

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
23 December 2004 (23.12.2004)

PCT

(10) International Publication Number
WO 2004/111202 A2

- (51) International Patent Classification⁷: **C12N** (US). BAUGHN, Mariah, R. [US/US]; 8547 W. Horner Street, Los Angeles, CA 90035 (US).
- (21) International Application Number: PCT/US2004/006969 (74) Agent: SIMKIN, Michele, M.; Foley & Lardner LLP, Washington Harbour, 300K Street N.W., Suite 500, Washington, DC 20007-5101 (US).
- (22) International Filing Date: 5 March 2004 (05.03.2004)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/453,108 7 March 2003 (07.03.2003) US
60/471,930 19 May 2003 (19.05.2003) US
60/490,749 28 July 2003 (28.07.2003) US
- (71) Applicant (for all designated States except US): INCYTE CORPORATION [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).
- (71) Applicant and
- (72) Inventor: RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): EMERLING, Brooke, M. [US/US]; 833 W. Buena Avenue, #2108, Chicago, IL 60613 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). BECHA, Shanya, D. [US/US]; 1 Saint Francis #5508, San Francisco, CA 94107 (US). CHAWLA, Narinder, K. [US/US]; 33 Union Square, #712, Union City, CA 94587
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, 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, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (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: NEUROTRANSMISSION-ASSOCIATION PROTEINS

(57) Abstract: Various embodiments of the invention provide human neurotransmission-associated proteins (NTRAN) and polynucleotides which identify and encode NTRAN. 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 NTRAN.

NEUROTRANSMISSION-ASSOCIATED PROTEINS

TECHNICAL FIELD

The invention relates to novel nucleic acids, neurotransmission-associated proteins encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of autoimmune/inflammatory, cardiovascular, neurological, developmental, cell proliferative, transport, psychiatric, metabolic, and endocrine disorders. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and neurotransmission-associated proteins.

BACKGROUND OF THE INVENTION

The human nervous system, which regulates all bodily functions, is composed of the central nervous system (CNS), consisting of the brain and spinal cord, and the peripheral nervous system (PNS), consisting of afferent neural pathways for conducting nerve impulses from sensory organs to the CNS, and efferent neural pathways for conducting motor impulses from the CNS to effector organs. The PNS can be further divided into the somatic nervous system, which regulates voluntary motor activity such as for skeletal muscle, and the autonomic nervous system, which regulates involuntary motor activity for internal organs such as the heart, lungs, and viscera. CNS-associated proteins function in neuronal signaling, cell adhesion, nerve regeneration, axon guidance, neurogenesis, and other processes.

The cerebral cortex or higher brain is the largest structure, consisting of a right and a left hemisphere interconnected by the corpus callosum. The cerebral cortex is involved in sensory, motor, and integrative functions related to perception, voluntary musculoskeletal movements, and the broad range of activities associated with consciousness, language, emotions, and memory. The cerebrum functions in association with the lower centers of the nervous system. The lower areas of the brain such as the medulla, pons, mesencephalon, cerebellum, basal ganglia, substantia nigra, hypothalamus, and thalamus control unconscious activities including arterial pressure and respiration, equilibrium, and feeding reflexes, such as salivation.

The central nervous system (CNS) is composed of more than 100 billion neurons at the spinal cord level, the lower brain level, and the higher brain or cortical level. Neurons transmit electric or chemical signals between cells. The spinal cord, a thin, tubular extension of the central nervous system within the bony spinal canal, contains ascending sensory and descending motor pathways, and is covered by membranes continuous with those of the brainstem and cerebral hemispheres. The spinal cord contains almost the entire motor output and sensory input systems of the trunk and limbs,

and neuronal circuits in the cord also control rhythmic movements, such as walking, and a variety of reflexes. The lower areas of the brain such as the medulla, pons, mesencephalon, cerebellum, basal ganglia, substantia nigra, hypothalamus, and thalamus control unconscious activities including arterial pressure and respiration, equilibrium, and feeding reflexes, such as salivation. Emotions, such as anger, excitement, sexual response, and reaction to pain or pleasure, originate in the lower brain. The cerebral cortex or higher brain is the largest structure, consisting of a right and a left hemisphere interconnected by the corpus callosum. The cerebral cortex is involved in sensory, motor, and integrative functions related to perception, voluntary musculoskeletal movements, and the broad range of activities associated with consciousness, language, emotions, and memory. The cerebrum functions in association with the lower centers of the nervous system.

Nervous system organization and development

A nerve cell (neuron) contains four regions, the cell body, axon, dendrites, and axon terminal. The cell body contains the nucleus and other organelles. The dendrites are processes which extend outward from the cell body and receive signals from sense organs or from the axons of other neurons. These signals are converted to electrical impulses and transmitted to the cell body. The axon, whose size can range from one millimeter to more than one meter, is a single process that conducts the nerve impulse away from the cell body. Cytoskeletal fibers, including microtubules and neurofilaments, run the length of the axon and function in transporting proteins, membrane vesicles, and other macromolecules from the cell body along the axon to the axon terminal. Some axons are surrounded by a myelin sheath made up of membranes from either an oligodendrocyte cell (CNS) or a Schwann cell (PNS). Myelinated axons conduct electrical impulses faster than unmyelinated ones of the same diameter. The axon terminal is at the tip of the axon away from the cell body. (See Lodish, H. et al. (1986) Molecular Cell Biology Scientific American Books New York NY, pp. 715-719.)

CNS-associated proteins have roles in neuronal signaling, cell adhesion, nerve regeneration, axon guidance, neurogenesis, and other functions. Certain CNS-associated proteins form an integral part of a membrane or are attached to a membrane. For example, neural membrane protein 35 (NMP35) is closely associated with neuronal membranes and is known to be highly expressed in the rat adult nervous system (Schweitzer, B. et al. (1998) *Mol. Cell. Neurosci.* 11:260-273). Synaptophysin (SY) is a major integral membrane protein of small synaptic vesicles. The chromosomal location of SY in human and mouse is on the X chromosome in subbands Xp11.22-p11.23. This region has been implicated in several inherited diseases including Wiskott-Aldrich syndrome, three forms of X-linked hypercalciuric nephrolithiasis, and the eye disorders retinitis pigmentosa 2, congenital stationary night blindness, and Aland Island eye disease (Fisher, S.E. et al. (1997) *Genomics* 45:340-347). Peripherin, or retinal degeneration slow protein (rds), is an integral

membrane glycoprotein that is present in the rims of photoreceptor outer segment disks. In mammals, rds is thought to stabilize the disk rim through heterophilic interactions with related nonglycosylated proteins. Rds is a mouse neurological mutation that is characterized by abnormal development of rod and cone photoreceptors followed by their slow degeneration (Kedzierski, W.J. et al. (1999)

5 Neurochem. 72:430-438).

Each of over a trillion neurons in adult humans connects with over a thousand target cells (Tessier-Lavigne, M. et al. (1996) Science 274:1123-1133). These neuronal connections form during embryonic development. Each differentiating neuron sends out an axon tipped at the leading edge by a growth cone. Aided by molecular guidance cues, the growth cone migrates through the embryonic
10 environment to its synaptic target. Progressive axon outgrowth occurs during neural development but not in the mature mammalian CNS. Following CNS injury, expression of growth-inhibiting molecules is enhanced while availability of their growth-promoting counterparts diminishes. Proteins governing developmental axon guidance contribute to the failure of injured central neurons to regenerate. These proteins include Semaphorin3A and the Semaphorin3A receptor proteins neuropilin-1 and plexin-A1
15 (Pasterkamp, R.J. and J. Verhaagen (2001) Brain Res. Brain Res. Rev. 35:36-54).

Semaphorins function during embryogenesis by providing local signals to specify territories inaccessible to growing axons (Puschel, A.W. et al. (1995) Neuron 14:941-948). They consist of at least 30 different members and are found in vertebrates, invertebrates, and even certain viruses. All semaphorins contain the sema domain which is approximately 500 amino acids in length. Neuropilin, a
20 semaphorin receptor, has been shown to promote neurite outgrowth *in vitro*. The extracellular region of neuropilins consists of three different domains: CUB, discoidin, and MAM domains. The CUB and the MAM motifs of neuropilin have been suggested to have roles in protein-protein interactions and are thought to be involved in the binding of semaphorins through the sema and the C-terminal domains (reviewed in Raper, J.A. (2000) Curr. Opin. Neurobiol. 10:88-94).

25 The guidance of axons during development involves both positive and negative effects (i.e., chemoattraction and chemorepulsion). The Slit family of proteins have been implicated in promoting axon branching, elongation, and repulsion. Members of the Slit family have been identified in a variety of organisms, including insects, amphibians, birds, rodents and humans (Guthrie, S. (1999) Current Biology 9:R432-R435). Slit proteins are ligands for the repulsive guidance receptor, Roundabout (Robo); however, Slit proteins also cause elongation in some assays. A post-translationally processed
30 form of Slit appears to be the active form of the protein (Guthrie, S. *supra*; and Brose, K. et al. (1999) Cell 96:795-806).

Axon growth is also guided in part by contact-mediated mechanisms involving cell surface and extracellular matrix (ECM) molecules. Many ECM molecules, including fibronectin, vitronectin,

members of the laminin, tenascin, collagen, and thrombospondin families, and a variety of proteoglycans, can act either as promoters or inhibitors of neurite outgrowth and extension (Tessier-Lavigne et al., *supra*). Receptors for ECM molecules include integrins, immunoglobulin superfamily members, and proteoglycans. ECM molecules and their receptors have also been implicated in the adhesion, maintenance, and differentiation of neurons (Reichardt, L.F. et al. (1991) *Ann. Rev. Neurosci.* 14:531-571). The proteoglycan testican is localized to the post-synaptic area of pyramidal cells of the hippocampus and may play roles in receptor activity, neuromodulation, synaptic plasticity, and neurotransmission (Bonnet, F. et al. (1996) *J. Biol. Chem.* 271:4373-4380).

Neurotrophins regulate development, maintenance, and function of vertebrate nervous systems. Neurotrophins activate two different classes of receptors, the Trk family of receptor tyrosine kinases and p75NTR, a member of the TNF receptor superfamily. Through these receptors, neurotrophins activate many signaling pathways, including those mediated by ras and members of the cdc-42/ras/rho G protein families, and by the MAP kinase, PI-3 kinase, and Jun kinase cascades. During development, limiting amounts of neurotrophins function as survival factors to ensure a match between the number of surviving neurons and the requirement for appropriate target innervation. They also regulate cell fate decisions, axon growth, dendrite pruning, the patterning of innervation, and the expression of proteins crucial for normal neuronal function, such as neurotransmitters and ion channels. These proteins also regulate many aspects of neural function. In the mature nervous system, they control synaptic function and synaptic plasticity, while continuing to modulate neuronal survival (Huang, E.J. and L.F. Reichardt (2001) *Ann. Rev. Neurosci.* 24:677-736). Neuritin is a protein induced by neural activity and by neurotrophins which promote neuritogenesis.

Neurexins are a large family of neuronal cell-surface proteins that function as cell adhesion molecules during synaptogenesis and in intercellular signaling and formation of intercellular junctions (Clarris, H.J. et al. (2002) *Int. J. Dev. Biol.* 46:649-652; Rowen, L. et al. (2002) *Genomics* 79:587-597). Neurexin IV is essential for axonal insulation in the PNS in embryos and larvae. Axonal insulation is of key importance for the proper propagation of action potentials. Caspr, a vertebrate homolog of Neurexin IV -- also named paranodin -- is found in septate-like junctional structures localized to the paranodal region of the nodes of Ranvier, between axons and Schwann cells. Caspr/paranodin is implicated in blood-brain barrier formation, and linkage of neuronal membrane components with the axonal cytoskeletal network (Bellen, H.J. et al. (1998) *Trends Neurosci.* 21:444-449).

The neurexophilins are neuropeptide-like proteins which are proteolytically processed after synthesis. They are ligands for the neuron-specific cell surface proteins, the α -neurexins. Neurexophilins and neurexins may participate in a neuron signaling pathway (Missler, M. and T.C.

Sudhof (1998) *J. Neurosci.* 18:3630-3638; Missler, M. et al. (1998) *J. Biol. Chem.* 273:34716-34723).
 Ninjurin is a neuron cell surface protein which plays a role in cell adhesion and in nerve regeneration
 following injury. Ninjurin is up-regulated after nerve injury in dorsal root ganglion neurons and in
 Schwann cells (Araki, T. and J. Milbrandt (1996) *Neuron* 17:353-361). Ninjurin2 is expressed in
 5 mature sensory and enteric neurons and promotes neurite outgrowth. Ninjurin2 is upregulated in
 Schwann cells surrounding the distal segment of injured nerve with a time course similar to that of
 ninjurin1, neural CAM, and L1 (Araki, T. and J. Milbrandt (2000) *J. Neurosci.* 20:187-195).

Mammalian Numb is a phosphotyrosine-binding (PTB) domain-containing protein which may
 be involved in cortical neurogenesis and cell fate decisions in the mammalian nervous system.

10 Numb's binding partner, the LNX protein, contains four PDZ domains and a ring finger domain and
 may participate in a signaling pathway involving Numb. PDZ domains have been found in proteins
 which act as adaptors in the assembly of multifunctional protein complexes involved in signaling events
 at surfaces of cell membranes (Ponting, C.P. (1997) *Bioessays* 19:469-479). LNX contains a tyrosine
 phosphorylation site which may be important for the binding of other PTB-containing proteins such as
 15 SHC, an adaptor protein which associates with tyrosine-phosphorylated growth factor receptors and
 downstream effectors (Dho, S.E. et al. (1998) *J. Biol. Chem.* 273:9179-9187).

Nogo has been identified as a component of the central nervous system myelin that prevents
 axonal regeneration in adult vertebrates. Cleavage of the Nogo-66 receptor and other
 glycoposphatidylinositol-linked proteins from axonal surfaces renders neurons insensitive to Nogo-66,
 20 facilitating potential recovery from CNS damage (Fournier, A.E. et al. (2001) *Nature* 409:341-346).

Homeobox transcription factors direct nerve-cell associated tissue patterning and
 differentiation. The presence and function of these proteins appears to be ubiquitous in nematodes,
 arthropods, and vertebrates. One example of these proteins is DRG11, a homeobox transcription
 factor expressed in mammalian sensory neurons, and which appears to be involved in neural crest
 25 development (Saito, T. et al. (1995) *Mol. Cell Neurosci.* 6:280-292). Cutaneous sensory neurons that
 detect noxious stimuli project to the dorsal horn of the spinal cord, while those innervating muscle
 stretch receptors project to the ventral horn. DRG11 is required for the formation of spatio-temporally
 appropriate projections from nociceptive sensory neurons to their central targets in the dorsal horn of
 the spinal cord (Chen, Z.F. et al. (2001) *Neuron* 31:59-73).

30 Synapses

Contact between one neuron and another occurs at a specialized site called the synapse.
 Many nervous system functions are regulated by diverse synaptic proteins such as synaptophysin, the
 synapsins, growth associated protein 43 (GAP-43), SV-2, and p65, which are distributed in subcellular
 compartments of the synapse. Synaptic terminals also contain many other proteins involved in calcium

transport, neurotransmission, signaling, growth, and plasticity. At this site, the axon terminal from one neuron (the presynaptic cell) sends a signal to another neuron (the postsynaptic cell). Synapses may be connected either electrically or chemically. An electrical synapse consists of gap junctions connecting the two neurons, allowing electrical impulses to pass directly from the presynaptic to the postsynaptic cell. In a chemical synapse, the axon terminal of the presynaptic cell contains membrane vesicles containing a particular neurotransmitter molecule. A change in electrical potential at the nerve terminal results in the influx of calcium ions through voltage-gated channels which triggers the release of the neurotransmitter from the synaptic vesicle by exocytosis. The neurotransmitter rapidly diffuses across the synaptic cleft separating the presynaptic nerve cell from the postsynaptic cell. The neurotransmitter then binds receptors and opens transmitter-gated ion channels located in the plasma membrane of the postsynaptic cell, provoking a change in the cell's electrical potential. This change in membrane potential of the postsynaptic cell may serve either to excite or inhibit further transmission of the nerve impulse.

Presynaptic calcium channel activity is modulated by cysteine-string proteins (CSPs). CSPs are secretory vesicle proteins that function in neurotransmission as well as in exocytosis in other cell-types. CSPs belong to the DnaJ/hsp40 (heat shock protein) chaperone family. The effect of CSPs on calcium levels is likely to be downstream of calcium release and is likely to involve exocytosis, possibly in connection with G-proteins (Braun, J.E. et al. (1995) *Neuropharmacology* 34:1361-9136; Magga, J.M. et al. (2000) *Neuron* 28:195-204; Dawson-Scully, K. et al. (2000) *J. Neurosci.* 20:6039-6047; and Chamberlain, L.H. et al. (2001) *J. Cell Sci.* 114:445-455). Neuregulins (NRGs) mediate between the electrical neural activity and molecular components by regulating the expression of ion channel receptors or transmitter release in synapses. NRGs may also be signaling factors involved in tuning locomotion or other higher functions by coordinating excitatory and inhibitory neurons (Ozaki, M. (2001) *Neuroscientist* 7:146-154).

N- and P/Q-type Ca^{2+} channels are localized in high density in presynaptic nerve terminals and are crucial elements in neuronal excitation-secretion coupling. In addition to mediating Ca^{2+} entry to initiate transmitter release, they are thought to interact directly with proteins of the synaptic vesicle docking/fusion machinery. N-type and P/Q-type Ca^{2+} channels are colocalized with syntaxin in high-density clusters in nerve terminals. The synaptic protein interaction (synprint) sites in the intracellular loop II-III (LII-III) of both alpha 1B and alpha 1A subunits of N-type and P/Q-type Ca^{2+} channels bind to syntaxin, SNAP-25, and synaptotagmin. Presynaptic Ca^{2+} channels not only provide the Ca^{2+} signal required by the exocytotic machinery, but also contain structural elements that are integral to vesicle docking, priming, and fusion processes (Catterall, W.A. (1999) *Ann. NY Acad. Sci.* 868:144-159). Synaptotagmins are a large family of proteins involved in both regulated and

constitutive vesicular trafficking. They include a neuronal type (synaptotagmin I-V, X, and XI) and a ubiquitous type (synaptotagmin VI-IX). Ca^{2+} -dependent synaptotagmin activation is involved in neurite outgrowth (Mikoshiba, K. et al. (1999) Chem. Phys. Lipids 98:59-67).

Proteins associated with the membranes of synaptic vesicles include vamp (synaptobrevin),
5 rab3A, synaptophysin, synaptotagmin (p65) and SV2. These membrane proteins function in regulated exocytosis by regulating neurotransmitter uptake, vesicle targeting, and fusion with the presynaptic plasma membrane (Elferink, L.A. and R.H. Scheller (1993) J. Cell Sci. Suppl.17:75-79).

Physophilin, also known as the Ac39 subunit of the V-ATPase, is an oligomeric protein that binds the synaptic vesicle protein synaptophysin, constituting a complex that may form the exocytotic
10 fusion pore. Ac39 is present in a synaptosomal complex which, in addition to synaptophysin, includes the bulk of synaptobrevin II, and subunits c and Ac115 of the V0 sector of the V-ATPase. *In situ* hybridization in rat brain reveals a largely neuronal distribution of Ac39/physophilin mRNA which correlates spatio-temporally with those of subunit c and synaptophysin. Immunohistochemical analysis shows that Ac39/physophilin is mostly concentrated in the neuropil with a pattern identical to subunit A
15 and very similar to synaptophysin. Double-labeling immunofluorescence shows a complete colocalization of Ac39/physophilin with subunit A and a partial colocalization with synaptophysin in the neuropil (Carrion-Vazquez M. et al. (1998) Eur. J. Neurosci.10:1153-1166).

The plasma membrane dopamine transporter (DAT) is essential for the reuptake of released dopamine from the synapse. Uptake of dopamine is temperature- and time-dependent, and is inhibited
20 by a variety of compounds, such as cocaine. DAT-knockout mice have been shown to exhibit extreme hyperactivity and resistance to both cocaine and amphetamine, consistent with the primary action of cocaine on DAT (Giros, B. et al. (1996) Nature 379:606-612). The perturbation of the tightly regulated DAT also predisposes neurons to damage by a variety of insults. Most notable is the selective degeneration of DAT-expressing dopamine nerve terminals in the striatum thought to
25 underlie Parkinson's disease. DAT expression can predict the selective vulnerability of neuronal populations, which suggests that therapeutic strategies aimed at altering DAT function could have significant benefits in a variety of disorders (Gary, W.M. et al. (1999) Trends Pharmacol. Sci. 20:424-429).

43 KD postsynaptic protein or acetylcholine receptor-associated 43 KD protein (RAPSIN) is
30 thought to play a role in anchoring or stabilizing the nicotinic acetylcholine receptor at synaptic sites. RAPSIN is involved in membrane association and may link the nicotinic acetylcholine receptor to the underlying postsynaptic cytoskeleton (Buckel, A. et al. (1996) Genomics 35:613-616). Neuritin is a protein whose gene is known to be induced by neural activity and by neurotrophins which promote neuritogenesis. Neuraxin is a structural protein of the rat central nervous system that is believed to be

immunologically related to microtubule-associated protein 5 (MAP5). Neuraxin is a novel type of neuron-specific protein which is characterized by an unusual amino acid composition, 12 central heptadecarepeats and putative protein and membrane interaction sites. The gene encoding neuraxin is unique in the haploid rat genome and is conserved in higher vertebrates. Neuraxin is implicated in neuronal membrane-microtubule interactions and is expressed throughout the rodent CNS (Rienitz, A. et al. (1989) EMBO J. 8:2879-2888).

Neurotransmitters and neurotransmitter transport proteins

Neurotransmitters comprise a diverse group of some 30 small molecules which include acetylcholine, monoamines such as serotonin, dopamine, and histamine, and amino acids such as gamma-aminobutyric acid (GABA), glutamate, and aspartate, and neuropeptides such as endorphins and enkephalins (McCance, K.L. and S.E. Huether (1994) PATHOPHYSIOLOGY, The Biologic Basis for Disease in Adults and Children, 2nd edition, Mosby, St. Louis, MO, pp. 403-404). Many of these molecules have more than one function and the effects may be excitatory, e.g. to depolarize the postsynaptic cell plasma membrane and stimulate nerve impulse transmission, or inhibitory, e.g. to hyperpolarize the plasma membrane and inhibit nerve impulse transmission.

Neurotransmitters and their receptors are targets of pharmacological agents aimed at controlling neurological function. For example, GABA is the major inhibitory neurotransmitter in the CNS, and GABA receptors are the principal target of sedatives such as benzodiazepines and barbiturates which act by enhancing GABA-mediated effects (Katzung, B.G. (1995) Basic and Clinical Pharmacology, 6th edition, Appleton & Lange, Norwalk, CT, pp. 338-339).

Two major classes of neurotransmitter transporters are essential to the function of the nervous system. The first class is uptake carriers in the plasma membrane of neurons and glial cells, which pump neurotransmitters from the extracellular space into the cell. This process relies on the Na^+ gradient across the plasma membrane, particularly the co-transport of Na^+ . Two families of proteins have been identified. One family includes the transporters for GABA, monoamines such as noradrenaline, dopamine, and serotonin, and amino acids such as glycine and proline. Common structural components include twelve putative transmembrane α -helical domains, cytoplasmic N- and C- termini, and a large glycosylated extracellular loop separating transmembrane domains three and four. This family of homologous proteins derives their energy from the co-transport of Na^+ and Cl^- ions with the neurotransmitter into the cell (Na^+/Cl^- neurotransmitter transporters). The second family includes transporters for excitatory amino acids such as glutamate. Common structural components include 6-10 putative transmembrane domains, cytoplasmic N- and C- termini, and glycosylations in the extracellular loops. The excitatory amino acid transporters are not dependent on Cl^- , and may require intracellular K^+ ions (Na^+/K^+ neurotransmitter transporters) (Liu, Y. et al. (1999) Trends Cell

Biol. 9:356-363).

The second class of neurotransmitter transporters is present in the vesicle membrane, and concentrates neurotransmitters from the cytoplasm into the vesicle, before exocytosis of the vesicular contents during synaptic transmission. Vesicular transport uses the electrochemical gradient across the vesicular membrane generated by a H⁺-ATPase. Two families of proteins are involved in the transport of neurotransmitters into vesicles. One family uses primarily proton exchange to drive transport into secretory vesicles and includes the transporters for monoamines and acetylcholine. For example, the monoamine transporters exchange two luminal protons for each molecule of cytoplasmic transmitter. The second family includes the GABA transporter, which relies on the positive charge inside synaptic vesicles. The two classes of vesicular transporters show no sequence similarity to each other and have structures distinct from those of the plasma membrane carriers (Schloss, P. et al. (1994) Curr. Opin. Cell Biol. 6:595-599; Liu et al., *supra*).

GABA is the predominant inhibitory neurotransmitter and is widely distributed in the mammalian nervous system. GABA is cleared from the synaptic cleft by specific, high-affinity, Na⁺- and Cl⁻- dependent transporters, which are thought to be localized to both pre- and postsynaptic neurons, as well as to surrounding glial cells. At least four GABA transporters (GAT1–GAT4) have been cloned (Liu, Q.-R. et al. (1993) J. Biol. Chem. 268:2106–2112). Studies of [³H]-GABA uptake into cultured cells and plasma-membrane vesicles isolated from various tissues revealed considerable differences in GABA transporter heterogeneity. GABA transporters exhibit differences in substrate affinity and specificity, distinct blocker pharmacologies, and different tissue localization. For example, the K_m values of GABA uptake of the expressed GAT1 to GAT4 are 6, 79, 18, and 0.8 mM, respectively. In addition to transporting GABA, GAT2 also transports betaine; GAT3 and GAT4 also transport β-alanine and taurine. Pharmacological studies revealed that GABA transport by GAT1 and GAT4 is more sensitive to 2,4-diaminobutyric acid and guavicine than that by GAT2 and GAT3. *In situ* hybridization showed that GAT1 and GAT4 expression is brain specific. GAT2 and GAT3 mRNAs were detected in tissues such as liver and kidney (Schloss et al., *supra*; Borden, L.A. (1996) Neurochem. Int. 29:335-356; Nelson, N. (1998) J. Neurochem. 71:1785-1803).

Human studies indicated that GABA transporter function is reduced in epileptic hippocampi. Decreased GABAergic neurotransmission has also been implicated in the pathophysiology of schizophrenia (Simpson, M.D. et al. (1992) Psychiatry Res. 42:273-282).

Diazepam binding inhibitor (DBI), also known as endozepine and acyl-Coenzyme (CoA)-binding protein, is an endogenous GABA receptor ligand which is thought to down-regulate the effects of GABA. DBI binds medium- and long-chain acyl-CoA esters with very high affinity and may function as an intracellular carrier of acyl-CoA esters (*125950 Diazepam Binding Inhibitor; DBI,

Online Mendelian Inheritance in Man (OMIM); PROSITE PDOC00686 Acyl-CoA-binding protein signature).

Glycine serves as one of the major inhibitory neurotransmitters in the mammalian nervous system by activating chloride-channel receptors, which are members of a ligand-gated ion-channel superfamily (Betz, H. (1990) Neuron 5:383-392). Glycine also facilitates excitatory transmission through an allosteric activation of the N-methyl-D-aspartate (NMDA) receptor (Johnson, J.W. and P. Ascher (1987) Nature 325:529-531). Forms of glycine transporter include GLYT 1 and GLYT 2. Variants of GLYT1 (GLYT1 a/b) are generated by alternative splicing (Liu, Q.-R. et al. (1993) J. Biol. Chem. 268:22802-22808). GLYT1a is transcribed in both neural and non-neural tissues, whereas GLYT1b was detected only in neural tissues (Borowsky, B. et al. (1993) Neuron 10:851-863). High levels of GLYT1a/b mRNA were found in hippocampus and cortex, implying its involvement in the regulation of excitatory synaptic transmission. It is not clear whether GLYT1a is expressed in neurons, in glia or in both. In contrast, GLYT1b is found almost exclusively in fiber tracts, suggesting its localization in glial cells (Schloss et al., *supra*). GLYT2 is expressed mainly in brainstem and spinal cord (Schloss et al., *supra*).

The second identified glycine transporter (GLYT2) differs from GLYT1a/b by its extended intracellular amino terminus. The predominant localization of its mRNA in brainstem and spinal cord and its insensitivity to N-methyl-aminoacetic acid suggests that GLYT2 terminates signal transduction at the strychnine-sensitive inhibitory glycine receptor. It has been proposed that, upon depolarization of cells harboring GLYT1b, the transporter runs backwards and releases glycine to act as a neuromodulatory amino acid at the NMDA receptor (Attwell, D. and M. Bouvier (1992) Curr. Biol. 2:541-543). Such a Ca^{2+} -independent, non-vesicular release of neurotransmitters by reverse transport was demonstrated for glutamate and serotonin. This evidence suggests that the transmitter transporters may be important for both the initiation and termination of neurotransmitter action (Schloss et al., *supra*).

Creatine transporters are strongly related to transporters for GABA. The primary sequence identity between creatine transporter species homologs is very high (98-99%). Pharmacological characterization demonstrated high affinity creatine uptake (27-43 mM), which was blocked by creatine analogs with high affinity. Creatine transporters are widely expressed in a variety of mammalian tissues, including brain, adrenal gland, intestine, colon, prostate, thymus, ovary, spleen, pancreas, placenta, umbilical cord, thyroid, tongue, pharynx, vertebral discs, jaw, and nasal epithelium. Genetic mapping in the mouse localizes the creatine transporter to a region on the X chromosome in linkage conservation with the human region Xq28, the location of the genes for several neuromuscular diseases (Nash, S.R. et al. (1994) Receptors Channels 2:165-174).

The substrates of a number of cDNA clones encoding proteins of the Na^+/Cl^- -dependent transporter families are still not identified. These are orphan transporters. Identification of the substrates for orphan transporters has been difficult because *in situ* hybridization and immunohistochemistry indicate that the transporters are synthesized by phenotypically different neuronal populations, for example glutaminergic, GABAergic, histaminergic, or serotonergic neurons. One of the transporters, NTT4, exhibits the highest homology to the creatine transporter. It differs structurally from other members of this family in having an unusually long loop between transmembranes seven and eight (Liu, Q.-R. et al. (1993) FEBS Lett. 315:114-118; Schloss et al., *supra*).

Glutamate is a major excitatory neurotransmitter in the mammalian central nervous system. Electrogenic (Na^+/K^+)-coupled glutamate transporters, located in the plasma membranes of nerve terminals and glial cells, mediate removal of glutamate released at excitatory synapses and maintain extracellular concentrations below neurotoxic levels. Glutamate transporters achieve this process by co-transport with three sodium ions and one proton, followed by translocation of a potassium ion in the opposite direction (Zerangue, N. and M.P. Kavanaugh (1996) Nature 383:634-637).

The membrane topology of the glutamate transporters reveals six membrane-spanning helices in the N-terminal part of the proteins (Slotboom, D.J. et al. (1999) Microbiol. Mol. Biol. Rev. 63:293-307). The C-terminal half of the glutamate transporters is well conserved and constitutes a major part of the translocation pathway and contains the binding sites for the substrate and co-transported ions (Zhang, Y. and B.I. Kanner (1999) Proc. Natl. Acad. Sci. USA 96:1710-1715).

Impaired re-uptake of synaptic glutamate, and a reduced expression of glutamate transporters have been found in the motor cortex of patients with amyotrophic lateral sclerosis (ALS). Inhibition of the synthesis of each glutamate transporter subtype using chronic antisense oligonucleotide administration, *in vitro* and *in vivo*, selectively and specifically reduced the protein expression and function of glutamate transporters. The loss of glial glutamate transporters produced elevated extracellular glutamate levels, neurodegeneration characteristic of excitotoxicity, and a progressive paralysis. The loss of the neuronal glutamate transporter did not elevate extracellular glutamate in the striatum but produced mild neurotoxicity and resulted in epilepsy (Rothstein, J.D. et al. (1996) Neuron 16:675-686).

The vesicular monoamine transporters (VMAT) package cytoplasmic monoamine neurotransmitters into secretory vesicles for regulated exocytotic release. VMAT acts as an electrogenic exchanger of protons and monoamines, using a proton electrochemical gradient. VMAT transporters include VMAT1 and VMAT2. The VMAT proteins possess twelve transmembrane segments, with both extremities lying on the cytoplasmic side. VMAT proteins are associated with

distinct vesicle populations in neurons and neuroendocrine cells (Henry, J.-P. et al. (1994) *J. Exp. Biol.* 196:251-262).

Vesicular transport is inhibited by the antihypertensive drug reserpine and the related but more centrally acting drug tetrabenazine. The mechanism of transport and the biochemistry of VMAT have
5 been analyzed with these drugs, using mainly the chromaffin granules from bovine adrenal glands as a source of transporters (Peter, D. et al. (1994) *J. Biol. Chem.* 269:7231-7237).

Human studies indicated that reserpine can cause a syndrome resembling depression, indicating the importance of vesicular transport activity for the control of mood and behavior. The psychostimulant amphetamine also disrupts the storage of amines in secretory vesicles, further
10 indicating that alterations in vesicular monoamine transport can affect behavior (Sulzer, D. and S. Rayport (1990) *Neuron* 5:797-808).

Another family of molecules that appear to be important for neurotransmission are the choline-transporter-like CTL1 proteins. The prototypic CTL1 was identified in yeast as a suppressor of a choline transport mutation; however, mammalian homologues have been identified. The proteins
15 comprise approximately ten putative transmembrane domains in addition to transporter-like motifs but do not appear to be canonical choline transporters. Choline transport is important to neurotransmission because choline is a precursor of acetylcholine, required in abundance by cholinergic neurons (O'Regan, S. et al. (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97:1835-1840).

Neuronal signals are transmitted across the neuromuscular junction (NMJ). Motor axons
20 release the molecule agrin to induce the formation of the postsynaptic apparatus in muscle fibers. Proteins such as dystroglycan, MuSK, and rapsyn participate in the transduction of agrin signals. Agrin also functions in the upregulation of gene transcription in myonuclei and the control of presynaptic differentiation (Ruegg, M.A. and J.L. Bixby (1998) *Trends Neurosci.* 21:22-27).

Neurological protein domains

25 CNS-associated proteins can be phosphoproteins. For example, ARPP-21 (cyclic AMP-regulated phosphoprotein) is a cytosolic neuronal phosphoprotein that is highly enriched in the striatum and in other dopaminoceptive regions of the brain. The steady-state level of ARPP-21 mRNA is developmentally regulated. But, in the neonatal and mature animal, ARPP-21 mRNA is not altered following 6-hydroxydopamine lesions of the substantia nigra or by pharmacologic treatments
30 that upregulate the D1- or D2-dopamine receptors (Ehrlich, M.E. et al. (1991) *Neurochem.* 57:1985-1991).

CNS-associated signaling proteins may contain PDZ domains. PDZ domains have been found in proteins which act as adaptors in the assembly of multifunctional protein complexes involved in signaling events at surfaces of cell membranes. PDZ domains are generally found in membrane-

associated proteins including neuronal nitric oxide synthase (NOS) and several dystrophin-associated proteins (Ponting, C.P. et al. (1997) Bioessays 19:469-479). PSD-95/SAP90 is a membrane-associated guanylate kinase found in neuronal cells at the postsynaptic density (PSD) (Takeuchi, M. et al. (1997) J. Biol. Chem. 272:11943-11951). PSD-95/SAP90 contains three PDZ domains, one SH3
 5 domain, and one guanylate kinase domain. The PDZ domains mediate interactions with NMDA receptors, Shaker-type potassium channels, and brain nitric oxide synthase. SAPAPs (SAP90/PSD-95-Associated Proteins) promote localization of PSD-95/SAP90 at the plasma membrane.

CNS-associated proteins may also contain epidermal growth factor (EGF) domains. The Notch proteins are transmembrane proteins which contain extracellular regions of repeated EGF
 10 domains. Notch proteins, such as the *Drosophila melanogaster* neurogenic protein Notch, are generally involved in the inhibition of developmental processes. Other members of the Notch family are the lin-12 and glp-1 genes of *Caenorhabditis elegans*. Genetic studies indicate that the lin-12 and glp-1 proteins act as receptors in specific developmental cell interactions which may be involved in certain embryonic defects (Tax, F. E. et al. (1994) Nature 368:150-154). Pecanex, a
 15 maternal-effect neurogenic locus of *D. melanogaster*, is believed to encode a large transmembrane protein. In the absence of maternal expression of the pecanex gene, an embryo develops severe hyperneuralization similar to that characteristic of Notch mutant embryos (LaBonne, S. G. et al. (1989) Dev. Biol. 136:1-116).

Other CNS-associated signaling proteins contain WW domains. The WW domain is a protein
 20 motif with two highly conserved tryptophans. It is present in a number of signaling and regulatory proteins, including Huntingtin interacting protein. Several fibroblast growth factor (FGF) homologous factors (i.e., FHF polypeptides) have also been implicated in nervous system development based on mRNA expression patterns in mouse and human tissues. Members of the FHF family of polypeptides are structurally distinct from prototypic FGFs, consistent with the unusual role of these FGF-related
 25 proteins (Smallwood, P.M. et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:9850-9857 and Hartung, H. et al. (1997) Mech. Dev. 64:31-39).

Disorders associated with neurological processes

Alzheimer's disease (AD) is a degenerative disorder of the CNS which causes progressive memory loss and cognitive decline during mid to late adult life. AD is characterized by a wide range
 30 of neuropathologic features including amyloid deposits and intra-neuronal neurofibrillary tangles. Although the pathogenic pathway leading to neurodegeneration and AD is not well understood, at least three genetic loci that confer genetic susceptibility to the disease have been identified (Schellenberg, G.D. (1995) Proc. Natl. Acad. Sci. 92:8552-8559; Sherrington, R. et al. (1995) Nature 375:754-760).

Familial British dementia (FBD), is an autosomal dominant disease featuring amyloid plaques

surrounded by astrocytes and microglia, neurofibrillary tangles, neuronal loss, and progressive dementia. The BRI gene on chromosome 13 encodes a 4 kD peptide, A-Bri. This membrane-anchored protein is a primary constituent of amyloid deposits, and its presence in lesions from the CNS of FBD patients may be a contributive factor of this disease (El-Agnaf, O.M.A. et al. (2001)

5 Biochemistry 40:3449-3457).

Astrocytomas, and the more malignant glioblastomas, are the most common primary tumors of the brain, accounting for over 65% of primary brain tumors. These tumors arise in glial cells of the astrocyte lineage. Following infection by pathogens, astrocytes function as antigen-presenting cells and modulate the activity of lymphocytes and macrophages. Astrocytomas constitutively express

10 many cytokines and interleukins that are normally produced only after infection by a pathogen (de Micco, C. (1989) J. Neuroimmunol. 25:93-108). In the course of identifying genes related to astrocyte differentiation, one cDNA was isolated from an astrocytoma cDNA library that encodes a protein structurally related to the plant pathogenesis-related (PR) proteins (Murphy, E.V. et al. (1995) Gene 159:131-135). The glioma pathogenesis-related protein (GliPR) is highly expressed in glioblastoma,

15 but not in fetal or adult brain, or in other nervous system tumors. PR proteins are a family of small (10-20 kDa), protease resistant proteins induced in plants by viral infections, such as tobacco mosaic virus. The synthesis of PR proteins is believed to be part of a primitive immunological response in plants (van Loon, L.C. (1985) Plant Mol. Biol. 4:111-116). GliPR shares up to 50% homology with the PR-1 protein family over a region that comprises almost two thirds of the protein, including a

20 conserved triad of amino acids, His-Glu-His, appropriately spaced to form a metal-binding domain (Murphy et al., *supra*).

Signaling initiated by the Trk family receptors plays a dynamic role in neurogenic tumors. The proto-oncogene Trks encode the high-affinity receptor tyrosine kinases for nerve growth factor (NGF) neurotrophins. A rearranged Trk oncogene is often observed in non-neuronal neoplasms such as

25 colon and papillary thyroid cancers. The proto-oncogene Trks regulates growth, differentiation and apoptosis of tumors of neuronal origin, such as neuroblastoma and medulloblastoma (Nakagawara, A. (2001) Cancer Lett. 169:107-114).

Neuronal thread proteins (NTP) are a group of immunologically related molecules found in the brain and neuroectodermal tumor cell lines. NTP expression is increased in neuronal cells during

30 proliferation, differentiation, brain development, in Alzheimer's disease (AD) brains, and in pathological states associated with regenerative nerve sprouting (de la Monte, S.M. et al. (1996) J. Neuropathol. Exp. Neurol. 55:1038-1050). Monoclonal antibodies generated to a recombinant NTP, AD7c-NTP, isolated from an end-stage AD brain library, showed high levels of NTP immunoreactivity in perikarya, neuropil fibers, and white matter fibers of AD brain tissue. *In vitro*

studies also demonstrated NTP upregulation, phosphorylation, and translocation from the perikarya to cell processes and growth cones during growth factor-induced neuritic sprouting and neuronal differentiation. Additionally, increased NTP immunoreactivity was found in Down syndrome brains beginning in the second decade, prior to establishment of widespread AD neurodegeneration, and at an age when a low-level or an absence of NTP expression was observed in control brains. These findings indicated that abnormal expression and accumulation of NTP in brain may be an early marker of AD neurodegeneration in Down syndrome (de la Monte, S.M. et al. (1996) *J. Neurol. Sci.* 135:118-125). Furthermore, the increased expression and accumulation of NTP in AD brain tissue was paralleled by corresponding elevations of NTP in cerebrospinal fluid (CSF), and elevated levels of NTP were detectable in the CSF early in the course of the disease.

Fe65-like protein (Fe65L2), a new member of the Fe65 protein family, is one of the ligands that interacts with the cytoplasmic domain of Alzheimer beta-amyloid precursor protein (APP). Transgenic mice expressing APP simulate some of the prominent behavioral and pathological features of Alzheimer's disease, including age-related impairment in learning and memory, neuronal loss, gliosis, neuritic changes, amyloid deposition, and abnormal tau phosphorylation (Duilio, A. et al. (1998) *Biochem. J.* 330:513-519).

Amyotrophic lateral sclerosis (ALS) is characterized by motor neuron death, altered peroxidase activity of mutant SOD1, changes in intracellular copper homeostasis, protein aggregation, and changes in the function of glutamate transporters leading to excitotoxicity. Neurofilaments and peripherin appear to play some part in motor neuron degeneration. ALS is occasionally associated with mutations of the neurofilament heavy chain gene (Al-Chalabi, A. and P.N. Leigh (2000) *Curr. Opin. Neurol.* 13:397-405). Cytoskeletal abnormalities such as abnormal inclusions containing neurofilaments (NFs) and/or peripherin, reduced mRNA levels for the NF light (NF-L) protein and mutations in the NF heavy (NF-H) gene have been observed in ALS. Intermediate filament inclusions containing peripherin may play a contributory role in ALS (Julien, J.P. and J.M. Beaulieu (2000) *J. Neurol. Sci.* 180:7-14).

Miller-Dieker syndrome (MDS) or isolated lissencephaly syndrome (ILS) are characterized by a smooth cerebral surface, a thickened cortex with four abnormal layers, and misplaced neurons. Both conditions may result from deletion or mutation in the LIS1 gene. The lissencephaly gene product Lis1 is a component of evolutionarily conserved intracellular multiprotein complexes essential for neuronal migration, and which may be components of the machinery for cell proliferation and intracellular transport (Leventer, R.J. et al. (2001) *Trends Neurosci.* 24:489-492). NudC, a nuclear movement protein, interacts with Lis1 (Morris, S.M. et al. (1998) *Curr. Biol.* 8:603-606).

Retinitis pigmentosa comprises a group of slowly progressive, inherited disorders of the retina

that cause loss of night vision and peripheral visual field in adolescence. A recessive nonsense mutation in the *Drosophila* opsin gene causes photoreceptor degeneration. In some families, genes encoding rhodopsin and peripherin/RDS map very close to the disease loci. Rhodopsin and peripherin/RDS mutations have been found in approximately 30% of all autosomal dominant cases

5 (Shastry, B.S. (1994) Am. J. Med. Genet. 52:467-474).

Synaptic proteins are involved in Alzheimer's disease (AD) and other disorders including ischemia, a variety of disorders where synapse-associated proteins are abnormally accumulated in the nerve terminals or synaptic proteins are altered after denervation, and neoplastic disorders (Masliah, E. and R. Terry (1993) Brain Pathol. 3:77-85). Synaptophysin (SY), a major integral membrane
10 protein of small synaptic vesicles, is on the X chromosome in subbands Xp11.22-p11.23, a region implicated in several inherited diseases including Wiskott-Aldrich syndrome, three forms of X-linked hypercalciuric nephrolithiasis, and the eye disorders retinitis pigmentosa 2, congenital stationary night blindness, and Aland Island eye disease (Fisher, S.E. et al. (1997) Genomics 45:340-347).

Mutations in the BRI2 isoform of the BRI gene family are associated with dementia in
15 humans (Vidal, R. et al. (2001) Gene 266:95-102).

Changes in the molecular and cellular components of neuronal signaling systems correlate with the effects on mood and cognition observed after long-term treatment with antidepressant drugs. Two serine/threonine kinases, Ca²⁺/calmodulin-dependent protein kinase II and cyclic AMP-dependent protein kinase, are activated in the brain following antidepressant treatment.
20 Associated changes in the phosphorylation of selected protein substrates in subcellular compartments including presynaptic terminals and microtubules may contribute to the modulation of synaptic transmission observed with antidepressants (Popoli, M. et al. (2001) Pharmacol. Ther. 89:149-170). Reserpine can cause a syndrome resembling depression, indicating the importance of vesicular transport activity for the control of mood and behavior. The psychostimulant amphetamine also
25 disrupts the storage of amines in secretory vesicles, further indicating that alterations in vesicular monoamine transport can affect behavior (Sulzer, D. and S. Rayport (1990) Neuron 5:797-808).

Decreased GABAergic neurotransmission has been implicated in the pathophysiology of CNS disorders such as epilepsy and schizophrenia. Impaired re-uptake of synaptic glutamate and a reduced expression of the glutamate transporter have been found in the motor cortex of patients with
30 amyotrophic lateral sclerosis (ALS). The loss of glial glutamate transporters produces elevated extracellular glutamate levels, neurodegeneration characteristic of excitotoxicity, and a progressive paralysis. The loss of neuronal glutamate transporters produces mild neurotoxicity and results in epilepsy (Rothstein, J.D. et al. (1996) Neuron 16:675-686). GABA transporter function is reduced in epileptic hippocampi. Transporters for dopamine, norepinephrine, and serotonin have particular

significance as targets for clinically relevant psychoactive agents including cocaine, antidepressants, and amphetamines. Cocaine and antidepressants are transporter antagonists that act with varying degrees of specificity to enhance synaptic concentrations of amines by limiting clearance.

Amphetamines enhance transporter mediated efflux in concert with a depletion of vesicular amine stores (Barker, E.L. and R.D. Blakely (1995) *Psychopharmacology* 28:321-333; Sulzer, D. and S. Rayport (1990) *Neuron* 5:797-808; Wall, S.C. et al. (1995) *Mol. Pharmacol.* 47:544-550).

The μ -opioid receptor (MOR) mediates the actions of analgesic agents including morphine, codeine, methadone, and fentanyl as well as heroin. MOR is functionally coupled to a G-protein-activated potassium channel (Mestek A. et al. (1995) *J. Neurosci.* 15:2396-2406). A variety of MOR subtypes exist. Alternative splicing has been observed with MOR-1 as with a number of G protein-coupled receptors including somatostatin 2, dopamine D2, prostaglandin EP3, and serotonin receptor subtypes 5-hydroxytryptamine₄ and 5-hydroxytryptamine₇ (Pan, Y.X. et al. (1999) *Mol. Pharm.* 56:396-403).

The central nervous system regulates the innate immune system by elaborating anti-inflammatory hormone cascades in response to bacterial products and immune mediators. The central nervous system also responds via acetylcholine-mediated efferent signals carried through the vagus nerve. Nicotinic cholinergic receptors expressed on macrophages detect these signals and respond with a dampened cytokine response (Tracey K.J. et al. (2001) *FASEB J.* 15:1575-1576).

Dysferlin is the protein product of the gene mutated in patients with an autosomal recessive limb-girdle muscular dystrophy type 2B (LGMD2B) and a distal muscular dystrophy, Miyoshi myopathy. Dysferlin is homologous to a *Caenorhabditis elegans* spermatogenesis factor, FER-1. Otoferlin, another human FER-1-like protein (ferlin), is responsible for autosomal recessive nonsyndromic deafness (DFNB9). All the ferlins are characterized by sequences corresponding to multiple C2 domains that share the highest level of homology with the C2A domain of rat synaptotagmin III (Britton S. et al. (2000) *Genomics* 68:313-321).

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

5 predisposition, condition, disease, or disorder.

Neurological disorders

Characterization of region-specific gene expression in the human brain provides a context and background for molecular neurobiology on a variety of neurological disorders. For example, Alzheimer's disease (AD) is a progressive, neurodestructive process of the human neocortex, characterized by the deterioration of memory and higher cognitive function. A progressive and irreversible brain disorder, AD is characterized by three major pathogenic episodes involving (a) an aberrant processing and deposition of beta-amyloid precursor protein (betaAPP) to form neurotoxic beta-amyloid (betaA) peptides and an aggregated insoluble polymer of betaA that forms the senile plaque, (b) the establishment of intraneuronal neuritic tau pathology yielding widespread deposits of agyrophilic neurofibrillary tangles (NFT) and (c) the initiation and proliferation of a brain-specific inflammatory response. These three seemingly disparate attributes of AD etiopathogenesis are linked by the fact that proinflammatory microglia, reactive astrocytes and their associated cytokines and chemokines are associated with the biology of the microtubule associated protein tau, betaA speciation and aggregation. Missense mutations in the presenilin genes PS1 and PS2, implicated in early onset familial AD, cause abnormal betaAPP processing with resultant overproduction of betaA42 and related neurotoxic peptides. Specific betaA fragments such as betaA42 can further potentiate proinflammatory mechanisms. Expression of the inducible oxidoreductase cyclooxygenase-2 and cytosolic phospholipase A2 (cPLA2) is strongly activated during cerebral ischemia and trauma, epilepsy and AD, indicating the induction of proinflammatory gene pathways as a response to brain injury. Neurotoxic metals such as aluminum and zinc, both implicated in AD etiopathogenesis, and arachidonic acid, a major metabolite of brain cPLA2 activity, each polymerize hyperphosphorylated tau to form NFT-like bundles. Studies have identified a reduced risk for AD in patients aged over 70 years who were previously treated with non-steroidal anti-inflammatory drugs for non-CNS afflictions that include arthritis. (For a review of the interrelationships between the mechanisms of PS1, PS2 and betaAPP gene expression, tau and betaA deposition and the induction, regulation and proliferation in AD of the neuroinflammatory response, see Lukiw, W.J, and Bazan, N.G. (2000) Neurochem. Res. 2000 25:1173-1184).

Colon cancer

While soft tissue sarcomas are relatively rare, more than 50% of new patients diagnosed with

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

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

Colon cancer is causally related to both genes and the environment. Several molecular pathways have been linked to the development of colon cancer, and the expression of key genes in any of these pathways may be lost by inherited or acquired mutation or by hypermethylation. There is a particular need to identify genes for which changes in expression may provide an early indicator of colon cancer or a predisposition for the development of colon cancer.

For example, it is well known that abnormal patterns of DNA methylation occur consistently in human tumors and include, simultaneously, widespread genomic hypomethylation and localized areas of increased methylation. In colon cancer in particular, it has been found that these changes occur early in tumor progression such as in premalignant polyps that precede colon cancer. Indeed, DNA methyltransferase, the enzyme that performs DNA methylation, is significantly increased in histologically normal mucosa from patients with colon cancer or the benign polyps that precede cancer, and this increase continues during the progression of colonic neoplasms (Wafik, S. et al. (1991) Proc. Natl. Acad. Sci. USA 88:3470-3474). Increased DNA methylation occurs in G+C rich areas of genomic DNA termed "CpG islands" that are important for maintenance of an "open" transcriptional conformation around genes, and hypermethylation of these regions results in a "closed" conformation that silences gene transcription. It has been suggested that the silencing or downregulation of differentiation genes by such abnormal methylation of CpG islands may prevent

differentiation in immortalized cells (Antequera, F. et al. (1990) Cell 62:503-514).

Familial Adenomatous Polyposis (FAP) is a rare autosomal dominant syndrome that precedes colon cancer and is caused by an inherited mutation in the adenomatous polyposis coli (APC) gene. FAP is characterized by the early development of multiple colorectal adenomas that progress to cancer at a mean age of 44 years. The APC gene is a part of the APC- β -catenin-Tcf (T-cell factor) pathway. Impairment of this pathway results in the loss of orderly replication, adhesion, and migration of colonic epithelial cells that results in the growth of polyps. A series of other genetic changes follow activation of the APC- β -catenin-Tcf pathway and accompanies the transition from normal colonic mucosa to metastatic carcinoma. These changes include mutation of the K-Ras proto-oncogene, changes in methylation patterns, and mutation or loss of the tumor suppressor genes p53 and Smad4/DPC4. While the inheritance of a mutated APC gene is a rare event, the loss or mutation of APC and the consequent effects on the APC- β -catenin-Tcf pathway is believed to be central to the majority of colon cancers in the general population.

Hereditary Nonpolyposis Colorectal Cancer (HNPCC) is another inherited autosomal dominant syndrome with a less well defined phenotype than FAP. HNPCC, which accounts for about 2% of colorectal cancer cases, is distinguished by the tendency to early onset of cancer and the development of other cancers, particularly those involving the endometrium, urinary tract, stomach and biliary system. HNPCC results from the mutation of one or more genes in the DNA mis-match repair (MMR) pathway. Mutations in two human MMR genes, MSH2 and MLH1, are found in a large majority of HNPCC families identified to date. The DNA MMR pathway identifies and repairs errors that result from the activity of DNA polymerase during replication. Furthermore, loss of MMR activity contributes to cancer progression through accumulation of other gene mutations and deletions, such as loss of the BAX gene which controls apoptosis, and the TGF β receptor II gene which controls cell growth. Because of the potential for irreparable damage to DNA in an individual with a DNA MMR defect, progression to carcinoma is more rapid than usual.

Although ulcerative colitis is a minor contributor to colon cancer, affected individuals have about a 20-fold increase in risk for developing cancer. Progression is characterized by loss of the p53 gene which may occur early, appearing even in histologically normal tissue. The progression of the disease from ulcerative colitis to dysplasia/carcinoma without an intermediate polyp state suggests a high degree of mutagenic activity resulting from the exposure of proliferating cells in the colonic mucosa to the colonic contents.

Almost all colon cancers arise from cells in which the estrogen receptor (ER) gene has been silenced. The silencing of ER gene transcription is age related and linked to hypermethylation of the ER gene (Issa, J-P. J. et al. (1994) Nature Genetics 7:536-540). Introduction of an exogenous ER

gene into cultured colon carcinoma cells results in marked growth suppression. The connection between loss of the ER protein in colonic epithelial cells and the consequent development of cancer has not been established.

Clearly there are a number of genetic alterations associated with colon cancer and with the development and progression of the disease, particularly the downregulation or deletion of genes, that potentially provide early indicators of cancer development, and which may also be used to monitor disease progression or provide possible therapeutic targets. The specific genes affected in a given case of colon cancer depend on the molecular progression of the disease. Identification of additional genes associated with colon cancer and the precancerous state would provide more reliable diagnostic patterns associated with the development and progression of the disease.

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

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, neurotransmission-associated proteins, referred to collectively as 'NTRAN' and individually as 'NTRAN-1,' 'NTRAN-2,' 'NTRAN-3,' 'NTRAN-4,' 'NTRAN-5,' 'NTRAN-6,' 'NTRAN-7,' and 'NTRAN-8' 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 neurotransmission-associated proteins 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 neurotransmission-associated proteins 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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-8.

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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-8. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:9-16.

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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. 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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. 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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

5 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:9-16, 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:9-16, c) a polynucleotide complementary to the
10 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
15 polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, 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:9-16, 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
20 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
25 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:9-16, b) a
30 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:9-16, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain

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

Another embodiment provides a composition comprising an effective amount of a polypeptide
5 selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and
10 d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, 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:1-8. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional NTRAN, comprising administering to a patient in need
15 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:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an
20 amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting agonist activity in the sample. Another
25 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 NTRAN, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness
30 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:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group

consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by
5 the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional NTRAN, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid
10 sequence selected from the group consisting of SEQ ID NO:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected
15 from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino
20 acid sequence selected from the group consisting of SEQ ID NO:1-8, 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:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected
25 from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence
30 of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, the method comprising a) contacting a sample comprising the target polynucleotide with a compound, b) detecting

altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, 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:9-16, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, 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:9-16, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

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

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

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

polypeptides.

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

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

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

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

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

30 DEFINITIONS

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

The term “agonist” refers to a molecule which intensifies or mimics the biological activity of

NTRAN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NTRAN either by directly interacting with NTRAN or by acting on components of the biological pathway in which NTRAN participates.

An "allelic variant" is an alternative form of the gene encoding NTRAN. Allelic variants may
5 result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times
10 in a given sequence.

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

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

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

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of NTRAN. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of NTRAN either by directly interacting with NTRAN or by acting on components of the biological pathway in which NTRAN 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 NTRAN 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
5 nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

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

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

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement,
25 3'-TCA-5'.

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

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to

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

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

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

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

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

The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains

at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

5 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.

10 “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 NTRAN or a polynucleotide encoding NTRAN which
15 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
20 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,
25 may be encompassed by the present embodiments.

A fragment of SEQ ID NO:9-16 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:9-16, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:9-16 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification
30 technologies and in analogous methods that distinguish SEQ ID NO:9-16 from related polynucleotides. The precise length of a fragment of SEQ ID NO:9-16 and the region of SEQ ID NO:9-16 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:1-8 is encoded by a fragment of SEQ ID NO:9-16. A fragment

of SEQ ID NO:1-8 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-8. For example, a fragment of SEQ ID NO:1-8 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-8. The precise length of a fragment of SEQ ID NO:1-8 and the region of SEQ ID NO:1-8 to which the fragment
5 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.

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

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical nucleotide matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way,
15 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
20 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,
25 window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at
30 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 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:

5 *Matrix: BLOSUM62*
 Reward for match: 1
 Penalty for mismatch: -2
 Open Gap: 5 and Extension Gap: 2 penalties
 Gap x drop-off: 50
 10 *Expect: 10*
 Word Size: 11
 Filter: on

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

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

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

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

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

Matrix: BLOSUM62
Open Gap: 11 and Extension Gap: 1 penalties
Gap x drop-off: 50
Expect: 10
 15 *Word Size: 3*
Filter: on

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

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

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

30 "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the

stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

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

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

5 An “immunogenic fragment” is a polypeptide or oligopeptide fragment of NTRAN 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 NTRAN which is useful in any of the antibody production methods disclosed herein or known in the art.

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

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

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

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

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

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

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

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

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

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a “mispriming library,” in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user’s specific needs.) The

PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

10 A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes
15 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
20 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,
25 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.

30 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 NTRAN,

nucleic acids encoding NTRAN, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

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

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

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

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

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

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

A “transgenic organism,” as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The

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). 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 for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

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

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

or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human neurotransmission-associated proteins (NTRAN), the polynucleotides encoding NTRAN, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/inflammatory, cardiovascular, neurological, developmental, cell proliferative, transport, psychiatric, metabolic, and endocrine disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to polypeptide embodiments 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 amino acid residues comprising signature sequences, domains, motifs, potential phosphorylation sites, and potential glycosylation sites. Column 5 shows analytical methods for protein structure/function analysis and in

some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are neurotransmission-associated proteins. For example, SEQ ID NO:2 is 99% identical, from residue M1 to residue G526, to human amyloid precursor-like protein 1 (GenBank ID g1709301) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 4.0e-289, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also has homology to proteins that are localized to the Golgi, cytoplasm, and plasma membrane, have heparin-binding ability, and are amyloid precursor-like protein 1, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:2 also contains an amyloid A4 extracellular domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:2 is an amyloid protein.

In an alternative example, SEQ ID NO:6 is 92% identical, from residue M1 to residue N1045, to *Rattus norvegicus* Robo2 (GenBank ID g6164831) 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:6 also has homology to proteins that have axonal signaling and guidance function, and are Robo receptors, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:6 also contains immunoglobulin and fibronectin type III domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM, INCY, and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLAST analyses against the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:6 is a Robo receptor protein.

In an alternative example, SEQ ID NO:8 is 88% identical, from residue E122 to residue T405, to human SEMB (semaphorin B) (GenBank ID g12248382) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.0e-300, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:8 also has homology to Semaphorin 4a (Semaphorin B) in the transmembrane type subfamily of semaphorins, which is a family of proteins involved in neuronal axon guidance during neural development, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:8 also contains a Sema domain as determined by searching for statistically significant matches in the

hidden Markov model (HMM)-based PFAM database of conserved protein families/domains. (See Table 3.) Data from BLAST analyses against the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:8 is a semaphorin. SEQ ID NO:1, SEQ ID NO:3-5, and SEQ ID NO:7 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-8 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:9-16 or that distinguish between SEQ ID NO:9-16 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_N₁N₂YYYYY_N₃N₄ represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYYY 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,”

5 “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).

10

15

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.

20

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.

25

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

30

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
5 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 NTRAN variants. Various embodiments of NTRAN variants can have at least about 70%, at least about 75%, at least about 80%, at least about 85%, at
10 least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% amino acid sequence identity to the NTRAN amino acid sequence, and can contain at least one functional or structural characteristic of NTRAN.

Various embodiments also encompass polynucleotides which encode NTRAN. In a particular
15 embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:9-16, which encodes NTRAN. The polynucleotide sequences of SEQ ID NO:9-16, 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.

20 The invention also encompasses variants of a polynucleotide encoding NTRAN. In particular, such a variant polynucleotide will have at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% polynucleotide sequence identity to a polynucleotide encoding NTRAN. A particular
25 aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:9-16 which has at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% polynucleotide sequence identity to a nucleic acid sequence
30 selected from the group consisting of SEQ ID NO:9-16. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of NTRAN.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding NTRAN. A splice variant may have portions which have significant

sequence identity to a polynucleotide encoding NTRAN, but will generally have a greater or lesser number of nucleotides due to additions or deletions of blocks of sequence arising from alternate splicing during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding NTRAN 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 NTRAN. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of NTRAN.

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 NTRAN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring NTRAN, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode NTRAN and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring NTRAN under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding NTRAN or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding NTRAN 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 NTRAN and NTRAN 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 NTRAN 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:9-16 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
5 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
10 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
15 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 NTRAN 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
20 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
25 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.
30 (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (BD 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
5 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
10 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
15 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 NTRAN may be cloned in recombinant DNA molecules that direct expression of NTRAN, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy
20 of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express NTRAN.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter NTRAN-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
25 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
30 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 NTRAN, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is

produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through “artificial” breeding and rapid
5 molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

10 In another embodiment, polynucleotides encoding NTRAN 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, NTRAN 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
15 solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of NTRAN, 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
20 variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

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

25 In order to express a biologically active NTRAN, the polynucleotides encoding NTRAN 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 polynucleotides encoding
30 NTRAN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding NTRAN. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding NTRAN 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 NTRAN and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

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

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding NTRAN. For example, routine cloning, subcloning, and propagation of polynucleotides encoding NTRAN can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding NTRAN into the vector's multiple cloning

site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of NTRAN are needed, e.g. for the production of antibodies, vectors which direct high level expression of NTRAN 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 NTRAN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

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

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

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers,

or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

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

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

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

In general, host cells that contain the polynucleotide encoding NTRAN and that express NTRAN may be identified by a variety of procedures known to those of skill in the art. These

procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

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

A wide variety of labels and conjugation techniques are known by those skilled in the art and
15 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 NTRAN include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding NTRAN, 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
20 available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or
25 chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

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

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation,

phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a “prepro” or “pro” form of the protein may also be used to specify protein targeting, folding, and/or activity.

Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding NTRAN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric NTRAN protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of NTRAN activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the NTRAN encoding sequence and the heterologous protein sequence, so that NTRAN 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 purification of fusion proteins.

In another embodiment, synthesis of radiolabeled NTRAN 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 example, ³⁵S-methionine.

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

In related embodiments, variants of NTRAN can be used to screen for binding of test compounds, such as antibodies, to NTRAN, a variant of NTRAN, or a combination of NTRAN and/or one or more variants NTRAN. In an embodiment, a variant of NTRAN can be used to screen for compounds that bind to a variant of NTRAN, but not to NTRAN having the exact sequence of a
5 sequence of SEQ ID NO:1-8. NTRAN variants used to perform such screening can have a range of about 50% to about 99% sequence identity to NTRAN, 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 NTRAN can be closely related to the natural ligand of NTRAN, e.g., a ligand or fragment thereof, a natural substrate,
10 a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor NTRAN (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 NTRAN can
15 be closely related to the natural receptor to which NTRAN 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 NTRAN which is capable of propagating a signal, or a decoy receptor for NTRAN 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-
20 336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG₁ (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

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

In an embodiment, anticalins can be screened for specific binding to NTRAN, fragments of NTRAN, or variants of NTRAN. 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 NTRAN involves producing appropriate cells which express NTRAN, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing NTRAN or cell membrane fractions which contain NTRAN are then contacted with a test compound and binding, stimulation, or inhibition of activity of either NTRAN 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 NTRAN, either in solution or affixed to a solid support, and detecting the binding of NTRAN 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).

NTRAN, fragments of NTRAN, or variants of NTRAN may be used to screen for compounds that modulate the activity of NTRAN. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive

for NTRAN activity, wherein NTRAN is combined with at least one test compound, and the activity of NTRAN in the presence of a test compound is compared with the activity of NTRAN in the absence of the test compound. A change in the activity of NTRAN in the presence of the test compound is indicative of a compound that modulates the activity of NTRAN. Alternatively, a test
5 compound is combined with an *in vitro* or cell-free system comprising NTRAN under conditions suitable for NTRAN activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of NTRAN 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 NTRAN or their mammalian homologs may
10 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
15 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).
20 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 NTRAN may also be manipulated *in vitro* in ES cells derived from
25 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 NTRAN can also be used to create “knockin” humanized animals
30 (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding NTRAN 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 NTRAN, e.g., by secreting NTRAN in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between
 5 regions of NTRAN and neurotransmission-associated proteins. In addition, examples of tissues
 expressing NTRAN can be found in Table 6 and can also be found in Example XI. Therefore,
 NTRAN appears to play a role in autoimmune/inflammatory, cardiovascular, neurological,
 developmental, cell proliferative, transport, psychiatric, metabolic, and endocrine disorders. In the
 treatment of disorders associated with increased NTRAN expression or activity, it is desirable to
 10 decrease the expression or activity of NTRAN. In the treatment of disorders associated with
 decreased NTRAN expression or activity, it is desirable to increase the expression or activity of
 NTRAN.

Therefore, in one embodiment, NTRAN or a fragment or derivative thereof may be
 administered to a subject to treat or prevent a disorder associated with decreased expression or
 15 activity of NTRAN. Examples of such disorders include, but are not limited to, an
 autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's
 disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia,
 asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune
 polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact
 20 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, syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic
 25 sclerosis, thromb osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome,
 rheumatoid arthritis, scleroderma, Sjögren's ocytopenic purpura, ulcerative colitis, uveitis, Werner
 syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial,
 fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as
 congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive
 30 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, 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, arteriovenous fistula,

atherosclerosis, 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; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, 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 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 degeneration, and familial frontotemporal dementia; a developmental disorder such as 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 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 cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer 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; a transport disorder such as

akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis,

5 myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrhythmia, tachyarrhythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological

10 disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's

15 disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a psychiatric disorder such as acute stress disorder, alcohol dependence, amphetamine dependence, anorexia nervosa, antisocial personality disorder, attention-deficit hyperactivity disorder, autistic disorder, anxiety, avoidant

20 personality disorder, bipolar disorder, borderline personality disorder, brief psychotic disorder, bulimia nervosa, cannabis dependence, cocaine dependence, conduct disorder, cyclothymic disorder, delirium, delusional disorder, dementia, dependent personality disorder, depression, dysthymic disorder, hallucinogen dependence, histrionic personality disorder, inhalant dependence, manic depression, multi-infarct dementia, narcissistic personality disorder, nicotine dependence, obsessive-compulsive

25 disorder, opioid dependence, oppositional defiant disorder, panic disorder, paranoid personality disorder, phencyclidine dependence, phobia, posttraumatic stress disorder, schizoaffective disorder, schizoid personality disorder, schizophrenia, sedative dependence, separation anxiety disorder, and sleep disorder; a metabolic disorder such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, fatty hepatocirrhosis,

30 fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, mannosidosis, neuraminidase deficiency, obesity, osteoporosis, phenylketonuria, pseudovitamin D-

deficiency rickets, disorders of carbohydrate metabolism such as congenital type II dyserythropoietic anemia, diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, galactose epimerase deficiency, glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, and inherited abnormalities of pyruvate metabolism, disorders of lipid metabolism such as

5 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, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis,

10 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, and lipid myopathies, and disorders of

15 copper metabolism such as Menke's disease, Wilson's disease, and Ehlers-Danlos syndrome type IX diabetes; and an endocrine disorder such as a disorder of the hypothalamus and/or pituitary resulting from lesions such as a primary brain tumor, adenoma, infarction associated with pregnancy, hypophysectomy, aneurysm, vascular malformation, thrombosis, infection, immunological disorder, and complication due to head trauma, a disorder associated with hypopituitarism including hypogonadism,

20 Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism, a disorder associated with hyperpituitarism including acromegaly, gigantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma, a disorder associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection,

25 subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism, a disorder associated with hyperthyroidism including thyrotoxicosis and its various forms, Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease, a disorder associated with hyperparathyroidism including Conn disease (chronic hypercalcemia), a pancreatic disorder such as Type I or Type II diabetes mellitus and associated

30 complications, a disorder associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease, a disorder associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbation of the menstrual cycle, polycystic ovarian

disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women, osteoporosis, and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, a hypergonadal disorder associated with Leydig cell tumors, androgen resistance associated with
5 absence of androgen receptors, syndrome of 5 α -reductase, and gynecomastia.

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

In a further embodiment, a composition comprising a substantially purified NTRAN in
10 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 NTRAN including, but not limited to, those provided above.

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

In a further embodiment, an antagonist of NTRAN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of NTRAN. Examples of such disorders include, but are not limited to, those autoimmune/inflammatory, cardiovascular, neurological, developmental, cell proliferative, transport, psychiatric, metabolic, and endocrine disorders described
20 above. In one aspect, an antibody which specifically binds NTRAN 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 NTRAN.

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

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

An antagonist of NTRAN may be produced using methods which are generally known in the art. In particular, purified NTRAN may be used to produce antibodies or to screen libraries of

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

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

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to NTRAN have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of NTRAN 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 NTRAN may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce NTRAN-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 population
5 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 NTRAN may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin
10 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
15 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 NTRAN and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering NTRAN epitopes is generally used, but a competitive binding assay may also be
20 employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for NTRAN. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of NTRAN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a
25 determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple NTRAN epitopes, represents the average affinity, or avidity, of the antibodies for NTRAN. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular NTRAN 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 NTRAN-
30 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 NTRAN, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons,

New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg
5 specific antibody/ml, is generally employed in procedures requiring precipitation of NTRAN-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 NTRAN, or any fragment
10 or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding NTRAN. 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
15 NTRAN (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) FASEB J. 9:1288-1296).
20 Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271-278; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in
25 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 NTRAN 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-
30 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.,
 5 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 NTRAN expression or regulation causes disease, the
 10 expression of NTRAN 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 NTRAN are treated by constructing mammalian expression vectors encoding NTRAN and introducing these vectors by mechanical means into NTRAN-deficient cells. Mechanical transfer
 15 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).

20 Expression vectors that may be effective for the expression of NTRAN include, but are not limited to, the pCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (BD Clontech, Palo Alto CA). NTRAN may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus
 25 (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
 30 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 NTRAN 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
5 standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to NTRAN expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding NTRAN 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
10 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
15 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
20 a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

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

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

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding NTRAN to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for NTRAN into the alphavirus genome in place of the capsid-coding region results in the production of a large number of NTRAN-coding RNAs and the synthesis of high levels of NTRAN 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 NTRAN 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 NTRAN.

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

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends

of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, 5 guanine, thymine, and uracil which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a 10 targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment 15 of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. siRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. 20 siRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

siRNA can be generated indirectly by introduction of dsRNA into the targeted cell. Alternatively, siRNA can be synthesized directly and introduced into a cell by transfection methods 25 and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable siRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide 30 length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other

sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected siRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

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

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

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

In various embodiments, one or more test compounds may be screened for effectiveness in

altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound

5 based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding NTRAN 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 NTRAN are assayed

10 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 NTRAN. 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

15 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.

20 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).

25 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-

30 466).

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

5 NTRAN, antibodies to NTRAN, and mimetics, agonists, antagonists, or inhibitors of NTRAN.

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

10 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
15 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 allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination
20 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 NTRAN or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, NTRAN or a fragment thereof may be joined to a short cationic N-
25 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,
30 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 NTRAN or fragments thereof, antibodies of NTRAN, and agonists, antagonists or inhibitors of

NTRAN, 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 $0.1 \mu\text{g}$ to $100,000 \mu\text{g}$, 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 NTRAN may be used for the diagnosis of disorders characterized by expression of NTRAN, or in assays to monitor patients being treated with NTRAN or agonists, antagonists, or inhibitors of NTRAN. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for NTRAN include methods which utilize the antibody and a label to detect NTRAN 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 NTRAN, including ELISAs, RIAs, and FACS, are

known in the art and provide a basis for diagnosing altered or abnormal levels of NTRAN expression. Normal or standard values for NTRAN expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to NTRAN under conditions suitable for complex formation. The amount of standard complex formation
5 may be quantitated by various methods, such as photometric means. Quantities of NTRAN 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 NTRAN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary
10 RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of NTRAN may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of NTRAN, and to monitor regulation of NTRAN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides,
15 including genomic sequences, encoding NTRAN or closely related molecules may be used to identify nucleic acid sequences which encode NTRAN. 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 NTRAN, allelic variants, or related sequences.

20 Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the NTRAN 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:9-16 or from genomic sequences including promoters, enhancers, and introns of the NTRAN gene.

Means for producing specific hybridization probes for polynucleotides encoding NTRAN
25 include the cloning of polynucleotides encoding NTRAN or NTRAN 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,
30 such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding NTRAN may be used for the diagnosis of disorders associated with expression of NTRAN. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia,

asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, 5 glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thromb osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's ocytopenic purpura, ulcerative colitis, uveitis, Werner 10 syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a cardiovascular disorder such as 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 15 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, arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, 20 balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, 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 25 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 30 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, 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 developmental disorder such as 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 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 cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and a cancer 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; a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrhythmia, tachyarrhythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a psychiatric disorder such

as acute stress disorder, alcohol dependence, amphetamine dependence, anorexia nervosa, antisocial personality disorder, attention-deficit hyperactivity disorder, autistic disorder, anxiety, avoidant personality disorder, bipolar disorder, borderline personality disorder, brief psychotic disorder, bulimia nervosa, cannabis dependence, cocaine dependence, conduct disorder, cyclothymic disorder, delirium,

5 delusional disorder, dementia, dependent personality disorder, depression, dysthymic disorder, hallucinogen dependence, histrionic personality disorder, inhalant dependence, manic depression, multi-infarct dementia, narcissistic personality disorder, nicotine dependence, obsessive-compulsive disorder, opioid dependence, oppositional defiant disorder, panic disorder, paranoid personality disorder, phencyclidine dependence, phobia, posttraumatic stress disorder, schizoaffective disorder, schizoid

10 personality disorder, schizophrenia, sedative dependence, separation anxiety disorder, and sleep disorder; a metabolic disorder such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism,

15 hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, lipid myopathies, lipodystrophies, lysosomal storage diseases, mannosidosis, neuraminidase deficiency, obesity, osteoporosis, phenylketonuria, pseudovitamin D-deficiency rickets, disorders of carbohydrate metabolism such as congenital type II dyserythropoietic anemia, diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus,

20 galactose epimerase deficiency, glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, and inherited abnormalities of pyruvate metabolism, disorders of lipid metabolism 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,

25 adrenoleukodystrophy, GM₂ gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol

30 acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, and lipid myopathies, and disorders of copper metabolism such as Menke's disease, Wilson's disease, and Ehlers-Danlos syndrome type IX diabetes; and an endocrine disorder such as a disorder of the hypothalamus and/or pituitary resulting from lesions such as a primary brain tumor, adenoma, infarction associated with pregnancy,

hypophysectomy, aneurysm, vascular malformation, thrombosis, infection, immunological disorder, and complication due to head trauma, a disorder associated with hypopituitarism including hypogonadism, Sheehan syndrome, diabetes insipidus, Kallman's disease, Hand-Schuller-Christian disease, Letterer-Siwe disease, sarcoidosis, empty sella syndrome, and dwarfism, a disorder associated with

5 hyperpituitarism including acromegaly, gigantism, and syndrome of inappropriate antidiuretic hormone (ADH) secretion (SIADH) often caused by benign adenoma, a disorder associated with hypothyroidism including goiter, myxedema, acute thyroiditis associated with bacterial infection, subacute thyroiditis associated with viral infection, autoimmune thyroiditis (Hashimoto's disease), and cretinism, a disorder associated with hyperthyroidism including thyrotoxicosis and its various forms,

10 Grave's disease, pretibial myxedema, toxic multinodular goiter, thyroid carcinoma, and Plummer's disease, a disorder associated with hyperparathyroidism including Conn disease (chronic hypercalcemia), a pancreatic disorder such as Type I or Type II diabetes mellitus and associated complications, a disorder associated with the adrenals such as hyperplasia, carcinoma, or adenoma of the adrenal cortex, hypertension associated with alkalosis, amyloidosis, hypokalemia, Cushing's

15 disease, Liddle's syndrome, and Arnold-Healy-Gordon syndrome, pheochromocytoma tumors, and Addison's disease, a disorder associated with gonadal steroid hormones such as: in women, abnormal prolactin production, infertility, endometriosis, perturbation of the menstrual cycle, polycystic ovarian disease, hyperprolactinemia, isolated gonadotropin deficiency, amenorrhea, galactorrhea, hermaphroditism, hirsutism and virilization, breast cancer, and, in post-menopausal women,

20 osteoporosis, and, in men, Leydig cell deficiency, male climacteric phase, and germinal cell aplasia, a hypergonadal disorder associated with Leydig cell tumors, androgen resistance associated with absence of androgen receptors, syndrome of 5 α -reductase, and gynecomastia. Polynucleotides encoding NTRAN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in

25 microarrays utilizing fluids or tissues from patients to detect altered NTRAN expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding NTRAN may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding NTRAN may be labeled by standard methods and added to a

30 fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding NTRAN in the sample indicates the presence of the associated disorder. Such assays may also be

used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of NTRAN, a normal or standard profile for expression is established. This may be accomplished by
5 combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding NTRAN, 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
10 obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

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

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

Additional diagnostic uses for oligonucleotides designed from the sequences encoding NTRAN may involve the use of PCR. These oligomers may be chemically synthesized, generated
25 enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding NTRAN, or a fragment of a polynucleotide complementary to the polynucleotide encoding NTRAN, 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.

30 In a particular aspect, oligonucleotide primers derived from polynucleotides encoding NTRAN 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 NTRAN are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

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

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

In another embodiment, NTRAN, fragments of NTRAN, or antibodies specific for NTRAN 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.

15 A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484;
20 hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The
25 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
30 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
5 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
10 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 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
15 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
20 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
25 of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis
30 in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test

compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

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

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

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

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

two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/25116; Shalon, D. et al. (1995) 5 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 NTRAN may be 10 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 15 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 20 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 25 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 NTRAN on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as 30 linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely

localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc.,

5 among normal, carrier, or affected individuals.

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

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

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

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

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

30 The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/453,108, U.S. Ser. No. 60/471,930, and U.S. Ser. No. 60/490,749, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

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

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

In some cases, Stratagene is provided with RNA and constructs the corresponding cDNA
15 libraries. Otherwise, cDNA is synthesized and cDNA libraries are constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription is initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters are ligated to double
20 stranded cDNA, and the cDNA is digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA is size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs are ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid
25 (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte, Palo Alto CA), pRARE (Incyte), or pINCY (Incyte), or derivatives thereof. Recombinant plasmids are transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Invitrogen.

30 II. Isolation of cDNA Clones

Plasmids obtained as described in Example I are recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids are purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid,

QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids are resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

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

10 III. Sequencing and Analysis

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

25 Polynucleotide sequences derived from Incyte cDNAs are validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof are then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART

(Schultz, J. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:5857-5864; Letunic, I. et al. (2002) *Nucleic Acids Res.* 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries are performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The

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

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

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

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

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

25 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).

30 The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences are also used to identify polynucleotide sequence fragments from SEQ ID NO:9-16. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

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

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

25 "Stitched" Sequences

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

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

“Stretched” Sequences

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

VI. Chromosomal Mapping of NTRAN Encoding Polynucleotides

The sequences used to assemble SEQ ID NO:9-16 are compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:9-16 are assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon are used to determine if any of the clustered sequences have been previously mapped. Inclusion of a mapped sequence in a cluster results in the assignment of all sequences of that cluster, including its particular SEQ ID NO:., to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map

position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by 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 (ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

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

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

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

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

other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding NTRAN are analyzed with respect to the tissue sources from which they are derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; 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 NTRAN. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ database (Incyte, Palo Alto CA).

VIII. Extension of NTRAN Encoding Polynucleotides

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

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

High fidelity amplification is obtained by PCR using methods well known in the art. PCR is performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contains DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps

2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ are as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

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

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

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

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

IX. Identification of Single Nucleotide Polymorphisms in NTRAN Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) are identified in SEQ ID NO:9-16 using the LIFESEQ database (Incyte). Sequences from the same gene
5 are clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters is used to distinguish SNPs from other sequence variants. Preliminary filters remove the majority of basecall errors by requiring a minimum Phred quality score of 15, and remove sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated
10 procedure of advanced chromosome analysis is applied to the original chromatogram files in the vicinity of the putative SNP. Clone error filters use statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters use statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human
15 sequences. A final set of filters removes duplicates and SNPs found in immunoglobulins or T-cell receptors.

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

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:9-16 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base
30 pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ - ^{32} P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25

superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

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

XI. Microarrays

 The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the
15 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 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
20 produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31).

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

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (BD Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 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-Aldrich, St. Louis MO) 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 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and
5 Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample
mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with
an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly
larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140
 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about
10 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1%
SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

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

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

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

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

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

15 Expression

For example, SEQ ID NO:9 showed differential expression, as determined by microarray analysis. Expression of SEQ ID NO:9 was down-regulated in diseased versus normal brain tissue. In this experiment, specific dissected brain regions from the brain of a female with severe AD were compared to dissected regions from a normal female and two normal male brains. The diagnosis of normal or severe AD was established by a certified neuropathologist based on microscopic examination of multiple sections throughout the brain. Expression of SEQ ID NO:9 was decreased at least two-fold in human brain tissue from the anterior hippocampal region from a patient with severe Alzheimer's disease versus anterior hippocampal tissue from the brain of a normal patient. Therefore, in various embodiments, SEQ ID NO:9 can be used for one or more of the following: i) monitoring treatment of Alzheimer's disease, ii) diagnostic assays for Alzheimer's disease, and iii) developing therapeutics and/or other treatments for Alzheimer's disease.

In a further example, changes in DNA methylation have been observed in tumors and implicated in microsatellite instability. Methylation of DNA in mammalian cells occurs on cytosines in CpG dinucleotides, which are found in CpG islands in the promoter regions of genes. Methylation of CpG islands is correlated with changes in gene expression. Studies in mice have shown that low levels of methyltransferase, the enzyme that methylates cytosines, can suppress intestinal polyp formation in mice lacking the tumor suppressor gene APC. This suggests a role for DNA methylation in the process of colon polyp formation. The compound 5-aza-2-deoxycytidine irreversibly inactivates the methyltransferase. Genes regulated by methylation in HT29 cells may be linked to the process of

polyp formation in the colon.

Expression of SEQ ID NO:16 was upregulated in HT29 cells treated with 5-aza-2-deoxycytidine versus untreated HT29 cells, as determined by microarray analysis. Gene expression profiles were obtained by comparing HT29 cells treated with 5-aza-2-deoxycytidine to untreated HT29
5 cells. HT29 cells (Huntsman Cancer Institute, Salt Lake City, UT) were treated with 500 nM 5-aza-2-deoxycytidine. Expression of SEQ ID NO:16 was increased at least two-fold in treated versus untreated cells. Therefore, in various embodiments, SEQ ID NO:16 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

10 XII. Complementary Polynucleotides

Sequences complementary to the NTRAN-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring NTRAN. 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
15 designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of NTRAN. 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 NTRAN-encoding transcript.

XIII. Expression of NTRAN

20 Expression and purification of NTRAN is achieved using bacterial or virus-based expression systems. For expression of NTRAN 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
25 element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express NTRAN upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of NTRAN 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
30 replaced with cDNA encoding NTRAN 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, NTRAN 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 NTRAN 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 NTRAN obtained by these methods can be used directly in the assays shown in Examples XVII and XVIII, where applicable.

XIV. Functional Assays

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

membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of NTRAN on gene expression can be assessed using highly purified
5 populations of cells transfected with sequences encoding NTRAN 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
10 art. Expression of mRNA encoding NTRAN and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of NTRAN Specific Antibodies

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

Alternatively, the NTRAN 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
20 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
25 complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-NTRAN activity by, for example, binding the peptide or NTRAN 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 NTRAN Using Specific Antibodies

Naturally occurring or recombinant NTRAN is substantially purified by immunoaffinity
30 chromatography using antibodies specific for NTRAN. An immunoaffinity column is constructed by covalently coupling anti-NTRAN 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 NTRAN are passed over the immunoaffinity column, and the column is

washed under conditions that allow the preferential absorbance of NTRAN (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/NTRAN 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 NTRAN is collected.

5 XVII. Identification of Molecules Which Interact with NTRAN

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

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

15 NTRAN 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 NTRAN Activity

20 An assay for NTRAN activity measures the expression of NTRAN on the cell surface. cDNA encoding NTRAN is transfected into an appropriate mammalian cell line. Cell surface proteins are labeled with biotin as described (de la Fuente, M.A. et al. (1997) *Blood* 90:2398-2405). Immunoprecipitations are performed using NTRAN-specific antibodies, and immunoprecipitated samples are analyzed using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)
25 and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of NTRAN expressed on the cell surface.

In the alternative, an assay for NTRAN activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding NTRAN is added
30 to quiescent 3T3 cultured cells using transfection methods well known in the art. The transiently transfected cells are then incubated in the presence of $[^3\text{H}]$ thymidine, a radioactive DNA precursor molecule. Varying amounts of NTRAN ligand are then added to the cultured cells. Incorporation of $[^3\text{H}]$ thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radioisotope counter, and the amount incorporated is directly proportional to the amount of newly

synthesized DNA. A linear dose-response curve over at least a hundred-fold NTRAN ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the concentration of NTRAN producing a 50% response level, where 100% represents maximal incorporation of [³H]thymidine into acid-precipitable DNA (McKay, I. and I. Leigh, eds. (1993)

5 Growth Factors: A Practical Approach, Oxford University Press, New York NY, p. 73.)

In a further alternative, the assay for NTRAN activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) *J. Biol. Chem.* 273:4990-4996). A plasmid encoding full length NTRAN is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human
10 embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods
15 well-known in the art. Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without ligand are proportional to the amount of NTRAN present in the transfected cells.

To measure changes in inositol phosphate levels, the cells are grown in 24-well plates containing 1×10^5 cells/well and incubated with inositol-free media and [³H]myoinositol, 2 mCi/well, for
20 48 hr. The culture medium is removed, and the cells washed with buffer containing 10 mM LiCl followed by addition of ligand. The reaction is stopped by addition of perchloric acid. Inositol phosphates are extracted and separated on Dowex AG1-X8 (Bio-Rad) anion exchange resin, and the total labeled inositol phosphates counted by liquid scintillation. Changes in the levels of labeled inositol phosphate from cells exposed to ligand compared to those without ligand are proportional to the
25 amount of NTRAN present in the transfected cells.

In a further alternative, the ion conductance capacity of NTRAN is demonstrated using an electrophysiological assay. NTRAN is expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding NTRAN. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to
30 those skilled in the art. A small amount of a second plasmid, which expresses any one of a number of marker genes such as β -galactosidase, is co-transformed into the cells in order to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after transformation under conditions appropriate for the cell line to allow expression and accumulation of NTRAN and β -galactosidase. Transformed cells expressing β -

galactosidase are stained blue when a suitable colorimetric substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance due to various ions by electrophysiological techniques that are well known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or β -galactosidase sequences alone, are used as controls and tested in parallel. The contribution of NTRAN to cation or anion conductance can be shown by incubating the cells using antibodies specific for either NTRAN. The respective antibodies will bind to the extracellular side of NTRAN, thereby blocking the pore in the ion channel, and the associated conductance. To study the dependence of NAP on external ions, sodium can be replaced by choline or N-methyl-D-glucamine and chloride by gluconate, NO_3 , or SO_4 (Kavanaugh, M.P. et al. (1992) J. Biol. Chem. 267:22007-22009).

In a further alternative, NTRAN transport activity is assayed by measuring uptake of labeled substrates into *Xenopus laevis* oocytes. Oocytes at stages V and VI are injected with NTRAN mRNA (10 ng per oocyte) and incubated for 3 days at 18°C in OR2 medium (82.5 mM NaCl, 2.5 mM KCl, 1 mM CaCl_2 , 1 mM MgCl_2 , 1 mM Na_2HPO_4 , 5 mM Hepes, 3.8 mM NaOH, 50 $\mu\text{g/ml}$ gentamycin, pH 7.8) to allow expression of NTRAN protein. Oocytes are then transferred to standard uptake medium (100 mM NaCl, 2 mM KCl, 1 mM CaCl_2 , 1 mM MgCl_2 , 10 mM Hepes/Tris pH 7.5). Uptake of various substrates (e.g., amino acids, sugars, drugs, and neurotransmitters) is initiated by adding a ^3H substrate to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in Na^+ -free medium, measuring the incorporated ^3H , and comparing with controls. NTRAN activity is proportional to the level of internalized ^3H substrate.

In a further alternative, NTRAN activity can be demonstrated using an electrophysiological assay for ion conductance. Capped NTRAN mRNA transcribed with T7 polymerase is injected into defolliculated stage V *Xenopus* oocytes, similar to the previously described method. Two to seven days later, transport is measured by two-electrode voltage clamp recording. Two-electrode voltage clamp recordings are performed at a holding potential of 50 mV. The data are filtered at 10 Hz and recorded with the MacLab digital-to-analog converter and software for data acquisition and analysis (AD Instruments, Castle Hill, Australia). To study the dependence of NTRAN on external ions, sodium can be replaced by choline or N-methyl-D-glucamine and chloride by gluconate, NO_3 , or SO_4 (Kavanaugh, M.P. et al. (1992) J. Biol. Chem. 267:22007-22009).

In the alternative, choline transporter activity or choline-transporter-like CTL1 protein activity of NTRAN is determined by measuring choline uptake by yeast transformed with expression vectors harboring polynucleotides encoding NTRAN. The assay is performed in nitrogen-free medium at 30°C for 10 or 30 min in the presence of 25 nM [^3H]choline. The cells are then filtered, and washed. The amount of [^3H]choline present in the cells is proportional to the activity of NTRAN in the cells

(O'Regan, S. *supra*).

In a further alternative, NTRAN protein kinase (PK) activity is measured by phosphorylation of a protein substrate using gamma-labeled [^{32}P]-ATP and quantitation of the incorporated radioactivity using a gamma radioisotope counter. NTRAN is incubated with the protein substrate,
5 [^{32}P]-ATP, and an appropriate kinase buffer. The ^{32}P incorporated into the product is separated from free [^{32}P]-ATP by electrophoresis and the incorporated ^{32}P is counted. The amount of ^{32}P recovered is proportional to the PK activity of NTRAN in the assay. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the hydrolyzed protein.

10 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.

15 Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of
20 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
7513114	1	7513114CD1	9	7513114CB1	
7512261	2	7512261CD1	10	7512261CB1	
7513222	3	7513222CD1	11	7513222CB1	
7522457	4	7522457CD1	12	7522457CB1	90187995CA2, 90188139CA2, 90188155CA2
7518601	5	7518601CD1	13	7518601CB1	90109918CA2
7520315	6	7520315CD1	14	7520315CB1	90190792CA2
7525536	7	7525536CD1	15	7525536CB1	
7520619	8	7520619CD1	16	7520619CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7513114CD1	g1857708	5.5E-271	[Homo sapiens] contactin associated protein Peles, E. et al. Identification of a novel contactin-associated transmembrane receptor with multiple domains implicated in protein-protein interactions. EMBO J. 16, 978-988 (1997)
1	7513114CD1	336668 CNTNAP1	4.3E-272	[Homo sapiens][Receptor (signaling)][Plasma membrane] Neurexin 4 (contactin associated protein 1), a neuronal paranodal transmembrane receptor containing epidermal growth factor-like and neurexin motifs, binds contactin (CNTN1), may have a role in neuronal intracellular signaling
1	7513114CD1	747875 Cntnap1	7.5E-252	Peles, E. et al. (1997) supra [Rattus norvegicus][Endoplasmic reticulum; Cytoplasmic; Cytoskeletal; Plasma membrane] Neurexin 4 (contactin associated protein 1), a neuronal paranodal transmembrane receptor containing epidermal growth factor-like and neurexin motifs, binds contactin (Rn.21397), may have a role in neuronal intracellular signaling
				Menegoz, M. et al. Paranodin, a glycoprotein of neuronal paranodal membranes. Neuron 19, 319-331 (1997)
2	7512261CD1	g1709301	4.0E-289	[Homo sapiens] amyloid precursor-like protein 1 Paliga, K. et al. Human amyloid precursor-like protein 1--cDNA cloning, ectopic expression in COS-7 cells and identification of soluble forms in the cerebrospinal fluid. Eur. J. Biochem. 250, 354-363 (1997)
2	7512261CD1	343962 APLP1	9.6E-291	[Homo sapiens][Structural protein][Plasma membrane] Amyloid precursor-like protein 1, membrane-associated heparin-binding glycoprotein expressed predominantly in the brain; constituent of senile plaques in Alzheimer's disease
				Lenkkeri, U. et al. Structure of the human amyloid-precursor-like protein gene APLP1 at 19q13.1. Hum. Genet. 102, 192-196 (1998)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
2	7512261CD1	583597 Ap1p1	5.2E-260	[Mus musculus][Structural protein][Golgi; Cytoplasmic; Plasma membrane] Amyloid precursor-like protein 1, member of the amyloid precursor protein family of heparin-binding glycoproteins
				Wasco, W. et al. Identification of a mouse brain cDNA that encodes a protein related to the Alzheimer disease-associated amyloid beta protein precursor. Proc. Natl. Acad. Sci. U. S. A. 89, 10758-10762 (1992)
3	7513222CD1	g4519503	1.4E-226	[Homo sapiens] semaphorin W
				Encinas, J. A. et al. Cloning, expression, and genetic mapping of Sema W, a member of the semaphorin family. Proc. Natl. Acad. Sci. U.S.A. 96, 2491-2496 (1999)
3	7513222CD1	341226 SEMA4F	1.1E-227	[Homo sapiens][Plasma membrane] Semaphorin 4F (semaphorin W), a member of the semaphorin family, may be involved in neuronal growth cone guidance
				Encinas, J. A. et al. (1999) supra
3	7513222CD1	609873 Sema4f	7.6E-203	[Rattus norvegicus][Plasma membrane] Semaphorin 4f (semaphorin W), a member of the semaphorin family, involved in neuronal growth cone guidance, induces the collapse of growth cones
				Schultze, W. et al. Semaphorin4F interacts with the synapse-associated protein SAP90/PSD-95. J. Neurochem. 78, 482-489 (2001)
4	7522457CD1	g6136788	3.9E-152	[Mus musculus] synaptotagmin VIII
				Fukuda, M. et al. Conserved N-terminal cysteine motif is essential for homo- and heterodimer formation of synaptotagmins III, V, VI, and X. J. Biol. Chem. 274, 31421-31427 (1999)
4	7522457CD1	608232 Syf8	2.7E-153	[Mus musculus] Synaptotagmin VIII, nonneuronal member of the synaptotagmin family of proteins, binds clathrin-AP2, syntaxins, and the synaptotagmin-binding, RNA-interacting protein Nsap1; functions in calcium-dependent exocytosis

Table 2

Polypeptide SEQ ID NO:	Incye Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
4				Li, C. et al. Ca(2+)-dependent and -independent activities of neural and non-neural synaptotagmins. <i>Nature</i> 375, 594-599 (1995)
4	7522457CD1	756736 Syt8	1.1E-151	[Rattus norvegicus] Synaptotagmin VIII, member of the synaptotagmin family of proteins, functions in calcium-dependent exocytosis and insulin secretion
				Desai, R. C. et al. The C2B domain of synaptotagmin is a Ca(2+)-sensing module essential for exocytosis. <i>J. Cell Biol.</i> 150, 1125-1136 (2000)
5	7518601CD1	g6634716	1.8E-56	[Homo sapiens] galanin-like peptide precursor
				Ohtaki, T. et al. Isolation and cDNA cloning of a novel galanin-like peptide (GALP) from porcine hypothalamus. <i>J. Biol. Chem.</i> 274, 37041-37045 (1999)
5	7518601CD1	732021 LOC85569	1.2E-57	[Homo sapiens] Member of the galanin family of peptide hormones, has high similarity to uncharacterized rat AF188491
				Cunningham, M. J. et al. Cloning and distribution of galanin-like peptide mRNA in the hypothalamus and pituitary of the macaque. <i>Endocrinology</i> 143, 755-763 (2002)
5	7518601CD1	657569 AF188491	1.2E-31	[Rattus norvegicus] Galanin-like peptide precursor, a hypothalamic neuropeptide that helps regulate gonadotropin secretion via release of luteinizing hormone-releasing hormone; may have a role in regulating feeding and reproduction
				Jureus, A. et al. Galanin-like peptide (GALP) is a target for regulation by leptin in the hypothalamus of the rat. <i>Endocrinology</i> 141, 2703-2706 (2000)
6	7520315CD1	g6164831	0.0	[Rattus norvegicus] Robo2
				Brose, K. et al. Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive axon guidance. <i>Cell</i> 96, 795-806 (1999)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
6	7520315CD1	735342 ROBO2	0.0	[Homo sapiens][Receptor (signalling)] Roundabout (axon guidance receptor, Drosophila) homolog 2, a member of the immunoglobulin superfamily that is involved in axon guidance Kidd, T. et al. Roundabout controls axon crossing of the CNS midline and defines a novel subfamily of evolutionarily conserved guidance receptors. Cell 92, 205-215 (1998)
6	7520315CD1	747041 Robo2	0.0	[Rattus norvegicus] Roundabout (axon guidance receptor, Drosophila) homolog 2, may serve as a receptor for Slit ligands which repel axonal outgrowth
				Nguyen Ba-Charvet, K. T. et al. Slit2-Mediated chemorepulsion and collapse of developing forebrain axons. Neuron 22, 463-473 (1999)
7	7525536CD1	g10801596	1.3E-66	[Mus musculus] Doc2gamma
				Fukuda, M. et al. Doc2gamma, a third isoform of double C2 protein, lacking calcium-dependent phospholipid binding activity. Biochem. Biophys. Res. Commun. 276, 626-632 (2000)
7	7525536CD1	624062 Doc2g	8.9E-68	[Mus musculus][Small molecule-binding protein] Double C2 gamma, contains a Munc13-1 interacting domain and two C2 domains but lacks calcium-dependent phospholipid binding activity, may help regulate vesicular trafficking, localizes to the nucleus and is highly expressed in heart
				Fukuda, M. et al. The C2A domain of double C2 protein gamma contains a functional nuclear localization signal. J. Biol. Chem. 276, 24441-24444 (2001)
7	7525536CD1	662807 Doc2a	4.0E-33	[Rattus norvegicus][Small molecule-binding protein] Double C2-like domains alpha, a putative calcium-dependent phospholipid binding protein, involved in calcium-dependent exocytosis
				Orita, S. et al. Doc2: a novel brain protein having two repeated C2-like domains. Biochem. Biophys. Res. Commun. 206, 439-448 (1995)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
8	7520619CD1	g12248382	2.0E-300	SEMB [Homo sapiens]
8	7520619CD1	587311 Sema4a	7.9E-214	[Mus musculus] Semaphorin 4a (Semaphorin B), is in the transmembrane type subfamily of semaphorins which is a family of proteins involved in neuronal axon guidance during neural development
				Williams-Hogarth, L. C. et al. Expression of semaphorins in developing and regenerating olfactory epithelium. J. Comp. Neurol. 423, 565-578 (2000)
8	7520619CD1	657699 FLJ12287	5.1E-104	[Homo sapiens] Protein containing two Sema domains and two plexin repeats, has a region of moderate similarity to a region of semaphorin 4D (mouse Sema4d), which is involved in immunity and is associated with defective B- and T-cell activation upon gene mutation

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7513114CD1	510	signal_cleavage: M1-G19	SPSCAN
			Signal Peptide: M2-G16	HMMER
			Signal Peptide: M1-G19	HMMER
			Signal Peptide: M1-G21	HMMER
			Signal Peptide: M1-G16	HMMER
			Coagulation factor 5/8 C-terminal domain.; G24-C168	HMMER_SMART
			Laminin G domain: L195-V332, G381-A497	HMMER_SMART
			F5/8 type C domain: E28-L165	HMMER_PFAM
			Laminin G domain: F203-A335, F389-N510	HMMER_PFAM
			Coagulation factor 5/8 type C domain (FA58C) IPB000421: Y22-P32, T87-N92, W94-W111, T122-Y166	BLIMPS_BLOCKS
			NEUREXIN IV CONTACTIN ASSOCIATED PROTEIN PRECURSOR SIGNAL PD022247: V332-G486	BLAST_PRODOM
			CONTACTIN ASSOCIATED PROTEIN NEUREXIN IV PD015853: M1-G39	BLAST_PRODOM
			CONTACTIN ASSOCIATED PROTEIN NEUREXIN IV PD151832: G164-S202	BLAST_PRODOM
			GLYCOPROTEIN PRECURSOR SIGNAL FACTOR REPEAT PROTEIN NEUROFILIN CELL DOMAIN COAGULATION PD000875: G31-L165	BLAST_PRODOM
			DISCOIDIN I N-TERMINAL DM00516 A42580 2085-2210: P50-C168 P21956 338-462: P50-C168 P12259 2095-2223: P50-C168 P00451 2221-2347: A49-C168	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1 cont.			Potential Phosphorylation sites: S15 S62 S184 S194 S202 S343 S386 S426 S459 S493 T142 T413	MOTIFS
			Potential Glycosylation sites: N120 N128 N276 N420	MOTIFS
			Coagulation factors 5/8 type C domain (FA58C) signature 1: G60-G89	MOTIFS
			Coagulation factors 5/8 type C domain (FA58C) signature 2: P150-C168	MOTIFS
2	7512261CD1	544	signal_cleavage: M1-A33	SPSCAN
			Signal Peptide: L20-G38	HMMER
			Signal Peptide: G16-P35	HMMER
			Signal Peptide: G16-G38	HMMER
			Signal Peptide: P15-S39	HMMER
			Signal Peptide: M1-G38	HMMER
			Signal Peptide: M1-A33	HMMER
			Signal Peptide: P15-G38	HMMER
			Signal Peptide: R13-A36	HMMER
			amyloid A4: G46-P211	HMMER_SMART
			Amyloid A4 extracellular domain: G46-P211	HMMER_PFAM
			Amyloidogenic glycoprotein (Amyloid A4) IPB001868: S81-H130, V135-A180, C181-P211, E246-A279, T285-M334, L358-A398, L399-K441	BLIMPS_BLOCKS
			Amyloidogenic glycoprotein signatures: Q184-S230	PROFILESKAN
			Amyloid A4 protein precursor signature PR00203: D200-P218, P301-Q324	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
2 cont.			PROTEIN PRECURSOR AMYLOID TRANSMEMBRANE PROTEASE SIGNAL SERINE INHIBITOR A4 GLYCOPROTEIN PD003339: L22-P211	BLAST_PRODUM
			PROTEIN PRECURSOR AMYLOID TRANSMEMBRANE PROTEASE SIGNAL SERINE INHIBITOR A4 GLYCOPROTEIN PD003344: K283-I456	BLAST_PRODUM
			AMYLOID-LIKE PROTEIN PRECURSOR APLP GLYCOPROTEIN TRANSMEMBRANE SIGNAL PD037194: E457-G526	BLAST_PRODUM
			AMYLOID-LIKE PROTEIN PRECURSOR APLP GLYCOPROTEIN TRANSMEMBRANE SIGNAL PD037193: P218-V258	BLAST_PRODUM
			AMYLOIDOGENIC GLYCOPROTEIN EXTRACELLULAR DOMAIN DM02422 P51693 187-619: L188-G526 Q06481 181-761: L188-E268 V284-A497 Q06481 181-761: E241-A279 P05067 342-768: K283-E488	BLAST_DOMO
			AMYLOIDOGENIC GLYCOPROTEIN EXTRACELLULAR DOMAIN DM04632 P51693 22-80: L23-R82	BLAST_DOMO
			Potential Phosphorylation sites: S12 S81 S170 S199 S227 S235 S305 S493 S503 S515 T71 T171 T287 T360 T427	MOTIFS
			Potential Glycosylation sites: N337 N461	MOTIFS
			Amyloidogenic glycoprotein extracellular domain signature: G204-P211	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	7513222CD1	421	signal_cleavage: M1-G34	SPSCAN
			Signal Peptide: T16-G34	HMMER
			Signal Peptide: Q14-G34	HMMER
			Signal Peptide: G13-G34	HMMER
			Signal Peptide: P12-R35	HMMER
			Signal Peptide: M1-G34	HMMER
			Signal Peptide: P12-G34	HMMER
			domain found in Plexins, Semaphorins and Integrins: N357-P408	HMMER_SMART
			Sema domain: Y65-W341	HMMER_PFAM
			SEMAPHORIN PROTEIN PRECURSOR	BLAST_PRODOR
			RECEPTOR KINASE SIGNAL TYROSINE	
			TYROSINE PROTEIN FAMILY HEPATOCYTE	
			PD001844: Y65-E119 E119-M335	
			SEMAPHORIN; FASCICLIN; COLLAPLIN; II;	BLAST_DOMO
			DM01606	
			I48745 I-619: F20-K117 E119-A413	
			I49069 I-646: L22-K117 E119-C407	
			I48744 I-639: L25-G120 P105-C407	
			I48747 I-646: L23-K117 E119-C407	
			Potential Phosphorylation sites: S33 S48 S177 S244 S291 S318 S346 S363 S378 T414 Y299	MOTIFS
			Potential Glycosylation sites: N64 N354	MOTIFS
4	7522457CD1	384	signal_cleavage: M1-C58	SPSCAN
			C2 domain: I120-R200, L245-H329	HMMER_PFAM
			Protein kinase C conserved region 2 (CaIB): E119-W215, R244-A354	HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4			Cytosolic domain: C58-S384	TMHMMER
cont.			Transmembrane domain: W35-C57	
			Non-cytosolic domain: M1-R34	
			C2 domain IPB000008: G149-F171, L245-F297, L311-G331	
			C2 domain signature and profile: L232-K283	PROFILESCAN
			Synaptotagmin signature PR00399: L232-V247, V247-A260, P300-D315	BLIMPS_PRINTS
			C2 domain proteins PF00168: L42-C51, L298-D308, L268-E293	BLIMPS_PFAM
			C2-DOMAIN	BLAST_DOMO
			DM00150 A55417 268-396: Q225-H348	MOTIFS
			DM00150 P29101 268-396: Q225-H348	
			DM00150 P46097 268-396: Q225-H348	
			DM00150 S33318 255-384: A223-H348	
			Potential Phosphorylation Sites: S77, S97, S191, S235, S242, T224, T280, T282, T286	
5	7518601CD1	173	Signal Peptide: M1-A25, M1-A23, M1-T21, M1-A27, M1-A19	HMMER
			signal_cleavage: M1-A19	SPSCAN
			Galanin family IPB001600: G33-E60	BLIMPS_BLOCKS
			Galanin signature PR00273: W34-V46	BLIMPS_PRINTS
			Potential Phosphorylation Sites: S17, S84, S117, S121, S129, T61, T108, Y77	MOTIFS
6	7520315CD1	1469	Signal Peptide: M1-G21	HMMER
			signal_cleavage: M1-G21	SPSCAN
			Fibronectin type III domain: P522-S607, L638-S724, P736-S826	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6 cont.			Immunoglobulin domain: G45-A112, G147-G205, E239-A295, G328-A393, D432-A490	HMMER_PFAM
			Fibronectin type 3 domain: P522-S604, L638-G721, P736-G823	HMMER_SMART
			Immunoglobulin: P37-A129, P139-F222, P231-R311, P320-T409, P424-T506	HMMER_SMART
			Immunoglobulin C-2 Type: S43-G117, A145-G210, L237-G300, A326-G398, A430-G495	HMMER_SMART
			Immunoglobulin V-Type: A241-A295, T330-A393, T434-A490	HMMER_SMART
			I type Ig domains from SCOP: P30-F135, Q137-L228, R229-V317, V318-R415, P416-S513	HMMER_INCY
			V type Ig domains from SCOP (anti: DI33-L228, P314-P418, I419-I512)	HMMER_INCY
			Ig superfamily from SCOP: I33-L130, F135-P225, F227-P314, F316-I419, I420-T497	HMMER_INCY
			Cytosolic domain: R883-L1469	TMHMMER
			Transmembrane domain: A860-W882	
			Non-cytosolic domain: M1-P859	
			TRANSMEMBRANE RECEPTOR ROBO1 DUTT1 PROTEIN PD143949: K963-N1044 P1016-G1254 V824-Q1062 P1194-P1321	BLAST_PRODOM
			TRANSMEMBRANE RECEPTOR ROBO1 DUTT1 PROTEIN PD112973: F1257-G1455	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6 cont.			BASIC FIBROBLAST GROWTH FACTOR RECEPTOR 1 DM01287 A39752 1-814:G147-V204 V325-E507 P18460 1-805:D141-C203 V325-E507 P16092 1-821:E240-G300 V325-E507 S1058-Q1095 A60350 1-800:E240-C293 V325-E507 P955-S1023 P16092 1-821: T164-G210 Potential Phosphorylation Sites: S43, S80, S101, S182, S285, S493, S847, S975, S1007, S1036, S1058, S1156, S1210, S1243, S1260, S1282, S1347, S1355, S1356, S1362, S1383, S1427, T73, T193, T220, T255, T277, T309, T413, T466, T581, T693, T728, T731, T980, T1085, T1370, T1253, T1288, T1385, Y108, Y166, Y201, Y486 Potential Glycosylation Sites: N123, N426, N752, N782, N789, N845, N907, N936, N986, N1057, N1065, N1274, N1447 Cell attachment sequence: R899-D901, R950-D952	BLAST_DOMO
				MOTIFS
				MOTIFS
				MOTIFS
7	7525536CD1	170	C2 domain IPB000008: T104-Y156 C2 domain signature and profile: L88-V142 C2 domain proteins PF00168: T83-L92, L127-E152	BLIMPS_BLOCKS PROFILES CAN BLIMPS_PFAM
			C2-DOMAIN DM00150 JC2473 87-213: S82-R170 P41885 740-871: A84-R170 Q06846 402-526: L85-R170	BLAST_DOMO
			Potential Phosphorylation Sites: S15, T141, T160, T168	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	7520619CD1	602	Signal Peptide: M1-A31	HMMER
			Signal_cleavage: M1-A31	SPSCAN
			Sema domain: F64-N448	HMMER_PFAM
			Cytosolic domain: A545-A602	TMHMMER
			Transmembrane domain: W522-V544	
			Non-cytosolic domain: M1-Y521	
			SEMAPHORIN PROTEIN PRECURSOR RECEPTOR KINASE SIGNAL TYROSINE TYROSINEPROTEIN FAMILY HEPATOCYTE PD001844:L67-P147 V184-C346 K118-T229 R301-T405	BLAST_PRODOME
			SEMAPHORIN B PRECURSOR SEM B SIGNAL IMMUNOGLOBULIN FOLD MULTIGENE FAMILY NEUROGENESIS DEVELOPMENTAL PROTEIN PD107003: M1-D63	BLAST_PRODOME
			SEMAPHORIN B PRECURSOR SEM B SIGNAL IMMUNOGLOBULIN FOLD MULTIGENE FAMILY NEUROGENESIS DEVELOPMENTAL PROTEIN PD116663: I406-A602	BLAST_PRODOME
			SEMAPHORIN; FASCICLIN; COLLAPLIN; II; DM01606 I48745 I-619:E122-T405 I406-L462 M1-L191, Q265-A307 DM01606 A49069 I-646:F64-K118 E133-V412 K407-W432 DM01606 I48747 I-646:L24-K118 E133-V412 P420-L453 DM01606 I48744 I-639:A31-K118 D134-V412	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8			Potential Phosphorylation Sites: S106, S111, S119, S143, S219, S309, S352, S378, S415, S579, S586, S588, T217, T229, T251, T328, T333, T405, T562	MOTIFS
cont.			Potential Glycosylation Sites: N448	MOTIFS
			Leucine zipper pattern: L403-L424	MOTIFS

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
9/7513114CBI/ 4744	1-302, 1-397, 1-481, 1-490, 1-524, 1-557, 1-4744, 123-719, 123-788, 123-820, 123-870, 123-871, 123-876, 123-879, 294-836, 299-796, 358-897, 397-919, 440-544, 509-959, 512-1119, 655-1119, 972-1508, 1046-1585, 1167-1389, 1530-2125, 1537-2103, 1569-1996, 1590-2072, 1592-2086, 1592-2097, 1592-2114, 1608-2232, 1655-2257, 1656-2300, 1659-2147, 1676-1855, 1676-1949, 1721-1945, 1723-2314, 1729-2006, 1755-2398, 1762-2293, 1838-2371, 1838-2424, 1838-2446, 1838-2480, 1893-2526, 1913-2151, 2006-2510, 2023-2267, 2026-2278, 2082-2446, 2091-2340, 2125-2563, 2132-2371, 2132-2737, 2214-2761, 2250-2486, 2256-2778, 2293-2950, 2297-2835, 2301-2920, 2349-2839, 2359-2859, 2364-2922, 2388-2948, 2395-2920, 2404-2927, 2426-2941, 2431-2920, 2434-2687, 2444-3040, 2463-3076, 2481-2732, 2557-2859, 2560-2808, 2587-2714, 2634-3239, 2638-2890, 2677-2925, 2707-2952, 2719-2920, 2819-3103, 2889-3411, 2916-3167, 2916-3176, 2952-3262, 2953-3817, 2963-3817, 2965-3492, 2977-3817, 2985-3817, 3021-3817, 3022-3817, 3026-3817, 3051-3758, 3059-3640, 3083-3586, 3086-3817, 3141-3817, 3175-3432, 3205-3976, 3252-3976, 3283-4000, 3386-3625, 3408-3667, 3464-3707, 3470-3770, 3476-4000, 3480-4084, 3499-3758, 3550-4118, 3563-3852, 3572-3817, 3661-3926, 3671-3921, 3671-3930, 3682-3912, 3682-3917, 3689-3934, 3696-3975, 3729-3867, 3775-4063, 3830-4017, 3841-4314, 3856-4113, 3931-4585, 4000-4251, 4002-4243, 4018-4283, 4026-4610, 4042-4495, 4072-4355, 4072-4365, 4086-4534, 4120-4643, 4138-4384, 4138-4476, 4155-4433, 4157-4418, 4159-4426, 4166-4433, 4167-4453, 4176-4514, 4194-4299, 4206-4469, 4216-4506, 4217-4423, 4237-4534, 4243-4528, 4255-4540, 4302-4550, 4315-4540, 4327-4536, 4331-4576, 4331-4581, 4331-4594, 4337-4583, 4338-4580, 4347-4592, 4356-4643, 4372-4604, 4378-4607, 4381-4635
10/7512261CBI/ 2294	1-481, 2-558, 2-594, 6-621, 7-622, 9-650, 13-2286, 15-659, 23-591, 23-659, 24-654, 27-622, 31-716, 33-520, 37-532, 43-521, 43-524, 43-552, 43-561, 43-583, 43-586, 43-595, 43-596, 43-628, 43-636, 43-655, 43-687, 43-703, 43-771, 48-691, 58-642, 67-711, 69-331, 80-823, 133-902, 141-346, 167-1034, 186-456, 202-433, 223-725, 232-886, 244-735, 245-394, 246-455, 246-501, 260-497, 260-679, 261-479, 261-673, 261-810, 300-523, 321-577, 350-554, 350-631, 359-606, 359-613, 360-612, 381-668, 383-967, 386-1015, 392-1076, 417-1166, 429-658, 431-702, 441-988, 451-708, 463-1085, 468-1157, 480-917, 484-839, 498-1147, 499-868, 510-851, 520-766, 522-1187, 546-1175, 550-890, 559-818, 561-1046, 565-715, 567-839, 573-1132, 578-978, 578-1099, 578-1166, 578-1220, 578-1247, 579-947, 581-962, 586-839, 601-1025, 607-1208, 607-1215, 613-1280, 615-1177, 624-1166, 639-1328, 650-1106, 654-905, 662-1059, 668-922, 677-961, 685-1208, 699-824, 699-1289, 700-997, 715-1040, 720-981, 730-1020, 760-976, 767-1052, 768-1324, 776-1020, 776-1076,

Table 4

Polynucleotide SEQ ID NO: / Incyte ID/ Sequence Length	Sequence Fragments
10 cont.	777-1042, 778-1052, 778-1346, 779-1342, 797-1198, 797-1422, 802-1089, 815-1490, 821-1094, 821-1228, 821-1265, 824-1401, 825-1110, 826-1381, 827-1420, 829-1162, 830-1083, 836-1111, 854-1144, 859-1410, 868-1500, 869-1125, 871-1328, 885-1067, 896-1138, 900-1023, 900-1211, 900-1224, 915-1141, 919-1142, 920-1407, 920-1452, 925-1124, 930-1137, 930-1138, 935-1091, 940-1186, 944-1430, 953-1180, 953-1208, 959-1630, 960-1187, 960-1545, 961-1106, 961-1552, 967-1231, 967-1235, 978-1218, 978-1287, 980-1298, 986-1184, 996-1262, 1001-1439, 1004-1248, 1010-1201, 1010-1246, 1010-1276, 1015-1198, 1015-1461, 1032-1204, 1039-1337, 1040-1273, 1040-1507, 1041-1301, 1043-1479, 1049-1552, 1050-1285, 1054-1282, 1055-1312, 1057-1209, 1057-1340, 1060-1450, 1062-1295, 1070-1205, 1070-1344, 1076-1486, 1076-1623, 1086-1324, 1088-1449, 1095-1630, 1124-1466, 1124-1551, 1124-1575, 1124-1630, 1129-1605, 1148-1410, 1150-1274, 1158-1407, 1158-1630, 1173-1463, 1195-1438, 1196-1501, 1205-1434, 1209-1408, 1214-1465, 1214-1491, 1229-1529, 1232-1404, 1239-1478, 1243-1562, 1253-1529, 1301-1561, 1313-1581, 1313-1585, 1323-1579, 1323-1580, 1327-1580, 1327-1591, 1327-1597, 1328-1605, 1333-1510, 1341-1605, 1348-1630, 1363-1571, 1363-1630, 1369-1501, 1369-1630, 1370-1626, 1406-1566, 1477-1740, 1626-1719, 1626-1820, 1626-1846, 1626-1850, 1626-1851, 1626-1862, 1626-1869, 1626-1874, 1626-1877, 1626-1885, 1626-1892, 1626-1898, 1626-1904, 1626-1907, 1626-1976, 1626-2125, 1626-2132, 1626-2185, 1626-2216, 1626-2281, 1627-1887, 1628-2176, 1630-1801, 1630-2294, 1631-1883, 1632-2012, 1635-1899, 1637-2182, 1638-1913, 1638-2128, 1640-2019, 1642-1902, 1642-1911, 1642-1916, 1646-1880, 1648-1941, 1649-1867, 1649-1889, 1649-1897, 1649-1946, 1649-1965, 1650-1940, 1650-2265, 1651-1985, 1654-2282, 1654-2287, 1657-1847, 1657-1903, 1657-1942, 1657-1949, 1658-1892, 1661-2285, 1663-2285, 1669-2234, 1670-1906, 1671-1904, 1671-2160, 1671-2212, 1673-2130, 1674-2246, 1682-2216, 1683-1994, 1690-1970, 1690-2286, 1691-2224, 1696-2262, 1698-1906, 1700-1887, 1707-1962, 1708-2228,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
10 cont.	1711-1976, 1713-2182, 1730-2244, 1734-2254, 1734-2259, 1734-2285, 1735-2220, 1741-2133, 1749-2023, 1753-2000, 1758-2286, 1763-2283, 1765-2032, 1768-2037, 1773-2283, 1781-2001, 1781-2182, 1781-2247, 1782-2013, 1784-2286, 1789-1918, 1792-2077, 1796-2079, 1803-2051, 1803-2077, 1810-2074, 1810-2285, 1811-2286, 1813-2283, 1813-2285, 1813-2294, 1822-2283, 1825-1942, 1825-2124, 1827-2217, 1827-2270, 1827-2286, 1828-2286, 1835-2286, 1843-2137, 1846-2283, 1847-2146, 1847-2285, 1856-2242, 1856-2286, 1860-2216, 1860-2286, 1863-2284, 1866-2283, 1867-2285, 1882-2283, 1889-2028, 1891-2286, 1893-2282, 1894-2286, 1896-2284, 1900-2284, 1920-2286, 1923-2208, 1925-2124, 1926-2287, 1927-2197, 1928-2083, 1930-2144, 1931-2201, 1931-2285, 1932-2163, 1938-2130, 1941-2288, 1946-2173, 1964-2241, 1971-2276, 1981-2248, 1981-2285, 1992-2282, 1993-2080, 1995-2286, 1996-2282, 1996-2284, 1997-2062, 2004-2238, 2005-2282, 2008-2294, 2013-2165, 2021-2282, 2021-2283, 2021-2285, 2022-2286, 2027-2094, 2027-2286, 2027-2293, 2032-2290, 2040-2242,
	2040-2247, 2044-2284, 2050-2294, 2057-2283, 2058-2287, 2062-2283, 2063-2280, 2077-2285, 2084-2288, 2090-2253, 2093-2286, 2097-2286, 2104-2290, 2123-2229, 2123-2285, 2123-2293, 2127-2283, 2127-2294, 2144-2286, 2149-2276, 2154-2286
11/7513222CB1/ 2067	1-684, 1-773, 1-2067, 114-366, 230-373, 372-923, 373-1079, 436-746, 453-984, 481-1077, 481-1078, 504-779, 524-1077, 538-1033, 543-1055, 547-1074, 567-643, 568-1077, 571-1077, 572-1077, 575-1077, 579-889, 579-1077, 584-1077, 586-922, 586-1049, 586-1132, 586-1202, 590-1077, 640-1077, 681-900, 681-921, 681-1125, 681-1184, 681-1194, 681-1219, 681-1250, 685-1252, 723-1077, 907-1041, 1126-1965, 1247-1958, 1272-1965, 1290-1965, 1297-1505, 1313-1965, 1342-2007, 1355-2062, 1389-2011, 1390-1999, 1393-2067, 1397-1816, 1478-2045, 1485-2013, 1522-2067, 1531-2067, 1597-2066, 1599-2067, 1612-2067, 1633-2067, 1635-1924, 1661-2062, 1665-2067, 1666-2067, 1711-2067, 1748-2063, 1768-2066, 1781-2067, 1820-2051, 1854-2067, 1999-2067
12/7522457CB1 1209	1-688, 22-1188, 638-1209
13/7518601CB1/ 635	1-635

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
14/7520315CB1/ 4453	1-658, 1-696, 1-700, 1-728, 1-764, 1-783, 2-4452, 334-1283, 337-1060, 684-1418, 684-1468, 684-1632, 709-1243, 713-1612, 717-1559, 765-1578, 767-1647, 1111-1826, 1114-1958, 1319-2115, 1332-2151, 1475-2385, 1563-2441, 1565-2343, 1693-2419, 1929-2647, 1934-2791, 2111-2843, 2111-2868, 2339-3197, 2342-3105, 2344-3063, 2398- 3172, 2709-3517, 2711-3708, 2990-3864, 2995-3874, 3096-3866, 3191-3875, 3301-4181, 3315-4002, 3603-4453, 3632-4453, 3779-4453, 3780-4453, 3783-4453, 4159-4453, 4161-4453, 4341-4453, 4353-4453
15/7525536CB1/ 1937	1-557, 1-585, 414-1329, 491-1328, 494-1329, 599-1329, 610-1328, 620-1328, 667-1328, 714-1328, 1232-1937, 1439-1702
16/7520619CB1/ 2608	1-803, 1-2608, 502-1441, 616-1441, 1184-2092, 1213-2075, 1250-2084, 1264-2070, 1283-2092, 1308-2111, 1337- 2078, 1369-2071, 1451-2069, 1453-2103, 1860-2608, 1882-2608

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
9	7513114CB1	HEAONOE01
10	7512261CB1	CORPNOT02
11	7513222CB1	PLACFER01
15	7525536CB1	TLYTXF03

Table 6

Library	Vector	Library Description
CORPNOT02	pINCY	Library was constructed using RNA isolated from diseased corpus callosum tissue removed from the brain of a 74-year-old Caucasian male who died from Alzheimer's disease.
HEAONOE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from the aorta of a 39-year-old Caucasian male, who died from a gunshot wound. Serology was positive for cytomegalovirus (CMV). Patient history included tobacco abuse (one pack of cigarettes per day for 25 years), and occasionally cocaine, marijuana, and alcohol use.
PLACFER01	pINCY	The library was constructed using RNA isolated from placental tissue removed from a Caucasian fetus, who died after 16 weeks' gestation from fetal demise and hydrocephalus. Patient history included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV. Family history included multiple pregnancies and live births, and an abortion.
TLYTJTF03	pRARE	This 5' cap isolated full-length library was constructed using RNA isolated from a treated Jurkat cell line derived from the T cells of a male. The cells were treated with 10ng/mL of anti-CD3 for 5 minutes. The cells were then fractionated to obtain the nuclei. Patient history included acute T-cell leukemia.

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.	
SPSscan	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
9	7513114	1233035H1	SNP00015848	172	4509	C	C	A	noncoding	n/a	n/a	n/a	n/a
9	7513114	1364738H1	SNP00015847	177	4333	C	C	T	noncoding	n/a	n/a	n/a	n/a
9	7513114	2460602H1	SNP00143047	218	2467	G	G	A	noncoding	n/a	n/a	n/a	n/a
9	7513114	2475872H1	SNP00115188	8	2139	T	T	C	noncoding	n/a	n/a	n/a	n/a
9	7513114	3423308T6	SNP00143048	121	3330	C	C	A	noncoding	n/a	n/a	n/a	n/a
9	7513114	3596632H1	SNP00075055	225	225	A	C	A	R51	0.71	n/a	n/a	n/a
9	7513114	7222848H1	SNP00115187	24	1141	T	T	C	L356	n/d	n/a	n/a	n/a
10	7512261	1269660H1	SNP00053532	59	1728	G	A	G	noncoding	n/d	n/a	n/a	n/a
10	7512261	1352819F6	SNP00123885	224	483	A	A	G	K144	n/a	n/a	n/a	n/a
10	7512261	1352819H1	SNP00123885	224	484	A	A	G	E144	n/a	n/a	n/a	n/a
10	7512261	1414310H1	SNP00053329	38	1556	G	G	A	G501	n/d	n/a	n/a	n/a
10	7512261	1414310H1	SNP00123380	71	1589	T	T	C	P512	n/a	n/a	n/a	n/a
10	7512261	1482250H1	SNP00129061	12	787	A	A	G	E245	n/a	n/a	n/a	n/a
10	7512261	1484663F6	SNP00029668	331	1373	G	G	C	E440	n/a	n/a	n/a	n/a
10	7512261	1484663T6	SNP00053532	510	1733	G	A	G	noncoding	n/d	n/a	n/a	n/a
10	7512261	383844R1	SNP00053330	332	1944	T	T	C	noncoding	n/d	0.34	0.09	0.06
10	7512261	5996926H1	SNP00123886	115	660	C	C	T	R203	n/d	n/d	n/d	n/d
10	7512261	7360054H1	SNP00123885	251	455	A	A	G	Q134	n/a	n/a	n/a	n/a
10	7512261	7457902H1	SNP00029668	32	1363	G	G	C	G437	n/a	n/a	n/a	n/a
10	7512261	7457902H1	SNP00053329	215	1546	G	G	A	R498	n/d	n/a	n/a	n/a
10	7512261	7457902H1	SNP00123380	248	1579	T	T	C	V509	n/a	n/a	n/a	n/a
10	7512261	7676541H1	SNP00129061	128	788	A	A	G	E245	n/a	n/a	n/a	n/a
11	7513222	2504535F6	SNP00154755	36	1855	T	T	C	noncoding	n/a	n/a	n/a	n/a
11	7513222	2581272T6	SNP00154755	158	1856	T	T	C	noncoding	n/a	n/a	n/a	n/a
12	7522457	1216677H1	SNP00035110	86	1096	C	C	T	R356	n/d	n/a	n/a	0.39
12	7522457	7256158H2	SNP00043145	489	414	G	G	A	A128	0.81	n/a	n/a	n/a
14	7520315	3082674H1	SNP00031233	160	3821	C	C	T	T1271	n/d	n/a	n/a	n/a
16	7520619	112902T6	SNP00008723	148	2420	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
16	7520619	112902T6	SNP00151397	352	2216	C	C	T	noncoding	n/a	n/a	n/a	n/a
16	7520619	1388790H1	SNP00008723	116	2419	T	C	T	noncoding	n/a	n/a	n/a	n/a
16	7520619	2191204H1	SNP00151397	22	2215	C	C	T	noncoding	n/a	n/a	n/a	n/a
16	7520619	2294056T6	SNP00008723	108	2456	C	C	T	noncoding	n/a	n/a	n/a	n/a
16	7520619	3115208T6	SNP00008723	130	2431	C	C	T	noncoding	n/a	n/a	n/a	n/a
16	7520619	636350T6	SNP00008723	41	2548	C	C	T	noncoding	n/a	n/a	n/a	n/a

What is claimed is:

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

- 5 a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8,
- 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:5 and SEQ ID NO:7,
- 10 c) a polypeptide comprising a naturally occurring amino acid sequence at least 97% identical to the amino acid sequence of SEQ ID NO:2,
- d) a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1 and SEQ ID NO:4,
- 15 e) a polypeptide consisting essentially of a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:6 and SEQ ID NO:8,
- f) a polypeptide comprising a naturally occurring amino acid sequence at least 98% 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:1-8, and
- 20 h) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the

25 group consisting of SEQ ID NO:1-8.

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:9-16.

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

7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
- 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:1-8.
- 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:9-16,
- 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:13 and SEQ ID NO:15,
- c) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-12, SEQ ID NO:14, and SEQ ID NO:16,
- 25 d) a polynucleotide complementary to a polynucleotide of a),
- e) a polynucleotide complementary to a polynucleotide of b),
- f) a polynucleotide complementary to a polynucleotide of c), and
- 30 g) an RNA equivalent of a)-f).
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
- 35 having a sequence of a polynucleotide of claim 12, the method comprising:

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

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

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

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

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

18. A composition of claim 17, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

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

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

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

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

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

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

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

10 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 NTRAN, comprising administering to a patient in need of such treatment a composition of claim 24.

15

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
- 20 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:

- 25 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
30 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.

35 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) contacting a sample comprising the target polynucleotide with a compound, under conditions suitable for the expression of the target polynucleotide,
- 5 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 screening for potential toxicity of a test compound, the method comprising:

- 10 a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a
- 15 polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample
- 20 indicates potential toxicity of the test compound.

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

- a) combining the biological sample with an antibody of claim 11, under conditions
- 25 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:

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

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

33. A method of diagnosing a condition or disease associated with the expression of NTRAN in a subject, comprising administering to said subject an effective amount of the composition of claim 5 32.

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

35. A method of diagnosing a condition or disease associated with the expression of NTRAN in a subject, comprising administering to said subject an effective amount of the composition of claim 10 34.

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

- 15 a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from the animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a 20 polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

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

25 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:

- 30 a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- 35 d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which specifically binds to a

polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

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

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

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

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

44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and
- 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:1-8 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

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

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

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled

polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and

- c) quantifying the expression of the polynucleotides in the sample.

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

10

 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.

15

 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.

20

 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.

25

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

30

 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.

35

 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.

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

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

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

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

10

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

64. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:9.

15 65. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:10.

66. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:11.

20

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

25 68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:13.

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:14.

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

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

35

<110> INCYTE CORPORATION; RAMKUMAR, Jayalaxmi;
EMERLING, Brooke M.; ELLIOTT, Vicki S.;
BECHA, Shanya D.; CHAWLA, Narinder K.;
BAUGHN, Mariah R.;

<120> NEUROTRANSMISSION-ASSOCIATED PROTEINS

<130> PF-1488 PCT

<140> To Be Assigned

<141> Herewith

<150> US 60/453,108

<151> 2003-03-07

<150> US 60/471,930

<151> 2003-05-19

<150> US 60/490,749

<151> 2003-07-28

<160> 16

<170> PERL Program

<210> 1

<211> 510

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7513114CD1

<400> 1

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

	215		220		225
Glu Leu Glu Gly	Ala His Leu Leu Leu	His Met Ser Leu Gly	Ser		
	230		235		240
Ser Pro Ile Gln	Pro Arg Pro Gly His	Thr Thr Val Ser Ala	Gly		
	245		250		255
Gly Val Leu Asn	Asp Gln His Trp His	Tyr Val Arg Val Asp	Arg		
	260		265		270
Phe Gly Arg Asp	Val Asn Phe Thr Leu	Asp Gly Tyr Val Gln	Arg		
	275		280		285
Phe Ile Leu Asn	Gly Asp Phe Glu Arg	Leu Asn Leu Asp Thr	Glu		
	290		295		300
Met Phe Ile Gly	Gly Leu Val Gly Ala	Ala Arg Lys Asn Leu	Ala		
	305		310		315
Tyr Arg His Asn	Phe Arg Gly Cys Ile	Glu Asn Val Ile Phe	Asn		
	320		325		330
Arg Val Asn Ile	Ala Gly Leu Ala Val	Arg Arg His Ser Arg	Ile		
	335		340		345
Thr Phe Glu Gly	Lys Val Ala Phe Arg	Cys Leu Asp Pro Val	Pro		
	350		355		360
His Pro Ile Asn	Phe Gly Gly Pro His	Asn Phe Val Gln Val	Pro		
	365		370		375
Gly Phe Pro Arg	Arg Gly Arg Leu Ala	Val Ser Phe Arg Phe	Arg		
	380		385		390
Thr Trp Asp Leu	Thr Gly Leu Leu Leu	Phe Ser Arg Leu Gly	Asp		
	395		400		405
Gly Leu Gly His	Val Glu Leu Thr Leu	Ser Glu Gly Gln Val	Asn		
	410		415		420
Val Ser Ile Ala	Gln Ser Gly Arg Lys	Lys Leu Gln Phe Ala	Ala		
	425		430		435
Gly Tyr Arg Leu	Asn Asp Gly Phe Trp	His Glu Val Asn Phe	Val		
	440		445		450
Ala Gln Glu Asn	His Ala Val Ile Ser	Ile Asp Asp Val Glu	Gly		
	455		460		465
Ala Glu Val Arg	Val Ser Tyr Pro Leu	Leu Ile Arg Thr Gly	Thr		
	470		475		480
Ser Tyr Phe Phe	Gly Gly Ala Ala Leu	Thr Cys Val Ser Met	Met		
	485		490		495
Asp Ala Ala Thr	Ser Leu Gly Met Thr	Ser Phe Ala Thr Ala	Asn		
	500		505		510

<210> 2

<211> 544

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7512261CD1

<400> 2

Met Gly Pro Ala Ser	Pro Ala Ala Arg Gly	Leu Ser Arg Arg Pro
1	5	10
Gly Gln Pro Pro Leu	Pro Leu Leu Leu Pro	Leu Leu Leu Leu Leu
	20	25
Leu Arg Ala Gln Pro	Ala Ile Gly Ser Leu	Ala Gly Gly Ser Pro
	35	40
Gly Ala Ala Glu Ala	Pro Gly Ser Ala Gln	Val Ala Gly Leu Cys
	50	55
Gly Arg Leu Thr Leu	His Arg Asp Leu Arg	Thr Gly Arg Trp Glu
	65	70
Pro Asp Pro Gln Arg	Ser Arg Arg Cys Leu	Arg Asp Pro Gln Arg
	80	85
		90

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

<210> 3
 <211> 421
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7513222CD1

<400> 3
 Met Pro Ala Ser Ala Ala Arg Pro Arg Pro Gly Pro Gly Gln Pro
 1 5 10 15
 Thr Ala Ser Pro Phe Pro Leu Leu Leu Leu Ala Val Leu Ser Gly
 20 25 30
 Pro Val Ser Gly Arg Val Pro Arg Ser Val Pro Arg Thr Ser Leu
 35 40 45
 Pro Ile Ser Glu Ala Asp Ser Cys Leu Thr Arg Phe Ala Val Pro
 50 55 60
 His Thr Tyr Asn Tyr Ser Val Leu Leu Val Asp Pro Ala Ser His
 65 70 75
 Thr Leu Tyr Val Gly Ala Arg Asp Thr Ile Phe Ala Leu Ser Leu
 80 85 90
 Pro Phe Ser Gly Glu Arg Pro Arg Arg Ile Asp Trp Met Val Pro
 95 100 105
 Glu Ala His Arg Gln Asn Cys Arg Lys Lys Gly Lys Lys Glu Gly
 110 115 120
 Asp Leu Gly Gly Arg Lys Thr Leu Gln Gln Arg Trp Thr Thr Phe
 125 130 135
 Leu Lys Ala Asp Leu Leu Cys Pro Gly Pro Glu His Gly Arg Ala
 140 145 150
 Ser Ser Val Leu Gln Asp Val Ala Val Leu Arg Pro Glu Leu Gly
 155 160 165
 Ala Gly Thr Pro Ile Phe Tyr Gly Ile Phe Ser Ser Gln Trp Glu
 170 175 180
 Gly Ala Thr Ile Ser Ala Val Cys Ala Phe Arg Pro Gln Asp Ile
 185 190 195
 Arg Thr Val Leu Asn Gly Pro Phe Arg Glu Leu Lys His Asp Cys
 200 205 210
 Asn Arg Gly Leu Pro Val Val Asp Asn Asp Val Pro Gln Pro Arg
 215 220 225
 Pro Gly Glu Cys Ile Thr Asn Asn Met Lys Leu Arg His Phe Gly
 230 235 240
 Ser Ser Leu Ser Leu Pro Asp Arg Val Leu Thr Phe Ile Arg Asp
 245 250 255
 His Pro Leu Met Asp Arg Pro Val Phe Pro Ala Asp Gly His Pro
 260 265 270
 Leu Leu Val Thr Thr Asp Thr Ala Tyr Leu Arg Val Val Ala His
 275 280 285
 Arg Val Thr Ser Leu Ser Gly Lys Glu Tyr Asp Val Leu Tyr Leu
 290 295 300
 Gly Thr Glu Asp Gly His Leu His Arg Ala Val Arg Ile Gly Ala
 305 310 315
 Gln Leu Ser Val Leu Glu Asp Leu Ala Leu Phe Pro Glu Pro Gln
 320 325 330
 Pro Val Glu Asn Met Lys Leu Tyr His Ser Trp Leu Leu Val Gly
 335 340 345
 Ser Arg Thr Glu Val Thr Gln Val Asn Thr Thr Asn Cys Gly Arg
 350 355 360
 Leu Gln Ser Cys Ser Glu Cys Ile Leu Ala Gln Asp Pro Val Cys
 365 370 375
 Ala Trp Ser Phe Arg Leu Asp Glu Cys Val Ala His Ala Gly Glu
 380 385 390
 His Arg Gly Leu Val Gln Asp Ile Glu Ser Ala Asp Val Ser Ser

				395						400					405
Leu	Cys	Pro	Lys	Glu	Pro	Gly	Ala	Thr	Ala	Thr	Glu	Gly	Thr	Ser	
				410					415					420	
Gly															

<210> 4
 <211> 384
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7522457CD1

<400> 4
 Met Gly His Pro Pro Val Ser Pro Ser Ala Pro Ala Pro Ala Gly
 1 5 10 15
 Thr Thr Ala Ile Pro Gly Leu Ile Pro Asp Leu Val Ala Gly Thr
 20 25 30
 Pro Trp Pro Arg Trp Ala Leu Ile Ala Gly Ala Leu Ala Ala Gly
 35 40 45
 Val Leu Leu Val Ser Cys Leu Leu Cys Ala Ala Cys Cys Cys Cys
 50 55 60
 Arg Arg His Arg Lys Lys Pro Arg Asp Lys Glu Ser Val Gly Leu
 65 70 75
 Gly Ser Ala Arg Gly Thr Thr Thr Thr His Leu Val Gln Pro Asp
 80 85 90
 Val Asp Gly Leu Glu Ser Ser Pro Gly Asp Ala Gln Gln Trp Gly
 95 100 105
 Arg Leu Gln Leu Ser Leu Glu Phe Asp Phe Gly Ser Gln Glu Ile
 110 115 120
 Arg Val Gly Leu Arg Gln Ala Ala Asp Leu Arg Pro Gly Gly Thr
 125 130 135
 Val Asp Pro Tyr Ala Arg Val Ser Val Ser Thr Gln Ala Gly His
 140 145 150
 Arg His Glu Thr Lys Val His Arg Gly Thr Leu Cys Pro Val Phe
 155 160 165
 Asp Glu Thr Cys Cys Phe His Ala Glu Leu Pro Gly Ala Thr Leu
 170 175 180
 Gln Val Gln Leu Phe Asn Phe Lys Arg Phe Ser Gly His Glu Pro
 185 190 195
 Leu Gly Glu Leu Arg Leu Pro Leu Gly Thr Val Asp Leu Gln His
 200 205 210
 Val Leu Glu His Trp Tyr Leu Leu Gly Pro Pro Ala Ala Thr Gln
 215 220 225
 Pro Glu Gln Val Gly Glu Leu Cys Phe Ser Leu Arg Tyr Val Pro
 230 235 240
 Ser Ser Gly Arg Leu Thr Val Val Val Leu Glu Ala Arg Gly Leu
 245 250 255
 Arg Pro Gly Leu Ala Glu Pro Tyr Val Lys Val Gln Leu Met Leu
 260 265 270
 Asn Gln Arg Lys Trp Lys Lys Arg Lys Thr Ala Thr Lys Lys Gly
 275 280 285
 Thr Ala Ala Pro Tyr Phe Asn Glu Ala Phe Thr Phe Leu Val Pro
 290 295 300
 Phe Ser Gln Val Gln Asn Val Asp Leu Val Leu Ala Val Trp Asp
 305 310 315
 Arg Ser Leu Pro Leu Arg Thr Glu Pro Val Gly Lys Val His Leu
 320 325 330
 Gly Ala Arg Ala Ser Gly Gln Pro Leu Gln His Trp Ala Asp Met
 335 340 345
 Leu Ala His Ala Arg Arg Pro Ile Ala Gln Arg His Pro Leu Arg

	350		355		360
Pro Ala Arg Glu Val Asp Arg Met Leu Ala Leu Gln Pro Arg Leu					
	365		370		375
Arg Leu Arg Leu Pro Leu Pro His Ser					
	380				

<210> 5
 <211> 173
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7518601CD1

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

<210> 6
 <211> 1469
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7520315CD1

<400> 6
 Met Ser Leu Leu Met Phe Thr Gln Leu Leu Leu Cys Gly Phe Leu
 1 5 10 15
 Tyr Val Arg Val Asp Gly Ser Arg Leu Arg Gln Glu Asp Phe Pro
 20 25 30
 Pro Arg Ile Val Glu His Pro Ser Asp Val Ile Val Ser Lys Gly
 35 40 45
 Glu Pro Thr Thr Leu Asn Cys Lys Ala Glu Gly Arg Pro Thr Pro
 50 55 60
 Thr Ile Glu Trp Tyr Lys Asp Gly Glu Arg Val Glu Thr Asp Lys
 65 70 75
 Asp Asp Pro Arg Ser His Arg Met Leu Leu Pro Ser Gly Ser Leu
 80 85 90

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

Val Ala Asn His	560	Val Lys Thr Thr Leu Tyr Thr Val Arg Gly Leu	565	570
Arg Pro Asn Thr	575	Ile Tyr Leu Phe Met Val Arg Ala Ile Asn Pro	580	585
Gln Gly Leu Ser	590	Asp Pro Ser Pro Met Ser Asp Pro Val Arg Thr	595	600
Gln Asp Ile Ser	605	Pro Pro Ala Gln Gly Val Asp His Arg Gln Val	610	615
Gln Lys Glu Leu	620	Gly Asp Val Leu Val Arg Leu His Asn Pro Val	625	630
Val Leu Thr Pro	635	Thr Thr Val Gln Ala Thr Trp Thr Val Asp Arg	640	645
Gln Pro Gln Phe	650	Ile Gln Gly Tyr Arg Val Met Tyr Arg Gln Thr	655	660
Ser Gly Leu Gln	665	Ala Thr Ser Ser Trp Gln Asn Leu Asp Ala Lys	670	675
Val Pro Thr Glu	680	Arg Ser Ala Val Leu Val Asn Leu Lys Lys Gly	685	690
Val Thr Tyr Glu	695	Ile Lys Val Arg Pro Tyr Phe Asn Glu Phe Gln	700	705
Gly Met Asp Ser	710	Glu Ser Lys Thr Val Arg Thr Thr Glu Glu Ala	715	720
Pro Ser Ala Pro	725	Pro Gln Ser Val Thr Val Leu Thr Val Gly Ser	730	735
Tyr Asn Ser Thr	740	Ser Ile Ser Val Ser Trp Asp Pro Pro Pro Pro	745	750
Asp His Gln Asn	755	Gly Ile Ile Gln Glu Tyr Lys Ile Trp Cys Leu	760	765
Gly Asn Glu Thr	770	Arg Phe His Ile Asn Lys Thr Val Asp Ala Ala	775	780
Ile Arg Ser Val	785	Ile Ile Gly Gly Leu Phe Pro Gly Ile Gln Tyr	790	795
Arg Val Glu Val	800	Ala Ala Ser Thr Ser Ala Gly Val Gly Val Lys	805	810
Ser Glu Pro Gln	815	Pro Ile Ile Ile Gly Arg Arg Asn Glu Val Val	820	825
Ile Thr Glu Asn	830	Asn Asn Ser Ile Thr Glu Gln Ile Thr Asp Val	835	840
Val Lys Gln Pro	845	Ala Phe Ile Ala Gly Ile Gly Gly Ala Cys Trp	850	855
Val Ile Leu Met	860	Gly Phe Ser Ile Trp Leu Tyr Trp Arg Arg Lys	865	870
Lys Arg Lys Gly	875	Leu Ser Asn Tyr Ala Val Thr Phe Gln Arg Gly	880	885
Asp Gly Gly Leu	890	Met Ser Asn Gly Ser Arg Pro Gly Leu Leu Asn	895	900
Ala Gly Asp Pro	905	Ser Tyr Pro Trp Leu Ala Asp Ser Trp Pro Ala	910	915
Thr Ser Leu Pro	920	Val Asn Asn Ser Asn Ser Gly Pro Asn Glu Ile	925	930
Gly Asn Phe Gly	935	Arg Gly Asp Val Leu Pro Pro Val Pro Gly Gln	940	945
Gly Asp Lys Thr	950	Ala Thr Met Leu Ser Asp Gly Ala Ile Tyr Ser	955	960
Ser Ile Asp Phe	965	Thr Thr Lys Thr Ser Tyr Asn Ser Ser Ser Gln	970	975
Ile Thr Gln Ala	980	Thr Pro Tyr Ala Thr Thr Gln Ile Leu His Ser	985	990
Asn Ser Ile His	995	Glu Leu Ala Val Asp Leu Pro Asp Pro Gln Trp	1000	1005
Lys Ser Ser Ile	1010	Gln Gln Lys Thr Asp Leu Met Gly Phe Gly Tyr	1015	1020
	1025		1030	1035

```

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

```

<210> 7

<211> Thr

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7525536CD1

<400> 7

```

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

```

<210> 8

<211> 602

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7520619CD1

<400> 8

```

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

```

	170		175		180
Ser Gln Pro Val	Leu Lys Thr Asp Asn	Phe Leu Arg Trp Leu	His		
	185		190		195
His Asp Ala Pro	Phe Val Ala Ala Ile	Pro Ser Thr Gln Val	Val		
	200		205		210
Tyr Phe Phe Phe	Glu Glu Thr Ala Ser	Glu Phe Asp Phe Phe	Glu		
	215		220		225
Arg Leu His Thr	Ser Arg Val Ala Arg	Val Cys Lys Asn Asp	Val		
	230		235		240
Gly Gly Glu Lys	Leu Leu Gln Lys Lys	Trp Thr Thr Phe Leu	Lys		
	245		250		255
Ala Gln Leu Leu	Cys Thr Gln Pro Gly	Gln Leu Pro Phe Asn	Val		
	260		265		270
Ile Arg His Ala	Val Leu Leu Pro Ala	Asp Ser Pro Thr Ala	Pro		
	275		280		285
His Ile Tyr Ala	Val Phe Thr Ser Gln	Trp Gln Val Gly Gly	Thr		
	290		295		300
Arg Ser Ser Ala	Val Cys Ala Phe Ser	Leu Leu Asp Ile Glu	Arg		
	305		310		315
Val Phe Lys Gly	Lys Tyr Lys Glu Leu	Asn Lys Glu Thr Ser	Arg		
	320		325		330
Trp Thr Thr Tyr	Arg Gly Pro Glu Thr	Asn Pro Arg Pro Gly	Ser		
	335		340		345
Cys Ser Val Gly	Pro Ser Ser Asp Lys	Ala Leu Thr Phe Met	Lys		
	350		355		360
Asp His Phe Leu	Met Asp Glu Gln Val	Val Gly Thr Pro Leu	Leu		
	365		370		375
Val Lys Ser Gly	Val Glu Tyr Thr Arg	Leu Ala Val Glu Thr	Ala		
	380		385		390
Gln Gly Leu Asp	Gly His Ser His Leu	Val Met Tyr Leu Gly	Thr		
	395		400		405
Ile Lys Glu Val	Leu Ala Val Pro Asn	Ser Ile Leu Glu Leu	Pro		
	410		415		420
Cys Pro His Leu	Ser Ala Leu Ala Ser	Tyr Tyr Trp Ser His	Gly		
	425		430		435
Pro Ala Ala Val	Pro Glu Ala Ser Ser	Thr Val Tyr Asn Gly	Ser		
	440		445		450
Leu Leu Leu Ile	Val Gln Asp Gly Val	Gly Gly Leu Tyr Gln	Cys		
	455		460		465
Trp Ala Thr Glu	Asn Gly Phe Ser Tyr	Pro Val Ile Ser Tyr	Trp		
	470		475		480
Val Asp Ser Gln	Asp Gln Thr Leu Ala	Leu Asp Pro Glu Leu	Ala		
	485		490		495
Gly Ile Pro Arg	Glu His Val Lys Val	Pro Leu Thr Arg Val	Ser		
	500		505		510
Gly Gly Ala Ala	Leu Ala Ala Gln Gln	Ser Tyr Trp Pro His	Phe		
	515		520		525
Val Thr Val Thr	Val Leu Phe Ala Leu	Val Leu Ser Gly Ala	Leu		
	530		535		540
Ile Ile Leu Val	Ala Ser Pro Leu Arg	Ala Leu Arg Ala Arg	Gly		
	545		550		555
Lys Val Gln Gly	Cys Glu Thr Leu Arg	Pro Gly Glu Lys Ala	Pro		
	560		565		570
Leu Ser Arg Glu	Gln His Leu Gln Ser	Pro Lys Glu Cys Arg	Thr		
	575		580		585
Ser Ala Ser Asp	Val Asp Ala Asp Asn	Asn Cys Leu Gly Thr	Glu		
	590		595		600
Val Ala					

<210> 9

<211> 4744

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7513114CB1

<400> 9

```

ggagcccaag ccagaactcg agccctagcc ggagccggtc acagggaggc ggctgccggg 60
accgtcagcc ctgcatgatg catctccggc tcttctgcat cctgctcgcc gcggtctcag 120
gagccgaggg ctggggctac tacggctgcg acgaggagct ggtgggtccc ctgtatgcac 180
gctccctggg cgctcctcc tactacagtc tccttactgc gccgagattc gccaggctgc 240
acggcataag cgggtgggta ccacggattg gggatccgaa tccctggctc cagatagact 300
taatgaagaa gcaccggatc cgggcccgtg ccacacaggg ctcccttaat tcttgggact 360
gggtcacacg ttacatgcta ctctacggcg accgagtggg cagctggaca ccgttctacc 420
agcgagggca caactcgacc ttctttggta acgtgaacga gtcggcgggtg gtgcgccatg 480
acctgcactt ccacttcact gcgcgctaca tccgcatcgt gccctggcc tggaaaccac 540
gcggcaagat cggcctgagg ctccggcctc atggctgccc atacaaggcc gacatactct 600
atttcgacgg cgacgatgcc atctcctacc gcttcccggc aggggtcagc cgaagcctgt 660
gggacgtgtt cgcttccagc ttcaagaccg aggagaagga cggctctctg ctgcacgccg 720
aggcgcccca gggcgactac gtgacgctcg agctggaggg ggcacacctg ctgctgcaca 780
tgagcctggg cagcagccct atccagccaa gaccagggtc caccaccgtg agcgaggcg 840
gagtcctcaa tgaccagcac tggcactatg tgcgggtgga ccgatttggc cgcatgtaa 900
atttcaccct ggacggctat gtgcagcgct ttatttctaa tggagacttc gagaggctga 960
acctggacac tgagatgttc atcggaggtc tgggtggcgc cgcgcggaag aacctggcct 1020
atcggcataa cttccgcggc tgcatagaaa acgtaatctt caaccgcgtc aacatcgag 1080
gcctggccgt gcggcgccat tcccgatca ccttcgaggg taagggtggc tttcgttgcc 1140
tggaaccggg accgcaccct atcaacttcg gaggccctca caacttcgtt caagtgcccg 1200
gtttcccacg ccgtggccgc ctggcagctc catttcgctt ccgcacctgg gacctaccg 1260
ggcttctcct tttctcccgct ctgggggacg ggctgggcca cgtggagctg acgctcagcg 1320
aagggcaggt caacgtgtcc atcgcgcaga gcggccgaaa gaagcttcag ttcgctgctg 1380
ggtaccgact gaatgacggc ttttggcacg aggtgaattt tgtggcacag gaaaaccatg 1440
cagttatcag cattgatgat gtggaagggg cagaggtcag ggtctcatac ccgttgctga 1500
tccggacagg gacctcaat tcttttgggg gtgcagccct aacatgtgtg agcatgatgg 1560
acgctgttac cagtcttggg atgacttcat ttgctactgc gaactgacgg gctacaaggg 1620
agagacctgc cacacacctt tgtataagga atcctgtgag gcttatcggc tcagtgggaa 1680
aacttctgga aacttcacca ttgatcctga tggcagtggt cccctgaagc catttgtagt 1740
gtactgtgat atccgagaga accgagcggt gacagttgtg cggcatgaca ggctgtggac 1800
aactcgagtg acaggttcca gcatggagcg gccattcctg ggggctatcc agtactggaa 1860
tgcatcctgg gaggaagtca gtgcccttgc caatgcttcc cagcattgtg aacagtggat 1920
cgagttctcc tgctacaatt cccggctgct caacactgca ggaggctacc cctacagctt 1980
ttggattggc cgaaatgagg agcagcactt ctactgggga ggctcccagc ctgggatcca 2040
gcgctgtgcc tgtggtctgg accggagctg tgtggaccct gccttgtact gcaactgtga 2100
cgctgaccag cccagtgga gaactgacaa gggactgctg accttgttg accatctgcc 2160
tgtactcag gtagtgatag gggatacgaa ccgctccact tctgaggccc agttcttctc 2220
gaggcctctg cgctgctatg gcgatcgaaa ttctgggaa accatttctc tccacaccgg 2280
ggctgcacta cgcttcccc caatccgtgc caaccacagc ctggatgtct ccttctactt 2340
caggacctct gctccctcgg ggtcttctc agagaatatg gggggccctt actgccagtg 2400
gcgccgacct tatgtgcggg tggaaactca cacatccgg gatgtggtct tcgctttga 2460
tgtggggaat ggggatgaga acctcacagt acactcagac gactttgagt tcaatgatga 2520
cgagtggcac ctggtccggg ctgaaatcaa cgtgaagcag gcccggtcc gagtggatca 2580
ccggccctgg gttctgcggc ctatgccact gcagacctac atctggatgg agtatgacca 2640
gccctctat gtgggatctg cagagcttaa gagacgccc tttgtgggtt gcttgagggc 2700
catgctctg aacggagtga ctctgaacct ggaggggcgt gccaatgcct ctgagggtac 2760
ctcaccacac tgcacaggcc actgtgccca ccctcggtc ccctgttctc atggaggccg 2820
ctgcgtggag cgctatagct actacacgtg tgactgtgac ctacggctt ttgatgggc 2880
atactgaac cacgatatgt gtggtttctt tgagccggg acctggatgc gctataacct 2940
acagtacgg ctgcgctctg cagccaggga gttctccac atgctgagcc ggccagtgc 3000
aggctatgag cctggctaca tccgggcta tgatactccg ggctatgtgc ctggctacca 3060
tggccccggg taccgcctgc ccgactacc ccggcctggg cggcctgtgc ccggttaccg 3120
tgggcctgtc tacaacgtta cgggagagga ggtctcctc agcttcagca ccagctccgc 3180
ccctgctgtc ctgctctacg tcagttcctt tgttcgtgac tacatggctg tgctcatcaa 3240
ggatgatggg acccttcagc tgcgatatca gctgggcacc agtccctaag tgtaccagct 3300
aaccactcga ccagtgaccg atggccagcc ccatagcatc aatatcacc gtgtttaccg 3360

```

```

gaacctcttc atccaggtgg actacttccc actgacagag cagaagttct cgctgttggt 3420
ggacagccag ttggactcac ccaaggcctt gtatttaggg cgtgtgatgg agacaggagt 3480
cattgaccgg gagatccagc gctacaacac cccaggtttc tcaggctgcc tgtctgggtg 3540
tcgattcaac aacgtggctc cctcaagac ccacttccga acccctcgac ccatgactgc 3600
tgagctagct gagggccttc gagttcaggg agaactgtcc gaatctaatt gcggagctat 3660
gccacgtctt gtttcagagg tgccacctga gcttgatccc tggatatctg cccagactt 3720
cccctactac catgatgaag gatgggttgc catactttta ggcttttttg tggcctttct 3780
gctgctgggg ctggtgggaa tgttgggtgc cttctatctg caaaatcatc gctataaggg 3840
ctcctaccat accaatgagc ccaaggctgc ccacgagtac catcctggca gcaaacctcc 3900
cctaccact tcaggccctg cccaggtccc caccctaca gcagctccca accaagctcc 3960
agcctcagcc ccagcccag ccccaactcc agcccagcc cctggccccc gggatcagaa 4020
cctacccag atcctggagg agtccaggtc tgaatgagtc agaagggtt ctgggaccaa 4080
ttccagctcc tgacattccc ccagtccctg ctctccccc tcctatcagg gacatttggc 4140
tctcttagc tggctctgct catccagagg atattcccc atccccccc catcaagttt 4200
gggtgggcaga gctacagatg ggacccaagg gagtggccga gcctcactgc ctaaaccaat 4260
gccccttctca tccctgtttc cccaggtccc tggctgttta tctgccccaa aggagaagcc 4320
tcatgggggtt gacatagggtc ctttctgcca tctctgttcc agctgctgtc agggattaac 4380
aacagagtgt aggggagatt aactgcctcc cttccaatag acactatcag caggagacaga 4440
tgtgtgggag tgcagggctg cagaggggat ggggggagga ggctgctaaa ccctatcccc 4500
cagcctcccc cctgccctga agatcttcca tttgcttcca ctcagctgga ggctcaagag 4560
ggcttgatgg ctgtcccctg cccccctcct tttgttttgt acacagagac caagaggctc 4620
cagtttagca ccttagtacc tcccgttgtt caatttgttt agcccaagcc ctaaaaatac 4680
tgcaacgtag agaaaataat gcagataccc tgactaggca ggctctaatt ccttcaagct 4740
tttg

```

<210> 10

<211> 2294

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7512261CB1

<400> 10

```

gggggcccgg cggggcggga gtgcagggga cgtgagggcg caagggccgg gacatggggc 60
ccgccagccc cgctgctcgc ggtctaagtc gccgcccggg ccagccggcg ctgccgctgc 120
tgctgccact attgctgctg cttctgcgcg cgcagcccgc catcgggagc ttggccgggtg 180
ggagccccgg cgcggccgag gccccggggt cggcccaggt ggctggacta tgcgggcgcc 240
taacccttca ccggggacctg cgcaccggcc gctgggaacc agaccacag cgctctcgac 300
gctgtctccg ggacccgcag cgcgtgctgg agtactgcag acagatgtac ccggagctgc 360
agattgcacg tgtggagcag gctacgcagg ccattcccat ggagcgctgg tgcgggggtt 420
cccgagcgcg cagctgcgcc cacccccacc accaggttgt gcccttcgc tgcctgcctg 480
gtgaatttgt gagtgaggcc ctgctggtgc ctgaaggctg ccggttcttg caccaggagc 540
gcatggacca atgtgagagt tcaacccgga ggcatcagga ggcacaggag gcctgcagct 600
cccagggcct catcctgcac ggctcgggca tgcctttacc ctgtggctcg gatcggttcc 660
tggtgtgga gtatgtgtgc tgtccccctc cagggacccc cgaccatct gggacagcag 720
ttggtgaccc ctccaccggc tccctggcccc cggggagcag agtagagggg gctgaggacg 780
aggaagagga ggaatccttc ccacagccag tagatgatta cttcgtggag cctccgcagg 840
ctgaagagga agaggaaacg gtcccacccc caagctccca tacacttgca gtggtcggca 900
aagtcactcc caccocgagg ccacagacg gtgtggatat ttactttggc atgcctgggg 960
aatcagtgga gcacgagggg ttccctgaggg ccaagatgga cctggaggag cgtaggatgc 1020
gccagattaa tgaggtgatg cgtgaatggg ccatggcaga caaccagtcc aagaacctgc 1080
ctaaagccga cagacaggcc ctgaatgagc acttccagtc cattctgcag actctggagg 1140
agcaggtgtc tggtagcgga cagcgcctgg tggaaaccca cgccaccgc gtcatcgccc 1200
ttatcaacga ccagcgccgg gctgccttgg agggcttcc tggcagccctg caggcagatc 1260
cgcctcaggg ggagcgtgtc ctgttgggcc tgcggcgcta cctgcgtgcg gagcagaagg 1320
aacagaggca cacgctgcgc cactaccagc atgtggccgc cgtggatccc gagaaggcac 1380
agcagatgcg cttccaggtg catacccacc ttcaagtgat tgaggagagg gtgaatcaga 1440
gcctgggcct gcttgaccag aacccccacc tggctcagga gctgcggccc caaatccagg 1500
aactcctcca ctctgaacac ctgggtccca gtgaattgga agcccctgcc cctgggggca 1560
gcagcgagga caaggtggg ctgcagcctc cagattccaa ggatgacacc cccatgacct 1620
ttccaaaagg tgaatgcgtc tgttccaagg ggtttccctt tccactcatc ggagattcag 1680

```

```

agggatgagc tggcaccagc tgggacaggg gtgtcccgtg aggctgtgtc gggctctgctg 1740
atcatgggag cgggcgggag ctcctctcatc gtctctctcca tgctgtctcct gcgcaggaag 1800
aagccctacg gggctatcag ccatggcggtg gtggaggtgg accccatgct gaccctggag 1860
gagcagcagc tccgcgaact gcagcggcac ggctatgaga accccactta ccgcttcctg 1920
gaggaacgac cctgacccgg ccccttcac cccttcagcc gagcccagac ctccctctctt 1980
cctggagccc cagaacccca actcccagcc tagggcagca gggagtcttg aagtgatcat 2040
ttcacaccct tttgtgagac ggctggaaat tcttatttcc cttttccaat tccaaaattc 2100
catccctaag aattcccaga tagtcccagc agcctcccca cgtggcacct cctcacctta 2160
atattttttt taagtttttt tatggctctt taagggtgacc gccaccttgg tcttagtgtc 2220
tattccctgg aattcacctt ctcatgtttc cctactaaca tcccaataaa gtcctcttcc 2280
ctaccaggcc aaaa 2294

```

<210> 11

<211> 2067

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7513222CB1

<400> 11

```

caggcggagc aaaagatgcc ggcctctgct gcgcggcccc gcccggtcc cgggcagcct 60
acagcctcgc ccttcccgcct actgctgctg gcggtgctga gcggcccggg atccggccgc 120
gtcccccgct cgggtgcccag aacctcgctt ccaatctctg aggctgactc ctgtctcacc 180
cgggttcgcag tccctcacac atacaattac tctgttctcc ttgtggatcc tgcctccac 240
acactttatg ttggcgcccc ggacaccatc ttcgctttat ccctgcccct ctcaggggag 300
agaccccgca ggattgactg gatggttcct gaggtcaca gacagaactg taggaagaaa 360
ggcaagaaag agggggacct cgggggcccgg aagacctcc agcagagatg gacgacgttt 420
ttgaaagctg acctgctctg tccagggcct gagcatggcc gggcctccag tgtcctgcag 480
gatgttgctg tgcttcgacc tgagcttggg gcagggactc ccactcttta tggcatcttt 540
tcttcccagt gggagggggc tactatctct gctgtctgtg ccttccgacc acaagacatt 600
cggacagtgc tgaatgggcc cttcagagaa ctaaaacatg actgcaacag aggactgcct 660
gtcgtggaca atgatgtgcc ccagcccaga cctggagagt gcatcaccaa caacatgaag 720
ctccggcact ttggctcatc tctctccctg cctgaccgcg tactcacctt catccgggac 780
caccactca tggacaggcc agtgtttcca gctgatggcc acccctgct ggtcactaca 840
gatacagcct atctcagagt cgtggcccac aggggtgacca gcctctcagg gaaagagtat 900
gatgtgctct acctggggac agaggatgga cacctccacc gagcagtgcg gatcggagct 960
cagctcagcg ttcttgaaga tctggcctta ttcccagagc cacagccagt tgagaacatg 1020
aaattgtacc acagctggct cctgggtggc tcccgtactg aggtgacaca agtgaataca 1080
accaactgtg gccgtctcca gagctgctca gactgcatcc tggcccagga cccagtctgt 1140
gcctggagct tccggctgga tgagtgtgtg gcccatgccg gggagcaccg agggttgggtc 1200
caagacatag agtcagcaga tgtctcctct ttgtgtccta aagagcctgg agcgacggcg 1260
acagagggaa cttctggcta gagacaaggt gggcctggac ctgggggctc caccttctgg 1320
gaccacaagc tacagccaag accctccctc cccctctcct gaagatgagc ggttgccgct 1380
ggccctggcc aagaggggca gtggcttttg tggattctca ccacccttcc tgcttgatcc 1440
ttgccaagc ccagcccaca ttccggctaac tggggctcct ctagccacat gtgatgaaac 1500
atccatctag agctgggcaa atgaccacta gtgtataagt gatcactgga acggagtgac 1560
actgagatg ctgggggtca ctgggcctgg aagaccatcc cagcctctga gttctctttg 1620
agtatgagt attacttgga ttttagtctc tgttctctct gagcctggat gggcttgggg 1680
ccagaccttt gcctgattcc tgattcccat gagaaatcag aactgctttc tgcagcaaat 1740
cagggcttcc ccctaacatc tgaactcctg taaaccttca tccctggccc cctatcttgg 1800
gcccattagt tttggggatg gggcacaggg catagctatg actttgcttt ctggttggag 1860
cctggccgga aggaagagcc ctggaggtgg ttgggggcaa atgtgccctg agtccttggg 1920
gtggttctgc ttattcttca agtttatctg aatctgtggg gagtgcata tccccatgtt 1980
gcaatatgga gtctctgccc tgagatcttc cccatctcag ttttccttcc atgaaagagt 2040
acgtgtaaat acatagtgtt cataaga 2067

```

<210> 12

<211> 1209

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7522457CB1

<400> 12

```

tccgcttctt ccaaaccagc aggtagaaag atgggggcacc caccagtctc tcccagtgcc 60
ccggcccccag ctggcaccac agctatacct gggcttattc cagacctgtg cgccggggacc 120
ccctggcccc gctgggctct cattgccggc gcccttgccg cgggctcct cctcgtctcc 180
tgctctctct gtgctgctg ctgctgctgc cgccgccaca ggaagaagcc cagggataag 240
gagtcctgtg gtctgggcag tgcccgccgc accaccacca cccacctggt gcaacctgat 300
gtggatggcc tggagtccag cccgggggat gctcagcaat gggggcgcc gcagctctcc 360
ctggagttcg actttggaag ccaggagatc aggggtgggc tgaggcaggc agccgacctg 420
aggcctgggg gcaccgtgga cccctatgcc cgggtcagcg tctccacca ggccggacac 480
agacatgaga caaaagtga ccgaggcacg ctctgcccc tgtttgacga gacctgctgc 540
ttccacgcgg agctgccagg gccaccctg caggtgcagc ttttcaactt caagcgcttc 600
tcggggcatg agccctggg tgagctccgt ctgccactgg gcaccgtgga tctgcagcat 660
gttctggagc actggtacct gctgggccc cggctgcca ctgagccga gcaggctggg 720
gagctgtgct tctctctcc gtacgtgcc agctcaggcc ggctgacct ggtggtgctg 780
gaggtcggg gcctgcgtcc aggacttgca gagccctacg tgaaggcca gctcatgctg 840
aaccagagga agtgaagaa gagaaagaca gccacaaaa agggcacggc ggccccctac 900
ttcaatgagg ccttcacctt cctggtgccc ttcagccagg tccagaatgt ggacctggtg 960
ctggctgtct gggaccgcag cctgccgctc cgaactgagc ccgtaggcaa ggtgcacctg 1020
ggtgcccggg cctcggggca gccctgcag cactgggcag acatgctggc ccacgcccgg 1080
cggccattg cccagcggca cccctgcgg ccagccagg aggtggaccg catgctggcc 1140
ctgcagcccc gccttcgcct gcgcctgcc ttgccccact cctgaatgca ccacatgcct 1200
ctgtctcca                                     1209

```

<210> 13

<211> 635

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7518601CB1

<400> 13

```

agctgtaggc acctgtcgtc ctgccttcga tggctcctcc ctccgtcccc ctggtcctcc 60
tctcgtcct cttgctgagc ctggcagaga ctccagcatc cgcacctgcc caccggggac 120
gaggaggctg gacctcaat agtgctggct acctctggg tcccgctct cacctcccc 180
aaatgggtga ccaagacgga aagagggaga cagcccttga gatcctagac ctgtggaagg 240
ccatcgacgg gctcccctac tcccaccctc cacagccctc caagaggaat gtgatggaga 300
cgtttgccaa accagagatt ggaggtaaag ccaggaaca cagaagagag acaccgacag 360
gagaggggga acaaggaagt ggcaggcaga gcttagagga tctgggcatg ctgagcatga 420
aaattcccaa ggaggaagt gtcctgaagt cacagatgtc ttcaaatccc tgttcctatc 480
cttcttctcc agctcctcct gctccctctg aaaccttttc taggtaccct atgctgagac 540
taagatcctg aagttaaata cccaggtctc atgaagacat tgtaaagcat ttgaattatt 600
ctaaaaaatt tctggagttt cttaggtttt attga                                     635

```

<210> 14

<211> 4453

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7520315CB1

<400> 14

```

tatgtcaaaa tgagtctgct gatgtttaca caactactgc tctgtggatt tttatatggt 60
cgggttgatg gatcgctct tcgccaggag gactttcccc cgcggattgt ggagcatcct 120
tccgatgtca tcgtctctaa gggcgagccc acgactctga actgcaaggc ggagggccgg 180
ccaacgcccc ccattgagtg gtacaaagat ggggagcgag tggagactga caaggacgat 240

```

```

ccccgggtccc acaggatgct tctgcccagc ggatccttat tcttcttgcg catcgtgcac 300
gggcgcagga gtaaacctga tgaagggaagc tacgtttgtg ttgcgaggaa ctatcttggg 360
gaagcagtga gtcgaaatgc gtctctggaa gtggcattgt tacgagatga cttccgacaa 420
aaccacacag atgtttagt ggagctgga gagcctgcaa tcctggagtg ccagcctccc 480
cggggacacc cagaaccac catctactgg aaaaagaca aagttcgaat tgatgacaag 540
gaagaaagaa taagtatccg tggtggaata ctgatgatct ccaataccag gaaaagtgat 600
gcagggatgt atacttgtgt tggtagcaat atgggtgggag aaagggacag tgaccacgca 660
gagctgactg tctttgaacg acccacatct ctcaggaggc caattaacca ggtggtactg 720
gaggaagaag ctgtagaatt tcgttgtcaa gtccaaggag atcctcaacc aactgtgagg 780
tggaaaaagg atgatgcaga cttgccaaga ggaaggatag acatcaaaga cgattacaca 840
ctaagaatta aaaagaccat gagtacagat gaaggcacct atatgtgtat tgctgagaat 900
cgggttggaa aaatggaagc ctctgttaca ctcaccgtcc gagctcccc acagtttgtg 960
gttcggccaa gagatcagat tggtgtctaa ggtcgaacag tgacgtttcc ctgtgaaact 1020
aaaggaacac cacagccagc tgttttttgg cagaagaag gcagccagaa cctacttttc 1080
ccaaaccaac ccagcagcc caacagtaga tgctcagtgt caccaactgg agacctcaca 1140
atcaccaaca ttcaacgttc cgacgcgggt tactacatct gccaggcttt aactgtggca 1200
ggaagcattt tagcaaaagc tcaactggag gttactgatg ttttgacaga tagacctcca 1260
cctataattc tacaaggccc agccaaccaa acgctggcag tggatggtac agcgttactg 1320
aaatgtaaag ccactggtga tcctcttccc gtaattagct ggttaaagga gggatttact 1380
tttccgggta gagatccaag agcaacaatt caagagcaag gcacactgca gattaagaat 1440
ttacggattt ctgatactgg cacttatact tgtgtggcta caagttcaag tggagagact 1500
tcctggagtg cagtgtgga tgtgacagaa tctggagcaa caatcagtaa aaactatgat 1560
ttaagtgacc tgccagggcc accatccaaa ccgcaggcca ctgatgttac taagaacagt 1620
gtcaccttgt cctggcagcc aggtacccct ggaacccttc cagcaagtgc atatatcatt 1680
gaggctttca gccaatcagt gagcaacagc tggcagaccg tggcaaacca tgtaaagacc 1740
accctctata ctgtaagagg actgcggccc aatacaatct acttattcat ggtcagagcg 1800
atcaaccccc aaggctctcag tgacccaagt cccatgtcag atcctgtgcg cacacaagat 1860
atcagcccac cagcacaagg agtggaccac aggcaagtgc agaaagagct aggagatgtc 1920
cttgtccgct ttcataatcc agttgtgtgt actcccacca cggttcaggc cacatggacg 1980
gttgatcgcc aaccccagtt tatccaaggc taccagatga tgtatcgta gacttcaggt 2040
ctgcaggcga catcttcgtg gcagaattta gatgccaaag tcccgactga acgaagtgtc 2100
gtcttagtca acctgaaaaa gggggtgact tatgaaatta aagtacggcc atattttaat 2160
gagttccaag gaatggatag tgaatctaaa acggttcgta ctactgaaga agcccaagt 2220
gccccaccac agtctgtcac tgtactgaca gttggaagct acaatagcac aagtattagt 2280
gtttcctggg atcctcctcc tcagatcac cagaatggaa ttatccaaga atacaagatc 2340
tggtgtctag gaaatgaaac gcgattccat atcaacaaaa ctgtggatgc agccattcgg 2400
tccgtaataa ttggtggact attcccaggt attcaatacc gggtagaggt tgcagctagt 2460
accagtgcag ggttgagat aaagagtga ccacagccaa taataatcgg gagacgcaat 2520
gaagttgtca ttactgaaaa caataacagc ataactgagc aaatcactga tgtggtgaag 2580
caaccagcct ttatagctgg tattggtggt gcctgctggg taattctgat gggtttttagc 2640
atatggttgt attggcgaag aaagaagagg aagggactca gtaattatgc tgttacgttt 2700
caaagaggag atggaggact aatgagcaat ggaagccgtc caggtcttct caatgctggg 2760
gatcccagct atccatggct tgctgattct tggccagcca cgagcttgcc agtaataaat 2820
agcaacagtg gcccaaatga gattggaaat tttggccgtg gagatgtgct gccaccagtt 2880
ccaggccaag gggataaaac agcaacgatg ctctcagatg gagccattta tagtagcatt 2940
gacttacta ccaaaaccag ttacaacagt tcacgcaaaa taacacaggc taccctatat 3000
gccacgacac agatcttgca ttccaacagc atacatgaat tggctgtcga tctgctgat 3060
ccacaatgga aaagctcaat tcagcaaaaa acagatctga tgggatttgg ttattctcta 3120
cctgatcaga acaaaggtaa caatggtggg aaaggtggaa aaaagaagaa aaataaaaac 3180
tcttctaaac cacagaaaaa caatggatcc acttgggcca atgtccctct acctcccccc 3240
ccagtccagc cccttccctgg cacggagctg gaacactatg cagtggaca acaagaaaat 3300
ggctatgaca gtgatagctg gtgcccacca ttgccagtac aaacttactt acaccaaggt 3360
ctggaagatg aactggaaga agatgatgat aggggtccaa cacctcctgt tcgaggcgtg 3420
gcttcttctc ctgtatctc ctttgacagc cagtcactg caactcttac tccatcccca 3480
cgggaagaga tgcaacccat gctgcaggct cactggatg agttgacaag agcctatcag 3540
tttgatatag caaaacaaac atggcacatt caaagcaata atcaacctcc acagcctcca 3600
gttccaccgt taggttatgt gtctggagcc ttgatttctg attttggaaac ggatgttgca 3660
gatgatgatg ccgacgacga agaggaagct ttagaaatcc ccaggcccct gagagcactg 3720
gaccagactc ctggatccag catggacaat ctagacagct ctgtgacagg aaaagccttt 3780
acctctctc aaagacctcg acctaccagc ccatthttct ctgacagtaa caccagtgc 3840
gccctgagtc aaagtcaag gctcggccc actaaaaaac acaagggagg gcggatggac 3900
caacaaccag cattgcctca tcgaaggga ggaatgacag atgatcttcc accaccacca 3960
gatccccgcg caggtcaggg ttttaaggcag caaataggcc cgagccagca ggctggtaac 4020

```

```

gtggaaaact cagcagagag aaaaggaagc tctctagaga gacaacatgc atccagctta 4080
gaagacacaa agagctcatt ggattgtcca gctagaacct ccctagagtgc gcagcgacaa 4140
accaggaat ggataagctc cacagaacga caagaagata tacggaaaagc cccacacaaa 4200
caaggtgtcg gatcagagga ggccttgggtg ccctatagca agcccagttt cccatctcca 4260
ggtggccaca gctcatcagg aacagcttct tctaagggat cctactggacc taggaaaacc 4320
gaggtgttga gagcaggcca ccagcgcaat gccagcgacc ttcttgacat aggatatatg 4380
ggctccaaca gtcaaggaca gtttacaggt gaattatagt aaatgagagg agacatacaa 4440
gctgctctga aga

```

<210> 15

<211> 1937

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7525536CB1

<400> 15

```

caaagttcca gtcaggaaga gtaagctgaa ggccctgttg cacagcatga caaacatccc 60
taacaatagc gccttgtgta ttccacagtt gcctgaagag cacattttta gtgttctttt 120
tttcattagc ctgactgact ctttctacaa tgcataattgt agaaattgca ttcttacaca 180
tatcacatca gaacaccata ttgtacccca tacacaccta cggttatggt gattaaaaat 240
taaaagaata actggggaga agagaggaga gggaatgctg gatgggttaa accagccaag 300
atcccctgca gggggtgacc cagaaaccaa aggtgagtag gagtggtcag gtcaaagggc 360
agggaggggtg gccagtttg ggaggggagc aggccgttac tgtgagatgc caaggggaca 420
ctccttgggtg aggtgaaaag gggccgtgcc tcaccaaagg attcagattt tattccaagc 480
aggggtggag gcgggatgat ctggttcacg ttccagagtg taaactggcc acccccacag 540
ggctccatca caggagcccc ctctgtggc tcccctcctc tccctcccag acctccctcc 600
cctcccacct tcccatgcac cccctccctt cctcctatgc cgtccctcct tcccacaaac 660
cacttatcac ctccacagcc ctctgtcacc tccggccac cagcctcctc acctgggagc 720
agggccacat gatggtggca aaaataactc cctttgacaa agtgtgaagg gggcagaggg 780
aggagggag ctgagcccca gcgctaggaa ggagctctga gagggttctg agctctgggt 840
caccctcact cactggggac acagcagtca cgggctctgc cctcatgagt gcgtaccac 900
tgccagcagc agcaactcac actgggtgaa gtagctocca gagcggtaag ggccgggagc 960
tggaactgact cccaacacac ccagtcccct ccaaagcctg ccacaaccct ccaggctgag 1020
gactccccc gccctccct caaccctgcg ctctgtcctc aggtccctgc atggtatcag 1080
cgatgcttct tcaccccatt tccgaggccg ggtgcgttcc cgccagcccc agtcctcagg 1140
catttctctg agtctccgcc caccgcccc cgtctgggtc cccatggcgg gcaactgcgc 1200
agcaggtggg cagcctcccc gggttagcat gcaggagcac atggccatcg atgtgagccc 1260
gggccccatc cggcccatcc gcctcatttc ccaactactc ccgcactttt acccttttgc 1320
ggagcctgcc ctgcaccctc cgaacctgcg ccccgagcgc gcgtccgcgc tccgctctgc 1380
accccagctg cagcccagac cagagccaga aggagactca gacgacagca ctgccctggg 1440
caccctggag ttcacacttc tttttgaagc ggacaacagt gccctgcatt gcacggctca 1500
tcgtgccaag ggctcaagc cattggcctc aggtccgcgc gatgcctatg tcaaagccaa 1560
tctgctgcca ggggccagca aggccagcca gcttcggaca cacaccgttc ggggcacgag 1620
ggtacctgtc tgggaggaga cactcaccta tcacggcttc acccgccagg atgctgagt 1680
caagaccctt aggtgaggat ctggcctggg ccacctgggg gtggagcaga gggatggggc 1740
tggcctggcc aggtctgacc tgggcggcca ccaggctgtg tgtgtgcgag gaccatggt 1800
acagcgacag tggcaggcac ctccctggg ggagctgcgc gtgccctga ggaagctggt 1860
gccaaaccga gccaggagct ttgacatctg tctggagaag cggaggctgg tgagtggggc 1920
tggagcacag gtgggac

```

<210> 16

<211> 2608

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7520619CB1

<400> 16

cctagaatcc	agacaccgag	tttcccaactg	tggctgggttc	aaggagctcc	ctggtgacag	60
tctgtggctg	agcatggccc	tcccagccct	gggcctggac	ccctggagcc	tcctgggcct	120
tttcctcttc	caactgcttc	agctgctgct	gccgacgacg	accgcggggg	gaggcgggca	180
ggggcccatg	cccaggggtca	gatactatgc	aggggatgaa	cgtagggcac	ttagcttctt	240
ccaccagaag	ggcctccagg	atlttgacac	tctgctcctg	agtggatgat	gaaatactct	300
ctacgtgggg	gctcgagaag	ccattctggc	cttgatatc	caggatccag	gggtccccag	360
gctaaagaac	atgataccgt	ggccagccag	tgacagaaaa	aagagtgaat	gtgcctttaa	420
gaagaagagc	aatgaggaac	ttcaagattc	ctacctgttg	cccatctcgg	aggacaaggt	480
catggaggga	aaaggccaaa	gcccccttga	ccccgctcac	aagcatacgg	ctgtcttggg	540
ggatgggatg	ctctattctg	gtactatgaa	caacttcctg	ggcagtgagc	ccatcctgat	600
gcgcacactg	ggatcccagc	ctgtcctcaa	gaccgacaac	ttcctccgct	ggctgcatca	660
tgacgcccc	tttgtggcag	ccatcccttc	gacccaggct	gtctacttct	tcttcgagga	720
gacagccagc	gagtttgact	tctttgagag	gctccacaca	tcgcgggtgg	ctagagtctg	780
caagaatgac	gtggggcggc	aaaagctgct	gcagaagaag	tggaccacct	tcctgaaggc	840
ccagctgctc	tgaccccagc	cggggcagct	gcccttcaac	gtcatccgcc	acgcggtcct	900
gctccccgcc	gattctccca	cagctcccca	catctacgca	gtcttcacct	ccagtgga	960
gggtggcggg	accaggagct	ctgcggtttg	tgcttctctt	ctcttggaca	ttgaactgtg	1020
ctttaagggg	aaatacaaag	agttgaacaa	agaaacttca	cgctggacta	cttatagggg	1080
ccctgagacc	aacccccggc	caggcagttg	ctcagtgggc	ccctcctctg	ataaggctct	1140
gaccttcatg	aaggaccatt	tcctgatgga	tgagcaagtg	gtggggacgc	ccctgctggg	1200
gaaatctggc	gtggagtata	cacggcttgc	agtggagaca	gcccagggcc	ttgatgggca	1260
cagccatctt	gtcatgtacc	tgggaaccat	taaagaagtc	ctggctgtcc	ccaactccat	1320
cctggagctc	ccctgcccc	acctgtcagc	cttggcctct	tattattgga	gtcatggccc	1380
agcagcagtc	ccagaagcct	cttccactgt	ctacaatggc	tccctcttgc	tgatagtgca	1440
ggatggagtt	gggggtctct	accagtgtcg	ggcaactgag	aatggctttt	cataccctgt	1500
gatctcctac	tgggtggaca	gccaggacca	gaccttgccc	ctggatcctg	aactggcagg	1560
catcccccg	gagcatgtga	aggtcccgtt	gaccagggct	agtgggtggg	ccgccctggc	1620
tgcccagcag	tcctactggc	cccactttgt	cactgtcact	gtcctctttg	ccttagtgct	1680
ttcaggagcc	ctcatcatcc	tcgtggcctc	cccattgaga	gcactccggg	ctcggggcaa	1740
ggttcagggc	tgtgagacc	tgccgctgg	ggagaaggcc	ccgttaagca	gagagcaaca	1800
cctccagtct	ccaaggaat	gcaggacctc	tgccagtgat	gtggacgctg	acaacaactg	1860
cctaggcact	gaggtagctt	aaactctagg	cacaggccgg	ggctgcggtg	caggcacctg	1920
gccatgctgg	ctgggcggcc	caagcacagc	cctgactagg	atgacagcag	cacaaaagac	1980
cacctttctc	ccctgagagg	agcttctgct	actctgcac	actgatgaca	ctcagcaggg	2040
tgatgcacag	cagtctgcct	cccctatggg	actcccttct	accaagcaca	tgagctctct	2100
aacagggtgg	gggctacccc	cagacctgct	cctacactga	tattgaagaa	cctggagagg	2160
atccttcagt	tctgccattc	cagggacctc	ccagaaacac	agtgtttcaa	gagaccctaa	2220
aaaacctgcc	tgtcccagga	ccctatggta	atgaacacca	aacatctaaa	caatcatatg	2280
ctaacatgcc	actcctggaa	actccactct	gaagctgccg	ctttggacac	caacactccc	2340
ttctcccagg	gtcatgcagg	gatctgctcc	ctcctgcttc	ccttaccagt	cgtgcacgcg	2400
tgactcccag	gaagtctttc	ctgaagtctg	accacctttc	ttcttgcttc	agttggggca	2460
gactctgatc	ccttctgccc	tggcagaatg	gcaggggtaa	tctgagcctt	cttcaactct	2520
ttaccctagc	tgacctcttc	acctctcccc	ctcccttttc	ctttgttttg	ggattcagaa	2580
aactgcttgt	cagagactgt	ttatttttt				2608