

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2011212165 B2**

(54) Title
Identification, assessment, and therapy of cancers with innate or acquired resistance to ALK inhibitors

(51) International Patent Classification(s)
G01N 33/68 (2006.01) **G01N 33/15** (2006.01)
C12Q 1/68 (2006.01) **G01N 33/574** (2006.01)

(21) Application No: **2011212165** (22) Date of Filing: **2011.02.04**

(87) WIPO No: **WO11/095894**

(30) Priority Data

(31)	Number	(32)	Date	(33)	Country
	61/337,465		2010.02.04		US

(43) Publication Date: **2011.08.11**

(44) Accepted Journal Date: **2016.03.03**

(71) Applicant(s)
Jichi Medical University

(72) Inventor(s)
Mano, Hiroyuki;Choi, Young L.;Soda, Manabu

(74) Agent / Attorney
Spruson & Ferguson, L 35 St Martins Tower 31 Market St, Sydney, NSW, 2000

(56) Related Art
UniProtKB/TrEMBL Accession No. D1MAM2 19 January 2010
WOOD, A.C. et al., Journal of Clinical Oncology, 2009, Vol. 27, No. 15S (May 20 Supplement), Abstract 10008b
US 2009/0156475
WO 2009/103061
McDermott, U. et al., Cancer Research, 2008, vol 68, p.3389-3395
WO 2010/132888
Lu, L. et al., Biochemistry, 2009, vol 48, p.3600-3609

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 August 2011 (11.08.2011)

PCT

(10) International Publication Number
WO 2011/095894 A3

(51) International Patent Classification:

G01N 33/68 (2006.01) *G01N 33/15* (2006.01)
C12Q 1/68 (2006.01) *G01N 33/574* (2006.01)

(21) International Application Number:

PCT/IB2011/000382

(22) International Filing Date:

4 February 2011 (04.02.2011)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/337,465 4 February 2010 (04.02.2010) US

(71) Applicant (for all designated States except US): **JICHI MEDICAL UNIVERSITY** [JP/JP]; 2-6-3 Hirakawa-cho, Chiyoda-ku, Tokyo 102-0093 (JP).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **MANO, Hiroyuki** [JP/JP]; 4-13-3 Hongo, Bunkyo-ku, Tokyo 113-0033 (JP). **CHOI, Young, I.** [KR/JP]; 2-3-4-204 Kojima, Taito-ku, Tokyo 111-0056 (JP). **SODA, Manabu** [JP/JP]; 5-5-4 East Gion, Shimotsuke-shi, Tochigi 329-0434 (JP).

(74) Agents: **YANAGIDA, Masashi** et al.; Yanagida & Associates, 7F, Shin-Yokohama KS Bldg., 3-18-3, Shin-Yokohama, Kohoku-ku, Yokohama-shi Kanagawa 222-0033 (JP).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(88) Date of publication of the international search report:

1 March 2012

(54) Title: IDENTIFICATION, ASSESSMENT, AND THERAPY OF CANCERS WITH INNATE OR ACQUIRED RESISTANCE TO ALK INHIBITORS

(57) Abstract: Described herein are compositions, kits, and methods for determining whether subjects having cancer(s) positive for ALK mutations are likely to respond to treatment with an ALK inhibitor and/or whether a patient having such cancer(s) is likely to have a relatively slower disease progression. Further described are methods for prognosing a time course of disease in a subject having such cancer.



WO 2011/095894 A3

***Identification, Assessment, and Therapy of Cancers with
Innate or Acquired Resistance to ALK Inhibitors***

RELATED APPLICATIONS

5 This application claims the benefit of priority to United States Provisional Patent Application serial number 61/337,465, filed February 4, 2010.

BACKGROUND OF THE INVENTION

 Tyrosine kinases are a class of enzymes that catalyze phosphorylation of tyrosine residues of protein substrates via a transfer of the terminal phosphate of adenosine
10 triphosphate. In many contexts, tyrosine kinases play critical roles in signal transduction for a number of cell functions, including cell proliferation, carcinogenesis, and cell differentiation.

 EML4-ALK is a fusion-type protein tyrosine kinase that is present in ~5% of non-small cell lung cancer (NSCLC) cases and which is generated as a result of a small
15 inversion within the short arm of human chromosome 2 (Soda, M. *et al.* (2007) *Nature* 448:561-566; Mano, H. (2008) *Cancer Sci.* 99:2349-2355). EML4-ALK undergoes constitutive dimerization as a result of interaction between the coiled-coil domain within the EML4 region of each monomer and thereby acquires pronounced oncogenic activity. Transgenic mice that express EML4-ALK, specifically in lung epithelial cells, develop
20 hundreds of adenocarcinoma nodules in both lungs soon after birth, and oral administration of a specific inhibitor of ALK tyrosine kinase activity rapidly eradicates such nodules from the lungs (Soda, M. *et al.* (2008) *Proc. Natl. Acad. Sci. USA* 105:19893-19897). These observations reveal the essential role of EML4-ALK in the carcinogenesis of NSCLC harboring this fusion kinase, and they further support the feasibility of molecularly targeted
25 therapy with ALK inhibitors for this cancer. For example, clinical trials of an inhibitor, PF-02341066, of the tyrosine kinase activity of both ALK and MET are under way for the treatment of EML4-ALK-positive NSCLC, and their interim results are promising (Kwak, E.L. *et al.* (2009) *J. Clin. Oncol.* 27(suppl):15s (abstract 3509)). A subset of EML4-ALK-positive tumors, however, do not respond to the inhibitor, with unknown molecular basis of
30 treatment failure.

In addition to PF-02341066, other tyrosine kinase inhibitors (TKIs) have been shown to possess pronounced therapeutic activity in cancer patients. Imatinib mesylate, a TKI for ABL1 and KIT, for instance, markedly improves the outcome of individuals with chronic myeloid leukemia positive for the BCR-ABL1 fusion kinase or with a gastrointestinal stromal tumor positive for activated KIT (Druker, B.J. et al. (2001) N. Engl. J. Med. 344: 1031-1037; Heinrich, M.C. et al. (2008) J. Clin. Oncol. 26:5360-5367). Furthermore, gefitinib and erlotinib, both of which are TKIs for the epidermal growth factor receptor (EGFR), are effective in the treatment of NSCLC associated with EGFR activation (Mok, T.S. et al. (2009) J. Clin. Oncol. 27:5080-2087; Mok, T.S. et al. (2009) N. Engl. J. Med. 361 :947-957). Unfortunately, a subset of target tumors are either refractory to corresponding TKIs from the start of treatment or become resistant after an initial response. Secondary mutations in the target kinases that directly or allosterically affect the shape of the ATP -binding pocket, resulting in hindrance of TKI binding, have been detected in some cases of treatment failure (Deininger, M. et al. (2005) Blood 105:2640-2653; Kobayashi, S. et al. (2005) N. Engl. J. Med. 352:786-792; Pao, W. et al. (2005) PLoS Med. 2:e73; Shah, N.P. et al. (2002) Cancer Cell 2: 117-125). Accordingly, there is an immediate need to identify mutations conferring resistance upon tyrosine kinases, such as EML4-ALK, in order to better develop compositions, kits, and methods for identifying, assessing, preventing, and treating disorders related to their aberrant expression and/or activity.

SUMMARY OF THE INVENTION

The present invention provides, at least, composition, methods, and kits for the identification, assessment and treatment of cancer based upon the identification of novel anaplastic lymphoma kinase (ALK) mutation(s) conferring resistance to known ALK inhibitors. Such ALK mutations are also clinically relevant for the identification of pharmaceutical compositions that are able to fit into the abnormal ATP -binding pocket of generated by the novel ALK mutation(s) and inhibit ALK activity.

In one aspect, the present invention provides a method for identifying a subject having cancer or at risk for developing cancer as having an increased risk of unresponsiveness to treatment with an ALK inhibitor, comprising:

analyzing a sample obtained from the subject to detect the presence of one or more mutant ALK polynucleotide molecules encoding a mutant ALK polypeptide comprising a mutation position 1156 or to detect the presence of one or more mutant ALK polypeptides comprising a mutation at position 1156,

wherein the presence of the one or more mutant ALK polynucleotide molecules or one or more mutant ALK polypeptides indicates that the subject has an increased risk of unresponsiveness to treatment with the ALK inhibitor.

In one aspect, the present invention provides a method of treating a patient having cancer, or at risk for developing cancer, comprising:

analyzing a sample obtained from the patient to detect the presence of one or more mutant ALK polynucleotide molecules encoding a mutant ALK polypeptide comprising a mutation at position 1156 or to detect the presence of one or more mutant ALK polypeptides comprising a mutation at position 1156; and

administering to said patient a therapeutically effective amount of an ALK inhibitor.

In one aspect, the present invention provides a kit for determining the chemosensitivity of a cancer patient to treatment with an ALK inhibitor, the kit comprising:

a reagent that specifically binds to one or more mutant ALK polypeptides comprising a mutation at position 1156, wherein the reagent comprises an antibody, an antibody derivative, or an antibody fragment that specifically binds to a mutant ALK polypeptide comprising a mutation at position 1156; and

instructions for use.

In one aspect, the present invention provides a method of determining whether a test compound modulates activity of one or more mutant ALK polypeptides comprising:

(a) contacting mammalian cells transfected with a construct encoding the one or more mutant ALK polypeptides comprising a mutation at position 1156 with the test compound; and

(b) assessing the mammalian cells for activity of the one or more mutant ALK polypeptides, wherein significantly modulated activity in the presence of the test compound relative to a control experiment identifies the test compound as a modulator of the one or more mutant ALK polypeptides.

In one aspect, the present invention provides a method for identifying a subject having cancer or at risk for developing cancer as having an increased risk of unresponsiveness to treatment with an ALK inhibitor, comprising collecting a sample from the patient and analyzing the sample to detect the presence of one or more mutant ALK polynucleotide molecules, wherein the presence of the one or more mutant ALK

2011212165 06 Jan 2016

polynucleotide molecules indicates that the subject has an increased risk of unresponsiveness to treatment with the ALK inhibitor.

In another aspect, the present invention provides a method for identifying a subject having cancer or at risk for developing cancer as having an increased risk of unresponsiveness to treatment with an ALK inhibitor, comprising collecting a sample from the patient and analyzing the sample to detect the expression level, structure, and/or activity of one or more mutant ALK polypeptides, wherein the presence of the one or more mutant ALK polypeptides indicates that the subject has an increased risk of unresponsiveness to treatment with the ALK inhibitor.

In some embodiments of any aspect of the present invention, the subject has not previously been treated with an ALK inhibitor, or has been previously treated with an ALK inhibitor and has developed at least partial resistance to the ALK inhibitor (*e.g.*, PF-02341066, PDD, 2-methyl-11-(2-methylpropyl)-4-oxo-4,5,6,11,12,13-hexahydro-2*H*-indazolo[5,4-*a*]pyrrolo[3,4-*c*]carbazol-8-yl [4-(dimethylamino) benzyl]carbamate, (1*S*,2*S*,3*R*,4*R*)-3-({5-chloro-2-[(1-ethyl-2,3,4,5-tetrahydro-6-methoxy-2-oxo-1*H*-1-benzazepin-7-yl)amino]-4-pyrimidinyl}amino)bicyclo[2.2.1]hept-5-ene-2-carboxamide, and NVP-TAE684). In other embodiments, the cancer is selected from the group consisting of anaplastic large cell lymphoma, neuroblastoma, breast cancer, colorectal cancer, inflammatory myofibroblastic tumors, and non-small cell lung cancers. In still other embodiments, the sample is selected from the group consisting of sputum, bronchoalveolar lavage, pleural effusion, tissue, whole blood, serum, plasma, buccal scrape, saliva, cerebrospinal fluid, urine, stool, circulating tumor cells, circulating nucleic acids, and bone marrow. In yet other embodiments, the sample comprises cells or tissue. In some embodiments the tissue is a tumor or cancer tissue. In other embodiments, the one or more mutant ALK polynucleotide molecules or polypeptides are selected from the group consisting of the mutant ALK polynucleotide molecules or polypeptides listed in Table 1. In still other embodiments, the one or more ALK mutations are assessed by a nucleic acid hybridization assay. In yet other embodiments, the one or more ALK mutations are assessed by polymerase chain reaction. In other embodiments, the expression level of the one or more ALK polypeptides is detected using a reagent which specifically binds to one or more ALK polypeptides (*e.g.*, antibody, an antibody derivative, and an antibody fragment). In still other embodiments, the amount, structure and/or activity of the one or more mutant ALK polypeptides is compared to a control sample. In yet other

embodiments, the one or more ALK mutations are assessed at a first point in time and at least one subsequent point in time. In other embodiments, the sample comprises germline or somatic genomic DNA.

In still another aspect, the present invention provides a method of treating a patient
5 having cancer, or at risk for developing cancer, comprising collecting a sample from the patient, analyzing the sample to detect the presence of one or more mutant ALK polynucleotide molecules set forth in Table 1, and administering to said patient a therapeutically effective amount of an ALK inhibitor. In some embodiments, the ALK inhibitor is selected from the group consisting of PF-02341066, PDD, 2-methyl-11-(2-
10 methylpropyl)-4-oxo-4,5,6,11,12,13-hexahydro-2*H*-indazolo[5,4-*a*]pyrrolo[3,4-*c*]carbazol-8-yl [4-(dimethylamino) benzyl]carbamate, (1*S*,2*S*,3*R*,4*R*)-3-({5-chloro-2-[(1-ethyl-2,3,4,5-tetrahydro-6-methoxy-2-oxo-1*H*-1-benzazepin-7-yl)amino]-4-
pyrimidinyl}amino)bicyclo[2.2.1]hept-5-ene-2-carboxamide, and NVP-TAE684. In other
embodiments, the subject has not previously been treated with an ALK inhibitor, or has
15 been previously treated with an ALK inhibitor and has developed at least partial resistance to the ALK inhibitor.

In yet another aspect, the present invention provides a kit for determining the chemosensitivity of a cancer patient to treatment with an ALK inhibitor, comprising: a reagent that specifically binds to one or more mutant ALK polynucleotide molecules or
20 polypeptides; and instructions for use. In some embodiments, the kit further comprises an ALK inhibitor. In other embodiments, the reagent comprises one or more polynucleotide probes, each of which comprises a polynucleotide sequence which is complementary to a nucleotide sequence listed in Table 1 or complementary to a nucleotide sequence encoding a polypeptide listed in Table 1 (*e.g.*, oligonucleotides, cDNA molecules, RNA molecules,
25 and synthetic gene probes comprising nucleobases). In still other embodiments, the probes comprise polynucleotides from about 50 to 10⁷ nucleotides in length. In yet other embodiments, the reagent comprises an antibody, and antibody derivative, and an antibody fragment to a polypeptide encoded by one or more polynucleotide sequences listed in Table 1.

30 In another aspect, the present invention provides a method of determining whether a test compound modulates activity of one or more mutant ALK polypeptides comprising contacting mammalian cells transfected with a construct encoding the one or more mutant ALK polypeptides with the test compound and assessing the mammalian cells for activity

of the one or more mutant ALK polypeptides, wherein significantly modulated activity in the presence of the test compound relative to a control experiment identifies the test compound as a modulator of the one or more mutant ALK polypeptides. In some embodiments, the

- 5 one or more mutant ALK polynucleotide molecules or polypeptides are selected from the group consisting of the mutant ALK polynucleotide molecules or polypeptides listed in Table 1. In other embodiments, the control comprises mammalian cells expressing a wild type ALK polypeptide selected from the group consisting of polypeptides listed in Table 1. In still other embodiments, activity of the one or more mutant ALK polypeptides is selected
- 10 from the group consisting of ATP binding, tyrosine kinase activity, cancer cell proliferation, tumor growth, tumor number, apoptosis, and tumor metastasis. In yet other embodiments, the control experiment comprises mammalian cells expressing the one or more mutant ALK polypeptides in the absence of the test compound as determined by, for example, activity of the one or more mutant ALK polypeptides (*e.g.*, ATP binding, tyrosine
- 15 kinase activity, cancer cell proliferation, tumor growth, tumor number, apoptosis, and tumor metastasis).

BRIEF DESCRIPTION OF THE FIGURES

- Figure 1** depicts novel ALK mutations of the present invention associated with
- 20 resistance to ALK tyrosine kinase inhibitors. Figure 1A shows a schematic representation of the EML4-ALK protein. Positions of two *de novo* mutations in the kinase domain are shown, and those of PCR primers for amplification of kinase-domain or fusion cDNAs are indicated by the closed and open arrows above, respectively. Figure 1B shows the results of deep sequencing of ALK kinase-domain cDNAs. PCR products of ~1000 bp from the
- 25 NSCLC cell line H2228 or from specimen IDs J-#1, J-#12, J-#113, J-#127 or LK-#33 were sequenced with the GAII system. The numbers for total read coverage (Total) and mismatched reads (Mismatch) are shown at each position of the kinase-domain cDNAs with blue and red diamonds, respectively. Insets show magnified views for the 5' region of the cDNAs for J-#1 and J-#113 (depicted by green rectangles). Figure 1C shows
- 30 electrophoretograms for the ALK cDNA clones surrounding G4374 and C4493 positions. PCR was performed with cDNAs prepared from sputum obtained before treatment (Initial) and from the cells in pleural effusion obtained after relapse (Relapse). Substituted A nucleotides are shown in red.

Figure 2 depicts genomic sequences surrounding the positions corresponding to G4374 and C4493 of ALK cDNA. Genomic DNA isolated from cells in the pleural effusion of the patient was subjected to PCR for 35 cycles of 94°C for 15 s, 60°C for 30 s and 68°C for 2 min, with Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA) and the following primers (5'-GGTAAGAAGTGGCTCACTCTTGAG-3' and 5'-CACAACAAGTGCAGCAAAGACTGG-3'), and the products were ligated into the pT7Blue-2 plasmid (Takara Bio). Inserts of the plasmids were then sequenced with the 3130xl Genetic Analyzer, resulting in the identification of PCR clones containing the G4374A (left panel) or C4493A (right panel) changes. Substituted A nucleotides are shown in red.

Figure 3 depicts the results of BA/F3 cells treated with PF-02341066. BA/F3 cells expressing EML4-ALK (wild type), EML4-ALK(C1156Y), EML4-ALK(L1196M), or the double mutant EML4-ALK(C1156Y/L1196M) were incubated in the presence of the indicated concentrations of PF-02341066 for 48 h, after which cell morphology was examined by phase-contrast microscopy. Scale bar, 20 μ m.

Figure 4 depicts properties of novel ALK mutations of the present invention associated with resistance to ALK tyrosine kinase inhibitors. Figure 4A shows the number of BA/F3 cells expressing EML4-ALK (wild type), EML4-ALK(C1156Y), EML4-ALK(L1196M), or the double mutant EML4-ALK(C1156Y/L1196M) counted after incubation of 5×10^5 cells for 48 h with the indicated concentrations of PF-02341066. The percentage of viable cells is shown relative to BA/F3 cells expressing the wild-type EML4-ALK. Data are means $\pm$ s.d. from three separate experiments. Figure 4B shows the effect of PF-02341066 on tyrosine phosphorylation of wild type or mutant forms of EML4-ALK. BA/F3 cells expressing FLAG-tagged wild-type EML4-ALK or its mutants were exposed to the indicated concentrations of PF-02341066 for 15 h, after which EML4-ALK was immunoprecipitated from cell lysates and subjected to immunoblot analysis with antibodies specific for Tyr¹⁶⁰⁴-phosphorylated ALK or for the FLAG epitope (ALK). Cells expressing an inactive mutant of EML4-ALK (KM) were examined as a negative control. Figure 4C shows an *in vitro* kinase assay for FLAG-tagged wild-type EML4-ALK or its mutants immunoprecipitated from BA/F3 cells (not exposed to an ALK inhibitor). The immunoprecipitates were incubated with [γ -³²P]ATP, a synthetic peptide, and the indicated concentrations of PF-02341066. Phosphorylation of the peptide substrate immunoprecipitates were separately subjected to immunoblot analysis with antibodies to

FLAG (lower panel).

Figure 5 depicts a three-dimensional structure model for the kinase domain of ALK. Amino acid positions of ALK were superimposed on the crystal structure of the insulin receptor with a bound ATP analog (ID "1ir3" in the Protein Data Bank of Japan, available
5 on the world wide web at pd bj.org/index.html). The right panel shows the protein structure observed from the left side of the model in the left panel. The α helices and β sheets are shown in magenta and orange, respectively. The positions of helix α C, Cys¹¹⁵⁶, and Leu¹¹⁹⁶ are also indicated.

DETAILED DESCRIPTION OF THE INVENTION

10 The present invention is based, at least in part, on the identification of specific regions of the genome, including for example, Anaplastic Lymphoma Kinase (ALK) mutations, associated with predicting efficacy of ALK inhibitors in treating cancer. In particular, novel ALK gene mutations (*e.g.*, EML4-ALK polypeptide encoding mutations) have been identified herein that can lead to polypeptides at least partially resistant to
15 therapy with ALK inhibitors. The present invention further provides methods for identifying such specific genomic regions using techniques known in the art, including, but not limited to, oligonucleotide-based microarrays (Brennan, *et al.* (2004) *Cancer Res.* 64(14):4744-8; Lucito, *et al.* (2003) *Genome Res.* 13:2291-2305; Bignell *et al.* (2004) *Genome Res.* 14:287-295; Zhao, *et al.* (2004) *Cancer Research*, 64(9):3060-71), and other
20 methods as described herein including, for example, polymerase chain reaction (PCR)- and direct sequencing-based methods. The present invention further provides diagnostic kits for use in the methods.

Various aspects of the present invention are described in further detail in the following subsections.

25 I. Definitions

The articles "a" and "an" are used herein to refer to one or to more than one (*i.e.*, to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

The term "altered amount" of a marker or "altered level" of a marker refers to
30 increased or decreased copy number of a marker or chromosomal region, such as ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1), and/or increased or

decreased expression level of a particular marker gene or genes in a cancer sample, as compared to the expression level or copy number of the marker in a control sample. The term "altered amount" of a marker also includes an increased or decreased protein level of a marker in a sample, *e.g.*, a cancer sample, as compared to the protein level of the marker in a normal, control sample.

The term "altered level of expression" of ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) refers to an expression level or copy number of a marker in a test sample such as a sample derived from a patient suffering from cancer, that is greater or less than the standard error of the assay employed to assess expression or copy number, and may be at least twice, at least twice three, at least twice four, at least twice five, or at least twice ten or more times the expression level or copy number of the ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) in a control sample (*e.g.*, a sample from a healthy subject not having the associated disease), or the average expression level or copy number of the ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) in several control samples. The altered level of expression is greater or less than the standard error of the assay employed to assess expression or copy number, and is at least twice, at least three, at least four, at least five, at least ten or more times the expression level or copy number of the ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) in a control sample (*e.g.*, a sample from a healthy subject not having the associated disease), or the average expression level or copy number of the ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) in several control samples.

The term "altered activity" of a marker refers to an activity of a marker which is increased or decreased in a disease state, *e.g.*, in a cancer sample, as compared to the activity of the marker in a normal, control sample. Altered activity of a marker may be the result of, for example, altered expression of the marker, altered protein level of the marker, altered structure of the marker, or, *e.g.*, an altered interaction with other proteins involved in the same or different pathway as the marker.

The term "altered structure" of a marker refers to the presence of mutations or mutations within the marker gene or marker protein, *e.g.*, mutations which affect expression or activity of the marker, as compared to the normal or wild-type gene or protein. For example, mutations include, but are not limited to inter- and intra-chromosomal

rearrangement, substitutions, deletions, and insertion mutations. Mutations may be present in the coding or non-coding region of the marker.

“Anaplastic lymphoma kinase” and “ALK” are used interchangeably herein and refer to native anaplastic lymphoma kinase, and certain variants and mutations thereof, derived from any source (*e.g.*, rodents, humans, and other mammals). In some embodiments, ALK protein is represented by NCBI Ref Seq identification number NP_004295. Unless indicated otherwise, the terms refer to the human protein. The gene encoding ALK may also be referred to herein as “ALK”. In some embodiments, ALK nucleotide sequences are represented by NCBI Ref Seq identification number NM_004304.3 and GenBank accession number 29029631, relevant sequences therein (*e.g.*, the coding, 5' UTR, 3'UTR, transcription start, translation start, transcription stop, translation stop, etc. sequences) of which can readily be identified by a skilled artisan.

In addition, “Anaplastic lymphoma kinase” and “ALK” are also used herein to include ALK fusion kinases and variants thereof, which are well known to a skilled artisan. Such ALK fusion kinases and variants thereof comprise ALK kinase activity and can harbor mutations as described herein rendering the ALK kinase activity resistant to ALK inhibitors. Representative examples include EML4-ALK Variant 1 (AB274722.1; BAF73611.1), EML4-ALK Variant 2 (AB275889.1; BAF73612.1), EML4-ALK Variant 3a (AB374361.1; BAG55003.1), EML4-ALK Variant 3b (AB374362.1; BAG55004.1), EML4-ALK Variant 4 (AB374363.1; BAG75147.1), EML4-ALK Variant 5a (AB374364.1; BAG75148.1), EML4-ALK Variant 5b (AB374365.1; BAG75149.1), EML4-ALK Variant 6 (AB462411.1; BAH57335.1), EML4-ALK Variant 7 (AB462412.1; BAH57336.1), KIF5B-ALK (AB462413.1; BAH57337.1), NPM-ALK, TPM3-ALK, TFGXL-ALK, TFGL-ALK, TFGS-ALK, ATIC-ALK, CLTC-ALK, MSN-ALK, TPM4-ALK, MYH9-ALK, RANBP2-ALK, ALO17-ALK, and CARS-ALK (see, for example, Pulford *et al.*, (2004) *J. Cell. Physiol.* 199:330-358, which is herein incorporated by reference in its entirety). In addition, a skilled artisan will understand that ALK kinase variants can arise depending upon the particular fusion event between an ALK kinase and its fusion partner (*e.g.*, EML4 can fuse at least exon 2, 6a, 6b, 13, 14, and/or 15, as described, for example, in Horn and Pao, (2009) *J. Clin. Oncol.* 27:4247-4253, which is herein incorporated by reference in its entirety). For example, representative ALK sequences are provided herein as follows:

Table 1**Wild Type ALK cDNA Sequence (NM_004304.3; GI:29029631):**

	1	gggggcgcca	gcggtggtag	cagctggtac	ctcccgccgc	ctctgttcgg	agggtcgcgg
5	61	ggcaccgagg	tgttttccgg	ccgccctctg	gtcggccacc	caaagccgcg	ggcgtgatg
	121	atgggtgagg	agggggcggc	aagatttcgg	gcgcccctgc	cctgaacgcc	ctcagctgct
	181	gccgccgggg	ccgctccagt	gcctgcgaac	tctgaggagc	cgaggcgccg	gtgagagcaa
	241	ggacgctgca	aacttgcgca	gcgcgggggc	tgggattcac	gcccagaagt	tcagcaggca
	301	gacagtccga	agccttcccc	cagcggagag	atagcttgag	ggtgcgcaag	acggcagcct
10	361	ccgccctcgg	ttcccgccca	gaccgggcag	aagagcttgg	aggagccaaa	aggaacgcaa
	421	aaggcggcca	ggacagcgtg	cagcagctgg	gagccgccgt	tctcagcctt	aaaagtgtgca
	481	gagattggag	gctgccccga	gaggggacag	accccagctc	cgactgcggg	gggcaggaga
	541	ggacggtacc	caactgccac	ctcccttcaa	ccatagtagt	tcctctgtat	cgagcgcagc
	601	gagctacaga	cgggggcgcg	gcactcttcg	cgagagcgcg	gaggctcaag	gtcccagcca
15	661	gtgagccag	tgtgcttgag	tgtctctgga	ctcgcccctg	agcttccagg	tctgtttcat
	721	ttagactcct	gctgcctcc	gtgcagttgg	gggaaagcaa	gagacttgcg	cgcacgcaca
	781	gtcctctgga	gatcaggtgg	aaggagccgc	tgggtacca	ggactgttca	gagcctcttc
	841	ccatctcggg	gagagcgaag	ggtgaggctg	ggcccgagag	gcagtgtaaa	cggcctcctc
	901	cggcgggatg	ggagccatcg	ggctcctgtg	gctcctgccg	ctgctgcttt	ccacggcagc
20	961	tgtgggctcc	gggatgggga	ccggccagcg	cgcgggctcc	ccagctgcgg	ggccgcgct
	1021	gcagcccg	gagccactca	gtactctcgc	cctgcagagg	aagagtctgg	cagttgactt
	1081	cgtggtgccc	tcgctcttcc	gtgtctacgc	ccgggacctt	ctgctgccac	catcctctc
	1141	ggagctgaag	gctggcaggc	ccgaggcccg	cggtctgcta	gctctggact	gcgcccgcct
	1201	gctcaggttg	ctggggcccg	cgccgggggt	ctcctggacc	gccggttcac	cagcccgggc
25	1261	agaggcccg	acgctgtcca	gggtgctgaa	ggcgcgctcc	gtgcgcaagc	tcggcgctgc
	1321	caagcagttg	gtgctggagc	tgggcgagga	ggcgatcttg	gaggggttgc	tcgggcccc
	1381	cggggagggc	gctgtggggc	tgtctcagtt	caatctcagc	gagctgttca	gttgggtgat
	1441	tcgccaaagg	gaaggcgac	tgaggatccg	cctgatgccc	gagaagaagg	cgtcgggaag
	1501	gggcagagag	ggaaggctgt	ccgcggcaat	tcgcgcctcc	cagccccgcc	ttctcttcca
30	1561	gatcttcggg	actggtcata	gtccttggga	atcaccaaca	aacatgcctt	ctccttctcc
	1621	tgattatttt	acatggaatc	tcacctggat	aatgaaagac	tccttccctt	tcctgtctca
	1681	tcgcagccga	tatggtctgg	agtgcagctt	tgacttcccc	tgtgagctgg	agtattcccc
	1741	tccactgcat	gacctcagga	accagagctg	gtcctggcgc	cgcattcccc	ccgaggaggg
	1801	ctcccagatg	gacttgctgg	atgggcctgg	ggcagagcgt	tctaaggaga	tgcccagagg
35	1861	ctcctttctc	cttctcaaca	cctcagctga	ctccaagcac	accatcctga	gtccgtggat
	1921	gtactgcagc	agtgcagact	gcacactggc	cgtctcgggt	cacaggcacc	gtcctctctc
	1981	tggaaaggtac	attgcccagc	tgtgcctcca	caacgaggct	gcaagagaga	tcctcctgat
	2041	gcccactcca	gggaagcatg	gttgacagct	gctccaggga	agaatcgggc	gtccagacaa
	2101	cccatttoga	gtggccctgg	aatacatctc	cagtggaaac	cgcagcttgt	ctgcagtggg
40	2161	cttctttgcc	ctgaagaact	gcagtgaagg	aacatcccca	ggctccaaga	tggccctgca
	2221	gagctccttc	acttgttgga	atgggacagt	cctccagctt	gggcaggcct	gtgacttcca
	2281	ccaggactgt	gcccaggag	aagatgagag	ccagatgtgc	cggaaactgc	ctgtgggttt
	2341	ttactgcaac	tttgaagatg	gttctgtgg	ctggacccaa	ggcacactgt	cacccacac
	2401	tcctcaatgg	caggtcagga	ccctaaagga	tgcgcggttc	caggaccacc	aagaccatgc
45	2461	tctattgctc	agtaccactg	atgtccccgc	ttctgaaagt	gctacagtga	ccagtgtctac
	2521	gtttcctgca	ccgatcaaga	gctctccatg	tgagctccga	atgtcctggc	tcattcgtgg
	2581	agtcttgagg	ggaaacgtgt	ccttggtgct	agtggagaac	aaaaccggga	aggagcaagg
	2641	caggatggtc	tggcatgtcg	ccgcctatga	aggtctgagc	ctgtggcagt	ggatggtgtt
	2701	gcctctcctc	gatgtgtctg	acaggttctg	gctgcagatg	gtcgcagtgg	ggggacaagg
50	2761	atccagagcc	atcgtggctt	ttgacaatat	ctccatcagc	ctggactgct	acctcaccat
	2821	tagcggagag	gacaagatcc	tgcagaatac	agcaccctaa	tcaagaaacc	tgtttgagag
	2881	aaacccaaac	aaggagctga	aaaccgggga	aaattcacca	agacagaccc	ccatctttga
	2941	ccctacagtt	cattggctgt	tcaccacatg	tggggccagc	gggccccatg	gccccacca
	3001	ggcacagtgc	aacaacgcct	accagaactc	caacctgagc	gtggaggtgg	ggagcgaggg
55	3061	ccccctgaaa	ggcatccaga	tctggaaggt	gccagccacc	gacacctaca	gcattctcgg
	3121	ctacggagct	gctggcgggg	aaggcgggga	gaacaccatg	atgcgggtccc	acggcggtgc
	3181	tgtgtcgggc	atcttcaacc	tggagaagga	tgacatgctg	tacatctcgg	tgggcagaca
	3241	gggagaggac	gcctgcccc	gtacaaacca	gttaatccag	aaagtctgca	ttggagagaa
	3301	caatgtgata	gaagaagaaa	tcctgtgtga	cagaagcgtg	catgagtggg	caggaggcgg
60	3361	aggaggaggg	ggtggagcca	cctacgtatt	taagatgaag	gatggagtgc	cgtgccccct

3421 gatcattgca gccggaggtg gtggcagggc ctacggggcc aagacagaca cgttccaccc
 3481 agagagactg gagaataact cctcggttct agggctaaac ggcaattccg gagccgcagg
 3541 tgggtggaggt ggctggaatg ataacacttc cttgctctgg gccggaaaaat ctttgcagga
 3601 ggggtgccacc ggaggacatt cctgccccca ggccatgaag aagtgggggt gggagacaag
 5 3661 aggggggttc ggagggggtg gaggggggtg ctccctcaggt ggaggaggcg gaggatatat
 3721 aggcggcaat gcagcctcaa acaatgaccc cgaaatggat ggggaagatg gggtttccct
 3781 catcagtcca ctgggcatcc tgtacacccc agcttttaaaa gtgatggaag gccacgggga
 3841 agtgaatatt aagcattatc taaactgcag tccactgtgag gtagacgaat gtacatgga
 3901 ccctgaaagc cacaagggtca tctgcttctg tgaccacggg acggtgctgg ctgaggatgg
 10 3961 cgtctcctgc attgtgtcac ccaccccgga gccacacctg ccactctcgc tgatcctctc
 4021 tgtgtgacc tctgccctcg tggcgccctg ggtcctggct ttctccggca tcatgattgt
 4081 tactcccgga aagcaccagg agctgcaagc catgcagatg gagctgcaga gccctgagta
 4141 caagctgagc aagctccgca cctcgacccat catgaccgac tacaacccca actactgctt
 4201 tgctggcaag acctcctcca tcagtgcact gaaggaggtg ccgcggaata acatcacctt
 15 4261 cattcgggggt ctgggccatg ggccttttgg ggaggtgtat gaaggccagg tgtccggaat
 4321 gcccaacgac ccaagccccc tgcaagtggc tgtgaagacg ctgcctgaag tgtgctctga
 4381 acaggacgaa ctggatttcc tcatggaagc cctgatcatc agcaaattca accaccagaa
 4441 catgtttcgc tgcatggggg gagacactca agtccctcct ccgagagacc cgccctcgcc ctgaccagcc
 4501 catggcgggg gccatgctgg accttctgca cgtggctcgg gacattgcct gtggctgtca
 20 4561 ctccctcctg gccatgctgg accttctgca cgtggctcgg gacattgcct gtggctgtca
 4621 gtatttggag gaaaaccact tcatccaccg agacattgct gccagaaact gcctcttgac
 4681 ctgtccaggc cctggaagag tggccaagat tggagacttc gggatggccc gagacatcta
 4741 cagggcgagc tactatagaa agggaggctg tgccatgctg ccagttaagt ggatgcccc
 4801 agaggccttc atggaaggaa tattcacttc taaaacagac acatggtcct ttggagtgtt
 25 4861 gctatgggaa atcttttctc ttggatatat gccatacccc agcaaaagca accaggaagt
 4921 tctggagttt gtcaccagtg gaggccggat ggaccaccc aagaactgcc ctggcctgtt
 4981 ataccgata atgactcagt gctggcaaca tcagcctgaa gacaggccca actttgccat
 5041 ctttttggag aggattgaat actgcaccca ggaccggat gtaatcaaca ccgctttgcc
 5101 gatagaatat ggtccacttg tggaaagagga agagaaagtg cctgtgaggc ccaaggaccc
 30 5161 tgagggggtt cctcctctcc tggctctctca acaggcaaaa cgggaggagg agcgagccc
 5221 agctgcccc aacacctctg ctaccacctc ctctggcaag gctgcaaaaga aaccacagc
 5281 tgcagagatc tctgttcgag tccctagagg gccggccgtg gaagggggac acgtgaatat
 5341 ggcattctct cagtccaacc ctcttcgga gttgcacaag gtccacggat ccagaaacaa
 5401 gccaccagc ttgtggaacc caacgtacgg ctctgggttt acagagaaac ccacaaaaaa
 35 5461 gaataatcct atagcaaaaga aggagccaca cgacaggggt aacctggggc tggagggaag
 5521 ctgtactgtc ccacctaacg ttgcaactgg gagacttccg ggggcctcac tgctcctaga
 5581 gccctcttcg ctgactgcca atatgaagga ggtacctctg ttcaggctac gtacttccc
 5641 ttgtgggaat gtcaattacg gctaccagca acagggcttg cccttagaag ccgctactgc
 5701 ccctggagct ggtcattacg aggataccat tctgaaaagc aagaatagca tgaaccagcc
 40 5761 tgggcccctga gctcggtcgc acactcactt ctcttctctg ggtccctaa gaccgtggag
 5821 gagagagagg caatggctcc ttcacaaacc agagacaaa tgtcacgttt tgttttgtgc
 5881 caacctatct tgaagtacca ccaaaaaagc tgtattttga aaatgcttta gaaaggtttt
 5941 gagcatgggt tcatcctatt ctttcgaaag aagaaaatat cataaaaatg agtgataaat
 6001 acaaggccca gatgtggttg cataaggttt ttatgcatgt ttgttgataa cttccttatg
 45 6061 cttctttcaa attgtgtgtg ctctgcttca atgtagtcag aattagctgc ttctatgttt
 6121 catagttggg gtcatagatg tttccttgcc ttgttgatgt ggacatgagc catttgaggg
 6181 gagagggaac ggaaataaag gatttatttg taatgactaa aa

50 Wild Type cDNA sequence TGC (4373 to 4375) codon mutation(s) encoding
 an amino acid other than cysteine or a corresponding mutation in a
 homolog thereof

55 Wild Type cDNA sequence CTG (4493 to 4495) codon mutation(s) encoding
 an amino acid other than leucine or a corresponding mutation in a
 homolog thereof

Wild Type cDNA sequence G4374A mutation or a corresponding mutation in
 a homolog thereof

60 Wild Type cDNA sequence C4493A mutation or a corresponding mutation in
 a homolog thereof

Wild Type ALK Protein Sequence (NP_004295.2; GI:29029632):

1 mgaigllwll plllstaavg sgmggtgqrag spaagpplqp replsysrlq rkslavdfvv
 61 pslfrvyard lllppsssel kagrpeargs laldcapllr llgpapgvsw tagspapaea
 121 rtlslrvlkgg svrklrrakq lvlelgeeeai legcvgppge aavglqlqfnl selfswwirq
 181 gegrlrlrlm pekkasevgr egrlsaaaira sqprllfqif gtghsslesp tnmppspdy
 241 ftwnltwimk dsfpflshrs ryglecsfdf pceleysppl hdlrnqswsw rripseeasq
 301 mdlldgpgae rskemprgsf lllntsadsk htllspwmrs ssehctlavs vhrhlqpsgr
 361 yiaqlphne aareillmpt pgkhgwtvlq grigrpdnpf rvaleyissg nrslsavdfff
 421 alknscgts pgskmalqss ftcwngtvlq lgqacdfhq d caggedesqm crklpvgfyc
 481 nfedgfcgwt qgtlsphtpq wqvrtlkdar fqdhqdhall lsttdvpase satvtsatfp
 541 apiksspcel rmswlrvgvl rgnvslvlve nktgkeqgrm vwhvaayegl slwgwmvlp
 601 ldvsdrfwlq mvawwgqgsr aivafdnisi sldcyltisg edkilqntsp ksrnlfernp
 661 nkelkpgens prqtpifdpt vhwlfittcga sgphgptqaq cnnayqnsnl svevgsegpl
 721 kgiqiwbkpa tdtysisgyg aaggkkgknt mmrshgvsvl gifnlekddm lyilvgqqge
 781 dacpstnqli qkvciqennv ieeerivnrs vhewaggggg gggatyvfk m kdgvpvplii
 841 aagggggrayg akttdtfhper lennssvlgl ngngsaagg ggwndntsll wagkslqega
 901 tggghscpam kkwgwetrgg fggggggcsc ggggggyig naasndp m dgedgvsfis
 961 plgilytpal kvmeghgevn ikhylvncshc evdechmdpe shkvicfcdh gtvlaedgvs
 1021 civsptpeph lplslilsvv tsalvaalvl afsgimivyr rkhqelqamq melqscpeykl
 1081 sklrtstimt dynpnycfag ktssisdike vprknitlir glghgafgev yegqvsmpn
 1141 dpsplqvavk tlpevcseqd eldfilmeali iskfnhqniv rcigvslqsl prfillelma
 1201 ggdlksflre trprpsqpss lamldllhva rdiacgcqyl eenhfihrdi aarnclltcp
 1261 gpgrvakigd fgmardiyra syyrkggcam lpvkwmppa fmegiftskt dtwsfgvllw
 1321 eifslgym py psksnqevle fvtsggrmdp pknpcgpyr imtqcwqhqp edrpnfaiil
 1381 erieyctqdp dvintalpie ygplveeeek vpvprkdpeg vppllvsgga kreeerspaa
 1441 ppplpttssg kaakkptaee isvrprgpa vegghvnmaf sqsnppselh kvhgsrnkpt
 1501 slwnptygsw ftekptkkn piakkephdr gnlglegsct vppnvatgrl pgasllleps
 1561 sltanmkevp lfrlrhfpcg nvnygyqqqg lpleaatapg aghyedtilk sknsmnqpgp

Wild Type protein sequence Cys1156Xaa mutation wherein Xaa is an amino acid other than cysteine or a corresponding mutation in a homolog thereof

Wild Type protein sequence Leu1196Xaa mutation wherein Xaa is an amino acid other than leucine or a corresponding mutation in a homolog thereof

Wild Type protein sequence Cys1156Tyr mutation or a corresponding mutation in a homolog thereof

Wild Type protein sequence Leu1196Met mutation or a corresponding mutation in a homolog thereof

EML4-ALK Variant 1 cDNA Sequence (AB274722.1; GI:152002652)

1 ggccggcgcg cgccggcgctc gcggctgctg cctgggaggg aggccgggca ggccgctgag
 61 cgccggcgct ctcaacgtga cggggaagtg gttcggcgcg ccgcggtta ctacccagg
 121 gcgaacggac ggacgacgga ggccggagcc ggtagccgag ccggcgac tagagaacga
 181 gcgggtcagg ctacgctcg gccactctgt cggctccgctg aatgaagtgc ccgccctct
 241 gagcccgag cccggcgctt tcccgcga atggacggtt tcgcccgcag tctcgatgat
 301 agtatttctg ctgcaagtac ttctgatgtt caagatcgcc tgtcagctct tgagtacga
 361 gttcagcaac aagaagatga aatcactgtg ctaaaggcgg ctttggtga tgttttgagg
 421 cgtcttgcaa tctctgaaga tcatgtggcc tcagtgaata aatcagctct aagtaaaggc
 481 caaccaagcc ctcgagcagt tattcccatg tctgtataa ccaatggaag tgggtgcaaac
 541 agaaaaccaa gtcataccag tgctgtctca attgcaggaa aagaaactct ttcattctgt
 601 gctaaaagtg gtacagaaaa aaagaaagaa aaaccacaa gacagagaga aaaaaagag
 661 gaatctcatt ctaatgatca aagtcacaaa attcgagcat caccttctcc ccagccctct
 721 tcacaacctc tccaaataca cagacaaact ccagaaagca agaatgctac tccacacaaa
 781 agcataaaac gaccatcacc agctgaaaag tcacataatt cttgggaaaaa ttcagatgat

841 agccgtaata aattgtcgaa aataccttca acacccaaat taataccaaa agttacccaaa
 901 actgcagaca agcataaaga tgatcatc acaccaagaag gagaatatat taaaatgttt
 961 atgcgcggtc ggccaattac catgttcatt ccttccgatg ttgacaacta tgatgacatc
 1021 agaacggaac tgcctcctga gaagctcaaa ctggagtgagg catatgggta tgcaggaaag
 5 1081 gactgtagag ctaatgttta ccttcttccg accggggaaa tagtttattt cattgcatca
 1141 gtagtagtac tttttaatta tgaggagaga actcagcgac actacctggg ccatacagac
 1201 tgtgtgaaat gccttgctat acatcctgac aaaattagga ttgcaactgg acagatagct
 1261 ggcgtggata aagatggaag gcctctacaa cccacgtca gagtgtggga ttctgttact
 1321 ctatccacac tgcagattat tggacttggc acttttgagc gtggagtagg atgcctggat
 10 1381 ttttcaaaag cagattcagg tgttcattta tgtgttattg atgactccaa tgagcatatg
 1441 cttactgtat gggactggca gaagaaagca aaaggagcag aaataaagac aacaaatgaa
 1501 gttgttttgg ctgtggagtt tcacccaaca gatgcaaata ccataattac atgcggtaaa
 1561 tctcataattt tcttctggac ctggagcggc aattcactaa caagaaaaca gggaattttt
 1621 gggaaatatg aaaagccaaa atttgtgcag tgttttagcat tcttggggaa tggagatggt
 15 1681 cttactggag actcaggtgg agtcatgctt atatggagca aaactactgt agagccaca
 1741 cctgggaaag gacctaagt gtaccgcccg aagcaccagg agctgcaagc catgcagatg
 1801 gagctgcaga gccctgagta caagctgagc aagctccgca cctcgacct catgaccgac
 1861 tacaacccca actactgctt tgcctggcag acctcctcca tcagtgcact gaaggaggtg
 1921 ccgcgcaaaa acatcacctt cattcggggg ctgggccatg gagcctttgg gagggtgtat
 20 1981 gaaggccagg tgtccggaat gcccaacgac ccaagccccc tgcaagtggc tgtgaagacg
 2041 ctgcctgaag tgtgctctga acaggacgaa ctggatttcc tcatggaagc cctgatcatc
 2101 agcaaatcca accaccagaa cattgttcgc tgcattgggg tgagcctgca atccctgccc
 2161 cggttcatcc tgctggagct catggcgggg ggagacctca agtccttcc ctagagagacc
 2221 cgccctcgcc cgagccagcc ctctcctctg gccatgctgg acctctgca cgtggctcgg
 25 2281 gacattgcct gtggctgtca gtatttggag gaaaaccact tcaccaccg agacattgct
 2341 gccagaaact gcctcttgac ctgtccaggc cctggaagag tggccaagat tggagacttc
 2401 gggatggccc gagacatcta caggcgagc tactatagaa agggaggtg tggcatgctg
 2461 ccagttaagt ggatgcccc agaggccttc atggaaggaa tattcacttc taaaacagac
 2521 acatggtcct ttggagtgtc gctatgggaa atcttttctc ttggatatat gccatacccc
 30 2581 agcaaaaagca accaggaagt tctggagttt gtcaccagtg gaggcgggat ggacccaccc
 2641 aagaactgcc ctgggctgt ataccggata atgaactcagt gctggcaaca tcagcctgaa
 2701 gacaggccca actttgcat cattttggag aggattgaat actgcacca ggacccggat
 2761 gccaatcaaca ccgctttgcc gatagaatat ggtccacttg tggaaagga agagaaagtg
 2821 cctgtgaggc ccaaggaccc tgaggggtt cctcctctcc tggctctca acaggcaaaa
 35 2881 cgggaggagg agcgcagccc agctgcccc caacctctgc ctaccacctc ctctggcaag
 2941 gctgcaaaga aaccacagc tgcagaggtc tctgttcgag tccctagagg gccggccgtg
 3001 gaagggggac acgtgaatat ggcattctct cagtccaacc ctcttcgga gttgcacagg
 3061 gtccacggat ccagaaacaa gccaccagc ttgtggaacc caacgtacg ctctgggtt
 3121 acagagaaac ccacaaaaa gaataatcct atagcaaaga aggagccaca cgagaggggt
 40 3181 aacctggggc tggaggaag ccttactgtc ccactaacg ttgcaactgg gagacttccg
 3241 ggggcctcac tgctcctaga gccctcttgc ctgactgcca atatgaagga ggtacctctg
 3301 ttcaggctac gtcacttccc ttgtgggaat gtcaattacg gctaccagca acagggttg
 3361 cccttagaag ccgtactgc cctggagct ggtcattacg aggataccat tctgaaagc
 3421 aagaatagca tgaaccagc tgggcccctga gctcggtcac acactcactt ctcttccctg
 45 3481 ggatccctaa gacgtggag gagagagagg caatcaatgg ctcttcaca aaccagagac
 3541 caaatgtcac gttttgttt gtgccaacct attttgaagt accacaaaaa aagctgtatt
 3601 ttgaaaatgc tttagaaagg ttttagacat gggttcatcc tattctttcg aaagaagaaa
 3661 atatcataaa atgagtgat aaatacaagg ccagatgtg gttgcataag gtttttatgc
 3721 atgtttgttg tatacttctt tatgttctt ttaaattgtg tgtgctctgc ttcaatgtag
 50 3781 tcagaattag ctgcttctat gtttcatagt tggggtcata gatgtttcct tgccttgttg
 3841 atgtggacat gagccatttg aggggagagg gaacggaaat aaaggagtta tttgtaatga
 3901 aaaaaaaaaa aaaaaaaaaa aaaaaa

**EML4-ALK Variant 1 Protein Sequence (BAF73611.1;
GI:152002653)**

55 1 mdgfagsldv sisaastsdv qdrlsalesr vqqgedeitv lkaaladvlr rlaisedhva
 61 svkksvsskg qpspravipm scitngsgan rkpshtsavs iagketlssa aksgtekkke
 121 kpqggrekke eshsndqspq iraspspqs sqplqihrt pesknaptk sikrpspaek
 181 shnswensdd srnklskips tpklipkvtk tadkhkdvii ngegeyikmf mrgrpitmfi
 60 241 psdvdnyddi rtelppekkl lewaygyrgk dcranvylp tgeivyfias vvvlfnyeer

301 tqrhylghtd cvkclaihp d kiriatgqia gvdkdgrplq phvrwdsvt lstlqiiglg
 361 tfergvgcld fskadsgvhl cviddsnehm ltvwdwqkka kgaeiktne vvlavefhpt
 421 dantiitcgk shiffwtwsg nsltrkqgif gkyekpkfvq clafllngdv ltgdsggvmll
 481 iwsktttvept pgkgpkvyrr khqelqamqm elqspeykls klrtsttimd ynpnycfagk
 5 541 tssisdlikev prknitlirg lghgafgevy egqvsgmpnd psplqvavkt lpevcseqde
 601 ldflmealii skfnhqniwr cigvslqslp rfillelmag gdlksflret rprpsqpssl
 661 amlldllhvar diacgcqyle enhfihrdia arnclltcbp pgrvakigdf gmardiyas
 721 yyrkggcaml pvkwmppeaf megiftsktd twsfgvllwe ifslgympyp sksnqevlef
 781 vtsggrmdpp knpcgpyvri mtqcwqhqpe drpnfaiile rieyctqdpd vintalpiey
 10 841 gplveeeekv pvrpkdpegv ppllvsqqak reeerspaap pplpttssgk aakktptaaev
 901 svrvprgpav egghvnmfafs qsnppselhr vhgssrnkpts lwnptygswf tekptkknnp
 961 iakkepherg nlglegsctv ppnvatgrlp gaslllepss ltanmkevpl frlrhfpogn
 1021 vnygyqqqgl pleaatapga ghyedtilks knsmnqpgp

15 **EML4-ALK Variant 2 cDNA Sequence (AB275889.1;
 GI:152002654)**

1 ggccggcgcg cgcggcgctc gcggctgctg cctgggaggg aggccgggca ggcggctgag
 61 cggcgcggtt ctcaacgtga cggggaagtg gtcggggcgg ccgcggttca ctaccccgag
 121 gcgaacggac ggacgacgga ggccggagcc ggtagccgag ccggcgacac tagagaacga
 20 181 gcgggtcagg ctcagcgtcg gccactctgt cggctccgctg aatgaagtgc ccgcccctct
 241 gagcccgag cccggcgctt tcccgcgaag atggacggtt tcgcccggag tctgatgat
 301 agtatttctg ctgcaagtac ttctgatgtt caagatcgcc tgtcagctct tgatgcacga
 361 gttcagcaac aagaagatga aatcactgtg ctaaaggcgg ctttggttga tgttttgagg
 421 cgtcttgcaa tctctgaaga tcatgtggcc tcagtgaata aatcagctct aagtaaaggc
 25 481 caaccaagcc ctgagcagc tattcccatg tctgtataa ccaatggaag tgggtcaaac
 541 agaaaaccaa gtcataccag tgctgtctca attgcaggaa aagaaaactct ttcactctgt
 601 gctaaaagtg gtacagaaaa aaagaaagaa aaaccacaag gacagagaga aaaaaagag
 661 gaatctcatt ctaatgatca aagtcacaaa attcagcagc caccttctcc ccagccctct
 721 tcacaacctc tccaaataca cagcaaaact ccagaaagca agaattgtac tcccacaaa
 30 781 agcataaaac gaccatcacc agctgaaaag tcacataatt cttgggaaaa ttcagatgat
 841 agccgtaata aattgtcgaa aataccttca acacccaaat taataccaaa agttacaaa
 901 actgcagaca agcataaaga tgctcatcct aaccaagaag gagaatatat taaaatgttt
 961 atgcgcggtc ggccaattac catgttcatt ccttccgatg ttgacaacta tgatgacatc
 1021 agaacggaac tgccctctga gaagctcaaa ctggagtggg catatgggta tcgaggaaaag
 35 1081 gactgtagag ctaattgtta ccttcttcga accggggaaa tagtttattt tcatgcatca
 1141 gtagtgtagc tatttaatta ttaggagaga actcagcgac actacctggg ccatacagac
 1201 tgtgtgaaat gccttgctat acatcctgac aaaattagga ttgcaactgg acagatagct
 1261 ggcgtggata aagatggaag gcctctacaa cccacgtcga gagggtggga ttctgttact
 1321 ctatccacac tgcagattat tggacttggc acttttgagc gtggagtagg atgcctggat
 40 1381 ttttcaaaag cagattcagg tgttcattta tgtgttattg atgactccaa tgagcatatg
 1441 cttactgtat gggactggca gaagaaagca aaaggagcag aaataaagac acaaatgaa
 1501 gttgttttgg ctgtggagtt tcacccaaca gatgcaata ccataattac atgcggtaaa
 1561 tctcatattt tcttctggac ctggagcggc aattcactaa caagaaaaca ggaattttt
 1621 gggaaatatg aaaagccaaa atttgtgcag tgttttagcat tcttggggaa tggagatgtt
 45 1681 cttactggag actcaggtgg agtcatgctt atatggagca aaactactgt agagcccaca
 1741 cctgggaaag gacctaaagg tgtatatcaa atcagcaaac aaatcaaagc tcatgatggc
 1801 agtgtgttca cactttgtca gatgagaaat gggatgttat taactggagg agggaaagac
 1861 agaaaaataa ttctgtggga tcatgatctg aatcctgaaa gagaaataga ggttctctgat
 1921 cagtatggca caatcagagc tgtagcagaa ggaaaggcag atcaattttt agtaggcaca
 50 1981 tcacgaaact ttattttacg aggaacattt aatgatggct tccaaataga agtacagggt
 2041 catacagatg agctttgggg tcttgccaca catcccttca aagatttgct cttgacatgt
 2101 gctcaggaca ggcaggtgtg cctgtggaac tcaatggaac acaggctgga atggaccagg
 2161 ctggtagatg aaccaggaca ctgtgcagat tttcatccaa gtggcacagt ggtggccata
 2221 ggaacgact caggcaggtg gtttgttctg gatgcagaaa ccagagatct agtttctatc
 55 2281 cacacagacg ggaatgaaca gctctctgtg atgogctact caatagatgg taccttctctg
 2341 gctgtaggat ctcagacaaa ctttattttac ctctatgtag tctctgaaaa tggaaagaaa
 2401 tatagcagat atggaaggtg cactggacat tccagctaca tcacacacct tgactggtcc
 2461 ccagacacaa agtatataat gtctaacctg ggagactatg aaatattgta cttgtaccgc
 2521 cggaagcacc aggagctgca agccatgcag atggagctgc agagccctga gtacaagctg
 60 2581 agcaagctcc gcacctcgac catcatgacc gactacaacc ccaactactg ctttctgtggc

2641 aagacctcct ccatcagtga cctgaaggag gtgccgcgga aaaacatcac cctcattcgg
 2701 ggtctggggc atggagccct tggggagggt tatgaaggcc aggtgtccgg aatgcccac
 2761 gacccaagcc cctgcaagt ggctgtgaag acgctgcctg aagtgtgctc tgaacaggac
 2821 gaactggatt tcctcatgga agccctgata atcagcaaat tcaaccacca gaacattgtt
 5 2881 cgctgcattg ggggtagcct gcaatccctg ccccggttca tcctgctgga gctcatggcg
 2941 gggggagacc tcaagtcctt cctccgagag acccgccctc gcccgagcca gccctcctcc
 3001 ctggccatgc tggaccttct gcacgtggct cgggacattg cctgtggctg tcagtatttg
 3061 gaggaaaacc acttcatcca ccgagacatt gctgccagaa actgcctctt gacctgtcca
 3121 ggccctggaa gagtggccaa gattggagac ttccgggatgg cccgagacat ctacagggcg
 10 3181 agctactata gaaagggagg ctgtgccatg ctgccagtta agtggatgcc cccagagggc
 3241 ttcatggaag gaatttcac gacacatggt cctttggagt gctgctatgg
 3301 gaaatccttt ctcttgata tatgccatac cccagcaaaa gcaaccagga agttctggag
 3361 tttgtcacca gtggaggccg gatggacca cccaagaact gccctgggccc tgtataccgg
 3421 ataatgactc agtgctggca acatcagcct gaagacaggc ccaactttgc catcattttg
 15 3481 gagaggattg aatactgcac ccaggaccog gatgtaatca acaccgcttt gccgatagaa
 3541 tatggccac ttgtggaaga ggaagagaaa gtgcctgtga ggcccaagga cctgagggg
 3601 gttcctcctc tcctggtctc tcaacaggca aaacgggagg aggagcgcag cccagctgcc
 3661 ccaccacctc tgcctaccac ctctctggc aaggtgcaa agaaaccac agctgcagag
 3721 gtctctgttc gagtccctag agggccggcc gtggaagggg gacacgtgaa tatggcattc
 20 3781 tctcagcca accctccttc ggagttgcac aggttccacg gatccagaaa caagcccacc
 3841 agcttgtgga acccaacgta cggctcctgg tttacagaga aaccaccaa aaagaataat
 3901 cctatagcaa agaaggagcc acacgagagg ggtaacctgg ggctggaggg aagctgtact
 3961 gtccaccta acgttgcaac tgggagactt cggggggcct cactgctcct agagccctct
 4021 tcgctgactg ccaatatgaa ggaggtacct ctgttcaggc tacgtcactt cccttgtggg
 25 4081 aatgtcaatt acggctacca gcaacagggc ttgcccttag aagccgctac tgcccctgga
 4141 gctggctatt acgaggatac cattctgaaa agcaagaata gcatgaacca ccttgggccc
 4201 tgagctcggc cacacactca cttctcttcc ttgggatccc taagaccgtg gaggagagag
 4261 aggcaatcaa tggctccttc acaaaccaga gaccaaagt cacgttttgt tttgtgcaa
 4321 cctattttga agtaccacca aaaaagctgt attttgaaaa tgcttttaga aggttttgag
 30 4381 catgggttca tcctattctt tcgaaagaag aaaatatcat aaaaatgagt gataaataca
 4441 aggccagat gtggttgcac aaggttttta tgcatgtttg ttgtatactt ccttatgott
 4501 cttttaaaatt gtgtgtgctc tgcttcaatg tagtcagaat tagctgcttc tatgtttcat
 4561 agttgggggc atagatgtt ccttgccttg ttgatgtgga catgagccat ttgaggggag
 4621 agggaaacgga aataaaggag ttatttgtaa tgaaaaaaa aaaaaaaa aaaaaaaa

35

**EML4-ALK Variant 2 Protein Sequence (BAF73612.1;
GI:152002655)**

1 mdgfagsldd sisaastsdv qdrlesalesr vqqgedeitv lkaaladvlr rlaisedhva
 61 svkksvsskg qpspravipm scitngsgan rkpshtsavs iagketlssa aksgtekkke
 40 121 kpqqgrekke eshsndqspq iraspspqpq sqplqihqrt pesknaptk sikrpspaek
 181 shnswensdd srnklskipk tpklipkvtk tadkhkdvii ngegeyikmf mrgrpitmfi
 241 psdvndyddi rtelppekik lewaygyrgk dcoranvllp tgeivyfias vvlfnyeer
 301 tqrhlyghtd cvkclaihpq kiriatgqia gvdkdgrplq phvrwdsvt lstlqiiglg
 361 tfergvgcld fskadsgvhl cviddsnehm ltvwdwqkka kgaeikttnv vvlavefhpt
 45 421 dantiitcgk shiffwtwsg nsltrkqgif gkyekpkfvq clafngngdv lgtsggvmml
 481 iwskttevep pgkgpkgyvq iskqikahdg svftlcqmrn gmltgggkd rkiilwdhdl
 541 npereievpd qygtiravae gkadqflvgt srnfilrgtf ndgfgievqg htdelwglat
 601 hpfkdlilltc aqdrqvclwn smehrlewtr lvdepgthead fhpsgtvvai gthsggrwfv
 661 daetrdlvsj htdgneqlsv mrysidgtfl avgshdnfiy lyvvsengrk ysrygrctgh
 50 721 ssyithldws pdnkyimsns gdeyilylyr rkqgelqamq melqspeytl sklrtstimt
 781 dynpnycfag ktssisdike vprknitlir glghgafgev yegqvsgmpn dpsplqvavk
 841 tlpevcseqd eldfmeali iskfnhqniv rcigvslqsl prfillelma ggdlksflre
 901 trprpsqpss lamldllhva rdiacgcqyl eenhfihrdi aarnclltcp gpgrvakigd
 961 fgmardiyra syyrkggcam lpvkwmppea fmegiftskt dtwsfgvllw eifslgympp
 55 1021 psksnqevle fvtsggrmdp pkncpgpvyr imtqcwqhqp edrpnfaiil erieyctqdp
 1081 dvintalpie ygplveeeek vpvprkdpeg vppllvsqqa kreeerspaa ppplpttssg
 1141 kaakkptaae vsrvprgpa vegghvnmaf sqsnppselh rvhgsnrkpt slwnptygsw
 1201 ftekpkknn piakkepher gnlglegscet vppnvatgrl pgaslllept sltanmkevp
 1261 lfrrlhfpcc nvnygyqqgg lpleaatapg aghyedtilk sknsnmqppg

60

**EML4-ALK Variant 3a Nucleic Acid Sequence (AB374361.1;
GI:194072592)**

1 actctgtcgg tccgctgaat gaagtgcgg ccctctaaag cccggagccc ggcgctttcc
 5 61 ccgcaagatg gacggtttcg ccggcagctc cgatgatagt atttctgctg caagtacttc
 121 tgatgttcaa gatcgctgt cagctcttga gtcacgagtt cagcaacaag aagatgaaat
 181 cactgtgcta aaggcggctt tggctgatgt ttgaggcgt cttgcaatct ctgaagatca
 241 tgtggcctca gtgaaaaaat cagctcgaag taaaggccaa ccaagccctc gagcagttat
 301 tcccatgtcc tgtataacca atggaagtgg tgcaaacaga aaaccaagtc ataccagtgc
 361 tgtctcaatt gcaggaaaag aaactctttc atctgctgct aaaagtggta cagaaaaaaa
 10 421 gaaagaaaaa ccacaaggac agagagaaaa aaaagaggaa tctcattcta atgatcaag
 481 tccacaaatt cgagcatcac cttctcccca gccctcttca caacctctcc aaatacacag
 541 acaaaactcca gaaagcaaga atgtactctc caccaaaaagc ataaaaagac catcaccagc
 601 tgaaaagtca cataattctt gggaaaaatt agatgatagc cgtaataaat tgtcgaaaat
 661 actttcaaca cccaaattaa taccaaaagt taccaaaact gcagacaagc ataaagatgt
 15 721 catcatcaac caagtgtacc gccggaagca ccaggagctg caagccatgc agatggagct
 781 gcagagccct gagtacaagc tgagcaagct ccgcacctcg accatcatga ccgactacaa
 841 ccccaactac tgctttgctg gcaagacctc ctccatcagt gacctgaagg aggtgccgcy
 901 gaaaaacatc accctcattc ggggtctggg ccatggagcc ttgggggagg tgatgaagg
 961 ccagggtgtcc ggaatgcccc acgacccaag cccctgcaa gtggctgtga agacgctgcc
 20 1021 tgaagtgtgc tctgaacagg acgaactgga tttctcatg gaagccctga tcatcagcaa
 1081 attcaaccac cagaacattg ttcgctgcac tgggggtgagc ctgcaatccc tgccccggtt
 1141 catcctgctg gagctcatgg cggggggaga cctcaagtcc ttcctccgag agaccgcccc
 1201 tcgcccagac cagccctcct cctggccat gctggacctt ctgcacgtgg ctccgggacat
 1261 tgcctgtggc tgtcagtatt tggaggaaaa ccacttcac caccgagaca ttgctgccag
 25 1321 aaactgcctc ttgacctgtc caggccctgg aagagtggcc aagattggag acttcgggat
 1381 ggcccagagc atctacaggc cgagctacta tagaaaggga ggctgtgcca tgcctccagt
 1441 taagtggatg cccccagagg ccttcattga aggaatattc acttctaaa cagacacatg
 1501 gtcctttgga gtgctgctat gggaaatctt ttctcttggga tatatgccat accccagcaa
 1561 aagcaaccag gaagttctgg agtttgtcac cagtggaggc cggatggacc caccagaaga
 30 1621 ctgccctggg cctgtatacc ggataatgac tcagtgtctg caacatcagc ctgaagacag
 1681 gcccaacttt gccatcattt tggagaggat tgaatactgc acccaggacc cggatgtaat
 1741 caacaccgct ttgccgatag aatatggctc acttgtggaa gaggaagaga aagtgcctgt
 1801 gaggcccaag gacctgagg ggggttctcc tctcctggtc tctcaacagg caaaacggga
 1861 ggaggagcgc agcccagctg cccaccacc tctgcctacc acctcctctg gcaaggctgc
 35 1921 aaagaaacc acagctgcag aggtctctgt tcgagtcctt agagggccgg ccgtggaagg
 1981 gggacacgtg aatatggcat tctctcagtc caacctcct tcggagttgc acaggggtcca
 2041 cggatccaga aacaagccca ccagcttctg gaaccaacg tacggctcct ggtttacaga
 2101 gaaaccacc aaaaagaata atcctatagc aaagaaggag ccacacgaga ggggtaacct
 2161 ggggctggag ggaagctgta ctgtccacc taacgttgca actgggagac ttccgggggc
 40 2221 ctcaactgct ctagagccct cttcgtgac tgccaatatg aaggaggtac ctctgttcag
 2281 gctacgtcac ttcccttctg ggaatgtcaa ttacggctac cagcaacagg gcttgccctt
 2341 agaagccgct actgcccctg gagctggctc ttacgaggat accattctga aaagcaagaa
 2401 tagcatgaac cagcctgggc cctgagctcg gtgcacact cacttctctt ccttgggatc
 2461 cctaagaccg tgg

45

**EML4-ALK Variant 3a Protein Sequence (BAG55003.1;
GI:194072593)**

1 mdgfagsldd sisaastsdv qdrlesalesr vqqgedeitv lkaaladvlr rlaisedhva
 50 61 svkksvsskg qpspravipm scitngsgan rkpshtsavs iagketlssa aksgtekkke
 121 kpqqgrekke eshsndqspq iraspspqqp sqplqihqrt pesknaptk sikrpspaek
 181 shnswensdd srnklskipk tpklikvkt tadkhkdvii nqvyrrkhqe lqamqmelqs
 241 peyklsklrl stimtdynpn ycfagktssi sdlkevprkn itlirlglgh afgevyegqv
 301 sgmpndpspl qvavktlpev cseqdeldfl mealiiskfn hqnivrcigv slqslprfil
 361 lelmaggdlk sflretrprp sqpsslamld llhvardiac gcqyleenhf ihrdiaarnc
 55 421 lltcpggprv akigdfgmar diyasyyrk ggcamlpvkw mppeafmegi ftsktdtwsf
 481 gvllweifsl gymypsksn qevlefvtsg grmdppkncp gpvyrimtqc wqhqpdrpn
 541 failleriey ctqdpdvint alpieygplv eeeekvpvrp kdpegvppll vsqgakreee
 601 rspaappplp ttssgkaakk ptaaevsrv prgpaveggh vnmafsqsnp pselhrvhgs
 661 rnkptslwnp tygswftekp tkknnpiakk ephergnlgl egscvppnv atgrlpgasl
 60 721 llepssltan mkevplflrl hfpcgnvnyg yqqqglplea atapgaghye dtiilksnsm

781 nqpgp

**EML4-ALK Variant 3b Nucleic Acid Sequence (AB374362.1;
GI:194072594)**

5 1 actctgtcgg tccgctgaat gaagtgcccg cccctctaag cccggagccc ggcgctttcc
 61 ccgcaagatg gacggtttcg cgggcagtct cgatgatagt atttctgctg caagtacttc
 121 tgatgttcaa gatcgctgt cagctcttga gtcacgagtt cagcaacaag aagatgaaat
 181 cactgtgcta aaggcggcct tggctgatgt tttgaggcgt cttgcaatct ctgaagatca
 241 tgtggcctca gtgaaaaaat cagtctcaag taaaggccaa ccaagccctc gagcagttat
 10 301 tcccatgtcc tgtataacca atggaagtgg tgcaaacaga aaaccaagtc ataccagtgc
 361 tgtctcaatt gcaggaaaag aaactctttc atctgctgct aaaagtggta cagaaaaaaa
 421 gaaagaaaaa ccacaaggac agagagaaaa aaaaggaggaa tctcatttca atgatcaaaag
 481 tccacaaatt cgagcatcac cttctcccca gccctcttca caacctctcc aaatacacag
 541 acaaaactca gaaagcaaga atgctactcc caccaaaagc ataaaacgac catcaccagc
 15 601 tgaaaagtca cataattctt gggaaaattc agatgatagc cgtaataaat tgtcgaaaat
 661 accttcaaca cccaaattaa taccaaaagt taccaaaact gcagacaagc ataaagatgt
 721 catcatcaac caagcaaaaa tgtcaactcg cgaaaaaaac agccaagtgt accgccggaa
 781 gcaccagag ctgcaagcca tgcagatgga gctgcagagc cctgagtaca agctgagcaa
 841 gctccgcacc tcgaccatca tgaccgacta caacccccaa tactgctttg ctggcaagac
 20 901 ctctccatc agtgacctga aggaggtgcc gcggaaaaac atcacctca ttcggggtct
 961 gggccatgga gcctttgggg aggtgtatga aggccaggtg tccggaatgc ccaacgaccc
 1021 aagccccctg caagtggctg tgaagacgct gcctgaagtg tgctctgaac aggacgaact
 1081 ggatttcctc atggaagccc tgatcatcag caaattcaac caccagaaca ttgttcgctg
 1141 cattgggggtg agcctgcaat cctgccccg gttcctcctg ctggagctca tggcgggggg
 25 1201 agacctcaag tccttcctcc gagagaccog cctcgccccg agccagccct cctccctggc
 1261 catgctggac cttctgcacg tggctcggga cattgcctgt ggctgtcagt atttgaggga
 1321 aaaccacttc atccaccgag acattgctgc cagaaactgc ctcttgacct gtccaggccc
 1381 tggaagagtg gccaagattg gagacttcgg gatggccga gacatctaca gggcgagcta
 1441 ctatagaaag ggaggtgtg ccatgctgcc agttaagtgg atgccccag aggccttcat
 30 1501 ggaaggaata ttcacttcta aaacagacac atggctcttt ggagtgtgc tatgggaaat
 1561 cttttctctt ggatatatgc cataccccag caaaagcaac caggaagttc tggagtttgt
 1621 caccagtgga ggccggatgg acccacccaa gaactgccct gggcctgtat accggataat
 1681 gactcagtgc tggcaacatc agcctgaaga caggcccaac tttgccatca ttttgagag
 1741 gattgaatac tgcacccagg acccggatgt aatcaacacc gctttgccga tagaatatgg
 35 1801 tccacttgtg gaagaggaag agaaagtgcc tgtgaggccc aaggacctg agggggttcc
 1861 tcctctcctg gtctctcaac aggcaaaacg ggaggaggag cgcagccca ctcgccacc
 1921 acctctgcct accacctcct ctggcaaggc tgcaaagaaa cccacagctg cagaggtctc
 1981 tgttcgagtc cctagagggc cggccgtgga agggggacac gtgaatatgg cattctctca
 2041 gtccaaccct ccttcggagt tgcacagggt ccacggatcc agaaacaagc ccaccagctt
 40 2101 gtggaacca acgtacggt cctggtttac agagaaaccc accaaaaaga ataactctat
 2161 agcaaagaag gagccacacg agaggggtaa cctggggctg gagggaagct gtactgtccc
 2221 acctaacgtt gcaactggga gacttcggg ggccctactg ctctagagc cctctctgct
 2281 gactgccaat atgaaggag tactctctgt caggctacgt cacttccct gtgggaatgt
 2341 caattacggc taccagcaac agggcttgcc cttagaagcc gctactgcc ctggagctgg
 45 2401 tcattacgag gataccattc tgaaaagcaa gaatagcatg aaccagcctg ggcctgagc
 2461 tcggtcgcac actcacttct cttccttggg atccctaaga ccgtgg

**EML4-ALK Variant 3b Protein Sequence (BAG55004.1;
GI:194072595)**

50 1 mdgfagsldd sisaastsdv qdrilsalesr vqqgedeitv lkaaladvlr rlaisedhva
 61 svkksvsskg qpspravipm scitngsgan rkpshtsavs iagketlssa aksgtekkke
 121 kpqgqrekke eshsndqspq iraspspqps sqplqihrt pesknaptk sikrpspaek
 181 shnswensdd srnklskips tpklipkvtk tadkhkdvii nqakmstrek nsqvyrrkhq
 241 elqamqmelq speyklsklr tstimtdynp nycfagktss isdlkevprk nitlirglgh
 55 301 gafgevyegq vsgmpndpsp lqvavktlpe vcseqdeldf lmealiiskf nhqnivrcig
 361 vslqslprfi llelmaggdl ksflretrpr psqpsslaml dlhvardia cgcqyleenh
 421 fihrdiaarn cltctpgpgr vakigdfgma rdiyrasyyr kggcamlpvk wmppeafmeg
 481 iftsktdtws fgvlleifs lgympysks nqevlefvtg ggrmdppknc pgpvyrmtq
 541 cwqhqpdrp nfaiilerie yctqdpdvin talpieygpl veeekvpvr pkdpegvppl
 60 601 lvsqqakree erspaapppl pttssgkaak kptaaevsvr vprgpavegg hvnmafsqsn

661 ppselhrvhg srnkptslwn ptygswftek ptkknnpiak kephergnlg legsctvppn
 721 vatgrlpgas lilepsslta nmkevplfrl rhfpcgnvny gyqqqgple aatapgaghy
 781 edtilkskns mnqpgp

5 **EML4-ALK Variant 4 Nucleic Acid Sequence (AB374363.1;
 GI:209837703)**

1 actctgtcgg tcgctgaat gaagtgccg ccctctaaag ccgagagccc ggcgctttcc
 61 cgcgaagatg gaagggttcg ccggcagctc cgatgatagt atttctgctg caagtacttc
 121 tgatgttcaa gatcgctgt cagctcttga gtcacgagtt cagcaacaag aagatgaaat
 10 181 cactgtgcta aaggcggctt tggctgatgt tttgaggcgt cttgcaatct ctgaagatca
 241 tgtggcctca gtgaaaaaat cagtctcaag taaaggccaa ccaagccctc gagcagttat
 301 tcccatgtcc tgtataacca atggaagtgg tgcaaacaga aaaccaagtc ataccagtgc
 361 tgtctcaatt gcaggaaaag aaactctttc atctgctgct aaaagtggta cagaaaaaaa
 421 gaaagaaaaa ccacaaggac agagagaaaa aaaagaggaa tctcattcta atgttcaaa
 15 481 tccacaaatt cgagcatcac cttctcccca gccctcttca caacctctcc aaatacacag
 541 acaaaactcca gaaagcaaga atgctactcc caccaaaagc ataaaacgac catcaccagc
 601 tgaaaagtca cataattctt gggaaaattc agatgatagc cgtaataaat tgcgaaaat
 661 accttcaaca cccaaattaa taccaaaagt taccaaaact gcagacaagc ataagatgt
 721 catcatcaac caagaaggag aatatattaa aatgtttatg cgcggtcggc caattacat
 20 781 gttcattcct tcgatgttg acaactatga tgacatcaga acggaactgc ctctgagaa
 841 gctcaaaactg gagtgggcat atggttatcg aggaaaggac tgtagagcta atgtttacct
 901 tcttccgacc ggggaaatag tttatttcat tgcacagta gtagtactat ttaattatga
 961 ggagagaact cagcgacact acctgggcca tacagactgt gtgaaatgcc ttgctataca
 1021 tctgacaaa attaggattg caactggaca gatagctggc gtggataaag atggaaggcc
 25 1081 tctacaacc cactcagag tgtgggattc tgttactcta tccactgc agattattgg
 1141 acttggcact tttgagcgtg gtagtaggag cctggatttt tcaaaagcag attcaggtgt
 1201 tcatttatgt gttattgatg actccaatga gcataatgctt actgtatggg actggcagag
 1261 gaaagcaaaa ggagcagaaa taaagacaac aaatgaagtt gttttggctg ttgagtttca
 1321 cccaacagat gcaaatacca taattacatg cggtaaatct catattttct tctggacctg
 30 1381 gagcggcaat tactaaciaa gaaaacaggg aatttttggg aaatatgaaa agccaaaatt
 1441 tgtgcagtgt ttagcattct tggggaatgg agatgttctt actggagact caggtggagt
 1501 catgcttata tggagcaaaa ctactgtaga gccacacctt gggaaaggac ctaaaggtgt
 1561 atatcaaatc agcaaacaaa tcaaagctca tgatggcagt gtgttcacac tttgtcagat
 1621 gagaaatggg atgttattaa ctggaggagg gaaagacaga aaaataattc tgtgggatca
 35 1681 tgacttgaat cctgaaagag aaatagagat atgctggatg agccctgagt acaagctgag
 1741 caagctcgc acctcgacca tcatgaccga ctacaacccc aactactgct tctgtggcaa
 1801 gacctctcc atcagtgacc tgaaggaggt gcccgggaaa aacatcacc tcatcgggg
 1861 tctgggccat ggagcctttg gggagggtga tgaaggccag gtgtccggaa tgcccaacga
 1921 cccaagcccc ctgcaagtgg ctgtgaagac gctgcctgaa gtgtgctctg aacaggacga
 40 1981 actggatttc ctcatggaag cctgatcat cagcaaattc aaccaccaga acattgttcg
 2041 ctgcattggg gtgagcctgc aatccctgcc ccggttcac cgtctggagc tcatggcggg
 2101 gggagacctc aagtccctcc tccagagagc ccgccctgc cggagccagc cctcctcct
 2161 ggccatgctg gaccttctgc acgtggctcg ggacattgcc tgtggtgtgc agtatttggg
 2221 ggaaaaccac ttcattccac gagacattgc tgccagaaac tgctcttga cctgtccagg
 45 2281 ccttgggaaga gtggccaaga ttggagactt cgggatggcc cgagacatct acagggcgag
 2341 ctactataga aaggagggtc gtgccatgct gccagttaag tggatgcccc cagaggcctt
 2401 catggaagga atattcactt ctaaaacaga cacatggtcc tttggagtgc tgcattggga
 2461 aatcttttct cttggatata tgccataccc cagcaaaagc aaccaagaag ttctggagtt
 2521 tgtcaccagt ggaggccgga tggaccacc caagaactgc cctgggcctg tataccggat
 50 2581 aatgactcag tgctggcaac atcagcctga agacaggccc aactttgcca tcattttggg
 2641 gaggattgaa tactgcacc aggaccgga tgtaataaac accgctttgc cgatagaata
 2701 tgggtccactt gtggaagagg aagagaaagt gcctgtgagg cccaaggacc ctgagggggt
 2761 tctcctcttc ctggtctctc aacaggcaaa acgggaggag gagcgcagcc cagctgcccc
 2821 accacctctg cctaccacct cctctggcaa ggctgcaaag aaaccacag ctgcagaggt
 55 2881 ctctgttcga gtccctagag ggccggcctt ggaaggggga cacgtgaata tggcattctc
 2941 tcagtccaac cctccttcgg agttgcacag ggtccacgga tccagaaaca agccaccag
 3001 cttgtggaac ccaacgtaag gctcctggtt tacagagaaa cccacaaaa agaataatcc
 3061 tatagcaaaag aaggagccac acgagagggg taacctggg ctggagggaa gctgtactgt
 3121 cccacctaac gttgcaactg ggagacttcc gggggcctca ctgctcctag agccctcttc
 60 3181 gctgactgcc aatatgaagg aggtacctct gttcaggcta cgtcacttcc cttgtgggaa

3241 tgtcaattac ggctaccagc aacagggcctt gcccttagaa gccgctactg cccttgagc
 3301 tggtcattac gaggatacca ttctgaaaag caagaatagc atgaaccagc ctgggccctg
 3361 agctcggtcg cacactcact tctcttcctt gggatcccta agaccgtgg

5 **EML4-ALK Variant 4 Protein Sequence (BAG75147.1;
 GI:209837704)**

1 mdgfgagsldd ssaastsdv qdrilsalesr vqqgedeitv lkaaladvlr rlaisedhva
 61 svkksvsskg qpspravipm scitngsgan rkpshtsavs iagketlssa aksgtekkke
 121 kpgggrekke eshsndqspq iraspspqpq sqplqihrtq pesknaptk sikrpspaek
 10 181 shnswensdd srnklskipk tpklipkvtk tadkhkdvii ngegeyikmf mrgrpitmfi
 241 psdvdnyddi rtelppelkl lewaygyrgk dcranvylp tgeivvyfias vvvlfnyeer
 301 tqrhylghtd cvkclaihpq kiriatgqia gvdkdgrplq phvrwvdsvt lstlqiiglg
 361 tfergvgcld fskadsgvhl cviddsnehm ltvwdwqrka kgaeiktne vvlavefhpt
 421 dantiitcgk shiffwtwsg nsltrkqgif gkyekpkfvq clafllngdv ltgdsqgvml
 15 481 iwskttept pgkpgkgyvq iskqikahdg svftlcqmrn gmltgggkd rkiilwdhdl
 541 npereieicw mspeyklskl rtstimtdyn pnycfagkts sisdlkevpr knitlirglg
 601 hgafgevyeg qvsgmpndps plqvavktp evcseqdeld flmealiisk fnhqniivrci
 661 gvslqslprf illelmaggd lksflretrp rpsqpslam ldlhvardi acgcqyleen
 721 hfihrdiaar nclltcpgpg rvakigdfgm ardiyrasyy rkggcamlpv kwmppeafme
 20 781 giftsktdtw sfgvllweif slgympypsk snqevlefvt sggrmdppkn cpgpvyrmt
 841 qcwqhqpdr pnfailleri eyctqdpdvi ntalpieygp lveeekvpv rpkdpegvvp
 901 llvsqqakre eerspaapp lpttssgkaa kkptaaevsv rvprgpaveg ghvnmfafsqs
 961 nppselhrvh gsrnkptslw nptygswfte kptkknnpia kkephergnl glegscvpp
 1021 nvatgrlpga sllelpslt anmkevplfr lrhfpcgnvn ygyqqqglpl eaatapgagh
 25 1081 yedtilkskn smnqpgp

**EML4-ALK Variant 5a Nucleic Acid Sequence (AB374364.1;
 GI:209837705)**

1 actctgtcgg tccgctgaat gaagtgcgcg cccctctaag cccggagccc ggcgctttcc
 30 61 ccgcaagatg gacggtttcg ccggcagctc cgtatgtagt atttctgctg caagtacttc
 121 tgatgttcaa gatcgctgtg cagctcttga gtcacagatt cagcaacaag aagatgaaat
 181 cactgtgcta aaggcggcct tggctgatgt tttgagcgt cttgcaatct ctgaagatca
 241 tgtggcctca gtgaaaaaat cagtctcaag taaagtgtac cgccggaagc accaggagct
 301 gcaagccatg cagatggagc tgcagagccc tgagtacaag ctgagcaagc tccgcacctc
 35 361 gaccatcatg accgactaca accccaacta ctgctttgct ggcaagacct cctccatcag
 421 tgacctgaag gaggtgcccg ggaaaaaaat caccctcatt cggggtcttg gccatggagc
 481 ctttggggag gtgtatgaag gccaggtgtc cggaatgccc aacgacccaa gccccctgca
 541 attggtgtgt aagacgctgc ctgaagtgtg ctctgaacag gacgaactgg atttctcat
 601 ggaagccctg atcatcagca aattcaacca ccagaacatt gtctgctgca ttggggtgag
 40 661 cctgcaatcc ctgccccggt tcatcctgct ggagctcatg gcggggggag acctcaagtc
 721 ctctctccga gagaccgcc ctcgcccag ccagccctcc tccctggcca tgctggacct
 781 tctgcacgtg gctcgggaca ttgcctgtgg ctgtcagtat ttggaggaaa accacttcat
 841 ccaccgagac attgctgcca gaaactgcct cttgacctgt ccaggccctg gaagagtggc
 901 caagattgga gacttcggga tggcccagga catctacagg gcgagctact atagaaaggg
 45 961 aggtgtgccc atgtgcccag ttaagtggat gccccagag gccttcatgg aaggaatatt
 1021 cacttctaaa acagacacat ggtcctttgg agtctgcta tgggaaatct tttctcttgg
 1081 atatatgcca taccacagca aaagcaacca ggaagttctg gagtttgtca ccagtggagg
 1141 ccgatgggac ccaccaaga actgcccctg gctgtatata cggataatga ctcagtgtctg
 1201 gcaacatcag cctgaagaca ggcccaactt tgccatcatt ttggagagga ttgaatactg
 50 1261 caccacaggac ccgatgtaa tcaacaccgc tttgccgata gaatatggtc cacttgtgga
 1321 agaggaaagag aaagtgcctg tgaggcccaa ggaccctgag ggggttctct ctctcctggt
 1381 ctctcaacag gcaaaacggg aggaggagcg cagcccagct gccccaccac ctctgcctac
 1441 cactctctct ggcaaggctg caaagaaacc cacagctgca gaggtctctg ttccagctcc
 1501 tagagggccg gccgtggaag ggggacacgt gaatatggca ttctctcagt ccaaccctcc
 55 1561 ttccggagttg cacagggtcc acggatccag aaacaagccc accagcttgt ggaaccaaac
 1621 gtacggctcc tggtttacag agaaaccac caaaaagaat aatcctatag caaagaagga
 1681 gccacacgag aggggtaacc tggggctgga gggaagctgt actgtccac ctacgttgcc
 1741 aactgggaga cttccggggg cctcactgct cctagagccc tcttcgctga ctgccaatat
 1801 gaaggaggta cctctgttca ggctacgtca cttcccttgt gggaatgtca attacggcta
 60 1861 ccagcaacag ggcttgccct tagaagccgc tactgcccct ggagctggtc attacgagga

1921 taccattctg aaaagcaaga atagcatgaa ccagcctggg ccctgagctc ggtcgcacac
1981 tcacttctct tccttgggat ccctaagacc gtgg

**EML4-ALK Variant 5a Protein Sequence (BAG75148.1;
GI:209837706)**

5 1 mdgfagsldd sisaastsdv qdrlsalesr vqqgedeitv lkaaladvlr rlaishedhva
61 svkksvsskv yrrkhqelqa mqmelqspey klslrtsti mtdynpnycf agktssisd1
121 kevprknitl irglghgafg evyegqvsgm pndpsplqva vktlpevcse qdeldflmea
181 liiskfnhqn ivrcigvslq slprfillel maggdlksfl retrprpsqp sslamlldllh
10 241 vardiacgcq yleenhfihf diaarncllt cpgrpgrvaki gdfgmardiy rasyyrkggc
301 amlpvkwmp eafmegifts ktdtwsfgvl lweifslgym pypsksnqev lefvtsggrm
361 dppkncpgpv yrimtqcwqh qpedrpnfai ilerieyctq dpdvintalp ieygplveee
421 ekvpvrpkdp egvppllvsq qakreersp aappplptts sgkaakkpta aevsvrvprg
481 pavegghvnm afsqsnppse lhrvhgsrnk ptslwnptyg swftekptkk nnpiakkeph
15 541 ergnlglegs ctvppnvatg rlpgasllle pssltanmke vplfrlrhfp cgnvnygyqq
601 qglpleaata pgaghyedti lksknsnmqp gp

**EML4-ALK Variant 5b Protein Sequence (AB374365.1;
GI:209837707)**

20 1 actctgtcgg tccgctgaat gaagtgcccg ccctctaaag cccggagccc ggcgctttcc
61 ccgcaagatg gacggtttcg ccggcagctc cgatgatagt atttctgctg caagtacttc
121 tgatgttcaa gatcgccgtg cagctcttga gtcacgagtt cagcaacaag aagatgaaat
181 cactgtgcta aaggcggctt tggctgatgt tttgagggct cttgcaatct ctgaagatca
241 tgtggcctca gtgaaaaaat cagctcaag taaaggttca gagctcaggg gaggatatgg
25 301 agatccaggg aggccttccg taggaagtgg cctgtgtagt gcttcaaggg ccaggctgcc
361 aggccatgtt gcagctgacc acccacctgc agtgtaccgc cggaagcacc aggagctgca
421 agccatgcag atggagctgc agagccctga gtacaagctg agcaagctcc gcacctcgac
481 catcatgacc gactacaacc ccaactactg ctttgcctgc aagacctcct ccatcagtga
541 cctgaaggag gtgcccgcga aaaacatcac cctcattcgg ggtctggggc atggagcctt
30 601 tggggaggtg tatgaaggcc aggtgtccgg aatgcccac gacccaagcc ccttgcaagt
661 ggctgtgaag acgctgcctg aagtgtgctc tgaacaggac gaactggatt tcctcatgga
721 agccctgac atcagcaaat tcaaccacca gaacattgtt cgctgcattg gggtagacct
781 gcaatccctg ccccggttca tcctgtgga gctcatggcg gggggagacc tcaagtcctt
841 cctccgagag acccgccctc gcccgagcca gccctcctcc ctggccatgc tggaccttct
35 901 gcacgtggtc cgggacattg cctgtggctg tcagtatttg gaggaaaacc actcatcca
961 ccgagacatt gctgccagaa actgcctcct gacctgtcca ggcctggaa gagtggccaa
1021 gattggagac ttcgggatgg ccgagacat ctacagggcg agctactata gaaagggagg
1081 ctgtgccatg ctgccagtta agtggatgcc ccagaggcc tcatggaa gaaattccac
1141 ttctaaaaca gacacatggt cctttggagt gctgctatgg gaaatctttt ctcttgata
40 1201 tatgccatac ccagcaaaa gcaaccagga agttctggag tttgtacca gtggaggccg
1261 gatggacca cccaagaact gccctgggcc tgtataccgg ataatgactc agtgctggca
1321 acatcagcct gaagacaggc ccaactttgc catcattttg gagaggattg aactactgcac
1381 ccaggacccg gatgtaatca acaccgcttt gccgatagaa tatgggtccac ttgtggaaga
1441 ggaagagaaa gtgcctgtga ggoccaaagga ccctgagggg gttcctcctc tcctggtctc
45 1501 tcaacaggca aaacgggagg aggagcgcag ccagctgcc ccaccacctc tgctaccac
1561 ctccctctggc aaggctgcaa agaaacccac agctgcagag gtctctgttc gagtccctag
1621 agggccggcc gtggaagggg gacacgtgaa tatggcattc tctcagtcca accctccttc
1681 ggagttgcac aggttccacg gatccagaaa caagcccacc agcttggtga acccaacgta
1741 cggctcctgg ttacagaga aaccaccaa aaagaataat cctatagcaa agaaggagcc
50 1801 acacgagagg ggtaacctgg ggctggaggg aagctgtact gtcccaccta acgttgcaac
1861 tgggagactt ccgggggacct cactgctcct agagccctct tcgctgactg ccaatatgaa
1921 ggaggtacct ctgttcaggc tacgtcactt cccttggtgg aatgtcaatt acggctacca
1981 gcaacagggc ttgcccttag aagccgctac tgccctgga gctggctatt acgaggatac
2041 cattctgaaa agcaagaata gcatgaacca gcctgggccc tgagctcggg cgcacactca
55 2101 cttctcttcc ttgggatccc taagaccgtg g

**EML4-ALK Variant 5b Protein Sequence (BAG75149.1;
GI:209837708)**

1 mdgfagsldd sisaastsdv qdrlsalesr vqqgedeitv lkaaladvlr rlaishedhva
60 61 svkksvsskg selrggygdp grlpvsglc sasrarlpgh vaadhppavy rrkhhqelqam

121 qmelqspeyk lsklrtstim tdympnycf gktssisdik evprknitli rglghgafge
 181 vyegqvsgmp ndpsplqvav ktlpevcseq deldflmeal iiskfnhqni vroigvslqs
 241 lprfillelm aggdlsflr etrprpsqps slamldllhv ardiacgcqy leenhfihrd
 301 iaarnclltc ppgprvakig dfgmardiyr asyyrkggca mlpvkwmpepe afmegiftsk
 5 361 tdtwsfgvll weifslgyp ypsksnqevl efvtsggrmd ppknpcgppy rmtqgcwqhq
 421 pedrpnfaii lerieyctqd pdvintalpi eygplveeee kvpvrpkdpe gvppllvsqq
 481 akreeerspa appplpttss gkaakkptaa evsvrvprgp avegghvnma fsqsnppsel
 541 hrvhgsrnkp tslwnptygs wftekptkkn npiakkephe rgnlglegsc tvppnvatgr
 601 lpgaslllep ssltanmkev plfrlrhfc gnvnygyqqq glpleaatap gaghyedtil
 10 661 ksknsnmqpg p

**EML4-ALK Variant 6 Nucleic Acid Sequence (AB462411.1;
 GI:227452648)**

1 tactctgtcg gtcgctgaa tgaagtgcc gccctctaa gcccgagcc cggcgcttcc
 15 61 cccgcaagat ggacggttcc gccggcagtc tcgatgatag tatttctgct gcaagtactt
 121 ctgatgttca agatcgccgt tcagctcttg agtcacgagt tcagcaacaa gaagtatgaa
 181 tcaactgtgt aaaggcggt ttggctgatg ttttgaggcg tcttgcaatc tctgaagatc
 241 atgtggcctc agtgaaaaaa tcagtctcaa gtaaggcca accaagccct cgagcagtta
 301 ttccccatgt ctgtataacc aatggaagt gtgcaaacag aaaaccaagt cataccagtg
 20 361 ctgtctcaat tgcaggaaaa gaaactcttt catctgctgc taaaagtgg acagaaaaaa
 421 agaaagaaaa accacaagga cagagagaaa aaaaagagga atctcattct aatgatcaaa
 481 gtccacaaat tcgagcatca ccttctcccc agccctcttc acaacctctc caaatacaca
 541 gacaaactcc agaaagcaag aatgctactc ccacaaaaag cataaaacga ccatcaccag
 601 ctgaaaagtc acataattct tgggaaaatt cagatgatag ccgtaataaa ttgtcgaaaa
 25 661 taccttcaac acccaaatat ataccaaaag ttaccaaaac tcagagacaag cataaagatg
 721 tcatcatcaa ccaagaagga gaatatatta aaatgtttat gcgcggctcg ccaattacca
 781 tgttcattcc ttccgatgtt gacaactatg atgacatcag aacggaactg cctcctgaga
 841 agctcaaaact ggagtgggca tatggttctc gaggaagga ctgtagagac aatgtttacc
 901 ttcttccgac cggggaata gtttatttca ttgcatcagt agtagacta tttattatg
 30 961 aggagagaa tcagcgacac tacctgggcc atacagactg tgtgaaatgc cttgctatac
 1021 atcctgacaa aattaggatt gcaactggac agatagctgg cgtggataaa gatggaaggc
 1081 ctctacaacc ccacgtcaga gtgtgggatt ctgttactct atccacactg cagattattg
 1141 gacttggcac ttttgagcgt ggagtaggat gcctggattt ttcaaaagca gattcagggtg
 1201 ttcatattat tgttattgat gactccaatg agcatatgct tactgtatgg gactggcaga
 35 1261 ggaaagcaaa aggagcagaa ataaagacaa caaatgaagt tgttttggct gtggagtttc
 1321 acccaacaga tgcaaatacc ataattacat gcggtaaatc tcataatttc tctggtacct
 1381 ggagcggcaa ttcactaaca agaaaacagg gaatttttgg gaaatatgaa aagccaaaat
 1441 ttgtgcagtg ttttagcattc ttggggaatg gagatgttct tactggagac tcagggtggag
 1501 tcatgcttat atggagcaaa actactgtag agccacacc tgggaaagga cctaaggaa
 40 1561 gtggcctgtg tagtgcttca agggccaggc tgccaggcca tgttgagct gaccacccac
 1621 ctgcagtgta ccgcccgaag caccaggagc tgcaagccat gcagatggag ctgcagagcc
 1681 ctgagtacaa gctgagcaag ctccgcacat cgaccatcat gaccgactac aaccccaact
 1741 actgctttgc tggcaagacc tctccatca gtgacctgaa ggaggtgccc ggaaaaaca
 1801 tcacctcat tcggggtctg ggccatggag cctttgggga ggtgtatgaa ggccagggtg
 45 1861 ccggaatgcc caacgacca agcccccctgc aagtggctgt gaagacgctg cctgaagtgt
 1921 gctctgaaca ggacgaactg gatttccctc tgggaagccct gatcatcagc aaattcaacc
 1981 accagaacat tgttcgctgc attggggtga gcctgcaatc cctgcccgg tcatcctgc
 2041 tggagctcat ggccggggga gacctcaagt ccttccctcc agagacccgc cctcgcccga
 2101 gccagccctc ctccctggcc atgctggacc ttctgcacgt ggctcgggac attgctgtg
 50 2161 gctgtcagta tttggaggaa aaccattca tccaccgaga cattgtctgc agaaactgcc
 2221 tcttgacctg tccaggccct ggaagagtgg ccaagattgg agacttcggg atggcccag
 2281 acatctacag ggcgagctac tatagaaagg gaggtgtgtc catgctgcca gttaaagtga
 2341 tgccccaga ggccttcatg gaaggaaat tcaacttctaa aacagacaca tggctctttg
 2401 gagtgtgct atgggaaatc ttttctcttg gatatatgcc ataccocagc aaagcaacc
 55 2461 aggaagtctt ggagtttgc accagtgagg gccggatgga cccacccaag aactgcccctg
 2521 ggctgtata ccggataatg actcagtgt ggcaacatca gcctgaagac aggcccaact
 2581 ttggcatcat tttggagagg attgaattc gcacccagga ccggatgta atcaacaccg
 2641 ctttgccgat agaataatgt ccacttgtgg aagaggaaga gaaagtgcct gtgaggccca
 2701 aggacctga gggggttcc cctctcctgg tctctcaaca ggcaaacgg gaggaggagc
 60 2761 gcagcccagc tgccccacca cctctgccta ccactctctc tggcaaggct gcaagaaac

2821 ccacagctgc agaggtctct gtctcgagtc ctagagggcc ggccgtggaa gggggacacg
 2881 tgaatatggc attctctcag tccaacctc cttcgaggtt gcacaggggtc cacggatcca
 2941 gaaacaagcc caccagcttg tggaacccaa cgtacggctc ctggtttaca gagaaaccca
 3001 ccaaaaagaa taatcctata gcaaagaagg agccacacga gaggggtaac ctggggctgg
 5 3061 aggggaagctg tactgtccca cctaacgttg caactgggag acttccgggg gcctcactgc
 3121 tcctagagcc ctcttcgctg actgccata tgaaggaggt acctctgttc aggtcactgc
 3181 acttcccttg tgggaatgtc aattacggct accagcaaca gggcttgccc ttagaagccg
 3241 ctactgccc tggagctggc cattacgagg ataccattct gaaaagcaag aatagcatga
 3301 accagcctgg gccctgagct cggtcgcaca ctcaattctc ttccttggga tcctaagac
 10 3361 cgtgg

**EML4-ALK Variant 6 Protein Sequence (BAH57335.1;
 GI:227452649)**

1 mdgfgalsldd sisaastsdv qdrilsalesr vqqgedeitv lkaaladvlr rlaisedhva
 15 61 svkksvsskg qpspravipm scitngsgan rkpshtsavs iagketlssa aksgtekkke
 121 kpggqrekke eshsndqspq iraspspqpq sqplqihrtt pesknaptk sikrpspaek
 181 shnswnsdd srnklskips tpklipkvtk tadkhkdvii ngegeyikmf mrgrpitmfi
 241 psdvdnyddi rtelppekik lewaygyrgk dcranvylp tgeivyfias vvvlfnyeer
 301 tqrhlyghtd cvkclaihpq kiriatgqia gvdkdgrplq phvrwvdsvt lstlqiiglg
 20 361 tfergvgcld fskadsgvhl cviddsnehm ltvwdwqrka kgaeiktne vvlavefhpt
 421 dantiitcgk shiffwtwsg nsltrkqgif gkyekpkfvq clafingdv ltgdsggvml
 481 iwskttept pgkgpksgsl csasrarlpq hvaadhppav yrrkhqelqa mqmelqspey
 541 klslrtsti mtdynpnycf agktssisl kevprknitl irglghgafg evyegqvsgm
 601 pndpsplqva vktlpevcse qdeldflmea liiskfnhqn ivrcigvslq slprfillel
 25 661 maggdlsfl retrprpsqp sslamldllh vardiacgcq yleenhfihr diaarncllt
 721 cpgpgrvaki gdfgmardiy rasyyrkggc amlpvkwmp eafmegifts ktdtwsfgvl
 781 lweifslgym pypsksnqev lefvtsggrm dppkncpgpv yrimtqcwqh qpdrpnfai
 841 ilerietctg dpdvintalp ieygplvee ekvpvrpkdp egvppllvsg qakreersp
 901 aappplptts sgkaakkpta aevsvrvprg pavegghvnm afsqsnppse lhrvhgsrnk
 30 961 ptslwnptyg swftekptk nnpiaakkeph ergnlglegs ctvppnvatg rlpgasllle
 1021 pssltanmke vplfrlrhfp cgnvnygyqq qglpleaata pgaghyedti lksknsnmqp
 1081 gp

**EML4-ALK Variant 7 Nucleic Acid Sequence (AB462412.1;
 GI:227452650)**

1 tactctgtcg gtccgctgaa tgaagtgcc gccctcttaa gcccgagacc cggcgctttc
 61 cccgcaagat ggacgggttc gccggcagtc tcgatgatag tatttctgct gcaagtactt
 121 ctgatgttca agatgcctg tcagctcttg agtcacgagt tcagcaacaa gaagatgaaa
 181 tcaactgtgt aaaggcggct ttggctgatg ttttgaggcg tcttgcaatc tctgaagatc
 40 241 atgtggcctc agtgaaaaaa tcagtctcaa gtaaaaggcca accaagccct cgagcagtta
 301 ttcccatgtc ctgtataacc aatggaagtg gtgcaaacag aaaaccaagt cataccagtg
 361 ctgtctcaat tgcaggaaaa gaaactcttt catctgctgc taaaagtggc acagaaaaaa
 421 agaaagaaaa accacaagga cagagagaaa aaaaagagga atctcattct aatgatcaaa
 481 gtccacaaat tcgagcatca ccttctcccc agccctcttc acaacctctc caaatacaca
 45 541 gacaaactcc agaaagcaag aatgctactc ccacaaaaag cataaaacga ccatcaccag
 601 ctgaaaagtc acataattct tgggaaaatt cagatgatag ccgtaataaa ttgtcgaaaa
 661 taccttcaac acccaaatta ataccaaaag ttaccaaaac tgcagacaag cataaagatg
 721 tcatcatcaa ccaagaagga gaatatatta aaatgtttat gcgcggctcg ccaattacca
 781 tgttcattcc ttccgatgtt gacaactatg atgacatcag aacggaactg cctcctgaga
 50 841 agctcaaaact ggagtgggca tatggtttatc gaggaaagga ctgtagagct aatgtttacc
 901 ttcttccgac cggggaaaata gtttatttca ttgcatcagt agtagtacta ttaattatg
 961 aggagagAAC tcagcgacac tacctgggcc atacagactg tgtgaaatgc cttgctatac
 1021 atcttgacaa aattaggatt gcaactggac agatagctgg cgtggataaa gatggaaaggc
 1081 ctctacaacc ccacgtcaga gtgtgggatt ctgttactct atccacactg cagattattg
 55 1141 gacttggcac ttttgagcgt ggagttaggt gcctggattt ttcaaaagca gattcagggtg
 1201 ttcatattat tgattattgat gactccaatg agcatatgct tactgtatgg gactggcaga
 1261 ggaaagcaaa aggagcagaa ataaagacaa caaatgaagt tgttttggct gtggagtttc
 1321 acccaacaga tgcaaatacc ataattacat gcggtaaatc tcatattttc ttctggacct
 1381 ggagcggcaa ttcactaaca agaaaacagg gaatttttgg gaaatatgaa aagccaaaat
 60 1441 ttgtgcagtg ttttagcattc ttggggaatg gagatgttct tactggagac tcagggtggag

1501 tcatgcttat atggagcaaa actactgtag agccacacacc tgggaaagga cctaaagggtg
 1561 tatatcaaat cagcaaaaca atcaaagctc atgatggcag tgtgttcaca ctttgtcaga
 1621 tgagaaatgg gatgttatta actggaggag ggaaagacag aaaaataatt ctgtgggac
 1681 atgatctgaa tcctgaaaaga gaaatagagc accaggagct gcaagccatg cagatggagc
 5 1741 tgcagagccc tgagtacaag ctgagcaagc tccgcacctc gaccatcatg accgactaca
 1801 accccaacta ctgctttgct ggcaagacct cctccatcag tgacctgaag gaggtgccgc
 1861 ggaaaaacat caccctcatt cgggggtctgg gccatggagc ctttggggag gtgtatgaag
 1921 gccaggtgtc cggaatgccc aacgacccaa gcccctgca agtggtgtg aagacgctgc
 1981 ctgaagtgtg ctctgaacag gacgaactgg atttctcat ggaagccctg atcatcagca
 10 2041 aattcaacca ccagaacatt gttcgtgca ttgggggtgag cctgcaatcc ctgccccggt
 2101 ctatctgct ggagctcatg ggggggggag acctcaagtc ctctccga gagaccgcc
 2161 ctgcccagc ccagccctcc tccctggcca tgctggacct tctgcacgtg ctcggggaca
 2221 ttgctgtgg ctgtcagtat ttggaggaaa accacttcat ccaccgagac attgctgcc
 2281 gaaactgcct cttgacctgt ccaggccctg gaagagtggc caagattgga gacttcggga
 15 2341 tggcccgaga catctacagg gogagctact atagaaagg aggctgtgcc atgtgccag
 2401 ttaagtggat gccccagag gccttcatgg aaggaatatt cacttctaaa acagacacat
 2461 ggtcctttgg agtgtgcta tgggaaatct tttctcttgg atatatgcca taccacagca
 2521 aaagcaacca ggaagtctg gagtttctca ccagtggag ccgatggac ccaccaaga
 2581 actgcccctg gcctgtatc cggataatga ctcaagtgtg gcaacatcag cctgaagaca
 20 2641 ggcccaactt tgccatcatt ttggagagga ttgaatactg caccaggac ccgatgtaa
 2701 tcaacaccgc tttgccgata gaatatggtc cacttgtgga agaggaagag aaagtgcctg
 2761 tgaggcccaa ggaccctgag ggggttcctc ctctcctggt ctctcaacag gcaaacggg
 2821 aggaggagcg cagcccagct gcccaccac ctctgcctac cactcctct ggcaaggctg
 2881 caaagaaacc cacagctgca gaggtctctg ttcgagtcct tagagggccg gccgtggaag
 25 2941 ggggacacgt gaatatggca ttctctcagt ccaacctcc ttcggagttg cacaaggctc
 3001 accgatccag aaacaagccc accagcttgt ggaacccaac gtacggctcc tggtttacag
 3061 agaaaccac caaaaagaat aatcctatag caaagaagga gccacacgac aggggtaacc
 3121 tggggctgga gggaagctgt actgtccac ctaacgttgc aactgggaga cttccggggg
 3181 cctcactgct cctagagccc tcttcgctga ctgccaatat gaaggaggta cctctgttca
 30 3241 ggctacgtca cttcccttgt gggaatgtca attacggcta ccagcaacag ggcttgcct
 3301 tagaagccgc tactgcccct ggagctggtc attacgagga taccattctg aaaagcaaga
 3361 atagcatgaa ccagcctggg cctgagctc ggtgcacac tcacttctct tccttgggat
 3421 ccctaagacc gtgga

35 **EML4-ALK Variant 7 Protein Sequence (BAH57336.1;
 GI:227452651)**

1 mdgfgalsldv ssaastsdv qdrleslsr vqqgedeitv lkaaladvlr rlaisedhva
 61 svkksvsskg qpspravipm scitngsgan rkpshtsavs iagketlssa aksgtekkke
 121 kpqqgrekke eshsndqspq iraspspqs sqplqihrt pesknaptk sikrspaek
 40 181 shnswensdd srnklskips tpklipkvtk tadkhkdvii ngegeyikmf mrgpmitmfi
 241 psdvdnyddi rtelppeklk lewaygyrgk dcranvylp tgeivyfias vvlfnyeer
 301 tqrhlyghd cvkclaihp kiriatgqia gvdkgdgrlp phrvwdsvt lstlqiiglg
 361 tfergvgld fskadsgvhl cviddsnehm ltvdwqrka kgaeiktne vvlavefhpt
 421 dantiitcgk shiffwtswg nsltrkqgif gkyekpkfvq clafngngdv ltgdsggvml
 45 481 iwskttept pgkgpkgyvq iskqikahdg svftlcqmrn gmltgggkd rkiilwdhdl
 541 npereiehqe lqamqmelqs peyklsklrt stimtdynpn ycfagktssi sdlkevprkn
 601 itlirglghg afgevyegqv sgmpndpspl qvavktlpev cseqdeldfl mealiskfn
 661 hqnivrcigv slqslprfil lelmaggdlk sflretrprp sqpsslamld llhvardiac
 721 gcqyleenhf ihrdiaarnc lltcpgpgrv akigdfgmar diyasyyrk ggcamlpvkw
 50 781 mppeafmegi ftsktdtwsf gvllweifsl gypypsksn qevlefvtsg grmdppkncp
 841 gpvyrimtgc wqhqpdrpn failerley ctqdpdvint alpieygplv eeeekvpvrp
 901 kdpegvppll vsqqakreee rsapaappplp ttssgkaakk ptaaevsrv prgpaveggh
 961 vnmafsqsnp pselkhvhs rnkptslwnp tygswftekp tkknnpiakk ephdrnlg
 1021 egscvtppnv atgrlpgasl llepssltan mkevplflrl hfpcgnvnyg yqqqglplea
 55 1081 atapgaghye dtilksknsm nqpgp

60 **KIF5B-ALK Nucleic Acid Sequence (AB462413.1; GI:227452652)**

1 tgcagaaaag atggcggacc tggccgagtg caacatcaaa gtgatgtgtc gcttcagacc
 61 tctcaacgag tctgaagtga accgcgcgca caagtacatc gccaagtttc agggagaaga
 121 caggtcgtg atcgcgctca agccttatgc atttgatcgg gtgttccagt caagcacatc

181 tcaagagcaa gtgtataatg actgtgcaaa gaagattggt aaagatgtac ttgaaggata
 241 taatggaaca atatttgc atggacaaac atcctctggg aagacacaca caatggaggg
 301 taaacttcat gatccagaag gcatgggaat tattccaaga atagtgaag atatttttaa
 361 ttatatattac tccatggatg aaaatttga atttcatatt aaggtttcat attttgaat
 5 421 atatttggat aagataaggg acctgttaga tgtttcaaag accaaccttt cagttcatga
 481 agacaaaaac cgagttccct atgtaaaggg gtgcacagag cgttttgtat gtagtccaga
 541 tgaagttatg gataccatag atgaaggaaa atccaacaga catgtagcag ttacaaatat
 601 gaatgaacat agctctagga gtcacagtat atttcttatt aatgtcaaac aagagaacac
 661 acaaacggaa caaaagctga gtggaaaact ttatctggtt gatttagctg gtagtgaaaa
 10 721 ggtagttaa actggagctg aagggtgctg gctggatgaa gctaaaaaca tcaacaagtc
 781 actttctgct cttggaaatg ttatttctgc ttggctgag gtagtagac atgttccata
 841 tccagatagt aaatgacaa gaatccttca agattcatta ggtggcaact gtagaacac
 901 tattgtaat tgcgtctctc catcatcata caatgagtc gaaacaaaat ctacactctt
 961 atttggccaa agggccaaaa caattaagaa cacagtttgt gtcaatgtgg agttaactgc
 15 1021 agaacagtgg aaaaagaagt atgaaaaaga aaaagaaaaa aataagatcc tgcggaacac
 1081 tattcagtgg cttgaaaatg agctcaacag atggcgtaat ggggagacgg tgcctattga
 1141 tgaacagttt gacaaagaga aagccaactt ggaagctttc acagtggata aagatattac
 1201 ctttaccat gataaaccag caaccgcaat tggagttata ggaaatttta ctgatgctga
 1261 aagaagaaag tgtgaagaag aaattgctaa attatacaaa cagcttgatg acaaggatga
 20 1321 agaaattaac cagcaaagtc aaactgtaga gaaactgaag acgcaaagt tggatcagga
 1381 ggagcttttg gcatctacca gaagggatca agacaatat caagctgagc tgaatgcct
 1441 tcaagcagaa aatgatgcct ctaaaagaag agtgaaagaa gttttacagg cctagaaga
 1501 acttgctgctc aattatgatc agaagtctca ggaagttgaa gacaaaacta aggaatatga
 1561 attgcttagt gatgaattga atcagaaatc ggcaacttta gcgagtatag atgctgagct
 25 1621 tcagaaactt aaggaaatga ccaaccacca gaaaaaacga gcagctgaga tgatggcatc
 1681 tttactaaaa gaccttgagc aaataggaat tgctgtggga aataatgatg taagcagcc
 1741 tgagggaact ggcattgatg atgaagagtt cactgttgca agactctaca ttagcaaaat
 1801 gaagtcagaa gtaaaaacca tgggtgaaacg ttgcaagcag ttagaaagca cacaaactga
 1861 gagcaacaaa aaaatggaag aaaatgaaaa ggagttagca gcatgtcagc ttcgtatctc
 30 1921 tcaacatgaa gccaaaatca agtcattgac tgaatacctt caaaatgtgg acaaaaagaa
 1981 aagacagttg gaggaatctg tcgatgcctt cagtgaagaa ctagtccagc ttcgagcaca
 2041 agagaaagtc catgaaatgg aaaaggagca cttaaataag gttcagactg caaatgaagt
 2101 taagcaagct gttgaacagc agatccagag ccatagagaa actcatcaaa taagcagcag
 2161 tagtttgaga gatgaagtag aagcaaaagc aaaacttatt actgatcttc aagacaaaaa
 35 2221 ccagaaaatg atgttagagc aggaacgtct aagagtagaa catgagaagt tgaagccac
 2281 agatcaggaa aagagcagaa aactacatga acttacggtt atgcaagata gacgagaaca
 2341 agcaagacaa gacttgaagg gtttggaaga gacagtggca aaagaacttc agactttaca
 2401 caacctgcgc aaactctttg ttcaggacct ggtacaaga gttaaaaaga gtgctgagat
 2461 tgattctgat gacaccggag gcagcgtgca tcagaagcaa aaaatctcct ttcttgaaaa
 40 2521 taatcttgaa cagctcacta aagtgcacaa acagttggta cgtgataatg cagatctccg
 2581 ctgtgaactt cctaagttgg aaaagcgaact tcgagctaca gctgagagag tgaagccttt
 2641 ggaatcagca ctgaaagaag ctaaaagaaaa tgcatctcgt gatcgcaaac gctatcagca
 2701 agaagtagat cgcataaagg aagcagtcag gtcaaagaat atggccagaa gagggcattc
 2761 tgcacagatt gtgtaccgcc ggaagcacca ggagctgcaa gccatgcaga tggagctgca
 45 2821 gagccctgag tacaagctga gcaagctccg cacctcgacc atcatgaccg actacaaccc
 2881 caactactgc tttgctggca agacctctc catcagtgac ctgaaggagg tgcgcgggaa
 2941 aaacatcacc ctcatcgagg gtctgggcca tggcgccctt ggggaggtgt atgaaggcca
 3001 ggtgtccgga atgcccacg acccaagccc cctgcaagt gctgtgaaga cgtgcctga
 3061 agtgtgctct gaacaggacg aaactggattt cctcatggaa gccctgatca tcagcaaat
 50 3121 caaccaccag aacattgttc gctgcattgg ggtgagcctg caatccctgc cccggttcat
 3181 cctgctggag ctcatggcgg ggggagacct caagtccttc ctccgagaga cccgccctcg
 3241 cccgagccag ccctcctccc tggccatgct ggacctctc cacgtggctc ggacattgc
 3301 ctgtggctgt cagtatttgg aggaaaaacca ctcatccac cgagacattg ctgccagaaa
 3361 ctgcctcttg acctgtccag gccctggaag agtggccaag attggagact tccggatggc
 55 3421 ccgagacatc tacaggcgca gctactatag aaaggagggc tgtgccatgc tgcagttaa
 3481 gtggatgcc ccagaggcct ccatggaagg aatattcact tctaaaacag acacatggtc
 3541 ctttggagtg ctgctatggg aaatcttttc tcttgatat atgccatacc ccagcaaaag
 3601 caaccaggaa gttctggagt ttgtaccag tggaggccgg atggaccac ccaagaactg
 3661 ccctgggcct gtataccgga taatgactca gtgctggcaa catcagcctg aagacaggcc
 60 3721 caactttgcc atcattttg agaggattga atactgcacc caggaccgg atgtaataca
 3781 caccgctttg ccgatagaat atggtccact tgtggaagag gaagagaaag tgcctgtgag

3841 gcccaaggac cctgaggggg ttctctctct cctgggtctct caacaggcaa aacgggagga
 3901 ggagcgcagc ccagctgccc caaccacctct gcctaccacc tcctctggca aggtgcaaaa
 3961 gaaaccacaca gctgcagagg tctctgttcg agtccctaga gggccggccg tggaaggggg
 4021 acacgtgaat atggcattct ctcagtcctaa cctccttcg gagttgcaca aggtccacgg
 5 4081 atccagaaac aagcccacca gcttgtggaa cccaacgtac ggctcctggg ttacagagaa
 4141 acccaccaaaa aagaataatc ctatagcaaa gaaggagcca cacgacaggg gtaacctggg
 4201 gctggagggg agctgtactg tcccacctaa cgttgcaact gggagacttc cgggggcctc
 4261 actgctccta gagccctctt cgctgactgc caatatgaag gaggtacctc tgttcaggct
 4321 acgtcacttc ccttgtggga atgtcaatta cggctaccag caacagggct tgccttaga
 10 4381 agccgctact gccctggag ctggtcatta cgaggatacc attctgaaaa gcaagaatag
 4441 catgaaccag cctgggccct gagctcggtc gcacactca

KIF5B-ALK Protein Sequence (BAH57337.1; GI:227452653)

1 madlaecnik vmcrfrplne sevnrgdkyi akfqqedtvv iaskpyafdr vfgsstsqeq
 15 61 vyndcakkiv kdvlegyngt ifaygqtssg kthtmegklh dpegmgii pr ivqdifnyiy
 121 smdenlefhi kvsyfeiyld kirdlldvsk tnlsvhedkn rvpvkgcte rfvcspdevm
 181 dtidegksnr hvavtnmneh ssrshsifli nvkqentqte qklsgklylv dlagsekvsk
 241 tgaegavlde akninkslsa lgnvisalae gstyvpyrds kmtrilqds l ggnrcrttivi
 301 ccspssynes etkstllfgg raktikntvc vnveltaeqw kkkyekekek nkilrntiqw
 20 361 lenelnrwrn getvpideqf dkekanleaf tvdkditltn dkpataigvi gnftdaerrk
 421 ceeeiaklyk qlddkdeein qqsqlevklk tqmldqeell astrrdqdnm qaelnrlqae
 481 ndaskeevke vlqaleelav nydqksqeve dktkeyells delnqksatl asidaelqkl
 541 kemtnhqkkr aaemmasllk dlaeigiavg nndvkqpegt gmideeftva rlyiskmkse
 601 vktmvkrckq lestqtesnk kmeenekela acqlrisqhe akikslteyl qnveqkkrql
 25 661 eesvdalsee lvqlraqekv hemekehlnc vqtanevkqa veqqiqshre thqkqisslr
 721 deveakakli tdlqdqngkm mlegelr lve heklkatdqe ksrklhelte mqrrearg
 781 dlkgleetva kelqtlhnlr klfvqdlatr vkksaeidsd dtggsaaqkq kisflennle
 841 qltkvhkqlv rdnadlrcl pklekrlrat aervkalesa lkeakenasr drkryqqevd
 901 rikeavrskn marrghsaqi vyrrkhqelq amqmelqspe yklsklrtst imtdynpnyc
 30 961 fagktssisd lkevprknit lirlglghgaf gevyeqqvsg mpndpsplqv avktlpevcs
 1021 eqdeldfme aliiskfnhq nivrcigvsl qslprfille lmaggdlksf lretrprpsq
 1081 psslamldll hvardiacgc qyleenhfih rdiaarncll tcpqpgrvak igdfgmardi
 1141 yrasyyrkkg camlpvkwmp peafmegift sktdtwsfgv llweifslgy mpypsksnqe
 1201 vlefvtsggr mdppkncpgp vyrimtcwq hqpedrpnfa ilerleyct qdpdaivntal
 35 1261 pieygp lvee eekvpvrpkd pegvppllv qgakreeers paappplptt ssgkaakkpt
 1321 aaevsvrvpr gpavegghvn mafsqsnpps elhkvhgern kptslwnpty gswftekptk
 1381 knnpiakkep hdrnlgleg sctvppnvat grlpgaslll epssltanmk evplfrlrhf
 1441 pcgnvnygyq qqglpleaat apgaghyedt ilksknsnmq pgp

40 **NPM-ALK Sequence (t(2;5) (p23;q35 chromosomal translocation) ***

TPM3-ALK Sequence (t(1;2) (p25;p23) chromosomal translocation) *

45 **TFGXL-ALK Nucleic Acid Sequence (AF390893.1; GI:20269389)**
 1 atgaacggac agttggatct aagtgggaag ctaatcatca aagctcaact tggggaggat
 61 attcggcgaa ttctatttca taatgaagat attacttatg atgaattagt gctaattgatg
 121 caacgagttt tcagaggaaa acttctgagt aatgatgaag taacaataaa gtataaagat
 50 181 gaagatggag atcttataac aatttttgat agttctgacc ttctctttgc aattcagtgc
 241 agtaggatac tgaaactgac attatttggt aatggccagc caagaccctc tgaatcaagt
 301 caggtgaaat atctccgtcg agaactgata gaacttcgaa ataaagtga tgcgttattg
 361 gatagcttgg aaccacctgg agaaccagga ccttcacca atattcctga aaatgatact
 421 gtggatggta gggaagaaaa gtctgcttct gattcttctg gaaaacagtc tactcaggtt
 55 481 atggcagcaa gtatgtctgc ttttgatcct ttaaaaaacc aagatgaaat caataaaaaat
 541 gttatgtcag cgtttggctt aacagatgat caggtttcag ggccaccag tgctcctgca
 601 gaagatcgtt caggaacacc cgacagcatt gcttctcct cctcagcagc tcaccacca
 661 ggcgttcagc cacagcagcc accatataca ggagctcaga ctcaagcagg tcagattgaa
 721 gtgtaccgcc ggaagcacca ggagctgcaa gccatgcaga tggagctgca gagcctgag
 60 781 tacaagctga gcaagctccg cacctcgacc atcatgaccg actacaaccc caactactgc

841 tttgctggca agacctcctc catcagtgac ctgaaggagg tgccggggaa aaacatcacc
 901 ctcattcggg gtctggggcca tggcgccctt ggggaggtgt atgaaggcca ggtgtccgga
 961 atgcccacg acccaagccc cctgcaagtg gctgtgaaga cgctgcctga agtgtgctct
 1021 gaacaggacg aactggattt cctcatggaa gccctgatca tcagcaaat caaccaccag
 5 1081 aacattgttc gctgcattgg ggtgagcctg caatccctgc cccggttcat cctgctggag
 1141 ctcatggcgg ggggagacct caagtccttc ctccgagaga cccgccctcg cccgagccag
 1201 cctcctccc tggccatgct ggaccttctg cacgtggctc gggacattgc ctgtggctgt
 1261 cagtatttgg aggaaaacca cttcatccac cgagacattg ctgccagaaa ctgcctcttg
 1321 acctgtccag gccctggaag agtggccaag attggagact tcgggatggc ccgagacatc
 10 1381 tacagggcga gctactatag aaagggaggc tgtgccatgc tgccagttaa gtgatgccc
 1441 ccagaggcct tcattggaag aatattcact tctaaaacag acacatggtc ctttggagtg
 1501 ctgctatggg aaatcttttc tcttgatat atgccatacc ccagcaaaag caaccaggaa
 1561 gttctggagt ttgtcaccag tggaggccgg atggaccac ccaagaactg cctgggcct
 1621 gtataccgga taatgactca gtgctggcaa catcagcctg aagacaggcc caactttgcc
 15 1681 atcatttttg agaggattga atactgcacc caggaccgg atgtaataca caccgctttg
 1741 ccgatagaat atggtccact tgtggaagag gaagagaaaag tgcctgtgag gcccaaggac
 1801 cctgaggggg ttctctctct cctggtctct caacaggcaa aacgggagga ggagcgcagc
 1861 ccagctgccc caccacctct gctaccacg tctctggca aggtgcaaa gaaaccaca
 1921 cctgcaagg tctctgttcg agtccctaga gggccggccg tggaggggg acacgtgaat
 20 1981 atggcattct ctcagtccaa cctcctctcg gagttgcaca aggtccacgg atccagaaac
 2041 aagcccacca gcttgtggaa cccaacgtac ggctcctgg ttacagagaa acccaccaaa
 2101 aagaataatc ctatagcaaa gaaggagcca cagcagagg gtaacctgg gctggaggga
 2161 agctgtactg tcccaccta cgttgcaact gggagacttc cgggggcctc actgctccta
 2221 gagccctctt cgctgactgc caatatgaag gaggtacctc tgttcaggct acgtcacttc
 25 2281 ccttgtggga atgtcaatta cggctaccag caacagggtc tgcccttaga agccgctact
 2341 gccctggag ctggtcatta cgaggatacc attctgaaa gcaagaatag catgaaccag
 2401 cctgggcct ga

TFGXL-ALK Protein Sequence (AAM17922.1; GI:20269390)*

30 1 mngqldlsgk liikaqlged irripihned itydelvlmm qrvfrgklls ndevtikykd
 61 edgdlitifd ssdlsfaiqc srilkltlfv ngqprpless qvkyllrreli elrnkvnrl1
 121 dsleppgepg pstnipendt vdgreeksas dssgkqstqv maasmsafdp lknqdeinkn
 181 vmsafgltd qvsgppsapa edrsgtpdsi assssaahpp gvqpqqppyt gaqtgagqie
 241 vyrrkhqelq amqmelqspe yklsklrlst imtdynpnyc fagktssisd lkevrprknt
 35 301 lirlglhgaf gevyegqvsq mpndpsplqv avktlpevcs eqdeldflme aliiskfnhq
 361 nivrcigvsl qslprfille lmaggdlskf lretrprpsq psslamldll hvardiacgc
 421 qyleenhfih rdiaarncll tpcpgprvak igdfgmardi yrasyyrkkg camlpvkwnmp
 481 peafmegift sktdtwsfgv llweifslgy mpypsksnqe vlefvtsggr mdpkncpgp
 541 vyrintqcwq hqpdrpnfa iilerieyct qdpdvintal pieygplvee eekvpvrpkd
 40 601 pegvpllvsv qqakreeers paappplptt ssgkaakkpt aaevsvrvpr gpavegghvn
 661 mafsqsnpps elhkvhgsrn kptslwnpty gswftekptk knnpiakkep hdrnlgleg
 721 sctvppnvat grlpgaslll epssltanmk evplflrlrh pcgnvnygyq qqglpleaat
 781 apgaghyedt ilksksnmnq pgp

TFGL-ALK Nucleic Acid Sequence (AF143407.1; GI:6739534)

45 1 cctccgcaag ccgtctttct ctgagattgt atatataaga catcctggag tccaccatga
 61 acggacagtt ggatctaagt gggaagctaa tcataaagc tcaacttggg gaggatattc
 121 ggcaatttcc tattcataat gaagatatta cttatgatga attagtgcta atgatgcaac
 181 gagttttcag aggaaaactt ctgagtaatg atgaagtaac aataaagtat aaagatgaag
 50 241 atggagatct tataacaatt tttgatagtt ctgacctttc ctttgcaatt cagtgcagta
 301 ggatactgaa actgacatta tttgttaatg gccagccaag accccttgaa tcaagtcagg
 361 tgaaatatct ccgtcgagaa ctgatagaac ttcgaaataa agtgaatcgt ttattggata
 421 gcttgaacc acctggagaa ccaggacctt ccaccaatat tcttgaaaat gatactgtgg
 481 atggtaggga agaaaagtct gcttctgatt cttctggaaa acagtctact caggttatgg
 55 541 cagcaagtat gtctgctttt gatcctttta aaaaccaaga tgaaatcaat aaaaatgtta
 601 tgtcagcgtt tggcttaaca gatgatcagg tttcagtgtg ccgcccgaag caccaggagc
 661 tgcaagccat gcagatggag ctgcagagcc ctgagtacaa gctgagcaag ctcgcacct
 721 cgaccatcat gaccgactac aaocccaact actgctttgc tggcaagacc tctccatca
 781 gtgacctgaa ggaggtgccg cggaaaaaca tcacctcat tcggggtctg ggccatggcg
 60 841 cctttgggga ggtgtatgaa ggccaggtgt ccggaatgcc caacgacca agccccctgc

901 aagtggctgt gaagacgctg cctgaagtgt gctctgaaca ggacgaactg gatttccctca
 961 tggaagccct gatcatcagc aaattcaacc accagaacat tgttcgctgc attgggggtga
 1021 gacctgaate cctgccccgg ttcatcctgc tggagctcat ggccgggggga gacctcaagt
 1081 ccttcctccg agagaccgc cctgccccga gccagccctc ctccctggcc atgctggacc
 5 1141 ttctgcacgt ggctcgggac attgcctgtg gctgtcagta tttggaggaa aaccacttca
 1201 tccaccgaga cattgctgcc agaaactgcc tcttgacctg tccaggccct ggaagagtgg
 1261 ccaagattgg agacttcggg atggcccgag acatctacag ggcgagctac tatagaaagg
 1321 gaggtctgtc catgctgcca gttaagtggg tgcgcccaaga ggccttcctg gaaggaatat
 1381 tcacttctaa aacagacaca tggctcctttg gagtgcctgt atgggaaatc ttttctcttg
 10 1441 gatatatgcc ataccccagc aaaagcaacc aggaagtctt ggagtttgtc accagtggag
 1501 gccggatgga cccacccaag cctgctgtata ccggataatg cctcagtgtc
 1561 ggcaacatca gcctgaagac aggcccaact ttgccatcat tttggaggag attgaatact
 1621 gcaccacagga cccggatgta atcaacaccg ctttgccgat agaatatggg ccacttgttg
 1681 aagaggaaga gaaagtgcct gtgaggccca aggaccctga ggggggttct cctctcctgg
 15 1741 tctctcaaca ggcaaacgg gaggaggagc gcagccagc tgcgccacca cctctgccta
 1801 ccacctctc tggcaaggct gcaaagaaac ccacagctgc agaggctctc gttcgagtcc
 1861 ctgaggggcc ggccgtggaa gggggacacg tgaatatggc attctctcag tccaacctc
 1921 cttcgaggtt gcacaaggc caccgatcca gaaacaagcc caccagcttg tggaaccaca
 1981 cgtacggctc ctggtttaca gagaaccoca ccaaaaagaa taatcctata gcaagaagg
 20 2041 agccacacga cagggttaac ctggggctgg agggaagctg tactgtccca cctaacgttg
 2101 caactgggag acttccgggg gctcactgc tcttagagcc ctcttcgctg actgccata
 2161 tgaaggaggt acctctgttc aggctacgtc acttcccttg tgggaatgtc aattacggct
 2221 accagcaaca gggcttgccc ttagaagccg ctactgcccc tggagctggg cattacgagg
 2281 ataccattct gaaaagcaag aatagcatga accagcctgg gccctgagct cggctgcaca
 25 2341 ctacttctc ttcttgga tccctaagac cgtggaggag agagaggcaa tggctccttc
 2401 acaaaccaga gaccaaattg cactgtttgt tttgtgcca cctattttga agtaccacca
 2461 aaaaagctgt attttgaaaa tgcttttaga aggttttgag catgggttca tctattctt
 2521 tcgaaagaag aaaatatcat aaaaatgagt gataaataca aggccagat gtggttgcac
 2581 aagggtttta tgcattgttg ttgtatactt ccttatgctt cttttaaat gtgtgtgctc
 30 2641 tgcttcaatg tagtcagaat tagctgcttc tatgtttcat agttgggggc atagatgttt
 2701 ccttgccctg ttgatgtgga catgagccat ttgaggggag agggaacgga aataaaggag
 2761 ttatttgtaa tgactaaaa

TFGL-ALK Protein Sequence (AAF27292.1; GI:6739535)*

35 1 mngqldlsgk liikaqlged irripihned itydelvlmm qrvfrgklls ndevtikyk
 61 edgdltitfd ssdlsfaiqc srilkltlfv ngqprpless qvkylrreli elrnkvnrl
 121 dsleppgepg pstnipendt vdgreeksas dssgkqstqv maasmsafdp lknqdeinkn
 181 vmsafgltd qsvyrrkhq elqamqmelq speyklsklr tstimtdynp nycfagktss
 241 isdlkevprk nitlirglgh gafgevyegq vsgmpndpsp lqvavktlpe vcseqdeldf
 40 301 lmealiiskf nhqnivrcig vslqslprfi llelmaggdl ksflretrpr psqpsslaml
 361 dllhvardia cgcqyleenh fihrdiaarn cltccpgpgr vakigdfgma rdiyrasyr
 421 kggcamlpvk wmppeafmeg iftsktdtws fgvlleifis lgympypsk nqevlefvt
 481 ggrmdppknc pgpvyrimtq cwqhqpdrp nfaiilerie yctqdpdvin talpieygp
 541 veeekvpvr pkdpegvppl lvsqqakree erspaapppl pttssgkaak kptaaevsvr
 45 601 vprgpavegg hvnmfafsqs ppselkhvhg srnkptslwn ptygswftek ptkknnpiak
 661 kephdrgnlg legscvppn vatgrlpgas lilepsslta nmkevplflr rhfpcgnvny
 721 gyqqqglple aatapgaghy edtilkskns mnqpgp

TFGS-ALK Nucleic Acid Sequence (AF125093.1; GI:7229260)

50 1 cctccgcaag ccgtctttct ctgagttgt atatatagaa catcctggag tccaccatga
 61 acggacagtt ggatctaagt gggaagctaa tcatcaaagc tcaactggg gaggatattc
 121 ggcaattcc tattcataat gaagatatta cttatgatga attagtgtc atgatgcaac
 181 gagttttcag aggaaaactt ctgagtaatg atgaagtaac aataaagtat aaagatgaag
 241 atggagatct tataacaatt tttgatagtt ctgacctttc ctttgcaatt cagtgcagta
 55 301 ggatactgaa actgacatta tttgttaatg gccagccaag accccttgaa tcaagtcagg
 361 tgaaatatct ccgtcgagaa ctgatagaac ttcgaaataa agtgaatcgt ttattggata
 421 gcttggaaac acctggagaa ccaggacctt coaccaatat tctgaaat gtgtaccgcc
 481 ggaagcaca ggagctgcaa gccatgcaga tggagctgca gagccctgag tacaagctga
 541 gcaagctccg cacctcgacc atcatgaccg actacaaccc caactactgc tttgtggca
 60 601 agacctctc catcagtgc ctgaaggagg tgccgcggaa aaacatcacc ctcatcggg

661 gtctgggcca tggcgccctt ggggaggtgt atgaaggcca ggtgtccgga atgccccaacg
 721 acccaagccc cctgcaagt gctgtgaaga cgctgcctga agtgtgtctt gaacaggagc
 781 aactggattt cctcatggaa gccctgatca tcagcaaat caaccaccag aacattgttc
 841 gctgcatttg ggtgagcctg caatccctgc ccgggttcat cctgctggag ctcatggcgg
 5 901 ggggagacct caagtccttc ctccgagaga ccgcccctcg ccgagccag cctcctccc
 961 tggccatgct ggaccttctg cagctggctc gggacattgc ctgtggctgt cagtatttgg
 1021 aggaaaacca cttcatccac cgagacattg ctgccagaaa ctgcctcttg acctgtccag
 1081 gccctggaag agtggccaag attggagact tggggatggc ccgagacatc tacagggcga
 1141 gctactatag aaagggaggc tgtgccatgc tgccagttaa gtggatgccc ccagaggcct
 10 1201 tcatggaagg aatattcact tctaaaacag acacatggtc ctttggagtgt ctgctatggg
 1261 aaatcttttc tcttgatat atgccatacc ccagcaaaag caaccaggaa gttctggagt
 1321 ttgtcaccag tggaggcccg atggaccac ccaagaactg cctgggccc gtataccgga
 1381 taatgactca gtgctggcaa catcagcctg aagacaggcc caactttgcc atcattttgg
 1441 agaggattga atactgcacc caggaccgg atgtaatcaa caccgctttg ccgatagaat
 15 1501 atggtccact tgtggaagag gaagagaaag tgcctgtgag gcccaggac cctgagggg
 1561 ttctctctct cctggtctct caacaggcaa aacgggagga ggagcgcagc ccagctgcc
 1621 caccacctct gcctaccacc tctctggca aggctgcaaa gaaaccaca gctgcagagg
 1681 tctctgttcg agtcctaga gggccggccg tggaggggg acacgtgaat atggcattct
 1741 ctcatgcaaa cctccttcg gagggtgaca aggtccacg atccagaaac aagcccacca
 20 1801 gcttgtggaa cccaacgtac ggtcctggt ttacagagaa acccaaaaaa aagaataatc
 1861 ctatagcaaa gaaggagcca cagcagagg gtaacctggg gctggaggga agctgtactg
 1921 tcccaccta cgttgcaact gggagacttc cgggggcctc actgctccta gagcctctt
 1981 cgctgactgc caatatgaag gaggtacctc tgttcaggct acgtcacttc cttgtggga
 2041 atgtcaatta cggctaccag caacagggct tgcccttaga agccgctact gccctgggag
 25 2101 ctggtcatta cgaggatacc attctgaaa gcaagaatag catgaaccag cctgggccc
 2161 gagctcggtc gcacactcac ttctcttct tgggatccct aagaccgtg aggagagaga
 2221 ggcaatggct ccttcacaaa ccagagacca aatgtcacgt tttgtttgt gccaacctat
 2281 tttgaagtac caccaaaaaa gctgtatttt gaaaatgctt tagaaagggt ttgagcatgg
 2341 gttcatccta ttctttcgaa agaagaaaat atcataaaaa tgagtataa atacaaggcc
 30 2401 cagatgtgtg tgcataggt ttttatgcat gtttgttgta tacttcttta tgctctttt
 2461 aaattgtgtg tgctctgctt caatgtagtc agaattagct gttctatgt tcatagttg
 2521 gggcataga tgtttcctt ccttgttgat gtggacatga gccatttgag gggagagga
 2581 acggaataa aggagttatt tgtaatgact aaaa

35 **TFGS-ALK Protein Sequence (AAF42734.1; GI:7229261)***

1 mngqldlsgk liikaqlged irripihned itydelvlmm qrvfrgklls ndevtikykd
 61 edgdltitfd ssdlsfaiqc srilkltlfv ngqprpless qvkylrreli elrnkvnrl1
 121 dsleppgepg pstnipenvy rrrkhqelqam qmelqspeyk lsklrtstim tdynpnycfa
 181 gktssisdik evprknitli rglghgafge vyegqvsgmp ndpsplqvav ktlpevcseq
 40 241 deldflmeal iiskfnhqn1 vrcigvslqs lprfillelm aggdlsf1r etrprpsqps
 301 slamldllhv ardiacgcqy leenhfihrd iaarnclltc ppgprvakig dfgmardiyr
 361 asyyrkggca mlpvkwmppc afmegiftsk tdtwsfgvll weifslgyp ypsksnqevl
 421 efvtsggrmd ppknpcgpyv rimtqcwqhq pedrpnfaii lerieyctqd pdvintalpi
 481 eygplveeee kvpvrpkdpe gvppllvsqq akreeerspa appplpttss gkaakkptaa
 45 541 evsvrvprgp avegghvnma fsqsnppsel hkvhgsrcnp tslwnptygs wftekptkkn
 601 npiakkephd rgnlglegsc tvppnvatgr lpgaslllep ssiltanmkev plfrlrfhpc
 661 gnvnygyqqq glpleaatap gaghyedtil ksknsmnqpg p

50 **ATIC-ALK Sequence (inv(2)(p23;q35) chromosomal translocation)***

CLTC-ALK Sequence (t(2;17)(p23;q23) chromosomal translocation)*

55 **MSN-ALK Nucleic Acid Sequence (AF295356.1; GI:14625823)**

1 aactccgctg cctttgccgc caccatgccc aaaacgatca gtgtgcgtgt gaccaccatg
 61 gatgcagagc tggagtttgc catccagccc aacaccaccg ggaagcagct atttgaccag
 121 gtggtgaaaa ctattggctt gaggaagtgt tggttctttg gtctgcagta ccaggacact
 181 aaaggtttct ccacctggct gaaactaat aagaagggtga ctgccccagga tgtcggaag
 241 gaaagccccc tgctctttaa gttcctgtgc aagttctacc ctgaggatgt gtccgaggaa
 60 301 ttgattcagg acatcactca gcgcctgttc tttctgcaag tgaagagggt cattctcaat

361 gatgatattt actgcccgcc tgagaccgct gtgctgctgg cctcgtatgc tgtccagtct
 421 aagtatggcg acttcaataa ggaagtgcac aagtctggct acctggccgg agacaagttg
 481 ctcccgcaga gaggctctga acagcacaaa ctcaacaagg accagtggga ggagcggatc
 541 caggtgtggc atgaggaaca ccgtggcatg ctcaaggagg atgctgtcct ggaatatctg
 5 601 aagattgctc aagatctgga gatgtatggt gtgaactact tcagcatcaa gaacaagaaa
 661 ggctcagagc tgtggctggg ggtggatgcc ctgggtctca acatctatga gcagaatgac
 721 agactaactc ccaagatagg cttcccctgg agtgaaatca ggaacatctc tttcaatgat
 781 aagaaatttg tcatcaagcc cattgacaaa aaagccccgg acttcgtctt ctatgctccc
 841 cggctgcgga ttaacaagcg gatcttggcc ttgtgcatgg ggaacatga actatacatg
 10 901 cgccgtcgca agcctgatac cattgagggt cagcagatga aggcacaggg ccgggaggag
 961 aagcaccaga agcagatgga gcgtgctatg ctggaaaatg agaagaagaa gcgtgaaatg
 1021 gcagagaagg agaaagagaa gattgaacgg gagaaggagg agctgatgga gaggctgaag
 1081 cagatcgagg aacagactaa gaaggctcag caagaactgg aagaacagac ccgtagggct
 1141 ctggaacttg agcaggaacg gaagcgtgcc cagagcgagg ctgaaaagct ggccaaggag
 15 1201 cgtcaagaag ctgaagaggc caaggaggcc ttgctgcagg cctcccggga ccagaaaaag
 1261 actcaggaac agctggcctt ggaaatggca gagctgacag ctogaatctc ccagctggag
 1321 atggcccgac agaagaagga gagtgagggt gtggagtggc agcagaagca ggagctgcaa
 1381 gccatgcaga tggagctgca gacccctgag tacaagctga gcaagctccg cacctcgacc
 1441 atcatgaccg actacaacc ccaactactgc tttgctggca agacctctc catcagtgac
 20 1501 ctgaaggagg tgccgcggaa aaacatcacc ctcatcggg gtctgggcca tggcgccctt
 1561 ggggaggtgt atgaaggcca ggtgtccgga atgcccacg acccaag

MSN-ALK Protein Sequence (AAK71522.1; GI:14625824)*

1 mpktisvrvt tmdaelefai qpnttgkqlf dqvvtiglr evwffglqyy dtkgfstwlk
 25 61 lnkkvtaqdv rkespllfkf rakfypedvs eeliqditqr lfflqvkegi lnddiycppe
 121 tavllasyav qskygdfnke vkhsgylagd kllpqrivleq hklndqwee riqvweehhr
 181 gmlredavle ylkiaqdlem ygvnyfsikn kkgsehlwlv dalglniyeq ndrlltpkigf
 241 pwseirnisf ndkkfvikpi dkkapdfvfy aprlrinkri lalcgmgnhel ymrrrkpdti
 301 evqqmkaqar eekhqkqmer amlenekkk emaeekeeki erekeelmer lkqieeqtkk
 30 361 aqqeleeqtr raleleqerk raqseaekla kerqaeek eallqasrdq kktqeqlale
 421 maeltarisq lemarqkkes eavewqqkqe lqamqmelqs peyklsklrt stimtdynpn
 481 ycfagktssi sdlkevprkn itlirglghg afgevyegqv sgmpndp

TPM4-ALK Minor Variant Nucleic Acid Sequence (AF362887.1; GI:14010353)

1 cgagaagttg agggagaaa gggggccgg gaacaggctg aggctgaggt ggcctccttg
 61 aaccgtagga tccagctggt tgaagaagag ctggaccgtg ctcaaggagc tgcggaagtg
 121 tctgaactaa aatgtggtga cctggaagaa gaactcaaga atgttactaa caatctgaaa
 181 tctctggagg ctgcatctga aaagtattct gaaaaggagg acaaatatga agaagaaatt
 40 241 aaacttctgt ctgacaaact gaaagaggct gagaccgtg ctgaatttgc agagagaacg
 301 gttgcaaaac tggaaaagac aattgatgac ctggaagtgt acctccggaa gcaccaagag
 361 ctgcaagcca tgcagatgga gctgcagagc cctgagtaca agctgagcaa gctccgcacc
 421 ctgcac

TPM4-ALK Minor Variant Protein Sequence (AAK51964.1; GI:14010354)

1 revegerrar eqaeaevasl nrriqlveee ldraqeraev selkcgdlle elknvtnnlk
 61 sleaasekys ekedkyeeei kllsdlkkea etraefaert vaklektidd levylrkhqe
 121 lqamqmelqs peyklsklrt ld

TPM4-ALK Major Variant Nucleic Acid Sequence (AF362886.1; GI:14010351)

1 ctggcagagt cccgttgccg agagatggat gacgagatta gactgatgga ccagaacctg
 61 aagtgtctga gtgctgctga agaaaagtac tctcaaaaag aagataaata tgaggaaagaa
 121 atcaagattc ttactgataa actcaaggag gcagagaccc gtgctgaatt tgcaagagaga
 181 acggttgcaa aactggaaaa gacaattgat gacctggaag tgtaccgccg gaagcaccag
 241 gagctgcaag ccattgcagat ggagctgcag agccctgagt acaagctgag caagctccgc
 301 acctcgac

**TPM4-ALK Major Variant Protein Sequence (AAK51963.1;
GI:14010352)**

1 laesrcremd eqirlmdqnl kolsaaee ky sqkedkyeee ikilt dklke aetraefaer
61 tvaklektid dlevyrrkhq elqamqmelq speyklsklr tst

**MYH9-ALK Sequence (t(2;22)(p23;q11.2) chromosomal
translocation)***

**RANBP2-ALK Sequence (t(2;2)(p23;q13) or inv(2)(p23;q11-13)
chromosomal translocations)***

**ALO17-ALK Sequence (t(2;17)(p23;q25) chromosomal
translocation)***

**CARS-ALK Sequence (t(2;11;2)(p23;p15;q31) chromosomal
translocation)***

*** With the exception of MSN-ALK and MYH-9, all of the
fusion proteins contain the final 563 amino acids of ALK.
MSN-ALK and MYH9 contain the final 567 and 566 amino
acids, respectively.**

“ALK mutations” generally refer to alterations in a nucleic acid and/or amino acid sequence relative to a reference anaplastic lymphoma kinase sequence. In some embodiments, however, “ALK mutations” can refer to specific anaplastic lymphomas kinase mutations predictive of response to treatment with ALK inhibiting agents (e.g., PF-02341066 and/or PDD). For example, mutations of the cysteine amino acid at position 1156 (C1156) and/or the leucine amino acid at position 1196 (L1196) of wild type ALK protein (NP_004295) to a different amino acid are described herein to confer resistance to ALK inhibiting agents. In one embodiment, the C1156 position comprises a tyrosine amino acid and/or the L1196 position comprises a methionine amino acid. A skilled artisan will also recognize that amino acid positions corresponding to the “C1156” and “L1196” mutations of wild type ALK protein will have different numbers relative to different reference sequences (e.g., ALK homologs, ALK fusion proteins, etc.) without affecting the predictive value of response to treatment with ALK inhibiting agents (e.g., PF-02341066 and/or PDD). A skilled artisan will further recognize that there is a known and definite correspondence between the amino acid sequence of a particular protein and the nucleotide sequences that can code for the protein, as defined by the genetic code (shown below). Likewise, there is a known and definite correspondence between the nucleotide sequence of a particular nucleic acid and the amino acid sequence encoded by that nucleic acid, as defined by the genetic code.

GENETIC CODE

	Alanine (Ala, A)	GCA, GCC, GCG, GCT
	Arginine (Arg, R)	AGA, ACG, CGA, CGC, CGG, CGT
5	Asparagine (Asn, N)	AAC, AAT
	Aspartic acid (Asp, D)	GAC, GAT
	Cysteine (Cys, C)	TGC, TGT
	Glutamic acid (Glu, E)	GAA, GAG
	Glutamine (Gln, Q)	CAA, CAG
10	Glycine (Gly, G)	GGA, GGC, GGG, GGT
	Histidine (His, H)	CAC, CAT
	Isoleucine (Ile, I)	ATA, ATC, ATT
	Leucine (Leu, L)	CTA, CTC, CTG, CTT, TTA, TTG
	Lysine (Lys, K)	AAA, AAG
15	Methionine (Met, M)	ATG
	Phenylalanine (Phe, F)	TTC, TTT
	Proline (Pro, P)	CCA, CCC, CCG, CCT
	Serine (Ser, S)	AGC, AGT, TCA, TCC, TCG, TCT
	Threonine (Thr, T)	ACA, ACC, ACG, ACT
20	Tryptophan (Trp, W)	TGG
	Tyrosine (Tyr, Y)	TAC, TAT
	Valine (Val, V)	GTA, GTC, GTG, GTT
	Termination signal (end)	TAA, TAG, TGA

An important and well known feature of the genetic code is its redundancy, whereby, for most of the amino acids used to make proteins, more than one coding nucleotide triplet may be employed (for example, illustrated above). Therefore, a number of different nucleotide sequences may code for a given amino acid sequence. Such nucleotide sequences are considered functionally equivalent since they result in the production of the same amino acid sequence in all organisms (although certain organisms may translate some sequences more efficiently than they do others). Moreover, occasionally, a methylated variant of a purine or pyrimidine may be found in a given nucleotide sequence. Such methylations do not affect the coding relationship between the trinucleotide codon and the corresponding amino acid. In addition, a skilled artisan will

understand how to mutate nucleotides of a specific codon so as to specifically alter an encoded amino acid based on the relevant codon chart. For example, the codon for Cys-1156 is "TGC" and that for Tyr may be "TAT" or "TAC". Thus, a single nucleotide G-to-A substitution at position 2 of the codon will encode tyrosine rather than cysteine. A skilled
5 artisan can perform similar manipulations to design other mutations.

"Binding compound" shall refer to a binding composition, such as a small molecule, an antibody, a peptide, a peptide or non-peptide ligand, a protein, an oligonucleotide, an oligonucleotide analog, such as a peptide nucleic acid, a lectin, or any other molecular entity that is capable of specifically binding to a target protein or molecule or stable
10 complex formation with an analyte of interest, such as a complex of proteins.

"Binding moiety" means any molecule to which molecular tags can be directly or indirectly attached that is capable of specifically binding to an analyte. Binding moieties include, but are not limited to, antibodies, antibody binding compositions, peptides, proteins, nucleic acids and organic molecules having a molecular weight of up to about
15 1000 daltons and containing atoms selected from the group consisting of hydrogen, carbon, oxygen, nitrogen, sulfur and phosphorus.

A "biomarker" or "marker" is a gene, mRNA, or protein which may be altered, wherein said alteration is associated with cancer. The alteration may be in amount, structure, and/or activity in a cancer tissue or cancer cell, as compared to its amount, structure, and/or activity, in a normal or healthy tissue or cell (*e.g.*, a control), and is
20 associated with a disease state, such as cancer. For example, a marker of the present invention which is associated with cancer or predictive of responsiveness to anti-cancer therapeutics may have an altered nucleotide sequence, amino acid sequence, chromosomal translocation, intra-chromosomal inversion, copy number, expression level, protein level,
25 protein activity, or methylation status, in a cancer tissue or cancer cell as compared to a normal, healthy tissue or cell. Furthermore, a "marker" includes a molecule whose structure is altered, *e.g.*, mutated (contains an mutation), *e.g.*, differs from the wild type sequence at the nucleotide or amino acid level, *e.g.*, by substitution, deletion, or insertion, when present in a tissue or cell associated with a disease state, such as cancer.

30 The terms "cancer" or "tumor" refer to the presence of cells possessing characteristics typical of cancer-causing cells, such as uncontrolled proliferation, immortality, metastatic potential, rapid growth and proliferation rate, and certain

characteristic morphological features. Cancer cells are often in the form of a tumor, but such cells may exist alone within an animal, or may be a non-tumorigenic cancer cell, such as a leukemia cell. As used herein, the term "cancer" includes premalignant as well as malignant cancers. Cancers include, but are not limited to, B cell cancer, *e.g.*, multiple myeloma, Waldenström's macroglobulinemia, the heavy chain diseases, such as, for example, alpha chain disease, gamma chain disease, mu chain disease, benign monoclonal gammopathy, immunocytic amyloidosis, melanomas, breast cancer, lung cancer (such as non-small cell lung carcinoma or NSCLC), bronchus cancer, colorectal cancer, prostate cancer, pancreatic cancer, stomach cancer, ovarian cancer, urinary bladder cancer, brain or central nervous system cancer, peripheral nervous system cancer, esophageal cancer, cervical cancer, uterine or endometrial cancer, cancer of the oral cavity or pharynx, liver cancer, kidney cancer, testicular cancer, biliary tract cancer, small bowel or appendix cancer, salivary gland cancer, thyroid gland cancer, adrenal gland cancer, osteosarcoma, chondrosarcoma, cancer of hematological tissues, adenocarcinomas, inflammatory myofibroblastic tumors, gastrointestinal stromal tumor (GIST), colon cancer, multiple myeloma (MM), myelodysplastic syndrome (MDS), myeloproliferative disorder (MPD), acute lymphocytic leukemia (ALL), acute myelocytic leukemia (AML), chronic myelocytic leukemia (CML), chronic lymphocytic leukemia (CLL), polycythemia Vera, Hodgkin lymphoma, non-Hodgkin lymphoma (NHL), soft-tissue sarcoma, fibrosarcoma, myxosarcoma, liposarcoma, osteogenic sarcoma, chordoma, angiosarcoma, endotheliosarcoma, lymphangiosarcoma, lymphangioendotheliosarcoma, synovioma, mesothelioma, Ewing's tumor, leiomyosarcoma, rhabdomyosarcoma, squamous cell carcinoma, basal cell carcinoma, adenocarcinoma, sweat gland carcinoma, sebaceous gland carcinoma, papillary carcinoma, papillary adenocarcinomas, stadenocarcinoma, medullary carcinoma, bronchogenic carcinoma, renal cell carcinoma, hepatoma, bile duct carcinoma, choriocarcinoma, seminoma, embryonal carcinoma, Wilms' tumor, bladder carcinoma, epithelial carcinoma, glioma, astrocytoma, medulloblastoma, craniopharyngioma, ependymoma, pinealoma, hemangioblastoma, acoustic neuroma, oligodendroglioma, meningioma, neuroblastoma, retinoblastoma, follicular lymphoma, diffuse large B-cell lymphoma, mantle cell lymphoma, hepatocellular carcinoma, thyroid cancer, gastric cancer, head and neck cancer, small cell cancers, essential thrombocythemia, agnogenic myeloid metaplasia, hypereosinophilic syndrome, systemic mastocytosis, familial

hypereosinophilia, chronic eosinophilic leukemia, neuroendocrine cancers, carcinoid tumors, and the like.

"Chemotherapeutic agent" means a chemical substance, such as a cytotoxic or cytostatic agent, that is used to treat a condition, *e.g.*, cancer.

5 "Complementary" refers to the broad concept of sequence complementarity between regions of two nucleic acid strands or between two regions of the same nucleic acid strand. It is known that an adenine residue of a first nucleic acid region is capable of forming specific hydrogen bonds ("base pairing") with a residue of a second nucleic acid region which is antiparallel to the first region if the residue is thymine or uracil. Similarly, it is
10 known that a cytosine residue of a first nucleic acid strand is capable of base pairing with a residue of a second nucleic acid strand which is antiparallel to the first strand if the residue is guanine. A first region of a nucleic acid is complementary to a second region of the same or a different nucleic acid if, when the two regions are arranged in an antiparallel fashion, at least one nucleotide residue of the first region is capable of base pairing with a residue of
15 the second region. In certain embodiments, the first region comprises a first portion and the second region comprises a second portion, whereby, when the first and second portions are arranged in an antiparallel fashion, at least about 50%, at least about 75%, at least about 90%, or at least about 95% of the nucleotide residues of the first portion are capable of base pairing with nucleotide residues in the second portion. In other embodiments, all
20 nucleotide residues of the first portion are capable of base pairing with nucleotide residues in the second portion.

 The "copy number of a gene" or the "copy number of a marker" refers to the number of DNA sequences in a cell encoding a particular gene product. Generally, for a given gene, a mammal has two copies of each gene. The copy number can be increased, however, by
25 gene amplification or duplication, or reduced by deletion.

 A marker is "fixed" to a substrate if it is covalently or non-covalently associated with the substrate such that the substrate can be rinsed with a fluid (*e.g.*, standard saline citrate, pH 7.4) without a substantial fraction of the marker dissociating from the substrate.

 "Hazard ratio", as used herein, refers to a statistical method used to generate an
30 estimate for relative risk. "Hazard ratio" is the ratio between the predicted hazard of one group versus another group. For example, patient populations treated with an ALK inhibiting agent versus without an ALK inhibiting agent can be assessed for whether or not

the ALK inhibiting agent is effective in increasing the time to distant recurrence of disease, particularly with regard to ALK mutation status. For example, treating subjects harboring ALK mutations in cancerous tissue, as described herein, results in increased therapeutic benefit from ALK inhibiting agents relative to subjects not having said ALK mutations in cancerous tissue.

"ALK inhibiting agent" or "ALK inhibitor," as used herein, refers to a compound that can inhibit the biological activity of ALK. Biological activities can also include patient response as set forth in this application. Exemplary ALK inhibiting agents include, but are not limited to, PF-02341066, PDD, 2-methyl-11-(2-methylpropyl)-4-oxo-4,5,6,11,12,13-hexahydro-2H-indazolo[5,4-a]pyrrolo[3,4-c]carbazol-8-yl [4-(dimethylamino)benzyl]carbamate, (1S,2S,3R,4R)-3-({5-chloro-2-[(1-ethyl-2,3,4,5-tetrahydro-6-methoxy-2-oxo-1H-1-benzazepin-7-yl)amino]-4-pyrimidinyl}amino)bicyclo[2.2.1]hept-5-ene-2-carboxamide, and NVP-TAE684 (see, for example, *PNAS* 104:270-275, 2007; Choi, Y.L. *et al.* (2008) *Cancer Res.* 68:4971-2976; and *Biochemistry* 48:3600-3609, 2009, which are hereby incorporated by reference).

The terms "homology" or "identity," as used interchangeably herein, refer to sequence similarity between two polynucleotide sequences or between two polypeptide sequences, with identity being a more strict comparison. The phrases "percent identity or homology" and "% identity or homology" refer to the percentage of sequence similarity found in a comparison of two or more polynucleotide sequences or two or more polypeptide sequences. "Sequence similarity" refers to the percent similarity in base pair sequence (as determined by any suitable method) between two or more polynucleotide sequences. Two or more sequences can be anywhere from 0-100% similar, or any integer value there between. Identity or similarity can be determined by comparing a position in each sequence that may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same nucleotide base or amino acid, then the molecules are identical at that position. A degree of similarity or identity between polynucleotide sequences is a function of the number of identical or matching nucleotides at positions shared by the polynucleotide sequences. A degree of identity of polypeptide sequences is a function of the number of identical amino acids at positions shared by the polypeptide sequences. A degree of homology or similarity of polypeptide sequences is a function of the number of amino acids at positions shared by the polypeptide sequences. The term "substantial homology,"

as used herein, refers to homology of at least 50%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% or more.

Cancer is "inhibited" if at least one symptom of the cancer is alleviated, terminated, slowed, or prevented. As used herein, cancer is also "inhibited" if recurrence or metastasis
5 of the cancer is reduced, slowed, delayed, or prevented.

A "marker nucleic acid" or "biomarker nucleic acid" is a nucleic acid (*e.g.*, DNA, mRNA, cDNA) encoded by or corresponding to a marker of the present invention. For example, such marker nucleic acid molecules include DNA (*e.g.*, genomic DNA and cDNA) comprising the entire or a partial sequence of any of the nucleic acid sequences set
10 forth in Table 1 or the complement or hybridizing fragment of such a sequence. The marker nucleic acid molecules also include RNA comprising the entire or a partial sequence of any of the nucleic acid sequences set forth in Table 1 or the complement of such a sequence, wherein all thymidine residues are replaced with uridine residues. A "marker protein" is a protein encoded by or corresponding to a marker of the present invention. A
15 marker protein comprises the entire or a partial sequence of a protein encoded by any of the sequences set forth in Table 1 or a fragment thereof. The terms "protein" and "polypeptide" are used interchangeably herein.

The "normal" copy number of a marker or "normal" level of expression of a marker is the level of expression, copy number of the marker, in a biological sample, *e.g.*, a sample
20 containing sputum, bronchoalveolar lavage, pleural effusion, tissue, whole blood, serum, plasma, buccal scrape, saliva, cerebrospinal fluid, urine, stool, and bone marrow, from a subject, *e.g.*, a human, not afflicted with cancer.

An "overexpression" or "significantly higher level of expression, copy number, and/or activity" of ALK gene mutations and/or gene products (*e.g.*, the markers set forth in
25 Table 1) refers to an expression level, copy number, and/or activity in a test sample that is greater than the standard error of the assay employed to assess expression or copy number, and may be at least two, at least three, at least four, at least five, or at least ten or more times the expression level or copy number of the ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) in a control sample (*e.g.*, a sample from a healthy
30 subject not afflicted with cancer), or the average expression level or copy number of the ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) in several control samples.

The term "probe" refers to any molecule which is capable of selectively binding to a specifically intended target molecule, for example a marker of the present invention.

Probes can be either synthesized by one skilled in the art, or derived from appropriate biological preparations. For purposes of detection of the target molecule, probes may be specifically designed to be labeled, as described herein. Examples of molecules that can be utilized as probes include, but are not limited to, RNA, DNA, proteins, antibodies, and organic monomers.

"RECIST" shall mean an acronym that stands for "Response Evaluation Criteria in Solid Tumours" and is a set of published rules that define when cancer patients improve ("respond"), stay the same ("stable") or worsen ("progression") during treatments. Response as defined by RECIST criteria have been published, for example, at Journal of the National Cancer Institute, Vol. 92, No. 3, Feb. 2, 2000 and RECIST criteria may include other similar published definitions and rule sets. One skilled in the art would understand definitions that go with RECIST criteria, as used herein, such as "PR," "CR," "SD" and "PD."

"Responsiveness," to "respond" to treatment, and other forms of this verb, as used herein, refer to the reaction of a subject to treatment with an ALK inhibiting agent. As an example, a subject responds to treatment with an ALK inhibiting agent if growth of a tumor in the subject is retarded about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90% or more. In another example, a subject responds to treatment with an ALK inhibiting agent if a tumor in the subject shrinks by about 5%, about 10%, about 20%, about 30%, about 40%, about 50% or more as determined by any appropriate measure, *e.g.*, by mass or volume. In another example, a subject responds to treatment with an ALK inhibiting agent if the subject experiences a life expectancy extended by about 5%, about 10%, about 20%, about 30%, about 40%, about 50% or more beyond the life expectancy predicted if no treatment is administered. In another example, a subject responds to treatment with an ALK inhibiting agent if the subject has an increased disease-free survival, overall survival or increased time to progression. Several methods may be used to determine if a patient responds to a treatment including the RECIST criteria, as set forth above.

"Sample," "tissue sample," "patient sample," "patient cell or tissue sample" or "specimen" each refer to a collection of similar cells obtained from a tissue of a subject or patient. The source of the tissue sample may be solid tissue as from a fresh, frozen and/or

preserved organ, tissue sample, biopsy, or aspirate; blood or any blood constituents; bodily fluids such as cerebral spinal fluid, amniotic fluid, peritoneal fluid or interstitial fluid; or cells from any time in gestation or development of the subject. The tissue sample may contain compounds that are not naturally intermixed with the tissue in nature such as
5 preservatives, anticoagulants, buffers, fixatives, nutrients, antibiotics or the like.

The amount of a marker, *e.g.*, expression or copy number of ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1), in a subject is "significantly" higher or lower than the normal amount of a marker, if the amount of the marker is greater or less, respectively, than the normal level by an amount greater than the standard error of
10 the assay employed to assess amount, or at least two, at least three, at least four, at least five, at least ten or more times that amount. Alternately, the amount of the marker in the subject can be considered "significantly" higher or lower than the normal amount if the amount is at least about two, at least about three, at least about four, or at least about five times, higher or lower, respectively, than the normal amount of the marker.

As used herein, "significant event" shall refer to an event in a patient's disease that is important as determined by one skilled in the art. Examples of significant events include, for example, without limitation, primary diagnosis, death, recurrence, the determination that a patient's disease is metastatic, relapse of a patient's disease or the progression of a patient's disease from any one of the above noted stages to another. A significant event may
20 be any important event used to assess OS, TTP and/or using the RECIST or other response criteria, as determined by one skilled in the art.

As used herein, the terms "subject" and "patient" are used interchangeably. As used herein, the terms "subject" and "subjects" refer to an animal, *e.g.*, a mammal including a non-primate (*e.g.*, a cow, pig, horse, donkey, goat, camel, cat, dog, guinea pig, rat, mouse,
25 sheep) and a primate (*e.g.*, a monkey, such as a cynomolgous monkey, gorilla, chimpanzee and a human).

As used herein, "time course" shall refer to the amount of time between an initial event and a subsequent event. For example, with respect to a patient's cancer, time course may relate to a patient's disease and may be measured by gauging significant events in the
30 course of the disease, wherein the first event may be diagnosis and the subsequent event may be metastasis, for example.

"Time to progression" or "TTP" refers to a time as measured from the start of the treatment to progression or a cancer or censor. Censoring may come from a study end or from a change in treatment. Time to progression can also be represented as a probability as, for example, in a Kaplan-Meier plot where time to progression may represent the probability of being progression free over a particular time, that time being the time between the start of the treatment to progression or censor.

A "transcribed polynucleotide" is a polynucleotide (e.g., an RNA, a cDNA, or an analog of one of an RNA or cDNA) which is complementary to or homologous with all or a portion of a mature RNA made by transcription of a marker of the present invention and normal post-transcriptional processing (e.g., splicing), if any, of the transcript, and reverse transcription of the transcript.

"Treat," "treatment," and other forms of this word refer to the administration of an ALK inhibiting agent to impede growth of a cancer, to cause a cancer to shrink by weight or volume, to extend the expected survival time of the subject and or time to progression of the tumor or the like.

An "underexpression" or "significantly lower level of expression, copy number, and/or activity" of ALK gene mutations and/or gene products (e.g., the markers set forth in Table 1) refers to an expression level or copy number in a test sample that is greater than the standard error of the assay employed to assess expression or copy number, for example, at least twice, at least three, at least four, at least five, or at least ten or more times less than the expression level, copy number, and/or activity of the ALK gene mutations and/or gene products (e.g., the markers set forth in Table 1) in a control sample (e.g., a sample from a healthy subject not afflicted with cancer), or the average expression level, copy number, and/or activity of the ALK gene mutations and/or gene products (e.g., the markers set forth in Table 1) in several control samples.

It will be understood that the term "comprising" is used herein in an inclusive manner unless the context clearly requires otherwise.

It will be understood that references herein to prior art should not be taken as an admission of common general knowledge in the field.

II. Exemplary Methods of the Invention

The present invention is based, at least in part, on the identification of specific regions of the genome, including, for example, ALK mutations, associated with predicting efficacy of ALK inhibitors in treating cancer. Analysis of ALK gene expression sequences has led to the identification of novel mutations to ALK polypeptides (e.g., biomarkers listed in Table 1 , including EML4-ALK polypeptides) that can render the polypeptides at least partially resistant to therapy with ALK inhibitors. Accordingly, the presence and/or

absence of one or more of such biomarkers in various methods described herein is within the scope of the present invention.

In some embodiments, methods of the present invention may be used to monitor the progression of cancer in a subject, wherein if a sample in a subject presents one or more
5 ALK mutations (*e.g.*, EML4-ALK mutations) identified herein during the progression of cancer, *e.g.*, at a first point in time and a subsequent point in time, then the cancer is less likely to respond to ALK inhibitor-mediated treatment and vice versa. In yet another embodiment, between the first point in time and a subsequent point in time, the subject has undergone treatment, *e.g.*, chemotherapy, radiation therapy, surgery, or any other
10 therapeutic approach useful for inhibiting cancer, has completed treatment, or is in remission.

As described further herein, one or more biomarkers of the present inventions (*e.g.*, ALK mutations, including EML4-ALK mutations) can be specifically identified by the presence in genomic (*e.g.*, germline and/or somatic) sequence when compared to a
15 reference sequence, such as SEQ ID NO:1. For example, the methods described herein can involve detecting biomarkers of the present invention by carrying out a target nucleic acid amplification reaction of a stretch of DNA comprising one or more mutations listed in Table 1 and analyzing the amplified target nucleic acid for the presence of the one or more mutations.

20 Various techniques for amplifying nucleic acid are known in the art, such as: PCR (Polymerase Chain Reaction), described in U.S. Pat. No. 4,683,195 (incorporated by reference), U.S. Pat. No. 4,683,202 (incorporated by reference) and U.S. Pat. No. 4,800,159 (incorporated by reference), and its RT-PCR alternative (Reverse Transcription PCR), particularly in its one-step format as disclosed in patent EP-B-0.569.272, LCR (Ligase
25 Chain Reaction), as described for example in patent application EP-A-0.201.184, RCR (Repair Chain Reaction), as described for example in international application WO-A-90/01069 (incorporated by reference), 3SR (Self Sustained Sequence Replication), as described for example in patent application WO-A-90/06995 (incorporated by reference), NASBA (Nucleic Acid Sequence-Based Amplification), as described for example instance
30 in EP-B-0.397.269 and U.S. Pat. No. 5,466,586 (incorporated by reference) using double stranded DNA as template, and TMA (Transcription Mediated Amplification), as described for example in U.S. Pat. No. 5,399,491 (incorporated by reference).

Detection of the presence of one or more of the mutations in the amplified product can be performed in various manners that are well known in the art, such as DNA sequencing methodologies like Sanger sequencing and deep sequencing, use of restriction enzymes, allele specific amplification, Peptide Nucleic Acid (PNA)-mediated PCR, 5 detection of conformational differences, like Single Strand Conformation Polymorphism (SSCP) and Denaturing Gradient Gel Electrophoresis (DGGE) assays with detection steps on membranes (dot blot) using labeled oligonucleotide probes, assays with detection steps in microtiter plates, like Reverse Hybridization, Oligonucleotide Ligation Assay (OLA, MLPA), First Nucleotide Change (FNC) technology, Cross-linking technology, Rapid cycle 10 PCR and simultaneous fluorescence analysis (*e.g.* 5'nuclease/Taqman), and PCR followed by mini-sequencing using mass spectrometry or capillary electrophoresis

III. Exemplary Isolated Nucleic Acid Molecules

One aspect of the present invention pertains to isolated nucleic acid molecules that correspond to a biomarker of the present invention, including nucleic acids which encode a 15 polypeptide corresponding to a marker of the present invention or a portion of such a polypeptide. The nucleic acid molecules of the present invention include those nucleic acid molecules which reside in ALK or ALK-related genomic (*e.g.*, germline and/or somatic) regions identified herein and/or encode ALK or ALK-related (*e.g.*, EML4-ALK) polypeptides. In some embodiments, the nucleic acid molecules of the present invention 20 comprise, consist essentially of, or consist of the nucleic sequences, or fragments thereof, presented in Table 1. Isolated nucleic acid molecules of the present invention also include nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules that correspond to a marker of the present invention, including nucleic acid molecules which encode a polypeptide corresponding to a marker of the present invention, 25 and fragments of such nucleic acid molecules, *e.g.*, those suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded; 30 in certain embodiments the nucleic acid molecule is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule. In certain embodiments, an "isolated" nucleic acid molecule is free of sequences (such as

protein-encoding sequences) which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, less than about 4 kB, less than about 3 kB, less than about 2 kB, less than about 1 kB, less than about 0.5 kB or less than about 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

The language "substantially free of other cellular material or culture medium" includes preparations of nucleic acid molecule in which the molecule is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, nucleic acid molecule that is substantially free of cellular material includes preparations of nucleic acid molecule having less than about 30%, less than about 20%, less than about 10%, or less than about 5% (by dry weight) of other cellular material or culture medium.

A nucleic acid molecule of the present invention, *e.g.*, ALK gene mutations set forth in Table 1), can be isolated using standard molecular biology techniques and the sequence information in the database records described herein. Using all or a portion of such nucleic acid sequences, nucleic acid molecules of the present invention can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, ed., *Molecular Cloning: A Laboratory Manual, 2nd ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid molecule of the present invention can be amplified using cDNA, mRNA, or genomic DNA (*e.g.*, germline and/or somatic genomic DNA) as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid molecules so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to all or a portion of a nucleic acid molecule of the present invention can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In another embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleic acid molecule which has a nucleotide sequence complementary to the

nucleotide sequence of a nucleic acid corresponding to a marker of the present invention or to the nucleotide sequence of a nucleic acid encoding a protein which corresponds to a marker of the present invention. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given
5 nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, a nucleic acid molecule of the present invention can comprise only a portion of a nucleic acid sequence, wherein the full length nucleic acid sequence comprises a marker of the present invention or which encodes a polypeptide corresponding to a
10 marker of the present invention. Such nucleic acid molecules can be used, for example, as a probe or primer. The probe/primer typically is used as one or more substantially purified oligonucleotides. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, at least about 8, at least about 9, at least about 10, at least about 11, at least about 12, at least about 13, at least about 14, at
15 least about 15, at least about 16, at least about 17, at least about 18, at least about 19, at least about 20, at least about 21, at least about 22, at least about 23, at least about 24, at least about 25, at least about 26, at least about 27, at least about 28, at least about 29, at least about 30, at least about 31, at least about 32, at least about 33, at least about 34, at least about 35, at least about 36, at least about 37, at least about 38, at least about 39, at
20 least about 40, at least about 45, at least about 50, at least about 55 at least about 60, at least about 65, at least about 70, at least about 75, at least about 80, at least about 85 kb, at least about 90, at least about 95, at least about 100 or more consecutive nucleotides of a nucleic acid of the present invention.

Probes based on the sequence of a nucleic acid molecule of the present invention
25 can be used to detect transcripts or genomic sequences corresponding to one or more markers of the present invention. The probe comprises a label group attached thereto, *e.g.*, a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues which mis-express the protein, such as by measuring levels of a nucleic acid molecule encoding the protein in
30 a sample of cells from a subject, *e.g.*, detecting mRNA levels or determining whether a gene encoding the protein has been mutated or deleted.

The invention further encompasses nucleic acid molecules that are substantially homologous to the ALK gene mutations and/or gene products (*e.g.*, the markers set forth in

Table 1) such that they are at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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%, at least 99%, at least 99.5% or greater. In other embodiments, the invention

5 further encompasses nucleic acid molecules that are substantially homologous to the ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) such that they differ by only or at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at

10 least 24, at least 25, at least 26, at least 27, at least 28, at least 29, at least 30, at least 31, at least 32, at least 33, at least 34, at least 35, at least 36, at least 37, at least 38, at least 39, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, at least 100 nucleotides or any range in between.

The term “single nucleotide polymorphism” (SNP) refers to a polymorphic site

15 occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele (*e.g.*, sequences that vary in less than 1/100 or 1/1000 members of a population). A SNP usually arises due to substitution of one nucleotide for another at the polymorphic site. SNPs can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative

20 to a reference allele. Typically the polymorphic site is occupied by a base other than the reference base. For example, where the reference allele contains the base “T” (thymidine) at the polymorphic site, the altered allele can contain a “C” (cytidine), “G” (guanine), or “A” (adenine) at the polymorphic site. SNP’s may occur in protein-coding nucleic acid sequences, in which case they may give rise to a defective or otherwise variant protein, or

25 genetic disease. Such a SNP may alter the coding sequence of the gene and therefore specify another amino acid (a “missense” SNP) or a SNP may introduce a stop codon (a “nonsense” SNP). When a SNP does not alter the amino acid sequence of a protein, the SNP is called “silent.” SNP’s may also occur in noncoding regions of the nucleotide sequence. This may result in defective protein expression, *e.g.*, as a result of alternative

30 splicing, or it may have no effect on the function of the protein.

In another embodiment, an isolated nucleic acid molecule of the present invention is at least 7, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at

least 90, at least 95, at least 100, at least 125, at least 150, at least 175, at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, at least 550, at least 650, at least 700, at least 800, at least 900, at least 1000, at least 1200, at least 1400, at least 1600, at least 1800, at least 2000, at least 2200, at least 2400, at least 2600, at least 2800, at least 3000, at least 3500, at least 4000, at least 4500, or more nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule corresponding to a marker of the present invention or to a nucleic acid molecule encoding a protein corresponding to a marker of the present invention. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, or at least 85% identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in sections 6.3.1-6.3.6 of *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989). Another, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50-65°C.

The invention also includes molecular beacon nucleic acid molecules having at least one region which is complementary to a nucleic acid molecule of the present invention, such that the molecular beacon is useful for quantitating the presence of the nucleic acid molecule of the present invention in a sample. A "molecular beacon" nucleic acid is a nucleic acid molecule comprising a pair of complementary regions and having a fluorophore and a fluorescent quencher associated therewith. The fluorophore and quencher are associated with different portions of the nucleic acid in such an orientation that when the complementary regions are annealed with one another, fluorescence of the fluorophore is quenched by the quencher. When the complementary regions of the nucleic acid molecules are not annealed with one another, fluorescence of the fluorophore is quenched to a lesser degree. Molecular beacon nucleic acid molecules are described, for example, in U.S. Patent 5,876,930 (incorporated by reference).

IV. Exemplary Isolated Proteins and Antibodies

One aspect of the present invention pertains to isolated proteins which correspond to individual markers of the present invention, and biologically active portions thereof. In one embodiment, the native polypeptide corresponding to a marker can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification

techniques. In another embodiment, polypeptides corresponding to a marker of the present invention are produced by recombinant DNA techniques. Alternative to recombinant expression, a polypeptide corresponding to a marker of the present invention can be synthesized chemically using standard peptide synthesis techniques.

5 An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the protein is separated from cellular
10 components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, less than about 20%, less than about 10%, or less than about 5% (by dry weight) of heterologous protein (also referred to herein as a "contaminating protein"). When the protein or biologically active portion thereof is recombinantly produced, it may
15 be substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the protein preparation. When the protein is produced by chemical synthesis, it may substantially be free of chemical precursors or other chemicals, *i.e.*, it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such
20 preparations of the protein have less than about 30%, less than about 20%, less than about 10%, less than about 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest.

 Biologically active portions of a polypeptide corresponding to a marker of the present invention include polypeptides comprising amino acid sequences sufficiently
25 identical to or derived from the amino acid sequence of the protein corresponding to ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) of the present invention, which include fewer amino acids than the full length protein, and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A
30 biologically active portion of a protein of the present invention can be a polypeptide which is, for example, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be

prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the present invention.

In certain embodiments, the polypeptide has an amino acid sequence of a protein encoded by a nucleic acid molecule listed in Table 1. Other useful proteins are
5 substantially identical (*e.g.*, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 86, at least 87, at least 88, at least 89, 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, at least 99, at least 99.5% or greater) to one of these sequences and retain the functional activity of the protein (*e.g.*, conferring resistance or sensitivity to an ALK inhibitor) of the corresponding full-length
10 protein yet differ in amino acid sequence.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or
15 nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = #
20 of identical positions/total # of positions (*e.g.*, overlapping positions) x 100). In one embodiment the two sequences are the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. Another, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and
25 Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a
30 nucleic acid molecules of the present invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to protein molecules of the present invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (1997)

Nucleic Acids Res. 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules. When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used (see the NCBI website on the world wide web at ncbi.nlm.nih.gov). Another non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) *Comput Appl Biosci*, 4:11-7. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Yet another useful algorithm for identifying regions of local sequence similarity and alignment is the FASTA algorithm as described in Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85:2444-2448. When using the FASTA algorithm for comparing nucleotide or amino acid sequences, a PAM120 weight residue table can, for example, be used with a *k*-tuple value of 2.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

An isolated polypeptide corresponding to a marker of the present invention, or a fragment thereof, can be used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. The full-length polypeptide or protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the present invention comprises at least 8 (or at least 10, at least 15, at least 20, or at least 30 or more) amino acid residues of the amino acid sequence of one of the polypeptides of the present invention, and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with a marker of the present invention to which the protein corresponds. Exemplary epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, *e.g.*, hydrophilic regions. Hydrophobicity sequence analysis, hydrophilicity sequence analysis, or similar analyses can be used to identify hydrophilic regions.

An immunogen typically is used to prepare antibodies by immunizing a suitable (*i.e.*, immunocompetent) subject such as a rabbit, goat, mouse, or other mammal or

vertebrate. An appropriate immunogenic preparation can contain, for example, recombinantly-expressed or chemically-synthesized polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or a similar immunostimulatory agent.

5 Accordingly, another aspect of the present invention pertains to antibodies directed against a polypeptide of the present invention. The terms "antibody" and "antibody substance" as used interchangeably herein refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds an antigen, such as a polypeptide of the
10 present invention. A molecule which specifically binds to a given polypeptide of the present invention is a molecule which binds the polypeptide, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an
15 enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope.

 Polyclonal antibodies can be prepared as described above by immunizing a suitable
20 subject with a polypeptide of the present invention as an immunogen. The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules can be harvested or isolated from the subject (*e.g.*, from the blood or serum of the subject) and further purified by well-known techniques, such as protein A
25 chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497, the human B cell hybridoma technique (see Kozbor *et al.*, 1983, *Immunol. Today*
30 4:72), the EBV-hybridoma technique (see Cole *et al.*, pp. 77-96 In *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., 1985) or trioma techniques. The technology for producing hybridomas is well known (see generally *Current Protocols in Immunology*, Coligan *et al.* ed., John Wiley & Sons, New York, 1994). Hybridoma cells producing a

monoclonal antibody of the present invention are detected by screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal
5 antibody directed against a polypeptide of the present invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with the polypeptide of interest. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia
10 *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409 (incorporated by reference); PCT
Publication No. WO 92/18619 (incorporated by reference); PCT Publication No. WO 91/17271 (incorporated by reference); PCT Publication No. WO 92/20791 (incorporated by
15 reference); PCT Publication No. WO 92/15679 (incorporated by reference); PCT Publication No. WO 93/01288 (incorporated by reference); PCT Publication No. WO 92/01047 (incorporated by reference); PCT Publication No. WO 92/09690 (incorporated by reference); PCT Publication No. WO 90/02809 (incorporated by reference); Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-
20 85; Huse *et al.* (1989) *Science* 246:1275- 1281; Griffiths *et al.* (1993) *EMBO J.* 12:725-734.

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the present invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA
25 techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671 (incorporated by reference); European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533 (incorporated by reference); U.S. Patent No. 4,816,567 (incorporated by reference); European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-
30 1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Cancer Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison (1985) *Science*

229:1202-1207; Oi *et al.* (1986) *Bio/Techniques* 4:214; U.S. Patent 5,225,539 (incorporated by reference); Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

Completely human antibodies are particularly desirable for therapeutic treatment of human subjects. Such antibodies can be produced using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, *e.g.*, all or a portion of a polypeptide corresponding to a marker of the present invention. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995) *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, see, *e.g.*, U.S. Patent 5,625,126 (incorporated by reference); U.S. Patent 5,633,425 (incorporated by reference); U.S. Patent 5,569,825 (incorporated by reference); U.S. Patent 5,661,016 (incorporated by reference); and U.S. Patent 5,545,806 (incorporated by reference). In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, *e.g.*, a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope (Jespers *et al.*, 1994, *Bio/technology* 12:899-903).

An antibody directed against a polypeptide corresponding to a marker of the present invention (*e.g.*, a monoclonal antibody) can be used to isolate the polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the marker (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the level and pattern of expression of the marker. The antibodies can also be used diagnostically to monitor protein levels in tissues or body fluids (*e.g.*, in a tumor

cell-containing body fluid) as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include, but are not limited to, various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include, but are not limited to, horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include, but are not limited to, streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include, but are not limited to, umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes, but is not limited to, luminol; examples of bioluminescent materials include, but are not limited to, luciferase, luciferin, and aequorin, and examples of suitable radioactive materials include, but are not limited to, ^{125}I , ^{131}I , ^{35}S or ^3H .

15 **V. Exemplary Recombinant Expression Vectors and Host Cells**

Another aspect of the present invention pertains to vectors, such as expression vectors, containing a nucleic acid encoding a polypeptide corresponding to a marker of the present invention (or a portion of such a polypeptide). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, namely expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the present invention comprise a nucleic acid of the present invention in a form suitable for expression of the nucleic acid in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is

5 operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is

10 intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, *Methods in Enzymology: Gene Expression Technology* vol.185, Academic Press, San Diego, CA (1991). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct

15 expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the present invention can be introduced into host cells to thereby produce proteins or

20 peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

The recombinant expression vectors of the present invention can be designed for expression of a polypeptide corresponding to a marker of the present invention in prokaryotic (*e.g.*, *E. coli*) or eukaryotic cells (*e.g.*, insect cells {using baculovirus

25 expression vectors}, yeast cells or mammalian cells). Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with

30 vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2)

to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988, *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, 1988, *Gene* 69:301-315) and pET 11d (Studier *et al.*, p. 60-89, In *Gene Expression Technology: Methods in Enzymology* vol.185, Academic Press, San Diego, CA, 1991). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a co-expressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21 (DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacterium with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, p. 119-128, In *Gene Expression Technology: Methods in Enzymology* vol. 185, Academic Press, San Diego, CA, 1990). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid may be among those utilized in *E. coli* (Wada *et al.*, 1992, *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the present invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari *et al.*, 1987, *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, 1982, *Cell* 30:933-943), pJRY88 (Schultz *et al.*, 1987, *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and pPicZ (Invitrogen Corp, San Diego, CA).

Alternatively, the expression vector is a baculovirus expression vector. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.*, 1983, *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers, 1989, *Virology* 170:31-39).

5 In yet another embodiment, a nucleic acid of the present invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987, *Nature* 329:840) and pMT2PC (Kaufman *et al.*, 1987, *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly
10 used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook *et al.*, *supra*.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid in a particular cell type (*e.g.*, tissue-specific
15 regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.*, 1987, *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton, 1988, *Adv. Immunol.* 43:235-275), such as promoters of T cell receptors (Winoto and Baltimore, 1989, *EMBO J.* 8:729-733) and
20 immunoglobulins (Banerji *et al.*, 1983, *Cell* 33:729-740; Queen and Baltimore, 1983, *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989, *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.*, 1985, *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 (incorporated by reference) and European
25 Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss, 1990, *Science* 249:374-379) and the α -fetoprotein promoter (Camper and Tilghman, 1989, *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA
30 molecule of the present invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA

molecule which is antisense to the mRNA encoding a polypeptide of the present invention. Regulatory sequences operably linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences
5 can be chosen which direct constitutive, tissue-specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid, or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of
10 gene expression using antisense genes see Weintraub *et al.*, 1986, *Trends in Genetics*, Vol. 1(1).

Another aspect of the present invention pertains to host cells into which a recombinant expression vector of the present invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood
15 that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic (*e.g.*, *E. coli*) or eukaryotic cell (*e.g.*, insect cells,
20 yeast or mammalian cells).

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-
25 precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may
30 integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Exemplary

selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

5 A host cell of the present invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce a polypeptide corresponding to a marker of the present invention. Accordingly, the invention further provides methods for producing a polypeptide corresponding to a marker of the present invention using the host cells of the present invention. In one embodiment, the method comprises culturing the host cell of
10 invention (into which a recombinant expression vector encoding a polypeptide of the present invention has been introduced) in a suitable medium such that the marker is produced. In another embodiment, the method further comprises isolating the marker polypeptide from the medium or the host cell.

 The host cells of the present invention can also be used to produce nonhuman
15 transgenic animals. For example, in one embodiment, a host cell of the present invention is a fertilized oocyte or an embryonic stem cell into which sequences encoding a polypeptide corresponding to a marker of the present invention have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous sequences encoding a marker protein of the present invention have been introduced into their genome
20 or homologous recombinant animals in which endogenous gene(s) encoding a polypeptide corresponding to a marker of the present invention sequences have been altered. Such animals are useful for studying the function and/or activity of the polypeptide corresponding to the marker, for identifying and/or evaluating modulators of polypeptide activity, as well as in pre-clinical testing of therapeutics or diagnostic molecules, for marker
25 discovery or evaluation, *e.g.*, therapeutic and diagnostic marker discovery or evaluation, or as surrogates of drug efficacy and specificity.

 As used herein, a "transgenic animal" is a non-human animal, *e.g.*, a mammal, such as a rodent, *e.g.*, a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs,
30 cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an

"homologous recombinant animal" is a non-human animal, such as a mammal, *e.g.*, a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

- 5 Transgenic animals also include inducible transgenic animals, such as those described in, for example, Chan I.T., *et al.* (2004) *J Clin Invest.* 113(4):528-38 and Chin L. *et al* (1999) *Nature* 400(6743):468-72.

A transgenic animal of the present invention can be created by introducing a nucleic acid encoding a polypeptide corresponding to a marker of the present invention into the
10 male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the polypeptide of the present
15 invention to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 (incorporated by reference) and 4,870,009 (incorporated by reference), U.S. Patent No. 4,873,191 (incorporated by reference) and in Hogan, *Manipulating the Mouse Embryo*,
20 Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of mRNA encoding the transgene in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover,
25 transgenic animals carrying the transgene can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a gene encoding a polypeptide corresponding to a marker of the present invention into which a deletion, addition or substitution has been introduced to thereby
30 alter, *e.g.*, functionally disrupt, the gene. In another embodiment, the vector is designed such that, upon homologous recombination, the endogenous gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the

endogenous gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous protein). In the homologous recombination vector, the altered portion of the gene is flanked at its 5' and 3' ends by additional nucleic acid of the gene to allow for
5 homologous recombination to occur between the exogenous gene carried by the vector and an endogenous gene in an embryonic stem cell. The additional flanking nucleic acid sequences are of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see, *e.g.*, Thomas and Capecchi, 1987, *Cell* 51:503 for a
10 description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous gene are selected (see, *e.g.*, Li *et al.*, 1992, *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (see, *e.g.*, Bradley, *Teratocarcinomas and*
15 *Embryonic Stem Cells: A Practical Approach*, Robertson, Ed., IRL, Oxford, 1987, pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the
20 transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication Nos. WO 90/11354 (incorporated by reference), WO 91/01140 (incorporated by reference), WO 92/0968 (incorporated by reference), and WO 93/04169 (incorporated by reference).

25 In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, *e.g.*, Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP
30 recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.*, 1991, *Science* 251:1351-1355). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double"

transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.* (1997) *Nature* 385:810-813 and PCT Publication Nos. WO 97/07668 (incorporated by reference) and WO 97/07669 (incorporated by reference).

V. Exemplary Kits

A kit is any manufacture (*e.g.*, a package or container) comprising at least one reagent, *e.g.*, a probe, for specifically detecting a marker of the present invention, the manufacture being promoted, distributed, or sold as a unit for performing the methods of the present invention. When the compositions, kits, and methods of the present invention are used for carrying out the methods of the present invention, the ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) of the present invention may be selected such that a positive result is obtained in at least about 20%, at least about 40%, at least about 60%, at least about 80%, at least about 90%, at least about 95%, at least about 99% or in 100% of subjects afflicted with cancer, of the corresponding stage, grade, histological type, or benign/premalignant/malignant nature. In certain embodiments, the marker or panel of markers of the present invention may be selected such that a PPV (positive predictive value) of greater than about 10% is obtained for the general population (*e.g.*, coupled with an assay specificity greater than 99.5%).

When a plurality of ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) of the present invention are used in the compositions, kits, and methods of the present invention, the amount, structure, and/or activity of each marker or level of expression or copy number can be compared with the normal amount, structure, and/or activity of each of the plurality of markers or level of expression or copy number, in non-cancerous samples of the same type, either in a single reaction mixture (*i.e.*, using reagents, such as different fluorescent probes, for each marker) or in individual reaction mixtures corresponding to one or more of the ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1). If a plurality of ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) is used, then 2, 3, 4, 5, 6, 7, 8, 9, 10, or more individual markers may be used or identified.

The invention includes compositions, kits, and methods for assaying cancer cells in a sample (*e.g.*, an archived tissue sample or a sample obtained from a subject). These compositions, kits, and methods are substantially the same as those described above, except that, where necessary, the compositions, kits, and methods are adapted for use with certain
5 types of samples.

The invention thus includes a kit for assessing the presence of cancer cells having or likely to have reduced responsiveness to ALK inhibitors (*e.g.*, in a sample such as a subject sample). The kit may comprise one or more reagents capable of identifying ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) of the present
10 invention, *e.g.*, binding specifically with a nucleic acid or polypeptide corresponding to ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) of the present invention. Suitable reagents for binding with a polypeptide corresponding to a marker of the present invention include antibodies, antibody derivatives, antibody fragments, and the like. Suitable reagents for binding with a nucleic acid (*e.g.*, a genomic
15 DNA, an mRNA, a spliced mRNA, a cDNA, or the like) include complementary nucleic acids. For example, the nucleic acid reagents may include oligonucleotides (labeled or non-labeled) fixed to a substrate, labeled oligonucleotides not bound with a substrate, pairs of PCR primers, molecular beacon probes, and the like. In some embodiments, the kits can comprise reagents useful for performing the methods described herein, such as comprising
20 at least one pair of primers recognizing and hybridizing to stretches of nucleic acid surrounding at least one stretch of nucleic acid comprising at least one mutation listed in Table 1 and means for detecting the amplified-target nucleic acid for the presence of said mutation.

The kit of the present invention may optionally comprise additional components
25 useful for performing the methods of the present invention. By way of example, the kit may comprise fluids (*e.g.*, SSC buffer) suitable for annealing complementary nucleic acids or for binding an antibody with a protein with which it specifically binds, one or more sample compartments, an instructional material which describes performance of a method of the present invention, a sample of normal cells, a sample of cancer cells, and the like.

30 A kit of the present invention may comprise a reagent useful for determining protein level or protein activity of a marker. In another embodiment, a kit of the present invention may comprise a reagent for determining methylation status of a marker, or may comprise a reagent for determining alteration of structure of a marker, *e.g.*, the presence of a mutation.

VI. Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, pharmacogenomics, and monitoring clinical trials are used for predictive purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to assays for determining the amount, structure, and/or activity of polypeptides or nucleic acids corresponding to one or more markers of the present invention, in order to determine whether an individual having cancer or at risk of developing cancer will be more likely to respond to ALK inhibitor-mediated therapy.

Accordingly, in one aspect, the invention is drawn to a method for determining whether a subject with a cancer is likely to respond to treatment with an ALK inhibiting agent. In another aspect, the invention is drawn to a method for predicting a time course of disease. In still another aspect, the method is drawn to a method for predicting a probability of a significant event in the time course of the disease. In certain embodiments, the method comprises detecting a biomarker or combination of biomarkers associated with responsiveness to treatment with an ALK inhibiting agent (*e.g.*, ALK mutations) as described herein, and determining whether the subject is likely to respond to treatment with the ALK inhibiting agent.

In some embodiments, the methods involve cytogenetic screening of biological tissue sample from a patient who has been diagnosed with or is suspected of having cancer (*e.g.*, presents with symptoms of cancer) to detect ALK mutations (*e.g.*, those listed in Table 1).

The results of the screening method and the interpretation thereof are predictive of the patient's response to treatment with ALK inhibiting agents (*e.g.*, PF-02341066 and/or PDD). According to the present invention, the presence of an ALK mutation is indicative that treatment with ALK inhibiting agents (*e.g.*, PF-02341066 and/or PDD) will provide enhanced therapeutic benefit against the cancer cells relative to those of patients not having an ALK mutation.

In one embodiment, the methods of the present invention comprise contacting a DNA sample, *e.g.*, a sample containing germline and/or somatic DNA, such as a chromosomal sample, obtained from cells isolated from the patient to polynucleotide probes that are specific for and hybridize under stringent conditions with genomic DNA in chromosomal regions associated with cytogenetic abnormalities (*e.g.*, ALK mutations

described herein) to determine the presence or absence of one or more of the abnormalities (e.g., mutations) in the cells of the patient. The results of the analysis are predictive of the patient's likely response to treatment with therapeutic agents, particularly agents that inhibit ALK (e.g., PF-02341066 and/or PDD).

5 In another embodiment, a time course is measured by determining the time between significant events in the course of a patient's disease, wherein the measurement is predictive of whether a patient has a long time course. In another embodiment, the significant event is the progression from primary diagnosis to death. In another embodiment, the significant event is the progression from primary diagnosis to metastatic disease. In another
10 embodiment, the significant event is the progression from primary diagnosis to relapse. In another embodiment, the significant event is the progression from metastatic disease to death. In another embodiment, the significant event is the progression from metastatic disease to relapse. In another embodiment, the significant event is the progression from relapse to death. In certain embodiments, the time course is measured with respect to
15 overall survival rate, time to progression and/or using the RECIST or other response criteria.

 In certain embodiments, a predetermined measure is created by dividing patient samples into at least two patient subgroups. In certain embodiments, the number of subgroups is two so that the patient sample is divided into a subgroup of patients having an
20 ALK mutation(s) and a subgroup not having an ALK mutation(s). In certain embodiments, the ALK mutation status in the subject is compared to either the subgroup having or not having an ALK mutation(s); if the patient has an ALK mutation(s), then the patient is unlikely to respond to an ALK inhibitor (e.g., PF-02341066 and/or PDD) and/or the patient is likely to have a long time course. In certain embodiments, the number of subgroups is
25 greater than two, including, without limitation, three subgroups, four subgroups, five subgroups and six subgroups, depending on stratification of predicted ALK inhibitor efficacy as correlated with particular ALK mutations. In certain embodiments, likeliness to respond is measured with respect to overall survival rate, time to progression and/or using the RECIST criteria. In certain embodiments, the ALK inhibitor is PF-02341066 and/or
30 PDD.

 In another aspect, the invention is drawn to a method for determining whether a subject with an ALK mutation positive cancer is likely to respond to treatment with an ALK inhibiting agent (e.g., PF-02341066 and/or PDD) and/or the time course of disease is

long. In another aspect, the invention is drawn to a method for predicting a time course of disease in a subject with an ALK mutation positive cancer. In another aspect, the invention is drawn to a method for predicting the probability of a significant event in a subject with an ALK mutation positive cancer.

5 **1. Methods for Detecting ALK Mutations**

Methods of evaluating ALK gene mutations and/or gene products (*e.g.*, the markers set forth in Table 1) are well known to those of skill in the art, including hybridization-based assays. For example, one method for evaluating the copy number of encoding nucleic acid in a sample involves a Southern Blot. In a Southern Blot, the genomic DNA
10 (typically fragmented and separated on an electrophoretic gel) is hybridized to a probe specific for the target region. Comparison of the intensity of the hybridization signal from the probe for the target region with control probe signal from analysis of normal genomic DNA (*e.g.*, a non-amplified portion of the same or related cell, tissue, organ, *etc.*) provides an estimate of the presence/absence and relative copy number of the target nucleic acid.
15 Alternatively, a Northern blot may be utilized for evaluating the copy number of encoding nucleic acid in a sample. In a Northern blot, mRNA is hybridized to a probe specific for the target region. Comparison of the intensity of the hybridization signal from the probe for the target region with control probe signal from analysis of normal mRNA (*e.g.*, a non-amplified portion of the same or related cell, tissue, organ, *etc.*) provides an estimate of the
20 presence/absence and relative copy number of the target nucleic acid.

An alternative means for determining the copy number is *in situ* hybridization (*e.g.*, Angerer (1987) *Meth. Enzymol* 152: 649). Generally, *in situ* hybridization comprises the following steps: (1) fixation of tissue or biological structure to be analyzed; (2) prehybridization treatment of the biological structure to increase accessibility of target
25 DNA, and to reduce nonspecific binding; (3) hybridization of the mixture of nucleic acids to the nucleic acid in the biological structure or tissue; (4) post-hybridization washes to remove nucleic acid fragments not bound in the hybridization and (5) detection of the hybridized nucleic acid fragments. The reagent used in each of these steps and the conditions for use vary depending on the particular application.

30 Exemplary hybridization-based assays include, but are not limited to, traditional "direct probe" methods such as Southern blots or *in situ* hybridization (*e.g.*, FISH and FISH plus SKY), and "comparative probe" methods such as comparative genomic hybridization

(CGH), *e.g.*, cDNA-based or oligonucleotide-based CGH. The methods can be used in a wide variety of formats including, but not limited to, substrate (*e.g.*, membrane or glass) bound methods or array-based approaches.

In one aspect, FISH analysis is used. Cell samples are obtained from patients
5 according to methods well known in the art in order to be tested by an appropriate cytogenetic testing method known in the art, for example, the FISH method. In one embodiment, FISH can be performed according to the Vysis™ system (Abbott Molecular), whose manufacturer's protocols are incorporated herein by reference.

Probes are used that contain DNA segments that are essentially complementary to
10 DNA base sequences existing in different portions of chromosomes. Examples of probes useful according to the invention, and labeling and hybridization of probes to samples are described in two U.S. patents to Vysis, Inc. U.S. Patent Nos. 5,491,224 (incorporated by reference) and 6,277,569 (incorporated by reference) to Bittner, *et al.*

Chromosomal probes are typically about 50 to about 10^5 nucleotides in length.
15 Longer probes typically comprise smaller fragments of about 100 to about 500 nucleotides in length. Probes that hybridize with centromeric DNA and locus-specific DNA are available commercially, for example, from Vysis, Inc. (Downers Grove, Ill.), Molecular Probes, Inc. (Eugene, Oreg.) or from Cytocell (Oxfordshire, UK). Alternatively, probes can be made non-commercially from chromosomal or genomic DNA through standard
20 techniques. For example, sources of DNA that can be used include genomic DNA, cloned DNA sequences, somatic cell hybrids that contain one, or a part of one, chromosome (*e.g.*, human chromosome) along with the normal chromosome complement of the host, and chromosomes purified by flow cytometry or microdissection. The region of interest can be isolated through cloning, or by site-specific amplification via the polymerase chain reaction
25 (PCR). See, for example, Nath and Johnson, *Biotechnic Histochem.*, 1998, 73(1):6-22, Wheelless *et al.*, *Cytometry* 1994, 17:319-326, and U.S. Patent No. 5,491,224 (incorporated by reference).

The probes to be used hybridize to a specific region of a chromosome to determine whether a cytogenetic abnormality is present in this region. One type of cytogenetic
30 abnormality is a deletion. Although deletions can be of one or more entire chromosomes, deletions normally involve loss of part of one or more chromosomes. If the entire region of a chromosome that is contained in a probe is deleted from a cell, hybridization of that probe

to the DNA from the cell will normally not occur and no signal will be present on that chromosome. If the region of a chromosome that is partially contained within a probe is deleted from a cell, hybridization of that probe to the DNA from the cell may still occur, but less of a signal may be present. For example, the loss of a signal is compared to probe
5 hybridization to DNA from control cells that do not contain the genetic abnormalities which the probes are intended to detect. In some embodiments, at least 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, or more cells are enumerated for presence of the cytogenetic abnormality.

Cytogenetic abnormalities to be detected can include, but are not limited to, non-
10 reciprocal translocations, intra-chromosomal inversions, point mutations, deletions, gene copy number changes, gene expression level changes, and germ line mutations. In particular, one type of cytogenetic abnormality is a duplication. Duplications can be of entire chromosomes, or of regions smaller than an entire chromosome. If the region of a chromosome that is contained in a probe is duplicated in a cell, hybridization of that probe
15 to the DNA from the cell will normally produce at least one additional signal as compared to the number of signals present in control cells with no abnormality of the chromosomal region contained in the probe. Although any probes that detect human chromosome 2p23 or ortholog thereof or any chromosomal region comprising a translocation with the ALK gene of 2p23 or ortholog thereof can be used. Suitable probes are well known in the art
20 (e.g., available from Vysis, Inc. (Downers Grove, IL).

Chromosomal probes are labeled so that the chromosomal region to which they hybridize can be detected. Probes typically are directly labeled with a fluorophore, an organic molecule that fluoresces after absorbing light of lower wavelength/higher energy. The fluorophore allows the probe to be visualized without a secondary detection molecule.
25 After covalently attaching a fluorophore to a nucleotide, the nucleotide can be directly incorporated into the probe with standard techniques such as nick translation, random priming, and PCR labeling. Alternatively, deoxycytidine nucleotides within the probe can be transaminated with a linker. The fluorophore then is covalently attached to the transaminated deoxycytidine nucleotides. See, U.S. Patent No. 5,491,224 (incorporated by
30 reference).

U.S. Patent No. 5,491,224 describes probe labeling as a number of the cytosine residues having a fluorescent label covalently bonded thereto. The number of fluorescently labeled cytosine bases is sufficient to generate a detectable fluorescent signal while the

individual so labeled DNA segments essentially retain their specific complementary binding (hybridizing) properties with respect to the chromosome or chromosome region to be detected. Such probes are made by taking the unlabeled DNA probe segment, transaminating with a linking group a number of deoxycytidine nucleotides in the segment,
5 covalently bonding a fluorescent label to at least a portion of the transaminated deoxycytidine bases.

Probes can also be labeled by nick translation, random primer labeling or PCR labeling. Labeling is done using either fluorescent (direct)-or haptene (indirect)-labeled nucleotides. Representative, non-limiting examples of labels include: AMCA-6-dUTP,
10 CascadeBlue-4-dUTP, Fluorescein-12-dUTP, Rhodamine-6-dUTP, TexasRed-6-dUTP, Cy3-6-dUTP, Cy5-dUTP, Biotin(BIO)-11-dUTP, Digoxigenin(DIG)-11-dUTP or Dinitrophenyl (DNP)-11-dUTP.

Probes also can be indirectly labeled with biotin or digoxigenin, or labeled with radioactive isotopes such as ^{32}P and ^3H , although secondary detection molecules or further
15 processing then is required to visualize the probes. For example, a probe labeled with biotin can be detected by avidin conjugated to a detectable marker. For example, avidin can be conjugated to an enzymatic marker such as alkaline phosphatase or horseradish peroxidase. Enzymatic markers can be detected in standard colorimetric reactions using a substrate and/or a catalyst for the enzyme. Catalysts for alkaline phosphatase include 5-
20 bromo-4-chloro-3-indolylphosphate and nitro blue tetrazolium. Diaminobenzoate can be used as a catalyst for horseradish peroxidase.

Probes can also be prepared such that a fluorescent or other label is not part of the DNA before or during the hybridization, and is added after hybridization to detect the probe hybridized to a chromosome. For example, probes can be used that have antigenic
25 molecules incorporated into the DNA. After hybridization, these antigenic molecules are detected using specific antibodies reactive with the antigenic molecules. Such antibodies can themselves incorporate a fluorochrome, or can be detected using a second antibody with a bound fluorochrome.

However treated or modified, the probe DNA is commonly purified in order to
30 remove unreacted, residual products (*e.g.*, fluorochrome molecules not incorporated into the DNA) before use in hybridization.

Prior to hybridization, chromosomal probes are denatured according to methods well known in the art. In general, hybridization steps comprise adding an excess of blocking DNA to the labeled probe composition, contacting the blocked probe composition under hybridizing conditions with the chromosome region to be detected, *e.g.*, on a slide
5 where the DNA has been denatured, washing away unhybridized probe, and detecting the binding of the probe composition to the chromosome or chromosomal region.

Probes are hybridized or annealed to the chromosomal DNA under hybridizing conditions. "Hybridizing conditions" are conditions that facilitate annealing between a probe and target chromosomal DNA. Since annealing of different probes will vary
10 depending on probe length, base concentration and the like, annealing is facilitated by varying probe concentration, hybridization temperature, salt concentration and other factors well known in the art.

Hybridization conditions are facilitated by varying the concentrations, base compositions, complexities, and lengths of the probes, as well as salt concentrations,
15 temperatures, and length of incubation. For example, *in situ* hybridizations are typically performed in hybridization buffer containing 1-2x SSC, 50-65% formamide and blocking DNA to suppress non-specific hybridization. In general, hybridization conditions, as described above, include temperatures of about 25° C to about 55° C, and incubation lengths of about 0.5 hours to about 96 hours.

20 Non-specific binding of chromosomal probes to DNA outside of the target region can be removed by a series of washes. Temperature and concentration of salt in each wash are varied to control stringency of the washes. For example, for high stringency conditions, washes can be carried out at about 65° C to about 80° C, using 0.2x to about 2x SSC, and about 0.1% to about 1% of a non-ionic detergent such as Nonidet P-40 (NP40). Stringency
25 can be lowered by decreasing the temperature of the washes or by increasing the concentration of salt in the washes. In some applications it is necessary to block the hybridization capacity of repetitive sequences. Thus, in some embodiments, tRNA, human genomic DNA, or Cot-I DNA is used to block non-specific hybridization.

After washing, the slide is allowed to drain and air dry, then mounting medium, a
30 counterstain such as DAPI, and a coverslip are applied to the slide. Slides can be viewed immediately or stored at -20° C. before examination.

For fluorescent probes used in fluorescence in situ hybridization (FISH) techniques, fluorescence can be viewed with a fluorescence microscope equipped with an appropriate filter for each fluorophore, or by using dual or triple band-pass filter sets to observe multiple fluorophores. See, for example, U.S. Patent No. 5,776,688 (incorporated by
5 reference). Alternatively, techniques such as flow cytometry can be used to examine the hybridization pattern of the chromosomal probes. FISH can be used to detect chromosome copy number or rearrangement of regions of chromosomes. These probes hybridize, or bind, to the complementary DNA and, because they are labeled with fluorescent tags, allow researchers to see the location of those sequences of DNA using a fluorescence microscope.
10 Unlike most other techniques used to study chromosomes, which require that the cells be actively dividing, FISH can also be performed on non-dividing cells, making it a highly versatile procedure. Therefore, FISH can be performed using interphase cells, or cells in metaphase of the cell division cycle. Many of the techniques involved in FISH analysis are described in U.S. Patent No. 5,447,841 (incorporated by reference) by Gray and Pinkel.

15 FISH results can be interpreted with reference to control cells that are known not to contain the specific cytogenetic abnormality the probe is designed to detect. The FISH hybridization pattern of the probe to DNA from the control cells is compared to hybridization of the same probe to the DNA from cells that are being tested or assayed for the specific cytogenetic abnormality. When a probe is designed to detect a deletion of a
20 chromosome or chromosomal region, there normally is less hybridization of the probe to DNA from the cells being tested than from the control cells. Normally, there is absence of a probe signal in the tested cells, indicative of loss of the region of a chromosome the probe normally hybridizes to. When a probe is designed to detect a chromosomal duplication or addition, there normally is more hybridization of the probe to DNA from the cells being
25 tested than from the control cells. Normally, there is addition of a probe signal in the tested cells, indicative of the presence of an additional chromosomal region to which the probe normally hybridizes.

In CGH methods, a first collection of nucleic acids (*e.g.*, from a sample, *e.g.*, a possible tumor) is labeled with a first label, while a second collection of nucleic acids (*e.g.*,
30 a control, *e.g.*, from a healthy cell/tissue) is labeled with a second label. The ratio of hybridization of the nucleic acids is determined by the ratio of the two (first and second) labels binding to each fiber in the array. Where there are chromosomal deletions or multiplications, differences in the ratio of the signals from the two labels will be detected

and the ratio will provide a measure of the copy number. Array-based CGH may also be performed with single-color labeling (as opposed to labeling the control and the possible tumor sample with two different dyes and mixing them prior to hybridization, which will yield a ratio due to competitive hybridization of probes on the arrays). In single color
5 CGH, the control is labeled and hybridized to one array and absolute signals are read, and the possible tumor sample is labeled and hybridized to a second array (with identical content) and absolute signals are read. Copy number difference is calculated based on absolute signals from the two arrays. Hybridization protocols suitable for use with the methods of the present invention are described, *e.g.*, in Albertson (1984) *EMBO J.* 3: 1227-1234; Pinkel (1988) *Proc. Natl. Acad. Sci. USA* 85: 9138-9142; EPO Pub. No. 430,402; *Methods in Molecular Biology*, Vol. 33: *In situ* Hybridization Protocols, Choo, cd., Humana Press, Totowa, N.J. (1994), *etc.* In one embodiment, the hybridization protocol of Pinkel, *et al.* (1998) *Nature Genetics* 20: 207-211, or of Kallioniemi (1992) *Proc. Natl Acad Sci USA* 89:5321-5325 (1992) is used. Array-based CGH is described in
10 U.S. Patent No. 6,455,258, the contents of each of which are incorporated herein by reference.

In still another embodiment, amplification-based assays can be used to measure presence/absence and copy number. In such amplification-based assays, the nucleic acid sequences act as a template in an amplification reaction (*e.g.*, Polymerase Chain Reaction
20 (PCR). In a quantitative amplification, the amount of amplification product will be proportional to the amount of template in the original sample. Comparison to appropriate controls, *e.g.*, healthy tissue, provides a measure of the copy number.

Methods of "quantitative" amplification are well known to those of skill in the art. For example, quantitative PCR involves simultaneously co-amplifying a known quantity of
25 a control sequence using the same primers. This provides an internal standard that may be used to calibrate the PCR reaction. Detailed protocols for quantitative PCR are provided in Innis, *et al.* (1990) *PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc. N.Y.). Measurement of DNA copy number at microsatellite loci using quantitative PCR analysis is described in Ginzinger, *et al.* (2000) *Cancer Research* 60:5405-5409. The
30 known nucleic acid sequence for the genes is sufficient to enable one of skill in the art to routinely select primers to amplify any portion of the gene. Fluorogenic quantitative PCR may also be used in the methods of the present invention. In fluorogenic quantitative PCR, quantitation is based on amount of fluorescence signals, *e.g.*, TaqMan and sybr green.

Other suitable amplification methods include, but are not limited to, ligase chain reaction (LCR) (see Wu and Wallace (1989) *Genomics* 4: 560, Landegren, *et al.* (1988) *Science* 241:1077, and Barringer *et al.* (1990) *Gene* 89: 117), transcription amplification (Kwoh, *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173), self-sustained sequence
5 replication (Guatelli, *et al.* (1990) *Proc. Nat. Acad. Sci. USA* 87: 1874), dot PCR, and linker adapter PCR, *etc.*

Loss of heterozygosity (LOH) mapping (Wang, Z.C., *et al.* (2004) *Cancer Res* 64(1):64-71; Seymour, A. B., *et al.* (1994) *Cancer Res* 54, 2761-4; Hahn, S. A., *et al.* (1995) *Cancer Res* 55, 4670-5; Kimura, M., *et al.* (1996) *Genes Chromosomes Cancer* 17,
10 88-93) may also be used to identify regions of amplification or deletion.

2. Methods for Assessing Gene Expression

Marker expression level can also be assayed. Expression of a marker of the present invention may be assessed by any of a wide variety of well known methods for detecting expression of a transcribed molecule or protein. Non-limiting examples of such methods
15 include immunological methods for detection of secreted, cell-surface, cytoplasmic, or nuclear proteins, protein purification methods, protein function or activity assays, nucleic acid hybridization methods, nucleic acid reverse transcription methods, and nucleic acid amplification methods.

In certain embodiments, activity of a particular gene is characterized by a measure
20 of gene transcript (*e.g.*, mRNA), by a measure of the quantity of translated protein, or by a measure of gene product activity. Marker expression can be monitored in a variety of ways, including by detecting mRNA levels, protein levels, or protein activity, any of which can be measured using standard techniques. Detection can involve quantification of the level of gene expression (*e.g.*, genomic DNA, cDNA, mRNA, protein, or enzyme activity),
25 or, alternatively, can be a qualitative assessment of the level of gene expression, in particular in comparison with a control level. The type of level being detected will be clear from the context.

Methods of detecting and/or quantifying the gene transcript (mRNA or cDNA made therefrom) using nucleic acid hybridization techniques are known to those of skill in the art
30 (see Sambrook *et al. supra*). For example, one method for evaluating the presence, absence, or quantity of cDNA involves a Southern transfer as described above. Briefly, the mRNA is isolated (*e.g.*, using an acid guanidinium-phenol-chloroform extraction method,

Sambrook *et al. supra.*) and reverse transcribed to produce cDNA. The cDNA is then optionally digested and run on a gel in buffer and transferred to membranes. Hybridization is then carried out using the nucleic acid probes specific for the target cDNA.

5 A general principle of such diagnostic and prognostic assays involves preparing a sample or reaction mixture that may contain a marker, and a probe, under appropriate conditions and for a time sufficient to allow the marker and probe to interact and bind, thus forming a complex that can be removed and/or detected in the reaction mixture. These assays can be conducted in a variety of ways.

10 For example, one method to conduct such an assay would involve anchoring the marker or probe onto a solid phase support, also referred to as a substrate, and detecting target marker/probe complexes anchored on the solid phase at the end of the reaction. In one embodiment of such a method, a sample from a subject, which is to be assayed for presence and/or concentration of marker, can be anchored onto a carrier or solid phase support. In another embodiment, the reverse situation is possible, in which the probe can be
15 anchored to a solid phase and a sample from a subject can be allowed to react as an unanchored component of the assay.

There are many established methods for anchoring assay components to a solid phase. These include, without limitation, marker or probe molecules which are immobilized through conjugation of biotin and streptavidin. Such biotinylated assay
20 components can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). In certain embodiments, the surfaces with immobilized assay components can be prepared in advance and stored.

25 Other suitable carriers or solid phase supports for such assays include any material capable of binding the class of molecule to which the marker or probe belongs. Well-known supports or carriers include, but are not limited to, glass, polystyrene, nylon, polypropylene, polyethylene, dextran, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite.

30 In order to conduct assays with the above-mentioned approaches, the non-immobilized component is added to the solid phase upon which the second component is anchored. After the reaction is complete, uncomplexed components may be removed (*e.g.*,

by washing) under conditions such that any complexes formed will remain immobilized upon the solid phase. The detection of marker/probe complexes anchored to the solid phase can be accomplished in a number of methods outlined herein.

5 In another embodiment, the probe, when it is the unanchored assay component, can be labeled for the purpose of detection and readout of the assay, either directly or indirectly, with detectable labels discussed herein and which are well-known to one skilled in the art.

It is also possible to directly detect marker/probe complex formation without further manipulation or labeling of either component (marker or probe), for example by utilizing the technique of fluorescence energy transfer (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169 (incorporated by reference); Stavrianopoulos, *et al.*, U.S. Patent No. 10 4,868,103 (incorporated by reference)). A fluorophore label on the first, 'donor' molecule is selected such that, upon excitation with incident light of appropriate wavelength, its emitted fluorescent energy will be absorbed by a fluorescent label on a second 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, spatial relationships between the molecules can be assessed. In a situation in 15 which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (*e.g.*, using a fluorimeter).

In another embodiment, determination of the ability of a probe to recognize a 25 marker can be accomplished without labeling either assay component (probe or marker) by utilizing a technology such as real-time Biomolecular Interaction Analysis (BIA) (see, *e.g.*, Sjolander, S. and Urbaniczky, C., 1991, *Anal. Chem.* 63:2338-2345 and Szabo *et al.*, 1995, *Curr. Opin. Struct. Biol.* 5:699-705). As used herein, "BIA" or "surface plasmon resonance" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a 30

detectable signal which can be used as an indication of real-time reactions between biological molecules.

Alternatively, in another embodiment, analogous diagnostic and prognostic assays can be conducted with marker and probe as solutes in a liquid phase. In such an assay, the complexed marker and probe are separated from uncomplexed components by any of a number of standard techniques, including but not limited to: differential centrifugation, chromatography, electrophoresis and immunoprecipitation. In differential centrifugation, marker/probe complexes may be separated from uncomplexed assay components through a series of centrifugal steps, due to the different sedimentation equilibria of complexes based on their different sizes and densities (see, for example, Rivas, G., and Minton, A.P., 1993, *Trends Biochem Sci.* 18(8):284-7). Standard chromatographic techniques may also be utilized to separate complexed molecules from uncomplexed ones. For example, gel filtration chromatography separates molecules based on size, and through the utilization of an appropriate gel filtration resin in a column format, for example, the relatively larger complex may be separated from the relatively smaller uncomplexed components. Similarly, the relatively different charge properties of the marker/probe complex as compared to the uncomplexed components may be exploited to differentiate the complex from uncomplexed components, for example, through the utilization of ion-exchange chromatography resins. Such resins and chromatographic techniques are well known to one skilled in the art (see, e.g., Heegaard, N.H., 1998, *J. Mol. Recognit.* Winter 11(1-6):141-8; Hage, D.S., and Tweed, S.A. *J Chromatogr B Biomed Sci Appl* 1997 Oct 10;699(1-2):499-525). Gel electrophoresis may also be employed to separate complexed assay components from unbound components (see, e.g., Ausubel *et al.*, ed., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1987-1999). In this technique, protein or nucleic acid complexes are separated based on size or charge, for example. In order to maintain the binding interaction during the electrophoretic process, non-denaturing gel matrix materials and conditions in the absence of reducing agent are typical. Appropriate conditions to the particular assay and components thereof will be well known to one skilled in the art.

In a particular embodiment, the level of mRNA corresponding to the marker can be determined both by *in situ* and by *in vitro* formats in a biological sample using methods known in the art. The term "biological sample" is intended to include tissues, cells, biological fluids and isolates thereof, isolated from a subject, as well as tissues, cells and fluids present within a subject. Many expression detection methods use isolated RNA. For

in vitro methods, any RNA isolation technique that does not select against the isolation of mRNA can be utilized for the purification of RNA from cells (see, *e.g.*, Ausubel *et al.*, ed., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York 1987-1999).

5 Additionally, large numbers of tissue samples can readily be processed using techniques well known to those of skill in the art, such as, for example, the single-step RNA isolation process of Chomczynski (1989, U.S. Patent No. 4,843,155 (incorporated by reference)).

The isolated nucleic acid can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One diagnostic method for the detection of mRNA levels
10 involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length cDNA, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding a marker of the present
15 invention. Other suitable probes for use in the diagnostic assays of the present invention are described herein. Hybridization of an mRNA with the probe indicates that the marker in question is being expressed.

In one format, the mRNA is immobilized on a solid surface and contacted with a probe, for example by running the isolated mRNA on an agarose gel and transferring the
20 mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probe(s) are immobilized on a solid surface and the mRNA is contacted with the probe(s), for example, in an Affymetrix gene chip array. A skilled artisan can readily adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the markers of the present invention.

25 The probes can be full length or less than the full length of the nucleic acid sequence encoding the protein. Shorter probes are empirically tested for specificity. Exemplary nucleic acid probes are 20 bases or longer in length (See, *e.g.*, Sambrook *et al.* for methods of selecting nucleic acid probe sequences for use in nucleic acid hybridization). Visualization of the hybridized portions allows the qualitative determination of the presence
30 or absence of cDNA.

An alternative method for determining the level of a transcript corresponding to a marker of the present invention in a sample involves the process of nucleic acid

amplification, *e.g.*, by rtPCR (the experimental embodiment set forth in Mullis, 1987, U.S. Patent No. 4,683,202 (incorporated by reference)), ligase chain reaction (Barany, 1991, *Proc. Natl. Acad. Sci. USA*, 88:189-193), self sustained sequence replication (Guatelli *et al.*, 1990, *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system
5 (Kwoh *et al.*, 1989, *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, 1988, *Bio/Technology* 6:1197), rolling circle replication (Lizardi *et al.*, U.S. Patent No. 5,854,033 (incorporated by reference)) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. Fluorogenic rtPCR may also be used in the methods of the present
10 invention. In fluorogenic rtPCR, quantitation is based on amount of fluorescence signals, *e.g.*, TaqMan and sybr green. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or *vice-versa*)
15 and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

20 For *in situ* methods, mRNA does not need to be isolated from the cells prior to detection. In such methods, a cell or tissue sample is prepared/processed using known histological methods. The sample is then immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the marker.

As an alternative to making determinations based on the absolute expression level of
25 the marker, determinations may be based on the normalized expression level of the marker. Expression levels are normalized by correcting the absolute expression level of a marker by comparing its expression to the expression of a gene that is not a marker, *e.g.*, a housekeeping gene that is constitutively expressed. Suitable genes for normalization include housekeeping genes such as the actin gene, or epithelial cell-specific genes. This
30 normalization allows the comparison of the expression level in one sample, *e.g.*, a subject sample, to another sample, *e.g.*, a non-cancerous sample, or between samples from different sources.

Alternatively, the expression level can be provided as a relative expression level. To determine a relative expression level of a marker, the level of expression of the marker is determined for 10 or more samples of normal versus cancer cell isolates, or even 50 or more samples, prior to the determination of the expression level for the sample in question.

5 The mean expression level of each of the genes assayed in the larger number of samples is determined and this is used as a baseline expression level for the marker. The expression level of the marker determined for the test sample (absolute level of expression) is then divided by the mean expression value obtained for that marker. This provides a relative expression level.

10 In certain embodiments, the samples used in the baseline determination will be from cancer cells or normal cells of the same tissue type. The choice of the cell source is dependent on the use of the relative expression level. Using expression found in normal tissues as a mean expression score aids in validating whether the marker assayed is specific to the tissue from which the cell was derived (*versus* normal cells). In addition, as more
15 data is accumulated, the mean expression value can be revised, providing improved relative expression values based on accumulated data. Expression data from normal cells provides a means for grading the severity of the cancer state.

In another embodiment, expression of a marker is assessed by preparing genomic DNA or mRNA/cDNA (*i.e.*, a transcribed polynucleotide) from cells in a subject sample,
20 and by hybridizing the genomic DNA or mRNA/cDNA with a reference polynucleotide which is a complement of a polynucleotide comprising the marker, and fragments thereof. cDNA can, optionally, be amplified using any of a variety of polymerase chain reaction methods prior to hybridization with the reference polynucleotide. Expression of one or more markers can likewise be detected using quantitative PCR (QPCR) to assess the level
25 of expression of the marker(s). Alternatively, any of the many known methods of detecting mutations or variants (*e.g.*, single nucleotide polymorphisms, deletions) of a marker of the present invention may be used to detect occurrence of a mutated marker in a subject.

In a related embodiment, a mixture of transcribed polynucleotides obtained from the sample is contacted with a substrate having fixed thereto a polynucleotide complementary
30 to or homologous with at least a portion (*e.g.*, at least 7, at least 10, at least 15, at least 20, at least 25, at least 30, at least 40, at least 50, at least 100, at least 500, or more nucleotide residues) of a marker of the present invention. If polynucleotides complementary to or homologous with a marker of the present invention are differentially detectable on the

substrate (*e.g.*, detectable using different chromophores or fluorophores, or fixed to different selected positions), then the levels of expression of a plurality of markers can be assessed simultaneously using a single substrate (*e.g.*, a "gene chip" microarray of polynucleotides fixed at selected positions). When a method of assessing marker
5 expression is used which involves hybridization of one nucleic acid with another, the hybridization may be performed under stringent hybridization conditions.

In another embodiment, a combination of methods to assess the expression of a marker is utilized.

Because the compositions, kits, and methods of the present invention rely on
10 detection of a difference in expression levels or copy number of one or more markers of the present invention, in certain embodiments the level of expression or copy number of the marker is significantly greater than the minimum detection limit of the method used to assess expression or copy number in at least one of normal cells and cancerous cells.

3. Methods for Assessing Expressed Protein

15 The activity or level of a marker protein can also be detected and/or quantified by detecting or quantifying the expressed polypeptide. The polypeptide can be detected and quantified by any of a number of means well known to those of skill in the art. These may include analytic biochemical methods such as electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC),
20 hyperdiffusion chromatography, and the like, or various immunological methods such as fluid or gel precipitin reactions, immunodiffusion (single or double), immunoelectrophoresis, radioimmunoassay (RIA), enzyme-linked immunosorbent assays (ELISAs), immunofluorescent assays, Western blotting, immunohistochemistry and the like. A skilled artisan can readily adapt known protein/antibody detection methods for use
25 in determining whether cells express a marker of the present invention.

Another agent for detecting a polypeptide of the present invention is an antibody capable of binding to a polypeptide corresponding to a marker of the present invention, *e.g.*, an antibody with a detectable label. Antibodies can be polyclonal or monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')₂) can be used. The term "labeled", with
30 regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another

reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin.

In another embodiment, the antibody is labeled, *e.g.*, a radio-labeled, chromophore-labeled, fluorophore-labeled, or enzyme-labeled antibody. In another embodiment, an antibody derivative (*e.g.*, an antibody conjugated with a substrate or with the protein or ligand of a protein-ligand pair {*e.g.*, biotin-streptavidin}), or an antibody fragment (*e.g.*, a single-chain antibody, an isolated antibody hypervariable domain, *etc.*) which binds specifically with a protein corresponding to the marker, such as the protein encoded by the open reading frame corresponding to the marker or such a protein which has undergone all or a portion of its normal post-translational modification, is used.

Immunohistochemistry or IHC refers to the process of localizing antigens (*e.g.* proteins) in cells of a tissue section exploiting the principle of antibodies binding specifically to antigens in biological tissues. Immunohistochemical staining is widely used in the diagnosis of abnormal cells such as those found in cancerous tumors. Specific molecular markers are characteristic of particular cellular events such as proliferation or cell death (apoptosis). IHC is also widely used in research to understand the distribution and localization of biomarkers and differentially expressed proteins in different parts of a biological tissue. Visualizing an antibody-antigen interaction can be accomplished in a number of ways. In the most common instance, an antibody is conjugated to an enzyme, such as peroxidase, that can catalyse a colour-producing reaction. Alternatively, the antibody can also be tagged to a fluorophore, such as fluorescein, rhodamine, DyLight Fluor or Alexa Fluor.

Proteins from cells can be isolated using techniques that are well known to those of skill in the art. The protein isolation methods employed can, for example, be such as those described in Harlow and Lane (Harlow and Lane, 1988, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York).

In one format, antibodies, or antibody fragments, can be used in methods such as Western blots or immunofluorescence techniques to detect the expressed proteins. In such uses, one may immobilize either the antibody or proteins on a solid support. Suitable solid phase supports or carriers include any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene,

dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros, and magnetite.

One skilled in the art will know many other suitable carriers for binding antibody or antigen, and will be able to adapt such support for use with the present invention. For example, protein isolated from cells can be run on a polyacrylamide gel electrophoresis and immobilized onto a solid phase support, such as nitrocellulose. The support can then be washed with suitable buffers followed by treatment with the detectably labeled antibody. The solid phase support can then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on the solid support can then be detected by conventional means. Means of detecting proteins using electrophoretic techniques are well known to those of skill in the art (see generally, R. Scopes (1982) *Protein Purification*, Springer-Verlag, N.Y.; Deutscher, (1990) *Methods in Enzymology* Vol. 182: *Guide to Protein Purification*, Academic Press, Inc., N.Y.).

In another embodiment, Western blot (immunoblot) analysis is used to detect and quantify the presence of a polypeptide in the sample. This technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the antibodies that specifically bind a polypeptide. The anti-polypeptide antibodies specifically bind to the polypeptide on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (*e.g.*, labeled sheep anti-human antibodies) that specifically bind to the anti-polypeptide.

In another embodiment, the polypeptide is detected using an immunoassay. As used herein, an immunoassay is an assay that utilizes an antibody to specifically bind to the analyte. The immunoassay is thus characterized by detection of specific binding of a polypeptide to an anti-antibody as opposed to the use of other physical or chemical properties to isolate, target, and quantify the analyte.

The polypeptide is detected and/or quantified using any of a number of well recognized immunological binding assays (see, *e.g.*, U.S. Patent Nos. 4,366,241 (incorporated by reference); 4,376,110 (incorporated by reference); 4,517,288 (incorporated by reference); and 4,837,168 (incorporated by reference)). For a review of the general immunoassays, see also Asai (1993) *Methods in Cell Biology* Volume 37: *Antibodies in*

Cell Biology, Academic Press, Inc. New York; Stites & Terr (1991) *Basic and Clinical Immunology* 7th Edition.

Immunological binding assays (or immunoassays) typically utilize a "capture agent" to specifically bind to and often immobilize the analyte (polypeptide or subsequence). The capture agent is a moiety that specifically binds to the analyte. In another embodiment, the capture agent is an antibody that specifically binds a polypeptide. The antibody (anti-peptide) may be produced by any of a number of means well known to those of skill in the art.

Immunoassays also often utilize a labeling agent to specifically bind to and label the binding complex formed by the capture agent and the analyte. The labeling agent may itself be one of the moieties comprising the antibody/analyte complex. Thus, the labeling agent may be a labeled polypeptide or a labeled anti-antibody. Alternatively, the labeling agent may be a third moiety, such as another antibody, that specifically binds to the antibody/polypeptide complex.

In one embodiment, the labeling agent is a second human antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be modified with a detectable moiety, *e.g.*, as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G may also be used as the label agent. These proteins are normal constituents of the cell walls of streptococcal bacteria. They exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of species (see, generally Kronval, *et al.* (1973) *J. Immunol.*, 111: 1401-1406, and Akerstrom (1985) *J. Immunol.*, 135: 2589-2542).

As indicated above, immunoassays for the detection and/or quantification of a polypeptide can take a wide variety of formats well known to those of skill in the art.

Exemplary immunoassays for detecting a polypeptide may be competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of captured analyte is directly measured. In one "sandwich" assay, for example, the capture agent (anti-peptide antibodies) can be bound directly to a solid substrate where they are immobilized. These immobilized antibodies then capture polypeptide present in the test

sample. The polypeptide thus immobilized is then bound by a labeling agent, such as a second human antibody bearing a label.

In competitive assays, the amount of analyte (polypeptide) present in the sample is measured indirectly by measuring the amount of an added (exogenous) analyte (polypeptide) displaced (or competed away) from a capture agent (anti-peptide antibody) by the analyte present in the sample. In one competitive assay, a known amount of, in this case, a polypeptide is added to the sample and the sample is then contacted with a capture agent. The amount of polypeptide bound to the antibody is inversely proportional to the concentration of polypeptide present in the sample.

In another embodiment, the antibody is immobilized on a solid substrate. The amount of polypeptide bound to the antibody may be determined either by measuring the amount of polypeptide present in a polypeptide/antibody complex, or alternatively by measuring the amount of remaining uncomplexed polypeptide. The amount of polypeptide may be detected by providing a labeled polypeptide.

The assays described herein are scored (as positive or negative or quantity of polypeptide) according to standard methods well known to those of skill in the art. The particular method of scoring will depend on the assay format and choice of label. For example, a Western Blot assay can be scored by visualizing the colored product produced by the enzymatic label. A clearly visible colored band or spot at the correct molecular weight is scored as a positive result, while the absence of a clearly visible spot or band is scored as a negative. The intensity of the band or spot can provide a quantitative measure of polypeptide.

Antibodies for use in the various immunoassays described herein, can be produced as described herein.

In another embodiment, level (activity) is assayed by measuring the enzymatic activity of the gene product. Methods of assaying the activity of an enzyme are well known to those of skill in the art.

In vivo techniques for detection of a marker protein include introducing into a subject a labeled antibody directed against the protein. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

Certain markers identified by the methods of the present invention may be secreted proteins. It is a simple matter for the skilled artisan to determine whether any particular marker protein is a secreted protein. In order to make this determination, the marker protein is expressed in, for example, a mammalian cell, *e.g.*, a human cell line, extracellular
5 fluid is collected, and the presence or absence of the protein in the extracellular fluid is assessed (*e.g.*, using a labeled antibody which binds specifically with the protein).

The following is an example of a method which can be used to detect secretion of a protein. About 8×10^5 293T cells are incubated at 37°C in wells containing growth medium (Dulbecco's modified Eagle's medium {DMEM} supplemented with 10% fetal
10 bovine serum) under a 5% (v/v) CO₂, 95% air atmosphere to about 60-70% confluence. The cells are then transfected using a standard transfection mixture comprising 2 micrograms of DNA comprising an expression vector encoding the protein and 10 microliters of LipofectAMINE™ (GIBCO/BRL Catalog no. 18342-012) per well. The transfection mixture is maintained for about 5 hours, and then replaced with fresh growth
15 medium and maintained in an air atmosphere. Each well is gently rinsed twice with DMEM which does not contain methionine or cysteine (DMEM-MC; ICN Catalog no. 16-424-54). About 1 milliliter of DMEM-MC and about 50 microcuries of Trans-³⁵S™ reagent (ICN Catalog no. 51006) are added to each well. The wells are maintained under the 5% CO₂ atmosphere described above and incubated at 37°C for a selected period.
20 Following incubation, 150 microliters of conditioned medium is removed and centrifuged to remove floating cells and debris. The presence of the protein in the supernatant is an indication that the protein is secreted.

It will be appreciated that subject samples, *e.g.*, a sample containing sputum,
bronchoalveolar lavage, pleural effusion, tissue, whole blood, serum, plasma, buccal scrape,
25 saliva, cerebrospinal fluid, urine, stool, and bone marrow, may contain cells therein, particularly when the cells are cancerous, and, more particularly, when the cancer is metastasizing, and thus may be used in the methods of the present invention. The cell sample can, of course, be subjected to a variety of well-known post-collection preparative and storage techniques (*e.g.*, nucleic acid and/or protein extraction, fixation, storage,
30 freezing, ultrafiltration, concentration, evaporation, centrifugation, *etc.*) prior to assessing the level of expression of the marker in the sample. Thus, the compositions, kits, and methods of the present invention can be used to detect expression of markers corresponding to proteins having at least one portion which is displayed on the surface of cells which

express it. It is a simple matter for the skilled artisan to determine whether the protein corresponding to any particular marker comprises a cell-surface protein. For example, immunological methods may be used to detect such proteins on whole cells, or well known computer-based sequence analysis methods (*e.g.*, the SIGNALP program; Nielsen *et al.*, 5 1997, *Protein Engineering* 10:1-6) may be used to predict the presence of at least one extracellular domain (*i.e.*, including both secreted proteins and proteins having at least one cell-surface domain). Expression of a marker corresponding to a protein having at least one portion which is displayed on the surface of a cell which expresses it may be detected without necessarily lysing the cell (*e.g.*, using a labeled antibody which binds specifically 10 with a cell-surface domain of the protein).

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid corresponding to a marker of the present invention in a biological sample, *e.g.*, a sample containing sputum, bronchoalveolar lavage, pleural effusion, tissue, whole blood, serum, plasma, buccal scrape, saliva, cerebrospinal fluid, urine, stool, and bone marrow. 15 Such kits can be used to determine if a subject is suffering from or is at increased risk of developing cancer. For example, the kit can comprise a labeled compound or agent capable of detecting a polypeptide or an mRNA encoding a polypeptide corresponding to a marker of the present invention in a biological sample and means for determining the amount of the polypeptide or mRNA in the sample (*e.g.*, an antibody which binds the 20 polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits can also include instructions for interpreting the results obtained using the kit.

For antibody-based kits, the kit can comprise, for example: (1) a first antibody (*e.g.*, attached to a solid support) which binds to a polypeptide corresponding to a marker of the 25 present invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable label.

For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, *e.g.*, a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the present invention or 30 (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the present invention. The kit can also comprise, *e.g.*, a buffering agent, a preservative, or a protein stabilizing agent. The kit can further comprise components necessary for detecting the detectable label (*e.g.*, an enzyme or a substrate). The kit can

also contain a control sample or a series of control samples which can be assayed and compared to the test sample. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

5 **4. Method for Assessing Structural Alterations**

The invention also provides a method for assessing the presence of a structural alteration, *e.g.*, mutation.

Another detection method is allele specific hybridization using probes overlapping the polymorphic site and having about 5, about 10, about 20, about 25, or about 30
10 nucleotides around the polymorphic region. In another embodiment of the present invention, several probes capable of hybridizing specifically to mutations are attached to a solid phase support, *e.g.*, a "chip". Oligonucleotides can be bound to a solid support by a variety of processes, including lithography. For example a chip can hold up to 250,000 oligonucleotides (GeneChip, Affymetrix™). Mutation detection analysis using these chips
15 comprising oligonucleotides, also termed "DNA probe arrays" is described *e.g.*, in Cronin *et al.* (1996) *Human Mutation* 7:244. In one embodiment, a chip comprises all the mutations of at least one polymorphic region of a gene. The solid phase support is then contacted with a test nucleic acid and hybridization to the specific probes is detected. Accordingly, the identity of numerous mutations of one or more genes can be identified in a
20 simple hybridization experiment. For example, the identity of the mutation of the nucleotide polymorphism in the 5' upstream regulatory element can be determined in a single hybridization experiment.

In other detection methods, it is necessary to first amplify at least a portion of a marker prior to identifying the mutation. Amplification can be performed, *e.g.*, by PCR
25 and/or LCR (see Wu and Wallace (1989) *Genomics* 4:560), according to methods known in the art. In one embodiment, genomic DNA of a cell is exposed to two PCR primers and amplification for a number of cycles sufficient to produce the required amount of amplified DNA. In certain embodiments, the primers are located between 150 and 350 base pairs apart.

30 Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177),

Q-Beta Replicase (Lizardi, P.M. *et al.*, (1988) *Bio/Technology* 6:1197), and self-sustained sequence replication (Guatelli *et al.*, (1989) *Proc. Nat. Acad. Sci.* 87:1874), and nucleic acid based sequence amplification (NABSA), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In one embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence at least a portion of a marker and detect mutations by comparing the sequence of the sample sequence with the corresponding reference (control) sequence. Exemplary sequencing reactions include those based on techniques developed by Maxam and Gilbert (*Proc. Natl Acad Sci USA* (1977) 74:560) or Sanger (Sanger *et al.* (1977) *Proc. Nat. Acad. Sci* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures may be utilized when performing the subject assays (*Biotechniques* (1995) 19:448), including sequencing by mass spectrometry (see, for example, U.S. Patent Number 5,547,835 (incorporated by reference) and international patent application Publication Number WO 94/16101 (incorporated by reference), entitled *DNA Sequencing by Mass Spectrometry* by H. Köster; U.S. Patent Number 5,547,835 (incorporated by reference) and international patent application Publication Number WO 94/21822 (incorporated by reference) entitled *DNA Sequencing by Mass Spectrometry Via Exonuclease Degradation* by H. Köster), and U.S. Patent Number 5,605,798 (incorporated by reference) and International Patent Application No. PCT/US96/03651 (incorporated by reference) entitled *DNA Diagnostics Based on Mass Spectrometry* by H. Köster; Cohen *et al.* (1996) *Adv Chromatogr* 36:127-162; and Griffin *et al.* (1993) *Appl Biochem Biotechnol* 38:147-159). It will be evident to one skilled in the art that, for certain embodiments, the occurrence of only one, two or three of the nucleic acid bases need be determined in the sequencing reaction. For instance, A-track or the like, *e.g.*, where only one nucleotide is detected, can be carried out.

Yet other sequencing methods are disclosed, *e.g.*, in U.S. Patent Number 5,580,732 (incorporated by reference) entitled "Method of DNA sequencing employing a mixed DNA-polymer chain probe" and U.S. Patent Number 5,571,676 (incorporated by reference) entitled "Method for mismatch-directed *in vitro* DNA sequencing."

In some cases, the presence of a specific allele of a marker in DNA from a subject can be shown by restriction enzyme analysis. For example, a specific nucleotide

polymorphism can result in a nucleotide sequence comprising a restriction site which is absent from the nucleotide sequence of another mutation.

In a further embodiment, protection from cleavage agents (such as a nuclease, hydroxylamine or osmium tetroxide and with piperidine) can be used to detect mismatched bases in RNA/RNA DNA/DNA, or RNA/DNA heteroduplexes (Myers, *et al.* (1985) *Science* 230:1242). In general, the technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing a control nucleic acid, which is optionally labeled, *e.g.*, RNA or DNA, comprising a nucleotide sequence of a marker mutation with a sample nucleic acid, *e.g.*, RNA or DNA, obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as duplexes formed based on basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digest the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine whether the control and sample nucleic acids have an identical nucleotide sequence or in which nucleotides they are different. See, for example, Cotton *et al* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al* (1992) *Methods Enzymol.* 217:286-295. In another embodiment, the control or sample nucleic acid is labeled for detection.

In another embodiment, an mutation can be identified by denaturing high-performance liquid chromatography (DHPLC) (Oefner and Underhill, (1995) *Am. J. Human Gen.* 57:Suppl. A266). DHPLC uses reverse-phase ion-pairing chromatography to detect the heteroduplexes that are generated during amplification of PCR fragments from individuals who are heterozygous at a particular nucleotide locus within that fragment (Oefner and Underhill (1995) *Am. J. Human Gen.* 57:Suppl. A266). In general, PCR products are produced using PCR primers flanking the DNA of interest. DHPLC analysis is carried out and the resulting chromatograms are analyzed to identify base pair alterations or deletions based on specific chromatographic profiles (see O'Donovan *et al.* (1998) *Genomics* 52:44-49).

In other embodiments, alterations in electrophoretic mobility are used to identify the type of marker mutation. For example, single strand conformation polymorphism

(SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA* 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal Tech Appl* 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids are denatured and
5 allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence and the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In
10 another embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment, the identity of a mutation of a polymorphic region is obtained by analyzing the movement of a nucleic acid comprising the polymorphic region
15 in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in
20 place of a denaturing agent gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:1275).

Examples of techniques for detecting differences of at least one nucleotide between two nucleic acids include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide probes
25 may be prepared in which the known polymorphic nucleotide is placed centrally (allele-specific probes) and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230; and Wallace *et al.* (1979) *Nucl. Acids Res.* 6:3543). Such allele specific oligonucleotide hybridization techniques may be used for the
30 simultaneous detection of several nucleotide changes in different polymorphic regions of marker. For example, oligonucleotides having nucleotide sequences of specific mutations are attached to a hybridizing membrane and this membrane is then hybridized with labeled

sample nucleic acid. Analysis of the hybridization signal will then reveal the identity of the nucleotides of the sample nucleic acid.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides
5 used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238; Newton *et al.* (1989) *Nucl. Acids Res.* 17:2503). This
10 technique is also termed "PROBE" for Probe Oligo Base Extension. In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al* (1992) *Mol. Cell Probes* 6:1).

In another embodiment, identification of the mutation is carried out using an oligonucleotide ligation assay (OLA), as described, *e.g.*, in U.S. Patent Number 4,998,617
15 and in Landegren, U. *et al.*, (1988) *Science* 241:1077-1080. The OLA protocol uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target. One of the oligonucleotides is linked to a separation marker, *e.g.*, biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut,
20 and create a ligation substrate. Ligation then permits the labeled oligonucleotide to be recovered using avidin, or another biotin ligand. Nickerson, D. A. *et al.* have described a nucleic acid detection assay that combines attributes of PCR and OLA (Nickerson, D. A. *et al.*, (1990) *Proc. Natl. Acad. Sci. (U.S.A.)* 87:8923-8927. In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

25 The invention further provides methods for detecting single nucleotide polymorphisms in a marker. Because single nucleotide polymorphisms constitute sites of variation flanked by regions of invariant sequence, their analysis requires no more than the determination of the identity of the single nucleotide present at the site of variation and it is unnecessary to determine a complete gene sequence for each subject. Several methods
30 have been developed to facilitate the analysis of such single nucleotide polymorphisms.

In one embodiment, the single base polymorphism can be detected by using a specialized exonuclease-resistant nucleotide, as disclosed, *e.g.*, in Mundy, C. R. (U.S.

Patent Number 4,656,127 (incorporated by reference)). According to the method, a primer complementary to the allelic sequence immediately 3' to the polymorphic site is permitted to hybridize to a target molecule obtained from a particular animal or human. If the polymorphic site on the target molecule contains a nucleotide that is complementary to the particular exonuclease-resistant nucleotide derivative present, then that derivative will be incorporated onto the end of the hybridized primer. Such incorporation renders the primer resistant to exonuclease, and thereby permits its detection. Since the identity of the exonuclease-resistant derivative of the sample is known, a finding that the primer has become resistant to exonucleases reveals that the nucleotide present in the polymorphic site of the target molecule was complementary to that of the nucleotide derivative used in the reaction. This method has the advantage that it does not require the determination of large amounts of extraneous sequence data.

In another embodiment of the present invention, a solution-based method is used for determining the identity of the nucleotide of a polymorphic site (Cohen, D. *et al.* French Patent 2,650,840; PCT Appln. No. WO91/02087 (incorporated by reference)). As in the Mundy method of U.S. Patent Number 4,656,127 (incorporated by reference), a primer is employed that is complementary to allelic sequences immediately 3' to a polymorphic site. The method determines the identity of the nucleotide of that site using labeled dideoxynucleotide derivatives, which, if complementary to the nucleotide of the polymorphic site will become incorporated onto the terminus of the primer.

An alternative method, known as Genetic Bit Analysis or GBA is described by Goelet, P. *et al.* (PCT Appln. No. 92/15712 (incorporated by reference)). The method of Goelet, P. *et al.* uses mixtures of labeled terminators and a primer that is complementary to the sequence 3' to a polymorphic site. The labeled terminator that is incorporated is thus determined by, and complementary to, the nucleotide present in the polymorphic site of the target molecule being evaluated. In contrast to the method of Cohen *et al.* (French Patent 2,650,840; PCT Appln. No. WO91/02087 (incorporated by reference)) the method of Goelet, P. *et al.* is a heterogeneous phase assay, in which the primer or the target molecule is immobilized to a solid phase.

Several primer-guided nucleotide incorporation procedures for assaying polymorphic sites in DNA have been described (Komher, J. S. *et al.*, (1989) *Nucl. Acids Res.* 17:7779-7784; Sokolov, B. P., (1990) *Nucl. Acids Res.* 18:3671; Syvanen, A. -C., *et al.*, (1990) *Genomics* 8:684-692; Kuppaswamy, M. N. *et al.*, (1991) *Proc. Natl. Acad. Sci.*

(U.S.A.) 88:1143-1147; Prezant, T. R. *et al.*, (1992) *Hum. Mutat.* 1:159-164; Ugozzoli, L. *et al.*, (1992) *GATA* 9:107-112; Nyren, P. (1993) *et al.*, *Anal. Biochem.* 208:171-175). These methods differ from GBA in that they all rely on the incorporation of labeled deoxynucleotides to discriminate between bases at a polymorphic site. In such a format, since the signal is proportional to the number of deoxynucleotides incorporated, polymorphisms that occur in runs of the same nucleotide can result in signals that are proportional to the length of the run (Syvanen, A.C., *et al.*, (1993) *Amer. J. Hum. Genet.* 52:46-59).

For determining the identity of the mutation of a polymorphic region located in the coding region of a marker, yet other methods than those described above can be used. For example, identification of a mutation which encodes a mutated marker can be performed by using an antibody specifically recognizing the mutant protein in, *e.g.*, immunohistochemistry or immunoprecipitation. Antibodies to wild-type markers or mutated forms of markers can be prepared according to methods known in the art.

Alternatively, one can also measure an activity of a marker, such as binding to a marker ligand. Binding assays are known in the art and involve, *e.g.*, obtaining cells from a subject, and performing binding experiments with a labeled ligand, to determine whether binding to the mutated form of the protein differs from binding to the wild-type of the protein.

VI. Exemplary Screening Methods Based on ALK-Inhibition

The present invention further provides methods for identifying substances that inhibit ALK polypeptides (*e.g.*, EML4-ALK polypeptides) to thereby inhibit cancer cell proliferation, growth, differentiation, apoptosis, and/or metastasis. The methods include contacting a test compound with an ALK polypeptide (*e.g.*, polypeptides listed in Table 1). In some embodiments, the ALK polypeptide comprises a variant (*e.g.*, polypeptides listed in Table 1) that increases the risk of partial or nonresponsiveness to inhibition by one or more ALK inhibitors. A compound that is an inhibitor of tumor metastasis may be identified by determining the effect of a test compound on activity of the ALK polypeptide variant (including, for example, ligand binding such as ATP binding and/or tyrosine kinase activity). In a particular example, a test compound that inhibits tyrosine kinase activity as compared to activity in the absence of the test compound identifies the test compound as an

inhibitor of tumor metastasis. If the compound inhibits activity of an ALK variant, it can further be evaluated for its ability to inhibit tumor growth or metastasis.

In particular, activating tyrosine kinase mutants, including the novel biomarkers of the present invention listed in Table 1 (*e.g.*, ALK mutants), are useful to identify
5 compounds that can be used to treat, ameliorate, or prevent neoplasms, for example by inhibiting or preventing cancer cell proliferation, growth, differentiation, apoptosis, and/or metastasis. Screening chemical libraries for molecules which modulate, *e.g.*, inhibit, antagonize, or agonize or mimic, are known in the art. The chemical libraries, for example, can be peptide libraries, peptidomimetic libraries, chemically synthesized libraries,
10 recombinant, *e.g.*, phage display libraries, and in vitro translation-based libraries, other non-peptide synthetic organic libraries.

The screening or creation, identification and selection of appropriate high affinity inhibitors of a novel biomarker of the present invention listed in Table 1 (*e.g.*, ALK mutants) can be accomplished by a variety of methods. Broadly speaking these may
15 include, but are not limited to, two general approaches. One approach is to use structural knowledge about the target protein to design a candidate molecule with which it will precisely interact. An example would be computer assisted molecular design, particularly based on novel structure-function information disclosed herein as Figure 6. A second approach is to use combinatorial or other libraries of molecules, whereby a large library of
20 molecules is screened for affinity with regard to the target enzyme, or ability to inhibit activity of the target enzyme. In a further example, a panel of antibodies may be screened for the ability to inhibit the target enzyme.

Some embodiments provided herein involve determining the ability of a given compound to inhibit a novel biomarker of the present invention listed in Table 1 (*e.g.*, ALK
25 mutants). Test compounds can be assessed for their probable ability to treat neoplastic lesions either directly, or indirectly by comparing their activities against compounds known to be useful for treating neoplasia. For example, the ability of test compounds to inhibit ligand binding such as ATP binding and/or tyrosine kinase activity against novel biomarkers of the present invention listed in Table 1 (*e.g.*, ALK mutants) can be compared
30 to that of known ALK inhibitors such as PF-02341066 and/or PDD. In one embodiment, such test compounds would have at least 100%, at least 99.9%, at least 99.8%, at least 99.7%, at least 99.6%, at least 99.5%, at least 99.4%, at least 99.3%, at least 99.2%, at least 99.1%, at least 99%, at least 98.5%, at least 98%, at least 97.5%, at least 97%, at least

96.5%, at least 96%, at least 95.5%, at least 94%, at least 93.5%, at least 93%, at least 92.5%, at least 92%, at least 91.5%, at least 91%, at least 90.5%, at least 90%, at least 89.5%, at least 89%, at least 88.5%, at least 88%, at least 87.5%, at least 87%, at least 86.5%, at least 86%, at least 85.5%, at least 85%, at least 84.5%, at least 84%, at least 83.5%, at least 83%, at least 82.5%, at least 82%, at least 81.5%, at least 81%, at least 80.5%, at least 80%, at least 79%, at least 78%, at least 77%, at least 76%, at least 75%, at least 74%, at least 73%, at least 72%, at least 71%, at least 70%, at least 69%, at least 68%, at least 67%, at least 66%, at least 65%, at least 64%, at least 63%, at least 62%, at least 61%, at least 60%, at least 59%, at least 58%, at least 57%, at least 56%, at least 55%, at least 54%, at least 53%, at least 52%, at least 51%, at least 50%, or any range in between, of inhibition of a novel biomarker of the present invention listed in Table 1 (*e.g.*, ALK mutants) relative to that of a known ALK inhibitor under the same assay conditions. In certain embodiments, cells can be transfected with a construct encoding a novel biomarker of the present invention listed in Table 1 (*e.g.*, ALK mutants), contacted with a test compound that is tagged or labelled with a detectable marker and analyzed for the presence bound test compound. In certain embodiments, the transfected cells are observed to bind the test compound as compared to cells that have not been transfected with a novel biomarker of the present invention listed in Table 1 (*e.g.*, ALK mutants), which is an indication that the test compound is binding to a novel biomarker of the present invention listed in Table 1 (*e.g.*, ALK mutants) expressed by those cells. The binding of the compound is typically determined by any one of a wide variety of assays known in the art such as ELISA, RIA, and/or BIAcore assays.

Compounds can be screened for inhibitory or other effects on the activity of a novel biomarker of the present invention listed in Table 1 (*e.g.*, ALK mutants) using an expressed recombinant version of the enzyme, or a homolog or ortholog isolated from another species. Alternatively, cells expressing one of these novel biomarker polypeptides can be treated with a test compound and the effect of the test compound on phosphorylation of a specific target can be determined, for instance using one of the techniques described herein. In one example, tyrosine kinase activity is determined. Methods for determining tyrosine kinase phosphorylation influencing activity (*e.g.*, inhibition) are well known to one of skill in the art. In some examples, tyrosine kinase activity may be determined by assessing incorporation of a labeled phosphate (such as ³²P-labeled phosphate) into a substrate which is capable of being phosphorylated by a novel biomarker of the present invention listed in

Table 1 (*e.g.*, ALK mutants) (*e.g.*, a protein or a peptide fragment, especially those of downstream signaling components). In other embodiments, tyrosine kinase activity can be measured using a universal tyrosine kinase activity kit (for example, Universal Tyrosine Kinase Assay Kit (Takara Bio, Inc., Madison, Wis.); Tyrosine Kinase Assay Kit (Millipore, 5 Billerica, Mass.)).

In another embodiment, screening methods are provided that involve further determining whether the compound reduces the growth of tumor cells, for instance, tumor cells known to express an activated tyrosine kinase mutation such as a novel biomarker of the present invention listed in Table 1 (*e.g.*, ALK mutants). Various cell lines can be used, 10 which may be selected based on the tissue to be tested that are well known to a skilled artisan (*e.g.*, BA/F3 cells). For example, many cell lines are well characterized, and are used for instance by the United States National Cancer Institute (NCI) in their screening program for new anti-cancer drugs.

Significant and statistically significant tumor cell growth inhibition, such as occurs 15 for greater than about 50% at a dose of 100 μ M, 90 μ M, 80 μ M, 70 μ M, 60 μ M, 50 μ M, 40 μ M, 30 μ M, 20 μ M, 10 μ M, 9 μ M, 8 μ M, 7 μ M, 6 μ M, 5 μ M, 4.5 μ M, 4 μ M, 3.5 μ M, 3 μ M, 2.5 μ M, 2 μ M, 1.5 μ M, 1 μ M, 900 nM, 850 nM, 800 nM, 750 nM, 700 nM, 650 nM, 600 nM, 550 nM, 500 nM, 450 nM, 400 nM, 350 nM, 300 nM, 250 nM, 200 nM, 150 nM, 100 nM, 95 nM, 90 nM, 85 nM, 80 nM, 75 nM, 70 nM, 65 nM, 60 nM, 55 nM, 50 nM, 45 nM, 20 40 nM, 35 nM, 30 nM, 25 nM, 20 nM, 15 nM, 10 nM, 5 nM, 4 nM, 3 nM, 2 nM, 1 nM or below, is further indicative that the compound is useful for treating neoplastic lesions. An IC_{50} value may be determined and used for comparative purposes. This value is the concentration of drug needed to inhibit tumor cell growth by 50% relative to the control.

These values can further be applied to other criteria. For example, in other 25 embodiments, screening methods provided herein further involve determining whether the test compound induces apoptosis in cultures of tumor cells. Two distinct forms of cell death may be described by morphological and biochemical criteria: necrosis and apoptosis. Necrosis is accompanied by increased permeability of the plasma membrane, whereby the cells swell and the plasma membrane ruptures within minutes. Apoptosis is characterized 30 by membrane blebbing, condensation of cytoplasm, and the activation of endogenous endonucleases.

Apoptosis occurs naturally during normal tissue turnover and during embryonic development of organs and limbs. Apoptosis also can be induced by various stimuli, including cytotoxic T-lymphocytes and natural killer cells, by ionizing radiation and by certain chemotherapeutic drugs. Inappropriate regulation of apoptosis is thought to play an important role in many pathological conditions including cancer, AIDS, or Alzheimer's disease, etc.

Test compounds can be screened for induction of apoptosis using cultures of tumor cells maintained under conditions as described above. In some examples of such screening methods, treatment of cells with test compounds involves either pre- or post-confluent cultures and treatment for one to seven days at various concentrations of the test compounds. Apoptotic cells can be measured in both the attached and "floating" portions of the cultures. Both are collected by removing the supernatant, trypsinizing the attached cells, and combining both preparations following a centrifugation wash step (10 minutes, 2000 rpm). Following treatment with a test compound, cultures can be assayed for apoptosis and necrosis, for instance by florescent microscopy following labeling with acridine orange and ethidium bromide. Many methods for measuring apoptotic cells are known to those of ordinary skill in the art; for instance, one method for measuring apoptotic cell number has been described by Duke & Cohen (Curr. Prot. Immuno., Coligan *et al.*, eds., 3.17.1-3.17.1, 1992). For example, floating and attached cells are collected by trypsinization and washed three times in PBS. Aliquots of cells are then centrifuged. The pellet is resuspended in media and a dye mixture containing acridine orange and ethidium bromide prepared in PBS and mixed gently. The mixture then can be placed on a microscope slide and examined for morphological features of apoptosis. Apoptosis also can be quantified by measuring an increase in DNA fragmentation in cells that have been treated with test compounds. Commercial photometric enzyme immunoassays (EIA) for the quantitative *in vitro* determination of cytoplasmic histone-associated-DNA-fragments (mono- and oligo-nucleosomes) are available (*e.g.*, Cell Death Detection ELISA, Boehringer Mannheim).

In additional embodiments, screening methods provided herein further include determining whether the test compound decreases tumor metastasis, for example in an animal model of metastasis. Methods for assessing tumor metastasis are known to one of skill in the art (see *e.g.* Khanna and Hunter, Carcinogenesis 26:513-523, 2005). One model of metastasis involves human-mouse xenografts, in which human cancer cell lines or tissues

are transplanted into immunocompromised mice (such as SCID mice or nude mice). In similar methods, a cell line that has been engineered to express a novel biomarker of the present invention listed in Table 1 (*e.g.*, ALK mutants) can be transplanted into an immunocompromised mouse. In one example, tumor cells or cell lines are injected directly
5 into the systemic circulation. The site of injection largely defines the site to which metastases develop in these experimental systems. The most common site of tumor cell injection employed for experimental metastasis models is the lateral tail vein in mice, which results primarily in pulmonary metastases. In contrast, intrasplenic or portal vein injection of tumor cells is the most common site employed for developing metastasis in the liver and
10 intracardiac injection of cells may result in metastases to several sites, including bone. Following injection of tumor cells or other cell lines into the circulation, development of metastases at the site of interest (such as lung) is monitored over a period of days or weeks.

Another model for assessing tumor metastasis utilizes orthotopic transplantation, wherein cancer cells are transplanted to the anatomic location or tissue from which a tumor
15 was derived (for example by direct injection or surgical implantation of tumor fragments). Spontaneous metastases that arise from the orthotopic tumor can be assessed over a period of days or weeks. The ability of a test compound to decrease or prevent tumor metastasis may be assessed by administering a test compound to an animal following injection of tumor cells subcutaneously, intramuscularly, or into the circulation or by orthotopic
20 transplantation. The number, size, or time of development of metastases may be assessed. A compound that inhibits tumor metastasis may decrease the number of metastases, for example by at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or even 100% as compared to a control sample. A compound that
25 inhibits tumor metastasis may also decrease the size of metastases as compared to a control sample. Similarly, a compound that inhibits tumor metastasis may delay the onset of development of metastases, for example by at least one week, two weeks, one month, six months, one year, or even indefinitely.

VII. Exemplary ALK Inhibitors

30 The methods disclosed herein include identifying a subject as a candidate for treatment with an inhibitor of a novel biomarker of the present invention listed in Table 1 (*e.g.*, ALK mutants) to induce tumor cell death, reduce tumor growth, or decrease risk of tumor metastasis. Inhibitors of ALK polypeptides are known to one of skill in the art. For

example, PF-02341066, PDD, 2-methyl-11-(2-methylpropyl)-4-oxo-4,5,6,11,12,13-hexahydro-2*H*-indazolo[5,4-*a*]pyrrolo[3,4-*c*]carbazol-8-yl [4-(dimethylamino)benzyl]carbamate, (1*S*,2*S*,3*R*,4*R*)-3-({5-chloro-2-[(1-ethyl-2,3,4,5-tetrahydro-6-methoxy-2-oxo-1*H*-1-benzazepin-7-yl)amino]-4-pyrimidinyl}amino)bicyclo[2.2.1]hept-5-ene-2-carboxamide, and NVP-TAE684 9 see, for example, *PNAS* 104:270-275, 2007; Choi, Y.L. *et al.* (2008) *Cancer Res.* 68:4971-2976; and *Biochemistry* 48:3600-3609, 2009, which are hereby incorporated by reference).

Incorporation by Reference

All publications, patents, and patent applications mentioned herein are hereby incorporated by reference in their entirety as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.

Also incorporated by reference in their entirety are any polynucleotide and polypeptide sequences which reference an accession number correlating to an entry in a public database, such as those maintained by The Institute for Genomic Research (TIGR) on the world wide web at tigr.org and/or the National Center for Biotechnology Information (NCBI) on the world wide web at ncbi.nlm.nih.gov.

EXEMPLIFICATION

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, figures, sequence listing, patents and published patent applications cited throughout this application are hereby incorporated by reference.

Example 1 - Materials and Methods for Examples 2-4

a. DNA sequencing

Oligo(dT)-primed cDNAs were generated from specimen RNAs extracted with the use of the EZ1 system (Qiagen, Valencia, CA) and were subjected to the polymerase chain reaction (PCR) of 30 cycles (consisting of 98°C for 10 s and 68°C for 1 min) with PrimeSTAR HS DNA polymerase (Takara Bio Inc., Shiga, Japan) and the primers ALK-

TK-F (5'-TACAACCCCAACTACTGCTTTGCT-3') and ALK-TK-R1 (5'-AGGCACTTTCTCTTCCTCTTCCAC-3'). The PCR products corresponding to the kinase domain of ALK were then fragmented and sequenced with an Illumina Genome Analyzer II (GAII) for 76 bases from both ends by the paired-end sequencing system (Illumina, San Diego, CA). Raw read data were quality-filtered on the basis of the presence of the PCR primer sequences and a Q value of ≥ 20 for all bases. The filter-passed reads were then aligned to the ALK cDNA sequence with the use of the Bowtie algorithm (available on the world wide web at bowtie-bio.sourceforge.net/index.shtml).

For capillary sequencing with a 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA), PCR products were prepared from cDNAs with the same primer set or with the combination of the EA-F-g-S (5'-CCACACCTGGGAAAGGACCTAAAG-3') and ALK-TK-R2 (5'-CCTCCAAATACTGACAGCCACAGG-3') primers.

b. Mutant EML4-ALK

A cDNA encoding FLAG epitope-tagged EML4-ALK variant 1 (Soda, M. *et al.* (2007) *Nature* 448:561-566) was inserted into the pMX-iresCD8 retroviral vector (Yamashita Y. *et al.* (2001) *J. Biol. Chem.* 276:39012-39020) for simultaneous expression of FLAG-tagged EML4-ALK and mouse CD8. Nucleotide changes corresponding to the C1156Y and L1196M mutations of ALK were introduced into the plasmid individually or in combination for expression of EML4-ALK(C1156Y), EML4-ALK(L1196M), or EML4-ALK(C1156Y/L1196M). Recombinant retroviruses based on these plasmids were generated with the use of the packaging cell line, BOSC23 (Pcar, W.S. *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:8392-8396), and were used to infect the mouse interleukin-3-dependent cell line BA/F3 (Palacios, R. *et al.* (1985) *Cell* 41:727-734). The resulting CD8-positive cells were purified with the use of a miniMACS cell separation column and magnetic beads conjugated with antibodies to CD8 (both from Miltenyi Biotec, Gladbach, Germany). PF-02341066 was obtained from Schleck.

For examination of the tyrosine phosphorylation of EML4-ALK, BA/F3 cells expressing the fusion protein were exposed to ALK inhibitors for 15 h, after which EML4-ALK was immunoprecipitated from cell lysates with antibodies to FLAG (Sigma-Aldrich, St. Louis, MO) and subjected to immunoblot analysis with antibodies to Tyr¹⁶⁰⁴-phosphorylated ALK (Cell Signaling Technology, Danvers, MA). An *in vitro* kinase assay was performed at room temperature for 30 min as described previously (Donella-Deana, A.

et al. (2005) *Biochemistry* 44:8533-8542) with the synthetic YFF peptide (Operon Biotechnologies, Huntsville, AL).

Example 2 - Novel ALK mutations associated with resistance to ALK tyrosine kinase inhibitors

The patient was a 28-year-old man without a history of smoking, and was diagnosed with lung adenocarcinoma at a clinical stage of T4N3M1 in April 2008. Given that the tumor did not harbor any EGFR mutations, the patient was treated by conventional chemotherapy, which resulted in disease progression with the formation of multiple metastases in the brain and bone. In November 2008, the presence of mRNA for EML4-ALK variant 1 in the tumor was confirmed by reverse transcription-PCR analysis of sputum as well as by fluorescence *in situ* hybridization analysis of a biopsy specimen. The patient was thus enrolled in a trial of PF-02341066 and experienced a marked improvement in his performance status (reduction from level 4 to 2). Although he showed a “partial response” to the treatment, his pleural effusion was not totally eradicated. After 5 months of treatment, however, the tumor abruptly started to grow again, resulting in an increase in pleural effusion and in the formation of multiple cancer nodules in both lungs. The patient was dropped from the trial in May 2009, and pleural effusion was then obtained for molecular analysis.

Given that the tumor resumed growth despite sustained administration of the ALK inhibitor, it was determined whether the tumor acquired secondary genetic changes conferring resistance to the drug. Furthermore, given that resistance to TKIs often results from acquired mutations within the target kinases, the possibility that EML4-ALK itself had undergone amino acid changes was examined.

Sputum (ID J-#1) and pleural effusion (ID J-#113) specimens were available for molecular analysis of the patient’s tumor before and after treatment, respectively. Given that the proportion of tumor cells in the two specimens may have differed, a next-generation sequencer was used to perform deep sequencing of EML4-ALK cDNAs derived from these specimens. The cDNAs corresponding to the tyrosine kinase domain of ALK were thus amplified from both specimens (Figure 1A), fragmented, and subjected to nucleotide sequencing with the GAII system. For comparison, the EML4-ALK-positive NSCLC cell line, H2228, and three other clinical specimens also positive for the fusion protein were similarly analyzed. A known single nucleotide polymorphism, rs3795850, was detected in

the cDNAs from the four specimens (Figure 1B). In addition, a T→C change at a position corresponding to nucleotide 4230 of human wild-type ALK cDNA (GenBank accession number, NM_004304) was detected at a low frequency (8.9%) in the J-#1 cDNAs.

Furthermore, two novel alterations, G→A and C→A changes at positions corresponding to nucleotides 4374 and 4493 of wild-type ALK cDNA, were detected at frequencies of 41.8 and 14.0%, respectively, in the J-#113 cDNAs. There were no other recurrent alterations (present in ≥5% of reads) in the kinase-domain cDNAs derived from any of the specimens.

These nucleotide changes were subsequently confirmed using a Sanger sequencer. To exclude the possibility that the mutations had occurred in endogenous wild-type ALK rather than in EML4-ALK, PCR was also performed with a forward primer targeted to EML4 cDNA so that only the fusion cDNA would be amplified (Figure 1A). The T4230C change was not detected among hundreds of fusion cDNAs derived from J-#1, indicating that it was an artifact that arose in the initial PCR or the GAII sequencing step.

However, both the G4374A and C4493A changes were readily confirmed by Sanger sequencing. Among 73 fusion cDNA clones sequenced for J-#113, 34 clones (46.6%) were positive for G4374A, 11 (15.1%) were positive for C4493A, and the remainder (38.4%) were wild type (Figure 1C). Whereas the PCR analysis covered both nucleotide positions in the same products, none of the products contained both mutations, indicating that each mutation occurred independently. Genomic fragments encompassing the G4374 or C4493 positions were also amplified by PCR and subjected to nucleotide sequencing, resulting in confirmation of each change in the tumor genome (Figure 2).

The G4374A and C4493A substitutions result in Cys→Tyr and Leu→Met changes at the positions corresponding to amino acids 1156 and 1196, respectively, of wild-type human ALK.

Example 3 - Novel ALK mutations confer resistance to ALK tyrosine kinase inhibitors

It was next examined whether such amino acid changes affect the sensitivity of EML4-ALK to ALK inhibitors. Wild-type EML4-ALK, the single mutants EML4-ALK(C1156Y) and EML4-ALK(L1196M), and the double mutant EML4-ALK(C1156Y/L1196M) were expressed individually in BA/F3 cells, and the cells were then exposed to ALK inhibitors. PF-02341066 inhibited in a concentration-dependent manner the growth of BA/F3 cells expressing wild-type EML4-ALK (Figure 4A). In

contrast, cells expressing either C1156Y or L1196M mutants manifested a markedly reduced sensitivity to this drug, with repeated experiments showing that BA/F3 cells expressing EML4-ALK(L1196M) were more resistant to PF-02341066 than were those expressing EML4-ALK(C1156Y) (Figure 3). The presence of both mutations did not result in an additive effect on the resistance of cells to PF-02341066. These data thus showed that C1156Y and L1196M mutations each confer resistance to this drug.

Tyrosine phosphorylation of EML4-ALK was examined by immunoblot analysis with antibodies specific for ALK phosphorylated at Tyr1604. Although exposure of BA/F3 cells to PF-02341066 markedly inhibited the tyrosine phosphorylation of wild-type EML4-ALK, it had no substantial effect on that of EML4-ALK(C1156Y) or EML4-ALK(L1196M) (Figure 4B). Consistent with these findings, an in vitro kinase assay revealed that the C1156Y and L1196M mutants of EML4-ALK were less sensitive to inhibition of enzymatic activity by PF-02341066 than was the wild-type protein (Figure 4C). As was the case for inhibition of cell growth (Figure 4A), the L1196M mutant was more refractory to inhibition of kinase activity by PF-02341066 than was the C1156Y mutant (Figure 4C).

Example 4 - Structure-function relationships between novel ALK mutations and resistance to ALK tyrosine kinase inhibitors

Figure 5 shows the positions of Cys1156 and Leu1196 in a three-dimensional structural model of the kinase domain of ALK based on the crystal structure of a related kinase, the insulin receptor. The former residue is positioned adjacent to the amino-terminus of the predicted helix α C as well as close to the upper lid of the ATP-binding pocket. No activating mutations have been reported at this position in other tyrosine kinases. Leu1196 of ALK corresponds to Thr315 of ABL1 and Thr790 of EGFR, each of which is the site of the most frequent acquired mutations that confer resistance to TKIs in these kinases (Deininger, M. *et al.* (2005) Blood 105:2640-2653; Linardou, H. *et al.* (2009) Nat. Rev. Clin. Oncol. 6:352-366). This "gatekeeper" site is located at the bottom surface of the ATP-binding pocket (Figure 5), and the presence of an amino acid with a bulky side chain at this position is known to interfere with the binding of many TKIs (Shah, N.P. *et al.* (2002) Cancer Cell 2:117-125; Tsao, M.S. *et al.* (2005) N. Engl. J. Med. 353:133-144).

Thus, two *de novo* mutations within the kinase domain of EML4-ALK that confer resistance to multiple ALK inhibitors were identified. Given that no EML4-ALK cDNAs were observed harboring both mutations, it is believed that each mutation developed

independently in distinct subclones of the tumor.

Without being bound by theory, given that cDNAs prepared from sputum of the patient before treatment did not contain nucleotide changes corresponding to the C1156Y or L1196M mutations, it is likely that the tumor subclones acquired the mutations *de novo* during treatment with PF-02341066. However, because pleural effusion could not be examined before treatment, the possibility that tumor cells harboring the C1156Y or L1196M mutants were already present in pleural effusion on initial admission of the patient cannot be completely excluded. If this were the case, the tumor might have acquired other, as yet unknown mutations during the 5-month period of treatment with PF-02341066 that allowed its subsequent rapid growth. However, the subclones of tumor cells with the C1156Y or L1196M mutations should have been refractory to the initial treatment and should have expanded during the treatment course. On the contrary, there were no signs of tumor expansion in the patient for at least 5 months, indicating that the C1156Y and L1196M mutations developed during treatment with PF-02341066. This notion is further supported by the fact that the T790M mutation of EGFR that confers resistance to gefitinib or erlotinib is frequently detected in patients previously treated with TKIs but rarely found in untreated individuals (Pao, W. *et al.* (2005) PLoS Med. 2:e73).

Amino acid substitutions at the gatekeeper position of several tyrosine kinases have been detected in tumors treated with TKIs (Kobayashi, S. *et al.* (2005) N. Engl. J. Med. 352:786-792; Pao, W. *et al.* (2005) PLoS Med. 2:e73; Shah, N.P. *et al.* (2002) Cancer Cell 2:117-125; Cools, J. *et al.* (2003) N. Engl. J. Med. 348:1201-1214; Tamborini, E. *et al.* (2004) Gastroenterology 127:294-299). Whereas no mutations at this site have previously been reported for EML4-ALK or ALK, the effects of various artificial amino acid substitutions at the gatekeeper position of NPM-ALK, another fusion-type oncokinase for ALK, were recently examined (Lu, L. *et al.* (2009) Biochemistry 48:3600-3609). Consistent with the present analysis of tumor cells *in vivo*, introduction of Met at this position was found to render NPM-ALK most resistant to multiple ALK inhibitors.

In contrast to gatekeeper substitutions, activating mutations at the position immediately amino-terminal to the α C helix (Cys1156 in ALK) have not been reported for other tyrosine kinases. Whereas a Thr→Ile change at the corresponding position of EGFR was described in one NSCLC case, its relevance to drug sensitivity was not examined (Tsao, M.S. *et al.* (2005) N. Engl. J. Med. 353:133-144). The importance of helix α C for allosteric regulation of enzymatic activity has been demonstrated for serine-threonine

kinases (Hindie, V. *et al.* (2009) Nat. Chem. Biol. 5:758-764). A change at Cys1156 of ALK might therefore interfere allosterically with TKI binding, or Cys1156 might be directly involved in the physical interaction between the kinase domain and TKIs.

5

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the present invention described herein. Such equivalents are intended to be encompassed by the

10 following claims.

CLAIMS

1. A method for identifying a subject having cancer or at risk for developing cancer as having an increased risk of unresponsiveness to treatment with an ALK inhibitor, comprising:

analyzing a sample obtained from the subject to detect the presence of one or more mutant ALK polynucleotide molecules encoding a mutant ALK polypeptide comprising a mutation at position 1156 or to detect the presence of one or more mutant ALK polypeptides comprising a mutation at position 1156,

wherein the presence of the one or more mutant ALK polynucleotide molecules or one or more mutant ALK polypeptides indicates that the subject has an increased risk of unresponsiveness to treatment with the ALK inhibitor.
2. The method of claim 1, wherein the subject has not previously been treated with an ALK inhibitor, or has been previously treated with an ALK inhibitor and has developed at least partial resistance to the ALK inhibitor.
3. The method of claim 1, wherein the cancer is selected from the group consisting of anaplastic large cell lymphoma, neuroblastoma, breast cancer, colorectal cancer, inflammatory myofibroblastic tumors, and non-small cell lung cancers.
4. The method of claim 1 or 2, wherein the ALK inhibitor is selected from the group consisting of PF-02341066, PDD, 2-methyl-11-(2-methylpropyl)-4-oxo-4,5,6,11,12,13-hexahydro-2H-indazolo[5,4-a]pyrrolo[3,4-c]carbazol-8-yl [4-(dimethylamino)benzyl]carbamate, (1S,2S,3R,4R)-3-({5-chloro-2-[(1-ethyl-2,3,4,5-tetrahydro-6-methoxy-2-oxo-1H-1-benzazepin-7-yl)amino]-4-pyrimidinyl}amino)bicyclo[2.2.1]hept-5-ene-2-carboxamide, and NVP-TAE684.
5. The method of claim 1, wherein the sample is selected from the group consisting of sputum, bronchoalveolar lavage, pleural effusion, tissue, whole blood, serum, plasma, buccal scrape, saliva, cerebrospinal fluid, urine, stool, circulating tumor cells, circulating nucleic acids, and bone marrow.
6. The method of claim 5, wherein the sample is tissue; and the tissue is a tumor or cancer tissue.
7. The method of claim 1, wherein the sample comprises cells.

8. The method of claim 1, wherein the one or more mutant ALK polynucleotide molecules encoding a mutant ALK protein comprising a mutation at position 1196 or the one or more mutant ALK polypeptides comprising a mutation at position 1196 are a mutant ALK polynucleotide molecule encoding a mutant ALK protein comprising a Cys1156Tyr mutation or a mutant ALK polypeptide comprising a Cys1156Tyr mutation.

9. The method of claim 1, wherein the one or more ALK mutations are assessed by a nucleic acid hybridization assay.

10. The method of claim 1, wherein the one or more ALK mutations are assessed by polymerase chain reaction.

11. The method of claim 1, wherein the presence of the one or more ALK polypeptides is detected using a reagent which specifically binds to one or more ALK polypeptides.

12. The method of claim 11, wherein the reagent is selected from the group consisting of an antibody, an antibody derivative, and an antibody fragment.

13. The method of claim 1, wherein the one or more ALK mutations are assessed at a first point in time and at least one subsequent point in time.

14. The method of claim 1, wherein the sample comprises germline or somatic genomic DNA.

15. A method of treating a patient having cancer, or at risk for developing cancer, comprising:

analyzing a sample obtained from the patient to detect the presence of one or more mutant ALK polynucleotide molecules encoding a mutant ALK polypeptide comprising a mutation at position 1156 or to detect the presence of one or more mutant ALK polypeptides comprising a mutation at position 1156; and

administering to said patient a therapeutically effective amount of an ALK inhibitor.

16. The method of claim 15, wherein the ALK inhibitor is selected from the group consisting of PF-02341066, PDD, 2-methyl-11-(2-methylpropyl)-4-oxo-4,5,6,11,12,13-hexahydro-2H-indazolo[5,4-a]pyrrolo[3,4-c]carbazol-8-yl [4-(dimethylamino)benzyl]carbamate, (1S,2S,3R,4R)-3-({5-chloro-2-[(1-ethyl-2,3,4,5-tetrahydro-6-methoxy-2-

oxo-1H-1-benzazepin-7-yl)amino]-4-pyrimidinyl)amino)bicyclo[2.2.1]hept-5-ene-2-carboxamide, and NVP-TAE684.

17. The method of claim 15, wherein the ALK inhibitor is PF-02341066.
18. The method of claim 15, wherein the subject has not previously been treated with an ALK inhibitor, or has been previously treated with an ALK inhibitor and has developed at least partial resistance to the ALK inhibitor.
19. A kit for determining the chemosensitivity of a cancer patient to treatment with an ALK inhibitor, the kit comprising:
 - a reagent that specifically binds to one or more mutant ALK polypeptides comprising a mutation at position 1156, wherein the reagent comprises an antibody, an antibody derivative, or an antibody fragment that specifically binds to a mutant ALK polypeptide comprising a mutation at position 1156; and
 - instructions for use.
20. The kit of claim 19, further comprising an ALK inhibitor.
21. The kit of claim 19 or 20, further comprising one or more polynucleotide probes, each of which comprises a polynucleotide sequence which is complementary to a nucleotide encoding a mutant ALK polypeptide comprising a mutation at position 1156.
22. The kit of claim 21, wherein the probes comprise polynucleotides from about 50 to 10^7 nucleotides in length.
23. The kit of claim 21, wherein the probes are selected from the group consisting of oligonucleotides, cDNA molecules, RNA molecules, and synthetic gene probes comprising nucleobases.
24. A method of determining whether a test compound modulates activity of one or more mutant ALK polypeptides comprising:
 - (a) contacting mammalian cells transfected with a construct encoding the one or more mutant ALK polypeptides comprising a mutation at position 1156 with the test compound; and

(b) assessing the mammalian cells for activity of the one or more mutant ALK polypeptides, wherein significantly modulated activity in the presence of the test compound relative to a control experiment identifies the test compound as a modulator of the one or more mutant ALK polypeptides.

25. The method of claim 24, wherein the one or more mutant ALK polypeptides are mutant ALK polypeptides comprising a Cys1156Tyr mutation.

26. The method of claim 24 or 25, wherein the control comprises mammalian cells expressing a wild type ALK polypeptide.

27. The method of claim 26, wherein activity of the one or more mutant ALK polypeptides is selected from the group consisting of: ATP binding, tyrosine kinase activity, cancer cell proliferation, tumor growth, tumor number, apoptosis, and tumor metastasis.

28. The method of claim 24 or 25, wherein the control experiment comprises mammalian cells expressing the one or more mutant ALK polypeptides in the absence of the test compound.

29. The method of claim 28, wherein activity of the one or more mutant ALK polypeptides is selected from the group consisting of: ATP binding, tyrosine kinase activity, cancer cell proliferation, tumor growth, tumor number, apoptosis, and tumor metastasis.

Jichi Medical University
Patent Attorneys for the Applicant/Nominated Person
SPRUSON & FERGUSON

Figure 1

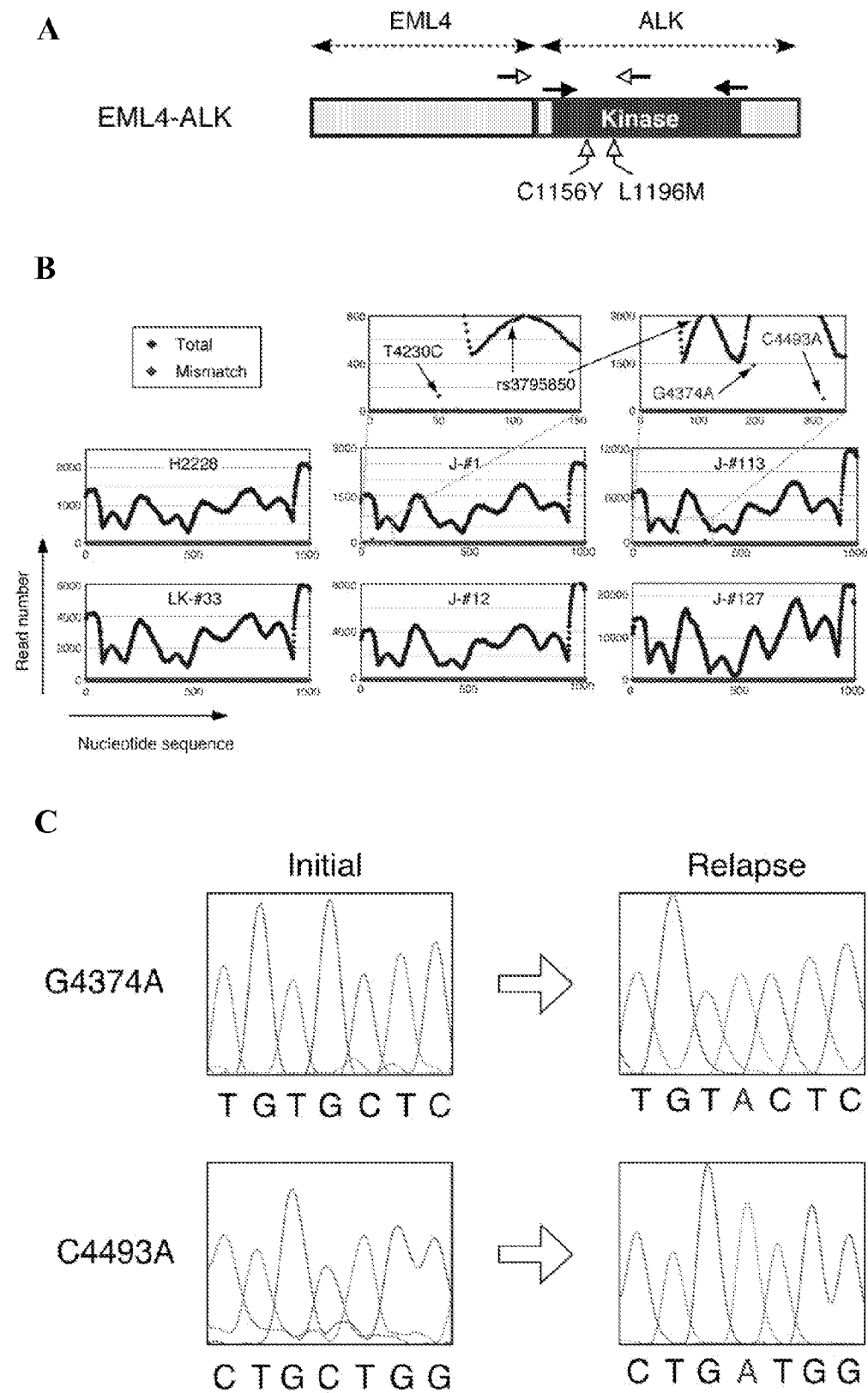


Figure 2

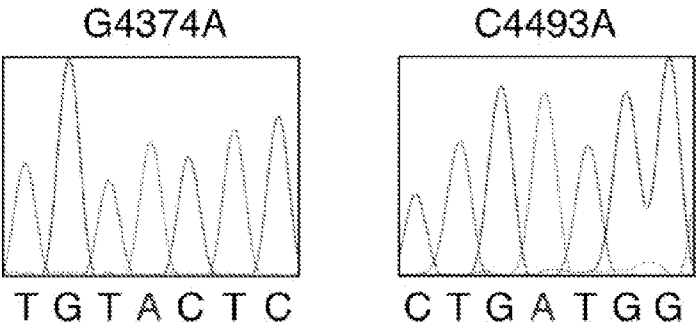


Figure 3

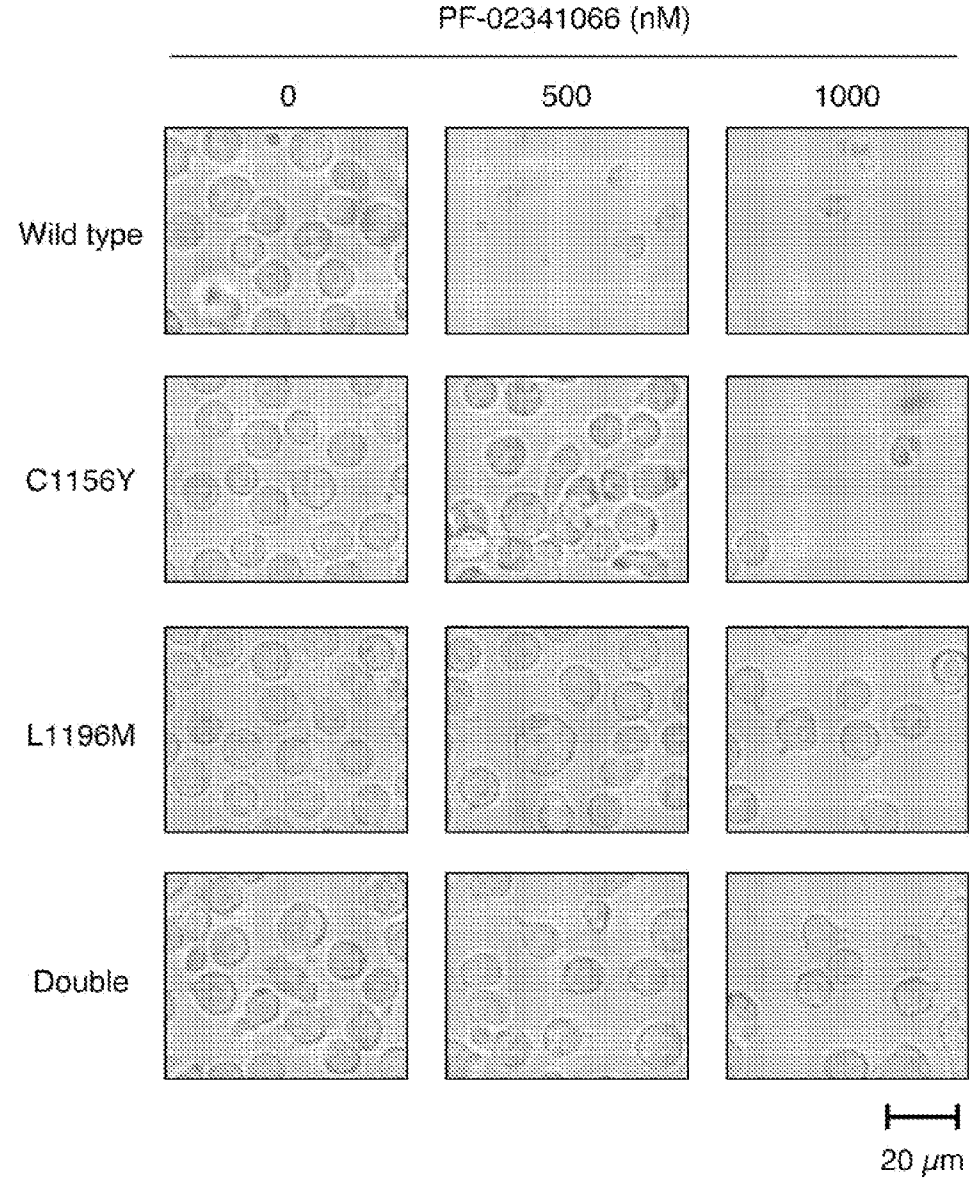


Figure 4

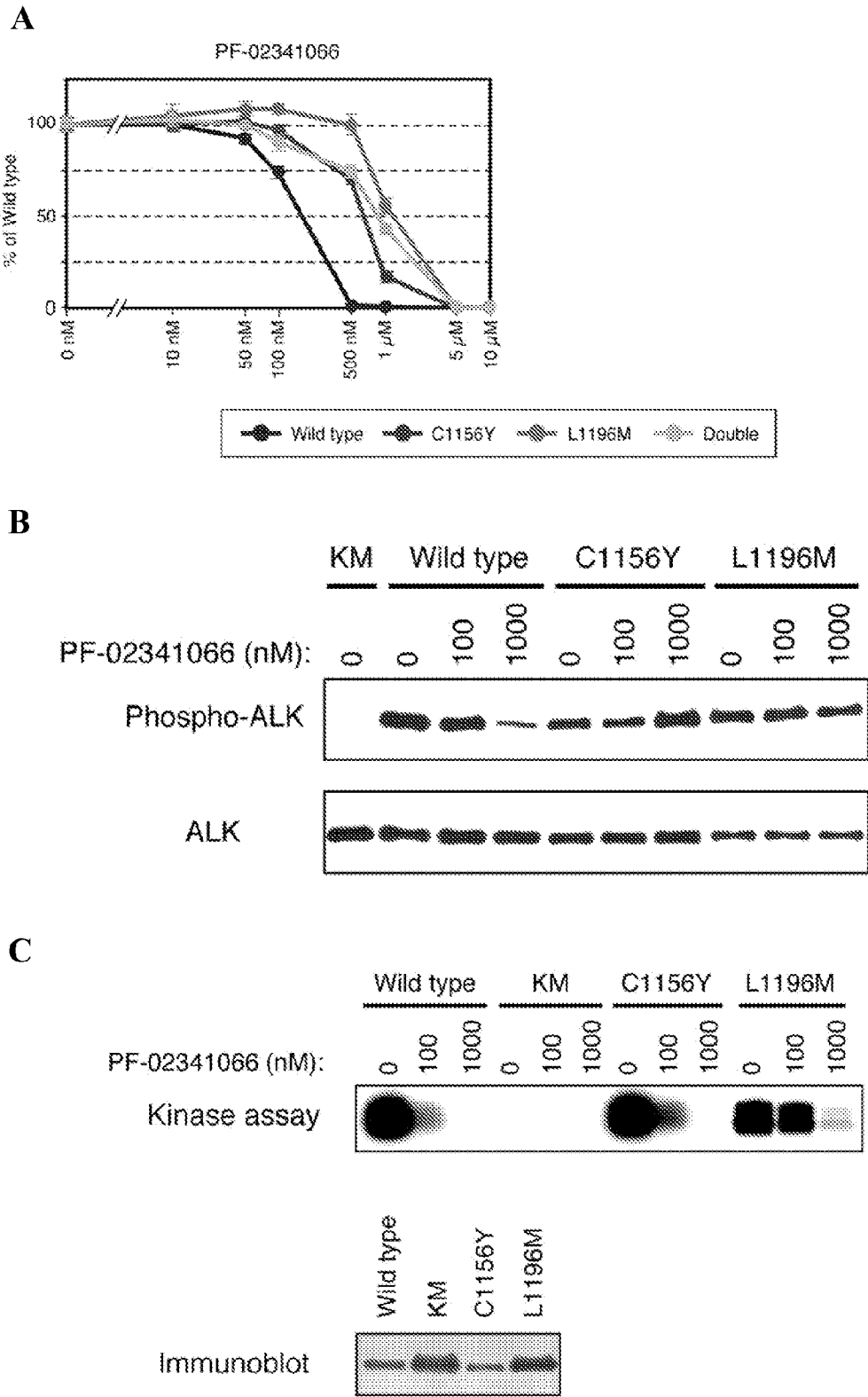
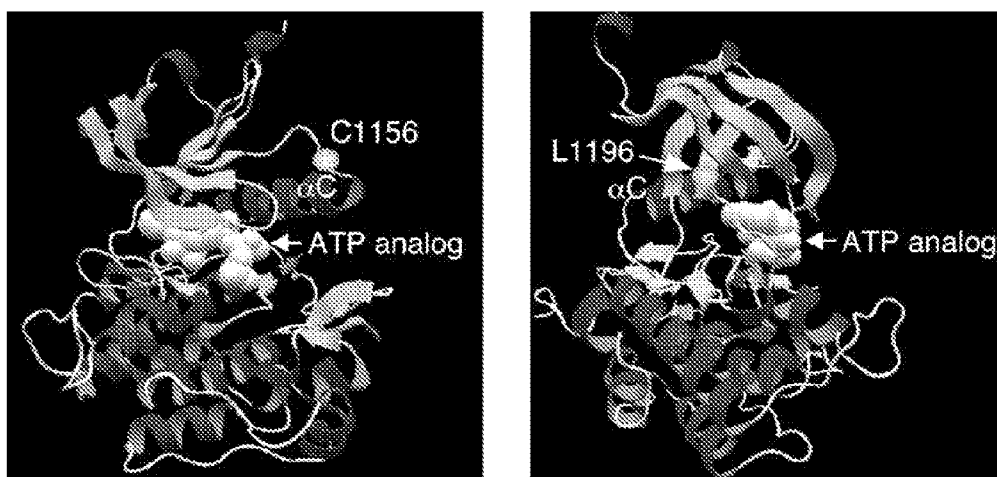


Figure 5



P042876Seq. TXT
SEQUENCE LISTING

<110> JI CHI MEDICAL UNIVERSITY

<120> IDENTIFICATION, ASSESSMENT, AND THERAPY OF CANCERS WITH
INNATE OR ACQUIRED RESISTANCE TO ALK INHIBITORS

<130> CGX-003.25

<140> PCT/IB2011/000382

<141> 2011-02-04

<150> 61/337,465

<151> 2010-02-04

<160> 40

<170> Patent In version 3.5

<210> 1

<211> 6222

<212> DNA

<213> Homo sapiens

<400> 1

gggggcggca gcggt ggt ag cagct ggt ac ct cccgccgc ct ct gt t cgg aggggt cgcgg	60
ggcaccgagg t gct t t ccgg ccgccct ct g gt cggccacc caaagccgcg ggcgct gat g	120
at gggg gagg agggggcggc aagat t t cgg gcgccct gc cct gaacgcc ct cagct gct	180
gccgccgggg ccgct ccagt gcct gcgaac t ct gaggagc cgaggcgccg gt gagagcaa	240
ggacgct gca aact t gcgca gcgcgggggc t gggat t cac gccagaagt t cagcaggca	300
gacagt ccga agcct t cccg cagcggagag at agct t gag ggt gcgcaag acggcagcct	360
ccgccct cgg t t cccgcca gaccgggcag aagagct t gg aggagccaaa aggaacgcaa	420
aaggcggcca ggacagcgt g cagcagct gg gagccgccgt t ct cagcct t aaaagt t gca	480
gagat t ggag gct gccccga gaggggacag accccagct c cgact gcggg gggcaggaga	540
ggacgggt acc caact gccac ct ccct t caa ccat agt agt t cct ct gt ac cgagcgcagc	600
gagct acaga cggggggcgcg gcaact cggcg cggagagcgg gaggct caag gt cccagcca	660
gt gagcccag t gt gct t gag t gt ct ct gga ct cgccct g agct t ccagg t ct gt t t cat	720
t t agact cct gct cgcct cc gt gcagt t gg gggaaagcaa gagact t gcg cgcacgcaca	780
gt cct ct gga gat caggt gg aaggagccgc t gggg accaa ggact gt t ca ggcct ct t c	840
ccat ct cggg gagagcgaag ggt gaggt g ggcccggaga gcagt gt aaa cggcct cct c	900
cggcgggat g ggagccat cg ggct cct gt g gct cct gccg ct gct gct t t ccacggcagc	960
t gt gggct cc gggat gggga ccggccagcg cgcgggct cc ccagct gcgg ggccgccgct	1020
gcagccccgg gagccact ca gct act cgcg cct gcagagg aagagt ct gg cagt t gact t	1080
cgt ggt gccc t cgt ct t cc gt gt ct acgc ccgggacct a ct gct gccac cat cct cct c	1140
ggagct gaag gct ggcaggc ccgagggccg cggct cgct a gct ct ggact gcgccccgct	1200
gct caggt t g ct ggggccgg cgcgggggt ct cct ggacc gccggt t cac cagccccggc	1260
agaggccccg acgt gt cca ggggt gct gaa gggcggt cc gt gcgcaagc t ccggcgt gc	1320

P042876Seq. TXT

caagcagt t g	gt gct ggagc	t gggcgagga	ggcgat ct t g	gaggggt t gcg	t cggggccccc	1380
cgggggagggcg	gct gt ggggc	t gct ccagt t	caat ct cagc	gagct gt t ca	gt t ggt ggat	1440
t cgccaaggc	gaagggcgac	t gaggat ccg	cct gat gccc	gagaagaagg	cgt cggaagt	1500
gggcagagag	ggaaggct gt	ccgcggcaat	t cgcgcct cc	cagccccgcc	t t ct ct t cca	1560
gat ct t cggg	act ggt cat a	gct cct t gga	at caccaaca	aacat gcct t	ct cct t ct cc	1620
t gat t at t t t	acat ggaat c	t cacct ggat	aat gaaagac	t cct t ccct t	t cct gt ct ca	1680
t cgcagccga	t at ggt ct gg	agt gcagct t	t gact t cccc	t gt gagct gg	agt at t cccc	1740
t ccaact gcat	gacct cagga	accagagct g	gt cct ggcgc	cgcct cccct	ccgaggaggc	1800
ct cccagat g	gact t gct gg	at gggcct gg	ggcagagcgt	t ct aaggaga	t gcccagagg	1860
ct cct t t ct c	ct t ct caaca	cct cagct ga	ct ccaagcac	acat cct ga	gt ccgt ggat	1920
gaggagcagc	agt gagcact	gcacact ggc	cgt ct cggg g	cacaggcacc	t gcagccct c	1980
t ggaaggt ac	at t gcccagc	t gct gcccga	caacgaggct	gcaagagaga	t cct cct gat	2040
gcccact cca	gggaagcat g	gt t ggacagt	gct ccaggga	agaat cgggc	gt ccagacaa	2100
cccat t t cga	gt ggccct gg	aat acat ct c	cagt ggaaac	cgcagct t gt	ct gcagt gga	2160
ct t ct t t gcc	ct gaagaact	gcagt gaagg	aacat cccca	ggct ccaaga	t ggccct gca	2220
gagct cct t c	act t gt t gga	at gggacagt	cct ccagct t	gggcaggcct	gt gact t cca	2280
ccaggact gt	gcccaggagg	aagat gagag	ccagat gt gc	cggaaact gc	ct gt gggg t t	2340
t t act gcaac	t t t gaagat g	gct t ct gt gg	ct ggacccaa	ggcacact gt	caccccacac	2400
t cct caat gg	caggt cagga	ccct aaagga	t gcccggt t c	caggaccacc	aagacat gc	2460
t ct at t gct c	agt accact g	at gt cccgc	t t ct gaaagt	gct acagt ga	ccagt gct ac	2520
gt t t cct gca	ccgat caaga	gct ct ccat g	t gagct ccga	at gt cct ggc	t cat t cgt gg	2580
agt ct t gagg	ggaaacgt gt	cct t ggt gct	agt ggagaac	aaaaccggga	aggagcaagg	2640
caggat ggt c	t ggcat gt cg	ccgcct at ga	aggct t gagg	ct gt ggcagt	ggat ggt gt t	2700
gcct ct cct c	gat gt gt ct g	acaggt t ct g	gct gcagat g	gt cgcct ggt	ggggacaagg	2760
at ccagagcc	at cgt ggct t	t t gacaat at	ct ccat cagc	ct ggact gct	acct caccat	2820
t agcggagag	gacaagat cc	t gcagaat ac	agcacccaaa	t caagaaacc	t gt t t gagag	2880
aaacccaaac	aaggagct ga	aaccggggga	aaat t cacca	agacagacc	ccat ct t t ga	2940
ccct acagt t	cat t ggct gt	t caccacat g	t ggggccagc	gggccccat g	gccccaccca	3000
ggcacagt gc	aacaacgcct	accagaact c	caacct gagg	gt ggaggt gg	ggagcgaggg	3060
ccccct gaaa	ggcat ccaga	t ct ggaaggt	gccagccacc	gacacct aca	gcct ct cggg	3120
ct acggagct	gct ggcggga	aaggcgggga	gaacacat g	at gcggt ccc	acggcgt gt c	3180
t gt gct gggc	at ct t caacc	t ggagaagga	t gacat gct g	t acat cct gg	t t gggcagca	3240
gggagaggac	gcct gcccga	gt acaaacca	gt t aat ccag	aaagt ct gca	t t ggagagaa	3300
caat gt gat a	gaagaagaaa	t ccgt gt gaa	cagaagcgt g	cat gagt ggg	caggaggcgg	3360

P042876Seq. TXT

aggaggaggg	ggt ggagcca	cct acgt at t	t aagat gaag	gat ggagt gc	cggt gccctt	3420
gat cat t gca	gccggaggt g	gt ggcagggc	ct acggggcc	aagacagaca	cgt t ccaccc	3480
agagagact g	gagaat aact	cct cggt t ct	agggct aaac	ggcaat t ccg	gagccgcagg	3540
t ggt ggaggt	ggct ggaat g	at aacact t c	ct t gct ct gg	gccggaaaat	ct t t gcagga	3600
gggt gccacc	ggaggacat t	cct gccccca	ggccat gaag	aagt gggggt	gggagacaag	3660
aggggggt t t c	ggaggggggt g	gaggggggt g	ct cct caggt	ggaggaggcg	gaggat at at	3720
aggcggcaat	gcagcct caa	acaat gaccc	cgaaat ggat	ggggaagat g	gggt t t cct t	3780
cat cagt cca	ct gggcat cc	t gt acacccc	agct t t aaaa	gt gat ggaag	gccacgggga	3840
agt gaat at t	aagcat t at c	t aaact gcag	t cact gt gag	gt agacgaat	gt cacat gga	3900
ccct gaaagc	cacaaggt ca	t ct gct t ct g	t gaccacggg	acggt gct gg	ct gaggat gg	3960
cgt ct cct gc	at t gt gt cac	ccaccccgga	gccacacct g	ccact ct cgc	t gat cct ct c	4020
t gt ggt gacc	t ct gccct cg	t ggccgccct	gggt cct ggct	t t ct ccggca	t cat gat t gt	4080
gt accgccgg	aagcaccagg	agct gcaagc	cat gcagat g	gagct gcaga	gccct gagt a	4140
caagct gagg	aagct ccgca	cct cgacat	cat gaccgac	t acaacccca	act act gct t	4200
t gct ggcaag	acct cct cca	t cagt gacct	gaaggaggt g	ccgcggaaaa	acat caccct	4260
cat t cggggt	ct gggccat g	gcgcct t t gg	ggaggt gt at	gaaggccagg	t gt ccggaat	4320
gccaacgac	ccaagcccc	t gcaagt ggc	t gt gaagacg	ct gcct gaag	t gt gct ct ga	4380
acaggacgaa	ct ggat t t cc	t cat ggaagc	cct gat cat c	agcaaat t ca	accaccagaa	4440
cat t gt t cgc	t gcat t gggg	t gaggct gca	at ccct gccc	cggt t cat cc	t gct ggagct	4500
cat ggcgggg	ggagacct ca	agt cct t cct	ccgagagacc	cgccct cgcc	cgagccagcc	4560
ct cct ccct g	gccat gct gg	acct t ct gca	cgt ggct cgg	gacat t gcct	gt ggct gt ca	4620
gt at t t ggag	gaaaaccact	t cat ccaccg	agacat t gct	gccagaaact	gcct ct t gac	4680
ct gt ccaggc	cct ggaagag	t ggccaagat	t ggagact t c	gggat ggccc	gagacat ct a	4740
cagggcgagc	t act at agaa	agggaggct g	t gccat gct g	ccagt t aagt	ggat gcccc	4800
agaggcct t c	at ggaaggaa	t at t cact t c	t aaaacagac	acat ggt cct	t t ggagt gct	4860
gct at gggaa	at ct t t t ct c	t t ggat at at	gccat acccc	agcaaaaagca	accaggaagt	4920
t ct ggagt t t	gt caccagt g	gaggccggat	ggaccacccc	aagaact gcc	ct gggcct gt	4980
at accggat a	at gact cagt	gct ggcaaca	t cagcct gaa	gacaggccca	act t t gccat	5040
cat t t t ggag	aggat t gaat	act gcaccca	ggacccggt	gt aat caaca	ccgct t t gcc	5100
gat agaat at	ggt ccact t g	t ggaagagga	agagaaagt g	cct gt gaggc	ccaaggaccc	5160
t gaggggggt t	cct cct ct cc	t ggt ct ct ca	acaggcaaaa	cgggaggagg	agcgcagccc	5220
agct gcccc	ccacct ct gc	ct accacct c	ct ct ggcaag	gct gcaaaga	aaccacagc	5280
t gcagagat c	t ct gt t cgag	t ccct agagg	gccggccgt g	gaaggggggac	acgt gaat at	5340
ggcat t ct ct	cagt ccaacc	ct cct t cgga	gt t gcacaag	gt ccacggat	ccagaaacaa	5400

P042876Seq. TXT

```

gccaccagc ttgtggaacc caacgt acgg ct cct ggt tt acagagaaac ccacaaaaaa 5460
gaat aat cct at agcaaaga aggagccaca cgacaggggt aacct ggggc tggagggaag 5520
ctgt actgt c ccacct aacg ttgcaact gg gagact tccg ggggcct cac t gct cct aga 5580
gccct ct t cg ct gact gcca at at gaagga ggt acct ct g t t caggct ac gt cact t ccc 5640
t t gt gggaaat gt caat t acg gct accagca acagggct t g ccct t agaag ccgct act gc 5700
ccct ggagct ggt cat t acg aggat accat t ct gaaaagc aagaat agca t gaaccagcc 5760
t gggccct ga gct cggcgc acact cact t ct ct t cct t g ggat ccct aa gaccgt ggag 5820
gagagagagg caat ggct cc t t caciaaac agagaccaa t gt cacgt t t t gt t t t gt gc 5880
caacct at t t t gaagt acca ccaaaaaagc t gt at t t t ga aaat gct t t a gaaaggt t t t 5940
gagcat ggg t cat cct at t ct t t cgaaag aagaaaat at cat aaaaat g agt gat aaat 6000
acaaggccca gat gt ggt t g cat aaggt t t t at gcat gt t t gt t gt at a ct t cct t at g 6060
ct t ct t t caa at t gt gt gt g ct ct gct t ca at gt agt cag aat t agct gc t t ct at gt t t 6120
cat agt t ggg gt cat agat g t t t cct t gcc t t gt t gat gt ggacat gagc cat t t gaggg 6180
gagaggaac ggaaat aaag gagt t at t t g t aat gact aa aa 6222

```

<210> 2
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic primer

<400> 2
 ggt aagaagt ggct cact ct t gag 24

<210> 3
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic primer

<400> 3
 cacaacaact gcagcaaaga ct gg 24

<210> 4
 <211> 1620
 <212> PRT
 <213> Homo sapiens

<400> 4
 Met Gly Ala Ile Gly Leu Leu Trp Leu Leu Pro Leu Leu Leu Ser Thr
 1 5 10 15

Ala Ala Val Gly Ser Gly Met Gly Thr Gly Gln Arg Ala Gly Ser Pro
 20 25 30

P042876Seq. TXT

Ala Ala Gly Pro Pro Leu Gln Pro Arg Glu Pro Leu Ser Tyr Ser Arg
35 40 45

Leu Gln Arg Lys Ser Leu Ala Val Asp Phe Val Val Pro Ser Leu Phe
50 55 60

Arg Val Tyr Ala Arg Asp Leu Leu Leu Pro Pro Ser Ser Ser Glu Leu
65 70 75 80

Lys Ala Gly Arg Pro Glu Ala Arg Gly Ser Leu Ala Leu Asp Cys Ala
85 90 95

Pro Leu Leu Arg Leu Leu Gly Pro Ala Pro Gly Val Ser Trp Thr Ala
100 105 110

Gly Ser Pro Ala Pro Ala Glu Ala Arg Thr Leu Ser Arg Val Leu Lys
115 120 125

Gly Gly Ser Val Arg Lys Leu Arg Arg Ala Lys Gln Leu Val Leu Glu
130 135 140

Leu Gly Glu Glu Ala Ile Leu Glu Gly Cys Val Gly Pro Pro Gly Glu
145 150 155 160

Ala Ala Val Gly Leu Leu Gln Phe Asn Leu Ser Glu Leu Phe Ser Trp
165 170 175

Trp Ile Arg Gln Gly Glu Gly Arg Leu Arg Ile Arg Leu Met Pro Glu
180 185 190

Lys Lys Ala Ser Glu Val Gly Arg Glu Gly Arg Leu Ser Ala Ala Ile
195 200 205

Arg Ala Ser Gln Pro Arg Leu Leu Phe Gln Ile Phe Gly Thr Gly His
210 215 220

Ser Ser Leu Glu Ser Pro Thr Asn Met Pro Ser Pro Ser Pro Asp Tyr
225 230 235 240

Phe Thr Trp Asn Leu Thr Trp Ile Met Lys Asp Ser Phe Pro Phe Leu
245 250 255

Ser His Arg Ser Arg Tyr Gly Leu Glu Cys Ser Phe Asp Phe Pro Cys
260 265 270

Glu Leu Glu Tyr Ser Pro Pro Leu His Asp Leu Arg Asn Gln Ser Trp
275 280 285

Ser Trp Arg Arg Ile Pro Ser Glu Glu Ala Ser Gln Met Asp Leu Leu
290 295 300

P042876Seq. TXT

Asp Gly Pro Gly Ala Glu Arg Ser Lys Glu Met Pro Arg Gly Ser Phe
 305 310 315 320
 Leu Leu Leu Asn Thr Ser Ala Asp Ser Lys His Thr Ile Leu Ser Pro
 325 330 335
 Trp Met Arg Ser Ser Ser Glu His Cys Thr Leu Ala Val Ser Val His
 340 345 350
 Arg His Leu Gln Pro Ser Gly Arg Tyr Ile Ala Gln Leu Leu Pro His
 355 360 365
 Asn Glu Ala Ala Arg Glu Ile Leu Leu Met Pro Thr Pro Gly Lys His
 370 375 380
 Gly Trp Thr Val Leu Gln Gly Arg Ile Gly Arg Pro Asp Asn Pro Phe
 385 390 395 400
 Arg Val Ala Leu Glu Tyr Ile Ser Ser Gly Asn Arg Ser Leu Ser Ala
 405 410 415
 Val Asp Phe Phe Ala Leu Lys Asn Cys Ser Glu Gly Thr Ser Pro Gly
 420 425 430
 Ser Lys Met Ala Leu Gln Ser Ser Phe Thr Cys Trp Asn Gly Thr Val
 435 440 445
 Leu Gln Leu Gly Gln Ala Cys Asp Phe His Gln Asp Cys Ala Gln Gly
 450 455 460
 Glu Asp Glu Ser Gln Met Cys Arg Lys Leu Pro Val Gly Phe Tyr Cys
 465 470 475 480
 Asn Phe Glu Asp Gly Phe Cys Gly Trp Thr Gln Gly Thr Leu Ser Pro
 485 490 495
 His Thr Pro Gln Trp Gln Val Arg Thr Leu Lys Asp Ala Arg Phe Gln
 500 505 510
 Asp His Gln Asp His Ala Leu Leu Leu Ser Thr Thr Asp Val Pro Ala
 515 520 525
 Ser Glu Ser Ala Thr Val Thr Ser Ala Thr Phe Pro Ala Pro Ile Lys
 530 535 540
 Ser Ser Pro Cys Glu Leu Arg Met Ser Trp Leu Ile Arg Gly Val Leu
 545 550 555 560
 Arg Gly Asn Val Ser Leu Val Leu Val Glu Asn Lys Thr Gly Lys Glu
 565 570 575

P042876Seq. TXT

G n G y A r g M e t V a l T r p H i s V a l A l a A l a T y r G u G y L e u S e r L e u
 580 585 590
 T r p G n T r p M e t V a l L e u P r o L e u L e u A s p V a l S e r A s p A r g P h e T r p
 595 600 605
 L e u G n M e t V a l A l a T r p T r p G y G n G y S e r A r g A l a I l e V a l A l a
 610 615 620
 P h e A s p A s n I l e S e r I l e S e r L e u A s p C y s T y r L e u T h r I l e S e r G y
 625 630 635 640
 G u A s p L y s I l e L e u G n A s n T h r A l a P r o L y s S e r A r g A s n L e u P h e
 645 650 655
 G u A r g A s n P r o A s n L y s G u L e u L y s P r o G y G u A s n S e r P r o A r g
 660 665 670
 G n T h r P r o I l e P h e A s p P r o T h r V a l H i s T r p L e u P h e T h r T h r C y s
 675 680 685
 G y A l a S e r G y P r o H i s G y P r o T h r G n A l a G n C y s A s n A s n A l a
 690 695 700
 T y r G n A s n S e r A s n L e u S e r V a l G u V a l G y S e r G u G y P r o L e u
 705 710 715 720
 L y s G y I l e G n I l e T r p L y s V a l P r o A l a T h r A s p T h r T y r S e r I l e
 725 730 735
 S e r G y T y r G y A l a A l a G y G y L y s G y G y L y s A s n T h r M e t M e t
 740 745 750
 A r g S e r H i s G y V a l S e r V a l L e u G y I l e P h e A s n L e u G u L y s A s p
 755 760 765
 A s p M e t L e u T y r I l e L e u V a l G y G n G n G y G u A s p A l a C y s P r o
 770 775 780
 S e r T h r A s n G n L e u I l e G n L y s V a l C y s I l e G y G u A s n A s n V a l
 785 790 795 800
 I l e G u G u G u I l e A r g V a l A s n A r g S e r V a l H i s G u T r p A l a G y
 805 810 815
 G y G y G y G y G y G y G y A l a T h r T y r V a l P h e L y s M e t L y s A s p
 820 825 830
 G y V a l P r o V a l P r o L e u I l e I l e A l a A l a G y G y G y G y A r g A l a
 835 840 845

P042876Seq. TXT

Tyr Gly Ala Lys Thr Asp Thr Phe His Pro Glu Arg Leu Glu Asn Asn
 850 855 860
 Ser Ser Val Leu Gly Leu Asn Gly Asn Ser Gly Ala Ala Gly Gly Gly
 865 870 875 880
 Gly Gly Trp Asn Asp Asn Thr Ser Leu Leu Trp Ala Gly Lys Ser Leu
 885 890 895
 Gln Glu Gly Ala Thr Gly Gly His Ser Cys Pro Gln Ala Met Lys Lys
 900 905 910
 Trp Gly Trp Glu Thr Arg Gly Gly Phe Gly Gly Gly Gly Gly Gly Cys
 915 920 925
 Ser Ser Gly Gly Gly Gly Gly Gly Tyr Ile Gly Gly Asn Ala Ala Ser
 930 935 940
 Asn Asn Asp Pro Glu Met Asp Gly Glu Asp Gly Val Ser Phe Ile Ser
 945 950 955 960
 Pro Leu Gly Ile Leu Tyr Thr Pro Ala Leu Lys Val Met Glu Gly His
 965 970 975
 Gly Glu Val Asn Ile Lys His Tyr Leu Asn Cys Ser His Cys Glu Val
 980 985 990
 Asp Glu Cys His Met Asp Pro Glu Ser His Lys Val Ile Cys Phe Cys
 995 1000 1005
 Asp His Gly Thr Val Leu Ala Glu Asp Gly Val Ser Cys Ile Val
 1010 1015 1020
 Ser Pro Thr Pro Glu Pro His Leu Pro Leu Ser Leu Ile Leu Ser
 1025 1030 1035
 Val Val Thr Ser Ala Leu Val Ala Ala Leu Val Leu Ala Phe Ser
 1040 1045 1050
 Gly Ile Met Ile Val Tyr Arg Arg Lys His Gln Glu Leu Gln Ala
 1055 1060 1065
 Met Gln Met Glu Leu Gln Ser Pro Glu Tyr Lys Leu Ser Lys Leu
 1070 1075 1080
 Arg Thr Ser Thr Ile Met Thr Asp Tyr Asn Pro Asn Tyr Cys Phe
 1085 1090 1095
 Ala Gly Lys Thr Ser Ser Ile Ser Asp Leu Lys Glu Val Pro Arg
 1100 1105 1110

P042876Seq. TXT

Lys Asn Ile Thr Leu Ile Arg Gly Leu Gly His Gly Ala Phe Gly
 1115 1120 1125
 Glu Val Tyr Gu Gly Gn Val Ser Gly Met Pro Asn Asp Pro Ser
 1130 1135 1140
 Pro Leu Gn Val Ala Val Lys Thr Leu Pro Glu Val Cys Ser Glu
 1145 1150 1155
 Gn Asp Gu Leu Asp Phe Leu Met Glu Ala Leu Ile Ile Ser Lys
 1160 1165 1170
 Phe Asn His Gn Asn Ile Val Arg Cys Ile Gly Val Ser Leu Gn
 1175 1180 1185
 Ser Leu Pro Arg Phe Ile Leu Leu Glu Leu Met Ala Gly Gly Asp
 1190 1195 1200
 Leu Lys Ser Phe Leu Arg Gu Thr Arg Pro Arg Pro Ser Gn Pro
 1205 1210 1215
 Ser Ser Leu Ala Met Leu Asp Leu Leu His Val Ala Arg Asp Ile
 1220 1225 1230
 Ala Cys Gly Cys Gn Tyr Leu Glu Glu Asn His Phe Ile His Arg
 1235 1240 1245
 Asp Ile Ala Ala Arg Asn Cys Leu Leu Thr Cys Pro Gly Pro Gly
 1250 1255 1260
 Arg Val Ala Lys Ile Gly Asp Phe Gly Met Ala Arg Asp Ile Tyr
 1265 1270 1275
 Arg Ala Ser Tyr Tyr Arg Lys Gly Gly Cys Ala Met Leu Pro Val
 1280 1285 1290
 Lys Trp Met Pro Pro Glu Ala Phe Met Glu Gly Ile Phe Thr Ser
 1295 1300 1305
 Lys Thr Asp Thr Trp Ser Phe Gly Val Leu Leu Trp Glu Ile Phe
 1310 1315 1320
 Ser Leu Gly Tyr Met Pro Tyr Pro Ser Lys Ser Asn Gn Glu Val
 1325 1330 1335
 Leu Glu Phe Val Thr Ser Gly Gly Arg Met Asp Pro Pro Lys Asn
 1340 1345 1350
 Cys Pro Gly Pro Val Tyr Arg Ile Met Thr Gn Cys Trp Gn His
 1355 1360 1365

P042876Seq. TXT

G n	Pro	G u	Asp	Arg	Pro	Asn	Phe	Al a	I l e	I l e	Leu	G u	Arg	I l e
	1370					1375					1380			
G u	Tyr	Oys	Thr	G n	Asp	Pro	Asp	Val	I l e	Asn	Thr	Al a	Leu	Pro
	1385					1390					1395			
I l e	G u	Tyr	G y	Pro	Leu	Val	G u	G u	G u	G u	Lys	Val	Pro	Val
	1400					1405					1410			
Arg	Pro	Lys	Asp	Pro	G u	G y	Val	Pro	Pro	Leu	Leu	Val	Ser	G n
	1415					1420					1425			
G n	Al a	Lys	Arg	G u	G u	G u	Arg	Ser	Pro	Al a	Al a	Pro	Pro	Pro
	1430					1435					1440			
Leu	Pro	Thr	Thr	Ser	Ser	G y	Lys	Al a	Al a	Lys	Lys	Pro	Thr	Al a
	1445					1450					1455			
Al a	G u	I l e	Ser	Val	Arg	Val	Pro	Arg	G y	Pro	Al a	Val	G u	G y
	1460					1465					1470			
G y	Hi s	Val	Asn	Met	Al a	Phe	Ser	G n	Ser	Asn	Pro	Pro	Ser	G u
	1475					1480					1485			
Leu	Hi s	Lys	Val	Hi s	G y	Ser	Arg	Asn	Lys	Pro	Thr	Ser	Leu	Trp
	1490					1495					1500			
Asn	Pro	Thr	Tyr	G y	Ser	Trp	Phe	Thr	G u	Lys	Pro	Thr	Lys	Lys
	1505					1510					1515			
Asn	Asn	Pro	I l e	Al a	Lys	Lys	G u	Pro	Hi s	Asp	Arg	G y	Asn	Leu
	1520					1525					1530			
G y	Leu	G u	G y	Ser	Oys	Thr	Val	Pro	Pro	Asn	Val	Al a	Thr	G y
	1535					1540					1545			
Arg	Leu	Pro	G y	Al a	Ser	Leu	Leu	Leu	G u	Pro	Ser	Ser	Leu	Thr
	1550					1555					1560			
Al a	Asn	Met	Lys	G u	Val	Pro	Leu	Phe	Arg	Leu	Arg	Hi s	Phe	Pro
	1565					1570					1575			
Oys	G y	Asn	Val	Asn	Tyr	G y	Tyr	G n	G n	G n	G y	Leu	Pro	Leu
	1580					1585					1590			
G u	Al a	Al a	Thr	Al a	Pro	G y	Al a	G y	Hi s	Tyr	G u	Asp	Thr	I l e
	1595					1600					1605			
Leu	Lys	Ser	Lys	Asn	Ser	Met	Asn	G n	Pro	G y	Pro			
	1610					1615					1620			

P042876Seq. TXT

<210> 5
 <211> 3926
 <212> DNA
 <213> Homo sapiens

<400> 5
 ggcggcgcgg cgcggcgcct c gcggct gct g cct gggaggg aggcgggca ggcggcct gag 60
 cgcgcgcggt ct caacgt ga cggggaagt g gt t cgggcgg ccgcggt t a ct accccagg 120
 gcgaacggac ggacgacgga ggcgggagcc ggt agccgag ccgggcgacc t agagaacga 180
 gcgggt cagg ct cagcgt cg gccact ct gt cggt ccgct g aat gaagt gc ccgcccct ct 240
 gagcccgag cccggcgct t t ccccgcaag at ggacggt t t cgccggcag t ct cgat gat 300
 agt at t t ct g ct gcaagt ac t t ct gat gt t caagat cgcc t gt cagct ct t gagt cacga 360
 gt t cagcaac aagaagat ga aat cact gt g ct aaaggcgg ct t t ggct ga t gt t t t gagg 420
 cgt ct t gcaa t ct ct gaaga t cat gt ggcc t cagt gaaaa aat cagt ct c aagt aaaggc 480
 caaccaagcc ct cgagcagt t at t cccat g t cct gt at aa ccaat ggaag t ggt gcaaac 540
 agaaaaccaa gt cat accag t gct gt ct ca at t gcaggaa aagaaact ct t t cat ct gct 600
 gct aaaagt g gt acagaaaa aaagaaagaa aaaccacaag gacagagaga aaaaaagag 660
 gaat ct cat t ct aat gat ca aagt ccacaa at t cgagcat cacct t ct cc ccagccct ct 720
 t cacaacct c t ccaaat aca cagacaaact ccagaaagca agaat gct ac t cccaccaa 780
 agcat aaaac gacct cacc agct gaaaag t cacat aat t ct t gggaaaa t t cagat gat 840
 agccgt aat a aat t gt cgaa aat acct t ca acacccaaat t aat accaaa agt t accaaa 900
 act gcagaca agcat aaaga t gt cat cat c aaccaagaag gagaat at at t aaaat gt t t 960
 at gcgcggt c ggccaat t ac cat gt t cat t cct t ccgat g t t gacaact a t gat gacat c 1020
 agaacggaac t gcct cct ga gaagct caaa ct ggagt ggg cat at ggt t a t cgaggaaag 1080
 gact gt agag ct aat gt t t a cct t ct t ccg accggggaaa t agt t t at t t cat t gcat ca 1140
 gt agt agt ac t at t t aat t a t gaggagaga act cagcgac act acct ggg ccat acagac 1200
 t gt gt gaaat gcct t gct at acat cct gac aaaat t agga t t gcaact gg acagat agct 1260
 ggcgt ggat a aagat ggaag gcct ct acaa cccacgt ca gagt gt ggga t t ct gt t act 1320
 ct at ccacac t gcagat t at t ggact t ggc act t t t gage gt ggagt agg at gcct ggat 1380
 t t t t caaaag cagat t cagg t gt t cat t t a t gt gt t at t g at gact ccaa t gagcat at g 1440
 ct t act gt at gggact ggca gaagaaagca aaaggagcag aaat aaagac aacaaat gaa 1500
 gt t gt t t t gg ct gt ggagt t t caccaca gat gcaaat a ccat aat t ac at gcggt aaa 1560
 t ct cat at t t t ct t ct ggac ct ggagcggc aat t cact aa caagaaaaca gggaat t t t t 1620
 gggaaat at g aaaagccaaa at t t gt gcag t gt t t agcat t ct t ggggaa t ggagat gt t 1680
 ct t act ggag act caggt gg agt cat gct t at at ggagca aaact act gt agagcccaca 1740
 cct gggaaag gacct aaagt gt accgccgg aagcaccagg agct gcaagc cat gcagat g 1800

P042876Seq. TXT						
gagct gcaga gccct gagt a caagct gage	aagct ccgca cct cgacat	cat gaccgac	1860			
t acaacccca act act gct t t gct ggcaag	acct cct cca t cagt gacct	gaaggaggt g	1920			
ccgcggaaaa acat caccct cat t cggggt	ct gggccat g gaggct t t gg	ggaggt gt at	1980			
gaaggccagg t gt ccggaat gcccacgac	ccaagccccc t gcaagt ggc	t gt gaagacg	2040			
ct gcct gaag t gt gct ct ga acaggacgaa	ct ggat t t cc t cat ggaagc	cct gat cat c	2100			
agcaaat t ca accaccagaa cat t gt t cgc	t gcat t gggg t gaggct gca	at ccct gccc	2160			
cggt t cat cc t gct ggagct cat ggcgggg	ggagacct ca agt cct t cct	ccgagagacc	2220			
cgccct cgcc cgagccagcc ct cct ccct g	gccat gct gg acct t ct gca	cgt ggct cgg	2280			
gacat t gcct gt ggct gt ca gt at t t ggag	gaaaaccact t cat ccaccg	agacat t gct	2340			
gccagaaact gcct ct t gac ct gt ccaggc	cct ggaagag t ggccaagat	t ggagact t c	2400			
gggat ggccc gagacat ct a cagggcgagc	t act at agaa agggaggct g	t gccat gct g	2460			
ccagt t aagt ggat gcccc agaggcct t c	at ggaaggaa t at t cact t c	t aaaacagac	2520			
acat ggt cct t t ggagt gct gct at gggaa	at ct t t t ct c t t ggat at at	gccat acccc	2580			
agcaaaaagca accaggaagt t ct ggagt t t	gt caccagt g gaggccggat	ggaccacccc	2640			
aagaact gcc ct gggcct gt at accggat a	at gact cagt gct ggcaaca	t cagcct gaa	2700			
gacaggccca act t t gccat cat t t t ggag	aggat t gaat act gcacca	ggaccgcat	2760			
gt aat caaca ccgct t t gcc gat aga at	ggg cact t g t ggaagagga	agagaaagt g	2820			
cct gt gaggc ccaaggaccc t gagggggt t	cct cct ct cc t ggt ct ct ca	acaggcaaaa	2880			
cgggaggagg agcgcagccc agct gcccc	ccacct ct gc ct accacct c	ct ct ggcaag	2940			
gct gcaaaga aaccacagc t gcagaggt c	t ct gt t cgag t ccct agagg	gccggccgt g	3000			
gaaggggggac acgt gaat at ggcat t ct ct	cagt ccaacc ct cct t cgga	gt t gcacagg	3060			
gt ccacggat ccagaaacaa gcccaccagc	t t gt ggaacc caacgt acgg	ct cct ggt t t	3120			
acagagaaac ccacaaaaa gaat aat cct	at agcaaaga aggagccaca	cgagaggggt	3180			
aacct ggggc t ggagggaag ct gt act gt c	ccacct aacg t t gcaact gg	gagact t ccg	3240			
ggggcct cac t gct cct aga gccct ct t cg	ct gact gcca at at gaagga	ggg acct ct g	3300			
t t caggct ac gt cact t ccc t t gt gggaat	gt caat t acg gct accagca	acagggct t g	3360			
ccct t agaag ccgct act gc ccct ggagct	ggg cat t acg aggat accat	t ct gaaaagc	3420			
aagaat agca t gaaccagcc t gggccct ga	gct cggg cac aact cact t	ct ct t cct t g	3480			
ggat ccct aa gaccgt ggag gagagagagg	caat caat gg ct cct t caca	aaccagagac	3540			
caaat gt cac gt t t t gt t t t	gt gccaacct at t t t gaagt	accacaaaaa	3600			
t t gaaaat gc t t t agaaagg t t t t gagcat	gggt t cat cc t at t ct t t cg	aaagaagaaa	3660			
at at cat aaa aat gagt gat aaat acaagg	cccagat gt g gt t gcat aag	gt t t t t at gc	3720			
at gt t t gt t g t at act t cct t at gct t ct t	t t aaat t gt g t gt gct ct gc	t t caat gt ag	3780			
t cagaat t ag ct gct t ct at gt t t cat agt	t ggggt cat a gat gt t t cct	t gcct t gt t g	3840			

P042876Seq. TXT

at gt ggacat gagccat t t t g aggggagagg gaacggaaat aaaggagt t a t t t gt aat ga 3900

aaaaaaaaa aaaaaaaaaa aaaaaa 3926

<210> 6

<211> 1059

<212> PRT

<213> Homo sapiens

<400> 6

Met Asp Gly Phe Ala Gly Ser Leu Asp Asp Ser Ile Ser Ala Ala Ser
1 5 10 15

Thr Ser Asp Val Gln Asp Arg Leu Ser Ala Leu Glu Ser Arg Val Gln
20 25 30

Gln Gln Glu Asp Glu Ile Thr Val Leu Lys Ala Ala Leu Ala Asp Val
35 40 45

Leu Arg Arg Leu Ala Ile Ser Glu Asp His Val Ala Ser Val Lys Lys
50 55 60

Ser Val Ser Ser Lys Gly Gln Pro Ser Pro Arg Ala Val Ile Pro Met
65 70 75 80

Ser Cys Ile Thr Asn Gly Ser Gly Ala Asn Arg Lys Pro Ser His Thr
85 90 95

Ser Ala Val Ser Ile Ala Gly Lys Glu Thr Leu Ser Ser Ala Ala Lys
100 105 110

Ser Gly Thr Glu Lys Lys Lys Glu Lys Pro Gln Gly Gln Arg Glu Lys
115 120 125

Lys Glu Glu Ser His Ser Asn Asp Gln Ser Pro Gln Ile Arg Ala Ser
130 135 140

Pro Ser Pro Gln Pro Ser Ser Gln Pro Leu Gln Ile His Arg Gln Thr
145 150 155 160

Pro Glu Ser Lys Asn Ala Thr Pro Thr Lys Ser Ile Lys Arg Pro Ser
165 170 175

Pro Ala Glu Lys Ser His Asn Ser Trp Glu Asn Ser Asp Asp Ser Arg
180 185 190

Asn Lys Leu Ser Lys Ile Pro Ser Thr Pro Lys Leu Ile Pro Lys Val
195 200 205

Thr Lys Thr Ala Asp Lys His Lys Asp Val Ile Ile Asn Gln Glu Gly
210 215 220

Glu Tyr Ile Lys Met Phe Met Arg Gly Arg Pro Ile Thr Met Phe Ile

225 230 235 240
 Pro Ser Asp Val Asp₂₄₅ Asn Tyr Asp Asp Ile₂₅₀ Arg Thr Glu Leu Pro₂₅₅ Pro
 Glu Lys Leu Lys₂₆₀ Leu Glu Trp Ala Tyr₂₆₅ Gly Tyr Arg Gly Lys₂₇₀ Asp Cys
 Arg Ala Asn₂₇₅ Val Tyr Leu Leu Pro₂₈₀ Thr Gly Glu Ile Val₂₈₅ Tyr Phe Ile
 Ala Ser₂₉₀ Val Val Val Leu Phe₂₉₅ Asn Tyr Glu Glu Arg₃₀₀ Thr Glu Arg His
 Tyr₃₀₅ Leu Gly His Thr Asp₃₁₀ Cys Val Lys Cys Leu₃₁₅ Ala Ile His Pro₃₂₀ Asp
 Lys Ile Arg Ile Ala₃₂₅ Thr Gly Glu Ile Ala₃₃₀ Gly Val Asp Lys Asp₃₃₅ Gly
 Arg Pro Leu Glu₃₄₀ Pro His Val Arg Val₃₄₅ Trp Asp Ser Val Thr₃₅₀ Leu Ser
 Thr Leu Glu₃₅₅ Ile Ile Gly Leu Gly₃₆₀ Thr Phe Glu Arg Gly₃₆₅ Val Gly Cys
 Leu Asp₃₇₀ Phe Ser Lys Ala Asp₃₇₅ Ser Gly Val His Leu₃₈₀ Cys Val Ile Asp
 Asp₃₈₅ Ser Asn Glu His Met₃₉₀ Leu Thr Val Trp Asp₃₉₅ Trp Glu Lys Lys Ala₄₀₀
 Lys Gly Ala Glu₄₀₅ Ile Lys Thr Thr Asn Glu₄₁₀ Val Val Leu Ala Val₄₁₅ Glu
 Phe His Pro Thr₄₂₀ Asp Ala Asn Thr Ile₄₂₅ Ile Thr Cys Gly Lys₄₃₀ Ser His
 Ile Phe Phe Trp Thr Trp Ser Gly₄₄₀ Asn Ser Leu Thr Arg₄₄₅ Lys Glu Gly
 Ile Phe₄₅₀ Gly Lys Tyr Glu Lys₄₅₅ Pro Lys Phe Val Glu₄₆₀ Cys Leu Ala Phe
 Leu₄₆₅ Gly Asn Gly Asp Val₄₇₀ Leu Thr Gly Asp Ser₄₇₅ Gly Gly Val Met Leu₄₈₀
 Ile Trp Ser Lys Thr₄₈₅ Thr Val Glu Pro Thr₄₉₀ Pro Gly Lys Gly Pro₄₉₅ Lys
 Val Tyr Arg Arg Lys His Glu Glu Leu Glu Ala Met Glu Met Glu Leu

G n	Ser	Pro	G u	Tyr	Lys	Leu	Ser	Lys	Leu	Arg	Thr	Ser	Thr	I l e	Met
		515					520					525			
Thr	Asp	Tyr	Asn	Pro	Asn	Tyr	Cys	Phe	Al a	G l y	Lys	Thr	Ser	Ser	I l e
	530					535					540				
Ser	Asp	Leu	Lys	G u	Val	Pro	Arg	Lys	Asn	I l e	Thr	Leu	I l e	Arg	G l y
545					550					555					560
Leu	G l y	H i s	G l y	Al a	Phe	G l y	G u	Val	Tyr	G u	G l y	G n	Val	Ser	G l y
				565					570					575	
Met	Pro	Asn	Asp	Pro	Ser	Pro	Leu	G n	Val	Al a	Val	Lys	Thr	Leu	Pro
			580					585					590		
G u	Val	Cys	Ser	G u	G n	Asp	G u	Leu	Asp	Phe	Leu	Met	G u	Al a	Leu
		595					600					605			
I l e	I l e	Ser	Lys	Phe	Asn	H i s	G n	Asn	I l e	Val	Arg	Cys	I l e	G l y	Val
	610					615					620				
Ser	Leu	G n	Ser	Leu	Pro	Arg	Phe	I l e	Leu	Leu	G u	Leu	Met	Al a	G l y
625					630					635					640
G l y	Asp	Leu	Lys	Ser	Phe	Leu	Arg	G u	Thr	Arg	Pro	Arg	Pro	Ser	G n
				645					650					655	
Pro	Ser	Ser	Leu	Al a	Met	Leu	Asp	Leu	Leu	H i s	Val	Al a	Arg	Asp	I l e
			660					665					670		
Al a	Cys	G l y	Cys	G n	Tyr	Leu	G u	G u	Asn	H i s	Phe	I l e	H i s	Arg	Asp
		675					680					685			
I l e	Al a	Al a	Arg	Asn	Cys	Leu	Leu	Thr	Cys	Pro	G l y	Pro	G l y	Arg	Val
	690					695					700				
Al a	Lys	I l e	G l y	Asp	Phe	G l y	Met	Al a	Arg	Asp	I l e	Tyr	Arg	Al a	Ser
705					710					715					720
Tyr	Tyr	Arg	Lys	G l y	G l y	Cys	Al a	Met	Leu	Pro	Val	Lys	Trp	Met	Pro
				725					730					735	
Pro	G u	Al a	Phe	Met	G u	G l y	I l e	Phe	Thr	Ser	Lys	Thr	Asp	Thr	Trp
			740					745					750		
Ser	Phe	G l y	Val	Leu	Leu	Trp	G u	I l e	Phe	Ser	Leu	G l y	Tyr	Met	Pro
		755					760					765			
Tyr	Pro	Ser	Lys	Ser	Asn	G n	G u	Val	Leu	G u	Phe	Val	Thr	Ser	G l y

770

775

Gly Arg Met Asp Pro Pro Lys Asn Cys Pro Gly Pro Val Tyr Arg Ile
785 790 795 800

Met Thr Gln Cys Trp Gln His Gln Pro Glu Asp Arg Pro Asn Phe Ala
805 810 815

Ile Ile Leu Glu Arg Ile Glu Tyr Cys Thr Gln Asp Pro Asp Val Ile
820 825 830

Asn Thr Ala Leu Pro Ile Glu Tyr Gly Pro Leu Val Glu Glu Glu Glu
835 840 845

Lys Val Pro Val Arg Pro Lys Asp Pro Glu Gly Val Pro Pro Leu Leu
850 855 860

Val Ser Gln Gln Ala Lys Arg Glu Glu Glu Arg Ser Pro Ala Ala Pro
865 870 875 880

Pro Pro Leu Pro Thr Thr Ser Ser Gly Lys Ala Ala Lys Lys Pro Thr
885 890 895

Ala Ala Glu Val Ser Val Arg Val Pro Arg Gly Pro Ala Val Glu Gly
900 905 910

Gly His Val Asn Met Ala Phe Ser Gln Ser Asn Pro Pro Ser Glu Leu
915 920 925

His Arg Val His Gly Ser Arg Asn Lys Pro Thr Ser Leu Trp Asn Pro
930 935 940

Thr Tyr Gly Ser Trp Phe Thr Glu Lys Pro Thr Lys Lys Asn Asn Pro
945 950 955 960

Ile Ala Lys Lys Glu Pro His Glu Arg Gly Asn Leu Gly Leu Glu Gly
965 970 975

Ser Cys Thr Val Pro Pro Asn Val Ala Thr Gly Arg Leu Pro Gly Ala
980 985 990

Ser Leu Leu Leu Glu Pro Ser Ser Leu Thr Ala Asn Met Lys Glu Val
995 1000 1005

Pro Leu Phe Arg Leu Arg His Phe Pro Cys Gly Asn Val Asn Tyr
1010 1015 1020

Gly Tyr Gln Gln Gln Gly Leu Pro Leu Glu Ala Ala Thr Ala Pro
1025 1030 1035

Gly Ala Gly His Tyr Glu Asp Thr Ile Leu Lys Ser Lys Asn Ser

1040

1045

1050

Met Asn G n Pro G y Pro
1055

<210> 7

<211> 4679

<212> DNA

<213> Homo sapiens

<400> 7

ggcggcgcgg	cgcggcgct c	gcggct gct g	cct gggaggg	aggccgggca	ggcggct gag	60
cggcgcggct	ct caacgt ga	cggggaagt g	gt t cgggcgg	ccgcggct t a	ct accccagg	120
gcgaacggac	ggacgacgga	ggcgggagcc	ggg agccgag	ccgggcgacc	t agagaacga	180
gcgggt cagg	ct cagcgt cg	gccact ct gt	cggt ccgct g	aat gaagt gc	ccgcccct ct	240
gagcccgag	cccggcgt t	t ccccgcaag	at ggacggg t	t cgccggcag	t ct cgat gat	300
agt at t t ct g	ct gcaagt ac	t t ct gat gt t	caagat cgcc	t gt cagct ct	t gagt cacga	360
gt t cagcaac	aagaagat ga	aat cact gt g	ct aaaggcgg	ct t t ggct ga	t gt t t t gagg	420
cgt ct t gcaa	t ct ct gaaga	t cat gt ggcc	t cagt gaaaa	aat cagt ct c	aagt aaaggc	480
caaccaagcc	ct cgagcagt	t at t cccat g	t cct gt at aa	ccaat ggaag	t ggt gcaaac	540
agaaaaccaa	gt cat accag	t gct gt ct ca	at t gcaggaa	aagaaact ct	t t cat ct gct	600
gct aaaagt g	gt acagaaaa	aaagaaagaa	aaaccacaag	gacagagaga	aaaaaaagag	660
gaat ct cat t	ct aat gat ca	aagt ccacaa	at t cgagcat	cacct t ct cc	ccagccct ct	720
t cacaacct c	t ccaaat aca	cagacaaact	ccagaaagca	agaat gct ac	t cccaccaaa	780
agcat aaaac	gacat cacc	agct gaaaag	t cacat aat t	ct t gggaaaa	t t cagat gat	840
agccgt aat a	aat t gt cgaa	aat acct t ca	acacccaaat	t aat accaaa	agt t accaaa	900
act gcagaca	agcat aaaga	t gt cat cat c	aaccaagaag	gagaat at at	t aaaat gt t t	960
at gcgcggt c	ggccaat t ac	cat gt t cat t	cct t ccgat g	t t gacaact a	t gat gacat c	1020
agaacggaac	t gcct cct ga	gaagct caaa	ct ggagt ggg	cat at ggt t a	t cgaggaaag	1080
gact gt agag	ct aat gt t t a	cct t ct t ccg	accggggaaa	t agt t t at t t	cat t gcat ca	1140
gt agt agt ac	t at t t aat t a	t gaggagaga	act cagcgac	act acct ggg	ccat acagac	1200
t gt gt gaaat	gcct t gct at	acat cct gac	aaaat t agga	t t gcaact gg	acagat agct	1260
ggcgt ggat a	aagat ggaag	gcct ct acaa	ccccacgt ca	gagt gt gggg	t t ct gt t act	1320
ct at ccacac	t gcagat t at	t ggact t ggc	act t t t gagc	gt ggagt agg	at gcct ggat	1380
t t t t caaaag	cagat t cagg	t gt t cat t t a	t gt gt t at t g	at gact ccaa	t gagcat at g	1440
ct t act gt at	gggact ggca	gaagaaagca	aaaggagcag	aaat aaagac	aacaaat gaa	1500
gt t gt t t t gg	ct gt ggagt t	t cacccaaca	gat gcaaat a	ccat aat t ac	at gcggt aaa	1560
t ct cat at t t	t ct t ct ggac	ct ggagcggc	aat t cact aa	caagaaaaca	gggaat t t t t	1620
gggaaat at g	aaaagccaaa	at t t gt gcag	t gt t t agcat	t ct t ggggaa	t ggagat gt t	1680

P042876Seq. TXT

ct t act ggag	act caggt gg	agt cat gct t	at at ggagca	aaact act gt	agagcccaca	1740
cct gggaaag	gacct aaagg	t gt at at caa	at cagcaaac	aaat caaagc	t cat gat ggc	1800
agt gt gt t ca	cact t t gt ca	gat gagaaat	gggat gt t at	t aact ggagg	agggaaagac	1860
agaaaaat aa	t t ct gt ggga	t cat gat ct g	aat cct gaaa	gagaaat aga	ggg t cct gat	1920
cagt at ggca	caat cagagc	t gt agcagaa	ggaaaggcag	at caat t t t t	agt aggcaca	1980
t cacgaaact	t t at t t t acg	aggaacat t t	aat gat ggct	t ccaaat aga	agt acaggggt	2040
cat acagat g	agct t t gggg	t ct t gccaca	cat ccct t ca	aagat t t gct	ct t gacat gt	2100
gct caggaca	ggcaggt gt g	cct gt ggaac	t caat ggaac	acaggct gga	at ggaccagg	2160
ct ggt agat g	aaccaggaca	ct gt gcagat	t t t cat ccaa	gt ggcacagt	ggg ggccat a	2220
ggaacgcact	caggcaggt g	gt t t gt t ct g	gat gcagaaa	ccagagat ct	agt t t ct at c	2280
cacacagacg	ggaat gaaca	gct ct ct gt g	at gcgct act	caat agat gg	t acct t cct g	2340
gct gt aggat	ct cat gacaa	ct t t at t t ac	ct ct at gt ag	t ct ct gaaaa	t ggaagaaaa	2400
t at agcagat	at ggaaggt g	cact ggacat	t ccagct aca	t cacacacct	t gact ggt cc	2460
ccagacaaca	agt at at aat	gt ct aact cg	ggagact at g	aaat at t gt a	ct t gt accgc	2520
cggaagcacc	aggagct gca	agccat gcag	at ggagct gc	agagccct ga	gt acaagct g	2580
agcaagct cc	gcacct cgac	cat cat gacc	gact acaacc	ccaact act g	ct t t gct ggc	2640
aagacct cct	ccat cagt ga	cct gaaggag	gt gccgcgga	aaaacat cac	cct cat t cgg	2700
ggg ct gggcc	at ggagcct t	t ggggaggt g	t at gaaggcc	aggt gt ccgg	aat gcccaac	2760
gacccaagcc	ccct gcaagt	ggct gt gaag	acgct gcct g	aagt gt gct c	t gaacaggac	2820
gaact ggat t	t cct cat gga	agccct gat c	at cagcaaat	t caaccacca	gaacat t gt t	2880
cgct gcat t g	gggt gagcct	gcaat ccct g	ccccggt t ca	t cct gct gga	gct cat ggcg	2940
gggggagacc	t caagt cct t	cct ccgagag	acccgccct c	gcccagacca	gccct cct cc	3000
ct ggccat gc	t ggacct t ct	gcacgt ggct	cgggacat t g	cct gt ggct g	t cagt at t t g	3060
gaggaaaacc	act t cat cca	ccgagacat t	gct gccagaa	act gcct ct t	gacct gt cca	3120
ggccct ggaa	gagt ggccaa	gat t ggagac	t t cgggat gg	cccagagacat	ct acagggcg	3180
agct act at a	gaaagggagg	ct gt gccat g	ct gccagt t a	agt ggat gcc	cccagaggcc	3240
t t cat ggaag	gaat at t cac	t t ct aaaaca	gacacat ggt	cct t t ggagt	gct gct at gg	3300
gaaat ct t t t	ct ct t ggat a	t at gccat ac	cccagcaaaa	gcaaccagga	agt t ct ggag	3360
t t t gt cacca	gt ggaggccg	gat ggacca	cccaagaact	gccct gggcc	t gt at accgg	3420
at aat gact c	agt gct ggca	acat cagcct	gaagacaggc	ccaact t t gc	cat cat t t t g	3480
gagaggat t g	aat act gcac	ccaggaccgg	gat gt aat ca	acaccgct t t	gccgat agaa	3540
t at ggt ccac	t t gt ggaaga	ggaagagaaa	gt gcct gt ga	ggcccaagga	ccct gagggg	3600
gt t cct cct c	t cct ggt ct c	t caacaggca	aaacgggagg	aggagcgcag	cccagct gcc	3660
ccaccacct c	t gcct accac	ct cct ct ggc	aaggct gcaa	agaaaccac	agct gcagag	3720

P042876Seq. TXT

```

gt ct ct gt t c  gagt ccct ag  agggccggcc  gt ggaagggg  gacacgt gaa t at ggcat t c      3780
t ct cagt cca  accct cct t c  ggagt t gcac  aggggt ccacg  gat ccagaaa  caagcccacc      3840
agct t gt gga  acccaacgt a  cggct cct gg  t t t acagaga  aaccaccaa  aaagaat aat      3900
cct at agcaa  agaaggagcc  acacgagagg  ggt aacct gg  ggct ggaggg  aagct gt act      3960
gt cccacct a  acgt t gcaac  t gggagact t  ccgggggacct  cact gct cct  agagccct ct      4020
t cgct gact g  ccaat at gaa  ggaggt acct  ct gt t caggc  t acgt cact t  ccct t gt ggg      4080
aat gt caat t  acggct acca  gcaacagggc  t t gccct t ag  aagccgct ac  t gccct gga      4140
gct ggt cat t  acgaggat ac  cat t ct gaaa  agcaagaat a  gcat gaacca  gcct gggccc      4200
t gagct cggc  cacacact ca  ct t ct ct t cc  t t gggat ccc  t aagaccgt g  gaggagagag      4260
aggcaat caa  t ggct cct t c  acaaaccaga  gaccaaat gt  cacgt t t t gt  t t t gt gccaa      4320
cct at t t t ga  agt accacca  aaaaagct gt  at t t t gaaaa  t gct t t agaa  aggt t t t gag      4380
cat ggg t t ca  t cct at t ct t  t cgaaagaag  aaaat at cat  aaaaat gagt  gat aaat aca      4440
aggcccagat  gt ggt t gcat  aagg t t t t a  t gcat gt t t g  t t gt at act t  cct t at gct t      4500
ct t t t aaat t  gt gt gt gct c  t gct t caat g  t agt cagaat  t agct gct t c  t at gt t t cat      4560
agt t ggggt c  at agat gt t t  cct t gcct t g  t t gat gt gga  cat gagccat  t t gaggggag      4620
agggaacgga  aat aaaggag  t t at t t gt aa  t gaaaaaaaa  aaaaaaaaaa  aaaaaaaaaa      4679

```

<210> 8
 <211> 1310
 <212> PRT
 <213> Homo sapiens

```

<400> 8
Met Asp Gly Phe Ala Gly Ser Leu Asp Asp Ser Ile Ser Ala Ala Ser
1      5      10
Thr Ser Asp Val Gln Asp Arg Leu Ser Ala Leu Glu Ser Arg Val Gln
20     25     30
Gln Gln Glu Asp Glu Ile Thr Val Leu Lys Ala Ala Leu Ala Asp Val
35     40     45
Leu Arg Arg Leu Ala Ile Ser Glu Asp His Val Ala Ser Val Lys Lys
50     55     60
Ser Val Ser Ser Lys Gly Gln Pro Ser Pro Arg Ala Val Ile Pro Met
65     70     75     80
Ser Cys Ile Thr Asn Gly Ser Gly Ala Asn Arg Lys Pro Ser His Thr
85     90     95
Ser Ala Val Ser Ile Ala Gly Lys Glu Thr Leu Ser Ser Ala Ala Lys
100    105    110

```

Ser Gly Thr Gu Lys Lys Lys Gu Lys Pro Gn Gly Gn Arg Gu Lys
 115 120 125
 Lys Gu Gu Ser His Ser Asn Asp Gn Ser Pro Gn Ile Arg Ala Ser
 130 135 140
 Pro Ser Pro Gn Pro Ser Ser Gn Pro Leu Gn Ile His Arg Gn Thr
 145 150 155 160
 Pro Gu Ser Lys Asn Ala Thr Pro Thr Lys Ser Ile Lys Arg Pro Ser
 165 170 175
 Pro Ala Gu Lys Ser His Asn Ser Trp Gu Asn Ser Asp Asp Ser Arg
 180 185 190
 Asn Lys Leu Ser Lys Ile Pro Ser Thr Pro Lys Leu Ile Pro Lys Val
 195 200 205
 Thr Lys Thr Ala Asp Lys His Lys Asp Val Ile Ile Asn Gn Gu Gly
 210 215 220
 Gu Tyr Ile Lys Met Phe Met Arg Gly Arg Pro Ile Thr Met Phe Ile
 225 230 235 240
 Pro Ser Asp Val Asp Asn Tyr Asp Asp Ile Arg Thr Gu Leu Pro Pro
 245 250 255
 Gu Lys Leu Lys Leu Gu Trp Ala Tyr Gly Tyr Arg Gly Lys Asp Cys
 260 265 270
 Arg Ala Asn Val Tyr Leu Leu Pro Thr Gly Gu Ile Val Tyr Phe Ile
 275 280 285
 Ala Ser Val Val Val Leu Phe Asn Tyr Gu Gu Arg Thr Gn Arg His
 290 295 300
 Tyr Leu Gly His Thr Asp Cys Val Lys Cys Leu Ala Ile His Pro Asp
 305 310 315 320
 Lys Ile Arg Ile Ala Thr Gly Gn Ile Ala Gly Val Asp Lys Asp Gly
 325 330 335
 Arg Pro Leu Gn Pro His Val Arg Val Trp Asp Ser Val Thr Leu Ser
 340 345 350
 Thr Leu Gn Ile Ile Gly Leu Gly Thr Phe Gu Arg Gly Val Gly Cys
 355 360 365
 Leu Asp Phe Ser Lys Ala Asp Ser Gly Val His Leu Cys Val Ile Asp
 370 375 380

Asp Ser Asn G u H i s Met Leu Thr Val Trp Asp Trp G n Lys Lys Al a
 385 390 395 400
 Lys G y Al a G u I l e Lys Thr Thr Asn G u Val Val Leu Al a Val G u
 405 410 415
 Phe H i s Pro Thr Asp Al a Asn Thr I l e I l e Thr Cys G y Lys Ser H i s
 420 425 430
 I l e Phe Phe Trp Thr Trp Ser G y Asn Ser Leu Thr Ar g Lys G n G y
 435 440 445
 I l e Phe G y Lys Tyr G u Lys Pro Lys Phe Val G n Cys Leu Al a Phe
 450 455 460
 Leu G y Asn G y Asp Val Leu Thr G y Asp Ser G y G y Val Met Leu
 465 470 475 480
 I l e Trp Ser Lys Thr Thr Val G u Pro Thr Pro G y Lys G y Pro Lys
 485 490 495
 G y Val Tyr G n I l e Ser Lys G n I l e Lys Al a H i s Asp G y Ser Val
 500 505 510
 Phe Thr Leu Cys G n Met Arg Asn G y Met Leu Leu Thr G y G y G y
 515 520 525
 Lys Asp Arg Lys I l e I l e Leu Trp Asp H i s Asp Leu Asn Pro G u Arg
 530 535 540
 G u I l e G u Val Pro Asp G n Tyr G y Thr I l e Arg Al a Val Al a G u
 545 550 555 560
 G y Lys Al a Asp G n Phe Leu Val G y Thr Ser Arg Asn Phe I l e Leu
 565 570 575
 Arg G y Thr Phe Asn Asp G y Phe G n I l e G u Val G n G y H i s Thr
 580 585 590
 Asp G u Leu Trp G y Leu Al a Thr H i s Pro Phe Lys Asp Leu Leu Leu
 595 600 605
 Thr Cys Al a G n Asp Arg G n Val Cys Leu Trp Asn Ser Met G u H i s
 610 615 620
 Arg Leu G u Trp Thr Arg Leu Val Asp G u Pro G y H i s Cys Al a Asp
 625 630 635 640
 Phe H i s Pro Ser G y Thr Val Val Al a I l e G y Thr H i s Ser G y Arg
 645 650 655

Trp Phe Val Leu Asp Ala Glu Thr Arg Asp Leu Val Ser Ile His Thr
 660 665 670
 Asp Gly Asn Glu Gln Leu Ser Val Met Arg Tyr Ser Ile Asp Gly Thr
 675 680 685
 Phe Leu Ala Val Gly Ser His Asp Asn Phe Ile Tyr Leu Tyr Val Val
 690 695 700
 Ser Glu Asn Gly Arg Lys Tyr Ser Arg Tyr Gly Arg Cys Thr Gly His
 705 710 715 720
 Ser Ser Tyr Ile Thr His Leu Asp Trp Ser Pro Asp Asn Lys Tyr Ile
 725 730 735
 Met Ser Asn Ser Gly Asp Tyr Glu Ile Leu Tyr Leu Tyr Arg Arg Lys
 740 745 750
 His Gln Glu Leu Gln Ala Met Gln Met Glu Leu Gln Ser Pro Glu Tyr
 755 760 765
 Lys Leu Ser Lys Leu Arg Thr Ser Thr Ile Met Thr Asp Tyr Asn Pro
 770 775 780
 Asn Tyr Cys Phe Ala Gly Lys Thr Ser Ser Ile Ser Asp Leu Lys Glu
 785 790 795 800
 Val Pro Arg Lys Asn Ile Thr Leu Ile Arg Gly Leu Gly His Gly Ala
 805 810 815
 Phe Gly Glu Val Tyr Glu Gly Gln Val Ser Gly Met Pro Asn Asp Pro
 820 825 830
 Ser Pro Leu Gln Val Ala Val Lys Thr Leu Pro Glu Val Cys Ser Glu
 835 840 845
 Gln Asp Glu Leu Asp Phe Leu Met Glu Ala Leu Ile Ile Ser Lys Phe
 850 855 860
 Asn His Gln Asn Ile Val Arg Cys Ile Gly Val Ser Leu Gln Ser Leu
 865 870 875 880
 Pro Arg Phe Ile Leu Leu Glu Leu Met Ala Gly Gly Asp Leu Lys Ser
 885 890 895
 Phe Leu Arg Glu Thr Arg Pro Arg Pro Ser Gln Pro Ser Ser Leu Ala
 900 905 910
 Met Leu Asp Leu Leu His Val Ala Arg Asp Ile Ala Cys Gly Cys Gln
 915 920 925

P042876Seq. TXT

Tyr Leu Gu Gu Asn His Phe Ile His Arg Asp Ile Ala Ala Arg Asn
930 935 940

Cys Leu Leu Thr Cys Pro Gly Pro Gly Arg Val Ala Lys Ile Gly Asp
945 950 955 960

Phe Gly Met Ala Arg Asp Ile Tyr Arg Ala Ser Tyr Tyr Arg Lys Gly
965 970 975

Gly Cys Ala Met Leu Pro Val Lys Trp Met Pro Pro Gu Ala Phe Met
980 985 990

Gu Gly Ile Phe Thr Ser Lys Thr Asp Thr Trp Ser Phe Gly Val Leu
995 1000 1005

Leu Trp Gu Ile Phe Ser Leu Gly Tyr Met Pro Tyr Pro Ser Lys
1010 1015 1020

Ser Asn Gn Gu Val Leu Gu Phe Val Thr Ser Gly Gly Arg Met
1025 1030 1035

Asp Pro Pro Lys Asn Cys Pro Gly Pro Val Tyr Arg Ile Met Thr
1040 1045 1050

Gn Cys Trp Gn His Gn Pro Gu Asp Arg Pro Asn Phe Ala Ile
1055 1060 1065

Ile Leu Gu Arg Ile Gu Tyr Cys Thr Gn Asp Pro Asp Val Ile
1070 1075 1080

Asn Thr Ala Leu Pro Ile Gu Tyr Gly Pro Leu Val Gu Gu Gu
1085 1090 1095

Gu Lys Val Pro Val Arg Pro Lys Asp Pro Gu Gly Val Pro Pro
1100 1105 1110

Leu Leu Val Ser Gn Gn Ala Lys Arg Gu Gu Gu Arg Ser Pro
1115 1120 1125

Ala Ala Pro Pro Pro Leu Pro Thr Thr Ser Ser Gly Lys Ala Ala
1130 1135 1140

Lys Lys Pro Thr Ala Ala Gu Val Ser Val Arg Val Pro Arg Gly
1145 1150 1155

Pro Ala Val Gu Gly Gly His Val Asn Met Ala Phe Ser Gn Ser
1160 1165 1170

Asn Pro Pro Ser Gu Leu His Arg Val His Gly Ser Arg Asn Lys
1175 1180 1185

P042876Seq. TXT

Pro Thr Ser Leu Trp Asn Pro Thr Tyr Gly Ser Trp Phe Thr Glu
 1190 1195 1200

Lys Pro Thr Lys Lys Asn Asn Pro Ile Ala Lys Lys Glu Pro His
 1205 1210 1215

Glu Arg Gly Asn Leu Gly Leu Glu Gly Ser Cys Thr Val Pro Pro
 1220 1225 1230

Asn Val Ala Thr Gly Arg Leu Pro Gly Ala Ser Leu Leu Leu Glu
 1235 1240 1245

Pro Ser Ser Leu Thr Ala Asn Met Lys Glu Val Pro Leu Phe Arg
 1250 1255 1260

Leu Arg His Phe Pro Cys Gly Asn Val Asn Tyr Gly Tyr Gln Gln
 1265 1270 1275

Gln Gly Leu Pro Leu Glu Ala Ala Thr Ala Pro Gly Ala Gly His
 1280 1285 1290

Tyr Glu Asp Thr Ile Leu Lys Ser Lys Asn Ser Met Asn Gln Pro
 1295 1300 1305

Gly Pro
 1310

<210> 9

<211> 2473

<212> DNA

<213> Homo sapiens

<400> 9

act ct gt cgg t ccgct gaat gaagt gcccg cccct ct aag cccggagccc ggcgct t t cc	60
ccgcaagat g gacggt t t cg ccggcagt ct cgat gat agt at t t ct gct g caagt act t c	120
t gat gt t caa gat cgcct gt cagct ct t ga gt cagcagt t cagcaacaag aagat gaaat	180
cact gt gct a aaggcggct t t ggct gat gt t t t gaggcgt ct t gcaat ct ct gaagat ca	240
t gt ggcct ca gt gaaaaaat cagt ct caag t aaaggccaa ccaagccct c gagcagt t at	300
t cccat gt cc t gt at aacca at ggaagt gg t gcaaacaga aaaccaagt c at accagt gc	360
t gt ct caat t gcaggaaaag aaact ct t t c at ct gct gct aaaagt ggt a cagaaaaaaa	420
gaaagaaaaa ccacaaggac agagagaaaa aaaagaggaa t ct cat t ct a at gat caaag	480
t ccacaaat t cgagcat cac ct t ct ccca gccct ct t ca caacct ct cc aaat acacag	540
acaaact cca gaaagcaaga at gct act cc caccaaaagc at aaaacgac cat caccagc	600
t gaaaagt ca cat aat t ct t gggaaaat t c agat gat agc cgt aat aaat t gt cgaaaat	660
acct t caaca cccaaat t aa t accaaaagt t accaaaact gcagacaagc at aaagat gt	720
cat cat caac caagt gt acc gccggaagca ccaggagct g caagccat gc agat ggagct	780

P042876Seq. TXT

gcagagccct	gagt acaagc	t gagcaagct	ccgcacct cg	accat cat ga	ccgact acaa	840
ccccaact ac	t gct t t gct g	gcaagacct c	ct ccat cagt	gacct gaagg	aggt gccgcg	900
gaaaaacat c	accct cat t c	gggggt ct ggg	ccat ggagcc	t t t ggggagg	t gt at gaagg	960
ccaggt gt cc	ggaat gccc	acgacccaag	ccccct gcaa	gt ggct gt ga	agacgct gcc	1020
t gaagt gt gc	t ct gaacagg	acgaact gga	t t t cct cat g	gaagccct ga	t cat cagcaa	1080
at t caaccac	cagaacat t g	t t cgct gcat	t ggggt gage	ct gcaat ccc	t gccccggt t	1140
cat cct gct g	gagct cat gg	cgggggggaga	cct caagt cc	t t cct ccgag	agaccgccc	1200
t cgccccgagc	cagccct cct	ccct ggccat	gct ggacct t	ct gcacgt gg	ct cgggacat	1260
t gcct gt ggc	t gt cagt at t	t ggaggaaaa	ccact t cat c	caccgagaca	t t gct gccag	1320
aaact gcct c	t t gacct gt c	caggccct gg	aagagt ggcc	aagat t ggag	act t cgggat	1380
ggccccgagac	at ct acaggg	cgagct act a	t agaaaggga	ggct gt gcca	t gct gccagt	1440
t aagt ggat g	ccccagagg	cct t cat gga	aggaat at t c	act t ct aaaa	cagacacat g	1500
gt cct t t gga	gt gct gct at	gggaaat ct t	t t ct ct t gga	t at at gccat	accccagcaa	1560
aagcaaccag	gaagt t ct gg	agt t t gt cac	cagt ggaggc	cggat ggacc	caccaagaa	1620
ct gccct ggg	cct gt at acc	ggat aat gac	t cagt gct gg	caacat cagc	ct gaagacag	1680
gccaact t t	gccat cat t t	t ggagaggat	t gaat act gc	acccaggacc	cggat gt aat	1740
caacaccgct	t t gccgat ag	aat at ggt cc	act t gt ggaa	gaggaagaga	aagt gcct gt	1800
gaggcccaag	gacct gagg	gggt t cct cc	t ct cct ggt c	t ct caacagg	caaaacggga	1860
ggaggagcgc	agcccagct g	ccccaccacc	t ct gcct acc	acct cct ct g	gcaaggct gc	1920
aaagaaaccc	acagct gcag	aggt ct ct gt	t cgagt ccct	agagggccgg	ccgt ggaagg	1980
gggacacgt g	aat at ggcat	t ct ct cagt c	caaccct cct	t cggagt t gc	acagggt cca	2040
cggat ccaga	aacaagccca	ccagct t gt g	gaaccaacg	t acggct cct	gggt t t acaga	2100
gaaaccacc	aaaaagaat a	at cct at agc	aaagaaggag	ccacacgaga	gggggt aacct	2160
ggggct ggag	ggaagct gt a	ct gt cccacc	t aacgt t gca	act gggagac	t t ccgggggc	2220
ct cact gct c	ct agagccct	ct t cgct gac	t gccaat at g	aaggaggt ac	ct ct gt t cag	2280
gct acgt cac	t t cct t gt g	ggaat gt caa	t t acggct ac	cagcaacagg	gct t gccct t	2340
agaagccgct	act gccct g	gagct ggt ca	t t acgaggat	accat t ct ga	aaagcaagaa	2400
t agcat gaac	cagcct gggc	cct gagct cg	gt cgcacact	cact t ct ct t	cct t gggat c	2460
cct aagaccg	t gg					2473

<210> 10
 <211> 785
 <212> PRT
 <213> Homo sapiens

<400> 10
 Met Asp Gly Phe Ala Gly Ser Leu Asp Asp Ser Ile Ser Ala Ala Ser
 1 5 10 15

P042876Seq. TXT

Thr Ser Asp Val Gln Asp Arg Leu Ser Ala Leu Glu Ser Arg Val Gln
20 25 30

Gln Gln Glu Asp Glu Ile Thr Val Leu Lys Ala Ala Leu Ala Asp Val
35 40 45

Leu Arg Arg Leu Ala Ile Ser Glu Asp His Val Ala Ser Val Lys Lys
50 55 60

Ser Val Ser Ser Lys Gly Gln Pro Ser Pro Arg Ala Val Ile Pro Met
65 70 75 80

Ser Cys Ile Thr Asn Gly Ser Gly Ala Asn Arg Lys Pro Ser His Thr
85 90 95

Ser Ala Val Ser Ile Ala Gly Lys Glu Thr Leu Ser Ser Ala Ala Lys
100 105 110

Ser Gly Thr Glu Lys Lys Lys Glu Lys Pro Gln Gly Gln Arg Glu Lys
115 120 125

Lys Glu Glu Ser His Ser Asn Asp Gln Ser Pro Gln Ile Arg Ala Ser
130 135 140

Pro Ser Pro Gln Pro Ser Ser Gln Pro Leu Gln Ile His Arg Gln Thr
145 150 155 160

Pro Glu Ser Lys Asn Ala Thr Pro Thr Lys Ser Ile Lys Arg Pro Ser
165 170 175

Pro Ala Glu Lys Ser His Asn Ser Trp Glu Asn Ser Asp Asp Ser Arg
180 185 190

Asn Lys Leu Ser Lys Ile Pro Ser Thr Pro Lys Leu Ile Pro Lys Val
195 200 205

Thr Lys Thr Ala Asp Lys His Lys Asp Val Ile Ile Asn Gln Val Tyr
210 215 220

Arg Arg Lys His Gln Glu Leu Gln Ala Met Gln Met Glu Leu Gln Ser
225 230 235 240

Pro Glu Tyr Lys Leu Ser Lys Leu Arg Thr Ser Thr Ile Met Thr Asp
245 250 255

Tyr Asn Pro Asn Tyr Cys Phe Ala Gly Lys Thr Ser Ser Ile Ser Asp
260 265 270

Leu Lys Glu Val Pro Arg Lys Asn Ile Thr Leu Ile Arg Gly Leu Gly
275 280 285

P042876Seq. TXT

H i s G l y A l a P h e G l y G u V a l T y r G u G l y G n V a l S e r G l y M e t P r o
 290 295 300
 A s n A s p P r o S e r P r o L e u G n V a l A l a V a l L y s T h r L e u P r o G l u V a l
 305 310 315 320
 C y s S e r G l u G n A s p G l u L e u A s p P h e L e u M e t G l u A l a L e u I l e I l e
 325 330 335
 S e r L y s P h e A s n H i s G n A s n I l e V a l A r g C y s I l e G l y V a l S e r L e u
 340 345 350
 G n S e r L e u P r o A r g P h e I l e L e u L e u G l u L e u M e t A l a G l y G l y A s p
 355 360 365
 L e u L y s S e r P h e L e u A r g G l u T h r A r g P r o A r g P r o S e r G n P r o S e r
 370 375 380
 S e r L e u A l a M e t L e u A s p L e u L e u H i s V a l A l a A r g A s p I l e A l a C y s
 385 390 395 400
 G l y C y s G n T y r L e u G l u G l u A s n H i s P h e I l e H i s A r g A s p I l e A l a
 405 410 415
 A l a A r g A s n C y s L e u L e u T h r C y s P r o G l y P r o G l y A r g V a l A l a L y s
 420 425 430
 I l e G l y A s p P h e G l y M e t A l a A r g A s p I l e T y r A r g A l a S e r T y r T y r
 435 440 445
 A r g L y s G l y G l y C y s A l a M e t L e u P r o V a l L y s T r p M e t P r o P r o G l u
 450 455 460
 A l a P h e M e t G l u G l y I l e P h e T h r S e r L y s T h r A s p T h r T r p S e r P h e
 465 470 475 480
 G l y V a l L e u L e u T r p G l u I l e P h e S e r L e u G l y T y r M e t P r o T y r P r o
 485 490 495
 S e r L y s S e r A s n G n G l u V a l L e u G l u P h e V a l T h r S e r G l y G l y A r g
 500 505 510
 M e t A s p P r o P r o L y s A s n C y s P r o G l y P r o V a l T y r A r g I l e M e t T h r
 515 520 525
 G n C y s T r p G n H i s G n P r o G l u A s p A r g P r o A s n P h e A l a I l e I l e
 530 535 540
 L e u G l u A r g I l e G l u T y r C y s T h r G n A s p P r o A s p V a l I l e A s n T h r
 545 550 555 560

P042876Seq. TXT

Ala Leu Pro Ile Glu Tyr Gly Pro Leu Val Glu Glu Glu Glu Lys Val
565 570 575

Pro Val Arg Pro Lys Asp Pro Glu Gly Val Pro Pro Leu Leu Val Ser
580 585 590

Gln Gln Ala Lys Arg Glu Glu Glu Arg Ser Pro Ala Ala Pro Pro Pro
595 600 605

Leu Pro Thr Thr Ser Ser Gly Lys Ala Ala Lys Lys Pro Thr Ala Ala
610 615 620

Glu Val Ser Val Arg Val Pro Arg Gly Pro Ala Val Glu Gly Gly His
625 630 635 640

Val Asn Met Ala Phe Ser Gln Ser Asn Pro Pro Ser Glu Leu His Arg
645 650 655

Val His Gly Ser Arg Asn Lys Pro Thr Ser Leu Trp Asn Pro Thr Tyr
660 665 670

Gly Ser Trp Phe Thr Glu Lys Pro Thr Lys Lys Asn Asn Pro Ile Ala
675 680 685

Lys Lys Glu Pro His Glu Arg Gly Asn Leu Gly Leu Glu Gly Ser Cys
690 695 700

Thr Val Pro Pro Asn Val Ala Thr Gly Arg Leu Pro Gly Ala Ser Leu
705 710 715 720

Leu Leu Glu Pro Ser Ser Leu Thr Ala Asn Met Lys Glu Val Pro Leu
725 730 735

Phe Arg Leu Arg His Phe Pro Cys Gly Asn Val Asn Tyr Gly Tyr Gln
740 745 750

Gln Gln Gly Leu Pro Leu Glu Ala Ala Thr Ala Pro Gly Ala Gly His
755 760 765

Tyr Glu Asp Thr Ile Leu Lys Ser Lys Asn Ser Met Asn Gln Pro Gly
770 775 780

Pro
785

<210> 11
<211> 2506
<212> DNA
<213> Homo sapiens

<400> 11
act ct gt cgg t ccgct gaat gaagt gcccg ccct ct aag cccggagccc ggcgct t t cc
Page 28

P042876Seq. TXT

ccgcaagat g	gacggt t t cg	ccggcagt ct	cgat gat agt	at t t ct gct g	caagt act t c	120
t gat gt t caa	gat cgcct gt	cagct ct t ga	gt cagcagt t	cagcaacaag	aagat gaaat	180
cact gt gct a	aaggcggct t	t ggct gat gt	t t t gaggcgt	ct t gcaat ct	ct gaagat ca	240
t gt ggcct ca	gt gaaaaaat	cagt ct caag	t aaaggccaa	ccaagccct c	gagcagt t at	300
t cccat gt cc	t gt at aacca	at ggaagt gg	t gcaaacaga	aaaccaagt c	at accagt gc	360
t gt ct caat t	gcaggaaaag	aaact ct t t c	at ct gct gct	aaaagt ggt a	cagaaaaaaa	420
gaaagaaaaa	ccacaaggac	agagagaaaa	aaaagaggaa	t ct cat t ct a	at gat caaag	480
t ccacaaat t	cgagcat cac	ct t ct ccca	gccct ct t ca	caacct ct cc	aaat acacag	540
acaaaact cca	gaaagcaaga	at gct act cc	cacaaaaagc	at aaaacgac	cat caccagc	600
t gaaaagt ca	cat aat t ct t	gggaaaat t c	agat gat agc	cgt aat aaat	t gt cgaaaat	660
acct t caaca	cccaaat t aa	t accaaaagt	t accaaaact	gcagacaagc	at aaagat gt	720
cat cat caac	caagcaaaaa	t gt caact cg	cgaaaaaaac	agccaagt gt	accgccggaa	780
gcaccaggag	ct gcaagcca	t gcagat gga	gct gcagagc	cct gagt aca	agct gagcaa	840
gct ccgcacc	t cgacat ca	t gaccgact a	caaccccaac	t act gct t t g	ct ggcaagac	900
ct cct ccat c	agt gacct ga	aggaggt gcc	gcggaaaaac	at caccct ca	t t cgggggt ct	960
gggccat gga	gcct t t gggg	aggt gt at ga	aggccaggt g	t ccggaat gc	ccaacgaccc	1020
aagccccct g	caagt ggct g	t gaagacgct	gcct gaagt g	t gct ct gaac	aggacgaact	1080
ggat t t cct c	at ggaagccc	t gat cat cag	caaat t caac	caccagaaca	t t gt t cgct g	1140
cat t ggggt g	agcct gcaat	ccct gccccg	gt t cat cct g	ct ggagct ca	t ggcgggggg	1200
agacct caag	t cct t cct cc	gagagacccg	ccct cgccccg	agccagccct	cct ccct ggc	1260
cat gct ggac	ct t ct gcacg	t ggct cggga	cat t gcct gt	ggct gt cagt	at t t ggagga	1320
aaaccact t c	at ccaccgag	acat t gct gc	cagaaact gc	ct ct t gacct	gt ccaggccc	1380
t ggaagagt g	gccaagat t g	gagact t cgg	gat ggccccga	gacat ct aca	gggcgagct a	1440
ct at agaaag	ggaggct gt g	ccat gct gcc	agt t aagt gg	at gccccag	aggcct t cat	1500
ggaaggaat a	t t cact t ct a	aaacagacac	at ggt cct t t	ggagt gct gc	t at gggaaat	1560
ct t t t ct ct t	ggat at at gc	cat accccag	caaaagcaac	caggaagt t c	t ggagt t t gt	1620
caccagt gga	ggccggat gg	accacccaa	gaact gccct	gggcct gt at	accggat aat	1680
gact cagt gc	t ggcaacat c	agcct gaaga	caggcccaac	t t t gccat ca	t t t t ggagag	1740
gat t gaat ac	t gcaccag	accggat gt	aat caacacc	gct t t gccga	t agaat at gg	1800
t ccaact t gt g	gaagaggaag	agaaagt gcc	t gt gaggccc	aaggaccct g	aggggggt t cc	1860
t cct ct cct g	gt ct ct caac	aggcaaaacg	ggaggaggag	cgcagcccag	ct gccccacc	1920
acct ct gcct	accacct cct	ct ggcaaggc	t gcaaagaaa	cccacagct g	cagaggt ct c	1980
t gt t cgagt c	cct agagggc	cggccgt gga	agggggacac	gt gaat at gg	cat t ct ct ca	2040
gt ccaaccct	cct t cggagt	t gcacagggt	ccacggat cc	agaaacaagc	ccaccagct t	2100

P042876Seq. TXT

gt ggaaccca acgt acggct cct ggt t t ac agagaaaccc accaaaaaga at aat cct at 2160
agcaaagaag gagccacacg agaggggt aa cct ggggct g gagggaagct gt act gt ccc 2220
acct aacgt t gcaact ggga gact t ccggg ggcct cact g ct cct agagc cct ct t cgct 2280
gact gccaat at gaaggagg t acct ct gt t caggct acgt cact t ccct t gt gggaat gt 2340
caat t acggc t accagcaac agggct t gcc ct t agaagcc gct act gccc ct ggagct gg 2400
t cat t acgag gat accat t c t gaaaagcaa gaat agcat g aaccagcct g ggccct gage 2460
t cggg cgcac act cact t ct ct t cct t ggg at ccct aaga ccgt gg 2506

<210> 12

<211> 796

<212> PRT

<213> Homo sapiens

<400> 12

Met Asp Gly Phe Ala Gly Ser Leu Asp Asp Ser Ile Ser Ala Ala Ser
1 5 10 15

Thr Ser Asp Val Gln Asp Arg Leu Ser Ala Leu Glu Ser Arg Val Gln
20 25 30

Gln Gln Glu Asp Glu Ile Thr Val Leu Lys Ala Ala Leu Ala Asp Val
35 40 45

Leu Arg Arg Leu Ala Ile Ser Glu Asp His Val Ala Ser Val Lys Lys
50 55 60

Ser Val Ser Ser Lys Gly Gln Pro Ser Pro Arg Ala Val Ile Pro Met
65 70 75 80

Ser Cys Ile Thr Asn Gly Ser Gly Ala Asn Arg Lys Pro Ser His Thr
85 90 95

Ser Ala Val Ser Ile Ala Gly Lys Glu Thr Leu Ser Ser Ala Ala Lys
100 105 110

Ser Gly Thr Glu Lys Lys Lys Glu Lys Pro Gln Gly Gln Arg Glu Lys
115 120 125

Lys Glu Glu Ser His Ser Asn Asp Gln Ser Pro Gln Ile Arg Ala Ser
130 135 140

Pro Ser Pro Gln Pro Ser Ser Gln Pro Leu Gln Ile His Arg Gln Thr
145 150 155 160

Pro Glu Ser Lys Asn Ala Thr Pro Thr Lys Ser Ile Lys Arg Pro Ser
165 170 175

Pro Ala Glu Lys Ser His Asn Ser Trp Glu Asn Ser Asp Asp Ser Arg
180 185 190

P042876Seq. TXT

Asn Lys Leu Ser Lys Ile Pro Ser Thr Pro Lys Leu Ile Pro Lys Val
 195 200 205
 Thr Lys Thr Ala Asp Lys His Lys Asp Val Ile Ile Asn Gl n Ala Lys
 210 215 220
 Met Ser Thr Arg Gl u Lys Asn Ser Gl n Val Tyr Arg Arg Lys His Gl n
 225 230 235 240
 Gl u Leu Gl n Ala Met Gl n Met Gl u Leu Gl n Ser Pro Gl u Tyr Lys Leu
 245 250 255
 Ser Lys Leu Arg Thr Ser Thr Ile Met Thr Asp Tyr Asn Pro Asn Tyr
 260 265 270
 Cys Phe Ala Gly Lys Thr Ser Ser Ile Ser Asp Leu Lys Gl u Val Pro
 275 280 285
 Arg Lys Asn Ile Thr Leu Ile Arg Gly Leu Gly His Gly Ala Phe Gly
 290 295 300
 Gl u Val Tyr Gl u Gly Gl n Val Ser Gly Met Pro Asn Asp Pro Ser Pro
 305 310 315 320
 Leu Gl n Val Ala Val Lys Thr Leu Pro Gl u Val Cys Ser Gl u Gl n Asp
 325 330 335
 Gl u Leu Asp Phe Leu Met Gl u Ala Leu Ile Ile Ser Lys Phe Asn His
 340 345 350
 Gl n Asn Ile Val Arg Cys Ile Gly Val Ser Leu Gl n Ser Leu Pro Arg
 355 360 365
 Phe Ile Leu Leu Gl u Leu Met Ala Gly Gly Asp Leu Lys Ser Phe Leu
 370 375 380
 Arg Gl u Thr Arg Pro Arg Pro Ser Gl n Pro Ser Ser Leu Ala Met Leu
 385 390 395 400
 Asp Leu Leu His Val Ala Arg Asp Ile Ala Cys Gly Cys Gl n Tyr Leu
 405 410 415
 Gl u Gl u Asn His Phe Ile His Arg Asp Ile Ala Ala Arg Asn Cys Leu
 420 425 430
 Leu Thr Cys Pro Gly Pro Gly Arg Val Ala Lys Ile Gly Asp Phe Gly
 435 440 445
 Met Ala Arg Asp Ile Tyr Arg Ala Ser Tyr Tyr Arg Lys Gly Gly Cys
 450 455 460

P042876Seq. TXT

Ala Met Leu Pro Val Lys Trp Met Pro Pro Glu Ala Phe Met Glu Gly
465 470 475 480

Ile Phe Thr Ser Lys Thr Asp Thr Trp Ser Phe Gly Val Leu Leu Trp
485 490 495

Glu Ile Phe Ser Leu Gly Tyr Met Pro Tyr Pro Ser Lys Ser Asn Gln
500 505 510

Glu Val Leu Glu Phe Val Thr Ser Gly Gly Arg Met Asp Pro Pro Lys
515 520 525

Asn Cys Pro Gly Pro Val Tyr Arg Ile Met Thr Gln Cys Trp Gln His
530 535 540

Gln Pro Glu Asp Arg Pro Asn Phe Ala Ile Ile Leu Glu Arg Ile Glu
545 550 555 560

Tyr Cys Thr Gln Asp Pro Asp Val Ile Asn Thr Ala Leu Pro Ile Glu
565 570 575

Tyr Gly Pro Leu Val Glu Glu Glu Glu Lys Val Pro Val Arg Pro Lys
580 585 590

Asp Pro Glu Gly Val Pro Pro Leu Leu Val Ser Gln Gln Ala Lys Arg
595 600 605

Glu Glu Glu Arg Ser Pro Ala Ala Pro Pro Pro Leu Pro Thr Thr Ser
610 615 620

Ser Gly Lys Ala Ala Lys Lys Pro Thr Ala Ala Glu Val Ser Val Arg
625 630 635 640

Val Pro Arg Gly Pro Ala Val Glu Gly Gly His Val Asn Met Ala Phe
645 650 655

Ser Gln Ser Asn Pro Pro Ser Glu Leu His Arg Val His Gly Ser Arg
660 665 670

Asn Lys Pro Thr Ser Leu Trp Asn Pro Thr Tyr Gly Ser Trp Phe Thr
675 680 685

Glu Lys Pro Thr Lys Lys Asn Asn Pro Ile Ala Lys Lys Glu Pro His
690 695 700

Glu Arg Gly Asn Leu Gly Leu Glu Gly Ser Cys Thr Val Pro Pro Asn
705 710 715 720

Val Ala Thr Gly Arg Leu Pro Gly Ala Ser Leu Leu Leu Glu Pro Ser
725 730 735

P042876Seq. TXT

Ser Leu Thr Ala Asn Met Lys Glu Val Pro Leu Phe Arg Leu Arg His
740 745 750

Phe Pro Cys Gly Asn Val Asn Tyr Gly Tyr Gln Gln Gln Gly Leu Pro
755 760 765

Leu Glu Ala Ala Thr Ala Pro Gly Ala Gly His Tyr Glu Asp Thr Ile
770 775 780

Leu Lys Ser Lys Asn Ser Met Asn Gln Pro Gly Pro
785 790 795

<210> 13

<211> 3409

<212> DNA

<213> Homo sapiens

<400> 13

act ct gt cgg t ccgct gaat gaagt gcccg cccct ct aag cccggagccc ggcgct t t cc	60
ccgcaagat g gacggt t t cg ccggcagt ct cgat gat agt at t t ct gct g caagt act t c	120
t gat gt t caa gat cgcct gt cagct ct t ga gt caccgagt t cagcaacaag aagat gaaat	180
cact gt gct a aaggcggct t t ggct gat gt t t t gaggcgt ct t gcaat ct ct gaagat ca	240
t gt ggcct ca gt gaaaaaat cagt ct caag t aaaggccaa ccaagccct c gagcagt t at	300
t cccat gt cc t gt at aacca at ggaagt gg t gcaaacaga aaaccaagt c at accagt gc	360
t gt ct caat t gcaggaaaag aaact ct t t c at ct gct gct aaaagt ggt a cagaaaaaaa	420
gaaagaaaaa ccacaaggac agagagaaaa aaaagaggaa t ct cat t ct a at gat caaag	480
t ccacaaat t cgagcat cac ct t ct ccca gccct ct t ca caacct ct cc aaat acacag	540
acaaact cca gaaagcaaga at gct act cc caccaaaagc at aaaacgac cat caccagc	600
t gaaaagt ca cat aat t ct t gggaaaat t c agat gat agc cgt aat aaat t gt cgaaaat	660
acct t caaca cccaaat t aa t accaaaagt t accaaaact gcagacaagc at aaagat gt	720
cat cat caac caagaaggag aat at at t aa aat gt t t at g cgcggt cggc caat t accat	780
gt t cat t cct t ccgat gt t g acaact at ga t gacat caga acggaact gc ct cct gagaa	840
gct caaact g gagt gggcat at ggt t at cg aggaaaggac t gt agagct a at gt t t acct	900
t ct t ccgacc ggggaaat ag t t t at t t cat t gcat cagt a gt agt act at t t aat t at ga	960
ggagagaact cagcgacact acct gggcca t acagact gt gt gaaat gcc t t gct at aca	1020
t cct gacaaa at t aggat t g caact ggaca gat agct ggc gt ggat aaag at ggaaggcc	1080
t ct acaacc cact cagag t gt gggat t c t gt t act ct a t ccacact gc agat t at t gg	1140
act t ggcact t t t gagcgt g gagt aggat g cct ggat t t t t caaaagcag at t caggt gt	1200
t cat t t at gt gt t at t gat g act ccaat ga gcat at gct t act gt at ggg act ggcagag	1260
gaaagcaaaa ggagcagaaa t aaagacaac aaat gaagt t gt t t t ggct g t ggagt t t ca	1320

P042876Seq. TXT

cccaacagat	gcaaat acca	t aat t acat g	cggt aaat ct	cat at t t t ct	t ct ggacct g	1380
gagcggcaat	t cact aacaa	gaaaacaggg	aat t t t t ggg	aaat at gaaa	agccaaaat t	1440
t gt gcagt gt	t t agcat t ct	t ggggaat gg	agat gt t ct t	act ggagact	caggt ggagt	1500
cat gct t at a	t ggagcaaaa	ct act gt aga	gcccacacct	gggaaaggac	ct aaaggt gt	1560
at at caaat c	agcaaacaaa	t caaagct ca	t gat ggcagt	gt gt t cacac	t t t gt cagat	1620
gagaaat ggg	at gt t at t aa	ct ggaggagg	gaaagacaga	aaaat aat t c	t gt gggat ca	1680
t gat ct gaat	cct gaaagag	aaat agagat	at gct ggat g	agccct gagt	acaagct gag	1740
caagct ccgc	acct cgacca	t cat gaccga	ct acaacccc	aact act gct	t t gct ggcaa	1800
gacct cct cc	at cagt gacc	t gaaggaggt	gccgcggaaa	aacat cacc	t cat t cgggg	1860
t ct gggccat	ggagcct t t g	gggaggt gt a	t gaaggccag	gt gt ccggaa	t gccaacga	1920
cccaagcccc	ct gcaagt gg	ct gt gaagac	gct gcct gaa	gt gt gct ct g	aacaggacga	1980
act ggat t t c	ct cat ggaag	ccct gat cat	cagcaaat t c	aaccaccaga	acat t gt t cg	2040
ct gcat t ggg	gt ggcct gc	aat ccct gcc	ccggt t cat c	ct gct ggagc	t cat ggcggg	2100
gggagacct c	aagt cct t cc	t ccgagagac	ccgccct cgc	ccgagccagc	cct cct ccct	2160
ggccat gct g	gacct t ct gc	acgt ggct cg	ggacat t gcc	t gt ggct gt c	agt at t t gga	2220
ggaaaaccac	t t cat ccacc	gagacat t gc	t gccagaaac	t gcct ct t ga	cct gt ccagg	2280
ccct ggaaga	gt ggccaaga	t t ggagact t	cgggat ggcc	cgagacat ct	acagggcgag	2340
ct act at aga	aagggaggct	gt gccat gct	gccagt t aag	t ggat gcccc	cagaggcct t	2400
cat ggaagga	at at t cact t	ct aaaacaga	cacat ggt cc	t t t ggagt gc	t gct at ggga	2460
aat ct t t t ct	ct t ggat at a	t gccat accc	cagcaaaaagc	aaccaagaag	t t ct ggagt t	2520
t gt caccagt	ggaggccgga	t ggaccacc	caagaact gc	cct gggcct g	t at accggat	2580
aat gact cag	t gct ggcaac	at cagcct ga	agacaggccc	aact t t gcca	t cat t t t gga	2640
gaggat t gaa	t act gcacc	aggacccgga	t gt aat caac	accgct t t gc	cgat agaat a	2700
t ggt ccaact t	gt ggaagagg	aagagaaagt	gcct gt gagg	cccaaggacc	ct gagggggt	2760
t cct cct ct c	ct ggt ct ct c	aacaggcaaa	acgggaggag	gagcgcagcc	cagct gcccc	2820
accacct ct g	cct accacct	cct ct ggcaa	ggct gcaaag	aaacccacag	ct gcagaggt	2880
ct ct gt t cga	gt ccct agag	ggccggccgt	ggaaggggga	cacgt gaat a	t ggcat t ct c	2940
t cagt ccaac	cct cct t cgg	agt t gcacag	ggc ccacgga	t ccagaaaca	agcccaccag	3000
ct t gt ggaac	ccaacgt acg	gct cct ggt t	t acagagaaa	cccacaaaaa	agaat aat cc	3060
t at agcaaag	aaggagccac	acgagagggg	t aacct gggg	ct ggagggaa	gct gt act gt	3120
cccacct aac	gt t gcaact g	ggagact t cc	gggggcct ca	ct gct cct ag	agccct ct t c	3180
gct gact gcc	aat at gaagg	aggt acct ct	gt t caggct a	cgt cact t cc	ct t gt gggaa	3240
t gt caat t ac	ggct accagc	aacagggct t	gccct t agaa	gccgct act g	cccct ggagc	3300
t ggt cat t ac	gaggat acca	t t ct gaaaag	caagaat agc	at gaaccagc	ct gggccct g	3360

agct cggg cgcacact cact tct ctt cct t gggat ccct a agaccgt gg

3409

<210> 14

<211> 1097

<212> PRT

<213> Homo sapiens

<400> 14

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

P042876Seq. TXT

Pro Ser Asp Val Asp₂₄₅ Asn Tyr Asp Asp Ile₂₅₀ Arg Thr Glu Leu Pro₂₅₅ Pro
 Glu Lys Leu Lys₂₆₀ Leu Glu Trp Ala Tyr₂₆₅ Gly Tyr Arg Gly Lys₂₇₀ Asp Cys
 Arg Ala Asn₂₇₅ Val Tyr Leu Leu Pro₂₈₀ Thr Gly Glu Ile₂₈₅ Val Tyr Phe Ile
 Ala Ser₂₉₀ Val Val Val Leu Phe₂₉₅ Asn Tyr Glu Glu Arg₃₀₀ Thr Gln Arg His
 Tyr₃₀₅ Leu Gly His Thr Asp₃₁₀ Cys Val Lys Cys Leu₃₁₅ Ala Ile His Pro Asp₃₂₀
 Lys Ile Arg Ile Ala₃₂₅ Thr Gly Gln Ile Ala₃₃₀ Gly Val Asp Lys Asp₃₃₅ Gly
 Arg Pro Leu Gln₃₄₀ Pro His Val Arg Val₃₄₅ Trp Asp Ser Val Thr₃₅₀ Leu Ser
 Thr Leu Gln₃₅₅ Ile Ile Gly Leu Gly₃₆₀ Thr Phe Glu Arg Gly₃₆₅ Val Gly Cys
 Leu Asp₃₇₀ Phe Ser Lys Ala Asp₃₇₅ Ser Gly Val His Leu Cys Val Ile Asp
 Asp₃₈₅ Ser Asn Glu His Met₃₉₀ Leu Thr Val Trp Asp₃₉₅ Trp Gln Arg Lys Ala₄₀₀
 Lys Gly Ala Glu₄₀₅ Ile Lys Thr Thr Asn Glu₄₁₀ Val Val Leu Ala₄₁₅ Glu
 Phe His Pro Thr₄₂₀ Asp Ala Asn Thr Ile₄₂₅ Ile Thr Cys Gly Lys₄₃₀ Ser His
 Ile Phe Phe₄₃₅ Trp Thr Trp Ser Gly₄₄₀ Asn Ser Leu Thr Arg₄₄₅ Lys Gln Gly
 Ile Phe₄₅₀ Gly Lys Tyr Glu Lys₄₅₅ Pro Lys Phe Val Gln₄₆₀ Cys Leu Ala Phe
 Leu Gly Asn Gly Asp₄₇₀ Val Leu Thr Gly Asp₄₇₅ Ser Gly Gly Val Met Leu₄₈₀
 Ile Trp Ser Lys Thr₄₈₅ Thr Val Glu Pro Thr₄₉₀ Pro Gly Lys Gly Pro₄₉₅ Lys
 Gly Val Tyr Gln₅₀₀ Ile Ser Lys Gln Ile₅₀₅ Lys Ala His Asp₅₁₀ Gly Ser Val

P042876Seq. TXT

Phe Thr Leu Cys Gln Met Arg Asn Gly Met Leu Leu Thr Gly Gly Gly
 515 520 525
 Lys Asp Arg Lys Ile Ile Leu Trp Asp His Asp Leu Asn Pro Glu Arg
 530 535 540
 Glu Ile Glu Ile Cys Trp Met Ser Pro Glu Tyr Lys Leu Ser Lys Leu
 545 550 555 560
 Arg Thr Ser Thr Ile Met Thr Asp Tyr Asn Pro Asn Tyr Cys Phe Ala
 565 570 575
 Gly Lys Thr Ser Ser Ile Ser Asp Leu Lys Glu Val Pro Arg Lys Asn
 580 585 590
 Ile Thr Leu Ile Arg Gly Leu Gly His Gly Ala Phe Gly Glu Val Tyr
 595 600 605
 Glu Gly Gln Val Ser Gly Met Pro Asn Asp Pro Ser Pro Leu Gln Val
 610 615 620
 Ala Val Lys Thr Leu Pro Glu Val Cys Ser Glu Gln Asp Glu Leu Asp
 625 630 635 640
 Phe Leu Met Glu Ala Leu Ile Ile Ser Lys Phe Asn His Gln Asn Ile
 645 650 655
 Val Arg Cys Ile Gly Val Ser Leu Gln Ser Leu Pro Arg Phe Ile Leu
 660 665 670
 Leu Glu Leu Met Ala Gly Gly Asp Leu Lys Ser Phe Leu Arg Glu Thr
 675 680 685
 Arg Pro Arg Pro Ser Gln Pro Ser Ser Leu Ala Met Leu Asp Leu Leu
 690 695 700
 His Val Ala Arg Asp Ile Ala Cys Gly Cys Gln Tyr Leu Glu Glu Asn
 705 710 715 720
 His Phe Ile His Arg Asp Ile Ala Ala Arg Asn Cys Leu Leu Thr Cys
 725 730 735
 Pro Gly Pro Gly Arg Val Ala Lys Ile Gly Asp Phe Gly Met Ala Arg
 740 745 750
 Asp Ile Tyr Arg Ala Ser Tyr Tyr Arg Lys Gly Gly Cys Ala Met Leu
 755 760 765
 Pro Val Lys Trp Met Pro Pro Glu Ala Phe Met Glu Gly Ile Phe Thr
 770 775 780

P042876Seq. TXT

Ser Lys Thr Asp Thr Trp Ser Phe Gly Val Leu Leu Trp Glu Ile Phe
 785 790 795 800
 Ser Leu Gly Tyr Met Pro Tyr Pro Ser Lys Ser Asn Glu Val Leu
 805 810 815
 Glu Phe Val Thr Ser Gly Gly Arg Met Asp Pro Pro Lys Asn Cys Pro
 820 825 830
 Gly Pro Val Tyr Arg Ile Met Thr Glu Cys Trp Glu His Glu Pro Glu
 835 840 845
 Asp Arg Pro Asn Phe Ala Ile Ile Leu Glu Arg Ile Glu Tyr Cys Thr
 850 855 860
 Glu Asp Pro Asp Val Ile Asn Thr Ala Leu Pro Ile Glu Tyr Gly Pro
 865 870 875 880
 Leu Val Glu Glu Glu Glu Lys Val Pro Val Arg Pro Lys Asp Pro Glu
 885 890 895
 Gly Val Pro Pro Leu Leu Val Ser Glu Glu Ala Lys Arg Glu Glu Glu
 900 905 910
 Arg Ser Pro Ala Ala Pro Pro Pro Leu Pro Thr Thr Ser Ser Gly Lys
 915 920 925
 Ala Ala Lys Lys Pro Thr Ala Ala Glu Val Ser Val Arg Val Pro Arg
 930 935 940
 Gly Pro Ala Val Glu Gly Gly His Val Asn Met Ala Phe Ser Glu Ser
 945 950 955 960
 Asn Pro Pro Ser Glu Leu His Arg Val His Gly Ser Arg Asn Lys Pro
 965 970 975
 Thr Ser Leu Trp Asn Pro Thr Tyr Gly Ser Trp Phe Thr Glu Lys Pro
 980 985 990
 Thr Lys Lys Asn Asn Pro Ile Ala Lys Lys Glu Pro His Glu Arg Gly
 995 1000 1005
 Asn Leu Gly Leu Glu Gly Ser Cys Thr Val Pro Pro Asn Val Ala
 1010 1015 1020
 Thr Gly Arg Leu Pro Gly Ala Ser Leu Leu Leu Glu Pro Ser Ser
 1025 1030 1035
 Leu Thr Ala Asn Met Lys Glu Val Pro Leu Phe Arg Leu Arg His
 1040 1045 1050

P042876Seq. TXT

Phe Pro Cys Gly Asn Val Asn Tyr Gly Tyr Gln Gln Gln Gly Leu
1055 1060 1065

Pro Leu Glu Ala Ala Thr Ala Pro Gly Ala Gly His Tyr Glu Asp
1070 1075 1080

Thr Ile Leu Lys Ser Lys Asn Ser Met Asn Gln Pro Gly Pro
1085 1090 1095

<210> 15
<211> 2014
<212> DNA
<213> Homo sapiens

<400> 15
act ct gt cgg t ccgct gaat gaagt gcccg cccct ct aag cccggagccc ggcgct t t cc 60
ccgcaagat g gacggt t t cg ccggcagt ct cgat gat agt at t t ct gct g caagt act t c 120
t gat gt t caa gat cgcct gt cagct ct t ga gt cagcagt t cagcaacaag aagat gaaat 180
cact gt gct a aaggcggct t t ggct gat gt t t t gaggcgt ct t gcaat ct ct gaagat ca 240
t gt gccct ca gt gaaaaaat cagt ct caag t aaagt gt ac cgccggaagc accaggagct 300
gcaagccat g cagat ggagc t gcagagccc t gagt acaag ct gagcaagc t ccgcacct c 360
gaccat cat g accgact aca accccaact a ct gct t t gct ggcaagacct cct ccat cag 420
t gacct gaag gaggt gccgc ggaaaaacat caccct cat t cgggggt ct gg gccat ggagc 480
ct t t ggggag gt gt at gaag gccaggt gt c cggaat gccc aacgacccaa gccccct gca 540
agt ggct gt g aagacgct gc ct gaagt gt g ct ct gaacag gacgaact gg at t t cct cat 600
ggaagccct g at cat cagca aat t caacca ccagaacat t gt t cgct gca t t ggggt gag 660
cct gcaat cc ct gccccggt t cat cct gct ggagct cat g gcgggggggag acct caagt c 720
ct t cct ccga gagaccgccc ct cgcccgag ccagccct cc t ccct ggcca t gct ggacct 780
t ct gcacgt g gct cgggaca t t gcct gt gg ct gt cagt at t t ggaggaaa accact t cat 840
ccaccgagac at t gct gcca gaaact gcct ct t gacct gt ccaggccct g gaagagt ggc 900
caagat t gga gact t cgga t ggcccgaga cat ct acagg gcgagct act at agaaaggg 960
aggct gt gcc at gct gccag t t aagt ggat gccccagag gcct t cat gg aaggaat at t 1020
cact t ct aaa acagacacat ggt cct t t gg agt gct gct a t gggaaat ct t t t ct ct t gg 1080
at at at gcc a t accccagca aaagcaacca ggaagt t ct g gagt t t gt ca ccagt ggagg 1140
ccggat ggac ccaccaaga act gccct gg gcct gt at ac cggat aat ga ct cagt gct g 1200
gcaacat cag cct gaagaca ggcccaact t t gccat cat t t t ggagagga t t gaat act g 1260
caccaggagc ccggat gt aa t caacaccgc t t t gccgat a gaat at ggt c cact t gt gga 1320
agaggaagag aaagt gcct g t gaggcccaa ggaccct gag ggggt t cct c ct ct cct ggt 1380
ct ct caacag gcaaaacggg aggaggagcg cagcccagct gccccaccac ct ct gcct ac 1440
cacct cct ct ggcaaggct g caaagaaacc cacagct gca gaggt ct ct g t t cgagt ccc 1500

P042876Seq. TXT

```

t agagggccg gccgt ggaag ggggacacgt gaat at ggca t t ct ct cagt ccaaccct cc      1560
t t cggagtt g cacaggggt cc acggat ccag aaacaagccc accagct t gt ggaacccaac      1620
gt acggct cc t ggt t t acag agaaaccac caaaaagaat aat cct at ag caaagaagga      1680
gccacacgag aggggt aacc t ggggct gga gggaagct gt act gt cccac ct aacgt t gc      1740
aact gggaga ct t ccggggg cct cact gct cct agagccc t ct t cgct ga ct gccaat at      1800
gaaggaggt a cct ct gt t ca ggct acgt ca ct t ccct t gt gggaat gt ca at t acggct a      1860
ccagcaacag ggct t gccct t agaagccgc t act gccct ggagct ggt c at t acgagga      1920
t accat t ct g aaaagcaaga at agcat gaa ccagcct ggg ccct gagct c ggt cgcacac      1980
t cact t ct ct t cct t gggat ccct aagacc gt gg                                2014

```

<210> 16
 <211> 632
 <212> PRT
 <213> Homo sapiens

<400> 16

Met Asp Gly Phe Ala Gly Ser Leu Asp Asp Ser Ile Ser Ala Ala Ser
1 5 10 15

Thr Ser Asp Val Gln Asp Arg Leu Ser Ala Leu Glu Ser Arg Val Gln
20 25 30

Gln Gln Glu Asp Glu Ile Thr Val Leu Lys Ala Ala Leu Ala Asp Val
35 40 45

Leu Arg Arg Leu Ala Ile Ser Glu Asp His Val Ala Ser Val Lys Lys
50 55 60

Ser Val Ser Ser Lys Val Tyr Arg Arg Lys His Gln Glu Leu Gln Ala
65 70 75 80

Met Gln Met Glu Leu Gln Ser Pro Glu Tyr Lys Leu Ser Lys Leu Arg
85 90 95

Thr Ser Thr Ile Met Thr Asp Tyr Asn Pro Asn Tyr Cys Phe Ala Gly
100 105 110

Lys Thr Ser Ser Ile Ser Asp Leu Lys Glu Val Pro Arg Lys Asn Ile
115 120 125

Thr Leu Ile Arg Gly Leu Gly His Gly Ala Phe Gly Glu Val Tyr Glu
130 135 140

Gly Gln Val Ser Gly Met Pro Asn Asp Pro Ser Pro Leu Gln Val Ala
145 150 155 160

Val Lys Thr Leu Pro Glu Val Cys Ser Glu Gln Asp Glu Leu Asp Phe
165 170 175

P042876Seq. TXT

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

P042876Seq. TXT

Ser Pro Ala Ala Pro Pro Pro Leu Pro Thr Thr Ser Ser Gly Lys Ala
450 455 460

Ala Lys Lys Pro Thr Ala Ala Glu Val Ser Val Arg Val Pro Arg Gly
465 470 475 480

Pro Ala Val Glu Gly Gly His Val Asn Met Ala Phe Ser Glu Ser Asn
485 490 495

Pro Pro Ser Glu Leu His Arg Val His Gly Ser Arg Asn Lys Pro Thr
500 505 510

Ser Leu Trp Asn Pro Thr Tyr Gly Ser Trp Phe Thr Glu Lys Pro Thr
515 520 525

Lys Lys Asn Asn Pro Ile Ala Lys Lys Glu Pro His Glu Arg Gly Asn
530 535 540

Leu Gly Leu Glu Gly Ser Cys Thr Val Pro Pro Asn Val Ala Thr Gly
545 550 555 560

Arg Leu Pro Gly Ala Ser Leu Leu Leu Glu Pro Ser Ser Leu Thr Ala
565 570 575

Asn Met Lys Glu Val Pro Leu Phe Arg Leu Arg His Phe Pro Cys Gly
580 585 590

Asn Val Asn Tyr Gly Tyr Glu Glu Glu Gly Leu Pro Leu Glu Ala Ala
595 600 605

Thr Ala Pro Gly Ala Gly His Tyr Glu Asp Thr Ile Leu Lys Ser Lys
610 615 620

Asn Ser Met Asn Glu Pro Gly Pro
625 630

<210> 17

<211> 2131

<212> DNA

<213> Homo sapiens

<400> 17

act ct gt cgg t ccgct gaat gaagt gcccg cccct ct aag cccggagccc ggcgct t t cc	60
ccgcaagat g gacggt t t cg ccggcagt ct cgat gat agt at t t ct gct g caagt act t c	120
t gat gt t caa gat cgcct gt cagct ct t ga gt cagcagt t cagcaacaag aagat gaaat	180
cact gt gct a aaggcggct t t ggct gat gt t t t gaggcgt ct t gcaat ct ct gaagat ca	240
t gt ggcct ca gt gaaaaaat cagt ct caag t aaagg t ca gagct caggg gaggat at gg	300
agat ccaggg aggct t cct g t aggaagt gg cct gt gt agt gct t caaggg ccaggct gcc	360

P042876Seq. TXT

```

aggccat gt t gcagct gacc acccacct gc agt gt accgc cggaagcacc aggagct gca 420
agccat gcag at ggagct gc agagccct ga gt acaagct g agcaagct cc gcacct cgac 480
cat cat gacc gact acaacc ccaact act g ct t t gct ggc aagacct cct ccat cagt ga 540
cct gaaggag gt gccgcgga aaaacat cac cct cat t cgg ggt ct gggcc at ggagcct t 600
t ggggaggt g t at gaaggcc aggt gt ccgg aat gcccaac gaccaagcc ccct gcaagt 660
ggct gt gaag acgct gcct g aagt gt gct c t gaacaggac gaact ggat t t cct cat gga 720
agccct gat c at cagcaaat t caaccacca gaacat t gt t cgct gcat t g gggg gagcct 780
gcaat ccct g ccccggt t ca t cct gct gga gct cat ggcg gggggagacc t caagt cct t 840
cct ccgagag acccgccct c gcccgagcca gccct cct cc ct ggccat gc t ggacct t ct 900
gcacgt ggct cgggacat t g cct gt ggct g t cagt at t t g gagggaaaacc act t cat cca 960
ccgagacat t gct gccagaa act gcct ct t gacct gt cca ggccct ggaa gagt ggccaa 1020
gat t ggagac t t cgggat gg cccgagacat ct acaggggcg agct act at a gaaagggagg 1080
ct gt gccat g ct gccagt t a agt ggat gcc cccagaggcc t t cat ggaag gaat at t cac 1140
t t ct aaaaca gacacat ggt cct t t ggagt gct gct at gg gaaat ct t t t ct ct t ggat a 1200
t at gccat ac cccagcaaaa gcaaccagga agt t ct ggag t t t gt cacca gt ggaggccg 1260
gat ggacca cccaagaact gccct gggcc t gt at accgg at aat gact c agt gct ggca 1320
acat cagcct gaagacaggc ccaact t t gc cat cat t t t g gagaggat t g aat act gcac 1380
ccaggacccg gat gt aat ca acaccgct t t gccgat agaa t at ggt ccac t t gt ggaaga 1440
ggaagagaaa gt gcct gt ga ggcccaagga ccct gagggg gt t cct cct c t cct ggt ct c 1500
t caacaggca aaacgggagg aggagcgcag cccagct gcc ccaccacct c t gcct accac 1560
ct cct ct ggc aaggct gcaa agaaaccac agct gcagag gt ct ct gt t c gagt ccct ag 1620
agggccggcc gt ggaagggg gacacgt gaa t at ggcat t c t ct cagt cca accct cct t c 1680
ggagt t gcac aggggt ccacg gat ccagaaa caagcccacc agct t gt gga acccaacgt a 1740
cggct cct gg t t t acagaga aaccaccaa aaagaat aat cct at agcaa agaaggagcc 1800
acacgagagg ggt aacct gg ggct ggaggg aagct gt act gt cccacct a acgt t gcaac 1860
t gggagact t ccggggggcct cact gct cct agagccct ct t cgct gact g ccaat at gaa 1920
ggaggt acct ct gt t caggc t acgt cact t ccct t gt ggg aat gt caat t acggct acca 1980
gcaacagggc t t gccct t ag aagccgct ac t gccct gga gct ggt cat t acgaggat ac 2040
cat t ct gaaa agcaagaat a gcat gaacca gcct gggccc t gagct cggg cgcacact ca 2100
ct t ct ct t cc t t gggat ccc t aagaccgt g g 2131

```

<210> 18

<211> 671

<212> PRT

<213> Homo sapiens

<400> 18

Mbt Asp Gly Phe Ala Gly Ser Leu Asp Asp Ser Ile Ser Ala Ala Ser

1		5		10		15
Thr	Ser	Asp	Val 20	Gln	Asp	Arg
						Leu
						Ser
						Ala
						Leu
						Glu
						Ser
						Arg
						Val
						Gln
Gln	Gln	Glu	Asp	Glu	Ile	Thr
		35				Val
						40
						Leu
						Lys
						Ala
						Ala
						Leu
						Ala
						Asp
						Val
Leu	Arg	Arg	Leu	Ala	Ile	Ser
	50					55
						Glu
						Asp
						His
						Val
						Ala
						Ser
						Val
						Lys
						Lys
Ser	Val	Ser	Ser	Lys	Gly	Ser
65					70	
						Glu
						Leu
						Arg
						Gly
						75
						Gly
						Tyr
						Gly
						Asp
						Pro
						80
Gly	Arg	Leu	Pro	Val	Gly	Ser
						Gly
						Leu
						Cys
						90
						Ser
						Ala
						Ser
						Arg
						Ala
						Arg
						95
Leu	Pro	Gly	His	Val	Ala	Ala
						Asp
						His
						105
						Pro
						Pro
						Ala
						Val
						Tyr
						110
						Arg
						Arg
Lys	His	Gln	Glu	Leu	Gln	Ala
						Met
						120
						Gln
						Met
						Glu
						Leu
						Gln
						Ser
						Pro
						Glu
						125
Tyr	Lys	Leu	Ser	Lys	Leu	Arg
						135
						Thr
						Ser
						Thr
						Ile
						Met
						140
						Thr
						Asp
						Tyr
						Asn
						130
Pro	Asn	Tyr	Cys	Phe	Ala	Gly
145						150
						Lys
						Thr
						Ser
						Ser
						155
						Ile
						Ser
						Asp
						Leu
						Lys
						160
Glu	Val	Pro	Arg	Lys	Asn	Ile
						Thr
						Leu
						Ile
						Arg
						Gly
						Leu
						Gly
						His
						Gly
						175
Ala	Phe	Gly	Glu	Val	Tyr	Glu
						Gly
						Gln
						Val
						Ser
						Gly
						Met
						Pro
						Asn
						Asp
						180
Pro	Ser	Pro	Leu	Gln	Val	Ala
						Val
						200
						Lys
						Thr
						Leu
						Pro
						Glu
						Val
						Cys
						Ser
						195
Glu	Gln	Asp	Glu	Leu	Asp	Phe
						215
						Leu
						Met
						Glu
						Ala
						Leu
						Ile
						Ile
						Ser
						Lys
						220
Phe	Asn	His	Gln	Asn	Ile	Val
225						230
						Arg
						Cys
						Ile
						Gly
						235
						Val
						Ser
						Leu
						Gln
						Ser
						240
Leu	Pro	Arg	Phe	Ile	Leu	Leu
						Glu
						Leu
						Met
						250
						Ala
						Gly
						Gly
						Asp
						Leu
						Lys
						255
Ser	Phe	Leu	Arg	Glu	Thr	Arg
						Pro
						Arg
						Pro
						Ser
						Gln
						Pro
						Ser
						270
						Ser
						Leu
						260
Ala	Met	Leu	Asp	Leu	Leu	His
						Val
						Ala
						Arg
						Asp
						Ile
						Ala
						Cys
						Gly
						Cys

275

280

285

G n Tyr Leu G u G u Asn H i s Phe I l e H i s Arg Asp I l e Al a Al a Arg
 290 295 300
 Asn Cys Leu Leu Thr Cys Pro G y Pro G y Arg Val Al a Lys I l e G y
 305 310 315 320
 Asp Phe G y M e t Al a Arg Asp I l e Tyr Arg Al a Ser Tyr Tyr Arg Lys
 325 330 335
 G y G y Cys Al a M e t Leu Pro Val Lys Trp M e t Pro Pro G u Al a Phe
 340 345 350
 M e t G u G y I l e Phe Thr Ser Lys Thr Asp Thr Trp Ser Phe G y Val
 355 360 365
 Leu Leu Trp G u I l e Phe Ser Leu G y Tyr M e t Pro Tyr Pro Ser Lys
 370 375 380
 Ser Asn G n G u Val Leu G u Phe Val Thr Ser G y G y Arg M e t Asp
 385 390 395 400
 Pro Pro Lys Asn Cys Pro G y Pro Val Tyr Arg I l e M e t Thr G n Cys
 405 410 415
 Trp G n H i s G n Pro G u Asp Arg Pro Asn Phe Al a I l e I l e Leu G u
 420 425 430
 Arg I l e G u Tyr Cys Thr G n Asp Pro Asp Val I l e Asn Thr Al a Leu
 435 440 445
 Pro I l e G u Tyr G y Pro Leu Val G u G u G u G u Lys Val Pro Val
 450 455 460
 Arg Pro Lys Asp Pro G u G y Val Pro Pro Leu Leu Val Ser G n G n
 465 470 475 480
 Al a Lys Arg G u G u G u Arg Ser Pro Al a Al a Pro Pro Pro Leu Pro
 485 490 495
 Thr Thr Ser Ser G y Lys Al a Al a Lys Lys Pro Thr Al a Al a G u Val
 500 505 510
 Ser Val Arg Val Pro Arg G y Pro Al a Val G u G y G y H i s Val Asn
 515 520 525
 M e t Al a Phe Ser G n Ser Asn Pro Pro Ser G u Leu H i s Arg Val H i s
 530 535 540
 G y Ser Arg Asn Lys Pro Thr Ser Leu Trp Asn Pro Thr Tyr G y Ser

545 550 555 560

Trp Phe Thr Glu Lys₅₆₅ Pro Thr Lys Lys Asn₅₇₀ Asn Pro Ile Ala Lys₅₇₅ Lys

Glu Pro His Glu₅₈₀ Arg Gly Asn Leu Gly₅₈₅ Leu Glu Gly Ser Cys₅₉₀ Thr Val

Pro Pro Asn₅₉₅ Val Ala Thr Gly Arg₆₀₀ Leu Pro Gly Ala Ser₆₀₅ Leu Leu Leu

Glu Pro Ser Ser Leu Thr Ala₆₁₅ Asn Met Lys Glu Val₆₂₀ Pro Leu Phe Arg

Leu Arg His Phe Pro Cys₆₃₀ Gly Asn Val Asn Tyr₆₃₅ Gly Tyr Glu Glu Glu₆₄₀

Gly Leu Pro Leu Glu₆₄₅ Ala Ala Thr Ala Pro₆₅₀ Gly Ala Gly His Tyr₆₅₅ Glu

Asp Thr Ile Leu Lys Ser Lys Asn Ser₆₆₅ Met Asn Glu Pro Gly₆₇₀ Pro

<210> 19
 <211> 3365
 <212> DNA
 <213> Homo sapiens

<400> 19

t act ct gt cg gt ccgct gaa t gaagt gcc	gccct ct aa gcccgagcc cggcgct t t c	60
cccgaagat ggacggt t t c gccggcagt c t cgat gat ag t at t t ct gct gcaagt act t		120
ct gat gt t ca agat cgcct g t cagct ct t g agt cacgagt t cagcaacaa gaagat gaaa		180
t cact gt gct aaaggcggct t t ggct gat g t t t t gaggcg t ct t gcaat c t ct gaagat c		240
at gt ggcct c agt gaaaaa t cagt ct caa gt aaaggcca accaagccct cgagcagt t a		300
t t cccat gt c ct gt at aacc aat ggaagt g gt gcaaacag aaaaccaagt cat accagt g		360
ct gt ct caat t gcaggaaaa gaaact ct t t cat ct gct gc t aaaagt ggt acagaaaaaa		420
agaaagaaaa accacaagga cagagagaaa aaaaagagga at ct cat t ct aat gat caaa		480
gt ccacaaat t cgagcat ca cct t ct cccc agccct ct t c acaacct ct c caaat acaca		540
gacaaact cc agaaagcaag aat gct act c ccacaaaag cat aaaacga ccat caccag		600
ct gaaaagt c acat aat t ct t gggaaaat t cagat gat ag ccgt aat aaa t t gt cgaaaa		660
t acct t caac acccaaat t a at accaaaag t t accaaaac t gcagacaag cat aaagat g		720
t cat cat caa ccaagaagga gaat at at t a aaat gt t t at gcgcggg cgg ccaat t acca		780
t gt t cat t cc t t ccgat gt t gacaact at g at gacat cag aacggaact g cct cct gaga		840
agct caaact ggagt gggca t at ggt t at c gaggaaagga ct gt agagct aat gt t t acc		900
t t ct t ccgac cgggggaaat a gt t t at t t ca t t gcat cagt agt agt act a t t t aat t at g		960

P042876Seq. TXT

aggagagaac t cagcgacac t acct gggcc at acagact g t gt gaaat gc ct t gct at ac	1020
at cct gacaa aat t aggat t gcaact ggac agat agct gg cgt ggat aaa gat ggaaggc	1080
ct ct acaacc ccacgt caga gt gt gggat t ct gt t act ct at ccacact g cagat t at t g	1140
gact t ggcac t t t t gagcgt ggagt aggat gcct ggat t t t t caaaagca gat t caggt g	1200
t t cat t t at g t gt t at t gat gact ccaat g agcat at gct t act gt at gg gact ggcaga	1260
ggaaagcaaa aggagcagaa at aaagacaa caaat gaagt t gt t t t ggct gt ggagt t t c	1320
acccaacaga t gcaaat acc at aat t acat gcggt aaat c t cat at t t t c t t ct ggacct	1380
ggagcggcaa t t cact aaca agaaaacagg gaat t t t t gg gaaat at gaa aagccaaaat	1440
t t gt gcagt g t t t agcat t c t t ggggaat g gagat gt t ct t act ggagac t caggt ggag	1500
t cat gct t at at ggagcaaa act act gt ag agcccacacc t gggaaagga cct aaaggaa	1560
gt ggcct gt g t agt gct t ca agggccaggc t gccaggcca t gt t gcagct gaccacccac	1620
ct gcagt gt a ccgccggaag caccaggagc t gcaagccat gcagat ggag ct gcagagcc	1680
ct gagt acaa gct gagcaag ct ccgcacct cgacct cat gaccgact ac aaccccaact	1740
act gct t t gc t ggcaagacc t cct ccat ca gt gacct gaa ggaggt gccg cggaaaaaca	1800
t caccct cat t cgggggt ct g ggccat ggag cct t t gggga ggt gt at gaa ggccaggt gt	1860
ccggaat gcc caacgacca agccccct gc aagt ggct gt gaagacgct g cct gaagt gt	1920
gct ct gaaca ggacgaact g gat t t cct ca t ggaagccct gat cat cagc aaat t caacc	1980
accagaacat t gt t cgct gc at t ggggt ga gcct gcaat c cct gccccgg t t cat cct gc	2040
t ggagct cat ggcgggggga gacct caagt cct t cct ccg agagaccgc cct cgccga	2100
gccagccct c ct ccct ggcc at gct ggacc t t ct gcagct ggct cgggac at t gcct gt g	2160
gct gt cagt a t t t ggaggaa aaccact t ca t ccaccgaga cat t gct gcc agaaact gcc	2220
t ct t gacct g t ccaggccct ggaagagt gg ccaagat t gg agact t cggg at ggcccagag	2280
acat ct acag ggcgagct ac t at agaaagg gaggt gt gc cat gct gcca gt t aagt gga	2340
t gccccaga ggcct t cat g gaaggaat at t cact t ct aa aacagacaca t ggt cct t t g	2400
gagt gct gct at gggaaat c t t t t ct ct t g gat at at gcc at accccagc aaaagcaacc	2460
aggaagt t ct ggagt t t gt c accagt ggag gccggat gga cccaccaag aact gccct g	2520
ggcct gt at a ccggat aat g act cagt gct ggcaacat ca gcct gaagac aggcccaact	2580
t t gccat cat t t t ggagagg at t gaat act gcaccagga cccggat gt a at caacaccg	2640
ct t t gccgat agaat at ggt ccact t gt gg aagaggaaga gaaagt gcct gt gaggccca	2700
aggacct ga gggggt t cct cct ct cct gg t ct ct caaca ggcaaaacgg gaggaggagc	2760
gcagcccagc t gccccacca cct ct gcct a ccacct cct c t ggcaaggct gcaaagaaac	2820
ccacagct gc agaggt ct ct gt t cgagt cc ct agagggcc ggccgt ggaa gggggacacg	2880
t gaat at ggc at t ct ct cag t ccaaccct c ct t cggagt t gcacagggt c cacggat cca	2940
gaaacaagcc caccagct t g t ggaaccaa cgt acggct c ct ggt t t aca gagaaaccca	3000

P042876Seq. TXT

ccaaaaagaa t aat cct at a gcaaagaagg agccacacga gaggggt aac ct ggggct gg 3060
 agggaagct g t act gt ccca cct aacgt t g caact gggag act t ccgggg gcct cact gc 3120
 t cct agagcc ct ct t cgct g act gccaat a t gaaggaggt acct ct gt t c aggct acgt c 3180
 act t ccct t g t gggaat gt c aat t acggct accagcaaca gggct t gccc t t agaagccg 3240
 ct act gcccc t ggagct ggt cat t acgagg at accat t ct gaaaagcaag aat agcat ga 3300
 accagcct gg gccct gagct cggcgcaca ct cact t ct c t t cct t ggga t ccct aagac 3360
 cgt gg 3365

<210> 20

<211> 1082

<212> PRT

<213> Homo sapiens

<400> 20

Met Asp Gly Phe Ala Gly Ser Leu Asp Asp Ser Ile Ser Ala Ala Ser
 1 5 10 15

Thr Ser Asp Val Gln Asp Arg Leu Ser Ala Leu Glu Ser Arg Val Gln
 20 25 30

Gln Gln Glu Asp Glu Ile Thr Val Leu Lys Ala Ala Leu Ala Asp Val
 35 40 45

Leu Arg Arg Leu Ala Ile Ser Glu Asp His Val Ala Ser Val Lys Lys
 50 55 60

Ser Val Ser Ser Lys Gly Gln Pro Ser Pro Arg Ala Val Ile Pro Met
 65 70 75 80

Ser Cys Ile Thr Asn Gly Ser Gly Ala Asn Arg Lys Pro Ser His Thr
 85 90 95

Ser Ala Val Ser Ile Ala Gly Lys Glu Thr Leu Ser Ser Ala Ala Lys
 100 105 110

Ser Gly Thr Glu Lys Lys Lys Glu Lys Pro Gln Gly Gln Arg Glu Lys
 115 120 125

Lys Glu Glu Ser His Ser Asn Asp Gln Ser Pro Gln Ile Arg Ala Ser
 130 135 140

Pro Ser Pro Gln Pro Ser Ser Gln Pro Leu Gln Ile His Arg Gln Thr
 145 150 155 160

Pro Glu Ser Lys Asn Ala Thr Pro Thr Lys Ser Ile Lys Arg Pro Ser
 165 170 175

Pro Ala Glu Lys Ser His Asn Ser Trp Glu Asn Ser Asp Asp Ser Arg
 180 185 190

P042876Seq. TXT

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

P042876Seq. TXT

Leu Gly Asn Gly Asp Val Leu Thr Gly Asp Ser Gly Gly Val Met Leu
 465 470 475 480
 Ile Trp Ser Lys Thr Thr Val Glu Pro Thr Pro Gly Lys Gly Pro Lys
 485 490 495
 Gly Ser Gly Leu Cys Ser Ala Ser Arg Ala Arg Leu Pro Gly His Val
 500 505 510
 Ala Ala Asp His Pro Pro Ala Val Tyr Arg Arg Lys His Gln Glu Leu
 515 520 525
 Gln Ala Met Gln Met Glu Leu Gln Ser Pro Glu Tyr Lys Leu Ser Lys
 530 535 540
 Leu Arg Thr Ser Thr Ile Met Thr Asp Tyr Asn Pro Asn Tyr Cys Phe
 545 550 555 560
 Ala Gly Lys Thr Ser Ser Ile Ser Asp Leu Lys Glu Val Pro Arg Lys
 565 570 575
 Asn Ile Thr Leu Ile Arg Gly Leu Gly His Gly Ala Phe Gly Glu Val
 580 585 590
 Tyr Glu Gly Gln Val Ser Gly Met Pro Asn Asp Pro Ser Pro Leu Gln
 595 600 605
 Val Ala Val Lys Thr Leu Pro Glu Val Cys Ser Glu Gln Asp Glu Leu
 610 615 620
 Asp Phe Leu Met Glu Ala Leu Ile Ile Ser Lys Phe Asn His Gln Asn
 625 630 635 640
 Ile Val Arg Cys Ile Gly Val Ser Leu Gln Ser Leu Pro Arg Phe Ile
 645 650 655
 Leu Leu Glu Leu Met Ala Gly Gly Asp Leu Lys Ser Phe Leu Arg Glu
 660 665 670
 Thr Arg Pro Arg Pro Ser Gln Pro Ser Ser Leu Ala Met Leu Asp Leu
 675 680 685
 Leu His Val Ala Arg Asp Ile Ala Cys Gly Cys Gln Tyr Leu Glu Glu
 690 695 700
 Asn His Phe Ile His Arg Asp Ile Ala Ala Arg Asn Cys Leu Leu Thr
 705 710 715 720
 Cys Pro Gly Pro Gly Arg Val Ala Lys Ile Gly Asp Phe Gly Met Ala
 725 730 735

P042876Seq. TXT

Arg Asp Ile Tyr Arg Ala Ser Tyr Tyr Arg Lys Gly Gly Cys Ala Met
 740 745 750
 Leu Pro Val Lys Trp Met Pro Pro Gu Ala Phe Met Gu Gly Ile Phe
 755 760 765
 Thr Ser Lys Thr Asp Thr Trp Ser Phe Gly Val Leu Leu Trp Gu Ile
 770 775 780
 Phe Ser Leu Gly Tyr Met Pro Tyr Pro Ser Lys Ser Asn Gn Gu Val
 785 790 800
 Leu Gu Phe Val Thr Ser Gly Gly Arg Met Asp Pro Pro Lys Asn Cys
 805 810 815
 Pro Gly Pro Val Tyr Arg Ile Met Thr Gn Cys Trp Gn His Gn Pro
 820 825 830
 Gu Asp Arg Pro Asn Phe Ala Ile Ile Leu Gu Arg Ile Gu Tyr Cys
 835 840 845
 Thr Gn Asp Pro Asp Val Ile Asn Thr Ala Leu Pro Ile Gu Tyr Gly
 850 855 860
 Pro Leu Val Gu Gu Gu Gu Lys Val Pro Val Arg Pro Lys Asp Pro
 865 870 875 880
 Gu Gly Val Pro Pro Leu Leu Val Ser Gn Gn Ala Lys Arg Gu Gu
 885 890 895
 Gu Arg Ser Pro Ala Ala Pro Pro Pro Leu Pro Thr Thr Ser Ser Gly
 900 905 910
 Lys Ala Ala Lys Lys Pro Thr Ala Ala Gu Val Ser Val Arg Val Pro
 915 920 925
 Arg Gly Pro Ala Val Gu Gly Gly His Val Asn Met Ala Phe Ser Gn
 930 935 940
 Ser Asn Pro Pro Ser Gu Leu His Arg Val His Gly Ser Arg Asn Lys
 945 950 955 960
 Pro Thr Ser Leu Trp Asn Pro Thr Tyr Gly Ser Trp Phe Thr Gu Lys
 965 970 975
 Pro Thr Lys Lys Asn Asn Pro Ile Ala Lys Lys Gu Pro His Gu Arg
 980 985 990
 Gly Asn Leu Gly Leu Gu Gly Ser Cys Thr Val Pro Pro Asn Val Ala
 995 1000 1005

P042876Seq. TXT

Thr Gly Arg Leu Pro Gly Ala Ser Leu Leu Leu Glu Pro Ser Ser
 1010 1015 1020

Leu Thr Ala Asn Met Lys Glu Val Pro Leu Phe Arg Leu Arg His
 1025 1030 1035

Phe Pro Cys Gly Asn Val Asn Tyr Gly Tyr Glu Glu Glu Gly Leu
 1040 1045 1050

Pro Leu Glu Ala Ala Thr Ala Pro Gly Ala Gly His Tyr Glu Asp
 1055 1060 1065

Thr Ile Leu Lys Ser Lys Asn Ser Met Asn Glu Pro Gly Pro
 1070 1075 1080

<210> 21
 <211> 3435
 <212> DNA
 <213> Homo sapiens

<400> 21
 t act ct gt cg gt ccgct gaa t gaagt gccc gccct ct aa gcccgagcc cggcgct t t c 60
 cccgcaagat ggacggt t t c gccggcagt c t cgat gat ag t at t t ct gct gcaagt act t 120
 ct gat gt t ca agat cgct g t cagct ct t g agt cagagt t cagcaaca gaagat gaaa 180
 t cact gt gct aaaggcggct t t ggct gat g t t t t gaggcg t ct t gcaat c t ct gaagat c 240
 at gt ggcct c agt gaaaaa t cagt ct caa gt aaaggcca accaagccct cgagcagt t a 300
 t t cccat gt c ct gt at aacc aat ggaagt g gt gcaaacag aaaaccaagt cat accagt g 360
 ct gt ct caat t gcaggaaaa gaaact ct t t cat ct gct gc t aaaagt ggt acagaaaaaa 420
 agaaagaaaa accacaagga cagagagaaa aaaaagagga at ct cat t ct aat gat caaa 480
 gt ccacaaat t cgagcat ca cct t ct cccc agccct ct t c acaacct ct c caaat acaca 540
 gacaaact cc agaaagcaag aat gct act c ccacaaaag cat aaaacga ccat caccag 600
 ct gaaaagt c acat aat t ct t gggaaaat t cagat gat ag ccgt aat aaa t t gt cgaaaa 660
 t acct t caac acccaaat t a at accaaaag t t accaaaac t gcagacaag cat aaagat g 720
 t cat cat caa ccaagaagga gaat at at t a aaat gt t t at gcgcggg cgg ccaat t acca 780
 t gt t cat t cc t t ccgat gt t gacaact at g at gacat cag aacggaact g cct cct gaga 840
 agct caaact ggagt gggca t at ggt t at c gaggaagga ct gt agagct aat gt t t acc 900
 t t ct t ccgac cggggaaat a gt t t at t t ca t t gcat cagt agt agt act a t t t aat t at g 960
 aggagagaac t cagcgacac t acct gggcc at acagact g t gt gaaat gc ct t gct at ac 1020
 at cct gacaa aat t aggat t gcaact ggac agat agct gg cgt ggat aaa gat ggaaggc 1080
 ct ct acaacc ccacgt caga gt gt gggat t ct gt t act ct at ccacact g cagat t at t g 1140
 gact t ggcac t t t t gagcgt ggagt aggat gcct ggat t t t t caaaagca gat t caggt g 1200

t t cat t t at g	t gt t at t gat	gact ccaat g	agcat at gct	t act gt at gg	gact ggcaga	1260
ggaaagcaaa	aggagcagaa	at aaagacaa	caaat gaagt	t gt t t t ggct	gt ggagt t t c	1320
acccaacaga	t gcaaat acc	at aat t acat	gcggt aaat c	t cat at t t t c	t t ct ggacct	1380
ggagcggcaa	t t cact aaca	agaaaacagg	gaat t t t t gg	gaaat at gaa	aagccaaaat	1440
t t gt gcagt g	t t t agcat t c	t t ggggaat g	gagat gt t ct	t act ggagac	t caggt ggag	1500
t cat gct t at	at ggagcaaa	act act gt ag	agcccacacc	t gggaaagga	cct aaaggt g	1560
t at at caaat	cagcaaacaa	at caaagct c	at gat ggcag	t gt gt t caca	ct t t gt caga	1620
t gagaaat gg	gat gt t at t a	act ggaggag	ggaaagacag	aaaaat aat t	ct gt gggat c	1680
at gat ct gaa	t cct gaaaga	gaaat agagc	accaggagct	gcaagccat g	cagat ggagc	1740
t gcagagccc	t gagt acaag	ct gagcaagc	t ccgcacct c	gacct cat g	accgact aca	1800
accccaact a	ct gct t t gct	ggcaagacct	cct ccat cag	t gacct gaag	gaggt gccgc	1860
ggaaaaacat	caccct cat t	cgggggt ct gg	gccat ggagc	ct t t ggggag	gt gt at gaag	1920
gccaggt gt c	cggaat gccc	aacgacccaa	gccccct gca	agt ggct gt g	aagacgct gc	1980
ct gaagt gt g	ct ct gaacag	gacgaact gg	at t t cct cat	ggaagccct g	at cat cagca	2040
aat t caacca	ccagaacat t	gt t cgct gca	t t ggggt gag	cct gcaat cc	ct gccccggt	2100
t cat cct gct	ggagct cat g	gcgggggggag	acct caagt c	ct t cct ccga	gagacccgcc	2160
ct cgccccgag	ccagccct cc	t cctt ggcca	t gct ggacct	t ct gcacgt g	gct cgggaca	2220
t t gcct gt gg	ct gt cagt at	t t ggaggaaa	accact t cat	ccaccgagac	at t gct gcca	2280
gaaact gcct	ct t gacct gt	ccaggccct g	gaagagt ggc	caagat t gga	gact t cggga	2340
t ggcccgaga	cat ct acagg	gcgagct act	at agaaaggg	aggct gt gcc	at gct gccag	2400
t t aagt ggat	gccccagag	gcct t cat gg	aaggaat at t	cact t ct aaa	acagacacat	2460
ggg cct t t gg	agt gct gct a	t gggaaat ct	t t t ct ct t gg	at at at gcca	t accccagca	2520
aaagcaacca	ggaagt t ct g	gagt t t gt ca	ccagt ggagg	ccggat ggac	ccaccaaga	2580
act gccct gg	gcct gt at ac	cggat aat ga	ct cagt gct g	gcaacat cag	cct gaagaca	2640
ggcccaact t	t gccat cat t	t t ggagagga	t t gaat act g	cacccaggac	ccggat gt aa	2700
t caacaccgc	t t t gccgat a	gaat at ggt c	cact t gt gga	agaggaagag	aaagt gcct g	2760
t gaggccccaa	ggaccct gag	gggggt t cct c	ct ct cct ggt	ct ct caacag	gcaaaacggg	2820
aggaggagcg	cagcccagct	gccccaccac	ct ct gcct ac	cacct cct ct	ggcaaggct g	2880
caaagaaacc	cacagct gca	gaggt ct ct g	t t cgagt ccc	t agagggccg	gccgt ggaag	2940
ggggacacgt	gaat at ggca	t t ct ct cagt	ccaaccct cc	t t cggagt t g	cacaaggt cc	3000
acggat ccag	aaacaagccc	accagct t gt	ggaacccaac	gt acggct cc	t ggt t t acag	3060
agaaaccac	caaaaagaat	aat cct at ag	caaagaagga	gccacacgac	aggggt aacc	3120
t ggggct gga	gggaagct gt	act gt cccac	ct aacgt t gc	aact gggaga	ct t ccggggg	3180
cct cact gct	cct agagccc	t ct t cgct ga	ct gccaat at	gaaggaggt a	cct ct gt t ca	3240

P042876Seq. TXT

ggct acgt ca ct t ccct t gt gggaat gt ca at t acggct a ccagcaacag ggct t gccct 3300
t agaagccgc t act gccct ggagct ggt c at t acgagga t accat t ct g aaaagcaaga 3360
at agcat gaa ccagcct ggg ccct gagct c ggt cgcacac t cact t ct ct t cct t gggat 3420
ccct aagacc gt gga 3435

<210> 22

<211> 1105

<212> PRT

<213> Homo sapiens

<400> 22

Met Asp Gly Phe Ala Gly Ser Leu Asp Asp Ser Ile Ser Ala Ala Ser
1 5 10 15

Thr Ser Asp Val Gln Asp Arg Leu Ser Ala Leu Glu Ser Arg Val Gln
20 25 30

Gln Gln Glu Asp Glu Ile Thr Val Leu Lys Ala Ala Leu Ala Asp Val
35 40 45

Leu Arg Arg Leu Ala Ile Ser Glu Asp His Val Ala Ser Val Lys Lys
50 55 60

Ser Val Ser Ser Lys Gly Gln Pro Ser Pro Arg Ala Val Ile Pro Met
65 70 75 80

Ser Cys Ile Thr Asn Gly Ser Gly Ala Asn Arg Lys Pro Ser His Thr
85 90 95

Ser Ala Val Ser Ile Ala Gly Lys Glu Thr Leu Ser Ser Ala Ala Lys
100 105 110

Ser Gly Thr Glu Lys Lys Lys Glu Lys Pro Gln Gly Gln Arg Glu Lys
115 120 125

Lys Glu Glu Ser His Ser Asn Asp Gln Ser Pro Gln Ile Arg Ala Ser
130 135 140

Pro Ser Pro Gln Pro Ser Ser Gln Pro Leu Gln Ile His Arg Gln Thr
145 150 155 160

Pro Glu Ser Lys Asn Ala Thr Pro Thr Lys Ser Ile Lys Arg Pro Ser
165 170 175

Pro Ala Glu Lys Ser His Asn Ser Trp Glu Asn Ser Asp Asp Ser Arg
180 185 190

Asn Lys Leu Ser Lys Ile Pro Ser Thr Pro Lys Leu Ile Pro Lys Val
195 200 205

Thr Lys Thr Ala Asp Lys His Lys Asp Val Ile Ile Asn Gln Glu Gly

210

215

Glu Tyr Ile Lys Met Phe Met Arg Gly Arg Pro Ile Thr Met Phe Ile
225 230 235 240

Pro Ser Asp Val Asp Asn Tyr Asp Asp Ile Arg Thr Glu Leu Pro Pro
245 250 255

Glu Lys Leu Lys Leu Glu Trp Ala Tyr Gly Tyr Arg Gly Lys Asp Cys
260 265 270

Arg Ala Asn Val Tyr Leu Leu Pro Thr Gly Glu Ile Val Tyr Phe Ile
275 280 285

Ala Ser Val Val Val Leu Phe Asn Tyr Glu Glu Arg Thr Gln Arg His
290 295 300

Tyr Leu Gly His Thr Asp Cys Val Lys Cys Leu Ala Ile His Pro Asp
305 310 315 320

Lys Ile Arg Ile Ala Thr Gly Gln Ile Ala Gly Val Asp Lys Asp Gly
325 330 335

Arg Pro Leu Gln Pro His Val Arg Val Trp Asp Ser Val Thr Leu Ser
340 345 350

Thr Leu Gln Ile Ile Gly Leu Gly Thr Phe Glu Arg Gly Val Gly Cys
355 360 365

Leu Asp Phe Ser Lys Ala Asp Ser Gly Val His Leu Cys Val Ile Asp
370 375 380

Asp Ser Asn Glu His Met Leu Thr Val Trp Asp Trp Gln Arg Lys Ala
385 390 395 400

Lys Gly Ala Glu Ile Lys Thr Thr Asn Glu Val Val Leu Ala Val Glu
405 410 415

Phe His Pro Thr Asp Ala Asn Thr Ile Ile Thr Cys Gly Lys Ser His
420 425 430

Ile Phe Phe Trp Thr Trp Ser Gly Asn Ser Leu Thr Arg Lys Gln Gly
435 440 445

Ile Phe Gly Lys Tyr Glu Lys Pro Lys Phe Val Gln Cys Leu Ala Phe
450 455 460

Leu Gly Asn Gly Asp Val Leu Thr Gly Asp Ser Gly Gly Val Met Leu
465 470 475 480

Ile Trp Ser Lys Thr Thr Val Glu Pro Thr Pro Gly Lys Gly Pro Lys

485

490

495

G y Val Tyr G n Ile Ser Lys G n Ile Lys Ala Hi s Asp G y Ser Val
 500 505 510
 Phe Thr Leu Cys G n Met Arg Asn G y Met Leu Leu Thr G y G y G y
 515 520 525
 Lys Asp Arg Lys Ile Ile Leu Trp Asp Hi s Asp Leu Asn Pro Gl u Arg
 530 535 540
 Gl u Ile Gl u Hi s G n Gl u Leu G n Ala Met G n Met Gl u Leu G n Ser
 545 550 555 560
 Pro Gl u Tyr Lys Leu Ser Lys Leu Arg Thr Ser Thr Ile Met Thr Asp
 565 570 575
 Tyr Asn Pro Asn Tyr Cys Phe Ala G y Lys Thr Ser Ser Ile Ser Asp
 580 585 590
 Leu Lys Gl u Val Pro Arg Lys Asn Ile Thr Leu Ile Arg G y Leu G y
 595 600 605
 Hi s G y Ala Phe G y Gl u Val Tyr Gl u G y G n Val Ser G y Met Pro
 610 615 620
 Asn Asp Pro Ser Pro Leu G n Val Ala Val Lys Thr Leu Pro Gl u Val
 625 630 635 640
 Cys Ser Gl u G n Asp Gl u Leu Asp Phe Leu Met Gl u Ala Leu Ile Ile
 645 650 655
 Ser Lys Phe Asn Hi s G n Asn Ile Val Arg Cys Ile G y Val Ser Leu
 660 665 670
 G n Ser Leu Pro Arg Phe Ile Leu Leu Gl u Leu Met Ala G y G y Asp
 675 680 685
 Leu Lys Ser Phe Leu Arg Gl u Thr Arg Pro Arg Pro Ser G n Pro Ser
 690 695 700
 Ser Leu Ala Met Leu Asp Leu Leu Hi s Val Ala Arg Asp Ile Ala Cys
 705 710 715 720
 G y Cys G n Tyr Leu Gl u Gl u Asn Hi s Phe Ile Hi s Arg Asp Ile Ala
 725 730 735
 Ala Arg Asn Cys Leu Leu Thr Cys Pro G y Pro G y Arg Val Ala Lys
 740 745 750
 Ile G y Asp Phe G y Met Ala Arg Asp Ile Tyr Arg Ala Ser Tyr Tyr

755

760

765

Arg Lys Gly Gly Cys Ala Met Leu Pro Val Lys Trp Met Pro Pro Glu
 770 775 780
 Ala Phe Met Glu Gly Ile Phe Thr Ser Lys Thr Asp Thr Trp Ser Phe
 785 790 795 800
 Gly Val Leu Leu Trp Glu Ile Phe Ser Leu Gly Tyr Met Pro Tyr Pro
 805 810 815
 Ser Lys Ser Asn Gln Glu Val Leu Glu Phe Val Thr Ser Gly Gly Arg
 820 825 830
 Met Asp Pro Pro Lys Asn Cys Pro Gly Pro Val Tyr Arg Ile Met Thr
 835 840 845
 Gln Cys Trp Gln His Gln Pro Glu Asp Arg Pro Asn Phe Ala Ile Ile
 850 855 860
 Leu Glu Arg Ile Glu Tyr Cys Thr Gln Asp Pro Asp Val Ile Asn Thr
 865 870 875 880
 Ala Leu Pro Ile Glu Tyr Gly Pro Leu Val Glu Glu Glu Glu Lys Val
 885 890 895
 Pro Val Arg Pro Lys Asp Pro Glu Gly Val Pro Pro Leu Leu Val Ser
 900 905 910
 Gln Gln Ala Lys Arg Glu Glu Glu Arg Ser Pro Ala Ala Pro Pro Pro
 915 920 925
 Leu Pro Thr Thr Ser Ser Gly Lys Ala Ala Lys Lys Pro Thr Ala Ala
 930 935 940
 Glu Val Ser Val Arg Val Pro Arg Gly Pro Ala Val Glu Gly Gly His
 945 950 955 960
 Val Asn Met Ala Phe Ser Gln Ser Asn Pro Pro Ser Glu Leu His Lys
 965 970 975
 Val His Gly Ser Arg Asn Lys Pro Thr Ser Leu Trp Asn Pro Thr Tyr
 980 985 990
 Gly Ser Trp Phe Thr Glu Lys Pro Thr Lys Lys Asn Asn Pro Ile Ala
 995 1000 1005
 Lys Lys Glu Pro His Asp Arg Gly Asn Leu Gly Leu Glu Gly Ser
 1010 1015 1020
 Cys Thr Val Pro Pro Asn Val Ala Thr Gly Arg Leu Pro Gly Ala

P042876Seq. TXT

1025

1030

1035

Ser Leu Leu Leu Glu Pro Ser Ser Leu Thr Ala Asn Met Lys Glu
 1040 1045 1050
 Val Pro Leu Phe Arg Leu Arg His Phe Pro Cys Gly Asn Val Asn
 1055 1060 1065
 Tyr Gly Tyr Gln Gln Gln Gly Leu Pro Leu Glu Ala Ala Thr Ala
 1070 1075 1080
 Pro Gly Ala Gly His Tyr Gu Asp Thr Ile Leu Lys Ser Lys Asn
 1085 1090 1095
 Ser Met Asn Gln Pro Gly Pro
 1100 1105

<210> 23
 <211> 4479
 <212> DNA
 <213> Homo sapiens

<400> 23
 tgcgagaaaag atggcggacc tggccgagt g caacat caaa gt gat gt gt c gct t cagacc 60
 tct caacgag tct gaagt ga accgcggcga caagt acat c gccaaagt t t c agggagaaga 120
 cacggg cgt g at cgcgt cca agcct t at gc at t t gat cgg gt gt t ccagt caagcacat c 180
 t caagagcaa gt gt at aat g act gt gcaaa gaagat t gt t aaagat gt ac t t gaaggat a 240
 t aat ggaaca at at t t gcat at ggacaaac at cct ct ggg aagacacaca caat ggaggg 300
 t aaact t cat gat ccagaag gcat gggaat t at t ccaaga at agt gcaag at at t t t t aa 360
 t t at at t t ac t ccat ggat g aaaat t t gga at t t cat at t aaggt t t cat at t t t gaaat 420
 at at t t ggat aagat aaggg acct gt t aga t gt t t caaag accaacct t t cagt t cat ga 480
 agacaaaaac cgagt t ccct at gt aaaggg gt gcacagag cgt t t t gt at gt agt ccaga 540
 t gaagt t at g gat accat ag at gaaggaaa at ccaacaga cat gt agcag t t acaaat at 600
 gaat gaacat agct ct agga gt cacagt at at t t ct t at t aat gt caaac aagagaacac 660
 acaaacggaa caaaagct ga gt ggaaaact t t at ct ggt t gat t t agct g gt agt gaaaa 720
 ggt t agt aaa act ggagct g aaggt gct gt gct ggat gaa gct aaaaaca t caacaagt c 780
 act t t ct gct ct t ggaaat g t t at t t ct gc t t t ggct gag ggt agt acat at gt t ccat a 840
 t cgagat agt aaaat gacaa gaat cct t ca agat t cat t a ggt ggcaact gt agaaccac 900
 t at t gt aat t t gct gct ct c cat cat cat a caat gagt ct gaaacaaaat ct aact ct t 960
 at t t ggccaa agggccaaaa caat t aagaa cacagt t t gt gt caat gt gg agt t aact gc 1020
 agaacagt gg aaaaagaagt at gaaaaaga aaaagaaaaa aat aagat cc t gcggaacac 1080
 t at t cagt gg ct t gaaaat g agct caacag at ggcgt aat ggggagacgg t gcct at t ga 1140
 t gaacagt t t gacaaagaga aagccaact t ggaagct t t c acagt ggat a aagat at t ac 1200

P042876Seq. TXT

t ct t accaat gat aaaccag caaccgcaat t ggagt t at a ggaaat t t t a ct gat gct ga	1260
aagaagaaag t gt gaagaag aaat t gct aa at t at acaaa cagct t gat g acaaggat ga	1320
agaaat t aac cagcaaagt c aact ggt aga gaaact gaag acgcaaat gt t ggat cagga	1380
ggagct t t t g gcat ct acca gaagggat ca agacaat at g caagct gagc t gaat cgcct	1440
t caagcagaa aat gat gcct ct aaagaaga agt gaaagaa gt t t t acagg ccct agaaga	1500
act t gct gt c aat t at gat c agaagt ct ca ggaagt t gaa gacaaaact a aggaat at ga	1560
at t gct t agt gat gaat t ga at cagaaat c ggcaact t t a gcgagt at ag at gct gagct	1620
t cagaaaact t aaggaaat ga ccaaccacca gaaaaaacga gcagct gaga t gat ggcct c	1680
t t t act aaaa gacct t gcag aaat aggaat t gct gt ggga aat aat gat g t aaagcagcc	1740
t gagggaaact ggcat gat ag at gaagagt t cact gt t gca agact ct aca t t agcaaaat	1800
gaagt cagaa gt aaaaacca t ggt gaaacg t t gcaagcag t t agaaagca cacaaact ga	1860
gagcaacaaa aaaat ggaag aaaat gaaaa ggagt t agca gcat gt cagc t t cgt at ct c	1920
t caacat gaa gccaaaat ca agt cat t gac t gaat acct t caaaat gt gg aacaaaagaa	1980
aagacagt t g gaggaat ct g t cgat gccct cagt gaagaa ct agt ccagc t t cgagcaca	2040
agagaaaagt c cat gaaat gg aaaaggagca ct t aaat aag gt t cagact g caaat gaagt	2100
t aagcaagct gt t gaacagc agat ccagag ccat agagaa act cat caaa aacagat cag	2160
t agt t t gaga gat gaagt ag aagcaaaagc aaaact t at t act gat ct t c aagaccaaaa	2220
ccagaaaat g at gt t agagc aggaacgt ct aagagt agaa cat gagaagt t gaaagccac	2280
agat caggaa aagagcagaa aact acat ga act t acggt t at gcaagat a gacgagaaca	2340
agcaagacaa gact t gaagg gt t t ggaaga gacagt ggca aaagaact t c agact t t aca	2400
caacct gcgc aaact ct t t g t t caggacct ggct acaaga gt t aaaaaga gt gct gagat	2460
t gat t ct gat gacaccggag gcagcgct gc t cagaagcaa aaaat ct cct t t ct t gaaaa	2520
t aat ct t gaa cagct cact a aagt gcacaa acagt t ggt a cgt gat aat g cagat ct ccg	2580
ct gt gaact t cct aagt t gg aaaagcgact t cgagct aca gct gagagag t gaaagct t t	2640
ggaat cagca ct gaaagaag ct aaagaaaa t gcat ct cgt gat cgcaaac gct at cagca	2700
agaagt agat cgcat aaagg aagcagt cag gt caaagaat at ggccagaa gagggcat t c	2760
t gcacagat t gt gt accgcc ggaagcacca ggagct gcaa gccat gcaga t ggagct gca	2820
gagccct gag t acaagct ga gcaagct ccg cacct cgacc at cat gaccg act acaaccc	2880
caact act gc t t t gct ggca agacct cct c cat cagt gac ct gaaggagg t gccgcggaa	2940
aaacat cacc ct cat t cggg gt ct gggcca t ggcgcct t t ggggaggt gt at gaaggcca	3000
ggt gt ccgga at gcccaacg acccaagccc cct gcaagt g gct gt gaaga cgct gcct ga	3060
agt gt gct ct gaacaggacg aact ggat t t cct cat ggaa gccct gat ca t cagcaaat t	3120
caaccaccag aacat t gt t c gct gcat t gg ggt gagcct g caat ccct gc cccggt t cat	3180
cct gct ggag ct cat ggcgg ggggagacct caagt cct t c ct ccgagaga cccgcct cg	3240

P042876Seq. TXT

```

cccgagccag ccct cct ccc t ggccat gct ggacct t ct g cacgt ggct c gggacat t gc 3300
ct gt ggct gt cagt at t t gg aggaaaacca ct t cat ccac cgagacat t g ct gccagaaa 3360
ct gcct ct t g acct gt ccag gccct ggaag agt ggccaag at t ggagact t cgggat ggc 3420
ccgagacat c t acagggcga gct act at ag aaagggaggc t gt gccat gc t gccagt t aa 3480
gt ggat gccc ccagaggcct t cat ggaagg aat at t cact t ct aaaacag acacat ggt c 3540
ct t t ggagt g ct gct at ggg aaat ct t t t c t ct t ggat at at gccat acc ccagcaaaag 3600
caaccaggaa gt t ct ggagt t t gt caccag t ggaggccgg at ggaccac ccaagaact g 3660
ccct gggcct gt at accgga t aat gact ca gt gct ggcaa cat cagcct g aagacaggcc 3720
caact t t gcc at cat t t t gg agaggat t ga at act gcacc caggaccgg at gt aat caa 3780
caccgct t t g ccgat agaat at ggt cact t gt ggaagag gaagagaaaag t gcct gt gag 3840
gccaaggac cct gaggggg t t cct cct ct cct ggt ct ct caacaggcaa aacgggagga 3900
ggagcgcagc ccagct gccc caccacct ct gcct accacc t cct ct ggca aggct gcaaa 3960
gaaaccaca gct gcagagg t ct ct gt t cg agt ccct aga gggccggccg t ggaaggggg 4020
acacgt gaat at ggcat t ct ct cagt ccaa ccct cct t cg gagt t gcaca aggt ccacgg 4080
at ccagaaac aagcccacca gct t gt ggaa cccaacgt ac ggct cct ggt t t acagagaa 4140
accacacaaa aagaat aat c ct at agcaaa gaaggagcca cagcacagg gt aacct ggg 4200
gct ggaggga agct gt act g t cccacct aa cgt t gcaact gggagact t c cgggggcct c 4260
act gct cct a gagccct ct t cgct gact gc caat at gaag gaggt acct c t gt t caggct 4320
acgt cact t c cct t gt gggga at gt caat t a cggct accag caacagggct t gccct t aga 4380
agccgct act gcccct ggag ct ggt cat t a cgaggat acc at t ct gaaaa gcaagaat ag 4440
cat gaaccag cct gggccct gagct cggc gcacact ca 4479

```

<210> 24

<211> 1483

<212> PRT

<213> Homo sapiens

<400> 24

Met Ala Asp Leu Ala Glu Cys Asn Ile Lys Val Met Cys Arg Phe Arg
1 5 10 15

Pro Leu Asn Glu Ser Glu Val Asn Arg Gly Asp Lys Tyr Ile Ala Lys
20 25 30

Phe Gln Gly Glu Asp Thr Val Val Ile Ala Ser Lys Pro Tyr Ala Phe
35 40 45

Asp Arg Val Phe Gln Ser Ser Thr Ser Gln Glu Gln Val Tyr Asn Asp
50 55 60

Cys Ala Lys Lys Ile Val Lys Asp Val Leu Glu Gly Tyr Asn Gly Thr
65 70 75 80

P042876Seq. TXT

I l e P h e A l a T y r G l y G n T h r S e r S e r G l y L y s T h r H i s T h r M e t G l u
85 90 95

G l y L y s L e u H i s A s p P r o G l u G l y M e t G l y I l e I l e P r o A r g I l e V a l
100 105 110

G n A s p I l e P h e A s n T y r I l e T y r S e r M e t A s p G l u A s n L e u G l u P h e
115 120 125

H i s I l e L y s V a l S e r T y r P h e G l u I l e T y r L e u A s p L y s I l e A r g A s p
130 135 140

L e u L e u A s p V a l S e r L y s T h r A s n L e u S e r V a l H i s G l u A s p L y s A s n
145 150 155 160

A r g V a l P r o T y r V a l L y s G l y C y s T h r G l u A r g P h e V a l C y s S e r P r o
165 170 175

A s p G l u V a l M e t A s p T h r I l e A s p G l u G l y L y s S e r A s n A r g H i s V a l
180 185 190

A l a V a l T h r A s n M e t A s n G l u H i s S e r S e r A r g S e r H i s S e r I l e P h e
195 200 205

L e u I l e A s n V a l L y s G n G l u A s n T h r G n T h r G l u G n L y s L e u S e r
210 215 220

G l y L y s L e u T y r L e u V a l A s p L e u A l a G l y S e r G l u L y s V a l S e r L y s
225 230 235 240

T h r G l y A l a G l u G l y A l a V a l L e u A s p G l u A l a L y s A s n I l e A s n L y s
245 250 255

S e r L e u S e r A l a L e u G l y A s n V a l I l e S e r A l a L e u A l a G l u G l y S e r
260 265 270

T h r T y r V a l P r o T y r A r g A s p S e r L y s M e t T h r A r g I l e L e u G n A s p
275 280 285

S e r L e u G l y G l y A s n C y s A r g T h r T h r I l e V a l I l e C y s C y s S e r P r o
290 295 300

S e r S e r T y r A s n G l u S e r G l u T h r L y s S e r T h r L e u L e u P h e G l y G n
305 310 315 320

A r g A l a L y s T h r I l e L y s A s n T h r V a l C y s V a l A s n V a l G l u L e u T h r
325 330 335

A l a G l u G n T r p L y s L y s L y s T y r G l u L y s G l u L y s G l u L y s A s n L y s
340 345 350

P042876Seq. TXT

I l e L e u A r g A s n T h r I l e G n T r p L e u G u A s n G u L e u A s n A r g T r p
 355 360 365
 A r g A s n G y G u T h r V a l P r o I l e A s p G u G n P h e A s p L y s G u L y s
 370 375 380
 A l a A s n L e u G u A l a P h e T h r V a l A s p L y s A s p I l e T h r L e u T h r A s n
 385 390 395 400
 A s p L y s P r o A l a T h r A l a I l e G y V a l I l e G y A s n P h e T h r A s p A l a
 405 410 415
 G u A r g A r g L y s C y s G u G u G u I l e A l a L y s L e u T y r L y s G n L e u
 420 425 430
 A s p A s p L y s A s p G u G u I l e A s n G n G n S e r G n L e u V a l G u L y s
 435 440 445
 L e u L y s T h r G n M e t L e u A s p G n G u G u L e u L e u A l a S e r T h r A r g
 450 455 460
 A r g A s p G n A s p A s n M e t G n A l a G u L e u A s n A r g L e u G n A l a G u
 465 470 475 480
 A s n A s p A l a S e r L y s G u G u V a l L y s G u V a l L e u G n A l a L e u G u
 485 490 495
 G u L e u A l a V a l A s n T y r A s p G n L y s S e r G n G u V a l G u A s p L y s
 500 505 510
 T h r L y s G u T y r G u L e u L e u S e r A s p G u L e u A s n G n L y s S e r A l a
 515 520 525
 T h r L e u A l a S e r I l e A s p A l a G u L e u G n L y s L e u L y s G u M e t T h r
 530 535 540
 A s n H i s G n L y s L y s A r g A l a A l a G u M e t M e t A l a S e r L e u L e u L y s
 545 550 555 560
 A s p L e u A l a G u I l e G y I l e A l a V a l G y A s n A s n A s p V a l L y s G n
 565 570 575
 P r o G u G y T h r G y M e t I l e A s p G u G u P h e T h r V a l A l a A r g L e u
 580 585 590
 T y r I l e S e r L y s M e t L y s S e r G u V a l L y s T h r M e t V a l L y s A r g C y s
 595 600 605
 L y s G n L e u G u S e r T h r G n T h r G u S e r A s n L y s L y s M e t G u G u
 610 615 620

P042876Seq. TXT

Asn 625 Glu Lys Glu Leu Ala 630 Ala Cys Gln Leu Arg 635 Ile Ser Gln His Glu 640
 Ala Lys Ile Lys Ser 645 Leu Thr Glu Tyr Leu 650 Gln Asn Val Glu Gln Lys 655
 Lys Arg Gln Leu 660 Glu Glu Ser Val Asp 665 Ala Leu Ser Glu Glu Leu Val 670
 Gln Leu Arg 675 Ala Gln Glu Lys Val 680 His Glu Met Glu Lys 685 Glu His Leu
 Asn 690 Lys Val Gln Thr Ala 695 Asn Glu Val Lys Gln Ala 700 Val Glu Gln Gln
 Ile 705 Gln Ser His Arg Glu 710 Thr His Gln Lys Gln 715 Ile Ser Ser Leu Arg 720
 Asp Glu Val Glu 725 Ala Lys Ala Lys Leu Ile 730 Thr Asp Leu Gln Asp 735 Gln
 Asn Gln Lys Met 740 Met Leu Glu Gln Glu 745 Arg Leu Arg Val Glu 750 His Glu
 Lys Leu Lys 755 Ala Thr Asp Gln Glu 760 Lys Ser Arg Lys Leu 765 His Glu Leu
 Thr 770 Val Met Gln Asp Arg Arg 775 Glu Gln Ala Arg Gln 780 Asp Leu Lys Gly
 Leu 785 Glu Glu Thr Val Ala 790 Lys Glu Leu Gln Thr 795 Leu His Asn Leu Arg 800
 Lys Leu Phe Val 805 Gln Asp Leu Ala Thr Arg 810 Val Lys Lys Ser Ala 815 Glu
 Ile Asp Ser Asp 820 Asp Thr Gly Gly Ser 825 Ala Ala Gln Lys Gln 830 Lys Ile
 Ser Phe Leu 835 Glu Asn Asn Leu Glu 840 Gln Leu Thr Lys Val 845 His Lys Gln
 Leu 850 Val Arg Asp Asn Ala Asp 855 Leu Arg Cys Glu 860 Leu Pro Lys Leu Glu
 Lys 865 Arg Leu Arg Ala Thr 870 Ala Glu Arg Val Lys 875 Ala Leu Glu Ser Ala 880
 Leu Lys Glu Ala 885 Lys Glu Asn Ala Ser Arg 890 Asp Arg Lys Arg Tyr 895 Gln

P042876Seq. TXT

G n G u Val Asp Arg I l e Lys G u A l a Val Arg Ser Lys Asn M e t A l a
 900 905 910
 Arg Arg G y H i s Ser A l a G n I l e Val Tyr Arg Arg Lys H i s G n G u
 915 920 925
 Leu G n A l a M e t G n M e t G u Leu G n Ser Pro G u Tyr Lys Leu Ser
 930 935 940
 Lys Leu Arg Thr Ser Thr I l e M e t Thr Asp Tyr Asn Pro Asn Tyr Cys
 945 950 955 960
 Phe A l a G y Lys Thr Ser Ser I l e Ser Asp Leu Lys G u Val Pro Arg
 965 970 975
 Lys Asn I l e Thr Leu I l e Arg G y Leu G y H i s G y A l a Phe G y G u
 980 985 990
 Val Tyr G u G y G n Val Ser G y M e t Pro Asn Asp Pro Ser Pro Leu
 995 1000 1005
 G n Val A l a Val Lys Thr Leu Pro G u Val Cys Ser G u G n Asp
 1010 1015 1020
 G u Leu Asp Phe Leu M e t G u A l a Leu I l e I l e Ser Lys Phe Asn
 1025 1030 1035
 H i s G n Asn I l e Val Arg Cys I l e G y Val Ser Leu G n Ser Leu
 1040 1045 1050
 Pro Arg Phe I l e Leu Leu G u Leu M e t A l a G y G y Asp Leu Lys
 1055 1060 1065
 Ser Phe Leu Arg G u Thr Arg Pro Arg Pro Ser G n Pro Ser Ser
 1070 1075 1080
 Leu A l a M e t Leu Asp Leu Leu H i s Val A l a Arg Asp I l e A l a Cys
 1085 1090 1095
 G y Cys G n Tyr Leu G u G u Asn H i s Phe I l e H i s Arg Asp I l e
 1100 1105 1110
 A l a A l a Arg Asn Cys Leu Leu Thr Cys Pro G y Pro G y Arg Val
 1115 1120 1125
 A l a Lys I l e G y Asp Phe G y M e t A l a Arg Asp I l e Tyr Arg A l a
 1130 1135 1140
 Ser Tyr Tyr Arg Lys G y G y Cys A l a M e t Leu Pro Val Lys Trp
 1145 1150 1155

P042876Seq. TXT

Met	Pro	Pro	Glu	Ala	Phe	Met	Glu	Gly	Ile	Phe	Thr	Ser	Lys	Thr
	1160					1165					1170			
Asp	Thr	Trp	Ser	Phe	Gly	Val	Leu	Leu	Trp	Glu	Ile	Phe	Ser	Leu
	1175					1180					1185			
Gly	Tyr	Met	Pro	Tyr	Pro	Ser	Lys	Ser	Asn	Gln	Glu	Val	Leu	Glu
	1190					1195					1200			
Phe	Val	Thr	Ser	Gly	Gly	Arg	Met	Asp	Pro	Pro	Lys	Asn	Cys	Pro
	1205					1210					1215			
Gly	Pro	Val	Tyr	Arg	Ile	Met	Thr	Gln	Cys	Trp	Gln	His	Gln	Pro
	1220					1225					1230			
Glu	Asp	Arg	Pro	Asn	Phe	Ala	Ile	Ile	Leu	Glu	Arg	Ile	Glu	Tyr
	1235					1240					1245			
Cys	Thr	Gln	Asp	Pro	Asp	Val	Ile	Asn	Thr	Ala	Leu	Pro	Ile	Glu
	1250					1255					1260			
Tyr	Gly	Pro	Leu	Val	Glu	Glu	Glu	Glu	Lys	Val	Pro	Val	Arg	Pro
	1265					1270					1275			
Lys	Asp	Pro	Glu	Gly	Val	Pro	Pro	Leu	Leu	Val	Ser	Gln	Gln	Ala
	1280					1285					1290			
Lys	Arg	Glu	Glu	Glu	Arg	Ser	Pro	Ala	Ala	Pro	Pro	Pro	Leu	Pro
	1295					1300					1305			
Thr	Thr	Ser	Ser	Gly	Lys	Ala	Ala	Lys	Lys	Pro	Thr	Ala	Ala	Glu
	1310					1315					1320			
Val	Ser	Val	Arg	Val	Pro	Arg	Gly	Pro	Ala	Val	Glu	Gly	Gly	His
	1325					1330					1335			
Val	Asn	Met	Ala	Phe	Ser	Gln	Ser	Asn	Pro	Pro	Ser	Glu	Leu	His
	1340					1345					1350			
Lys	Val	His	Gly	Ser	Arg	Asn	Lys	Pro	Thr	Ser	Leu	Trp	Asn	Pro
	1355					1360					1365			
Thr	Tyr	Gly	Ser	Trp	Phe	Thr	Glu	Lys	Pro	Thr	Lys	Lys	Asn	Asn
	1370					1375					1380			
Pro	Ile	Ala	Lys	Lys	Glu	Pro	His	Asp	Arg	Gly	Asn	Leu	Gly	Leu
	1385					1390					1395			
Glu	Gly	Ser	Cys	Thr	Val	Pro	Pro	Asn	Val	Ala	Thr	Gly	Arg	Leu
	1400					1405					1410			

P042876Seq. TXT

Pro Gly Ala Ser Leu Leu Leu Glu Pro Ser Ser Leu Thr Ala Asn
 1415 1420 1425

Met Lys Glu Val Pro Leu Phe Arg Leu Arg His Phe Pro Cys Gly
 1430 1435 1440

Asn Val Asn Tyr Gly Tyr Gln Gln Gln Gly Leu Pro Leu Glu Ala
 1445 1450 1455

Ala Thr Ala Pro Gly Ala Gly His Tyr Glu Asp Thr Ile Leu Lys
 1460 1465 1470

Ser Lys Asn Ser Met Asn Gln Pro Gly Pro
 1475 1480

<210> 25
 <211> 2412
 <212> DNA
 <213> Homo sapiens

<400> 25
 at gaacggac agt t ggat ct aagt gggaag ct aat cat ca aagct caact t ggggaggat 60
 at t cggcgaa t t cct at t ca t aat gaagat at t act t at g at gaat t agt gct aat gat g 120
 caacgagt t t t cagaggaaa act t ct gagt aat gat gaag t aacaat aaa gt at aaagat 180
 gaagat ggag at ct t at aac aat t t t t gat agt t ct gacc t t t cct t t gc aat t cagt gc 240
 agt aggat ac t gaaact gac at t at t t gt t aat ggccagc caagaccct t gaat caagt 300
 caggt gaaat at ct ccgt cg agaact gat a gaact t cgaa at aaagt gaa t cgt t t at t g 360
 gat agct t gg aaccacct gg agaaccagga cct t ccacca at at t cct ga aaat gat act 420
 gt ggat ggt a gggaagaaaa gt ct gct t ct gat t ct t ct g gaaaacagt c t act caggt t 480
 at ggcagcaa gt at gt ct gc t t t t gat cct t t aaaaaacc aagat gaaat caat aaaaat 540
 gt t at gt cag cgt t t ggct t aacagat gat caggt t t cag ggccaccag t gct cct gca 600
 gaagat cgt t caggaacacc cgacagcat t gct t cct cct cct cagcagc t caccacca 660
 ggcgt t cagc cacagcagcc accat at aca ggagct caga ct caagcagg t cagat t gaa 720
 gt gt accgcc ggaagcacca ggagct gcaa gccat gcaga t ggagct gca gagccct gag 780
 t acaagct ga gcaagct ccg cacct cgacc at cat gaccg act acaacc caact act gc 840
 t t t gct ggca agacct cct c cat cagt gac ct gaaggagg t gccgcggaa aaacat cacc 900
 ct cat t cggg gt ct gggcca t ggcgccct t t ggggaggt gt at gaaggcca ggt gt ccgga 960
 at gcccaacg acccaagccc cct gcaagt g gct gt gaaga cgct gcct ga agt gt gct ct 1020
 gaacaggacg aact ggat t t cct cat ggaa gccct gat ca t cagcaaat t caaccaccag 1080
 aacat t gt t c gct gcat t gg ggt gagcct g caat ccct gc cccggt t cat cct gct ggag 1140
 ct cat ggcgg ggggagacct caagt cct t c ct ccgagaga cccgccct cg cccgagccag 1200

P042876Seq. TXT

```

ccct cct ccc t ggccat gct ggacct t ct g cacgt ggct c gggacat t gc ct gt ggct gt 1260
cagt at t t gg aggaaaacca ct t cat ccac cgagacat t g ct gccagaaa ct gcct ct t g 1320
acct gt ccag gccct ggaag agt ggccaag at t ggagact t cgggat ggc ccgagacat c 1380
t acagggcga gct act at ag aaagggaggc t gt gccat gc t gccagt t aa gt ggat gccc 1440
ccagaggcct t cat ggaagg aat at t cact t ct aaaacag acacat ggt c ct t t ggagt g 1500
ct gct at ggg aaat ct t t t c t ct t ggat at at gccat acc ccagcaaaag caaccaggaa 1560
gt t ct ggagt t t gt caccag t ggaggccgg at ggaccac ccaagaact g ccct gggcct 1620
gt at accgga t aat gact ca gt gct ggcaa cat cagcct g aagacaggcc caact t t gcc 1680
at cat t t t gg agaggat t ga at act gcacc caggaccgg at gt aat caa caccgct t t g 1740
ccgat agaat at ggt ccact t gt ggaagag gaagagaaag t gcct gt gag gccaaggac 1800
cct gaggggg t t cct cct ct cct ggt ct ct caacaggcaa aacgggagga ggagcgcagc 1860
ccagct gccc caccacct ct gcct accacc t cct ct ggca aggct gcaaa gaaaccacaa 1920
gct gcagagg t ct ct gt t cg agt ccct aga gggccggccg t ggaaggggg acacgt gaat 1980
at ggcat t ct ct cagt ccaa ccct cct t cg gagt t gcaca aggt ccacgg at ccagaaac 2040
aagcccacca gct t gt ggaa cccaacgt ac ggct cct ggt t t acagagaa acccaccaaa 2100
aagaat aat c ct at agcaaa gaaggagcca cagcagagg gt aacct ggg gct ggagggga 2160
agct gt act g t cccacct aa cgt t gcaact gggagact t c cgggggcct c act gct cct a 2220
gagccct ct t cgct gact gc caat at gaag gaggt acct c t gt t caggct acgt cact t c 2280
cct t gt ggga at gt caat t a cggct accag caacagggt t gccct t aga agccgct act 2340
gcccct ggag ct ggt cat t a cgaggat acc at t ct gaaaa gcaagaat ag cat gaaccag 2400
cct gggccct ga 2412

```

<210> 26
 <211> 803
 <212> PRT
 <213> Homo sapiens

<400> 26
 Met Asn Gly Gln Leu Asp Leu Ser Gly Lys Leu Ile Ile Lys Ala Gln
 1 5 10 15

Leu Gly Glu Asp Ile Arg Arg Ile Pro Ile His Asn Glu Asp Ile Thr
 20 25 30

Tyr Asp Glu Leu Val Leu Met Met Gln Arg Val Phe Arg Gly Lys Leu
 35 40 45

Leu Ser Asn Asp Glu Val Thr Ile Lys Tyr Lys Asp Glu Asp Gly Asp
 50 55 60

Leu Ile Thr Ile Phe Asp Ser Ser Asp Leu Ser Phe Ala Ile Gln Cys
 65 70 75 80

P042876Seq. TXT

Ser Arg Ile Leu Lys₈₅ Leu Thr Leu Phe Val₉₀ Asn Gly Gln Pro Arg₉₅ Pro

Leu Gu Ser Ser₁₀₀ Gln Val Lys Tyr Leu₁₀₅ Arg Arg Gu Leu Ile₁₁₀ Gu Leu

Arg Asn Lys₁₁₅ Val Asn Arg Leu Leu₁₂₀ Asp Ser Leu Gu Pro₁₂₅ Pro Gly Gu

Pro Gly₁₃₀ Pro Ser Thr Asn Ile₁₃₅ Pro Gu Asn Asp Thr₁₄₀ Val Asp Gly Arg

Gu₁₄₅ Gu Lys Ser Ala Ser₁₅₀ Asp Ser Ser Gly Lys₁₅₅ Gln Ser Thr Gln Val₁₆₀

Met Ala Ala Ser Met₁₆₅ Ser Ala Phe Asp Pro₁₇₀ Leu Lys Asn Gln Asp₁₇₅ Gu

Ile Asn Lys Asn₁₈₀ Val Met Ser Ala Phe₁₈₅ Gly Leu Thr Asp Asp₁₉₀ Gln Val

Ser Gly₁₉₅ Pro Pro Ser Ala Pro Ala₂₀₀ Gu Asp Arg Ser Gly₂₀₅ Thr Pro Asp

Ser Ile₂₁₀ Ala Ser Ser Ser Ser₂₁₅ Ala Ala His Pro₂₂₀ Pro Gly Val Gln Pro

Gln₂₂₅ Gln Pro Pro Tyr Thr₂₃₀ Gly Ala Gln Thr Gln₂₃₅ Ala Gly Gln Ile Gu₂₄₀

Val Tyr Arg Arg Lys₂₄₅ His Gln Gu Leu Gln₂₅₀ Ala Met Gln Met₂₅₅ Gu Leu

Gln Ser Pro Gu₂₆₀ Tyr Lys Leu Ser Lys₂₆₅ Leu Arg Thr Ser Thr₂₇₀ Ile Met

Thr Asp Tyr₂₇₅ Asn Pro Asn Tyr Cys₂₈₀ Phe Ala Gly Lys Thr₂₈₅ Ser Ser Ile

Ser Asp₂₉₀ Leu Lys Gu Val Pro₂₉₅ Arg Lys Asn Ile Thr₃₀₀ Leu Ile Arg Gly

Leu Gly His Gly Ala Phe₃₁₀ Gly Gu Val Tyr Gu₃₁₅ Gly Gln Val Ser Gly₃₂₀

Met Pro Asn Asp Pro₃₂₅ Ser Pro Leu Gln Val₃₃₀ Ala Val Lys Thr Leu₃₃₅ Pro

Gu Val Cys Ser₃₄₀ Gu Gln Asp Gu Leu₃₄₅ Asp Phe Leu Met Gu₃₅₀ Ala Leu

P042876Seq. TXT

I l e I l e S e r L y s P h e A s n H i s G n A s n I l e V a l A r g C y s I l e G l y V a l
355 360 365

S e r L e u G n S e r L e u P r o A r g P h e I l e L e u L e u G l u L e u M e t A l a G l y
370 375 380

G l y A s p L e u L y s S e r P h e L e u A r g G l u T h r A r g P r o A r g P r o S e r G n
385 390 395 400

P r o S e r S e r L e u A l a M e t L e u A s p L e u L e u H i s V a l A l a A r g A s p I l e
405 410 415

A l a C y s G l y C y s G n T y r L e u G l u G l u A s n H i s P h e I l e H i s A r g A s p
420 425 430

I l e A l a A l a A r g A s n C y s L e u L e u T h r C y s P r o G l y P r o G l y A r g V a l
435 440 445

A l a L y s I l e G l y A s p P h e G l y M e t A l a A r g A s p I l e T y r A r g A l a S e r
450 455 460

T y r T y r A r g L y s G l y G l y C y s A l a M e t L e u P r o V a l L y s T r p M e t P r o
465 470 475 480

P r o G l u A l a P h e M e t G l u G l y I l e P h e T h r S e r L y s T h r A s p T h r T r p
485 490 495

S e r P h e G l y V a l L e u L e u T r p G l u I l e P h e S e r L e u G l y T y r M e t P r o
500 505 510

T y r P r o S e r L y s S e r A s n G n G l u V a l L e u G l u P h e V a l T h r S e r G l y
515 520 525

G l y A r g M e t A s p P r o P r o L y s A s n C y s P r o G l y P r o V a l T y r A r g I l e
530 535 540

M e t T h r G n C y s T r p G n H i s G n P r o G l u A s p A r g P r o A s n P h e A l a
545 550 555 560

I l e I l e L e u G l u A r g I l e G l u T y r C y s T h r G n A s p P r o A s p V a l I l e
565 570 575

A s n T h r A l a L e u P r o I l e G l u T y r G l y P r o L e u V a l G l u G l u G l u G l u
580 585 590

L y s V a l P r o V a l A r g P r o L y s A s p P r o G l u G l y V a l P r o P r o L e u L e u
595 600 605

V a l S e r G n G n A l a L y s A r g G l u G l u G l u A r g S e r P r o A l a A l a P r o
610 615 620

P042876Seq. TXT

Pro Pro Leu Pro Thr Thr Ser Ser Gly Lys Ala Ala Lys Lys Pro Thr
625 630 635 640

Ala Ala Gu Val Ser Val Arg Val Pro Arg Gly Pro Ala Val Gu Gly
645 650 655

Gly His Val Asn Met Ala Phe Ser Gln Ser Asn Pro Pro Ser Gu Leu
660 665 670

His Lys Val His Gly Ser Arg Asn Lys Pro Thr Ser Leu Trp Asn Pro
675 680 685

Thr Tyr Gly Ser Trp Phe Thr Gu Lys Pro Thr Lys Lys Asn Asn Pro
690 695 700

Ile Ala Lys Lys Gu Pro His Asp Arg Gly Asn Leu Gly Leu Gu Gly
705 710 715 720

Ser Cys Thr Val Pro Pro Asn Val Ala Thr Gly Arg Leu Pro Gly Ala
725 730 735

Ser Leu Leu Leu Gu Pro Ser Ser Leu Thr Ala Asn Met Lys Gu Val
740 745 750

Pro Leu Phe Arg Leu Arg His Phe Pro Cys Gly Asn Val Asn Tyr Gly
755 760 765

Tyr Gln Gln Gln Gly Leu Pro Leu Gu Ala Ala Thr Ala Pro Gly Ala
770 775 780

Gly His Tyr Gu Asp Thr Ile Leu Lys Ser Lys Asn Ser Met Asn Gln
785 790 795 800

Pro Gly Pro

<210> 27
<211> 2779
<212> DNA
<213> Homo sapiens

<400> 27
cct ccgcaag ccgt ctttct ct agagt tgt at at at agaa cat cct ggag t ccacat ga 60
acggacagt t ggat ct aagt gggaagct aa t cat caaagc t caact t ggg gaggat at t c 120
ggcgaat t cc t at t cat aat gaagat at t a ct t at gat ga at t agt gct a at gat gcaac 180
gagt ttt cag aggaaaact t ct gagt aat g at gaagt aac aat aaagt at aaagat gaag 240
at ggagat ct t at aacaat t ttt gat agt t ct gacct ttc cttt gcaat t cagt gcagt a 300
ggat act gaa act gacat t a ttt gtt aat g gccagccaag acccct t gaa t caagt cagg 360
t gaaat at ct ccgt cgagaa ct gat agaac ttc gaaat aa agt gaat cgt t t at t ggat a 420

P042876Seq. TXT

gct t ggaacc	acct ggagaa	ccaggacct t	ccaccaat at	t cct gaaaat	gat act gt gg	480
at ggt agggga	agaaaagt ct	gct t ct gat t	ct t ct ggaaa	acagt ct act	caggt t at gg	540
cagcaagt at	gt ct gct t t t	gat cct t t aa	aaaaccaaga	t gaaat caat	aaaaat gt t a	600
t gt cagcgt t	t ggct t aaca	gat gat cagg	t t t cagt gt a	ccgccggaag	caccaggagc	660
t gcaagccat	gcagat ggag	ct gcagagcc	ct gagt acaa	gct gagcaag	ct ccgcacct	720
cgacct cat	gaccgact ac	aaccccaact	act gct t t gc	t ggcaagacc	t cct ccat ca	780
gt gacct gaa	ggaggt gccg	cggaaaaaca	t caccct cat	t cgggggt ct g	ggccat ggcg	840
cct t t gggga	gggt gt at gaa	ggccaggt gt	ccggaat gcc	caacgaccca	agccccct gc	900
aagt ggct gt	gaagacgct g	cct gaagt gt	gct ct gaaca	ggacgaact g	gat t t cct ca	960
t ggaagccct	gat cat cagc	aaat t caacc	accagaacat	t gt t cgct gc	at t ggggt ga	1020
gcct gcaat c	cct gccccg	t t cat cct gc	t ggagct cat	ggcgggggga	gacct caagt	1080
cct t cct ccg	agagaccgc	cct cgccccga	gccagccct c	ct ccct ggcc	at gct ggacc	1140
t t ct gcacgt	ggct cgggac	at t gcct gt g	gct gt cagt a	t t t ggaggaa	aaccact t ca	1200
t ccaccgaga	cat t gct gcc	agaaact gcc	t ct t gacct g	t ccaggccct	ggaagagt gg	1260
ccaagat t gg	agact t cggg	at ggcccag	acat ct acag	ggcgagct ac	t at agaaagg	1320
gaggct gt gc	cat gct gcca	gt t aagt gga	t gccccaga	ggcct t cat g	gaaggaat at	1380
t cact t ct aa	aacagacaca	t ggt cct t t g	gagt gct gct	at gggaaat c	t t t t ct ct t g	1440
gat at at gcc	at accccagc	aaaagcaacc	aggaagt t ct	ggagt t t gt c	accagt ggag	1500
gccggat gga	cccaccaag	aact gccct g	ggcct gt at a	ccggat aat g	act cagt gct	1560
ggcaacat ca	gcct gaagac	aggcccaact	t t gccat cat	t t t ggagagg	at t gaat act	1620
gcaccagga	cccggat gt a	at caacaccg	ct t t gccgat	agaat at ggt	ccact t gt gg	1680
aagaggaaga	gaaagt gcct	gt gaggccca	aggacct ga	gggggt t cct	cct ct cct gg	1740
t ct ct caaca	ggcaaacgg	gaggaggagc	gcagcccagc	t gccccacca	cct ct gcct a	1800
ccacct cct c	t ggcaaggct	gcaaagaaac	ccacagct gc	agaggt ct ct	gt t cgagt cc	1860
ct agagggcc	ggccgt ggaa	gggggacacg	t gaat at ggc	at t ct ct cag	t ccaaccct c	1920
ct t cggagt t	gcacaaggt c	cacggat cca	gaaacaagcc	caccagct t g	t ggaacccaa	1980
cgt acggct c	ct ggt t t aca	gagaaacca	ccaaaaagaa	t aat cct at a	gcaaagaagg	2040
agccacacga	caggggt aac	ct ggggct gg	agggaagct g	t act gt ccca	cct aacgt t g	2100
caact gggag	act t ccgggg	gcct cact gc	t cct agagcc	ct ct t cgct g	act gccaat a	2160
t gaaggaggt	acct ct gt t c	aggct acgt c	act t ccct t g	t gggaat gt c	aat t acggct	2220
accagcaaca	gggt t t gcc	t t agaagccg	ct act gccc	t ggagct ggt	cat t acgagg	2280
at accat t ct	gaaaagcaag	aat agcat ga	accagcct gg	gccct gagct	cggt cgcaca	2340
ct cact t ct c	t t cct t ggga	t ccct aagac	cgt ggaggag	agagaggcaa	t ggct cct t c	2400
acaaaccaga	gaccaaat gt	cacgt t t t gt	t t t gt gccaa	cct at t t t ga	agt accacca	2460

P042876Seq. TXT

aaaaagctgt attttgaaaa tgctttagaa aggttttgag catgggttca tcctattctt 2520
 tcgaaagaag aaaatatcat aaaaatgagt gataaataca aggcccagat gtggttgcatt 2580
 aagggttttta tgcattgttg ttgtatactt ccttatgctt cttttaaat t gtgtgtgctc 2640
 tgcttcaatg tagtcagaat tagctgcttc tatgtttcat agttgggggc atagatgttt 2700
 ccttgccttg ttgatgtgga catgagccat ttgagggggag agggaacgga aat aaaggag 2760
 ttatttgt aa tgactaaaa 2779

<210> 28
 <211> 756
 <212> PRT
 <213> Homo sapiens

<400> 28
 Met Asn Gly Gln Leu Asp Leu Ser Gly Lys Leu Ile Ile Lys Ala Gln
 1 5 10 15

Leu Gly Glu Asp Ile Arg Arg Ile Pro Ile His Asn Glu Asp Ile Thr
 20 25 30

Tyr Asp Glu Leu Val Leu Met Met Gln Arg Val Phe Arg Gly Lys Leu
 35 40 45

Leu Ser Asn Asp Glu Val Thr Ile Lys Tyr Lys Asp Glu Asp Gly Asp
 50 55 60

Leu Ile Thr Ile Phe Asp Ser Ser Asp Leu Ser Phe Ala Ile Gln Cys
 65 70 75 80

Ser Arg Ile Leu Lys Leu Thr Leu Phe Val Asn Gly Gln Pro Arg Pro
 85 90 95

Leu Glu Ser Ser Gln Val Lys Tyr Leu Arg Arg Glu Leu Ile Glu Leu
 100 105 110

Arg Asn Lys Val Asn Arg Leu Leu Asp Ser Leu Glu Pro Pro Gly Glu
 115 120 125

Pro Gly Pro Ser Thr Asn Ile Pro Glu Asn Asp Thr Val Asp Gly Arg
 130 135 140

Glu Glu Lys Ser Ala Ser Asp Ser Ser Gly Lys Gln Ser Thr Gln Val
 145 150 155 160

Met Ala Ala Ser Met Ser Ala Phe Asp Pro Leu Lys Asn Gln Asp Glu
 165 170 175

Ile Asn Lys Asn Val Met Ser Ala Phe Gly Leu Thr Asp Asp Gln Val
 180 185 190

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

Pro Tyr Pro Ser Lys Ser Asn Gln Glu Val Leu Glu Phe Val Thr Ser
 465 470 475 480
 Gly Gly Arg Met Asp Pro Pro Lys Asn Cys Pro Gly Pro Val Tyr Arg
 485 490 495
 Ile Met Thr Gln Cys Trp Gln His Gln Pro Glu Asp Arg Pro Asn Phe
 500 505 510 515
 Ala Ile Ile Leu Glu Arg Ile Glu Tyr Cys Thr Gln Asp Pro Asp Val
 515 520 525
 Ile Asn Thr Ala Leu Pro Ile Glu Tyr Gly Pro Leu Val Glu Glu Glu
 530 535 540
 Glu Lys Val Pro Val Arg Pro Lys Asp Pro Glu Gly Val Pro Pro Leu
 545 550 555 560
 Leu Val Ser Gln Gln Ala Lys Arg Glu Glu Glu Arg Ser Pro Ala Ala
 565 570 575
 Pro Pro Pro Leu Pro Thr Thr Ser Ser Gly Lys Ala Ala Lys Lys Pro
 580 585 590
 Thr Ala Ala Glu Val Ser Val Arg Val Pro Arg Gly Pro Ala Val Glu
 595 600 605
 Gly Gly His Val Asn Met Ala Phe Ser Gln Ser Asn Pro Pro Ser Glu
 610 615 620
 Leu His Lys Val His Gly Ser Arg Asn Lys Pro Thr Ser Leu Trp Asn
 625 630 635 640
 Pro Thr Tyr Gly Ser Trp Phe Thr Glu Lys Pro Thr Lys Lys Asn Asn
 645 650 655
 Pro Ile Ala Lys Lys Glu Pro His Asp Arg Gly Asn Leu Gly Leu Glu
 660 665 670
 Gly Ser Cys Thr Val Pro Pro Asn Val Ala Thr Gly Arg Leu Pro Gly
 675 680 685
 Ala Ser Leu Leu Leu Glu Pro Ser Ser Leu Thr Ala Asn Met Lys Glu
 690 695 700
 Val Pro Leu Phe Arg Leu Arg His Phe Pro Cys Gly Asn Val Asn Tyr
 705 710 715 720
 Gly Tyr Gln Gln Gln Gly Leu Pro Leu Glu Ala Ala Thr Ala Pro Gly
 725 730 735

P042876Seq. TXT

Ala	Gly	His	Tyr	Glu	Asp	Thr	Ile	Leu	Lys	Ser	Lys	Asn	Ser	Met	Asn
			740					745					750		

Gln Pro Gly Pro
755

<210> 29
 <211> 2614
 <212> DNA
 <213> Homo sapiens

<400> 29	cct ccgcaag ccgt ct t t ct ct agagt t gt at at at agaa cat cct ggag t ccacat ga	60
	acggacagt t ggat ct aagt gggaagct aa t cat caaagc t caact t ggg gaggat at t c	120
	ggcgaat t cc t at t cat aat gaagat at t a ct t at gat ga at t agt gct a at gat gcaac	180
	gagt t t t cag aggaaaact t ct gagt aat g at gaagt aac aat aaagt at aaagat gaag	240
	at ggagat ct t at aacaat t t t t gat agt t ct gacct t t c ct t t gcaat t cagt gcagt a	300
	ggat act gaa act gacat t a t t t gt t aat g gccagccaag acccct t gaa t caagt cagg	360
	t gaaat at ct ccgt cgagaa ct gat agaac t t cgaaat aa agt gaat cgt t t at t ggat a	420
	gct t ggaacc acct ggagaa ccaggacct t ccaccaat at t cct gaaaat gt gt accgcc	480
	ggaagcacca ggagct gcaa gccat gcaga t ggagct gca gagccct gag t acaagct ga	540
	gcaagct ccg cacct cgacc at cat gaccg act acaaccc caact act gc t t t gct ggca	600
	agacct cct c cat cagt gac ct gaaggagg t gccgcggaa aaacat cacc ct cat t cggg	660
	gt ct gggcca t ggcgcct t t ggggaggt gt at gaaggcca ggt gt ccgga at gccaacg	720
	acccaagccc cct gcaagt g gct gt gaaga cgct gcct ga agt gt gct ct gaacaggacg	780
	aact ggat t t cct cat ggaa gccct gat ca t cagcaaat t caaccaccag aacat t gt t c	840
	gct gcat t gg ggt gacct g caat ccct gc cccggt t cat cct gct ggag ct cat ggcgg	900
	ggggagacct caagt cct t c ct ccgagaga cccgccct cg cccgagccag ccct cct ccc	960
	t ggccat gct ggacct t ct g cacgt ggct c gggacat t gc ct gt ggct gt cagt at t t gg	1020
	aggaaaacca ct t cat ccac cgagacat t g ct gccagaaa ct gcct ct t g acct gt ccag	1080
	gccct ggaag agt ggccaag at t ggagact t cgggat ggc ccgagacat c t acagggcga	1140
	gct act at ag aaagggaggc t gt gccat gc t gccagt t aa gt ggat gcc ccagaggcct	1200
	t cat ggaagg aat at t cact t ct aaaacag acacat ggt c ct t t ggagt g ct gct at ggg	1260
	aaat ct t t t c t ct t ggat at at gccat acc ccagcaaaag caaccaggaa gt t ct ggagt	1320
	t t gt caccag t ggaggccgg at ggaccac ccaagaact g ccct gggcct gt at accgga	1380
	t aat gact ca gt gct ggcaa cat cagcct g aagacaggcc caact t t gcc at cat t t t gg	1440
	agaggat t ga at act gcacc caggaccgg at gt aat caa caccgct t t g ccgat agaat	1500
	at ggt ccaact t gt ggaagag gaagagaaag t gcct gt gag gcccaaggac cct gaggggg	1560
	t t cct cct ct cct ggt ct ct caacaggcaa aacgggagga ggagcgcagc ccagct gcc	1620

P042876Seq. TXT

caccacct ct gcct accacc t cct ct ggca aggct gcaaa gaaacccaca gct gcagagg 1680
t ct ct gt t cg agt ccct aga gggccggccg t ggaaggggg acacgt gaat at ggcat t ct 1740
ct cagt ccaa ccct cct t cg gagt t gcaca aggt ccacgg at ccagaaac aagcccacca 1800
gct t gt ggaa cccaacgt ac ggct cct ggt t t acagagaa acccaccaaa aagaat aat c 1860
ct at agcaaa gaaggagcca cacgacaggg gt aacct ggg gct ggaggga agct gt act g 1920
t cccacct aa cgt t gcaact gggagact t c cgggggcct c act gct cct a gagccct ct t 1980
cgct gact gc caat at gaag gaggt acct c t gt t caggct acgt cact t c cct t gt ggga 2040
at gt caat t a cggct accag caacagggct t gccct t aga agccgct act gccct ggag 2100
ct ggt cat t a cgaggat acc at t ct gaaaa gcaagaat ag cat gaaccag cct gggccct 2160
gagct cggc gcacact cac t t ct ct t cct t gggat ccct aagaccgt gg aggagagaga 2220
ggcaat ggct cct t caaaa ccagagacca aat gt cacgt t t t gt t t t gt gccaacct at 2280
t t t gaagt ac caccaaaaaa gct gt at t t t gaaaat gct t t agaaaggt t t t gagcat gg 2340
gt t cat cct a t t ct t t cgaa agaagaaaat at cat aaaaa t gagt gat aa at acaaggcc 2400
cagat gt ggt t gcat aagg t t t t at gcat gt t t gt t gt a t act t cct t a t gct t ct t t t 2460
aaat t gt gt g t gct ct gct t caat gt agt c agaat t agct gct t ct at gt t t cat agt t g 2520
gggt cat aga t gt t t cct t g cct t gt t gat gt ggacat ga gccat t t gag gggagagggga 2580
acggaaat aa aggagt t at t t gt aat gact aaaa 2614

<210> 30

<211> 701

<212> PRT

<213> Homo sapiens

<400> 30

Met Asn Gly Gln Leu Asp Leu Ser Gly Lys Leu Ile Ile Lys Ala Gln
1 5 10 15

Leu Gly Gu Asp Ile Arg Arg Ile Pro Ile His Asn Glu Asp Ile Thr
20 25 30

Tyr Asp Glu Leu Val Leu Met Met Gln Arg Val Phe Arg Gly Lys Leu
35 40 45

Leu Ser Asn Asp Glu Val Thr Ile Lys Tyr Lys Asp Glu Asp Gly Asp
50 55 60

Leu Ile Thr Ile Phe Asp Ser Ser Asp Leu Ser Phe Ala Ile Gln Cys
65 70 75 80

Ser Arg Ile Leu Lys Leu Thr Leu Phe Val Asn Gly Gln Pro Arg Pro
85 90 95

Leu Glu Ser Ser Gln Val Lys Tyr Leu Arg Arg Glu Leu Ile Glu Leu
100 105 110

P042876Seq. TXT

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

P042876Seq. TXT

G y I l e P h e T h r S e r L y s T h r A s p T h r T r p S e r P h e G y V a l L e u L e u
 385 390 395 400
 T r p G u I l e P h e S e r L e u G y T y r M e t P r o T y r P r o S e r L y s S e r A s n
 405 410 415
 G n G u V a l L e u G u P h e V a l T h r S e r G y G y A r g M e t A s p P r o P r o
 420 425 430
 L y s A s n C y s P r o G y P r o V a l T y r A r g I l e M e t T h r G n C y s T r p G n
 435 440 445
 H i s G n P r o G u A s p A r g P r o A s n P h e A l a I l e I l e L e u G u A r g I l e
 450 455 460
 G u T y r C y s T h r G n A s p P r o A s p V a l I l e A s n T h r A l a L e u P r o I l e
 465 470 475 480
 G u T y r G y P r o L e u V a l G u G u G u G u L y s V a l P r o V a l A r g P r o
 485 490 495
 L y s A s p P r o G u G y V a l P r o P r o L e u L e u V a l S e r G n G n A l a L y s
 500 505 510
 A r g G u G u G u A r g S e r P r o A l a A l a P r o P r o P r o L e u P r o T h r T h r
 515 520 525
 S e r S e r G y L y s A l a A l a L y s L y s P r o T h r A l a A l a G u V a l S e r V a l
 530 535 540
 A r g V a l P r o A r g G y P r o A l a V a l G u G y G y H i s V a l A s n M e t A l a
 545 550 555 560
 P h e S e r G n S e r A s n P r o P r o S e r G u L e u H i s L y s V a l H i s G y S e r
 565 570 575
 A r g A s n L y s P r o T h r S e r L e u T r p A s n P r o T h r T y r G y S e r T r p P h e
 580 585 590
 T h r G u L y s P r o T h r L y s L y s A s n A s n P r o I l e A l a L y s L y s G u P r o
 595 600 605
 H i s A s p A r g G y A s n L e u G y L e u G u G y S e r C y s T h r V a l P r o P r o
 610 615 620
 A s n V a l A l a T h r G y A r g L e u P r o G y A l a S e r L e u L e u L e u G u P r o
 625 630 635 640
 S e r S e r L e u T h r A l a A s n M e t L y s G u V a l P r o L e u P h e A r g L e u A r g
 645 650 655

P042876Seq. TXT

His Phe Pro Cys Gly Asn Val Asn Tyr Gly Tyr Gln Gln Gln Gly Leu
660 665 670

Pro Leu Glu Ala Ala Thr Ala Pro Gly Ala Gly His Tyr Glu Asp Thr
675 680 685

Ile Leu Lys Ser Lys Asn Ser Met Asn Gln Pro Gly Pro
690 695 700

<210> 31
<211> 1607
<212> DNA
<213> Homo sapiens

<400> 31
aact ccgct g cct t t gccgc caccat gccc aaaacgat ca gt gt gcgt gt gaccacat g 60
gat gcagagc t ggagt t t gc cat ccagccc aacaccaccg ggaagcagct at t t gaccag 120
gt ggt gaaaa ct at t ggct t gaggggaagt t t ggt t ct t t g gt ct gcagt a ccaggacact 180
aaaggt t t ct ccacct ggct gaaact caat aagaaggt ga ct gcccagga t gt gcggaag 240
gaaagccccc t gct ct t t aa gt t ccgt gcc aagt t ct acc ct gaggat gt gt ccgaggaa 300
t t gat t cagg acat cact ca gcgcct gt t c t t t ct gcaag t gaaagaggg cat t ct caat 360
gat gat at t t act gcccgcc t gagaccgt gt gct gct gg cct cgt at gc t gt ccagt ct 420
aagt at ggcg act t caat aa ggaagt gcat aagt ct ggct acct ggccgg agacaagt t g 480
ct cccgcaga gagt cct gga acagcacaaa ct caacaagg accagt ggga ggagcggat c 540
caggt gt ggc at gaggaaca ccgt ggcat g ct cagggagg at gct gt cct ggaat at ct g 600
aagat t gct c aagat ct gga gat gt at ggt gt gaact act t cagcat caa gaacaagaaa 660
ggct cagagc t gt ggct ggg ggt ggat gcc ct ggggt ct ca acat ct at ga gcagaat gac 720
agact aact c ccaagat agg ct t cccct gg agt gaaat ca ggaacat ct c t t t caat gat 780
aagaaat t t g t cat caagcc cat t gacaaa aaagccccgg act t cgt ct t ct at gct ccc 840
cggct gcgga t t aacaagcg gat ct t ggcc t t gt gcat gg ggaacct ga act at acat g 900
cgccgt cgca agcct gat ac cat t gaggt g cagcagat ga aggcacaggc ccgggaggag 960
aagcaccaga agcagat gga gcgt gct at g ct ggaaaat g agaagaagaa gcgt gaaat g 1020
gcagagaagg agaaagagaa gat t gaacgg gagaaggagg agct gat gga gaggct gaag 1080
cagat cgagg aacagact aa gaaggct cag caagaact gg aagaacagac ccgt agggct 1140
ct ggaact t g agcaggaacg gaagcgt gcc cagagcgagg ct gaaaagct ggccaaggag 1200
cgt caagaag ct gaagaggc caaggaggcc t t gct gcagg cct cccggga ccagaaaaag 1260
act caggaac agct ggcct t ggaaat ggca gagct gacag ct cgaat ct c ccagct ggag 1320
at ggcccgcac agaagaagga gagt gaggct gt ggagt ggc agcagaagca ggagct gcaa 1380
gcat gcaga t ggagct gca gagccct gag t acaagct ga gcaagct ccg cacct cgacc 1440
at cat gaccg act acaaccc caact act gc t t t gct ggca agacct cct c cat cagt gac 1500

P042876Seq. TXT

ct gaaggagg t gccgcggaa aaacat cacc ct cat t cggg gt ct gggcca t ggcgcct t t 1560
ggggaggt gt at gaaggcca ggt gt ccgga at gcccaacg acccaag 1607

<210> 32
<211> 527
<212> PRT
<213> Homo sapiens

<400> 32
Met Pro Lys Thr Ile Ser Val Arg Val Thr Thr Met Asp Ala Glu Leu
1 5 10 15

Glu Phe Ala Ile Gln Pro Asn Thr Thr Gly Lys Gln Leu Phe Asp Gln
20 25 30

Val Val Lys Thr Ile Gly Leu Arg Glu Val Trp Phe Phe Gly Leu Gln
35 40 45

Tyr Gln Asp Thr Lys Gly Phe Ser Thr Trp Leu Lys Leu Asn Lys Lys
50 55 60

Val Thr Ala Gln Asp Val Arg Lys Glu Ser Pro Leu Leu Phe Lys Phe
65 70 75 80

Arg Ala Lys Phe Tyr Pro Glu Asp Val Ser Glu Glu Leu Ile Gln Asp
85 90 95

Ile Thr Gln Arg Leu Phe Phe Leu Gln Val Lys Glu Gly Ile Leu Asn
100 105 110

Asp Asp Ile Tyr Cys Pro Pro Glu Thr Ala Val Leu Leu Ala Ser Tyr
115 120 125

Ala Val Gln Ser Lys Tyr Gly Asp Phe Asn Lys Glu Val His Lys Ser
130 135 140

Gly Tyr Leu Ala Gly Asp Lys Leu Leu Pro Gln Arg Val Leu Glu Gln
145 150 155 160

His Lys Leu Asn Lys Asp Gln Trp Glu Glu Arg Ile Gln Val Trp His
165 170 175

Glu Glu His Arg Gly Met Leu Arg Glu Asp Ala Val Leu Glu Tyr Leu
180 185 190

Lys Ile Ala Gln Asp Leu Glu Met Tyr Gly Val Asn Tyr Phe Ser Ile
195 200 205

Lys Asn Lys Lys Gly Ser Glu Leu Trp Leu Gly Val Asp Ala Leu Gly
210 215 220

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

Pro Arg Lys Asn Ile Thr Leu Ile Arg Gly Leu Gly His Gly Ala Phe
 500 505 510

Gly Gu Val Tyr Gu Gly Gn Val Ser Gly Met Pro Asn Asp Pro
 515 520 525

<210> 33
 <211> 426
 <212> DNA
 <213> Homo sapiens

<400> 33
 cgagaagt t g agggagaaag gcgggcccgg gaacaggct g aggct gaggt ggcct cct t g 60
 aaccgt agga t ccagct ggt t gaagaagag ct ggaccgt g ct caggagcg t gcggaggt g 120
 t ct gaact aa aat gt ggt ga cct ggaagaa gaact caaga at gt t act aa caat ct gaaa 180
 t ct ct ggagg ct gcat ct ga aaagt at t ct gaaaaggagg acaa at ga agaagaaat t 240
 aaact t ct gt ct gacaaact gaaagaggct gagaccgt g ct gaat t t gc agagagaacg 300
 gt t gcaaaac t ggaaaagac aat t gat gac ct ggaagt gt acct ccggaa gcaccaagag 360
 ct gcaagcca t gcagat gga gct gcagagc cct gagt aca agct gagcaa gct ccgcacc 420
 ct cgac 426

<210> 34
 <211> 142
 <212> PRT
 <213> Homo sapiens

<400> 34
 Arg Gu Val Gu Gly Gu Arg Arg Ala Arg Gu Gn Ala Gu Ala Gu
 1 5 10 15

Val Ala Ser Leu Asn Arg Arg Ile Gn Leu Val Gu Gu Gu Leu Asp
 20 25 30

Arg Ala Gn Gu Arg Ala Gu Val Ser Gu Leu Lys Cys Gly Asp Leu
 35 40 45

Gu Gu Gu Leu Lys Asn Val Thr Asn Asn Leu Lys Ser Leu Gu Ala
 50 55 60

Ala Ser Gu Lys Tyr Ser Gu Lys Gu Asp Lys Tyr Gu Gu Gu Ile
 65 70 75 80

Lys Leu Leu Ser Asp Lys Leu Lys Gu Ala Gu Thr Arg Ala Gu Phe
 85 90 95

Ala Gu Arg Thr Val Ala Lys Leu Gu Lys Thr Ile Asp Asp Leu Gu
 100 105 110

Val Tyr Leu Arg Lys His Gn Gu Leu Gn Ala Met Gn Met Gu Leu
 115 120 125

P042876Seq. TXT

Gln Ser Pro Glu Tyr Lys Leu Ser Lys Leu Arg Thr Leu Asp
130 135 140

<210> 35
<211> 308
<212> DNA
<213> Homo sapiens

<400> 35
ctggcagagt cccgttgccg agagatggat gaggcagatt a gactgatgga ccagaacctg 60
aagtgtctga gtgctgctga agaaaagtac tctcaaaaag aagataaat a t gaggagaaga 120
atcaagattc ttactgat aa actcaaggag gcagagaccc gtgctgaatt tgcagagaga 180
acggttgcaa aactggaaaa gacaattgat gacctggaag tgtaccgccg gaagcaccag 240
gagctgcaag ccatgcagat ggagctgcag agccctgagt acaagctgag caagctccgc 300
acctcgac 308

<210> 36
<211> 103
<212> PRT
<213> Homo sapiens

<400> 36
Leu Ala Glu Ser Arg Cys Arg Glu Met Asp Glu Gln Ile Arg Leu Met
1 5 10 15

Asp Gln Asn Leu Lys Cys Leu Ser Ala Ala Glu Glu Lys Tyr Ser Gln
20 25 30

Lys Glu Asp Lys Tyr Glu Glu Glu Ile Lys Ile Leu Thr Asp Lys Leu
35 40 45

Lys Glu Ala Glu Thr Arg Ala Glu Phe Ala Glu Arg Thr Val Ala Lys
50 55 60

Leu Glu Lys Thr Ile Asp Asp Leu Glu Val Tyr Arg Arg Lys His Gln
65 70 75 80

Glu Leu Gln Ala Met Gln Met Glu Leu Gln Ser Pro Glu Tyr Lys Leu
85 90 95

Ser Lys Leu Arg Thr Ser Thr
100

<210> 37
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic primer

<400> 37

tacaacccca actactgctt tgct P042876Seq. TXT 24

<210> 38
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 38
aggcactttc tcttcctctt ccac 24

<210> 39
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 39
ccacacctgg gaaaggacct aaag 24

<210> 40
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
primer

<400> 40
cctccaaat a ctgacagcca cagg 24