcctcacta 6540
gaaccctaca gcacaggatt gcactccacg tagctgatgg tttagatac ctccactcag 6600
ccatgattat ataccgagac ctgaaacccc acaatgtgct gcttttcaca ctgtatccca 6660
atgctgccat cattgcaaag attgctgact acggcattgc tcagtactgc tgtagaatgg 6720
ggataaaaac atcagagggc acaccaggtt ttcgtgcacc tgaagttgcc agaggaaatg 6780
tcatttataa ccaacaggct gatgtttatt catttggttt actactctat gacattttga 6840
caactggagg tagaatagta gagggtttga agtttccaaa tgagtttgat gaattagaaa 6900
tacaaggaaa attacctgat ccagttaaag aatatggttg tgccccatgg cctatggttg 6960
aaaaattaat taaacagtgt ttgaaagaaa atcctcaaga aaggcctact tctgcccagg 7020
tctttgacat tttgaattca gctgaattag tctgtctgac gagacgcatt ttattaccta 7080
aaaacgtaat tgttgaatgc atggttgcta cacatcaca cagcaggaat gcaagcattt 7140
ggctgggctg tgggcacacc gacagaggac agctctcatt tcttgactta aatactgaag 7200
gatacacttc tgaggaagtt gctgatagta gaatatttg cttagccttg gtgcatcttc 7260
ctgttgaaaa ggaaagctgg attgtgtctg ggacacagtc tggtagcttc ctggtcatca 7320
ataccgaaga tgggaaaaag agacataccc tagaaaagat gactgattct gtcactgtt 7380
tgtattgcaa ttccttttcc aagcaaagca acaaaaaaaa tttcttttgg gttggaaccg 7440
ctgatggcaa gtttagcaatt tttgaagata agactgttaa gcttaaagga gctgctcctt 7500
tgaagatact aaatatagga aatgtcagta ctccattgat gtgtttgagt gaatccacaa 7560
attcaacgga aagaaatgta atgtggggag gatgtggcac aaagattttc tccttttcta 7620
atgatttcac cattcagaaa ctcatcgaga caagaacaag ccaactgttt tcttatgcag 7680
ctttcagtgat ttccaacatc ataacagtgg tggtagacac tgctctctat attgctaagc 7740
aaaatagccc tgttgttgaa gtgtgggata agaaaactga aaaactctgt ggactaatag 7800
actgctgca ctttttaagg gaggtaacgg taaaagaaaa caaggaatca aaacacaaaa 7860
tgtcttattc tgggagagtg aaaacctctt gccttcagaa gaacactgct ctttggatag 7920
gaactggagg agggcatatt ttactcctgg atctttcaac tcgtcgactt atacgtgtaa 7980
tttacaactt ttgtaattcg gtcagagtc tgaatgacagc acagctagga agccttaaaa 8040
atgtcatgct ggtattgggc tacaaccgga aaaatactga aggtacacaa aagcagaaag 8100
agatacaatc ttgcttgacc gtttgggaca tcaatcttcc acatgaagtg caaaatttag 8160
aaaaacacat tgaagtgaga aaagaattag ctgaaaaaat gagacgaaca tctgttgagt 8220
aagagagaaa taggaattgt ctttggatag gaaaattatt ctctcctctt gtaaatattt 8280
attttaaaaa tgttcacatg gaaagggtag tcacattttt tgaaatagct cgtgtgtatg 8340
aaggaaatgtt attattttta atttaaatat atgtaaaaat acttaccagt aaatgtgtat 8400
tttaaaagaac tatttaaaac acaatgttat atttcttata aataccagtt actttcgttc 8460
attaattaat gaaaataaat ctgtgaagta cctaatttaa gtactcatac taaaatttat 8520
aaggccgata attttttggt ttcttgtctg taatggaggt aaactttatt ttaaattctg 8580
tgcttcagac aggactattg cttgttgatt tttctagaaa tctgcacggt ataatgaaaa 8640
tattaagaca gtttcccatg taatgtattc cttcttagat tgcacgaaa tgcactatca 8700
tatatgcttg taaatattca aatgaatttg cactaataaa gtcctttgcg gc 8752

```

<210> 25

<211> 868

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7499961CB1

<400> 25

```

tccctcaacc tacttataga ctatttttct tgctctgcag catggaccaa agagaaattc 60
tgcaagaagt cctggatgag gccccaaagca agaaaattac taaagaggag tttgccaatg 120
aatttctgaa gctgaaaagg caatctacca agtacaaggc agacaaaacc taccctacaa 180

```

```

ctgtggctga gaagcccaag aatatcaaga aaaacagata taaggatatt ttgccctatg 240
attatagccg ggtagaacta tccctgataa cctctgatga ggattccagc tacatcaatg 300
ccaacttcct taaggaggtt tatggacca aggcttatat tgccaccag ggtcctttat 360
ctacaacctt cctggacttc tggaggatga tttgggaata tagtgtcctt agtttgaaaa 420
ttttgcgaaa catgaaaaag agtatctgta attcttgccc accaaacaag cctgcagaat 480
ctgttcagtc aaataactcc agctcatttc tgaattttgg ttttgcaaac cgtttttcaa 540
aacccaaagg accaaggaat ccaccaccaa cttggaatat ttaataaaac tccagattta 600
taataatatg ggctgcaagt acacctgcaa ataaaactac tagaatactg ctagttaaaa 660
taagtgtctt atatgcataa tatcaaatat gaagatatgc taatgtgtta atagctttta 720
aaagaaaagc aaaatgccaa taagtgccag ttttgcattt tcatatcatt tgcattgagt 780
tgaaaactgc aaataaaagt ttgtcacttg agcttatgta cagaatgcta tatgagaaac 840
acttttagaa tggattttatt tttcattt

```

<210> 26

<211> 3165

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500309CB1

<400> 26

```

gaggggtgaaa agctcggatc tgtgttttggg gaaggccagg cttgcgctcc tcgccggggtt 60
ccgcgaagggt taaccttggc tgacttggct cgcgagcaaa gggcagcgtc tgagctcccg 120
gcgttccagg agtggcctct tttgtaggag cacctgaaat gcagcgtctg gtgcactaag 180
ccgtagcggc agcagcagcc acagcgacag cgctggggcc ctgtgtagaa gctccatccc 240
ccttgtcttt gtgcttgcct gcgtccccag actcagagat tatcttagaa gacctaggac 300
tccaaaaatg tttccctga aggacgctga aatgggagcc tttaccttct ttgcctcggc 360
tctgccacat gatgtttgtg gaagcaatgg acttctctc acaccaaatt ccatcaaaat 420
tttagggcgc tttcaaactc ttaaaacat caccatccc agactctgcc agtatgtgga 480
tatttctagg ggaaagcatg aacgactagt ggtcgtggct gaacattgtg aacgtagtct 540
ggaagacttg cttcgagaaa ggaaacctgt gagctgttca acggttttgt gtatagcatt 600
tgaggttctt cagggcttgc agtatatgaa caaacatggg atagtacaca gggcattgtc 660
tcttcataat atcctgttgg accgaaaggg acatattaaa ttggctaaat ttggacttta 720
tcacatgaca gctcatggtg atgatgttga tttcccaata gggatccct cgtacttggc 780
ccctgaggta attgcacagg gaattttcaa aacctgat cacatgcaa gtaaaaaacc 840
attgccttct ggcccaaat cagatgtatg gtctcttggg atcattttat ttgagctttg 900
tgtgggaaga aaattatttc agagcttggg tatttctgaa agactaaaat ttttgcttac 960
tttgatttgt gtagatgaca ctttaatagt tctggctgaa gagcatgggt gtttggacat 1020
tataaaggag cttcctgaaa ctgtgataga tcttttgaat aagtgcctta ccttccatcc 1080
ttctaaggag ccaaccccag atgaattaat gaaggacaaa gtattcagtg aggtatcacc 1140
tttatatacc ccttttacca aacctgccag tctgttttca tcttctctga gatgtgctga 1200
tttaactctg cctgaggata tcagtcagtt gtgtaaagat ataaataatg attacctggc 1260
agaaagatct attgaagaag tgtattacct ttggtgtttg gctggagggt acttggagaa 1320
agagcttgtc aacaaggaaa tcattcgatc caaacacct atctgcacac tcccaattt 1380
tctctttgag gatggtgaaa gctttggaca aggtcgagat agaagctcg ttttagatga 1440
taccactgtg acatttgtct tatgccagct aagaaataga ttgaaagat ttggtggaga 1500
agcattttac ccattacttg aagatgacca gtctaattta cctcattcaa acagcaataa 1560
tgagttgtct gcagctgcca cgctcccttt aatcatcaga gagaaggata cacagtacca 1620
actaaataga attattctct tcgacaggct gctaaaggct tatccatata aaaaaacca 1680
aatctggaaa gaagcaagag ttgacattcc tctcttatg agaggtttaa cctgggctgc 1740
tcttctggga gttgagggtg ctattcatgc caagtacgat gcaattgata aagacactcc 1800
aattcctaca gatagacaaa ttgaagtggg tattcctcgc tgtcatcagt acgatgaact 1860
gttatcatca ccagaagggtc atgcaaaatt taggcgtgta ttaaaagcct gggtagtgtc 1920
tcacctgat cttgtgtatt ggcaagggtc tgactcactt tgtgtccat tcctatatct 1980
aaacttcaat aatgaagcct tggcttatgc atgtatgtct gcttttattc ccaaatacct 2040
gtataacttc ttcttaaaag acaactcaca tgtaatacaa gagtatctga ctgtcttctc 2100
tcagatgatt gcatctcatg atccagagct gagtaatcat ctcaatgaga ttggtttcat 2160
tccagatctc tatgccatcc cttgggtttct taccatgttt actcatgtat ttccactaca 2220
caaaattttc cacctctggg ataccttact acttgggaat tctctttcc cattctgtat 2280
tgagtagca attcttcagc agctgcggga ccggcttttg gctaatggct ttaatgagt 2340
tattcttctc ttctccgatt taccagaaat tgacattgaa cgctgtgtga gagaatctat 2400

```

```

caacctgttt tgttggaact ctaaaagtgc tacttacaga cagcatgctc aacctccaaa 2460
gccatcttct gacagcagtg gaggcagaag ttcggcacct tatttctctg ctgagtgtcc 2520
agatcctcca aagacagatc tgtcaagaga atccatccca ttaaatgacc tgaagtcaga 2580
agtatcacca cggatttcag cagaggacct gattgacttg tgtgagctca cagtgcagg 2640
ccacttcaaa acaccagca agaaaacaaa gtccagtaaa ccaaagctcc tgggtggtga 2700
catccggaat agtgaagaga agcatcaaca ttccattcag tgctgccttc actgcagaag 2760
gggagcttac ccagggccct taaactgcta tgctccagaa cttcaaaggg aaggtcattg 2820
tcatcgtggg gcatgtggca aaacacacag ctgagtttgc agctcacctt gtgaagatga 2880
aatatccaag aatctgtatt ctagatgggt gcattaataa aataaagcca acaggcctcc 2940
tcaccatccc atctcctcaa atatgaagaa ccaagagtgt gactgccaaa acttagtgtg 3000
gcatcagcac caacagcaca gttcttcata tccacgccac tctcagacaa aactagatgt 3060
ccagattggt gcatttcctg aagttgtcac gagacatttt ttaaaatctc ataaccaaac 3120
atgttcagtt atccatgcaa gaaactgaca ctactggccc agcac 3165

```

<210> 27

<211> 1630

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2446376CB1

<400> 27

```

ctccccgggca cctaggatgg agcctatggc aagggttttag cacagtgact acaggaatca 60
cagcccagac acaaaagcag gaaacccttt gaccgggctc cttcctattg caccaacagc 120
cttgtgttgc tgcaatgaaa acacttcccc aagcagctgt ggccaagaga cgcagaaact 180
gccttggtcca cgggccccgc ctcagactcc aacactcaca agagagcaga ggagccccc 240
gtcttggggga ccacagaaga tgccatgtgc tccacgatgt cggccccccac ctgcctggcc 300
cacttgccctc cctgcttccct gctgctggca ctgggtccttg tccccctcaga tgcctctggg 360
cagagcagca ggaatgactg gcaggtgcta cagcccaggg gccccatgct ggtggcagaa 420
ggtgagacac ttctactgag gtgtatggtg gtcggctcct gcaactgatg tatgataaaa 480
tgggtgaagg tgagcactca ggaccaacag gaaatttata acttttaaagc tggctccttc 540
cctggggtaa tgcccatgat ccaacggaca tcagaaccac tgaattgtga ttattccatc 600
tataatccaca atgtcaccag ggagcacact ggaacctacc actgtgtgag gtttgatgg 660
ttgagtgaac actcagaaat gaaatcggat gaaggcacct cagtgcctgt gaagggagct 720
ggggaccctg aaccagacct gtggatcatc cagccccagg aattgggtgt ggggaccact 780
ggagacactg tctttctgaa ctgcacagtg cttggagacg gtccccctgg acccatcagg 840
tggttccagg gagctggtct gagccgggag gccatttaca actttggagg catctcccac 900
cccaaggcga cagcgttgca ggcctccaac aatgacttca gcattcttct gcaaaacgtc 960
tccagtggag atgcaggcac ctattactgt gtaaagtttc agaggaaacc caacaggcaa 1020
tacctgtctg gacagggcac cagcctgaaa gtgaaagcaa aatctacctc ttccaaagag 1080
gcagaattca ccagtgaacc tgcaactgag atgtctccaa caggcctcct ggttgtgttc 1140
gcacctgtgg tcttggggct gaaggcaatt accttggtg cactcctact ggccctggct 1200
acctctcgga ggagccctgg gcaagaagat gtcaagacca caggcccagc aggagccatg 1260
aacaccttag catggagcaa ggtcaagag tgaggggtca gcccagagt gaggaccctc 1320
tgagttggag aggagccagg gctcctcaac catttcccta cctccagtc cagcctctag 1380
gtgccccag gcctcatgac aaactcctag atccctacat ctgggttttg tccacctagt 1440
gaaattccct tctttgcacc gggcttccct ctaaaatgtc tccctttctc tttttggcct 1500
gttcaagacc tccttgcttt tcagtcctgg ctcagtcctc cctcaacacc cttgcccctg 1560
ctgcagccct ttctggtgcg ccctgcccct ttccccacct cgctacatcc ttcttggcct 1620
ccaacaacgc 1630

```

<210> 28

<211> 1802

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5562847CB1

<400> 28

```

ctggcccgcca tgtgcaccgt agtggaccct cgcattgtcc ggagatacct actcaggcgg 60
cagctcgggc agggggcccta tggcattgtg tgggaaggcag tggaccggag gactgggtgag 120
gtcgtggcca tcaagaaaaat ctttcatgtc ttttagggata agacagatgc ccagagaaca 180
ttccgggaaa tcacgctcct ccaggagtctt ggggaccatc ccaacatcat cagcctcctt 240
gacgtgatcc gggcagagaa cgacagggac atttacctgg tgtttgagtt tatggacact 300
gacctgaacg cagtcacccg gaagggcggc ctgctgcagg acgtccacgt gcgctccatc 360
ttctaccagc tcctgcgggc caccgggttc ctccactcgg ggcacgttgt gcaccgggac 420
cagaagccgt cccatgtgct cctggatgcc acttgacag tgaagctgtg tgactttggc 480
ctggcccgt ccttgggcca cctccctgag gggcctgagg accaggccgt gacagagtac 540
gtggccacac gctgggtaccg agcaccggag gtgctgtctt cttcgcaccg atacaccctt 600
gggggtggaca tgtggagtct gggctgtatc ctgggggaga tgctgcgggg gagaccctt 660
ttccccggca gtcacaccct ccaccagctg gagctgatcc tggagaccat cccaccgcca 720
tctgaggagg acctcctggc tctcggctca ggctgccgtg cctctgtgct gcaccagctg 780
gggtcccggc cagcacagac gctggatgcc ctccaccgc cagacacctc cccagaggcc 840
ttggacctcc ttaggcgact cctggtgttc gccccggaca agcggttaag cgcgacccag 900
gcactgcagc acccctacgt gcagagggtc cactgcccc cgcacgagtg ggcacgagag 960
gcagatgtgc ggccccgggc acacgaaggg gtccagctct ctgtgcctga gtaccgcagc 1020
cgcgtctatc agatgatcct ggagtgtgga ggcagcagcg gcacctcgag agagaagggc 1080
ccggagggtg tctcccaag ccaggcacac ctgcacaaac ccagagccga cctcagctg 1140
ccttctagga cacctgtgca ggggtcccaga cccaggcccc agagcagccc aggccatgac 1200
cctgccgagc acgagtcccc ccgtgcagcc aagaacgttc ccaggcagaa ctccgctccc 1260
ctgtccaaa ctgctctcct agggaatggg gaaaggcccc ctggggcgaa ggaagcgccc 1320
cccttgacac tctcgttgtt gaagccaagc gggaggggag ctgcgccctc cctgacctcc 1380
caggttgcgg ctcaggtggc caaccaggcc ctgatccggg gtgactggaa ccggggcggt 1440
gggggtgagg tggccagcgt acaacaggtc cctccccggc ttcctccgga ggccccggcc 1500
ggccggagga tgttcagcac ctctgccttg caggtgtccc aggggggtgc cagggctttg 1560
cttgagggtt actcccaagc ctacgggact gtctgccact cggcactggg ccacctgccc 1620
ctgttggagg ggcaccatgt gtgagccgcc ctactccctt cacctggccc tctgttcttg 1680
ccccagcccc ttcccagac cctctccag tctctgcac cccttagccc tccctgcttt 1740
gcctggcccc ttgaagtccc agggagcttg cccgggtctc ctcgggggag cagatgaggg 1800
cc 1802

```

<210> 29

<211> 835

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7493671CB1

<400> 29

```

ctctggatct atcaataaaa gaagtccttc gtgtgggcta catatataga tgttttcatg 60
aagaggagtg aaaagccaga aggatataga caaatgaggc ctaagacctt tcctgccagt 120
aactatactg tcagtagccg gcaaatgtta caagaaattc gggaatccct taggaattta 180
tctaaaccat ctgatgctgc taaggctgag cataacatga gtaaaatgtc aaccgaagat 240
cctcgacaag tcagaaatcc acccaaattt gggacgcata ataaagcctt gcaggaaatt 300
cgaaactctc tgcttccatt tgcaaatgaa acaaattctt ctcgagtagc ttcagaagtt 360
aatccacaaa tgcttcaaga cttgcaagct gctggatttg atgaggatat ggttatacaa 420
gctcttcaga aaactaacia cagaagtata gaagcagcaa ttgaattcat tagtaaaatg 480
agttaccaag atcctcgacg agagcagatg gctgcagcag ctgccagacc tattaatgcc 540
agcatgaaac caggttggtat tatctcaaga tgcccaggat caaatgagaa agatgctttg 600
ccaaaaagaa tctaattaca tccgtcttaa aagggtctaa atggacaagt ctatgtttgt 660
gaagataaag acactaggaa taggagcatt tgggtgaagtc tgtctagcaa gaaaagtaga 720
tactaaggct ttgtatgcaa caaaaactct tcgaaagaaa gatgttcttc ttcgaaatca 780
agtcgctcat gttaaggctg agagagatat cctgggtgaa gctgacaatg aatgg 835

```

<210> 30

<211> 4814

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7499991CB1

<400> 30

```

gctgcgccca gacccacggg cagccggggcg ggcggacgcc tccaagacca gtctctgcct 60
ccgcagcgtc agcgggagcg ggctgcctcc gaagaaagga gaatggcggt cagtaaagga 120
ttccggatct atcacaaatt ggatcccca cctttcagcc tcatagtggg aactaggcat 180
aagggaagaat gtctcatgtt cgagtctggg gctgtcgctg tgctctcatc tgcagaaaaa 240
gaggcaatca agggtagata ctccaaagta ctggatgcat atggactctt aggtgttctg 300
cggttaaatc ttggtgatac tatgttacat tatctggtcc tagtcactgg atgtatgtct 360
gttggaaaaa ttcaagaatc tgaagttttc cgagttactt ccactgagtt tatatcactg 420
cgaatcgatt cttcagatga ggatcgcatc tcagaagtgc ggaaagtgtt gaattcagga 480
aacttttatt ttgcatgggtc tgcactgtgc atcagtttag atttgagtct taatgcgcat 540
cgtagcatgc aagaacagac aactgataat agatttttct ggaatcagtc tttgcatttg 600
catctcaaac actatggcgt gaattgtgat gactggttat tacgtcttat gtgtggagga 660
gtagaaatca gaacaattta tgctgctcat aaacaggcga aggcttgccct catttcaaga 720
ttaagctgtg aacgagctgg gaccagggtt aatgtccggg gaacaaatga tgatgggtcat 780
gttgccaatt ttgtagaaac agaacagggt gtgtacttag atgactcagt ttcttccttc 840
atacaaatcc gaggatctgt tccattgttc tgggagcaac cagggttgca agtgggatct 900
catcgtgtcc gtatgtcaag gggatttgaa gccaatgcac ctgcttttga caggcatttt 960
agaacactta agaacttata tggtaacaa ataatagtaa atttgcttgg atctaaggaa 1020
ggtgaacata tgctaagtaa agctttccag agtcatttga aagcttctga acatgctgct 1080
gatatccaga tgggtgaattt tgactatcat caaatgggta agggaggaaa ggcagaaaaa 1140
ttacatagtg ttcttaaac tcaagtccag aagtttctag attatggatt tttttatttc 1200
aatggaagtg aagttcaaag atgccagagt ggtacagttc gaacaaactg cttggattgt 1260
cttgatagaa caaatagtgt gcaggcattt cttggcttag agatgctagc taaacagttg 1320
gaagctcttg gtttagctga aaagcctcag ttggtgactc gctttcaaga agtttttcgg 1380
tcaatgtggg ccgtgaatgg cgattcaatc agtaagatat atgcaggaac tggagctctt 1440
gaagggaaag cgaagttaaa agatggtgct cgctctgtta cccgaacaat tcagaataac 1500
ttctttgaca gctccaagca agaggccatt gatgttttgc tactgggaaa tactctgaat 1560
agtgtattag ctgacaaagc tcgagcactt ttaactactg gaagtttgcg tgtttctgag 1620
cagacattac agtcagcatc ttctaaagta ctaaagagca tgtgtgagaa tttctacaaa 1680
tattcaaagc ctaagaaaaat tcgagtatgt gtcggaacct ggaatgtgaa tgggtgggaa 1740
caatttcgca gcatagcttt taagaatcag acactcactg actggcttct tgatgcacc 1800
aagttagctg gcatccagga gtttcaagat aaaagaagta agccaactga tatatttgca 1860
attggttttg aagaaatggg agaattgaat gctggaaaca ttgtgaatgc aagcacaaca 1920
aatcagaagc tctgggctgt agaacttcag aagacaatct ccagagacaa caagtatgtg 1980
ctgctggctt ctgaacagtt ggtgggctgc tgtttgtttg tttttatcag accacagcat 2040
gctcctttta tcagggatgt tgcagttgat actgtgaaga ctggaatggg aggtgcaact 2100
ggaaataagg gagcagttgc aatccgaatg ctcttccata caaccagcct ttgcttcgtc 2160
tgtagccact ttgctgcagg gcagtcacaa gtcaaagaaa gaaatgaaga ttttatagaa 2220
atagcacgaa aattgagttt tcctatggga aggatgctat ttcccatga ctatgtattt 2280
tggtgtgggtg atttcaacta tcgaatcgat ctccctaacg aagaagttaa agagctcata 2340
agacagcaaa attgggattc tcttatagca ggagatcaac ttatcaatca gaaaaatgct 2400
ggacaggttt ttagaggatt tttagaagga aaggtaacct ttgctccgac atataagtat 2460
gacttgtttt ctgacgacta tgacaccagt gaaaagtgc gcacctgc ctggacagac 2520
cgtgtccttt ggagaaggag gaaatggcct tttgatagat cagctgaaga tctagatctt 2580
ctaaatgcta gttttcaaga tgaaagcaaa attctgtaca cgtggactcc aggcactttg 2640
ctgcaactatg gaagagctga gctgaagact tctgaccaca ggcctgtcgt tgccctgatt 2700
gatatagata tatttgaagt tgaagctgaa gagaggcaaa acatttataa agaagtaatt 2760
gcagttcagg gtccaccaga tggtagagta ttggtctcaa tcaaaagttc tttaccagaa 2820
aataattttt ttgatgatgc cttgattgat gagcttctgc agcagtttgc aagttttggg 2880
gaagttatac ttataagatt tgtagaagat aaaatgtggg ttacattttt ggagggaagc 2940
tctgccttga atgttctgag cctaaatggg aaagagttat tgaatcggac tataactatt 3000
gctttaaaaa gtccagactg gatcaaaaat ttggaagaag aaatgagttt agagaaaatt 3060
agcattgcat tgccatcatc aacaagctct accctgcttg gtgaagatgc agaggttgca 3120
gcagattttg atatggaagg tgatgttgat gactatagtg ctgaagtgga ggaacttctt 3180
cctcagcatc tccagccatc ttcaagttcc ggcctatggta ctccccag ctcttcacc 3240
cgaactagtc cctgccagtc acctacaata tcagaggggc ctgtaccttc ccttcccatc 3300
agaccaagcc gagcacgctc aagaactcct gggcctccca gtgcacagag ttctcctatt 3360
gacgcgcagc cagcaacgcc gctgccgcag aaagaccccg cccagccctt ggagcccaag 3420
cggccgcgcg cgcgccgcgc ggtcgccctt cccacacgcc cggctcccc acagagacct 3480
cctccgcctt caggacgcag tcagccttca cctcaagcag gacttgcagg cccaggacct 3540

```

```

gctggatata gtacagccag accgacgatt cctcctcgtg ctggagttat cagtgcacca 3600
cagagccacg cgcgggcatc tgctggaaga ctgactcctg aaagccaaag caaaacatca 3660
gaaacgtcga aaggttcaac tttccttcct gaaccactga agcctcaggc tgcttttctt 3720
ccgcagtctt ctttgccccc gcctgctcaa aggttgcaag agcctcttgt ccctgtggca 3780
gcacctatgc ctgagtctgg cccccagcca aatttggaag cccaccaca accaccacct 3840
cgaagcaggt catcccatag ctgaccttca gaagcttcct cacaaccgca acaggagcaa 3900
ccatcagggg aacaggtaaa aacaaatgga atctctgatg gcaaaagaga atcaccatta 3960
aagattgacc catttgaaga tctgtcattt aatctgcttg ctgtatcaaa ggctcagcta 4020
tctgttcaaa cgtcacctgt tcccacccca gacccaaaga ggttgattca gttgccttct 4080
gcaacgcaaa gtaatgttaa tactttgagt tctgtaagtt gcatgccaac aatgcctcca 4140
attccagctc ggagtcaatc ccaggaaaat atgcgaagtt ctccaaaccc atttattact 4200
ggcttgacca ggacaaatcc tttcagtgac aggactgctg ctcttggaag cccatttaga 4260
gccaaagtct aagaatcaga ggcaacttca tgggttctcca aagaagagcc cgttactatc 4320
agtccttttc cttctctgca gcctcttggt cataacaaaa gcagggcttc atcttcaact 4380
gatggcttta aggacagttt tgatctacag ggccagtcta cattaataat tagcaacccg 4440
aaaggatggg taaccttcga ggaagaagag gattttggtg tgaaagggaa gtcaaagtca 4500
gcttggttcag acttactggg taatcagcca agtttcattt ctggctccaa cctgacattg 4560
aatgatgact ggaataaaagg tacaatgttc tcttctgtg tgttgccgtc aagaagacct 4620
cctccacctc ctgtccctct gctcccgccc ggcaccagcc ctccagtaga tcctttcacg 4680
accttgacct ctaaggcttc acccacactg gactttacag aaagataacg ccatgcaata 4740
gaaaacagtg ggtacttgct tttggcagga tagagctaag agaattgggc attagtattt 4800
cattatgtgc aata 4814

```

<210> 31

<211> 1076

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7499977CB1

<400> 31

```

ctgcccgtca ctccccttgt cacctcttgg agccccctcc taaccagcgg ccagtggggt 60
tcccataccc caggatgtga gcctctttaa cctgtaatgc tgtggctagc ccttggcccc 120
tttcctgcca tggagaacca ggtgctggta attcgcatca agatcccaa tagtggcgcg 180
gtggactgga cagtgcactc cgggcccgcag ttactcttca gggatgtgct ggatgtgata 240
ggccagggtt tgcctgaagc aacaactaca gcatttgaat atgaagatga agatgggtgat 300
cgaattacag tgagaagtga tgaggaaatg aaggcaatgc tgcatatta ttattccaca 360
gtaatggaac agcaagtaaa tggacagtta atagagcctc tgcagatatt tccaagagcc 420
tgcaagcctc ctggggaacg gaacatacat ggccctgaagc ttctgcagtg cattgttgat 480
gaggattcgc cgtccttcc agttggagag ttctcggagc catttgtaga tttcatcact 540
cagtgtatgc gaaaacagcc aaaagaaagg ccagcacctg aagaattgat gggccacccg 600
ttcatcgtgc agttcaatga tggaaatgcc gccgtggtgt ccatgtgggt gtgccgggcg 660
ctggaggaga ggccggagcca gcagggggccc ccgtgaggct gccgcagggc actgaaagcc 720
caggaccagt aaccaaggag aacaacccac ccgtgcctct tctccgtatg ctgcctgcgc 780
cagaagagct ttgctgggcc ctggcttccc tgccctcgcc ttcacctctg tcagcaggtg 840
gccttgccctg gggagcccca tgtgtggccc accccaccag gccatcccca taccttcttg 900
tttgaaggcg ctgacactgg cagagaggta aaggggtggg cattgagaat ggaggctccc 960
agggctccctg cccacttctg ttttcctaata gttttctct ataaagggtc agggccgtca 1020
gcatcactga tgggaataaa agtattaatg ctttgtgaca aaaaaaaaaa aaaaaa 1076

```

<210> 32

<211> 4683

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 8043337CB1

<400> 32

```

cttagtccgc attcgctcca ggttttggga ccctaggttg cggagtcctt accctaccct 60

```

```

ggcctctcga gcagttgtcc ccataactcg gaatctagag cgcgtgttgc gaggcaggag 120
cacgtggcag tcaagtagct tcccagtcce gaacgccgcc cgtccccacc ccgccgtggc 180
cactagcaac gacctctgtg aagttggaga ggcggtaacg gaggcactcc cctgctgca 240
ccccgccgtt tctacggggc tcagaaacca gtttgtttgt ttctgcgggg tagtgtcgac 300
ctgtcttacg ggcgtcgccc gagacaggac ggagtcaaac cctgtgtatc aactgaagac 360
gagtgtcagg tgtggagagt ctacgtgccc cctttcagtc tggactgtga gctgctgctg 420
gttagacagt cttggtttct ctttcaggat gtcattttca aaatgcggga tggtagctct 480
gctttattaa gccccgtagg aagactgcc aacactagac gatgcttatt agtcatcacc 540
gttattccta ctaacgtcct gtgtcactga gtttttttaa tgtctagcat atctgtaaag 600
atgccttaga aaaagaatca tggagaagta tgttagacta cagaagattg gagaaggttc 660
atltggaaaa gccattcttg ttaaatctac agaagatggc agacagtatg ttatcaagga 720
aattaacatc tcaagaatgt ccagtaaaga aagagaagaa tcaaggagag aagttgcagt 780
attggcaaac atgaagcatc caaatattgt ccagtataga gaatcatttg aagaaaatgg 840
ctctctctac atagtaatgg attactgtga gggaggggat ctgtttaagc gaataaatgc 900
tcagaaaggc gttttgtttc aagaggatca gattttggac tggtttgtac agatagtgtt 960
ggccctgaaa catgtacatg atagaaaaat tcttcacatga gacattaaat ctacagaacat 1020
atlttttaact aaagatggaa cagtacaact tggagatttt ggaattgcta gatttcttaa 1080
tagtactgta gagctggctc gaacttgcac agggacccca tactacttgt cacctgaaat 1140
ctgtgaaaac aaaccttata ataataaaag tgacatttgg gctctggggg gtgtccttta 1200
tgagctgtgt acacttaaac atgcttttga agctggcagt atgaaaaacc tggtagtgaa 1260
gataatatct ggatcttttc cacctgtgtc tttgcattat tcctatgatc tccgcagttt 1320
ggtgtctcag ttatttaaaa gaaatcctag ggatagacca tcagtcaact ccatattgga 1380
gaaaggtttt atagccaaac gcattgaaaa gtttctctct cctcagctta ttgcagaaga 1440
atlttgtcta aaaacatttt cgaagtttgg atcacagcct ataccagcta aaagaccagc 1500
ttcaggacaa aactcgattt ctgttatgcc tgctcagaaa attacaaagc ctgccgctaa 1560
atatggaata ccttttagcat ataagaaata tggagataaa aaattacacg aaaagaaacc 1620
actgcaaaaa cataaacagg cccatcaaac tccagagaag agagtgaata ctggagtaaga 1680
aaggaggaaa atatctgagg aagcagcaag aaagagaagg ctggaattta ttgaaaaaga 1740
aaagaaacaa aaggatcaga ttattagtgt aatgaaggct gaacaaatga aaaggcaaga 1800
aaaggaaagg ttggaaagaa taaatagggc cagggaacaa ggatggagaa atgtgctaag 1860
tgctggtgga agtgggtgaag taaaggctcc ttttctgggc agtggatgga ctatagctcc 1920
atcatctttt tcttctcgag gacagtatga acattaccat gccatttttg accaaatgca 1980
gcaacaaaga gcagaagata atgaagctaa atggaaaaga gaaatatatg gtcgaggtct 2040
tccagaaagg caaaaagggc agctagctgt agtaagagct aaacaagtag aagagttcct 2100
gcagcgaaaa cgggaagcta tgcagaataa agctcgagcc gaaggacata tggtttatct 2160
ggcaagactg aggcacaata gactacagaa tttcaatgag cgccaacaga ttaaagccaa 2220
acttcgtggg gaaaagaaag aagctaatac ttctgaagga caagaaggaa gtgaagaggc 2280
tgacatgagg cgcaaaaaaa tccaatcact gaaggcccat gcaaatgcac gtgctgctgt 2340
actaaaagaa caactagaac gaaagagaaa ggaggcttat gagagagaaa aaaaagtgtg 2400
ggaagagcat ttggtggcta aaggagttaa gagttctgat gtttctccac ctttgggaca 2460
gcatgaaaca ggtggctctc catcaaagca acagatgaga tctgttattt ctgtaacttc 2520
agctttgaaa gaagttggcg tggacagtat ttttaactgat acccgggaaa cttcagaaga 2580
gatgcaaaaag accaacaatg ctatttcaag taagcgagaa atacttcgta gattaaatga 2640
aaatcttaaa gctcaagaag atgaaaaagg aaagcagaat ctctctgata cttttgagat 2700
aaatgttcat gaagatgcca aagagcatga aaaagaaaaa tcagtttcat ctgatcgcaa 2760
gaagtgggag gcaggagggtc aacttgtgat tctcttggtg gagttaacac tagatacatc 2820
cttctctaca actgaaagac atacagtggg agaagttatt aaattagggt ctaatggatc 2880
tccaagaaga gcctggggga aaagtccgac agattctgtt ctaaagatac ttggagaagc 2940
tgaactacaa cttcagacag aactattaga aaatacaact attagaagtg agatttctcc 3000
cgaaggggaa aagtacaaac ccttaattac tggagaaaaa aaagtacaat gtatttcaca 3060
tgaaataaac ccatcagcta ttgttgatc tctgttgag acaaaaagtc ccgagttcag 3120
tgaggcatct ccacagatgt cattgaaact ggaaggaaat ttagaagaac ctgatgattt 3180
ggaaacagaa attctacaag agccaagtgg aacaaacaaa gatgagagct tgccatgcac 3240
tattactgat gtgtggatta gtgaggaaaa agaaacaaag gaaactcagt cggcagatag 3300
gatcaccatt caggaaaatg aagtttctga agatggagtc tcgagtactg tggaccaact 3360
tagtgacatt catatagagc ctggaaccaa tgattctcag cactctaaat gtgatgtaga 3420
taagtctgtg caaccgggaa catttttcca taaggtgggt cattctgaac acttgaactt 3480
agtccctcaa gttcaatcag ttcagtgttc accagaagaa tcttttgcat ttcatctca 3540
ctcgcattta ccaccaaaaa ataaaaacaa gaattccttg ctgattggac tttcaactgg 3600
tctgtttgat gcaaacaccc caaagatgtt aaggacatgt tcaattccag atctctcaaa 3660
gctgttcaga acccttatgg atgttcccac cgtaggagat gttcgtcaag acaatcttga 3720
aatagatgaa attgaagatg aaaacattaa agaaggacct tctgattctg aagacattgt 3780
gtttgaagaa actgacacag atttacaaga gctgcaggcc tcgatggaac agttacttag 3840

```

```

ggaacaacct ggtgaagaat acagtgaaga agaagagtca gtcttgaaga acagtgatgt 3900
ggagccaact gcaaatggga cagatgtggc agatgaagat gacaatccca gcagtgaag 3960
tgccctgaac gaagaatggc actcagataa cagtgatggg gaaattgcta gtgaatgtga 4020
atgcgtagt gtctttaacc atttagagga actgagactt catctggagc aggaaatggg 4080
ctttgaaaaa ttctttgagg tttatgagaa aataaaggct attcatgaag atgaagatga 4140
aaatattgaa atttgttcaa aaatagttca aaatatattg ggaaatgaac atcagcatct 4200
ttatgccaaag attcttcatt tagtcatggc agatggagcc taccaagaag ataatgatga 4260
ataatcctca aaatgttttt taatcctcaa ctatatgaaa gcatttgaat ttggcttatc 4320
agaataacaa gcttcagtgg gaaatacagc aattatattt ttaaaaaatc agatttaaga 4380
tggacttttct tattgcatga aaaagatgga gaaacatgcc atttttcaat gaagattcta 4440
atattttatc tattttgttc attgaattcc atggttaaat ctcataaaat atatacttta 4500
ttaaatcatc caaccaagc ataggaaaca ttgaccaga acctgactta atgggttttga 4560
agatttacta tgcaataggg taaactttga gtctcagcaa atgtcttttag gttgaaggaa 4620
ttacctatgt catgaaggac ctgtctgtgg tttttcaatg gagtctttaa gcatgatctt 4680
ttt 4683

```

<210> 33

<211> 3129

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5947007CB1

<400> 33

```

ctcggatcac tagtaacggc gcagtgtgtg gaaaggcgcc tccgttggaa acttgggctg 60
agtaccgagg cgggcgcgag cgaggcgccc tagacatctt ctccctcccg ttgcctcagg 120
ggcttgggtt cggaccagac ctatgtgcag tctgccccat catgctctta agactgttga 180
agattcttac ttttgggtcat agggctggta aaagtgcaga tttattgcta aacatgggtg 240
catttttggg taaacccaaa actgaaaaac ataatgctca tgggtgctggg aatggtttac 300
gttatggcct gagcagcatg caaggatgga gagtggaaat ggaagatgca cacacagctg 360
ttgtagggtat tcctcacggc ttggaagact ggtcattttt tgcagtttat gatggtcag 420
ctggatcccc agtggcaaat tactgctcaa cacatttatt agaacacatc actactaacg 480
aagacttttag ggcagctgga aaatcaggat ctgctcttga gctttcagtg gaaaatgtta 540
agaatgggtat cagaactgga tttttgaaaa ttgatgaata catgcgtaac ttttcagacc 600
tcagaaaagg gatggacagg agtgggttcaa ctgcagtggg agttatgatt tcacctaagc 660
atatctactt tatcaactgt ggtgattcac gtgctgttct gtataggaat ggacaagtct 720
gcttttctac ccaggatcac aaaccttgca atccaaggga aaaggagcga atccaaaatg 780
caggaggcag cgtgatgata caacgtgtta atgggttcatt agcagtatct cgtgctctgg 840
gggactatga ttacaagtgt gttgatggca agggcccaac agaacaactt gtttctccag 900
agcctgaggt ttatgaaatt ttaagagcag aagaggatga atttatcatc ttggcttgtg 960
atgggatctg ggatgttatg agtaatgagg agctctgtga atatgttaaa tctaggcttg 1020
aggtatctga tgacctggaa aatgtgtgca attgggtagt ggacacttgt ttacacaagg 1080
gaagtgcaga taacatgagt attgtactag tttgcttttc aaatgctccc aaggctcag 1140
atgaagcggg gaaaaaagat tcagagttgg ataagcactt ggaatcacgg gttgaagaga 1200
ttatggagaa gtctggcgag gaaggaatgc ctgatcttgc ccatgtcatg cgcacttgt 1260
ctgcagaaaa tatcccaaatt ttgcctcctg ggggaggtct tgctggcaag cgtaatgtta 1320
ttgaagctgt ttatagtaga ctgaatccac atagagaaag tgatgggggt gctggagatc 1380
tagaagacct atggtagcct taaaaacctt ctaaaatgct tttgattctg aaaattgggg 1440
gaaaaaactt ttaatcacia ttttcttcaa tacaagggga aaatattctt gcggattccc 1500
aacgttttgt gatatgagca gaaaatcatt agcatttccc atcatttgtt catatttgtg 1560
ttttctgaca gttgccactt gtagcattgc ctgtactaca gtattttttg ccaacctcag 1620
gcatactcgt tacatctgta ttgaactttc ggccctagaa accagtggag ttatttcacc 1680
acaaatcaac aatgtgcctg aggtgcatgg gaaatatagt tagctatact ctgaaaatac 1740
attatgtttt ttttctttta acaaaacaca caacatgtaa gcatgtaaga gtaaagaatt 1800
gtatgatatg ttcctttttt cagttcacca agttggaagc cttttgcagc tctgtggctt 1860
ggaatttcat ttgagcaatt tctataggat atgtatttat tattgattgt tatttaattt 1920
ttttccaatt ttacctgtat taccaaactg ggttctccaa taatgtccaa attgtaattg 1980
tgcttgcctt caagataaag tgtatttggg aataatatta taaacctta caaattttat 2040
gcatgtatct actgcatcct tcaactctca ctagaaaatc ttttgaaacc aaatggatta 2100
atztatggct atttataatt tgctttgaca tctcactgtt ggaaattttt taaagatgag 2160
atttgccttt ataatgtaaa ttgtgatttt tgttttacat gtgggtttct atagttttta 2220

```

```

ttttttcagc ttttaagata cgagttttgt gtaatttggg atttttaatc atttatgtta 2280
ttttaaaagc tcagaatatc acattgaaat tactataaat acatttaaaa ttatctat 2340
tagatctaag gaaatactac agagatat  tcatgggttc agtaactttt cattttataa 2400
cattgggcac ggtacagagt gattgtcaca taagggtact gaagatttat tagtttaatt 2460
ctatttttac agtaaccttg aattcttctg agttttgcat gtattaaatt caattaatgc 2520
tgaacatgaa gagtaaagta tttatctgaa agaagtttct ggggttaggag aagtaatgaa 2580
tgtatccatt tgtacatggg ttacatgttg tggatgcttt gtaaaccatt tcctgtatgt 2640
ttaaattgtg tttcagcagg atgtaattgc cttgtgtgtg agttaaaatg agtcatcatc 2700
tggtcctttg tgaaatggaa ttcatgggat tttctgtaac gttttcctga agctgtttct 2760
ggagagccac acatttaaat acagacagct ttctgatca tttgatttat tgtgcacctg 2820
atttttgggc taaaaggaat tattgccaca atatatttta tttattcttt agatttttagc 2880
cttgttaagtt aaagtgcctt acatgatgat gtgaaaagct gtttgtccct ttactgggtt 2940
tgggggggtt ttaaaagata gggaatgaag aatgcaaaat ggtttatcgt tcaaactgtc 3000
cactctgata caaccctgta ctgatagtag tcccagtag gatattgtga tgtttcatac 3060
aatgcagtga acataaccaa cttgttacct aaataaagaa ttgataaaaa cagtgtgaca 3120
tattaaaaa 3129

```

<210> 34

<211> 2580

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510576CB1

<400> 34

```

agattgcccc tgacttgatt tggctgacct gcctagaaat attatgttga ataatgatga 60
gttggaaatt gaacaagctc cagagttttc tcctaggtga tggcagtttt ggatcagttt 120
accgagcagc ctatgaagga gaagaagtgg ctgtgaagat ttttaataaa catacatcac 180
tcaggctgtt aagacaagag cttgtgggtgc tttgccacct ccaccacccc agtttgatat 240
ctttgctggc agctgggatt cgtccccgga tgttgggtgat ggagttagcc tccaagggtt 300
ccttggatcg cctgcttcag caggacaaag ccagcctcac tagaacccta cagcacagga 360
ttgcaactca cgtagctgat ggtttgagat acctccactc agccatgatt atataccgag 420
acctgaaacc ccacaatgtg ctgcttttca cactgtatcc caatgctgcc atcattgcaa 480
agattgctga ctacggcatt gctcagtagt gctgtagaat ggggataaaa acatcagagg 540
gcacaccagg gtttcgtgca cctgaagttg ccagaggaaa tgtcatttat aaccaacagg 600
ctgatgttta ttcatattgg ttactactct atgacatttt gacaactgga ggtagaatag 660
tagagggttt gaagtttcca aatgagtttg atgaattaga aatacaagga aaattacctg 720
gtctttgaca ttttgaattc agctgaatta gtctgtctga cgagacgcat tttattacct 780
aaaaacgtaa ttgttgaatg catggttgct acacatcaca acagcaggaa tgcaagcatt 840
tggctgggct gtgggcacac cgacagagga cagctctcat ttcttgactt aaatactgaa 900
ggatacactt ctgaggaagt tgctgatagt agaataattg gcttagcctt ggtgcatctt 960
cctgttgaaa aggaaagctg gatttgtgtc gggacacagt ctgggtactc cctgggtcatc 1020
aataccgaag atgggaaaaa gagacatacc ctagaaaaga tgactgattc tgtcacttgt 1080
ttgtattgca attccttttc caagcaaagc aaacaaaaaa attttctttt ggttggaaac 1140
gctgatggca agttagcaat ttttgaagat aagactgtta agcttaaagg agctgctcct 1200
ttgaagatac taaatatagg aaatgtcagt actccattga tgtgtttgag tgaatccaca 1260
aattcaacgg aaagaaatgt aatgtgggga ggaatgtggc caaagatttt ctccttttct 1320
aatgatttca ccattcagaa actcattgag acaagaacaa gccaaactgt ttcttatgca 1380
gctttcagtg attccaacat cataacagtg gtggtagaca ctgctctcta tattgctaag 1440
caaaatagcc ctgttgtgga agtgtgggat aagaaaactg aaaaactctg tggactaata 1500
gactgcgtgc actttttaag ggaggtaacg gtaaaagaaa acaaggaatc aaaacacaaa 1560
atgtcttatt ctgggagagt gaaaaccctc tgccttcaga agaactgc tctttggata 1620
ggaactggag gaggccatat tttactcctg gatctttcaa ctcgctgact tatacgtgta 1680
atttacaact tttgtaattc ggtcagagtc atgatgacag cacagctagg aagccttaaa 1740
aatgtcatgc tggatttggg ctacaaccgg aaaaatactg aaggtagaca aaagcagaaa 1800
gagatacaat cttgcttgac cgtttgggac atcaatcttc cacatgaagt gcaaaattta 1860
gaaaaacaca ttgaagttag aaaagaatta gctgaaaaaa tgagacgaac atctgttgag 1920
taagagagaa ataggaattg tctttggata ggaaaattat tctctcctct tgtaaatatt 1980
tattttaaaa atgttcacat ggaaagggtg ctcacatttt ttgaaatagc tctgtgtgat 2040
gaaggaatgt tattattttt aatttaataa tatgtaaaaa tacttaccag taaatgtgta 2100
ttttaaagaa ctattttaaa cacaatgtta tatttcttat aaataccagt tactttcgtt 2160

```

```

cattaattaa tgaaaataaa tctgtgaagt acctaattta agtactcata ctaaaattta 2220
taaggccgat aattttttgt tttcttgtct gtaatggagg taaactttat tttaaattct 2280
gtgcttcaga caggactatt gcttggtgat ttttctagaa atctgcacgg tataatgaaa 2340
atattaagac agtttcccat gtaatgtatt ccttccttaga ttgcatcgaa atgcactatc 2400
atatatgctt gtaaataattc aaatgaattt gcactaataa agtcctttgt tggatgtga 2460
attctctttg ttgctgttgc aaacagtgc tcttacacaa cttcactcaa ttcaaaagaa 2520
aactccatta aaagtactaa tgaaaaaaca tgacatactg tcaaagtcct catatctagg 2580

```

<210> 35

<211> 1301

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 90111008CB1

<400> 35

```

gttttttagca cagtgactac aggaatcaca gccagacac aaaagcagga aaccctttga 60
ccgggctcct tcctattgca ccaacagcct tgtgttgctg caatgaaaac acttcccca 120
gcagctgtgg ccaagagacg cagaaactgc cttgtccacg ggccccgcct cagactccaa 180
cactcacaag agagcagagg agccccaaagt cttgggggacc acagaagatg ccatgtgctc 240
cacgatgtcg gccccacct gcctggccca cttgcctccc tgcttctctg tgctggcact 300
ggtccttgct ccctcagatg cctctgggca gagcagcagg aatgactggc aggtgctaca 360
gcccgagggc cccatgctgg tggcagaagg agctggggac cctgaaccag acctgtggat 420
catccagccc caggaattgg tgttggggac cactggagac actgtcttct tgaactgcac 480
agtgtcttga gacgggtccc ctggacccat caggtggttc cagggagctg gtctgagccg 540
ggaggccatt tacaactttg gaggcattct ccacccaag gcgacagcgg tgcaggcctc 600
caacaatgac ttcagcattc ttctgcaaaa cgtctccagt gaggatgcag gcacctatta 660
ctgtgtaaaag tttcagagga aacccaacag gcaataacctg tctggacagg gcaccagcct 720
gaaagtgaiaa gcaaaatcta cctcttccaa agaggcagaa ttcaccagtg aacctgcaac 780
tgagatgtct ccaacaggcc tcctggttgt gttcgcacct gtggtcctgg ggctgaaggc 840
aattaccttg gctgcactcc tactggccct ggctacctct cggaggagcc cttgggcaaga 900
agatgtcaag accacaggcc cagcaggagc catgaacacc ttagcatgga gcaagggtca 960
agagtgaggg gtacgccccg gagtgaggac cctctgagtt ggagaggagc cagggtcctc 1020
caaccatttc cctacctcca gtcccagcct ctagggtgcc ccaggcctca tgacaaactc 1080
ctagatccct acatctggtt ttggtccacc tagtgaaatt cccttctttg caccgggctt 1140
ccctctaaaaa tgtctccctt tctctttttg gcctgttcaa gacctccttg cttttcagtc 1200
cctggctcag tctctctca acacccttgc ccctgttgca gccctttctg gtgcgccctg 1260
cccctttccc cactctgcta catccttctt ggctccaaa c 1301

```

<210> 36

<211> 1675

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503305CB1

<400> 36

```

ggacactgag tggaaggcaa aatggcggcg gcggcggcgg tggcctggtg ttaaggggag 60
agccagggtc ttacgacccc tgggacgggc cgcgctggcc cgcggcagcc cccccgttcg 120
tctccccgct ctgccccacc agggatactt ggggttgctg ggacggactc tggccgcctc 180
agcgtccgcc ctacggcccc tggccgctgt ccaggagctc tgctctcccc tccagagtta 240
attatttata ttgtaaagaa ttttaacagt cctggggact tccttgaagc atcattttca 300
cttttgctca gaagaaagct ctggatctat caaataaaga agtccttcgt gtgggctaca 360
tatatagatg ttttcatgaa gaggagtga aagccagaag gatatagaca aatgaggcct 420
aagacctttc ctgccagtaa ctatactgtc agtagccggc aaatgttaca agaaattcgg 480
gaatccctta ggaatttatc taaaccatct gatgctgcta aggtgagca taacatgagt 540
aaaatgtcaa ccgaagatcc tcgacaagtc agaaatccac ccaaatttgg gacgcatcat 600
aaagccttgc aggaatttgc aaactctctg cttccatttg caaatgaaac aaattcttct 660
cggagtactt cagaagttaa tccacaaatg cttcaagact tgcaagctgc tggattttag 720

```

```

gaggtgacca tccacggcaa gatagcatgg atttcagtaa tgaatggggg gatccctcaa 780
gctgtcgatg tggagacaga ctgaagccat tagagcggag agctgcacgc cagcaccagc 840
gatgtctagc acattctttg gttgggactc ccaattatat tgcacctgaa gtgttgctac 900
gaacaggata cacacagttg tgtgattggg ggagtgttgg tgttattctt tttgaaatgt 960
tggtgggaca acctcctttc ttggcacaaa caccattaga aacacaaatg aaggtaaggt 1020
tatcaactgg caaacatctc ttcacattcc accacaagct aaactcagtc ctgaagcttc 1080
tgatcttatt attaaacttt gccgaggacc cgaagatcgc ttaggcaaga atgggtgctga 1140
tgaaataaaa gctcatccat tttttaaaac aattgatttc tccagtgacc tgagacagca 1200
gtctgcttca tacattccta aaatcacaca cccaacagat acatcaaatt ttgatcctgt 1260
tgatcctgat aaattatgga gtgatgataa cgaggaagaa aatgtaaata acactctcaa 1320
tggatgggat aaaaatggaa agcatcctga acatgcattc tatgaattta ccttcggaag 1380
gttttttgat gacaatggct acccatataa ttatccgaag cctattgaat atgaatacat 1440
taattcacia ggctcagagc agcagtcgga tgaagatgat caaacacag gctcagagat 1500
taaaaatcgc gatctagtat atgtttaaca cactagtaaa taaatgtaat gaggatttgt 1560
aaaagggcct gaaatgcgag gtgttttgag gttctgagag taaaattatg caaatatgac 1620
agagctatat atgtgtgctc tgtgtacaat attttatttt cctaaattat gggaa 1675

```