

(19) **DANMARK**

(10) **DK/EP 2423331 T3**



Patent- og
Varemærkestyrelsen

(12) Oversættelse af
europæisk patentskrift

-
- (51) Int.Cl.: **C 07 H 21/04 (2006.01)**
- (45) Oversættelsen bekendtgjort den: **2019-07-29**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2019-05-08**
- (86) Europæisk ansøgning nr.: **11178200.9**
- (86) Europæisk indleveringsdag: **2005-03-31**
- (87) Den europæiske ansøgnings publiceringsdag: **2012-02-29**
- (30) Prioritet: **2004-03-31 US 558218 P** **2004-04-09 US 561095 P**
2004-04-27 US 565753 P **2004-04-27 US 565985 P**
2004-05-25 US 574035 P **2004-06-07 US 577916 P**
2004-07-29 US 592287 P
- (62) Stamansøgningsnr: **05757467.5**
- (84) Designerede stater: **AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU MC NL PL PT RO SE SI SK TR**
- (73) Patenthaver: **The General Hospital Corporation, 55 Fruit Street, Boston, MA 02114, USA**
Dana-Farber Cancer Institute, Inc., 450 Brookline Avenue, Boston, MA 02215, USA
- (72) Opfinder: **Bell, Daphne Winifred, 4450 South Park Avenue, Apt. 1706, Chevy Chase, MD 20515, USA**
Haber, Daniel A., 34 Monadnock Road, Chestnut Hill, MA 02467, USA
Janne, Pasi Antero, 128 Roundwood Road, Newton, MA 02464, USA
Johnson, Bruce E., 80 Seaver Street, Brookline, MA 02445, USA
MEYERSON, Matthew, 307 Independence Road, Concord, MA 01742, USA
Paez, Juan Guillermo, 125 Pembroke Street, Apt. 2, Boston, MA 02118, USA
Sellers, William R., 80 Seaver Street, Brookline, MA 02445, USA
SORDELLA, Raffaella, 1640 Moors Hill Road, Syosset, NY 11791, USA
Lynch, Thomas J., 21 Highland Street, New Haven, CT 06511, USA
Settleman, Jeffrey E., 355 Magee Avenue, Mill Valley, CA 94941, USA
- (74) Fuldmægtig i Danmark: **Chas. Hude A/S, H.C. Andersens Boulevard 33, 1780 København V, Danmark**
- (54) Benævnelse: **FREMGANGSMÅDE TIL AT BESTEMME REAKTIVITET AF CANCER PÅ BEHANDLINGER MÅLRETTET MOD EPIDERMAL VÆKSTFAKTOR-RECEPTOR**
- (56) Fremdragne publikationer:
WO-A2-02/102976
HAN YUCHUN ET AL: "Tyrophostin AG 1478 preferentially inhibits human glioma cells expressing truncated rather than wild-type epidermal growth factor receptors", CANCER RESEARCH, vol. 56, no. 17, 1996, pages 3859-3861, XP002533062, ISSN: 0008-5472
HIRSCH F R ET AL: "Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis", JOURNAL OF CLINICAL ONCOLOGY, AMERICAN SOCIETY OF CLINICAL ONCOLOGY, US, vol. 21, no. 20, 15 October 2003 (2003-10-15), pages 3798-

Fortsættes ...

3807, XP002397277, ISSN: 0732-183X, DOI: 10.1200/JCO.2003.11.069

DANCEY JANET E: "Predictive factors for epidermal growth factor receptor inhibitors - The bull's-eye hits the arrow", CANCER CELL, vol. 5, no. 5, May 2004 (2004-05), pages 411-415, XP002533061, ISSN: 1535-6108

GUILLERMO PAEZ J ET AL: "EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy", SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, WASHINGTON, DC; US, vol. 304, 4 June 2004 (2004-06-04), pages 1497-1500, XP008136813, ISSN: 0036-8075, DOI: 10.1126/SCIENCE.1099314

LYNCH T J ET AL: "Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib", NEW ENGLAND JOURNAL OF MEDICINE, MASSACHUSETTS MEDICAL SOCIETY, BOSTON, MA, US, vol. 350, no. 21, 20 May 2004 (2004-05-20) , pages 2129-2139, XP002359960, ISSN: 1533-4406

FREDERICK LORI ET AL: "Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas", CANCER RESEARCH, vol. 60, no. 5, 1 March 2000 (2000-03-01), pages 1383-1387, XP002533730, ISSN: 0008-5472

HAN YUCHUN ET AL: "Preferential inhibition of glioblastoma cells with wild-type epidermal growth factor receptors by a novel tyrosine kinase inhibitor ethyl-2,5-dihydroxycinnamate", ONCOLOGY RESEARCH, vol. 9, no. 11-12, 1997, pages 581-587, XP002929150, ISSN: 0965-0407

PEDERSEN MIKKEL W ET AL: "Expression of a naturally occurring constitutively active variant of the epidermal growth factor receptor in mouse fibroblasts increases motility.", INTERNATIONAL JOURNAL OF CANCER, vol. 108, no. 5, 20 February 2004 (2004-02-20), pages 643-653, XP002533063, ISSN: 0020-7136

Description

BACKGROUND

[0001] Epithelial cell cancers, for example, prostate cancer, breast cancer, colon cancer, lung cancer, pancreatic cancer, ovarian cancer, cancer of the spleen, testicular cancer, cancer of the thymus, etc., are diseases characterized by abnormal, accelerated growth of epithelial cells. This accelerated growth initially causes a tumor to form. Eventually, metastasis to different organ sites can also occur. Although progress has been made in the diagnosis and treatment of various cancers, these diseases still result in significant mortality.

[0002] Lung cancer remains the leading cause of cancer death in industrialized countries. Cancers that begin in the lungs are divided into two major types, non-small cell lung cancer and small cell lung cancer, depending on how the cells appear under a microscope. Non-small cell lung cancer (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) generally spreads to other organs more slowly than does small cell lung cancer. About 75 percent of lung cancer cases are categorized as non-small cell lung cancer (e.g., adenocarcinomas), and the other 25 percent are small cell lung cancer. Non-small cell lung cancer (NSCLC) is the leading cause of cancer deaths in the United States, Japan and Western Europe. For patients with advanced disease, chemotherapy provides a modest benefit in survival, but at the cost of significant toxicity, underscoring the need for therapeutic agents that are specifically targeted to the critical genetic lesions that direct tumor growth (Schiller JH et al., *N Engl J Med*, 346: 92-98, 2002).

[0003] Epidermal growth factor receptor (EGFR) is a 170 kilodalton (kDa) membrane-bound protein expressed on the surface of epithelial cells. EGFR is a member of the growth factor receptor family of protein tyrosine kinases, a class of cell cycle regulatory molecules. (W. J. Gullick et al., 1986, *Cancer Res.*, 46:285-292). EGFR is activated when its ligand (either EGF or TGF- α) binds to the extracellular domain, resulting in autophosphorylation of the receptor's intracellular tyrosine kinase domain (S. Cohen et al., 1980, *J. Biol. Chem.*, 255:4834-4842; A. B. Schreiber et al., 1983, *J. Biol. Chem.*, 258:846-853).

[0004] EGFR is the protein product of a growth promoting oncogene, *erbB* or *ErbB1*, that is but one member of a family, i.e., the ERBB family of protooncogenes, believed to play pivotal roles in the development and progression of many human cancers. In particular, increased expression of EGFR has been observed in breast, bladder, lung, head, neck and stomach cancer as well as glioblastomas. The ERBB family of oncogenes encodes four, structurally-related transmembrane receptors, namely, EGFR, HER-2/neu (*erbB2*), HER-3 (*erbB3*) and HER-4 (*erbB4*). Clinically, ERBB oncogene amplification and/or receptor overexpression in tumors have been reported to correlate with disease recurrence and poor patient prognosis, as well as with responsiveness in therapy. (L. Harris et al., 1999, *Int. J. Biol. Markers*, 14:8-15; and J. Mendelsohn and J. Baselga, 2000, *Oncogene*, 19:6550-6565).

[0005] EGFR is composed of three principal domains, namely, the extracellular domain (ECD), which is glycosylated and contains the ligand-binding pocket with two cysteine-rich regions; a short transmembrane domain, and an intracellular domain that has intrinsic tyrosine kinase activity. The transmembrane region joins the ligand-binding domain to the intracellular domain. Amino acid and DNA sequence analysis, as well as studies of nonglycosylated forms of EGFR, indicate that the protein backbone of EGFR has a mass of 132 kDa, with 1186 amino acid residues (A. L. Ullrich et al., 1984, *Nature*, 307:418-425; J. Downward et al., 1984, *Nature*, 307:521-527; C. R. Carlin et al., 1986, *Mol. Cell. Biol.*, 6:257-264; and F. L. V. Mayes and M. D. Waterfield, 1984, *The EMBO J.*, 3:531-537).

[0006] The binding of EGF or TGF- α to EGFR activates a signal transduction pathway and results in cell proliferation. The dimerization, conformational changes and internalization of EGFR molecules function to transmit intracellular signals leading to cell growth regulation (G. Carpenter and S. Cohen, 1979, *Ann. Rev. Biochem.*, 48:193-216). Genetic alterations that affect the regulation of growth factor receptor function, or lead to overexpression of receptor and/or ligand, result in cell proliferation. In addition, EGFR has been determined to play a role in cell differentiation, enhancement of cell motility, protein secretion, neovascularization, invasion, metastasis and resistance of cancer cells to chemotherapeutic agents and radiation. (M.-J. Oh et al., 2000, *Clin. Cancer Res.*, 6:4760-4763).

[0007] A variety of inhibitors of EGFR have been identified, including a number already undergoing clinical trials for treatment of various cancers. For a recent summary, see de Bono, J. S. and Rowinsky, E. K. (2002), "The ErbB Receptor Family: A Therapeutic Target For Cancer", *Trends in Molecular Medicine*, 8, S19-26.

[0008] A promising set of targets for therapeutic intervention in the treatment of cancer includes the members of the HER-kinase axis. They are frequently upregulated in solid epithelial tumors of, by way of example, the prostate, lung and breast, and are also upregulated in glioblastoma tumors. Epidermal growth factor receptor (EGFR) is a member of the HER-kinase axis, and has been the target of choice for the development of several different cancer therapies. EGFR tyrosine kinase inhibitors (EGFR-TKIs) are among these therapies, since the reversible phosphorylation of tyrosine residues is required for activation of the EGFR pathway. In other words, EGFR-TKIs block a cell surface receptor responsible for triggering and/or maintaining the cell signaling pathway that induces tumor cell growth and division. Specifically, it is believed that these inhibitors interfere with the EGFR kinase domain, referred to as HER-1. Among the more promising EGFR-TKIs are three series of compounds: quinazolines, pyridopyrimidines and pyrrolopyrimidines.

[0009] Two of the more advanced compounds in clinical development include Gefitinib (compound ZD1839 developed by AstraZeneca UK Ltd.; available under the tradename IRESSA; hereinafter "IRESSA") and Erlotinib (compound OSI-774 developed by Genentech, Inc. and OSI Pharmaceuticals, Inc.; available under the tradename TARCEVA; hereinafter "TARCEVA"); both have generated encouraging clinical results. Conventional cancer treatment with both IRESSA and

IRESSA involves the daily, oral administration of no more than 500 mg of the respective compounds. In May, 2003, IRESSA became the first of these products to reach the United States market, when it was approved for the treatment of advanced non-small cell lung cancer patients.

[0010] IRESSA is an orally active quinazoline that functions by directly inhibiting tyrosine kinase phosphorylation on the EGFR molecule. It competes for the adenosine triphosphate (ATP) binding site, leading to suppression of the HER-kinase axis. The exact mechanism of the IRESSA response is not completely understood, however, studies suggest that the presence of EGFR is a necessary prerequisite for its action.

[0011] A significant limitation in using these compounds is that recipients thereof may develop a resistance to their therapeutic effects after they initially respond to therapy, or they may not respond to EGFR-TKIs to any measurable degree at all. In fact, only 10-15 percent of advanced non-small cell lung cancer patients respond to EGFR kinase inhibitors. Thus, a better understanding of the molecular mechanisms underlying sensitivity to IRESSA and TARCEVA would be extremely beneficial in targeting therapy to those individuals whom are most likely to benefit from such therapy.

[0012] There is a significant need in the art for a satisfactory treatment of cancer, and specifically epithelial cell cancers such as lung, ovarian, breast, brain, colon and prostate cancers, which incorporates the benefits of TKI therapy and overcoming the non-responsiveness exhibited by patients. Such a treatment could have a dramatic impact on the health of individuals, and especially older individuals, among whom cancer is especially common.

SUMMARY

[0013] Tyrosine kinase inhibitor (TKI) therapy such as gefitinib (IRESSA®) is not effective in the vast majority of individuals that are affected with the cancers noted above. The present inventors have surprisingly discovered that the presence of somatic mutations in the kinase domain of EGFR substantially increases sensitivity of the EGFR to TKI such as IRESSA, TARCEVA. For example less than 30% of patients having such cancer are susceptible to treatment by current TKIs, whereas greater than 50%, more preferably 60, 70, 80, 90 % of patients having a mutation in the EGFR kinase domain are susceptible. In addition, these mutations confer increased kinase activity of the EGFR. Thus, patients having these mutations will likely be responsive to current tyrosine kinase inhibitor (TKI) therapy, for example, gefitinib.

[0014] Accordingly, the present invention provides a novel method to determine the likelihood of effectiveness of an epidermal growth factor receptor (EGFR) targeting treatment in a human patient affected with cancer. The method comprises detecting the presence or absence of at least one nucleic acid variance in the kinase domain of the erbB1 gene of said patient relative to the wildtype erbB1 gene. The presence of at least one variance indicates that the EGFR targeting treatment is likely to be effective. Preferably, the nucleic acid variance increases the kinase activity of the EGFR. The patient can then be treated with an EGFR targeting treatment. In one embodiment of the present invention, the EGFR targeting treatment is a tyrosine kinase inhibitor. In a preferred embodiment, the tyrosine kinase inhibitor is an anilinoquinazoline. The anilinoquinazoline may be a synthetic anilinoquinazoline. Preferably, the synthetic anilinoquinazoline is either gefitinib or erlotinib. In another embodiment, the EGFR targeting treatment is an irreversible EGFR inhibitor, including 4-dimethylamino-but-2-enoic acid [4-(3-chloro-4-fluoro-phenylamino)-3-cyano-7-ethoxy-quinolin-6-yl]-amide ("EKB-569", sometimes also referred to as "EKI-569", see for example WO/2005/018677 and Torrance et al., Nature Medicine, vol. 6, No. 9, Sept. 2000, p. 1024) and/or HKI-272 or HKI-357 (Wyeth; see Greenberger et al., Proc. 11th NCI EORTC-AACR Symposium on New Drugs in Cancer Therapy, Clinical Cancer Res. Vol. 6 Supplement, Nov. 2000, ISSN 1078-0432; in Rabindran et al., Cancer Res. 64: 3958-3965 (2004); Holbro and Hynes, Ann. Rev. Pharm. Tox. 44:195-217 (2004); Tsou et al, J. Med. Chem. 2005, 48,1107-1131; and Tejpar et al., J. Clin. Oncol. ASCO Annual Meeting Proc. Vol. 22, No. 14S: 3579 (2004)).

[0015] In one embodiment of the present invention, the EGFR is obtained from a biological sample from a patient with or at risk for developing cancer. The variance in the kinase domain of EGFR (or the erbB1 gene) effects the conformational structure of the ATP-binding pocket. Preferably, the variance in the kinase domain of EGFR is an in frame deletion or a substitution in exon 18, 19, 20 or 21.

[0016] In one embodiment, the in frame deletion is in exon 19 of EGFR (erbB1). The in frame deletion in exon 19 preferably comprises at deletion of at least amino acids leucine, arginine, glutamic acid and alanine, at codons 747, 748, 749, and 750. In one embodiment, the in-frame deletion comprises nucleotides 2235 to 2249 and deletes amino acids 746 to 750 (the sequence glutamic acid, leucine, arginine, glutamic acid, and alanine), see Table 2, Table S2, Figure 2B, Figure 4A, Figure 5, SEQ ID NO: 511, Figure 6C, and Figure 8C. In another embodiment, the in-frame deletion comprises nucleotides 2236 to 2250 and deletes amino acids 746 to 750, see Table S2, Figure 5, SEQ ID NO: 511, and Figure 6C. Alternatively, the in-frame deletion comprises nucleotides 2240 to 2251, see Table 2, Figure 2C, Figure 4A, Figure 5, SEQ ID NO: 511, or nucleotides 2240 to 2257, see Table 2, Table S3A, Figure 2A, Figure 4A, Figure 5, SEQ

ID NO: 511, Figure 6C, and Figure 8E. Alternatively, the in-frame deletion comprises nucleotides 2239 to 2247 together with a substitution of cytosine for guanine at nucleotide 2248, see Table S3A and Figure 8D, or a deletion of nucleotides 2238 to 2255 together with a substitution of thymine for adenine at nucleotide 2237, see Table S3A and Figure 8F, or a deletion of nucleotides 2254 to 2277, see Table S2. Alternatively, the in-frame deletion comprises nucleotides 2239-2250delTTAAGAGAAGCA; 2251A>C, or 2240-2254delTTAAGAGAAGCA, or 2257-2271delCCGAAAGCCAACAAG, as shown in Table S3B.

[0017] In another embodiment, the substitution is in exon 21 of EGFR. The substitution in exon 21 comprises at least one amino acid. In one embodiment, the substitution in exon 21 comprises a substitution of a guanine for a thymine at nucleotide 2573, see Figure 4A and Figure 5, SEQ ID NO: 511. This substitution results in an amino acid substitution, where the wildtype Leucine is replaced with an Arginine at amino acid 858, see Figure 5, Table 2, Table S2, Table S3A, Figure 2D, Figure 6A, figure 8B, and SEQ ID NO: 512. Alternatively, the substitution in exon 21 comprises a substitution of an adenine for a thymine at nucleotide 2582, see Figure 4A and Figure 5, SEQ ID NO: 511. This substitution results in an amino acid substitution, where the wildtype Leucine is replaced with a Glutamine at amino acid 861, see Figure 5, Table 2, Figure 2E, Table S3B, and SEQ ID NO: 512.

[0018] The substitution may also be in exon 18 of EGFR. In one embodiment, the substitution is in exon 18 is a thymine for a guanine at nucleotide 2155, see Figure 4A and Figure 5, SEQ ID NO: 511. This substitution results in an amino acid substitution, where the wildtype Glycine is substituted with a Cysteine at codon 719, see Figure 5, SEQ ID NO: 512. In another embodiment, the substitution in exon 18 is an adenine for a guanine at nucleotide 2155 resulting in an amino acid substitution, where the wildtype Glycine is substituted for a Serine at codon 719, see Table S2, Figure 6B, Figure 8A, Figure 5, SEQ ID NO: 511 and 512.

[0019] In another embodiment, the substitution is an insertion of guanine, guanine and thymine (GGT) after nucleotide 2316 and before nucleotide 2317 of SEQ ID NO: 511 (2316_2317 ins GGT). This can also be described as an insertion of valine (V) at amino acid 772 (P772_H733 insV). Other mutations are shown in Table S3B and include, for example, and insertion of CAACCCGG after nucleotide 2309 and before nucleotide 2310 of SEQ ID NO 511 and an insertion of GCGTGGACA after nucleotide 2311 and before nucleotide 2312 of SEQ ID NO 511. The substitution may also be in exon 20 and in one embodiment is a substitution of AA for GG at nucleotides 2334 and 2335, see Table S3B.

[0020] In summary, in preferred embodiments, the nucleic acid variance of the erbB1 gene is a substitution of a thymine for a guanine or an adenine for a guanine at nucleotide 2155 of SEQ ID NO 511, a deletion of nucleotides 2235 to 2249, 2240 to 2251, 2240 to 2257, 2236 to 2250, 2254 to 2277, or 2236 to 2244 of SEQ ID NO 511, an insertion of nucleotides guanine, guanine, and thymine (GGT) after nucleotide 2316 and before nucleotide 2317 of SEQ ID NO 511, and a substitution of a guanine for a thymine at nucleotide 2573 or an adenine for a thymine at nucleotide 2582 of SEQ ID NO 511.

[0021] The detection of the presence or absence of at least one nucleic acid variance can be determined by amplifying a segment of nucleic acid encoding the receptor. The segment to be amplified is 1000 nucleotides in length, preferably, 500 nucleotides in length, and most preferably 100 nucleotides in length or less. The segment to be amplified can include a plurality of variances.

[0022] In another embodiment, the detection of the presence or absence of at least one variance provides for contacting EGFR nucleic acid containing a variance site with at least one nucleic acid probe. The probe preferentially hybridizes with a nucleic acid sequence including a variance site and containing complementary nucleotide bases at the variance site under selective hybridization conditions. Hybridization can be detected with a detectable label.

[0023] In yet another embodiment, the detection of the presence or absence of at least one variance comprises sequencing at least one nucleic acid sequence and comparing the obtained sequence with the known erbB1 nucleic acid sequence. Alternatively, the presence or absence of at least one variance comprises mass spectrometric determination of at least one nucleic acid sequence.

[0024] In a preferred embodiment, the detection of the presence or absence of at least one nucleic acid variance comprises performing a polymerase chain reaction (PCR). The erbB1 nucleic acid sequence containing the hypothetical variance is amplified and the nucleotide sequence of the amplified nucleic acid is determined. Determining the nucleotide sequence of the amplified nucleic acid comprises sequencing at least one nucleic acid segment. Alternatively, amplification products can be analyzed by using any method capable of separating the amplification products according to their size, including automated and manual gel electrophoresis and the like.

[0025] Alternatively, the detection of the presence or absence of at least one variance comprises determining the haplotype of a plurality of variances in a gene.

[0026] In another embodiment, the presence or absence of an EGFR variance can be detected by analyzing the erbB1 gene product (protein). In this embodiment, a probe that specifically binds to a variant EGFR is utilized. In a preferred embodiment, the probe is an antibody that preferentially binds to a variant EGFR. The presence of a variant EGFR predicts the likelihood of effectiveness of an EGFR targeting treatment. Alternatively, the probe may be an antibody fragment, chimeric antibody, humanized antibody or an aptamer.

[0027] The present invention further provides a probe which specifically binds under selective binding conditions to a nucleic acid sequence comprising at least one nucleic acid variance in the EGFR gene (erbB1). In one embodiment,

the variance is a mutation in the kinase domain of erbB1 that confers a structural change in the ATP-binding pocket.

[0028] The probe of the present invention may comprise a nucleic acid sequence of about 500 nucleotide bases, preferably about 100 nucleotides bases, and most preferably about 50 or about 25 nucleotide bases or fewer in length. The probe may be composed of DNA, RNA, or peptide nucleic acid (PNA). Furthermore, the probe may contain a detectable label, such as, for example, a fluorescent or enzymatic label.

[0029] The present invention additionally provides a novel method to determine the likelihood of effectiveness of an epidermal growth factor receptor (EGFR) targeting treatment in a patient affected with cancer. The method comprises determining the kinase activity of the EGFR in a biological sample from a patient. An increase in kinase activity following stimulation with an EGFR ligand, compared to a normal control, indicates that the EGFR targeting treatment is likely to be effective.

[0030] The present invention further provides a novel method for treating a patient affected with or at risk for developing cancer. The method involves determining whether the kinase domain of the EGFR of a patient contains at least one nucleic acid variance. Preferably, the EGFR is located at the site of the tumor or cancer and the nucleic acid variance is somatic. The presence of such a variance indicates that an EGFR targeted treatment will be effective. If the variance is present, the tyrosine kinase inhibitor is administered to the patient.

[0031] As above, the tyrosine kinase inhibitor administered to an identified patient may be an anilinoquinazoline or an irreversible tyrosine kinase inhibitor, such as for example, EKB-569, HKI-272 and/or HKI-357 (Wyeth). Preferably, the anilinoquinazoline is a synthetic anilinoquinazoline and most preferably the synthetic anilinoquinazoline is gefitinib and erlotinib.

[0032] The cancer to be treated by the methods of the present invention include, for example, but are not limited to, gastrointestinal cancer, prostate cancer, ovarian cancer, breast cancer, head and neck cancer, lung cancer, non-small cell lung cancer, cancer of the nervous system, kidney cancer, retina cancer, skin cancer, liver cancer, pancreatic cancer, genital-urinary cancer and bladder cancer. In a preferred embodiment, the cancer is non-small cell lung cancer.

[0033] A kit for implementing the PCR methods of the present invention is also encompassed. The kit includes at least one degenerate primer pair designed to anneal to nucleic acid regions bordering the genes that encode for the ATP-binding pocket of the EGFR kinase domain. Additionally, the kit contains the products and reagents required to carry out PCR amplification, and instructions.

[0034] In a preferred embodiment, the primer pairs contained within the kit are selected from the group consisting of SEQ ID NO: 505, SEQ ID NO: 506, SEQ ID NO: 507, and SEQ ID NO: 508. Also preferred are the primers listed in Table 6 and 7 in the examples.

[0035] In yet another embodiment, the present invention discloses a method for selecting a compound that inhibits the catalytic kinase activity of a variant epidermal growth factor receptor (EGFR). As a first step, a variant EGFR is contacted with a potential compound. The resultant kinase activity of the variant EGFR is then detected and a compound is selected that inhibits the kinase activity of the variant EGFR. In one embodiment, the variant EGFR is contained within a cell. The method can also be used to select a compound that inhibits the kinase activity of a variant EGFR having a secondary mutation in the kinase domain that confers resistance to a TKI, e.g., gefitinib or erlotinib.

[0036] In one embodiment, the variant EGFR is labeled. In another embodiment, the EGFR is bound to a solid support. In a preferred embodiment, the solid support is a protein chip.

[0037] In yet another embodiment of the present invention, a pharmaceutical composition that inhibits the catalytic kinase activity of a variant epidermal growth factor receptor (EGFR) is disclosed. The compound that inhibits the catalytic kinase activity of a variant EGFR is selected from the group consisting of an antibody, antibody fragment, small molecule, peptide, protein, antisense nucleic acid, ribozyme, PNA, siRNA, oligonucleotide aptamer, and peptide aptamer.

[0038] A method for treating a patient having an EGFR mediated disease is also disclosed. In accordance with the method, the patient is administered the pharmaceutical composition that inhibits the catalytic kinase activity of a variant epidermal growth factor receptor (EGFR).

[0039] In one embodiment, the EGFR mediated disease is cancer. In a preferred embodiment, the cancer is of epithelial origin. For example, the cancer is gastrointestinal cancer, prostate cancer, ovarian cancer, breast cancer, head and neck cancer, lung cancer, non-small cell lung cancer, cancer of the nervous system, kidney cancer, retina cancer, skin cancer, liver cancer, pancreatic cancer, genital-urinary cancer and bladder cancer. In a preferred embodiment, the cancer is non-small cell lung cancer.

[0040] In another embodiment, a method for predicting the acquisition of secondary mutations (or selecting for mutations) in the kinase domain of the erbB1 gene is disclosed. A cell expressing a variant form of the erbB1 gene is contacted with an effective, yet sub-lethal dose of a tyrosine kinase inhibitor. Cells that are resistant to a growth arrest effect of the tyrosine kinase inhibitor are selected and the erbB1 nucleic acid is analyzed for the presence of additional mutations in the erbB1 kinase domain. In one embodiment, the cell is *in vitro*. In another embodiment, the cell is obtained from a transgenic animal. In one embodiment, the transgenic animal is a mouse. In this mouse model, cells to be studied are obtained from a tumor biopsy. Cells containing a secondary mutation in the erbB1 kinase domain selected by the present invention can be used in the above methods to select a compound that inhibits the kinase activity of the variant EGFR

having a secondary mutation in the kinase domain.

[0041] In an alternative embodiment for predicting the acquisition of secondary mutations in the kinase domain of the *erbB1* gene, cells expressing a variant form of the *erbB1* gene are first contacted with an effective amount of a mutagenizing agent. The mutagenizing is, for example, ethyl methanesulfonate (EMS), N-ethyl-N-nitrosourea (ENU), N-methyl-N-nitrosourea (MNU), phorbacarbazine hydrochloride (Prc), methyl methanesulfonate (MeMS), chlorambucil (Chl), melphalan, porcarbazine hydrochloride, cyclophosphamide (Cp), diethyl sulfate - (Et_2SO_4), acrylamide monomer (AA), triethylene melamin (TEM), nitrogen mustard, vincristine, dimethylnitrosamine, N-methyl-N'-nitro-Nitrosoguanidine (MNNG), 7,12 dimethylbenz(a)anthracene (DMBA), ethylene oxide, hexamethylphosphoramide, bisulfan, or ethyl methanesulfonate (EtMs). The cell is then contacted with an effective, yet sub-lethal dose of a tyrosine kinase inhibitor. Cells that are resistant to a growth arrest effect of the tyrosine kinase inhibitor are selected and the *erbB1* nucleic acid is analyzed for the presence of additional mutations in the *erbB1* kinase domain.

BRIEF DESCRIPTION OF THE FIGURES

[0042]

Figures 1A-1B show a representative illustration of Gefitinib response in refractory non-small cell lung cancer (NSCLC). Chest CT scan of case 6 (Table 1), demonstrating (Figure 1A) a large mass in the right lung before treatment with gefitinib, and (Figure 1B) marked improvement six weeks after Gefitinib was initiated.

Figure 2 shows EGFR mutations in Gefitinib-responsive tumors.

Figures 2A -2C show nucleotide sequence of the *EGFR* gene in tumor specimens with heterozygous in-frame deletions within the kinase domain (double peaks) (SEQ ID NOS 643-654, respectively, in order of appearance). Tracings in both sense and antisense directions are shown to demonstrate the two breakpoints of the deletion; wild-type nucleotide sequence is shown in capital letters, and the mutant sequence is in lowercase letters. The 5' breakpoint of the delL747-T751insS mutation is preceded by a T to C substitution that does not alter the encoded amino acid. Figure 2D and Figure 2E show heterozygous missense mutations (arrows) resulting in amino acid substitutions within the tyrosine kinase domain (SEQ ID NOS 656 & 658). The double peaks represent two nucleotides at the site of heterozygous mutations. For comparison, the corresponding wild-type sequence is also shown (SEQ ID NOS 655 & 657).

Figure 2F is a schematic representation of dimerized EGFR molecules bound by the EGF ligand. The extracellular domain (containing two receptor ligand [L]-domains and a furin-lilce domain), transmembrane region, and the cytoplasmic domain (containing the catalytic kinase domain) are highlighted. The position of tyrosine¹⁰⁶⁸ (Y-1068), a site of autophosphorylation used as a marker of receptor activation, is indicated, along with downstream effectors activated by EGFR autophosphorylation (STAT3, MAP Kinase (MAPK), and AKT). The location of tumor-associated mutations, all within the tyrosine kinase domain, is shown.

Figure 3 demonstrates enhanced EGF-dependent activation of mutant EGFR and increased sensitivity of mutant EGFR to Gefitinib.

Figure 3A shows a time course of ligand-induced activation of the delL747-P753insS and L858R mutants, compared with wild type EGFR, following addition of EGF to serum starved cells. EGFR autophosphorylation is used as a marker of receptor activation, using Western blotting with an antibody that specifically recognizes the phosphorylated tyrosine¹⁰⁶⁸ residue of EGFR (left panel), compared with the total levels of EGFR expressed in Cos-7 cells (control; right panel). Autophosphorylation of EGFR is measured at intervals following addition of EGF (10 ng/ml).

Figure 3B is a graphical representation of EGF-induction of wild-type and mutant receptor phosphorylation (see panel A). Autoradiographs from three independent experiments were quantified using the NIH image software; intensity of EGFR phosphorylation is normalized to total protein expression, and shown as percent activation of the receptor, with standard deviation.

Figure 3C shows a dose-dependent inhibition of EGFR activation by Gefitinib. Autophosphorylation of EGFR tyrosine¹⁰⁶⁸ is demonstrated by Western blotting analysis of Cos-7 cells expressing wild-type or mutant receptors, and stimulated with 100 ng/ml of EGF for 30 min. Cells were untreated (U) or pretreated for 3 hrs with increasing concentrations of Gefitinib as shown (left panel). Total amounts of EGFR protein expressed are shown as control (right panel).

Figure 3D shows the quantification of results from two experiments described for panel 3C (NIH image software). Concentrations of phosphorylated EGFR were normalized to protein expression levels and expressed as percent activation of the receptor.

Figure 4 demonstrates clustering of mutations at critical sites within the ATP- binding pocket of EGFR.

Figure 4A shows the position of overlapping in-frame deletions in exon 19 and missense mutations in exon 21 of the *EGFR* gene, in multiple cases of NSCLC (SEQ ID NOS 495-504 (DNA)). Partial nucleotide sequence is shown for each exon, with deletions marked by dashed lines and missense mutations highlighted and underlined; the wild-

type *EGFR* nucleotide and amino acid sequences are shown (SEQ ID NOS 493 & 494 (DNA) & 509-510 (amino acid)). Figure 4B shows the tridimensional structure of the EGFR ATP cleft flanked by the amino (N) and carboxy (C) lobes of the kinase domain (coordinates derived from PDB 1M14, and displayed using Cn3D software). The inhibitor, representing Gefitinib, is pictured occupying the ATP cleft. The locations of the two missense mutations are shown, within the activating loop of the kinase; the three in-frame deletions are all present within another loop, which flanks the ATP cleft.

Figure 4C is a close-up of the EGFR kinase domain, showing the critical amino acid residues implicated in binding to either ATP or to the inhibitor. Specifically, 4-anilinoquinazoline compounds such as gefitinib inhibit catalysis by occupying the ATP-binding site, where they form hydrogen bonds with methionine⁷⁹³ (M793) and cysteine⁷⁷⁵ (C775) residues, whereas their anilino ring is close to methionine⁷⁶⁵ (M766), lysine⁷⁴⁵ (K745), and leucine⁷⁸⁸ (L788) residues. In-frame deletions within the loop that is targeted by mutations are predicted to alter the position of these amino acids relative to the inhibitor. Mutated residues are shown within the activation loop of the tyrosine kinase.

Figure 5 shows the nucleotide and amino acid sequence of the *erbB1* gene. The amino acids are depicted as single letters, known to those of skill in the art. Nucleotide variances in the kinase domain are highlighted by patient number, see Table 2. SEQ ID NO: 511 includes nucleotides 1 through 3633. SEQ ID NO: 512 includes amino acids 1 through 1210.

Figures 6A-6C: Sequence alignment of selected regions within the EGFR and B-Raf kinase domains. Depiction of EGFR mutations in human NSCLC. EGFR (gb:X00588;) mutations in NSCLC tumors are highlighted in gray. B-Raf (gb:M95712) mutations in multiple tumor types (5) are highlighted in black. Asterisks denote residues conserved between EGFR and B-Raf. FIG. 6A depicts L858R mutations in the activation loop (SEQ ID NOS 477-479). FIG. 6B depicts the G719S mutant in the P-loop (SEQ ID NOS 480-482). FIG. 6C depicts deletion mutants in EGFR exon 19 (SEQ ID NOS 483-489).

Figure 7: Positions of missense mutations G719S and L858R and the Del-1 deletion in the three-dimensional structure of the EGFR kinase domain. The activation loop is shown in yellow, the P-loop is in blue and the C-lobe and N-lobe are as indicated. The residues targeted by mutation or deletion are highlighted in red. The Del-1 mutation targets the residues ELREA in codons 746 to 750. The mutations are located in highly conserved regions within kinases and are found in the p-loop and activation loop, which surround the region where ATP and also gefitinib and erlotinib are predicted to bind.

Figures 8A - 8F. Representative chromatograms of EGFR DNA from normal tissue and from tumor tissues. The locations of the identified mutations are as follows. Fig. 8A depicts the Exon 18 Kinase domain P loop (SEQ ID NOS 659-660). Fig. 8B depicts the Exon 21 Kinase domain A-loop (SEQ ID NOS 661-662). Fig. 8C depicts the Exon 19 Kinase domain Del-1 (SEQ ID NOS 663-665). Fig. 8D depicts the Exon 19 Kinase domain Del-3 (SEQ ID NOS 666-668). Fig. 8E depicts the Exon 19 Kinase domain Del-4 (SEQ ID NOS 669-671). Fig. 8F depicts the Exon 19 Kinase domain Del-5 (SEQ ID NOS 672-674).

FIG. 9: Sequence alignment of the EGFR and BCR-ABL polypeptides and the location of residues conferring a drug resistant phenotype. The EGFR polypeptide (SEQ ID NO:492) encoded by the nucleotide sequence disclosed in GenBank accno. NM_005228 and the BCR-ABL polypeptide (SEQ ID NO:491) encoded by the nucleotide sequence disclosed in GenBank accno. M14752 are aligned and conserved residues are shaded. BCR-ABL mutations conferring resistance to the tyrosine kinase inhibitor imatinib (STI571, Glivec/Gleevec) are denoted by asterisks.

Figure 10 shows the decision making process for patient with metastatic NSCLC undergoing EGFR testing.

Figure 11 shows a diagram of EGFR exons 18-24 (not to scale). Arrows depict the location of identified mutations. Asterisks denote the number of patients with mutations at each location. The blow-up diagram depicts the overlap of the exon 19 deletions, and the number of patients (n) with each deletion. Note that these are the results are not meant to be inclusive of all the EGFR mutations to date.

DETAILED DESCRIPTION

[0043] The present invention provides a novel method to determine the likelihood of effectiveness of an epidermal growth factor receptor (EGFR) targeting treatment in a patient affected with cancer. The method comprises detecting the presence or absence of at least one nucleic acid variance in the kinase domain of the *erbB1* gene of said patient. The presence of at least one variance indicates that the EGFR targeting treatment is likely to be effective. Preferably, the nucleic acid variance increases the kinase activity of the EGFR. The patient can then be treated with an EGFR targeting treatment. In one embodiment of the present invention, the EGFR targeting treatment is a tyrosine kinase inhibitor. In a preferred embodiment, the tyrosine kinase inhibitor is an anilinoquinazoline. The anilinoquinazoline may be a synthetic anilinoquinazoline. Preferably, the synthetic anilinoquinazoline is either gefitinib or erlotinib.

Definitions:

[0044] The terms "ErbB 1", "epidermal growth factor receptor", and "EGFR" are used interchangeably herein and refer to native sequence EGFR as disclosed, for example, in Carpenter et al. Ann. Rev. Biochem. 56:881-914 (1987), including variants thereof (e.g. a deletion mutant EGFR as in Humphrey et al. PNAS (USA) 87:4207-4211 (1990)). erbB1 refers to the gene encoding the EGFR protein product.

[0045] The term "kinase activity increasing nucleic acid variance" as used herein refers to a variance (i.e. mutation) in the nucleotide sequence of a gene that results in an increased kinase activity. The increased kinase activity is a direct result of the variance in the nucleic acid and is associated with the protein for which the gene encodes.

[0046] The term "drug" or "compound" as used herein refers to a chemical entity or biological product, or combination of chemical entities or biological products, administered to a person to treat or prevent or control a disease or condition. The chemical entity or biological product is preferably, but not necessarily a low molecular weight compound, but may also be a larger compound, for example, an oligomer of nucleic acids, amino acids, or carbohydrates including without limitation proteins, oligonucleotides, ribozymes, DNAzymes, glycoproteins, siRNAs, lipoproteins, aptamers, and modifications and combinations thereof.

[0047] The term "genotype" in the context of this invention refers to the particular allelic form of a gene, which can be defined by the particular nucleotide(s) present in a nucleic acid sequence at a particular site(s).

[0048] The terms "variant form of a gene", "form of a gene", or "allele" refer to one specific form of a gene in a population, the specific form differing from other forms of the same gene in the sequence of at least one, and frequently more than one, variant sites within the sequence of the gene. The sequences at these variant sites that differ between different alleles of the gene are termed "gene sequence variances" or "variances" or "variants". Other terms known in the art to be equivalent include mutation and polymorphism, although mutation is often used to refer to an allele associated with a deleterious phenotype. In preferred aspects of this invention, the variances are selected from the group consisting of the variances listed in the variance tables herein.

[0049] In the context of this invention, the term "probe" refers to a molecule which can detectably distinguish between target molecules differing in structure. Detection can be accomplished in a variety of different ways depending on the type of probe used and the type of target molecule. Thus, for example, detection may be based on discrimination of activity levels of the target molecule, but preferably is based on detection of specific binding. Examples of such specific binding include antibody binding and nucleic acid probe hybridization. Thus, for example, probes can include enzyme substrates, antibodies and antibody fragments, and preferably nucleic acid hybridization probes.

[0050] As used herein, the terms "effective" and "effectiveness" includes both pharmacological effectiveness and physiological safety. Pharmacological effectiveness refers to the ability of the treatment to result in a desired biological effect in the patient. Physiological safety refers to the level of toxicity, or other adverse physiological effects at the cellular, organ and/or organism level (often referred to as side-effects) resulting from administration of the treatment. "Less effective" means that the treatment results in a therapeutically significant lower level of pharmacological effectiveness and/or a therapeutically greater level of adverse physiological effects.

[0051] The term "primer", as used herein, refers to an oligonucleotide which is capable of acting as a point of initiation of polynucleotide synthesis along a complementary strand when placed under conditions in which synthesis of a primer extension product which is complementary to a polynucleotide is catalyzed. Such conditions include the presence of four different nucleotide triphosphates or nucleoside analogs and one or more agents for polymerization such as DNA polymerase and/or reverse transcriptase, in an appropriate buffer ("buffer" includes substituents which are cofactors, or which affect pH, ionic strength, etc.), and at a suitable temperature. A primer must be sufficiently long to prime the synthesis of extension products in the presence of an agent for polymerase. A typical primer contains at least about 5 nucleotides in length of a sequence substantially complementary to the target sequence, but somewhat longer primers are preferred. Usually primers contain about 15-26 nucleotides, but longer primers may also be employed.

[0052] A primer will always contain a sequence substantially complementary to the target sequence, that is the specific sequence to be amplified, to which it can anneal. A primer may, optionally, also comprise a promoter sequence.. The term "promoter sequence" defines a single strand of a nucleic acid sequence that is specifically recognized by an RNA polymerase that binds to a recognized sequence and initiates the process of transcription by which an RNA transcript is produced. In principle, any promoter sequence may be employed for which there is a known and available polymerase that is capable of recognizing the initiation sequence. Known and useful promoters are those that are recognized by certain bacteriophage polymerases, such as bacteriophage T3, T7 or SP6.

[0053] A "microarray" is a linear or two-dimensional array of preferably discrete regions, each having a defined area, formed on the surface of a solid support. The density of the discrete regions on a microarray is determined by the total numbers of target polynucleotides to be detected on the surface of a single solid phase support, preferably at least about 50/cm², more preferably at least about 100/cm², even more preferably at least about 500/cm², and still more preferably at least about 1,000/cm². As used herein, a DNA microarray is an array of oligonucleotide primers placed on a chip or other surfaces used to amplify or clone target polynucleotides. Since the position of each particular group of primers in

the array is known, the identities of the target polynucleotides can be determined based on their binding to a particular position in the microarray.

[0054] The term "label" refers to a composition capable of producing a detectable signal indicative of the presence of the target polynucleotide in an assay sample. Suitable labels include radioisotopes, nucleotide chromophores, enzymes, substrates, fluorescent molecules, chemiluminescent moieties, magnetic particles, bioluminescent moieties, and the like. As such, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means.

[0055] The term "support" refers to conventional supports such as beads, particles, dipsticks, fibers, filters, membranes and silane or silicate supports such as glass slides.

[0056] The term "amplify" is used in the broad sense to mean creating an amplification product which may include, for example, additional target molecules, or target-like molecules or molecules complementary to the target molecule, which molecules are created by virtue of the presence of the target molecule in the sample. In the situation where the target is a nucleic acid, an amplification product can be made enzymatically with DNA or RNA polymerases or reverse transcriptases.

[0057] As used herein, a "biological sample" refers to a sample of tissue or fluid isolated from an individual, including but not limited to, for example, blood, plasma, serum, tumor biopsy, urine, stool, sputum, spinal fluid, pleural fluid, nipple aspirates, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, cells (including but not limited to blood cells), tumors, organs, and also samples of *in vitro* cell culture constituent. In a preferred embodiment, the sample is from a resection, bronchoscopic biopsy, or core needle biopsy of a primary or metastatic tumor, or a cellblock from pleural fluid. In addition, fine needle aspirate samples are used. Samples may be either paraffin-embedded or frozen tissue.

[0058] The term "antibody" is meant to be an immunoglobulin protein that is capable of binding an antigen. Antibody as used herein is meant to include antibody fragments, e.g. F(ab)₂, Fab', Fab, capable of binding the antigen or antigenic fragment of interest. Preferably, the binding of the antibody to the antigen inhibits the activity of a variant form of EGFR.

[0059] The term "humanized antibody" is used herein to describe complete antibody molecules, i.e. composed of two complete light chains and two complete heavy chains, as well as antibodies consisting only of antibody fragments, e.g. Fab, Fab', F(ab)₂, and Fv, wherein the CDRs are derived from a non-human source and the remaining portion of the Ig molecule or fragment thereof is derived from a human antibody, preferably produced from a nucleic acid sequence encoding a human antibody.

[0060] The terms "human antibody" and "humanized antibody" are used herein to describe an antibody of which all portions of the antibody molecule are derived from a nucleic acid sequence encoding a human antibody. Such human antibodies are most desirable for use in antibody therapies, as such antibodies would elicit little or no immune response in the human patient.

[0061] The term "chimeric antibody" is used herein to describe an antibody molecule as well as antibody fragments, as described above in the definition of the term "humanized antibody." The term "chimeric antibody" encompasses humanized antibodies. Chimeric antibodies have at least one portion of a heavy or light chain amino acid sequence derived from a first mammalian species and another portion of the heavy or light chain amino acid sequence derived from a second, different mammalian species.

[0062] Preferably, the variable region is derived from a non-human mammalian species and the constant region is derived from a human species. Specifically, the chimeric antibody is preferably produced from a 9 nucleotide sequence from a non-human mammal encoding a variable region and a nucleotide sequence from a human encoding a constant region of an antibody.

[0063] Table 2 is a partial list of DNA sequence variances in the kinase domain of erbB1 relevant to the methods described in the present invention. These variances were identified by the inventors in studies of biological samples from patients with NSCLC who responded to gefitinib and patients with no exposure to gefitinib.

[0064] Nucleic acid molecules can be isolated from a particular biological sample using any of a number of procedures, which are well-known in the art, the particular isolation procedure chosen being appropriate for the particular biological sample. For example, freeze-thaw and alkaline lysis procedures can be useful for obtaining nucleic acid molecules from solid materials; heat and alkaline lysis procedures can be useful for obtaining nucleic acid molecules from urine; and proteinase K extraction can be used to obtain nucleic acid from blood (Rolff, A et al. PCR: Clinical Diagnostics and Research, Springer (1994).

Detection Methods

[0065] Determining the presence or absence of a particular variance or plurality of variances in the kinase domain of the erbB1 gene in a patient with or at risk for developing cancer can be performed in a variety of ways. Such tests are commonly performed using DNA or RNA collected from biological samples, e.g., tissue biopsies, urine, stool, sputum, blood, cells, tissue scrapings, breast aspirates or other cellular materials, and can be performed by a variety of methods

including, but not limited to, PCR, hybridization with allele-specific probes, enzymatic mutation detection, chemical cleavage of mismatches, mass spectrometry or DNA sequencing, including minisequencing. In particular embodiments, hybridization with allele specific probes can be conducted in two formats: (1) allele specific oligonucleotides bound to a solid phase (glass, silicon, nylon membranes) and the labeled sample in solution, as in many DNA chip applications, or (2) bound sample (often cloned DNA or PCR amplified DNA) and labeled oligonucleotides in solution (either allele specific or short so as to allow sequencing by hybridization). Diagnostic tests may involve a panel of variances, often on a solid support, which enables the simultaneous determination of more than one variance.

[0066] In another aspect, determining the presence of at least one kinase activity increasing nucleic acid variance in the erbB1 gene may entail a haplotyping test. Methods of determining haplotypes are known to those of skill in the art, as for example, in WO 00/04194.

[0067] Preferably, the determination of the presence or absence of a kinase activity increasing nucleic acid variance involves determining the sequence of the variance site or sites by methods such as polymerase chain reaction (PCR). Alternatively, the determination of the presence or absence of a kinase activity increasing nucleic acid variance may encompass chain terminating DNA sequencing or minisequencing, oligonucleotide hybridization or mass spectrometry.

[0068] The methods of the present invention may be used to predict the likelihood of effectiveness (or lack of effectiveness) of an EGFR targeting treatment in a patient affected with or at risk for developing cancer. Preferably, cancers include cancer of epithelial origin, including, but are not limited to, gastrointestinal cancer, prostate cancer, ovarian cancer, breast cancer, head and neck cancer, lung cancer, non-small cell lung cancer, cancer of the nervous system, kidney cancer, retina cancer, skin cancer, liver cancer, pancreatic cancer, genital-urinary cancer and bladder cancer. In a preferred embodiment, the cancer is non-small cell lung cancer.

[0069] The present invention generally concerns the identification of variances in the kinase domain of the erbB1 gene which are indicative of the effectiveness of an EGFR targeting treatment in a patient with or at risk for developing cancer. Additionally, the identification of specific variances in the kinase domain of EGFR, in effect, can be used as a diagnostic or prognostic test. For example, the presence of at least one variance in the kinase domain of erbB1 indicates that a patient will likely benefit from treatment with an EGFR targeting compound, such as, for example, a tyrosine kinase inhibitor.

[0070] Methods for diagnostic tests are well known in the art and disclosed in patent application WO 00/04194. In an exemplary method, the diagnostic test comprises amplifying a segment of DNA or RNA (generally after converting the RNA to cDNA) spanning one or more known variances in the kinase domain of the erbB1 gene sequence. This amplified segment is then sequenced and/or subjected to polyacrylamide gel electrophoresis in order to identify nucleic acid variances in the amplified segment.

PCR

[0071] In one embodiment, the invention provides a method of screening for variants in the kinase domain of the erbB1 gene in a test biological sample by PCR or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran, et al., 1988. Science 241: 1077-1080; and Nakazawa, et al., 1994. Proc. Natl. Acad. Sci. USA 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the EGFR-gene (see, Abravaya, et al., 1995. Nucl. Acids Res. 23: 675-682). The method comprises the steps of designing degenerate primers for amplifying the target sequence, the primers corresponding to one or more conserved regions of the gene, amplifying reaction with the primers using, as a template, a DNA or cDNA obtained from a test biological sample and analyzing the PCR products. Comparison of the PCR products of the test biological sample to a control sample indicates variances in the test biological sample. The change can be either and absence or presence of a nucleic acid variance in the test biological sample.

[0072] Alternative amplification methods include: self sustained sequence replication (see, Guatelli, et al., 1990. Proc. Natl. Acad. Sci. USA 87: 1874-1 878), transcriptional amplification system (see, Kwoh, et al., 1989. Proc. Natl. Acad. Sci. USA 86: 1173-1177); Qb Replicase (see, Lizardi, et al, 1988. BioTechnology 6: 1197), 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.

[0073] Primers useful according to the present invention are designed using amino acid sequences of the protein or nucleic acid sequences of the kinase domain of the erbB1 gene as a guide, e.g. SEQ ID NO: 493, SEQ ID NO: 494, SEQ ID NO: 509, and SEQ ID NO: 510. The primers are designed in the homologous regions of the gene wherein at least two regions of homology are separated by a divergent region of variable sequence, the sequence being variable either in length or nucleic acid sequence.

[0074] For example, the identical or highly, homologous, preferably at least 80%-85% more preferably at least 90-99% homologous amino acid sequence of at least about 6, preferably at least 8-10 consecutive amino acids. Most preferably, the amino acid sequence is 100% identical. Forward and reverse primers are designed based upon the maintenance of codon degeneracy and the representation of the various amino acids at a given position among the known gene family

members. Degree of homology as referred to herein is based upon analysis of an amino acid sequence using a standard sequence comparison software, such as protein-BLAST using the default settings (<http://www.ncbi.nlm.nih.gov/BLAST/>).

Table 3 below represents the usage of degenerate codes and their standard symbols:

	T	C	A	G
T	TTT Phe (F) TTC " TTA Leu (L) TTG "	TCT Ser (S) TCC " TCA " TCG "	TAT Tyr (Y) TAC TAA Ter TAG Ter	TGT Cys (C) TGC TGA Ter TGG Trp (W)
C	CTT Leu (L) CTC " CTA " CTG "	CCT Pro (P) CCC " CCA " CCG "	CAT His (H) CAC " CAA Gln (Q) CAG "	CGT Arg (R) CGC " CGA " CGG "
A	ATT Ile (I) ATC " ATA " ATG Met	ACT Thr (T) ACC " ACA " ACG "	AAT Asn (N) AAC " AAA Lys AAG "	AGT Ser (S) AGC " AGA Arg (R) AGG "
G	GTT Val (V) GTC " GTA " GTG "	GCT Ala (A) GCC " GCA " GCG "	GAT Asp (D) GAC " GAA Glu (E) GAG "	GGT Gly (G) GGC " GGA " GGG "

[0075] Preferably any 6-fold degenerate codons such as L, R and S are avoided since in practice they will introduce higher than 6-fold degeneracy. In the case of L, TTR and CTN are compromised YTN (8-fold degeneracy), in the case of R, CGN and AGR compromises at MGN (8-fold degeneracy), and finally S, TCN and AGY which can be compromised to WSN (16-fold degeneracy). In all three cases on 6 of these will match the target sequence. To avoid this loss of specificity, it is preferable to avoid these regions, or to make two populations, each with the alternative degenerate codon, e.g. for S include TCN in one pool, and AGY in the other.

[0076] Primers may be designed using a number of available computer programs, including, but not limited to Oligo Analyzer3.0; Oligo Calculator; NetPrimer; Methprimer; Primer3; WebPrimer; PrimerFinder; Primer9; Oligo2002; Pride or GenomePride; Oligos; and Codehop. Detailed information about these programs can be obtained, for example, from www.molbiol.net.

[0077] Primers may be labeled using labels known to one skilled in the art. Such labels include, but are not limited to radioactive, fluorescent, dye, and enzymatic labels.

[0078] Analysis of amplification products can be performed using any method capable of separating the amplification products according to their size, including automated and manual gel electrophoresis, mass spectrometry, and the like.

[0079] Alternatively, the amplification products can be separated using sequence differences, using SSCP, DGGE, TGGE, chemical cleavage or restriction fragment polymorphisms as well as hybridization to, for example, a nucleic acid arrays.

[0080] The methods of nucleic acid isolation, amplification and analysis are routine for one skilled in the art and examples of protocols can be found, for example, in the Molecular Cloning: A Laboratory Manual (3-Volume Set) Ed. Joseph Sambrook, David W. Russel, and Joe Sambrook, Cold Spring Harbor Laboratory; 3rd edition (January 15, 2001), ISBN: 0879695773. Particularly useful protocol source for methods used in PCR amplification is PCR (Basics: From Background to Bench) by M. J. McPherson, S. G. Møller, R. Beynon, C. Howe, Springer Verlag; 1st edition (October 15, 2000), ISBN: 0387916008.

[0081] Preferably, exons 19 and 21 of human EGFR are amplified by the polymerase chain reaction (PCR) using the following primers: Exon19 sense primer, 5'- GCAATATCAGCCTTAGGTGCGGCTC-3' (SEQ ID NO: 505); Exon 19 antisense primer, 5'-CATAGAAAGTGAACATTTAGGATGTG-3' (SEQ ID NO: 506); Exon21 sense primer, 5'-CTAACGT-TCG CCAGCCATAAGTCC-3' (SEQ ID NO: 507); and Exon21 antisense primer, 5'- GCTGCGAGCTACCCAG AAT-GTCTGG-3' (SEQ ID NO: 508).

[0082] In an alternative embodiment, mutations in a EGFR gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample

DNA. Moreover, the use of sequence specific ribozymes (see, e.g., U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

[0083] Other methods for detecting mutations in the EGFR gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. See, e.g., Myers, et al., 1985. Science 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type EGFR sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to 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 digesting 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 the site of mutation. See, e.g., Cotton, et al., 1988. Proc. Natl. Acad. Sci. USA 85: 4397; Saleeba, et al., 1992. Methods Enzymol. 2 17: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

[0084] In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in EGFR cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. See, e.g., Hsu, et al., 1994. Carcinogenesis 15: 1657-1662. According to an exemplary embodiment, a probe based on a mutant EGFR sequence, e.g., a DEL-1 through DEL-5, G719S, G857V, L883S or L858R EGFR sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, e.g., U.S. Patent No. 5,459,039.

[0085] In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in EGFR genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. See, e.g., Orita, et al., 1989. Proc. Natl. Acad. Sci. USA: 86: 2766; Cotton, 1993. Mutat. Res. 285: 125-144; Hayashi, 1992. Genet. Anal. Tech. Appl. 9: 73-79. Single-stranded DNA fragments of sample and control EGFR nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, 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 one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. See, e.g., Keen, et al., 1991. Trends Genet. 7: 5.

[0086] In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). See, e.g., 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 place of a denaturing gradient to identify differences in the mobility of control and sample DNA. See, e.g., Rosenbaum and Reissner, 1987. Biophys. Chem. 265: 12753.

[0087] Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. See, e.g., Saiki, et al., 1986. Nature 324: 163; Saiki, et al., 1989. Proc. Natl. Acad. Sci. USA 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

[0088] Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides 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; see, e.g., Gibbs, et al., 1989. Nucl. Acids Res. 17: 2437-24-48) or at the extreme 3-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (see, e.g., Prossner, 1993. Tibtech. 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. See, e.g., Gasparini, et al., 1992. Mol. Cell Probes 6: 1. It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification. See, e.g., Barany, 1991. Proc. Natl. Acad. Sci. USA 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

Solid Support and Probe

[0089] In an alternative embodiment, the detection of the presence or absence of the at least one nucleic acid variance involves contacting a nucleic acid sequence corresponding to the desired region of the erbB1 gene, identified above, with a probe. The probe is able to distinguish a particular form of the gene or the presence or a particular variance or variances, e.g., by differential binding or hybridization. Thus, exemplary probes include nucleic acid hybridization probes, peptide nucleic acid probes, nucleotide-containing probes which also contain at least one nucleotide analog, and antibodies, e.g., monoclonal antibodies, and other probes as discussed herein. Those skilled in the art are familiar with the preparation of probes with particular specificities. Those skilled in the art will recognize that a variety of variables can be adjusted to optimize the discrimination between two variant forms of a gene, including changes in salt concentration, temperature, pH and addition of various compounds that affect the differential affinity of GC vs. AT base pairs, such as tetramethyl ammonium chloride. (See Current Protocols in Molecular Biology by F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, K. Struhl and V. B. Chanda (Editors), John Wiley & Sons.)

[0090] Thus, in preferred embodiments, the detection of the presence or absence of the at least one variance involves contacting a nucleic acid sequence which includes at least one variance site with a probe, preferably a nucleic acid probe, where the probe preferentially hybridizes with a form of the nucleic acid sequence containing a complementary base at the variance site as compared to hybridization to a form of the nucleic acid sequence having a non-complementary base at the variance site, where the hybridization is carried out under selective hybridization conditions. Such a nucleic acid hybridization probe may span two or more variance sites. Unless otherwise specified, a nucleic acid probe can include one or more nucleic acid analogs, labels or other substituents or moieties so long as the base-pairing function is retained.

[0091] The probe may be designed to bind to, for example, at least three continuous nucleotides on both sides of the deleted region of SEQ ID NO: 495, SEQ ID NO: 497, or SEQ ID NO: 499. Such probes, when hybridized under the appropriate conditions, will bind to the variant form of EGFR, but will not bind to the wildtype EGFR.

[0092] Such hybridization probes are well known in the art (see, e.g., Sambrook et al., Eds., (most recent edition), Molecular Cloning: A Laboratory Manual, (third edition, 2001), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.). Stringent hybridization conditions will typically include salt concentrations of less than about 1M, more usually less than about 500 mM and preferably less than about 200 mM. Hybridization temperatures can be as low as 5°C, but are typically greater than 22°C, more typically greater than about 30°C, and preferably in excess of about 37°C. Longer fragments may require higher hybridization temperatures for specific hybridization. Other factors may affect the stringency of hybridization, including base composition and length of the complementary strands, presence of organic solvents and extent of base mismatching; the combination of parameters used is more important than the absolute measure of any one alone. Other hybridization conditions which may be controlled include buffer type and concentration, solution pH, presence and concentration of blocking reagents (e.g., repeat sequences, CotI DNA, blocking protein solutions) to decrease background binding, detergent type(s) and concentrations, molecules such as polymers which increase the relative concentration of the polynucleotides, metal ion(s) and their concentration(s), chelator(s) and their concentrations, and other conditions known or discoverable in the art. Formulas may be used to predict the optimal melting temperature for a perfectly complementary sequence for a given probe, but true melting temperatures for a probe under a set of hybridization conditions must be determined empirically. Also, a probe may be tested against its exact complement to determine a precise melting temperature under a given set of condition as described in Sambrook et al, "Molecular Cloning," 3rd edition, Cold Spring Harbor Laboratory Press, 2001. Hybridization temperatures can be systematically altered for a given hybridization solution using a support associated with target polynucleotides until a temperature range is identified which permits detection of binding of a detectable probe at the level of stringency desired, either at high stringency where only target polynucleotides with a high degree of complementarity hybridize, or at lower stringency where additional target polynucleotides having regions of complementarity with the probe detectably hybridize above the background level provided from nonspecific binding to noncomplementary target polynucleotides and to the support. When hybridization is performed with potential target polynucleotides on a support under a given set of conditions, the support is then washed under increasing conditions of stringency (typically lowered salt concentration and/or increased temperature, but other conditions may be altered) until background binding is lowered to the point where distinct positive signals may be seen. This can be monitored in progress using a Geiger counter where the probe is radiolabeled, radiographically, using a fluorescent imager, or by other means of detecting probe binding. The support is not allowed to dry during such procedures, or the probe may become irreversibly bound even to background locations. Where a probe produces undesirable background or false positives, blocking reagents are employed, or different regions of the probe or different probes are used until positive signals can be distinguished from background. Once conditions are found that provide satisfactory signal above background, the target polynucleotides providing a positive signal are isolated and further characterized. The isolated polynucleotides can be sequenced; the sequence can be compared to databank entries or known sequences; where necessary, full-length clones can be obtained by techniques known in the art; and the polynucleotides can be expressed using suitable vectors and hosts to determine if the polynucleotide identified

encodes a protein having similar activity to that from which the probe polynucleotide was derived. The probes can be from 10-50 nucleotides. However, much oarger probes can also be employed, e.g., 50-500 nucleotides or larger. Solid Phase Support

[0093] The solid phase support of the present invention can be of any solid materials and structures suitable for supporting nucleotide hybridization and synthesis. Preferably, the solid phase support comprises at least one substantially rigid surface on which oligonucleotides or oligonucleotide primers can be immobilized. The solid phase support can be made of, for example, glass, synthetic polymer, plastic, hard non-mesh nylon or ceramic. Other suitable solid support materials are known and readily available to those of skill in the art. The size of the solid support can be any of the standard microarray sizes, useful for DNA microarray technology, and the size may be tailored to fit the particular machine being used to conduct a reaction of the invention. Methods and materials for derivatization of solid phase supports for the purpose of immobilizing oligonucleotides are known to those skill in the art and described in, for example, U.S. Pat. No. 5,919,523.

[0094] The solid support can be provided in or be part of a fluid containing vessel. For example, the solid support can be placed in a chamber with sides that create a seal along the edge of the solid support so as to contain the polymerase chain reaction (PCR) on the support. In a specific example the chamber can have walls on each side of a rectangular support to ensure that the PCR mixture remains on the support and also to make the entire surface useful for providing the primers.

[0095] The oligonucleotide or oligonucleotide primers of the invention are affixed, immobilized, provided, and/or applied to the surface of the solid support using any available means to fix, immobilize, provide and/or apply the oligonucleotides at a particular location on the solid support. For example, photolithography (Affymetrix, Santa Clara, Calif.) can be used to apply the oligonucleotide primers at particular position on a chip or solid support, as described in the U.S. Pat. Nos. 5,919,523, 5,837,832, 5,831,070, and 5,770, 722. The oligonucleotide primers may also be applied to a solid support as described in Brown and Shalon, U.S. Pat. No. 5,807,522 (1998). Additionally, the primers may be applied to a solid support using a robotic system, such as one manufactured by Genetic MicroSystems (Woburn, Mass.), GeneMachines (San Carlos, Calif.) or Cartesian Technologies (Irvine, Calif.).

[0096] In one aspect of the invention, solid phase amplification of target polynucleotides from a biological sample is performed, wherein multiple groups of oligonucleotide primers are immobilized on a solid phase support. In a preferred embodiment, the primers within a group comprises at least a first set of primers that are identical in sequence and are complementary to a defined sequence of the target polynucleotide, capable of hybridizing to the target polynucleotide under appropriate conditions, and suitable as initial primers for nucleic acid synthesis (i.e., chain elongation or extension). Selected primers covering a particular region of the reference sequence are immobilized, as a group, onto a solid support at a discrete location. Preferably, the distance between groups is greater than the resolution of detection means to be used for detecting the amplified products. In a preferred embodiment, the primers are immobilized to form a microarray or chip that can be processed and analyzed via automated, processing. The immobilized primers are used for solid phase amplification of target polynucleotides under conditions suitable for a nucleic acid amplification means. In this manner, the presence or absence of a variety of potential variances in the kinase domain of the erbB1 gene can be determined in one assay.

[0097] A population of target polynucleotides isolated from a healthy individual can used as a control in determining whether a biological source has at least one kinase activity increasing variance in the kinase domain of the erb1 gene. Alternatively, target polynucleotides isolated from healthy tissue of the same individual may be used as a control as above.

[0098] An in situ-type PCR reactions on the microarrays can be conducted essentially as described in e.g. Embretson et al, Nature 362:359-362 (1993); Gosden et al, BioTechniques 15(1):78-80 (1993); Heniford et al Nuc. Acid Res. 21(14):3159-3166 (1993); Long et al, Histochemistry 99:151-162 (1993); Nuovo et al, PCR Methods and Applications 2(4):305-312 (1993); Patterson et al Science 260:976-979 (1993).

[0099] Alternatively, variances in the kinase domain of erbB1 can be determined by solid phase techniques without performing PCR on the support. A plurality of oligonucleotide probes, each containing a distinct variance in the kinase domain of erbB1, in duplicate, triplicate or quadruplicate, may be bound to the solid phase support. The presence or absence of variances in the test biological sample may be detected by selective hybridization techniques, known to those of skill in the art and described above.

Mass Spectrometry

[0100] In another embodiment, the presence or absence of kinase activity increasing nucleic acid variances in the kinase domain of the erbB1 gene are determined using mass spectrometry. To obtain an appropriate quantity of nucleic acid molecules on which to perform mass spectrometry, amplification may be necessary. Examples of appropriate amplification procedures for use in the invention include: cloning (Sambrook et al., Molecular Cloning: A Laboratory Manual, 3rd Edition, Cold Spring Harbor Laboratory Press, 2001), polymerase chain reaction (PCR) (C. R. Newton and A. Graham, PCR, BIOS Publishers, 1994), ligase chain reaction (LCR) (Wiedmann, M., et al., (1994) PCR Methods

Appl. Vol. 3, Pp. 57-64; F. Barnay Proc. Natl. Acad. Sci USA 88, 189-93 (1991), strand displacement amplification (SDA) (G. Terrance Walker et al., Nucleic Acids Res. 22, 2670-77 (1994)) and variations such as RT-PCR (Higuchi, et al., Bio/Technology 11:1026-1030 (1993)), allele-specific amplification (ASA) and transcription based processes.

[0101] To facilitate mass spectrometric analysis, a nucleic acid molecule containing a nucleic acid sequence to be detected can be immobilized to a solid support. Examples of appropriate solid supports include beads (e.g. silica gel, controlled pore glass, magnetic, Sephadex/Sepharose, cellulose), flat surfaces or chips (e.g. glass fiber filters, glass surfaces, metal surface (steel, gold, silver, aluminum, copper and silicon), capillaries, plastic (e.g. polyethylene, polypropylene, polyamide, polyvinylidenedifluoride membranes or microtiter plates)); or pins or combs made from similar materials comprising beads or flat surfaces or beads placed into pits in flat surfaces such as wafers (e.g. silicon wafers).

[0102] Immobilization can be accomplished, for example, based on hybridization between a capture nucleic acid sequence, which has already been immobilized to the support and a complementary nucleic acid sequence, which is also contained within the nucleic acid molecule containing the nucleic acid sequence to be detected. So that hybridization between the complementary nucleic acid molecules is not hindered by the support, the capture nucleic acid can include a spacer region of at least about five nucleotides in length between the solid support and the capture nucleic acid sequence. The duplex formed will be cleaved under the influence of the laser pulse and desorption can be initiated. The solid support-bound base sequence can be presented through natural oligoribo- or oligodeoxyribonucleotide as well as analogs (e.g. thio-modified phosphodiester or phosphotriester backbone) or employing oligonucleotide mimetics such as PNA analogs (see e.g. Nielsen et al., Science, 254, 1497 (1991)) which render the base sequence less susceptible to enzymatic degradation and hence increases overall stability of the solid support-bound capture base sequence.

[0103] Prior to mass spectrometric analysis, it may be useful to "condition" nucleic acid molecules, for example to decrease the laser energy required for volatilization and/or to minimize fragmentation. Conditioning is preferably performed while a target detection site is immobilized. An example of conditioning is modification of the phosphodiester backbone of the nucleic acid molecule (e.g. cation exchange), which can be useful for eliminating peak broadening due to a heterogeneity in the cations bound per nucleotide unit. Contacting a nucleic acid molecule with an alkylating agent such as alkyl iodide, iodoacetamide, β -iodoethanol, 2,3-epoxy-1-propanol, the monothio phosphodiester bonds of a nucleic acid molecule can be transformed into a phosphotriester bond. Likewise, phosphodiester bonds may be transformed to uncharged derivatives employing trialkylsilyl chlorides. Further conditioning involves incorporating nucleotides which reduce sensitivity for depurination (fragmentation during MS) such as N7- or N9-deazapurine nucleotides, or RNA building blocks or using oligonucleotide triesters or incorporating phosphorothioate functions which are alkylated or employing oligonucleotide mimetics such as PNA.

[0104] For certain applications, it may be useful to simultaneously detect more than one (mutated) loci on a particular captured nucleic acid fragment (on one spot of an array) or it may be useful to perform parallel processing by using oligonucleotide or oligonucleotide mimetic arrays on various solid supports. "Multiplexing" can be achieved by several different methodologies. For example, several mutations can be simultaneously detected on one target sequence by employing corresponding detector (probe) molecules (e.g. oligonucleotides or oligonucleotide mimetics). However, the molecular weight differences between the detector oligonucleotides D1, D2 and D3 must be large enough so that simultaneous detection (multiplexing) is possible. This can be achieved either by the sequence itself (composition or length) or by the introduction of mass-modifying functionalities M1-M3 into the detector oligonucleotide.

[0105] Preferred mass spectrometer formats for use in the invention are matrix assisted laser desorption ionization (MALDI), electrospray (ES), ion cyclotron resonance (ICR) and Fourier Transform. Methods of performing mass spectrometry are known to those of skill in the art and are further described in Methods of Enzymology, Vol. 193: "Mass Spectrometry" (J. A. McCloskey, editor), 1990, Academic Press, New York.

Sequencing

[0106] In other preferred embodiments, determining the presence or absence of the at least one kinase activity increasing nucleic acid variance involves sequencing at least one nucleic acid sequence. The sequencing involves the sequencing of a portion or portions of the kinase domain of erbB1 which includes at least one variance site, and may include a plurality of such sites. Preferably, the portion is 500 nucleotides or less in length, more preferably 100 nucleotides or less, and most preferably 45 nucleotides or less in length. Such sequencing can be carried out by various methods recognized by those skilled in the art, including use of dideoxy termination methods (e.g., using dye-labeled dideoxy nucleotides), minisequencing, and the use of mass spectrometric methods.

Immunodetection

[0107] In one embodiment, determining the presence or absence of the at least one kinase activity increasing nucleic acid variance involves determining the activation state of downstream targets of EGFR.

[0108] The inventors of the present application have compared the phosphorylation status of the major downstream

targets of EGFR. For example, the EGF-induced activation of Erk1 and Erk2, via Ras, of Akt via PLC γ /PI3K, and of STAT3 and STAT5 via JAK2, has been examined. Erk1 and Erk2, via Ras, Akt via PLC γ /PI3K, and STAT3 and STAT5 via JAK2 are essential downstream pathways mediating oncogenic effects of EGFR (R. N. Jorissen et al., *Exp. Cell Res.* 284, 31 (2003)).

[0109] The inventors of the present application have shown that EGF-induced Erk activation is indistinguishable among cells expressing wild-type EGFR or either of the two activating EGFR mutants.

[0110] In contrast, phosphorylation of both Akt and STAT5 was substantially elevated in cells expressing either of the mutant EGFRs. Increased phosphorylation of STAT3 was similarly observed in cells expressing mutant EGFRs. Thus, the selective EGF-induced autophosphorylation of C-terminal tyrosine residues within EGFR mutants is well correlated with the selective activation of downstream signaling pathways.

[0111] In one embodiment of the present application, the presence of EGFR mutations can be determined using immunological techniques well known in the art, e.g., antibody techniques such as immunohistochemistry, immunocytochemistry, FACS scanning, immunoblotting, radioimmunoassays, western blotting, immunoprecipitation, enzyme-linked immunosorbent assays (ELISA), and derivative techniques that make use of antibodies directed against activated downstream targets of EGFR. Examples of such targets include, for example, phosphorylated STAT3, phosphorylated STAT5, and phosphorylated Akt. Using phospho-specific antibodies, the activation status of STAT3, STAT5, and Akt can be determined. Activation of STAT3, STAT5, and Akt are useful as a diagnostic indicator of activating EGFR mutations.

[0112] In one embodiment of the present invention, the presence of activated (phosphorylated) STAT5, STAT3, or Akt indicates that an EGFR targeting treatment is likely to be effective.

[0113] The invention provides a method of screening for variants in the kinase domain of the erbB1 gene in a test biological sample by immunohistochemical or immunocytochemical methods.

[0114] Immunohistochemistry ("IHC") and immunocytochemistry ("ICC") techniques, for example, may be used. IHC is the application of immunochemistry to tissue sections, whereas ICC is the application of immunochemistry to cells or tissue imprints after they have undergone specific cytological preparations such as, for example, liquid-based preparations. Immunochemistry is a family of techniques based on the use of a specific antibody, wherein antibodies are used to specifically target molecules inside or on the surface of cells. The antibody typically contains a marker that will undergo a biochemical reaction, and thereby experience a change color, upon encountering the targeted molecules. In some instances, signal amplification may be integrated into the particular protocol, wherein a secondary antibody, that includes the marker stain, follows the application of a primary specific antibody.

[0115] Immunohistochemical assays are known to those of skill in the art (e.g., see Jalkanen, et al., *J. Cell. Biol.* 101:976-985 (1985); Jalkanen, et al., *J. Cell. Biol.* 105:3087-3096 (1987).

[0116] Antibodies, polyclonal or monoclonal, can be purchased from a variety of commercial suppliers, or may be manufactured using well-known methods, e. g., as described in Harlow et al., *Antibodies: A Laboratory Manual*, 2nd Ed; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1988). In general, examples of antibodies useful in the present invention include anti-phospho-STAT3, anti-phospho-STAT5, and anti-phospho-Akt antibodies. Such antibodies can be purchased, for example, from Upstate Biotechnology (Lake Placid, NY), New England Biolabs (Beverly, MA), NeoMarkers (Fremont, CA)

[0117] Typically, for immunohistochemistry, tissue sections are obtained from a patient and fixed by a suitable fixing agent such as alcohol, acetone, and paraformaldehyde, to which is reacted an antibody. Conventional methods for immunohistochemistry are described in Harlow and Lane (eds) (1988) In "Antibodies A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor, New York; Ausbel et al (eds) (1987), in *Current Protocols In Molecular Biology*, John Wiley and Sons (New York, NY). Biological samples appropriate for such detection assays include, but are not limited to, cells, tissue biopsy, whole blood, plasma, serum, sputum, cerebrospinal fluid, breast aspirates, pleural fluid, urine and the like.

[0118] For direct labeling techniques, a labeled antibody is utilized. For indirect labeling techniques, the sample is further reacted with a labeled substance.

[0119] Alternatively, immunocytochemistry may be utilized. In general, cells are obtained from a patient and fixed by a suitable fixing agent such as alcohol, acetone, and paraformaldehyde, to which is reacted an antibody. Methods of immunocytological staining of human samples is known to those of skill in the art and described, for example, in Brauer et al., 2001 (*FASEB J*, 15, 2689- 2701), Smith-Swintosky et al., 1997.

[0120] Immunological methods of the present invention are advantageous because they require only small quantities of biological material. Such methods may be done at the cellular level and thereby necessitate a minimum of one cell. Preferably, several cells are obtained from a patient affected with or at risk for developing cancer and assayed according to the methods of the present invention.

Other Diagnostic Methods

[0121] An agent for detecting mutant EGFR protein is an antibody capable of binding to mutant EGFR protein, preferably

an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., F_{ab} or $F_{(ab)2}$) can be used. The term "labeled", with 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. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mutant EGFR mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of mutant EGFR mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of mutant EGFR protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of mutant EGFR genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of mutant EGFR protein include introducing into a subject a labeled anti- mutant EGFR protein antibody. 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.

[0122] In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject.

[0123] In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting mutant EGFR protein, mRNA, or genomic DNA, such that the presence of mutant EGFR protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of mutant EGFR protein, mRNA or genomic DNA in the control sample with the presence of mutant EGFR protein, mRNA or genomic DNA in the test sample.

[0124] In a different embodiment, the diagnostic assay is for mutant EGFR activity. In a specific embodiment, the mutant EGFR activity is a tyrosine kinase activity. One such diagnostic assay is for detecting EGFR-mediated phosphorylation of at least one EGFR substrate. Levels of EGFR activity can be assayed for, e.g., various mutant EGFR polypeptides, various tissues containing mutant EGFR, biopsies from cancer tissues suspected of having at least one mutant EGFR, and the like. Comparisons of the levels of EGFR activity in these various cells, tissues, or extracts of the same, can optionally be made. In one embodiment, high levels of EGFR activity in cancerous tissue is diagnostic for cancers that may be susceptible to treatments with one or more tyrosine kinase inhibitor. In related embodiments, EGFR activity levels can be determined between treated and untreated biopsy samples, cell lines, transgenic animals, or extracts from any of these, to determine the effect of a given treatment on mutant EGFR activity as compared to an untreated control.

Method of Treating a Patient

[0125] In one embodiment, the invention provides a method for selecting a treatment for a patient affected by or at risk for developing cancer by determining the presence or absence of at least one kinase activity increasing nucleic acid variance in the kinase domain of the *erbB1* gene. In another embodiment, the variance is a plurality of variances, whereby a plurality may include variances from one, two, three or more gene loci.

[0126] In certain embodiments, the presence of the at least one variance is indicative that the treatment will be effective or otherwise beneficial (or more likely to be beneficial) in the patient. Stating that the treatment will be effective means that the probability of beneficial therapeutic effect is greater than in a person not having the appropriate presence of the particular kinase activity increasing nucleic acid variance(s) in the kinase domain of the *erbB1* gene.

[0127] The treatment will involve the administration of a tyrosine kinase inhibitor. The treatment may involve a combination of treatments, including, but not limited to a tyrosine kinase inhibitor in combination with other tyrosine kinase inhibitors, chemotherapy, radiation, etc..

[0128] Thus, in connection with the administration of a tyrosine kinase inhibitor, a drug which is "effective against" a cancer indicates that administration in a clinically appropriate manner results in a beneficial effect for at least a statistically significant fraction of patients, such as a improvement of symptoms, a cure, a reduction in disease load, reduction in tumor mass or cell numbers, extension of life, improvement in quality of life, or other effect generally recognized as positive by medical doctors familiar with treating the particular type of disease or condition.

[0129] In a preferred embodiment, the compound is an anilinoquinazoline or synthetic anilinoquinazoline. European Patent Publication No. 0566226 discloses anilinoquinazolines which have activity against epidermal growth factor (EGF) receptor tyrosine kinase. It is also known from European Patent Applications Nos. 0520722 and 0566226 that certain 4- anilinoquinazoline derivatives are useful as inhibitors of receptor tyrosine kinases. The very tight structure-activity relationships shown by these compounds suggests a clearly-defined binding mode, where the quinazoline ring binds in the adenine pocket and the anilino ring binds in an adjacent, unique lipophilic pocket. Three 4-anilinoquinazoline analogues (two reversible and one irreversible inhibitor) have been evaluated clinically as anticancer drugs. Denny, *Farmaco* January-February 2001;56(1-2):51-6. Alternatively, the compound is EKB-569, an inhibitor of EGF receptor kinase (Tor-

rance et al., *Nature Medicine*, vol. 6, No. 9, Sept. 2000, p. 1024). In a most preferred embodiment, the compound is gefitinib (IRESSA®) or erlotinib (TARCEVA®).

[0130] Treatment targeting cancer cells containing at least one mutant EGFR described herein may be administered alone or in combination with any other appropriate anti-cancer treatment and/or therapeutic agent known to one skilled in the art. In one embodiment, treatment of a pathology, such as a cancer, is provided comprising administering to a subject in need thereof therapeutically effective amounts of a compound that inhibits EGFR kinase activity, such as gefitinib, erlotinib, etc., administered alone or in combination with at least one other anti-cancer agent or therapy. Inhibition of activated protein kinases through the use of targeted small molecule drugs or antibody-based strategies has emerged as an effective approach to cancer therapy. See, e.g., G. D. Demetri et al., *N. Engl. J. Med.* 347, 472 (2002); B. J. Druker et al., *N. Engl. J. Med.* 344, 1038 (2001); D.J. Slamon et al., *N. Engl. J. Med.* 344, 783 (2001).

[0131] In one embodiment, the anti-cancer agent is at least one chemotherapeutic agent. In a related embodiment, the anti-cancer agent is at least one radiotherapy. In a variant embodiment, the anti-cancer therapy is an antiangiogenic therapy (e.g., endostatin, angiostatin, TNP-470, Caplostatin (Stachi-Fainaro et al., *Cancer Cell* 7(3), 251 (2005))

[0132] The therapeutic agents may be the same or different, and may be, for example, therapeutic radionuclides, drugs, hormones, hormone antagonists, receptor antagonists, enzymes or proenzymes activated by another agent, autocrines, cytokines or any suitable anti-cancer agent known to those skilled in the art. In one embodiment, the anti-cancer agent is Avastin, an anti-VEGF antibody proven successful in anti-angiogenic therapy of cancer against both solid cancers and hematological malignancies. See, e.g., Ribatti et al. 2003 *J Hematother Stem Cell Res.* 12(1), 11-22. Toxins also can be used in the methods of the present invention. Other therapeutic agents useful in the present invention include anti-DNA, anti-RNA, radiolabeled oligonucleotides, such as antisense oligonucleotides, anti-protein and anti-chromatin cytotoxic or antimicrobial agents. Other therapeutic agents are known to those skilled in the art, and the use of such other therapeutic agents in accordance with the present invention is specifically contemplated.

[0133] The antitumor agent may be one of numerous chemotherapy agents such as an alkylating agent, an antime-tabolite, a hormonal agent, an antibiotic, an antibody, an anti-cancer biological, gleevec, colchicine, a vinca alkaloid, L-asparaginase, procarbazine, hydroxyurea, mitotane, nitrosoureas or an imidazole carboxamide. Suitable agents are those agents that promote depolarization of tubulin or prohibit tumor cell proliferation. Chemotherapeutic agents contemplated as within the scope of the invention include, but are not limited to, anti-cancer agents listed in the Orange Book of Approved Drug Products With Therapeutic Equivalence Evaluations, as compiled by the Food and Drug Administration and the U.S. Department of Health and Human Services. Nonlimiting examples of chemotherapeutic agents include, e.g., carboplatin and paclitaxel. Treatments targeting EGFR kinase activity can also be administered together with radiation therapy treatment. Additional anti-cancer treatments known in the art are contemplated as being within the scope of the invention.

[0134] The therapeutic agent may be a chemotherapeutic agent. Chemotherapeutic agents are known in the art and include at least the taxanes, nitrogen mustards, ethylenimine derivatives, alkyl sulfonates, nitrosoureas, triazenes; folic acid analogs, pyrimidine analogs, purine analogs, vinca alkaloids, antibiotics, enzymes, platinum coordination complexes, substituted urea, methyl hydrazine derivatives, adrenocortical suppressants, or antagonists. More specifically, the chemotherapeutic agents may be one or more agents chosen from the non-limiting group of steroids, progestins, estrogens, antiestrogens, or androgens. Even more specifically, the chemotherapy agents may be azaribine, bleomycin, bryostatin-1, busulfan, carmustine, chlorambucil, carboplatin, cisplatin, CPT-11, cyclophosphamide, cytarabine, dacarbazine, dactinomycin, daunorubicin, dexamethasone, diethylstilbestrol, doxorubicin, ethinyl estradiol, etoposide, fluorouracil, fluoxymesterone, gemcitabine, hydroxyprogesterone caproate, hydroxyurea, L-asparaginase, leucovorin, lomustine, mechlorethamine, medroprogesterone acetate, megestrol acetate, melphalan, mercaptopurine, methotrexate, methotrexate, mithramycin, mitomycin, mitotane, paclitaxel, phenyl butyrate, prednisone, procarbazine, semustine streptozocin, tamoxifen, taxanes, taxol, testosterone propionate, thalidomide, thioguanine, thiotepa, uracil mustard, vinblastine, or vincristine. The use of any combinations of chemotherapy agents is also contemplated. The administration of the chemotherapeutic agent may be before, during or after the administration of a treatment targeting EGFR activity.

[0135] Other suitable therapeutic agents are selected from the group consisting of radioisotope, boron addend, immunomodulator, toxin, photoactive agent or dye, cancer chemotherapeutic drug, antiviral drug, antifungal drug, antibacterial drug, antiprotozoal drug and chemosensitizing agent (See, U.S. Patent Nos. 4,925,648 and 4,932,412). Suitable chemotherapeutic agents are described in REMINGTON'S PHARMACEUTICAL SCIENCES, 19th Ed. (Mack Publishing Co. 1995), and in Goodman and Gilman's *The Pharmacological Basis of Therapeutics* (Goodman et al., Eds. Macmillan Publishing Co., New York, 1980 and 2001 editions). Other suitable chemotherapeutic agents, such as experimental drugs, are known to those of skill in the art. Moreover a suitable therapeutic radioisotope is selected from the group consisting of α -emitters, β -emitters, γ -emitters, Auger electron emitters, neutron capturing agents that emit α -particles and radioisotopes that decay by electron capture. Preferably, the radioisotope is selected from the group consisting of ^{225}Ac , ^{198}Au , ^{32}P , ^{125}I , ^{131}I , ^{90}Y , ^{186}Re , ^{188}Re , ^{67}Cu , ^{177}Lu , ^{213}Bi , ^{10}B , and ^{211}At .

[0136] Where more than one therapeutic agent is used, they may be the same or different. For example, the therapeutic agents may comprise different radionuclides, or a drug and a radionuclide. In a preferred embodiment, treatment targeting

EGFR activity inhibits mutant EGFR kinase activity.

[0137] In another embodiment, different isotopes that are effective over different distances as a result of their individual energy emissions are used as first and second therapeutic agents. Such agents can be used to achieve more effective treatment of tumors, and are useful in patients presenting with multiple tumors of differing sizes, as in normal clinical circumstances.

[0138] Few of the available isotopes are useful for treating the very smallest tumor deposits and single cells. In these situations, a drug or toxin may be a more useful therapeutic agent. Accordingly, in preferred embodiments of the present invention, isotopes are used in combination with non-isotopic species such as drugs, toxins, and neutron capture agents. Many drugs and toxins are known which have cytotoxic effects on cells, and can be used in connection with the present invention. They are to be found in compendia of drugs and toxins, such as the Merck Index, Goodman and Gilman, and the like, and in the references cited above.

[0139] Drugs that interfere with intracellular protein synthesis can also be used in the methods of the present invention; such drugs are known to those skilled in the art and include puromycin, cycloheximide, and ribonuclease.

[0140] The therapeutic methods of the invention may be used for cancer therapy. It is well known that radioisotopes, drugs, and toxins can be conjugated to antibodies or antibody fragments which specifically bind to markers which are produced by or associated with cancer cells, and that such antibody conjugates can be used to target the radioisotopes, drugs or toxins to tumor sites to enhance their therapeutic efficacy and minimize side effects. Examples of these agents and methods are reviewed in Wawrzynczak and Thorpe (in *Introduction to the Cellular and Molecular Biology of Cancer*, L. M. Franks and N. M. Teich, eds, Chapter 18, pp. 378-410, Oxford University Press, Oxford, 1986), in *Immunoconjugates: Antibody Conjugates in Radioimaging and Therapy of Cancer* (C. W. Vogel, ed., 3-300, Oxford University Press, N.Y., 1987), in Dillman, R. O. (*CRC Critical Reviews in Oncology/Hematology* 1:357, CRC Press, Inc., 1984), in Pastan et al. (*Cell* 47:641, 1986), in Vitetta et al. (*Science* 238:1098 -1104, 1987) and in Brady et al. (*Int. J. Rad. Oncol. Biol. Phys.* 13:1535-1544, 1987). Other examples of the use of immunoconjugates for cancer and other forms of therapy have been disclosed, inter alia, in U.S. Pat. Nos. 4,331,647, 4,348,376, 4,361,544, 4,468,457, 4,444,744, 4,460,459, 4,460,561, 4,624,846, 4,818,709, 4,046,722, 4,671,958, 4,046,784, 5,332,567, 5,443,953, 5,541,297, 5,601,825, 5,635,603, 5,637,288, 5,677,427, 5,686,578, 5,698,178, 5,789,554, 5,922,302, 6,187,287, and 6,319,500.

[0141] Additionally, the treatment methods of the invention can be used in combination with other compounds or techniques for preventing, mitigating or reversing the side effects of certain cytotoxic agents. Examples of such combinations include, e.g., administration of IL-1 together with an antibody for rapid clearance, as described in e.g., U.S. Pat. No. 4,624,846. Such administration can be performed from 3 to 72 hours after administration of a primary therapeutic treatment targeting EGFR activity in combination with an anti-cancer agent (e.g., with a radioisotope, drug or toxin as the cytotoxic component). This can be used to enhance clearance of the conjugate, drug or toxin from the circulation and to mitigate or reverse myeloid and other hematopoietic toxicity caused by the therapeutic agent.

[0142] In another aspect of the invention, cancer therapy may involve a combination of more than one tumoricidal agent, e.g., a drug and a radioisotope, or a radioisotope and a Boron-10 agent for neutron-activated therapy, or a drug and a biological response modifier, or a fusion molecule conjugate and a biological response modifier. The cytokine can be integrated into such a therapeutic regimen to maximize the efficacy of each component thereof.

[0143] Similarly, certain antileukemic and antilymphoma antibodies conjugated with radioisotopes that are β or α emitters may induce myeloid and other hematopoietic side effects when these agents are not solely directed to the tumor cells. This is observed particularly when the tumor cells are in the circulation and in the blood-forming organs. Concomitant and/or subsequent administration of at least one hematopoietic cytokine (e.g., growth factors, such as colony stimulating factors, such as G-CSF and GM-CSF) is preferred to reduce or ameliorate the hematopoietic side effects, while augmenting the anticancer effects.

[0144] It is well known in the art that various methods of radionuclide therapy can be used for the treatment of cancer and other pathological conditions, as described, e.g., in Harbert, "Nuclear Medicine Therapy", New York, Thieme Medical Publishers, 1987, pp. 1-340. A clinician experienced in these procedures will readily be able to adapt the cytokine adjuvant therapy described herein to such procedures to mitigate any hematopoietic side effects thereof. Similarly, therapy with cytotoxic drugs, administered with treatment targeting EGFR activity, can be used, e.g., for treatment of cancer or other cell proliferative diseases. Such treatment is governed by analogous principles to radioisotope therapy with isotopes or radiolabeled antibodies. The ordinary skilled clinician will be able to adapt the administration of the additional anti-cancer therapy before, during and/or after the primary anti-cancer therapy.

KITS

[0145] The present invention therefore also provides predictive, diagnostic, and prognostic kits comprising degenerate primers to amplify a target nucleic acid in the kinase domain of the erbB1 gene and instructions comprising amplification protocol and analysis of the results. The kit may alternatively also comprise buffers, enzymes, and containers for performing the amplification and analysis of the amplification products. The kit may also be a component of a screening,

diagnostic or prognostic kit comprising other tools such as DNA microarrays. Preferably, the kit also provides one or more control templates, such as nucleic acids isolated from normal tissue sample, and/or a series of samples representing different variances in the kinase domain of the erbB1 gene.

[0146] In one embodiment, the kit provides two or more primer pairs, each pair capable of amplifying a different region of the erbB1 gene (each region a site of potential variance) thereby providing a kit for analysis of expression of several gene variances in a biological sample in one reaction or several parallel reactions.

[0147] Primers in the kits may be labeled, for example fluorescently labeled, to facilitate detection of the amplification products and consequent analysis of the nucleic acid variances.

[0148] In one embodiment, more than one variance can be detected in one analysis. A combination kit will therefore comprise of primers capable of amplifying different segments of the kinase domain of the erbB1 gene. The primers may be differentially labeled, for example using different fluorescent labels, so as to differentiate between the variances.

[0149] The primers contained within the kit may include the following primers: Exon 19 sense primer, 5'-GCAATAT-CAGCCTTAGGTGCGGCTC-3' (SEQ ID NO: 505); Exon 19 antisense primer, 5'-CATAGAA AGTGAACATTTAGGAT-GTG-3' (SEQ ID NO: 506); Exon 21 sense primer, 5'-CTAACGTTTCG CCAGCCATAAGTCC-3' (SEQ ID NO: 507); and Exon 21 antisense primer, 5'-GCTGCGAGCTCACCCAG AATGTCTGG-3' (SEQ ID NO: 508).

[0150] In a preferred embodiment, the primers are selected from the group consisting of SEQ ID NOS 646-673 (see Tables 5 and 6). These primers have SEQ ID NO 645 on the 5' end of the forward primer and SEQ ID NO 674 on the 5' end of the reverse primers.

Immunodetection Kits

[0151] In further embodiments, the invention provides immunological kits for use in detecting the activation levels of downstream EGFR targets (i.e. STAT3, STAT5, and Akt). Such kits will generally comprise one or more antibodies that have immunospecificity for the phosphorylated form of STAT3, STAT5, or Akt.

[0152] A kit comprising an antibody capable of immunospecifically binding a phosphorylated protein in a mammalian cell selected from the group consisting of phosphorylated Akt, STAT3, and STAT5 proteins and instructions for using the antibody to examine the mammalian cell for Akt, STAT3 or STAT5 pathway activation is provided in the present invention. In preferred methods, the kit comprises different antibodies, each of which is capable of immunospecifically binding phosphorylated proteins in a mammalian cell selected from the group consisting of phosphorylated Akt, STAT3 or STAT5 proteins.

[0153] The kit generally comprises, a) a pharmaceutically acceptable carrier; b) an antibody directed against phosphorylated STAT3, STAT5, or Akt, in a suitable container means; and c) an immunodetection reagent. Antibodies (monoclonal or polyclonal) are commercially available and may also be prepared by methods known to those of skill in the art, for example, in Current Protocols in Immunology, John Wiley & Sons, Edited by: John E. Coligan, Ada M. Kruisbeek, David H. Margulies, Ethan M. Shevach, Warren Strober, 2001.

[0154] In certain embodiments, the antigen or the antibody may be bound to a solid support, such as a column matrix or well of a microtitre plate. The immunodetection reagents of the kit may take any one of a variety of forms, including those detectable labels that are associated with, or linked to, the given antibody or antigen itself. Detectable labels that are associated with or attached to a secondary binding ligand are also contemplated. Exemplary secondary ligands are those secondary antibodies that have binding affinity for the first antibody or antigen.

[0155] Suitable assay labels are known in the art and include enzyme labels, such as, glucose oxidase; radioisotopes, such as iodine (¹³¹I, ¹²⁵I, ¹²³I, ¹²¹I), carbon (¹⁴C), sulfur (³⁵S), tritium (³H), indium (^{115m}In, ^{113m}In, ¹¹²In, ¹¹¹In), and technetium (⁹⁹Tc, ^{99m}Tc), thallium (²⁰¹Tl), gallium (⁶⁸Ga, ⁶⁷Ga), palladium (¹⁰³Pd), molybdenum (⁹⁹Mo), xenon (¹³³Xe), fluorine (¹⁸F), ¹⁵³Sm, ¹⁷⁷Lu, ¹⁵⁹Gd, ¹⁴⁹Pm, ¹⁴⁰La, ¹⁷⁵Yb, ¹⁶⁶Ho, ⁹⁰Y, ⁴⁷Sc, ¹⁸⁶Re, ¹⁸⁸Re, ¹⁴²Pr, ¹⁰⁵Rh, ⁹⁷Ru; luminescent labels, such as luminol; and fluorescent labels, such as fluorescein and rhodamine, and biotin.

[0156] Further suitable immunodetection reagents for use in the present kits include the two-component reagent that comprises a secondary antibody that has binding affinity for the first antibody or antigen, along with a third antibody that has binding affinity for the second antibody, wherein the third antibody is linked to a detectable label.

[0157] A number of exemplary labels are known in the art and all such labels may be employed in connection with the present invention. Radiolabels, nuclear magnetic spin-resonance isotopes, fluorescent labels and enzyme tags capable of generating a colored product upon contact with an appropriate substrate are suitable examples.

[0158] The kits may contain antibody-label conjugates either in fully conjugated form, in the form of intermediates, or as separate moieties to be conjugated by the user of the kit.

[0159] The kits may further comprise a suitably aliquoted composition of an antigen whether labeled or unlabeled, as may be used to prepare a standard curve for a detection assay or as a positive control.

[0160] The kits of the invention, regardless of type, will generally comprise one or more containers into which the biological agents are placed and, preferably, suitably aliquoted. The components of the kits may be packaged either in aqueous media or in lyophilized form.

[0161] The immunodetection kits of the invention may additionally contain one or more of a variety of other cancer marker antibodies or antigens, if so desired. Such kits could thus provide a panel of cancer markers, as may be better used in testing a variety of patients. By way of example, such additional markers could include, other tumor markers such as PSA, SeLe (X), HCG, as well as p53, cyclin D1, p16, tyrosinase, MAGE, BAGE, PAGE, MUC18, CEA, p27, [bgr]HCG or other markers known to those of skill in the art.

[0162] The container means of the kits will generally include at least one vial, test tube, flask, bottle, or even syringe or other container means, into which the antibody or antigen may be placed, and preferably, suitably aliquoted. Where a second or third binding ligand or additional component is provided, the kit will also generally contain a second, third or other additional container into which this ligand or component may be placed.

[0163] The kits of the present invention will also typically include a means for containing the antibody, antigen, and any other reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into which the desired vials are retained.

[0164] The methods of the present invention also encompass the identification of compounds that interfere with the kinase activity of a variant form of the EGFR. The variant EGFR comprises at least one variance in its kinase domain. Such compounds may, for example, be tyrosine kinase inhibitors. Methods for identifying compounds that interfere with the kinase activity of a receptor are generally known to those of skill in the art and are further described in, for example, for example, Dhanabal et al., Cancer Res. 59:189-197 (1999); Xin et al., J. Biol. Chem. 274:9116-9121 (1999); Sheu et al., Anticancer Res. 18:4435-4441; Ausprunk et al., Dev. Biol. 38:237-248 (1974); Gimbrone et al., J. Natl. Cancer Inst. 52:413-427; Nicosia et al., In vitro 18:538-549. In general, compounds are identified, using the methods disclosed herein, that interfere with the enhanced kinase activity characteristic of at least one variance in the kinase domain of the erbB1 gene.

Solid Support

[0165] In another embodiment, the invention provides a kit for practicing the methods of the invention. In one embodiment, a kit for the detection of variances in the kinase domain of erbB1 gene on a solid support is described. The kit can include, e.g. the materials and reagents for detecting a plurality of variances in one assay. The kit can include e.g. a solid support, oligonucleotide primers for a specific set of target polynucleotides, polymerase chain reaction reagents and components, e.g. enzymes for DNA synthesis, labeling materials, and other buffers and reagents for washing. The kit may also include instructions for use of the kit to amplify specific targets on a solid support. Where the kit contains a prepared solid support having a set of primers already fixed on the solid support, e.g. for amplifying a particular set of target polynucleotides, the design and construction of such a prepared solid support is described above. The kit also includes reagents necessary for conducting a PCR on a solid support, for example using an in situ-type or solid phase type PCR procedure where the support is capable of PCR amplification using an in situ-type PCR machine. The PCR reagents, included in the kit, include the usual PCR buffers, a thermostable polymerase (e.g. Taq DNA polymerase), nucleotides (e.g. dNTPs), and other components and labeling molecules (e.g. for direct or indirect labeling as described above). The kits can be assembled to support practice of the PCR amplification method using immobilized primers alone or, alternatively, together with solution phase primers.

[0166] Alternatively, the kit may include a solid support with affixed oligonucleotides specific to any number of EGFR variances, further defined in Figures 4A-4C and Figures 7 and 8. A test biological sample may be applied to the solid support, under selective hybridization conditions, for the determination of the presence or absence of variances in the kinase domain of erbB1.

[0167] The methods of the present invention also encompass the identification of compounds that interfere with the kinase activity of a variant form of the EGFR. The variant EGFR comprises at least one variance in its kinase domain. However, in an alternative embodiment, the variant EGFR comprises a secondary mutation that confers resistance to a first TKI e.g., gefitinib or erlotinib. Such compounds may, for example, be tyrosine kinase inhibitors. Methods for identifying compounds that interfere with the kinase activity of a receptor are generally known to those of skill in the art and are further described in, for example, for example, Dhanabal et al., Cancer Res. 59:189-197 (1999); Xin et al., J. Biol. Chem. 274:9116-9121 (1999); Sheu et al., Anticancer Res. 18:4435-4441; Ausprunk et al., Dev. Biol. 38:237-248 (1974); Gimbrone et al., J. Natl. Cancer Inst. 52:413-427; Nicosia et al., In vitro 18:538-549. In general, compounds are identified, using the methods disclosed herein, that interfere with the enhanced kinase activity characteristic of at least one variance in the kinase domain of the erbB1 gene. Such known variances are described in Figures 4, 7, 8 and Table 2.

[0168] Once identified, such compounds are administered to patients in need of EGFR targeted treatment, for example, patients affected with or at risk for developing cancer.

[0169] The route of administration may be intravenous (I.V.), intramuscular (I.M.), subcutaneous (S.C.), intradermal (I.D.), intraperitoneal (I.P.), intrathecal (I.T.), intrapleural, intrauterine, rectal, vaginal, topical, intratumor and the like. The compounds of the invention can be administered parenterally by injection or by gradual infusion over time and can be delivered by peristaltic means.

[0170] Administration may be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays, for example, or using suppositories. For oral administration, the compounds of the invention are formulated into conventional oral administration forms such as capsules, tablets and tonics.

[0171] For topical administration, the pharmaceutical composition (inhibitor of kinase activity) is formulated into ointments, salves, gels, or creams, as is generally known in the art.

[0172] The therapeutic compositions of this invention are conventionally administered intravenously, as by injection of a unit dose, for example. The term "unit dose" when used in reference to a therapeutic composition of the present invention refers to physically discrete units suitable as unitary dosage for the subject, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required diluent; i.e., carrier, or vehicle.

[0173] The compositions are administered in a manner compatible with the dosage formulation, and in a therapeutically effective amount. The quantity to be administered and timing depends on the subject to be treated, capacity of the subject's system to utilize the active ingredient, and degree of therapeutic effect desired. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner and are peculiar to each individual.

[0174] The tyrosine kinase inhibitors useful for practicing the methods of the present invention are described herein. Any formulation or drug delivery system containing the active ingredients, which is suitable for the intended use, as are generally known to those of skill in the art, can be used. Suitable pharmaceutically acceptable carriers for oral, rectal, topical or parenteral (including inhaled, subcutaneous, intraperitoneal, intramuscular and intravenous) administration are known to those of skill in the art. The carrier must be pharmaceutically acceptable in the sense of being compatible with the other ingredients of the formulation and not deleterious to the recipient thereof.

[0175] As used herein, the terms "pharmaceutically acceptable", "physiologically tolerable" and grammatical variations thereof, as they refer to compositions, carriers, diluents and reagents, are used interchangeably and represent that the materials are capable of administration to or upon a mammal without the production of undesirable physiological effects.

[0176] Formulations suitable for parenteral administration conveniently include sterile aqueous preparation of the active compound which is preferably isotonic with the blood of the recipient. Thus, such formulations may conveniently contain distilled water, 5% dextrose in distilled water or saline. Useful formulations also include concentrated solutions or solids containing the compound which upon dilution with an appropriate solvent give a solution suitable for parental administration above.

[0177] For enteral administration, a compound can be incorporated into an inert carrier in discrete units such as capsules, cachets, tablets or lozenges, each containing a predetermined amount of the active compound; as a powder or granules; or a suspension or solution in an aqueous liquid or non-aqueous liquid, e.g., a syrup, an elixir, an emulsion or a draught. Suitable carriers may be starches or sugars and include lubricants, flavorings, binders, and other materials of the same nature.

[0178] A tablet may be made by compression or molding, optionally with one or more accessory ingredients. Compressed tablets may be prepared by compressing in a suitable machine the active compound in a free-flowing form, e.g., a powder or granules, optionally mixed with accessory ingredients, e.g., binders, lubricants, inert diluents, surface active or dispersing agents. Molded tablets may be made by molding in a suitable machine, a mixture of the powdered active compound with any suitable carrier.

[0179] A syrup or suspension may be made by adding the active compound to a concentrated, aqueous solution of a sugar, e.g., sucrose, to which may also be added any accessory ingredients. Such accessory ingredients may include flavoring, an agent to retard crystallization of the sugar or an agent to increase the solubility of any other ingredient, e.g., as a polyhydric alcohol, for example, glycerol or sorbitol.

[0180] Formulations for rectal administration may be presented as a suppository with a conventional carrier, e.g., cocoa butter or Witepsol S55 (trademark of Dynamite Nobel Chemical, Germany), for a suppository base.

[0181] Formulations for oral administration may be presented with an enhancer. Orally-acceptable absorption enhancers include surfactants such as sodium lauryl sulfate, palmitoyl carnitine, Laureth-9, phosphatidylcholine, cyclodextrin and derivatives thereof; bile salts such as sodium deoxycholate, sodium taurocholate, sodium glycocholate, and sodium fusidate; chelating agents including EDTA, citric acid and salicylates; and fatty acids (e.g., oleic acid, lauric acid, acyl-carnitines, mono- and diglycerides). Other oral absorption enhancers include benzalkonium chloride, benzethonium chloride, CHAPS (3-(3-cholamidopropyl)-dimethylammonio-1-propanesulfonate), Big-CHAPS (N, N-bis(3-D-gluconamidopropyl)-cholamide), chlorobutanol, octoxynol-9, benzyl alcohol, phenols, cresols, and alkyl alcohols. An especially preferred oral absorption enhancer for the present invention is sodium lauryl sulfate.

[0182] Alternatively, the compound may be administered in liposomes or microspheres (or microparticles). Methods for preparing liposomes and microspheres for administration to a patient are well known to those of skill in the art. U.S. Pat. No. 4,789,734, describes methods for encapsulating biological materials in liposomes. Essentially, the material is

dissolved in an aqueous solution, the appropriate phospholipids and lipids added, along with surfactants if required, and the material dialyzed or sonicated, as necessary. A review of known methods is provided by G. Gregoriadis, Chapter 14, "Liposomes," *Drug Carriers in Biology and Medicine*, pp. 287-341 (Academic Press, 1979).

[0183] Microspheres formed of polymers or proteins are well known to those skilled in the art, and can be tailored for passage through the gastrointestinal tract directly into the blood stream. Alternatively, the compound can be incorporated and the microspheres, or composite of microspheres, implanted for slow release over a period of time ranging from days to months. See, for example, U.S. Pat. Nos. 4,906,474, 4,925,673 and 3,625,214, and Jein, *TIPS* 19:155-157 (1998).

[0184] In one embodiment, the tyrosine kinase inhibitor of the present invention can be formulated into a liposome or microparticle which is suitably sized to lodge in capillary beds following intravenous administration. When the liposome or microparticle is lodged in the capillary beds surrounding ischemic tissue, the agents can be administered locally to the site at which they can be most effective. Suitable liposomes for targeting ischemic tissue are generally less than about 200 nanometers and are also typically unilamellar vesicles, as disclosed, for example, in U.S. Pat. No. 5,593,688 to Baldeschweiler, entitled "Liposomal targeting of ischemic tissue,".

[0185] Preferred microparticles are those prepared from biodegradable polymers, such as polyglycolide, polylactide and copolymers thereof. Those of skill in the art can readily determine an appropriate carrier system depending on various factors, including the desired rate of drug release and the desired dosage.

[0186] In one embodiment, the formulations are administered via catheter directly to the inside of blood vessels. The administration can occur, for example, through holes in the catheter. In those embodiments wherein the active compounds have a relatively long half life (on the order of 1 day to a week or more), the formulations can be included in biodegradable polymeric hydrogels, such as those disclosed in U.S. Pat. No. 5,410,016 to Hubbell et al. These polymeric hydrogels can be delivered to the inside of a tissue lumen and the active compounds released over time as the polymer degrades. If desirable, the polymeric hydrogels can have microparticles or liposomes which include the active compound dispersed therein, providing another mechanism for the controlled release of the active compounds.

[0187] The formulations may conveniently be presented in unit dosage form and may be prepared by any of the methods well known in the art of pharmacy. All methods include the step of bringing the active compound into association with a carrier which constitutes one or more accessory ingredients. In general, the formulations are prepared by uniformly and intimately bringing the active compound into association with a liquid carrier or a finely divided solid carrier and then, if necessary, shaping the product into desired unit dosage form.

[0188] The formulations may further include one or more optional accessory ingredient(s) utilized in the art of pharmaceutical formulations, e.g., diluents, buffers, flavoring agents, binders, surface active agents, thickeners, lubricants, suspending agents, preservatives (including antioxidants) and the like.

[0189] Compounds of the present methods may be presented for administration to the respiratory tract as a snuff or an aerosol or solution for a nebulizer, or as a microfine powder for insufflation, alone or in combination with an inert carrier such as lactose. In such a case the particles of active compound suitably have diameters of less than 50 microns, preferably less than 10 microns, more preferably between 2 and 5 microns.

[0190] Generally for nasal administration a mildly acid pH will be preferred. Preferably the compositions of the invention have a pH of from about 3 to 5, more preferably from about 3.5 to about 3.9 and most preferably 3.7. Adjustment of the pH is achieved by addition of an appropriate acid, such as hydrochloric acid.

[0191] The preparation of a pharmacological composition that contains active ingredients dissolved or dispersed therein is well understood in the art and need not be limited based on formulation. Typically such compositions are prepared as injectables either as liquid solutions or suspensions, however, solid forms suitable for solution, or suspensions, in liquid prior to use can also be prepared. The preparation can also be emulsified.

[0192] The active ingredient can be mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient and in amounts suitable for use in the therapeutic methods described herein. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol or the like and combinations thereof. In addition, if desired, the composition can contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents and the like which enhance the effectiveness of the active ingredient.

[0193] The kinase inhibitor of the present invention can include pharmaceutically acceptable salts of the components therein. Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the polypeptide) that are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, tartaric, mandelic and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine and the like.

[0194] Physiologically tolerable carriers are well known in the art. Exemplary of liquid carriers are sterile aqueous solutions that contain no materials in addition to the active ingredients and water, or contain a buffer such as sodium phosphate at physiological pH value, physiological saline or both, such as phosphate-buffered saline. Still further, aqueous carriers can contain more than one buffer salt, as well as salts such as sodium and potassium chlorides, dextrose, polyethylene glycol and other solutes.

[0195] Liquid compositions can also contain liquid phases in addition to and to the exclusion of water. Exemplary of such additional liquid phases are glycerin, vegetable oils such as cottonseed oil, and water-oil emulsions.

Predicting Mutations

[0196] In another embodiment, the present invention discloses a method to predict variances in the erbB1 gene following treatment with a tyrosine kinase inhibitor. It is generally known that response to cancer treatment with a tyrosine kinase inhibitor is often followed by resistance to that or other similar compounds. Such resistance is thought to arise through the acquisition of mutations in the drug target, for example in the EGFR. The ability to predict (and select) such mutations will allow for better treatment options and fewer relapses.

[0197] In one embodiment of the present invention, DNA encoding the EGFR kinase domain is isolated and sequenced from a tumor sample of cancer patients that have responded to gefitinib (or a similar EGFR targeting treatment) but have subsequently relapsed. The relapse in such patients is expected to involve the acquisition of secondary mutations within the EGFR kinase domain. Compounds that target, and inhibit the kinase activity of, these newly defined mutations are then identified using methods disclosed herein. Such compounds may be used alone, or in combination with other known EGFR targeting treatments, to treat cancer patients with primary or secondary (as above) mutations in the kinase domain of EGFR.

[0198] In one embodiment, predicting variances in the kinase (catalytic) domain of the EGFR (erbB1 gene) is done *in vitro*. In this method, cells, e.g. fibroblast cells, are stably transfected with cDNAs containing kinase domain mutations that have been identified in human cancer cell lines. For example, the cells may be transfected with an EGFR that bears a mutation such as SEQ ID NO:495, further described in Figure 4A, or with any number of identified or as yet unidentified kinase domain-mutated EGFRs. The transfection of kinase domain-mutated EGFRs into cells will result in aberrant proliferation of the cells in culture. Methods of stable transfection are known to those of skill in the art and are further defined in Current Protocols in Molecular Biology by F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, K. Struhl and V. B. Chanda (Editors), John Wiley & Sons., 2004. The transfected cells are then given an effective, yet sub-lethal, dose of a drug, preferably a tyrosine kinase inhibitor, predicted to inhibit cellular proliferation. In a preferred embodiment, the drug is an anilinoquinazoline, synthetic anilinoquinazoline, gefitinib or erlotinib. The cells are serially passaged in the presence of drug and subclones that survive are selected. Over many generations, cells that survive (i.e. are resistant to the compound), are selected and analyzed for variances in the erbB1 gene. Secondary variances can thus be predicted to occur following repeated treatment with a tyrosine kinase inhibitor *in vivo*.

[0199] Alternatively, cells are transfected with gefitinib-resistant mutant cDNA derived from human NSCLC cell lines, for example, NCI-1650 and NCI-1975. Each cell line has a heterozygous mutation with the kinase domain of EGFR, and is, therefore, expected to be sensitive to gefitinib. The EGFR mutation in NCI-1650 consists of an in-frame deletion of 15 nucleotides at position 2235-2249 (delE746-A750) within exon 19, while NCI-1975 has a missense mutation within exon 21 that substitutes a G for T at nucleotide 2573 (L858R). As shown herein, the L858R mutation in NCI-H1975 is activating and confers increased sensitivity to gefitinib *in vitro*. Other cancer cell lines that harbor EGFR kinase domain mutations may be utilized. The cancer cell lines may include lung cancer as well as other cancers that are found to harbor such mutations.

[0200] The cells may be treated with a mutagen in order to increase the frequency with which cells acquire secondary mutations. A mutagen may induce mutations at different frequencies depending upon the dosage regimen, mode of delivery, and the developmental stage of the organism or cell upon mutagen administration, all parameters of which are disclosed in the prior art for different mutagens or mutagenesis techniques. The mutagen may be an alkylating agent, such as ethyl methanesulfonate (EMS), N-ethyl-N-nitrosourea (ENU) or N-methyl-N-nitrosourea (MNU). Alternatively, the mutagen may be, for example, phocarbazine hydrochloride (Prc), methyl methanesulfonate (MeMS), chlorambucil (Chl), melphalan, porcarbazine hydrochloride, cyclophosphamide (Cp), diethyl sulfate (Et₂SO₄), acrylamide monomer (AA), triethylene melamin (TEM), nitrogen mustard, vincristine, dimethylnitrosamine, N-methyl-N'-nitro-Nitrosoguanidine (MNNG), 7,12 dimethylbenz(a)anthracene (DMBA), ethylene oxide, hexamethylphosphoramide, bisulfan, and ethyl methanesulfonate (EtMs). Methods of treating cells with mutagens is described, for example, in U.S. 6,015,670. Following mutagenesis, cells (i.e. transfected with variant EGFR or human cancer cell line derived) can be cultured in gefitinib-supplemented medium to select for the outgrowth of resistant clones. Subcultivation of individual clones can be followed, for example, by nucleotide sequence determination of the EGFR gene following specific PCR-mediated amplification of genomic DNA corresponding to the EGFR kinase domain.

[0201] In another embodiment, cells (with an EGFR variance) are serially passaged in the presence of gradually increasing concentrations of gefitinib (or a similar tyrosine kinase inhibitor) over a course of several weeks or months in order to select for the spontaneous acquisition of mutations within the EGFR gene that confer resistance to gefitinib. Selected cells (that continue to proliferate at relatively high gefitinib concentration) can be isolated as colonies, and mutations will be identified as described above. Such variances can thus be predicted to occur following repeated treatment with a tyrosine kinase inhibitor *in vivo*. See, for example, Scappini et al., Cancer, April 1, 2004, Vol. 100, pg. 1459.

[0202] In yet another embodiment, a variant form of the EGFR gene can be propagated in a DNA repair-deficient bacterial strain before re-introducing it into stably selected cell lines. Replication in such bacteria will enhance the frequency of mutagenesis. Alternatively, "error-prone" PCR can be utilized to enhance the frequency of mutations in the cloned EGFR DNA in vitro, using standard methods, known to those of skill in the art.

[0203] In another embodiment, predicting variances in the kinase domain of the erbB1 gene is done *in vivo*. For example, a kinase activity increasing variant form of the erbB1 gene is transfected into an animal, i.e. a mouse, generating a cancer model. The animal is then treated with an effective dose of a compound, preferably an anilinoquinazoline, synthetic anilinoquinazoline, gefitinib or erlotinib. Upon repeated exposure to the compound, the cancer is initially inhibited. As in humans treated with such compounds, tumor cells in the animal acquire mutations which make them resistant to such treatment. The methods of the present invention allow for the isolation and characterization of the erbB1 gene in such resistant tumors. Compounds that specifically target these newly characterized variances are useful in the treatment of patients suspected of carrying such a mutated erbB1 gene. Such patients include, for example, patients who initially respond to therapy with a tyrosine kinase inhibitor, but subsequently fail to respond to the same or similar compound.

[0204] Methods of creating an animal model are known to those of skill in the art and are further defined in e.g., Ohashi et al., Cell, 65:305-317 (1991); Adams et al., Nature, 325:223-228 (1987); and Roman et al., Cell, 61:383-396 (1990). In the case of fertilized oocytes, the preferred method of transgene introduction is by microinjection, see, e.g., Leder et al., U.S. Pat. Nos. 4,736,866 and 5,175,383 whereas in the case of embryonic stem (ES) cells, the preferred method is electroporation. However, other methods including viral delivery systems such as retroviral infection, or liposomal fusion can be used. The isolation and characterization of nucleic acid is described above and in the examples.

[0205] The above-identified kinase activity increasing variances in the erbB1 gene may be screened for in patients (diagnostically or prognostically), using the methods of the present invention. The presence or absence of such mutations may then be used as a criteria for determining ones sensitivity to treatment with an EGFR targeting compound, such as, for example, a tyrosine kinase inhibitor.

[0206] Compounds that specifically target these newly defined variances, whether detected in vivo or in vitro, can be selected using techniques known in the art and discussed herein. Candidate drug screening assays may be used to identify bioactive candidate agents that inhibit the activity of variant forms of EGFR. Of particular interest are screening assays for agents that have a low toxicity for human cells. A wide variety of assays may be used for this purpose, including labeled in vitro protein-protein binding assays, electrophoretic mobility shift assays, enzyme activity assays, immunoassays for protein binding, and the like. The purified mutant EGFR protein may also be used for determination of three-dimensional crystal structure, which can be used for modeling intermolecular interactions, transporter function, etc. Such compounds may be, for example, tyrosine kinase inhibitors, antibodies, aptamers, siRNAs, and vectors that inhibit the kinase activity of EGFR.

[0207] In another embodiment, compounds useful in the method of the present invention are antibodies which interfere with kinase signaling via the mutant EGFR, including monoclonal, chimeric humanized, and recombinant antibodies and fragment thereof which are characterized by their ability to inhibit the kinase activity of the EGFR and which have low toxicity.

[0208] Neutralizing antibodies are readily raised in animals such as rabbits or mice by immunization with an EGFR with at least one nucleic acid variance in its kinase domain. Immunized mice are particularly useful for providing sources of B cells for the manufacture of hybridomas, which in turn are cultured to produce large quantities of anti-EGFR monoclonal antibodies. Chimeric antibodies are immunoglobulin molecules characterized by two or more segments or portions derived from different animal species. Generally, the variable region of the chimeric antibody is derived from a non-human mammalian antibody, such as murine monoclonal antibody, and the immunoglobulin constant region is derived from a human immunoglobulin molecule. Preferably, both regions and the combination have low immunogenicity as routinely determined. Humanized antibodies are immunoglobulin molecules created by genetic engineering techniques in which the murine constant regions are replaced with human counterparts while retaining the murine antigen binding regions. The resulting mouse-human chimeric antibody should have reduced immunogenicity and improved pharmacokinetics in humans. Preferred examples of high affinity monoclonal antibodies and chimeric derivatives thereof, useful in the methods of the present invention, are described in the European Patent Application EP 186,833; PCT Patent Application WO 92/16553; and US Patent No. 6,090,923.

[0209] Existing or newly identified compounds as described above are useful in the treatment of patients carrying primary and/or secondary EGFR mutations.

[0210] In a preferred embodiment, the compound is an inhibitor of the tyrosine kinase activity of an EGFR with at least one variance in its kinase domain, particularly small molecule inhibitors having selective action on "mutated" EGFRs as compared to other tyrosine kinases. Inhibitors of EGFR include, but are not limited to, tyrosine kinase inhibitors such as quinazolines, such as PID 153035, 4-(3-chloroanilino) quinazoline, or CP- 358,774, pyridopyrimidines, pyrimidopyrimidines, pyrrolopyrimidines, such as CGP 59326, CGP 60261 and CGP 62706, and pyrazolopyrimidines, 4-(phenylamino)-7H-pyrrolo[2,3-d]pyrimidines (Traxler et al., (1996) J. Med Chem 39:2285-2292), curcumin (diferuloyl methane) (Laxmin

arayana, et al., (1995), *Carcinogen* 16:1741-1745), 4,5-bis (4-fluoroanilino) phthalimide (Buchdunger et al. (1995) *Clin. Cancer Res.* 1:813-821; Dinney et al. (1997) *Clin. Cancer Res.* 3:161-168); tyrphostins containing nitrothiophene moieties (Brunton et al. (1996) *Anti Cancer Drug Design* 11:265-295); the protein kinase inhibitor ZD-1 839 (AstraZeneca); CP-358774 (Pfizer, Inc.); PD-01 83805 (Warner-Lambert), EKB-569 (Torrance et al., *Nature Medicine*, Vol. 6, No. 9, Sept. 2000, p. 1024), HKI-272 and HKI-357 (Wyeth); or as described in International patent application WO99/09016 (American Cyanamid); WO98/43960 (American Cyanamid); WO97/38983 (Warener Labert); WO99/06378 (Warner Lambert); WO99/06396 (Warner Lambert) ; WO96/30347 (Pfizer, Inc.); WO96/33978 (Zeneca); WO96/33977 (Zeneca); and WO96/33980) Zeneca.

[0211] In another embodiment, an antisense strategy may be used to interfere with the kinase activity of a variant EGFR. This approach may, for instance, utilize antisense nucleic acids or ribozymes that block translation of a specific mRNA, either by masking that mRNA with an antisense nucleic acid or cleaving it with a ribozyme. For a general discussion of antisense technology, see, e.g., *Antisense DNA and RNA*, (Cold Spring Harbor Laboratory, D. Melton, ed., 1988).

[0212] Reversible short inhibition of variant EGFR gene transcription may also be useful. Such inhibition can be achieved by use of siRNAs. RNA interference (RNAi) technology prevents the expression of genes by using small RNA molecules such as small interfering RNAs (siRNAs). This technology in turn takes advantage of the fact that RNAi is a natural biological mechanism for silencing genes in most cells of many living organisms, from plants to insects to mammals (McManus et al., *Nature Reviews Genetics*, 2002, 3(10) p. 737). RNAi prevents a gene from producing a functional protein by ensuring that the molecule intermediate, the messenger RNA copy of the gene is destroyed. siRNAs can be used in a naked form and incorporated in a vector, as described below. One can further make use of aptamers to specifically inhibit variant EGFR gene transcription, see, for example, U.S. Patent 6,699,843. Aptamers useful in the present invention may be identified using the SELEX process. The methods of SELEX have been described in, for example, U.S. Patent Nos. 5,707,796, 5,763,177, 6,011,577, 5,580,737, 5,567,588, and 5,660,985.

[0213] An "antisense nucleic acid" or "antisense oligonucleotide" is a single stranded nucleic acid molecule, which, on hybridizing under cytoplasmic conditions with complementary bases in a RNA or DNA molecule, inhibits the latter's role. If the RNA is a messenger RNA transcript, the antisense nucleic acid is a countertranscript or mRNA-interfering complementary nucleic acid. As presently used, "antisense" broadly includes RNA-RNA interactions, RNA- DNA interactions, ribozymes, RNAi, aptamers and Rnase-H mediated arrest.

[0214] Ribozymes are RNA molecules possessing the ability to specifically cleave other single stranded RNA molecules in a manner somewhat analogous to DNA restriction endonucleases. Ribozymes were discovered from the observation that certain mRNAs have the ability to excise their own introns. By modifying the nucleotide sequence of these ribozymes, researchers have been able to engineer molecules that recognize specific nucleotide sequences in an RNA molecule and cleave it (Cech, 1989, *Science* 245(4915) p. 276). Because they are sequence-specific, only mRNAs with particular sequences are inactivated.

[0215] Antisense nucleic acid molecules can be encoded by a recombinant gene for expression in a cell (e.g., U.S. patent No 5,814,500; U.S. 5,811, 234), or alternatively they can be prepared synthetically (e.g., U.S. patent No 5,780,607).

[0216] The present invention further provides methods of treating patients with cancer. In particular, patients with at least one nucleic acid variance in the kinase domain of EGFR. The treatment method comprises administering an siRNA-containing composition to a patient within an appropriate time window. The siRNAs may be chemically synthesized, produced using in vitro transcription, etc. In addition, the siRNA molecule can be customized to individual patients in such a way as to correspond precisely to the mutation identified in their tumor. Since siRNA can discriminate between nucleotide sequences that differ by only a single nucleotide, it is possible to design siRNAs that uniquely target a mutant form of the EGFR gene that is associated with either a single nucleotide substitution or a small deletion of several nucleotides-both of which have been identified in tumors as described herein. SiRNAs have been described in Brummelkamp et al., *Science* 296: 550-553, 2002, Jaque et al., *Nature* 418: 435-438, 2002, Elbashir S. M. et al. (2001) *Nature*, 411: 494-498, McCaffrey et al. (2002), *Nature*, 418: 38-39; Xia H. et al. (2002), *Nat. Biotech.* 20: 1006-1010, Novina et al. (2002), *Nat. Med.* 8: 681-686, and U.S. Application No. 20030198627.

[0217] An important advantage of such a therapeutic strategy relative to the use of drugs such as gefitinib, which inhibit both the mutated receptor and the normal receptor, is that siRNA directed specifically against the mutated EGFR should not inhibit the wildtype EGFR. This is significant because it is generally believed that the "side effects" of gefitinib treatment, which include diarrhea and dermatitis, are a consequence of inhibition of EGFR in normal tissues that require its function.

[0218] The delivery of siRNA to tumors can potentially be achieved via any of several gene delivery "vehicles" that are currently available. These include viral vectors, such as adenovirus, lentivirus, herpes simplex virus, vaccinia virus, and retrovirus, as well as chemical-mediated gene delivery systems (for example, liposomes), or mechanical DNA delivery systems (DNA guns). The oligonucleotides to be expressed for such siRNA-mediated inhibition of gene expression would be between 18 and 28 nucleotides in length.

[0219] In another embodiment, the compounds are antisense molecules specific for human sequences coding for an

EGFR having at least one variance in its kinase domain. The administered therapeutic agent may be an antisense oligonucleotides, particularly synthetic oligonucleotides; having chemical modifications from native nucleic acids, or nucleic acid constructs that express such anti-sense molecules as RNA. The antisense sequence is complementary to the mRNA of the targeted EGFR genes, and inhibits expression of the targeted gene products (see e.g. Nyce et al. (1997) *Nature* 385:720). Antisense molecules inhibit gene expression by reducing the amount of mRNA available for translation, through activation of RNase H or steric hindrance. One or a combination of antisense molecules may be administered, where a combination may comprise multiple different sequences from a single targeted gene, or sequences that complement several different genes.

[0220] A preferred target gene is an EGFR with at least one nucleic acid variance in its kinase domain. The gene sequence is incorporated herein, such as, for example, in Figure 5. Generally, the antisense sequence will have the same species of origin as the animal host.

[0221] Antisense molecules may be produced by expression of all or a part of the target gene sequence in an appropriate vector, where the vector is introduced and expressed in the targeted cells. The transcriptional initiation will be oriented such that the antisense strand is produced as an RNA molecule.

[0222] The anti-sense RNA hybridizes with the endogenous sense strand mRNA, thereby blocking expression of the targeted gene. The native transcriptional initiation region, or an exogenous transcriptional initiation region may be employed. The promoter may be introduced by recombinant methods in vitro, or as the result of homologous integration of the sequence into a chromosome. Many strong promoters that are active in muscle cells are known in the art, including the O-actin promoter, SV40 early and late promoters, human cytomegalovirus promoter, retroviral LTRs, etc. Transcription vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences. Transcription cassettes may be prepared comprising a transcription initiation region, the target gene or fragment thereof, and a transcriptional termination region. The transcription cassettes may be introduced into a variety of vectors, e.g. plasmid; retrovirus, e.g. lentivirus; adenovirus; and the like, where the vectors are able to transiently or stably be maintained in cells, usually for a period of at least about one day, more usually for a period of at least about several days.

[0223] Aptamers are also useful. Aptamers are a promising new class of therapeutic oligonucleotides or peptides and are selected in vitro to specifically bind to a given target with high affinity, such as for example ligand receptors. Their binding characteristics are likely a reflection of the ability of oligonucleotides to form three dimensional structures held together by intramolecular nucleobase pairing. Aptamers are synthetic DNA, RNA or peptide sequences which may be normal and modified (e.g. peptide nucleic acid (PNA), thiophosphorylated DNA, etc) that interact with a target protein, ligand (lipid, carbohydrate, metabolite, etc). In a further embodiment, RNA aptamers specific for a variant EGFR can be introduced into or expressed in a cell as a therapeutic.

[0224] Peptide nucleic acids (PNAs) are compounds that in certain respects are similar to oligonucleotides and their analogs and thus may mimic DNA and RNA. In PNA, the deoxyribose backbone of oligonucleotides has been replaced by a pseudo-peptide backbone (Nielsen et al. 1991 *Science* 254, 1457-1500). Each subunit, or monomer, has a naturally occurring or non-naturally occurring nucleobase attached to this backbone. One such backbone is constructed of repeating units of N-(2-aminoethyl) glycine linked through amide bonds. PNA hybridises with complementary nucleic acids through Watson and Crick base pairing and helix formation. The Pseudo-peptide backbone provides superior hybridization properties (Egholm et al. *Nature* (1993) 365, 566-568), resistance to enzymatic degradation (Demidov et al. *Biochem. Pharmacol.* (1994) 48, 1310-1313) and access to a variety of chemical modifications (Nielsen and Haaima *Chemical Society Reviews* (1997) 73-78). PNAs specific for a variant EGFR can be introduced into or expressed in a cell as a therapeutic. PNAs have been described, for example, in U.S. Application No. 20040063906.

[0225] Patients to be treated with a compound which targets a variant EGFR include, for example, patients diagnosed with a primary or secondary mutation in their EGFR, patients who initially respond to therapy with a tyrosine kinase inhibitor, but subsequently fail to respond to the same or similar compound. Alternatively, compounds that target secondary EGFR mutations may be given to cancer patients in combination with compounds that target primary EGFR mutations, for example, gefitinib, as a combination therapy. By combining compounds that target both primary and secondary EGFR mutations, the likelihood of resistance will be reduced.

[0226] Additional EGFR mutations that confer resistance to currently known anti-cancer therapeutics, including but not limited to EGFR tyrosine kinase inhibitors gefitinib, erlotinib and the like, are within the scope of the invention. Resistant EGFR mutants are predicted to have mutants analogous to mutants identified in kinase domains of related tyrosine kinase domain containing proteins that have high homology in this kinase region. Papers describing mutations in analogous proteins include those known in the art for BCR-ABL. See, e.g., Bradford et al. *Blood*. 2003 Jul 1;102(1):276-83, Epub 2003 Mar 06; Hochhaus et al., *Leukemia*. 2002 Nov;16(11):2190-6; and Al-Ali et al., *Hematol J*. 2004;5(1):55-60.

[0227] A mutant EGFR resistant to known EGFR tyrosine kinase inhibitors includes any one or more EGFR polypeptides, or a nucleotide encoding the same, with a non-wild type residue at one or more positions analogous to c-abl (BCR-ABL) residues that confirm an imatinib resistant phenotype. The residues that when mutated in EGFR confer drug

resistance include especially those residues from the kinase domain, including but not limited to, e.g., the P-loop and the activation loop, wherein the mutated residues in the EGFR polypeptide are analogous to c-abl residues. Contemplated resistant EGFR mutants have non-wild type residues at the amino acids positions that are analogous to at least positions Met 244, Leu 248, Gly 250, Gln 252, Tyr 253, Glu 255, Asp 276, Thr 315, Phe 317, Met 351, Glu 355, Phe 359, His 396, Ser 417, and Phe 486 of BCR-ABL, see, for example Table S3C and FIG. 9. These BCL-ABL residues correspond to residues Lys 714, Leu 718, Ser 720, Ala 722, Phe 723, Thr 725, Ala 750, Thr 790, Leu 792, Met 825, Glu 829, Leu 833, His 870, Thr 892, Phe 961, respectively, in EGFR. See, e.g., Table S3C, FIG. 9.

Prognostic Testing

[0228] The methods of the present invention are used as a prognostic indicator of the development of cancer. Alternatively, the methods are used to detect cancer that is present but has not yet been diagnosed or is at a stage that is undetectable. Patients at risk for developing cancer are screened, using the methods of the present invention, for the presence of kinase activity increasing nucleic acid variation in the *erbB1* gene. The presence of a variance or variances in the kinase domain of the *erbB1* gene indicate the presence or imminent presence of cancer. Thus, the presence of variances in the kinase domain of the *erbB1* gene suggest that a patient would benefit from an EGFR targeted treatment. As described herein, an EGFR targeted treatment is preferably treatment with a tyrosine kinase inhibitor.

[0229] In a preferred embodiment of the present invention, a patient is screened for the presence or absence of nucleic acid variances in the kinase domain of the *erbB1* gene by obtaining a biological sample. The sample may be any sample from the patient including tissue, e.g., from the tongue, mouth, cheek, trachea, bronchial tube, lungs, etc. or fluid, e.g., from sputum or lung aspirates. Methods of obtaining these biological specimens are well known to those of skill in the art.

[0230] Thus, the invention provides a method for identifying a disease or disorder associated with aberrant mutant EGFR expression or activity in which a test sample is obtained from a subject and mutant EGFR protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of mutant EGFR protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant mutant EGFR expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue, especially a tissue biopsy sample.

[0231] Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant mutant EGFR expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant mutant EGFR expression or activity in which a test sample is obtained and mutant EGFR protein or nucleic acid is detected (e.g., wherein the presence of mutant EGFR protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with mutant EGFR expression or activity).

EXAMPLES

Example 1

Nucleotide Sequence Analysis of Tumor Specimens

[0232] Tumor specimens from initial diagnostic or surgical procedures were collected from patients with NSCLC who were subsequently treated with Gefitinib, under an IRB-approved protocol. Frozen tumor specimens, along with matched normal tissue, were available for four cases, and paraffin-embedded material was used for the remaining specimens. In addition, 25 unselected cases of primary NSCLC (15 bronchioalveolar, 7 adenocarcinoma, and 3 large cell lung cancers), with matched normal tissues, were obtained from the Massachusetts General Hospital tumor bank. For mutational analysis of the entire *EGFR* coding sequence, DNA was extracted from specimens, followed by amplification of all 28 exons, automated sequencing of uncloned PCR fragments, and analysis of electropherograms in both sense and antisense direction for the presence of heterozygous mutations. All sequence variants were confirmed by multiple independent PCR amplifications. Primer sequences and amplification conditions are provided in Supplementary Material. *EGFR* mutations in exons 19 and 21 were also sought in primary tumors of the breast (15 cases), colon (20 cases), kidney (16 cases), and brain (4 cases), along with a panel of 78 cancer-derived cell lines representing diverse histologies (listed below).

Functional Analysis of Mutant EGFR Constructs

[0233] The L858R and delL747-P753insS mutations were introduced into the full length *EGFR* coding sequence using site-directed mutagenesis and inserted into a cytomegalovirus-driven expression construct (pUSE, Upstate). Cos-7 cells were transfected (Lipofectamine 2000, Invitrogen) using 1 µg of the expression constructs, followed after 18 hrs by replating at 5x10⁴ cells/ well (12-well plates, Costar) in DMEM lacking fetal calf serum. After 16 hrs of serum starvation, cells were stimulated with 10 ng/ml of EGF (SIGMA). To demonstrate Gefitinib inhibition, the drug was added to the culture medium 3 hrs prior to the addition of EGF (30 min stimulation with 100 ng/ml of EGF). Cell lysates were prepared in 100 µL of Laemmli lysis buffer, followed by resolution of proteins on 10% SDS-PAGE, transfer to PVDF membranes, and Western blot analysis using enhanced chemiluminescence reagent (Amersham). Autophosphorylation of EGFR was measured using antibody to phosphotyrosine Y-1068, and comparable protein expression was shown using anti-EGFR antibody (working concentration of 1:1000; Cell Signaling Technology).

MUTATIONAL ANALYSIS

[0234] The polymerase chain reaction was used to amplify the 28 exons comprising the *EGFR* gene using DNA isolated from primary tumor tissue or tumor-derived cell-lines. Primer pairs used were: Exon 1, CAGATTTGGCTCGACCTGGACATAG (sense) (SEQ ID NO: 513) and CAGCTGATCTCAAGGAAACAGG (antisense) (SEQ ID NO: 514); Exon 2, GTATTATCAGTCAC TAAAGCTCAC (sense) (SEQ ID NO: 515) and CACACTTCAAGTGGGAATTCTGC (SEQ ID NO: 516); Exon 3, CTCGTG TGCATTAGGGTTCAACTGG (sense) (SEQ ID NO: 517) and CCTTCTCCGAGGTGGAATTGAGTGAC (antisense) (SEQ ID NO: 518); Exon 4, GCTAATTGCGGGACTCTTGTTTCGCAC (sense) (SEQ ID NO: 519) and TACATGC TTTTCTAGTGGTCAG (antisense) (SEQ ID NO: 520); Exon 5, GGTCTCAAGTGATTCTACAAACCAG (sense) (SEQ ID NO: 521) and CCTTCACCTACTGGTTCACATCTG (antisense) (SEQ ID NO: 522); Exon 6, CATGGT TTGACTTAGTTTGAATGTGG (sense) (SEQ ID NO: 523) and GGATACTAAAGATACTTTGTCAC CAGG(antisense) (SEQ ID NO: 524); Exon 7, GAACACTAGGCTGCAAAGACAGTAAC (sense) (SEQ ID NO: 525) and CCAAGCAAG-GCAAACACATCCACC(antisense) (SEQ ID NO: 526); Exon 8, GGAGGATGGAGCC TTTCCATCAC (sense) (SEQ ID NO: 527) and GAAGAGGAAGATGTGTTCTTTGG (antisense) (SEQ ID NO: 528); Exons 9 and 10, GAATGAAGGATGATGTGGCAGTGG (sense) (SEQ ID NO: 529) and CAAAACATCAGCC ATTAACGG (antisense) (SEQ ID NO: 530); Exon 11, CCACTTACTGTTCATATAATACAGAG (sense) (SEQ ID NO: 531) and CATGTGAGATAGCATTGGGAATGC (antisense) (SEQ ID NO: 532); Exon 12, CATGACCT ACCATCATTGGAAAGCAG (sense) (SEQ ID NO: 533) and GTAATTTACAGTTAGGAATC (sense) (SEQ ID NO: 534); Exon 13, GTCACCCAAGGTCATGGAGCACAGG (sense) (SEQ ID NO: 535) and CAGAATGC CTGTAAAGCTATAAC (antisense) (SEQ ID NO: 536); Exon 14, GTCCTGGAGTC-CCAACCTCTTGAC (sense) (SEQ ID NO: 537) and GGAAGTGGCTCTGA TGGCCGTCCTG (antisense) (SEQ ID NO: 538); Exon 15, CCAC TCACACACACTAAATATTTTAAG (sense) (SEQ ID NO: 539) and GACCAAAACACCTTAAGTAA CTGACTC (antisense) (SEQ ID NO: 540); Exon 16, CCAA TCCAACATCCAGACACATAG (sense) (SEQ ID NO: 541) and CCAGAGCCATAGAACTTGATCAG (antisense) (SEQ ID NO: 542); Exon 17, GTATGGACTATGGC ACT-TCAATTGCATGG (sense) (SEQ ID NO: 543) and CCAGAGAACATGGCAACCAGCACAGGAC (antisense) (SEQ ID NO: 544); Exon 18, CAAATGAGCTGGCAAGTGCCGTGTC (sense) (SEQ ID NO: 545) and GAGTTT CCAAACACT-CAGTGAAAC (antisense) (SEQ ID NO: 546) or CAAGTGCCGTGTCCTGGCACCCAAGC (sense) (SEQ ID NO: 675) and CCAAACACTCAGTGAAACAAAGAG (antisense) (SEQ ID NO: 676); Exon 19, GCAATATCAGCC TTAGG TGCG-GCTC (sense) (SEQ ID NO: 547) and CATAGAAAGTGAACATTTAGGATGTG (antisense) (SEQ ID NO: 548); Exon 20, CCATGAGTACGTATTTTGAACTC (sense) (SEQ ID NO: 549) and CATATCC CCATGGC AAACCTCTTGC (antisense) (SEQ ID NO: 550); Exon 21, CTAACGTTCGCCAG CCATAAGTCC (sense) (SEQ ID NO: 551) and GCT-GCGAGCTCACCCAGAATGTCTGG (antisense) (SEQ ID NO: 552); Exon 22, GACGGG TCCTGGGGTGATCTGGCTC (sense) (SEQ ID NO: 553) and CTCAGTACAATAGATAGACAGCAATG (antisense) (SEQ ID NO: 684); Exon 23, CAG-GACTACAGAAATGTAGGTTTC (sense) (SEQ ID NO: 555) and GTGCCCTG CCTTAAGTAATGTGATGAC (antisense) (SEQ ID NO: 556); Exon 24, GACTGG AAGTGTCGCA TCACCAATG (sense) (SEQ ID NO: 557) and GGTTTAATAAT-GCGATCTGGGACAC (antisense) (SEQ ID NO: 558); Exon 25, GCAGCTATAATTTAGAGAACCAAGG (sense) (SEQ ID NO: 559) and GGTT AAAATTGACTTC ATTTCCATG (antisense) (SEQ ID NO: 560); Exon 26, CCTAGTTGCTCTAAA ACTAACG (sense) (SEQ ID NO: 561) and CTGTGAGGCGTGACAGCCGTGCAG (antisense) (SEQ ID NO: 562); Exon 27, CAACCTACTAATCAG AACCAGCATC (sense) (SEQ ID NO: 563) and CCTTCACTGTGTCTGC AAATCTGC (antisense) (SEQ ID NO: 564); Exon 28, CCTGTCATAAGTCTCCTTGTTGAG (sense) (SEQ ID NO: 565) and CAGTCT-GTGGGTCTAAG AGCTAATG (antisense) (SEQ ID NO: 566). Annealing temperatures were 58°C (exons 1, 3, 4, 7-10, 12-25, 27, and 28), 56°C (exons 2, 5, 6, and 26), or 52°C (exon 11).

[0235] Nested PCR amplification of DNA extracted from archival tumor tissue was performed as follows. An initial PCR for exons 2, 5, 6, 7, 11, 12, 14, 16, 18, 19, 20, 21, 23, 24, 25, 26, and 27 was generated using primers and conditions described above. Subsequently, 2 µl of this reaction was amplified in a secondary PCR using the following internal primer pairs: Exon 2, CAGGAATGGGTGAGTCTCTGTGTG (sense) (SEQ ID NO: 567) and GTGGAATTCTGCCAG-

GCCTTTC (antisense) (SEQ ID NO: 568) ; Exon 5, GATTCTACAAACCA GCCAGCCAAAC (sense) (SEQ ID NO: 569) and CCTACTGGTTCACATCTGACCCTG (antisense) (SEQ ID NO: 570) ; Exon 6, GTTTGAATGTGGTTTCGTTGGAAG (sense) (SEQ ID NO: 571) and CTTTGTCAACCAGG CAGAGG GCAATATC (antisense) (SEQ ID NO: 572) ; Exon 7, GACAGTAACTTGGGCTTTCTGAC (sense) (SEQ ID NO: 573) and CATCCACCCAAAGACTCTCCAAG (antisense) (SEQ ID NO: 574) ; Exon11, CTGTTTCATA TAATAC AGAGTCCCTG (sense) (SEQ ID NO: 575) and GAGAGATGCAG-GAGCTCTGTGC (antisense) (SEQ ID NO: 576) ; Exon12, GCAGTTTGTAGTCAATCAAAGGTGG (sense) (SEQ ID NO: 577) and GTAATTTAAATGGGAAT AGCCC (antisense) (SEQ ID NO: 578) ; Exon14, CAACTCCTTGACCATTAC-CTCAAG (sense) (SEQ ID NO: 579) and GATGGCCGTCCTGCCACACAGG (antisense) (SEQ ID NO: 580) ; Exon16, GAGTAGTTTACGCA TATATTGC (sense) (SEQ ID NO: 581) and GACAGTCAGAAATGCAGGAAAGC (antisense) (SEQ ID NO: 582) ; Exon18, CAAGTGCCGTGTCCTGGCACCCAAGC (sense) (SEQ ID NO: 583) and CCAAACACTCA GT-GAAACAAAGAG (antisense) (SEQ ID NO: 584) or GCACCCAAGCCCATGCCGTGGCTGC (sense) (SEQ ID NO: 677) and GAAACAAAGAGTAAAGTAGATGATGG (antisense) (SEQ ID NO: 678) ; Exon 19, CCTTAGGTGCGGCTC-CACAGC (sense) (SEQ ID NO: 585) and CATTTAGGATGTGGAGATGAGC (antisense) (SEQ ID NO: 586) ; Exon 20, GAAACTCAAG ATCGCATTCATGC (sense) (SEQ ID NO: 587) and GCAAACCTTGCTATCCCAGGAG (antisense) (SEQ ID NO: 588) ; Exon 21, CAGCCATAAGTCCTCGACGTGG (sense) (SEQ ID NO: 589) and CATCCTCCCCT GCATGTGTTAAAC (antisense) (SEQ ID NO: 590) ; Exon 23, GTAGTTTCTAAACATCAAGAAAC (sense) (SEQ ID NO: 591) and GTGATGACATTTCTCCAGGGATGC (antisense) (SEQ ID NO: 592) ; Exon 24, CATCACCA ATGCCT-TCTTTAAGC (sense) (SEQ ID NO: 593) and GCTGGAGGGTTTAATAATGCGATC (antisense) (SEQ ID NO: 594) ; Exon 25, GCAAACACACAGGCACCTGCTGGC (sense) (SEQ ID NO: 595) and CATTTTC CATGTGAGTTTCACTAGAT-GG (antisense) (SEQ ID NO: 596) ; Exon 26, CACCTTCACAATATACCCTCCATG (sense) (SEQ ID NO: 679) and GACAGCCGTGCAGGGAAAAACC (antisense) (SEQ ID NO: 680) ; Exon 27, GAACCAGCATCTCAAGGAGATCTC (sense) (SEQ ID NO: 681) and GAGCACCTGGCTTGACACTGGAG (antisense) (SEQ ID NO: 682).

[0236] Nested PCR amplifications for the remaining exons consisted of primary PCR using the following primers. Exon 1, GACCGGACGACAGGCCACCTCGTC (sense) (SEQ ID NO: 597) and GAAGAACGAAACGTCCCGTTCCTCC (antisense) (SEQ ID NO: 598) ; Exon 3, GTTGAGCACT CGTGTGCATTAGG (sense) (SEQ ID NO: 599) and CTCAGT-GCACGTGTACTGGGTA (antisense) (SEQ ID NO: 600) ; Exon 4, GTTCACTGGGCTAATTGCGGGACTCTTGTTCG-CAC (sense) (SEQ ID NO: 601) and GGTA AATACATGCTTTTCTAGTGGTCAG (antisense) (SEQ ID NO: 602) ; Exon 8, GGAGGATGGA GCCTTTCCATCAC (sense) (SEQ ID NO: 603) and GAAGAGGAAGATGTGTTCCCTTTGG (anti-sense) (SEQ ID NO: 604) ; Exon 9, GAATGAAGGATGATGTGGCAGTGG (sense) (SEQ ID NO: 605) and GTATGT-GTGAAGGAG TCACTGAAAC (antisense) (SEQ ID NO: 606) ; Exon 10, GGTGAGTCACAGGTTCACTTGC (sense) (SEQ ID NO: 607) and CAAACATCAGCCATTAACGG (antisense) (SEQ ID NO: 608) ; Exon 13, GTAGCCAGCATGTC TGTGTAC (sense) (SEQ ID NO: 609) and CAGAATGCCTGTAAAGCTATAAC (antisense) (SEQ ID NO: 610) ; Exon 15, CATTTGGCTTTCCCACTCACAC (sense) (SEQ ID NO: 611) and GACCAAAACACCTTAA GTAAGTACTC (anti-sense) (SEQ ID NO: 612) ; Exon 17, GAAGCTACATAGTGTCTCACTTTCC (sense) (SEQ ID NO: 613) and CACAAC-TGCTAATGGCCGTTCTCG (antisense) (SEQ ID NO: 614) ; Exon 22, GAGCAGCCCTGAACTCCGTCAGACTG (sense) (SEQ ID NO: 683) and CTCAGTACAATAGATAGACAGCAATG (antisense) (SEQ ID NO: 684) ; Exon 28a GCTCC TGCTCCCTGTCTAAGTC (sense) (SEQ ID NO: 615) and GAAGTCCTGCTGGTAGTCAGGGTTG (antisense) (SEQ ID NO: 616) ; Exon 28b, CTGCAGTGGGCAACCCCGAGTATC (sense) (SEQ ID NO: 617) and CAGTC TGTGGGTCTAA-GAGCTAATG (antisense) (SEQ ID NO: 618). Secondary PCR amplification was carried out using primer pairs: Exon 1, GACAGGCCACCTCGTCGGCGTC (sense) (SEQ ID NO: 619) and CAGCTGATCTCAAGGAAACAGG (antisense) (SEQ ID NO: 620) ; Exon 3, CTCGTG TGCATTA GGGTTCAACTGG (sense) (SEQ ID NO: 621) and CCTTCTCCGAG-GTGAATTGAGTGAC (antisense) (SEQ ID NO: 622) ; Exon 4, GCTAATTGCGGGACTCTTGTTCGCAC (sense) (SEQ ID NO: 623) and TACATGCTTT TCTAGTGGTCAG (antisense) (SEQ ID NO: 624) ; Exon 8, CCTTTCCATCACCCCT-CAAGAGG (sense) (SEQ ID NO: 625) and ' GATGTGTTCTTTGGAGGTGGCATG (antisense) (SEQ ID NO: 626) ; Exon 9, GATGTGG CAGTGGCGGTTCCGGTG (sense) (SEQ ID NO: 627) and GGAGTCACTGAAACAAACAACAGG (antisense) (SEQ ID NO: 628) ; Exon 10, GGTTCACTGCTTGTATAAAG (sense) (SEQ ID NO: 629) and CCATTAACG-GT AAAATTTTCAAGAG (antisense) (SEQ ID NO: 630) ; Exon 13, CCAAGGTCATGGAGCACAGG (sense) (SEQ ID NO: 631) and CTGTAAAGCTATAACAACAACCTGG (antisense) (SEQ ID NO: 632) ; Exon 15, CCACTCACA CACTAAATATTTTAAAG (sense) (SEQ ID NO: 633) and GTAAGTACTCAAATACAAACCAC (antisense) (SEQ ID NO: 634) ; Exon 17, GAAGCTACATAGTGTCTCACTTTCC (sense) (SEQ ID NO: 635) and CACAA CTGCTAAT-GGCCGTTCTCG (antisense) (SEQ ID NO: 636) ; Exon 22, GACGGGTCCTGGGGTGATCTGGCTC (sense) (SEQ ID NO: 685) and CTCAGTACAATAGATAGACAGCAATG (antisense) (SEQ ID NO: 686) ; Exon 28a, CCTGTCTAAG TCTCCTTGTGAG (sense) (SEQ ID NO: 637) and GGTAGTCAGGGTTGTCCAGG (antisense) (SEQ ID NO: 638) ; Exon 28b, CGAGTATCTCAACACTGTCCAGC (sense) (SEQ ID NO: 639) and CTAAGAGCTAATGCGGGC ATGGCTG (antisense) (SEQ ID NO: 640) . Annealing temperature for exon 1 amplifications was 54°. Annealing temperatures for both primary and secondary amplifications were 58°C (exons 3, 4, 7-10, 12-17, 19-25, 27, and 28), 56°C (exons 2, 5, 6, and 26), or 52°C (exons 11 and 18).

[0237] PCR amplicons were purified using exonuclease I (United States Biochemical, Cleveland, OH), and shrimp

alkaline phosphatase (United States Biochemical, Cleveland, OH) prior to sequencing. Purified DNA was diluted and cycle-sequenced using the ABI BigDye Terminator kit v1.1 (ABI, Foster City, CA) according to manufacturer's instructions. Sequencing reactions were electrophoresed on an ABI3100 genetic analyzer. Electropherograms were analyzed in both sense and antisense direction for the presence of mutations, using Sequence Navigator software in combination with

5 Factura to mark heterozygous positions. All sequence variants were confirmed in multiple independent PCR amplifications and sequencing reactions.

Cancer-Derived Cell Lines:

10 **[0238]** A panel of 14 lung cancer-derived cell lines was analyzed for *EGFR* mutations. These were derived from tumors of NSCLC (N=5), small cell lung cancer (N=6), adenosquamous (N=1), bronchial carcinoid (N=1), and unknown histology (N=1). Specific cell lines were: NCI-H460, NCI-522, HOP-92, NCIH841, NCIH734, NCIH2228, NCIH596, NCIH727, NCIH446, NCIH1781, NCIH209, NCIH510, NCIH82, NCIH865. In addition, 64 cancer-derived cell lines were screened for mutations in exons 19 and 21. These represented the following histologies: breast cancer (BT549, BT483, UACC893,

15 HS467T, HS578T, MCF7, MCF7-ADR, MDA-MB-15, MDA-MB-175, MDA-MB-231, MDA-MB-415, MDA-MB-436, MDA-MB-453, MDA-MB-468, T47D), ovarian cancer (ES-2, IGROV-1, MDAH2774, OV1063, OVCAR3, OVCAR4, OVCAR5, SKOV3, SW626), CNS cancers (SF-295, SNB-19, U-251, CCF-STTG1, SW-1088, SW-1783, T98G, M059K, A172, SK-N-DZ, SK-N-MC), leukemia (CCRF-CEM, K562, MOLT-4, RPMI8226, SR), prostate cancer (DU-145, PC-3), colon cancer (COLO-205, HCT-116, HCT-15, HT-29, SW-620), renal cancer (786-O, ACHN, CAKI-1, SN-12C, UO-31), melanoma

20 (LOX-IMVI, M14, SKMEL2, UACC-62), osteosarcoma (SAOS-2), and head and neck cancers (O11, O13, O19, O28, O22, O29, O12). The head and neck cancer cell-lines were provided by Dr. James Rocco, Massachusetts General Hospital/Massachusetts Eye and Ear Infirmary. All other cell-lines are available through the American Type Culture Collection (Manassas, VA).

[0239] Genomic DNA was isolated from snap-frozen tumor specimens. Tumor specimens were first crushed to a fine powder using a pre-chilled and sterilized mortar and pestle. Tumor tissue was immediately transferred into a DNA extraction solution consisting of 100mM sodium chloride, 10mM Tris pH7.5, 25mM EDTA (disodium ethylenediamine tetraacetate) pH8.0, and 0.5% (w/v) sodium dodecyl sulfate, and 100µg/ml fresh proteinase K and incubated overnight at 37°C or for 3 hours at 50°C. DNA was then extracted using standard phenol-chloroform methods, ethanol precipitated,

25 washed with 70% ethanol, air-dried and resuspended in TE buffer. The DNA concentration was determined spectrophotometrically. Exons 19 and 21 of human *EGFR* were amplified by the polymerase chain reaction using the following primer pairs: Exon19 sense primer, 5'-GCAATATCAGCCTTAGGTGCGGCTC-3' (SEQ ID NO: 505); Exon 19 antisense primer, 5'-CATAGAA AGTGAACATTTAGGATGTG-3' (SEQ ID NO: 506); Exon 21 sense primer, 5'-CTAACGTTCCG CCAGCCATAAGTCC-3' (SEQ ID NO: 507); Exon21 antisense primer, 5'- GCTGCGAGCTCACCCAG AATGTCTGG-3' (SEQ ID NO: 508). For each sample, 20ng of genomic DNA was amplified in a PCR reaction consisting of 1X Expand Long Template buffer 1 (Roche, Mannheim Germany), 50µM sequencing grade dATP (Amersham Biosciences, Cleveland OH), 50µM sequencing grade dCTP (Amersham Biosciences, Cleveland OH), 50µM sequencing grade dGTP (Amersham Biosciences, Cleveland OH), 50µM sequencing grade dTTP (Amersham Biosciences, Cleveland OH), 0.2µM sense primer, 0.2µM antisense primer, 1.25 units Expand Long Template enzyme mix (Taq DNA polymerase/Tgo DNA polymerase) (Roche, Mannheim Germany) that has been preincubated for 5 minutes on ice with 1/6 volume of TaqStart Antibody (1.1 µg/µl) (Clontech, Palo Alto, CA) and water to final volume of 25µl. Each series of amplifications also

30 includes a negative control for which the DNA template is omitted. PCR cycling conditions for both exons were 95°C for 2 min followed by 40 cycles of 95°C for 30 s, 58°C for 30 s and 72°C for 45 sec; and a final extension of 72°C for 10 min followed by holding at 4°C on an MJ-Research PTC-200 or PTC-225 thermal-cycler (MJ-Research, Waltham MA).

[0240] PCR products were resolved by electrophoresis through a 0.8% agarose gel to ensure amplification from patient material and no amplification in the negative control. PCR products were purified prior to sequencing by mixing 10µl each PCR amplicon with 0.5µl exonuclease I (10U/µl) (United States Biochemical, Cleveland, OH), and 1 µl shrimp alkaline phosphatase (1U/µl) (United States Biochemical, Cleveland, OH) and incubating at 37°C for 20 minutes followed by inactivation at 80°C for 15 minutes on a thermal-cycler (MJ-Research, Waltham, MA). Purified DNA was diluted in water, according to the intensity of the amplicon, and cycle-sequencing was performed using the ABI BigDye Terminator kit v1.1 (Applied Biosystems, Foster City, CA) according to manufacturer's instructions. Cycle-sequencing was performed on an MJ-Research thermal-cycler using the following cycling conditions: Primers used for sequencing were: Exon19 sense primer, 5'-GCAATATCAGCCTTAGGTGCGGCTC-3' (SEQ ID NO: 505); Exon 19 antisense primer, 5'- CATAG AAAGTGAACATTTAGGATGTG-3' (SEQ ID NO: 506); Exon21 sense primer, 5'-CTAACGTTCCGCCAG CCATAAGTCC-3' (SEQ ID NO: 507) or 5'-CGTGGAGAGGCTCAGAGCCTGGCATG-3' (SEQ ID NO: 687); Exon 21 antisense primer,

35 5'-GCTGCGAGCTCACCCAGAATGTCTGG-3' (SEQ ID NO: 508). Sequencing reactions were electrophoresed on an ABI3100 genetic analyzer (Applied Biosystems, Foster City, CA). Factura and Sequence Navigator (Applied Biosystems, Foster City, CA) software programs were used to mark potential heterozygous positions and display them for evaluation. Nucleotide positions at which the height of the secondary peak was greater than, or equal to, 30% the height of the

50

primary peak were marked as heterozygous and were confirmed by analysis of both sense and antisense reads. Samples with sequence indicative of the presence of a mutation were re-amplified and sequenced for confirmation.

Position of Primers Used in Sequence Analysis Relative to Exons 19 and 21

Intronic primers are shown in lower case and underlined.

5 Intronic Sequence is Shown in Lowercase.

Exonic Sequence is Shown in Uppercase.

EGFR Exon 19 (5'-3') (SEQ ID NO: 641)

10 gcaatatcagccttaggtgcccgtccacagccccagtgtccctcaccttcgggggtgcatcgc
 tggtaacatccaccagatcactgggcagcatgtggcaccatctcacaattgccagttaacg
 tcttccttctctctctgtcatagGGACTCTGGATCCCAGAAGGTGAGAAAATTAAATTTCCC
 15 GTCGCTATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGCCAACAAGGAAATCCTCGATgt
 gagtttctgctttgctgtgtgggggtccatggctctgaacctcaggcccaccttttctcatg
 tctggcagctgctctgctctagacctgctcatctccacatcctaaatgttcactttctatg

20 EGFR Exon 21 (5'-3') (SEQ ID NO: 642) or (SEQ ID NO: 687)

ctaacgttcgcccagccataagtcctcgacgtggagaggctcagagcctggcatgaacatgac
 25 cctgaattcggatgcagagcttcttcccatgatgatctgtccctcacagcagggtcttctct
 gtttcagGGCATGAACTACTTGGAGGACCGTCGCTTGGTGACCGCGACCTGGCAGCCAGGA
 ACGTACTGGTGAAAAACCCGCAGCATGTCAAGATCACAGATTTTGGGCTGGCCAAACTGCTG
 30 GGTGCGGAAGAGAAA GAATACCATGCAGAAGGAGGCAAAGtaaggaggtggcttttaggtcag
 ccagcattttctga caccagggaccaggtgccttcccactagctgtattgtttaacacat
 gcaggggaggatgctctccagacattctgggtgagctcgcagc

35 RESULTS

Clinical Characteristics of Gefitinib Responders

40 **[0241]** Patients with advanced, chemotherapy-refractory NSCLC have been treated with single agent Gefitinib since 2000 at Massachusetts General Hospital. A total of 275 patients were treated, both prior to its approval on May 2003 by the FDA, as part of a compassionate use expanded access program, and following that date using commercial supply. During this period, 25 patients were identified by clinicians as having significant clinical responses. A significant clinical response was defined either as a partial response using RECIST criteria for patients with measurable disease, or for patients whose tumor burden could not be quantified using these criteria, an evaluable response was assessed by two
 45 physicians. Table 1 shows clinical characteristics of 9 cases for whom tumor specimens obtained at the time of initial diagnosis were available. For the other Gefitinib-responders, tissue was not available, most commonly because diagnostic specimens were limited to cytology from needle aspirates. As a group, the 9 patients experienced substantial benefit from Gefitinib. The median survival from the start of drug treatment is in excess of 18 months, and the median duration of therapy is greater than 16 months. Consistent with previous reports, Gefitinib-responders have a high prevalence of
 50 female sex, absence of smoking history, and tumors with bronchioalveolar histology (11, 12). Case 6 is representative of the Gefitinib-responsive cohort. This patient is a 32 year-old man, without smoking history, who presented with multiple brain lesions and disease in the right lung diagnosed as bronchioalveolar carcinoma. He was treated with whole brain radiotherapy, followed by a series of chemotherapy regimens to which his tumor did not respond (carboplatin and gemcitabine; docetaxel; vinorelbine). With a declining functional status and progressive lung tumor burden, he started
 55 therapy with 250 mg per day of Gefitinib. His shortness of breath promptly improved and a lung CT scan 6 weeks after initiation of treatment revealed the dramatic improvement shown in Figure 1.

EGFR Mutations in Gefitinib Responders

[0242] We hypothesized that cases of NSCLC with striking responses to Gefitinib might harbor somatic mutations in *EGFR*, indicating an essential role played by this growth factor signaling pathway in these tumors. To search for such mutations, we first tested for rearrangements within the extracellular domain of EGFR that are characteristic of gliomas (15): none were detected. We therefore sequenced the entire coding region of the gene using PCR-amplification of individual exons. Heterozygous mutations were observed in 8/9 cases, all of which were clustered within the kinase domain of EGFR (Table 2 and Figure 2). Four tumors had in-frame deletions removing amino acids 746-750 (delE756-A750; case 1), 747 to 750 (delL747-T751insS; case 2), and 747 to 752 (delL747-P753insS; cases 3 and 4). The latter two deletions were associated with the insertion of a serine residue, resulting from the generation of a novel codon at the deletion breakpoint. Remarkably, these four deletions were overlapping, with the deletion of four amino acids (leucine, arginine, glutamic acid and alanine, at codons 747 to 750) within exon 19 shared by all cases (see Figure 4a). Another three tumors had amino acid substitutions within exon 21: leucine to arginine at codon 858 (L858R; cases 5 and 6), and leucine to glutamine at codon 861 (L861Q; case 7). The L861Q mutation is of particular interest, since the same amino acid change in the mouse *egfr* gene is responsible for the Dark Skin (dsk5) trait, associated with altered EGFR signaling (18). A fourth missense mutation in the kinase domain resulted in a glycine to cysteine substitution at codon 719 within exon 18 (G719C; case 8). Matched normal tissue was available for cases 1, 4, 5 and 6, and showed only wild-type sequence, indicating that the mutations had arisen somatically, during tumor formation. No mutations were observed in seven cases of NSCLC that failed to respond to Gefitinib ($P=0.0007$; 2-sided Fisher's exact test).

Prevalence of Specific EGFR Mutations in NSCLC and Other Cancer Types

[0243] Unlike gliomas, in which rearrangements affecting the EGFR extracellular domain have been extensively studied (15), the frequency of *EGFR* mutations in NSCLC has not been defined. We therefore sequenced the entire coding region of the gene in 25 primary cases of NSCLC unrelated to the Gefitinib study, including 15 with bronchioalveolar histology, which has been associated with Gefitinib-responsiveness in previous clinical trials (11, 12). Heterozygous mutations were detected in two bronchioalveolar cancers. Both cases had in-frame deletions in the kinase domain identical to those found in Gefitinib responders, namely delL747-P753insS and delE746-A750 (Table 2). Given the apparent clustering of *EGFR* mutations, we sequenced exons 19 and 21 in a total of 55 primary tumors and 78 cancer-derived cell lines, representing diverse tumor types (see Supplementary Material). No mutations were detected, suggesting that these arise only in a subset of cancers, in which EGFR signaling may play a critical role in tumorigenesis.

Increase in EGF-Induced Activation and Gefitinib Inhibition of Mutant EGFR Proteins

[0244] To study the functional properties encoded by these mutations, the L747-S752insS deletion and the L858R missense mutants were expressed in cultured cells. Transient transfection of wild-type and mutant constructs into Cos-7 cells demonstrated equivalent expression levels, indicating that the mutations do not affect protein stability. EGFR activation was quantified by measuring phosphorylation of the tyrosine¹⁰⁶⁸ residue, commonly used as a marker of receptor autophosphorylation (19). In the absence of serum and associated growth factors, neither wild-type nor mutant EGFR demonstrated autophosphorylation (Figure 3a). However, addition of EGF led to a 2-3 fold increase in receptor activation for both the missense and deletion *EGFR* mutants, compared with the wild-type receptor. Moreover, whereas normal EGFR activation was downregulated after 15 min, consistent with receptor internalization, the two mutant receptors demonstrated continued activation for up to 3 hrs (Figure 3a). Similar results were obtained with antibodies measuring total EGFR phosphorylation following addition of EGF (not shown).

[0245] Since 7/8 EGFR kinase mutations reside near the ATP cleft, which is targeted by Gefitinib, we determined whether the mutant proteins have altered sensitivity to the inhibitor. EGF-induced receptor autophosphorylation was measured in cells pretreated with variable concentrations of Gefitinib. Remarkably, both mutant receptors displayed increased sensitivity to inhibition by Gefitinib. Wild-type EGFR had an IC_{50} of 0.1 μ M and showed complete inhibition of autophosphorylation at 2 μ M Gefitinib, whereas the two mutant proteins had an IC_{50} of 0.015 μ M and abrogation of autophosphorylation at 0.2 μ M (Figure 3b). This difference in drug sensitivity may be clinically relevant, since pharmacokinetic studies indicate that daily oral administration of 400-600 mg of Gefitinib results in a mean steady-state trough plasma concentration of 1.1-1.4 μ M, while the currently recommended daily dose of 250 mg leads to a mean trough concentration of 0.4 μ M (20).

Example 2

[0246] Tumor cells harboring mutations within the kinase domain of the EGFR, and are therefore sensitive to growth inhibition by gefitinib treatment, can undergo "second-site" mutations, also within the kinase domain, that confer resistance

to gefitinib but are still "activating" in the sense that they exhibit increased EGFR signaling relative to wild-type EGFR. Such gefitinib-resistant mutants are generated from two sporadic human NSCLC cell lines namely NCI-1650 and NCI-1975. Each cell line contains a heterozygous mutation with the kinase domain of EGFR, and is, therefore, expected to be sensitive to gefitinib. The EGFR mutation in NCI-1650 consists of an in-frame deletion of 15 nucleotides at position 2235-2249 (delLE746-A750) within exon 19, while NCI-1975 has a missense mutation within exon 21 that substitutes a G for T at nucleotide 2573 (L858R). The L858R mutation in NCI-H1975 has been shown herein to be activating and to confer increased sensitivity to gefitinib in vitro.

[0247] Gefitinib-resistant cell lines, derived from both NCI-1650 and NCI-1975 are isolated, following random chemical mutagenesis using EMS (ethyl methanesulfonate) followed by culture in gefitinib-supplemented medium to select for the outgrowth of resistant clones. Subcultivation of individual clones is followed by nucleotide sequence determination of the EGFR gene following specific PCR-mediated amplification of genomic DNA corresponding to the EGFR kinase domain.

[0248] A variation of this strategy involves the serial passage of these two cell lines in the presence of gradually increasing concentrations of gefitinib over a course of several weeks or months in order to select for the spontaneous acquisition of mutations within the EGFR gene that confer resistance to gefitinib. Selected cells (that continue to proliferate at relatively high gefitinib concentration) are isolated as colonies, and mutations are identified as described above.

Example 3

[0249] To determine whether mutation of receptor tyrosine kinases plays a causal role in NSCLC, we searched for somatic genetic alterations in a set of 119 primary NSCLC tumors, consisting of 58 samples from Nagoya City University Hospital in Japan and 61 from the Brigham and Women's Hospital in Boston, Massachusetts. The tumors included 70 lung adenocarcinomas and 49 other NSCLC tumors from 74 male and 45 female patients, none of whom had documented treatment with EGFR kinase inhibitors.

[0250] As an initial screen, we amplified and sequenced the exons encoding the activation loops of 47 of the 58 human receptor tyrosine kinase genes (*) (Table S1) from genomic DNA from a subset of 58 NSCLC samples including 41 lung adenocarcinomas. Three of the tumors, all lung adenocarcinomas, showed heterozygous missense mutations in *EGFR* not present in the DNA from normal lung tissue from the same patients (Table S2; S0361, S0388, S0389). No mutations were detected in amplicons from other receptor tyrosine kinase genes. All three tumors had the *same EGFR* mutation, predicted to change leucine ("L") at position 858 to arginine ("R") (FIG. 6A; CTG→CGG; "L858R"), wherein all numbering refers to human EGFR.

[0251] We next examined exons 2 through 25 of *EGFR* in the complete collection of 119 NSCLC tumors. Exon sequencing of genomic DNA revealed missense and deletion mutations of *EGFR* in a total of 16 tumors, all within exons 18 through 21 of the kinase domain. All sequence alterations in this group were heterozygous in the tumor DNA; in each case, paired normal lung tissue from the same patient showed wild-type sequence, confirming that the mutations are somatic in origin. The distribution of nucleotide and protein sequence alterations, and the patient characteristics associated with these abnormalities, are summarized in Table S2.

[0252] Substitution mutations G719S and L858R were detected in two and three tumors, respectively. The "G719S" mutation changes the glycine (G) at position 719 to serine (S) (FIG. 6B). These mutations are located in the GXGXXG motif (SEQ ID NO:490) of the nucleotide triphosphate binding domain or P-loop and adjacent to the highly conserved DFG motif in the activation loop (52), respectively. See, e.g., FIG. 7. The mutated residues are nearly invariant in all protein kinases and the analogous residues (G463 and L596) in the B-Raf protein serine-threonine kinase are somatically mutated in colorectal, ovarian and lung carcinomas (41, 53) (FIG. 6A, 6B).

[0253] We also detected multiple deletion mutations clustered in the region spanning codons 746 to 759 within the kinase domain of EGFR. Ten tumors carried one of two overlapping 15-nucleotide deletions eliminating *EGFR* codons 746 to 750, starting at either nucleotide 2235 or 2236 (Del-1; FIGS. 6C and 8C; Table S2). *EGFR* DNA from another tumor displayed a heterozygous 24-nucleotide gap leading to the deletion of codons 752 to 759 (Del-2; FIG. 6C). Representative chromatograms are shown in FIGS. 8A-8F.

[0254] The positions of the substitution mutations and the Del-1 deletion in the three-dimensional structure of the active form of the EGFR kinase domain (54) are shown in FIG. 7. Note that the sequence alterations cluster around the active site of the kinase, and that the substitution mutations lie in the activation loop and glycine-rich P-loop, structural elements known to be important for autoregulation in many protein kinases (52).

[0255] Two additional *EGFR* mutations in two different tumor types have been identified. Namely, we have identified the *EGFR* mutation G857V in Acute Myelogenous Leukemia (AML) and the *EGFR* mutation L883S in a metastatic sarcoma. The "G857V" mutation has the glycine (G) at position 857 substituted with a valine (V), while the "L883S" mutation has the leucine (L) at position 883 substituted with a serine (S). These findings suggest that mutations in EGFR occur in several tumor types and, most importantly, that EGFR inhibitors would be efficacious in the treatment of patients harboring such mutations. This expands the use of kinase inhibitors such as, e.g., the tyrosine kinase inhibitors gefitinib

(marketed as Iressa™), erlotinib (marketed as Tarceva™), and the like in treating tumor types other than NSCLC.

[0256] The *EGFR* mutations show a striking correlation with the differential patient characteristics described in Japanese and U.S. patient populations. As noted above, clinical trials reveal significant variability in the response to the tyrosine kinase inhibitor gefitinib (Iressa™), with higher responses seen in Japanese patients than in a predominantly European-derived population (27.5% vs. 10.4%, in a multi-institutional phase II trial) (48); and with partial responses seen more frequently in the U.S. in women, non-smokers, and patients with adenocarcinomas (49-51). We show that *EGFR* mutations were more frequent in adenocarcinomas (15/70 or 21%) than in other NSCLCs (1/49 or 2%); more frequent in women (9/45 or 20%) than in men (7/74 or 9%), and more frequent in the patients from Japan (15/58 or 26%, and 14/41 adenocarcinomas or 32%) than in those from the US (1/61 or 2%, and 1/29 adenocarcinomas or 3%). The highest fraction of *EGFR* mutations was observed in Japanese women with adenocarcinoma (8/14 or 57%). Notably, the patient characteristics that correlate with the presence of *EGFR* mutations appear to be those that correlate with clinical response to gefitinib treatment.

[0257] To investigate whether *EGFR* mutations might be a determinant of gefitinib sensitivity, pre-treatment NSCLC samples were obtained from 5 patients who responded and 4 patients who progressed during treatment with gefitinib, where these patients were identified out of more than 125 patients treated at the Dana-Farber Cancer Institute either on an expanded access program or after regulatory approval of gefitinib (49). Four of the patients had partial radiographic responses ($\geq 50\%$ tumor regression in a CT scan after 2 months of treatment) while the fifth patient experienced dramatic symptomatic improvement in less than two months. All of the patients were from the United States and were Caucasian.

[0258] While sequencing of the kinase domain (exons 18 through 24) revealed no mutations in tumors from the four patients whose tumors progressed on gefitinib, all five tumors from gefitinib-responsive patients harbored *EGFR* kinase domain mutations. The Chi-squared test revealed the difference in *EGFR* mutation frequency between gefitinib responders (5/5) and non-responders (0/4) to be statistically significant with $p = 0.0027$, while the difference between the gefitinib-responders and unselected U.S. NSCLC patients (5/5 vs. 1/61) was also significant with $p < 10^{-12}$ (*). The *EGFR* L858R mutation, previously observed in the unselected tumors, was identified in one gefitinib-sensitive lung adenocarcinoma (FIG. 6A; Table S3A, IR3T). Three gefitinib-sensitive tumors contained heterozygous in-frame deletions (Fig. 6C and Tables S3A and S3B, Del-3 in two cases and Del-4 in one) and one contained a homozygous in-frame deletion (Fig. 6C and Tables S3A and S3B, Del-5). Each of these deletions was within the codon 746 to 753 region of *EGFR* where deletions were also found in unselected tumors. Each of these three deletions is also associated with an amino acid substitution (Tables S3A-S3C). In all four samples where matched normal tissue was available, these mutations were confirmed as somatic.

Example 3A: Primer design

[0259] The cDNA sequences of receptor tyrosine kinases were obtained from GenBank (accession numbers listed in Table S1), and were to the human genome assembly (<http://genome.ucsc.edu>) using the BLAT alignment to identify exon/intron boundaries. External gene specific primer pairs were designed to amplify exon sequences and at least 250 bp of flanking intronic sequence or adjacent exonic sequence on each side using the Primer3 program (http://frodo.wi.mit.edu/primer3/primer3_code.html). The resulting predicted amplicons were then used to design internal primers flanking the exon (generally greater than 50 bp from the exon/intron boundary) and containing appended M13 forward or reverse primer tails. These nested primer sets were tested for appropriate amplicon size and high-quality sequence from control DNA. Amplicons encompassing exons encoding the receptor tyrosine kinase activation loop of 47 tyrosine kinases were amplified and sequenced in a set of 58 primary lung cancer samples from Nagoya City University Medical School. In addition, amplicons covering the full length *EGFR* were also amplified.

Example 3B: PCR and sequencing methods for genomic DNA

[0260] Tyrosine kinase exons and flanking intronic sequences were amplified using specific primers in a 384-well format nested PCR setup. Each PCR reaction contained 5 ng of DNA, 1X HotStar Buffer, 0.8 mM dNTPs, 1 mM MgCl₂, 0.2U HotStar Enzyme (Qiagen, Valencia, CA), and 0.2 μ M forward and reverse primers in a 10 μ L reaction volume. PCR cycling parameters were: one cycle of 95°C for 15 min, 35 cycles of 95°C for 20s, 60°C for 30s and 72°C for 1 min, followed by one cycle of 72°C for 3 min.

[0261] The resulting PCR products were purified by solid phase reversible immobilization chemistry followed by bi-directional dye-terminator fluorescent sequencing with universal M13 primers. Sequencing fragments were detected via capillary electrophoresis using ABI Prism 3700 DNA Analyzer (Applied Biosystems, Foster City, CA). PCR and sequencing were performed by Agencourt Bioscience Corporation (Beverly, MA).

Example 3B: Sequence analysis and validation

[0262] Forward (F) and reverse (R) chromatograms were analyzed in batch by Mutation Surveyor 2.03 (SoftGenetics, State College, PA), followed by manual review. High quality sequence variations found in one or both directions were scored as candidate mutations. Exons harboring candidate mutations were reamplified from the original DNA sample and re-sequenced as above.

Example 3C: Patients

[0263] Lung tumor specimens were obtained from patients with non-small cell lung cancer treated at Nagoya City University Hospital and the Brigham and Women's Hospital (unselected Japanese tumors and gefitinib-treated U.S. tumors, respectively) and from the Brigham and Women's Hospital anonymized tumor bank (unselected U.S. samples) under Institutional Review Board approved studies. Information on gender, age, and histology was available for most samples. Patient samples were also obtained from patients treated on an open-label clinical trial of gefitinib at Dana-Farber Cancer Institute (13). Responses to gefitinib were defined using standard criteria (See, e.g., A. B. Miller, B. Hoogstraten, M. Staquet, A. Winkler, 1981 Cancer 47, 207-14). IRB approval was obtained for these studies.

[0264] Of the gefitinib-responsive patients, there were two patients who had been previously treated with at least one cycle of chemotherapy, one patient previously treated with radiation therapy, one patient concurrently treated with chemotherapy, and one patient who received no other treatment. For gefitinib-insensitive patients, treatment failure was defined as the appearance of new tumor lesions or the growth of existing tumor lesions in a CT scan after 2 months of gefitinib treatment compared to a baseline CT scan.

Example 3D: cDNA sequencing of patient samples

[0265] Total RNA is isolated from tissue samples using Trizol™ (Invitrogen, Carlsbad, CA) and is purified using an RN easy™ mini-elute cleanup kit (Qiagen, Valencia, CA). cDNA is transcribed from 2 µg of total RNA with Superscript II Reverse Transcriptase (Invitrogen Life technologies, Carlsbad, CA), according to the manufacturer's recommendations. The cDNA is used as template for subsequent PCR amplifications of EGFR.

[0266] The components of the PCR are: 20mM Tris-HCl(pH 8.4), 50mM KCl, 1.5mM MgCl₂, 0.1mM each of dATP, dCTP, dGTP, dTTP, 0.2 µM of each primer, and 0.05 units/µl Taq polymerase (Taq Platinum, GIBCO BRL, Gaithersburg, MD). Amplification of fragment "a" requires addition of 4% DMSO to the reaction. The primer sequences are listed in Table S4. Forward and reverse primers are synthesized with 18 base pairs of an overhanging M13 forward and reverse sequences respectively. The thermocycling conditions are: 94°C, 4min; followed by 11 cycles, with denaturing step at 94°C for 20", extension step at 72°C for 20", and with a 20" annealing step that decreased 1°C/ cycle, from 60°C at cycle one to 50°C at cycle 11; cycle 11 was then repeated 25 times. A 6 minute incubation at 72°C followed by a 4°C soak completes the program.

[0267] An aliquot of the PCR reaction is diluted 1:50 with water. The diluted PCR product is sequenced using an M13 Forward Big Dye Primer kit (Perkin-Elmer/ Applied Biosystems, Foster City, CA), according to the manufacturer's recommendations. The sequencing products are separated on a fluorescent sequencer (model 3100 from Applied Biosystems, Foster City, CA). Base calls are made by the instrument software, and reviewed by visual inspection. Each sequence is compared to the corresponding normal sequence using Sequencher 4.1 software (Gene Codes Corp.).

Example 3E: Tumor types expressing mutant EGFR

[0268] Two additional mutations in EGFR were found in two different tumor types. An EGFR mutation that substitutes a glycine (G) for a valine (V) at position 857 ("G857V") was identified in Acute Myelogenous Leukemia (AML). An EGFR mutation that substitutes a leucine (L) with a serine (S) at position 883 ("L883S") in a metastatic sarcoma.

Example 3F: Cell lines

[0269] The effects of gefitinib on NSCLC cell lines in vitro were examined. One cell line, H3255, was particularly sensitive to gefitinib, with an IC₅₀ of 40 nM. Other cell lines had much higher IC₅₀s. For example, a wild type cell line H1666 has an IC₅₀ of 2 µM, which is 50 fold higher than for the mutant cell line. When the EGFR from this cell line was sequenced, it contained the L858R missense mutation, while the other cell lines were wild type for EGFR. Much lower concentrations of gefitinib were required to turn off EGFR and also AKT and ERK phosphorylation by EGFR as compared to EGFR wild type cells, which required at least 100 times higher concentrations of gefitinib to achieve the same effect. These findings suggest that the mutant receptor is more sensitive to the effects of gefitinib. Also note here,

Example 3G: Combination therapies

[0270] Tumor specimens were analyzed from patients with advanced NSCLC treated on the randomized trial of carboplatin/paclitaxel with or without erlotinib. The clinical portion of this trial demonstrated equivalent survival in the two treatment arms. Tumor specimens were available for sequencing from 228 of the 1076 patients. The preliminary clinical characteristics of these patients is not different from the group as a whole with respect to baseline demographics, response rate, median and overall survival.

[0271] Exons 18-21 of the tyrosine kinase domain were sequenced and 29 mutations, for a mutation frequency of 12.7 percent, were identified.

[0272] As a whole the patients with EGFR mutations have a better survival regardless of whether they received treatment with chemotherapy alone or in combination with erlotinib. These differences are statistically significant with a p value of less than 0.001. These findings raise the possibility the EGFR mutations, in addition to being predictors of response to gefitinib and erlotinib, may also be prognostic for an improved survival.

[0273] (*) Note that the frequency of EGFR mutation in the unselected US patients, 1 of 61, appears to be low when compared to the frequency of reported gefitinib response at 10.4%. This difference has a modest statistical significance ($p = 0.025$ by the chi-squared test). Thus this result could still be due to chance, could be due to a fraction of responders who do not have EGFR mutations, or could be due to failure to detect EGFR mutations experimentally in this tumor collection. If the frequency of EGFR mutation in gefitinib-responsive US patients (5/5) is compared to the expected frequency of gefitinib response (10.4%), the chi-squared probability is again less than 10-12.

EXAMPLE 4

Study Design:

[0274] We performed a retrospective cohort study of NSCLC patients referred for somatic *EGFR* kinase domain sequencing from August 2004 to January 2005 at Massachusetts General Hospital (MGH), Dana-Farber Cancer Institute (DFCI), and Brigham and Women's Hospital (BWH). These three institutions comprise Dana-Farber/Partners CancerCare (DF/PCC), an academic joint venture cancer center that cares for approximately 1,200 lung cancer patients per year. In August 2004, *EGFR* kinase domain sequencing was made available for clinical use at DF/PCC. Clinicians could select which patients to refer for testing, however patients needed to have sufficient and appropriate tumor specimens available. Tumor cells had to comprise at least 50% of the specimen based on histologic examination by MGH and BWH reference pathologists, and the specimen had to be from a resection, bronchoscopic biopsy, or core needle biopsy of a primary or metastatic tumor, or a cellblock from pleural fluid. In rare cases, fine needle aspirate samples were determined adequate. Samples could be either paraffin-embedded or frozen tissue. Due to the low incidence of *EGFR* mutations in squamous cell tumors (62) patients with this diagnosis were not eligible for testing.

[0275] We identified patients undergoing *EGFR* testing using the EGFR case log maintained at the Laboratory for Molecular Medicine (LMM), of the Harvard Medical School/Partners HealthCare Center for Genetics and Genomics (CLIA# 22D1005307), the diagnostic testing facility where all sequencing was performed and interpreted. We included all patients referred for *EGFR* testing from DF/PCC with a diagnosis of NSCLC during the study period.

[0276] Patient age, gender, and race were collected from the electronic medical record system. Smoking status, cancer history, *EGFR* kinase domain sequencing results, and subsequent EGFR-TKI treatment plans were documented using structured physician chart review. Specifically, the smoking status and cancer history were obtained from physician and nursing notes. Former smokers were defined as patients who had quit smoking at least one year before their diagnosis of lung cancer and never-smokers were defined as patients who had smoked less than 100 cigarettes in their lifetime. Smokers who had quit within a year of their diagnosis or who were smoking at the time of diagnosis were classified as current smokers. Pack-years of smoking were calculated by multiplying the number of packs smoked per day by the number of years of smoking. Tumor histology and *EGFR* kinase domain sequencing results were obtained from pathology reports. All pathology specimens were centrally reviewed at either MGH or BWH and histology was categorized using the World Health Organization (WHO) classification system (63). Subsequent treatment plans were obtained from physician notes.

[0277] Complete data were available for age, gender, tumor histology, and *EGFR* mutation status. There were missing data for race (12%), tumor stage at time of testing (4%), smoking status (6%), prior treatments (5%), and subsequent EGFR-TKI treatment plans (11%). This study was approved by the Institutional Review Board at DF/PCC.

EGFR Gene Sequencing:

[0278] Serial sections of either frozen or formalin-fixed, paraffin-embedded (FFPE) tumor tissue were cut and placed on a glass slide. A region of tumor tissue consisting of at least 50% viable tumor cells was identified by a pathologist.

FFPE samples were extracted with xylene and ethanol to remove paraffin. Both FFPE and frozen tissue samples were digested with proteinase K overnight. Genomic deoxyribonucleic acid (DNA) was extracted from tissue and peripheral whole blood using standard procedures. Genomic DNA was extracted from saliva samples using the DNA Genotek-Oragene™ saliva kit.

[0279] The kinase domain of *EGFR* (exons 18-24 and flanking intronic regions) was amplified in a set of individual nested polymerase chain reaction (PCR) reactions. The primers used in the nested PCR amplifications are described in Table S1A and B and SEQ ID 1-424 with the addition of universal sequences to the 5' ends of the primers (5' tgtaaacgacggccagt) (SEQ ID NO. 645). The PCR products were directly sequenced bi-directionally by dye-terminator sequencing. PCR was performed in a 384-well plate in a volume of 15 µl containing 5 ng genomic DNA, 2 mM MgCl₂, 0.75 µl DMSO, 1 M Betaine, 0.2 mM dNTPs, 20 pmol primers, 0.2 µl AmpliTaq Gold® (Applied Biosystems), 1X buffer (supplied with AmpliTaq Gold). Thermal cycling conditions were as follows: 95°C for 10 minutes; 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minute for 30 cycles; and 72°C for 10 minutes. PCR products were purified with Ampure® Magnetic Beads (Agencourt).

[0280] Sequencing products were purified using Cleanseq™ Magnetic Beads (Agencourt) and separated by capillary electrophoresis on an ABI 3730 DNA Analyzer (Applied Biosystems). Sequence analysis was performed by Mutation Surveyor (SoftGenetics, State College, PA) and manually by two reviewers. Non-synonymous DNA sequence variants were confirmed by analysis of 3-5 independent PCR reactions of the original genomic DNA sample. Blood or saliva samples from individuals with non-synonymous DNA sequence variants were analyzed to determine whether the sequence changes were unique to tumor tissue.

Statistical Analysis:

[0281] We constructed logistic regression models to assess the univariate association between patient demographic and clinical characteristics and *EGFR* mutation status. To identify significant predictors of mutation positive status, we constructed a multivariable logistic regression model including independent variables identified in prior studies as predictive of mutations, specifically gender, race, histology, and smoking status. Six patients were excluded from these analyses due to missing *EGFR* mutation data as a result of PCR failure. All analyses were performed using SAS statistical software (version 8.02, SAS Institute, Cary, NC).

RESULTS:

Patient Characteristics:

[0282] Among the 100 patients with NSCLC referred for somatic *EGFR* kinase domain sequencing as part of clinical cancer care during the study period, the mean age was 60.7 years and 63% were female (Table 4). The majority of patients were white (76%) or Asian (7%), and had metastatic disease at the time the test was ordered (67%). Nearly all patients (94%) tested for *EGFR* mutations had adenocarcinoma, adenocarcinoma with bronchioloalveolar carcinoma (BAC) features, or pure BAC. Approximately one third of the patients were never-smokers. Therapy administered prior to the referral for *EGFR* testing included surgery (50%), chest radiotherapy (22%), chemotherapy (47%), and *EGFR* directed targeted therapy (11%).

Mutations Identified:

[0283] The average length of time from referral for testing to result availability was 12 business days. The majority of specimens submitted were paraffin-embedded (74%). Six of the 74 (8%) paraffin-embedded specimens failed PCR amplification, while all of the 26 frozen specimens were successfully amplified. Among the 94 patients with interpretable results, 23 (24%) were found to have at least one mutation in the *EGFR* kinase domain, with two of these patients demonstrating two point mutations each, for a total of 25 mutations identified (Table 5). Among the 23 patients with mutations, 9 (39%) had one or more point mutations, 12 (52%) had in-frame overlapping deletions in exon 19 and two patients (9%) had duplications in exon 20. The point mutations were in exons 18 and 21, and included five 2573T>G (L858R), and one each of 2126A>T (E709V), 2155G>A (G719S), 2156G>C (G719A), 2327G>A (R776H), 2543C>T (P848L), and 2582T>A (L861Q). One of the point mutations (P848L) was detected in both the tumor specimen and in mononuclear cells obtained from a buccal swab. No mutations were detected in exons 22, 23, or 24.

Predictors of Mutations:

[0284] In our sample, there was no significant association between *EGFR* mutation status and age ($p = 0.61$), female gender ($p = 0.92$), Asian race ($p = 0.08$), or metastatic disease at the time of referral ($p = 0.43$, Table 4). None of the 6

patients with non-adenocarcinoma tumor histology were found to have mutations. Among the patients with adenocarcinoma, adenocarcinoma with BAC features and pure BAC, there was no association between BAC/BAC features and *EGFR* mutation status ($p = 0.35$).

[0285] None of the 17 current smokers were found to have a mutation. Never-smokers were significantly more likely to have an *EGFR* mutation than former smokers (odds ratio [OR] = 3.08, 95% confidence interval [CI] 1.09-8.76). The mean number of pack-years smoked was significantly lower among *EGFR* mutation-positive patients (0.7 pack-years) compared to *EGFR* mutation-negative patients (25.0 pack-years, $p < 0.001$). For each additional pack-year smoked, there was a 4% decrease in the likelihood of having a mutation (OR = 0.96, 95% CI 0.93-0.99).

[0286] The number of pack-years of smoking remained a significant predictor of mutation status after controlling for gender, race, and tumor histology (OR = 0.96, 95% CI 0.93-0.99).

Subsequent Use of Test Information:

[0287] *EGFR* mutation-positive patients were significantly more likely to have documented plans to receive subsequent *EGFR*-TKI treatment (86%) than *EGFR* mutation-negative patients (11%, $p < 0.001$). Clinicians documented that the *EGFR* results affected their prioritization of recommended therapies in 38% of cases. These cases included 14 (61%) of the 23 mutation-positive patients for whom *EGFR*-TKI therapy was recommended earlier than it would have been had the test been negative, and 24 (34%) of the 71 mutation-negative patients for whom *EGFR*-TKI therapy was not recommended, or was recommended later than it would have been had the test been positive.

[0288] *EGFR* mutation status was more likely to change prioritization of treatment options in patients with metastatic disease (54%) than in patients with local or locally advanced disease (19%, $p = 0.003$). Given this finding, we further analyzed the decision-making process in metastatic patients (Figure 10). Among the 31 patients with metastatic disease whose test results affected treatment recommendations, five mutation-positive patients were offered first-line *EGFR*-TKI treatment and six mutation-positive patients were offered second-line *EGFR*-TKI treatment in lieu of chemotherapy. Twenty mutation-negative patients were encouraged to defer *EGFR*-TKI treatment until third-line treatment or beyond based on their negative *EGFR* test results. Among the 26 patients with metastatic disease whose test results did not affect treatment recommendations, two mutation-negative patients received first-line *EGFR*-TKI treatment despite their negative results, nine patients including four mutation-positive patients received second or third-line *EGFR*-TKI treatment, and 15 patients including two mutation-positive patients did not receive a recommendation for an *EGFR*-TKI. Three of the patients with metastatic disease were participating in trials evaluating first-line *EGFR*-TKI therapy. Nine of the patients with metastatic disease had previously received or were receiving *EGFR*-TKIs at the time of *EGFR* testing.

DISCUSSION:

[0289] We studied the first 100 patients with NSCLC to undergo screening for somatic *EGFR* mutations as part of clinical cancer care at our institution and found that testing was feasible and significantly impacted the treatment of NSCLC patients. Patients harboring *EGFR* mutations were significantly more likely to receive recommendations for *EGFR*-TKI therapy than patients without mutations. Physicians adjusted their treatment recommendations based on the test results in over one-third of the cases, and were more likely to do so in patients with metastatic disease. In our patient sample, physicians used positive *EGFR* test results to help make the decision to prioritize *EGFR*-TKIs over chemotherapy for some patients, especially for first or second-line treatment. However, negative *EGFR* test results did not prevent physicians from administering *EGFR*-TKIs to selected patients. Many of the patients in whom the test result did not impact clinical decision-making had early stage, resected disease or were already receiving an *EGFR*-TKI for metastatic disease at the time of testing. This is reasonable since the utility of *EGFR*-TKIs as adjuvant therapy is not known and there is a benefit to *EGFR*-TKI therapy in a small number of patients without an identified *EGFR* mutation (65, 66-70, 71).

[0290] Our study also provides evidence that molecular diagnostics can enhance the clinical ability to identify patients with *EGFR* mutations. Many oncologists currently use the clinical characteristics associated with *EGFR* mutations and response to *EGFR*-TKIs to guide the decision-making process for patients with NSCLC. Indeed, our population of patients referred for *EGFR* testing demonstrated an increased prevalence of such characteristics. For example, 95% of referred patients had adenocarcinoma or BAC tumor histology, compared to 45% in the general NSCLC population (72). While never-smokers comprised 29% of our population, the incidence of never-smokers in the general NSCLC population has been reported as 2-10%, and may be as high as 27% in women with NSCLC (73-75). Similarly, our population consisted of only 17% current smokers, compared to the 38-75% rate of current smoking among newly diagnosed NSCLC patients (75, 78-80). Our clinically selected population consequently had an *EGFR* mutation rate of 24%, which is substantially higher than rates documented by our and other U.S. groups that tested unselected available NSCLC tumor samples (65-66, 81). However, it is important to note that while clinicians appeared to be attempting to select patients for testing that had the clinical characteristics predictive of *EGFR* mutations, the mutation frequency was still only 24%, highlighting the fact that molecular diagnostics increase the information available to make clinical decisions.

[0291] Smoking status was the strongest predictor of *EGFR* mutation status in our patients, with an increase in smoking history associated with a significantly decreased likelihood of harboring an *EGFR* mutation, after controlling for previously described predictors of mutation status. Our results are consistent with other case series documenting the importance of smoking status in the likelihood of *EGFR* mutations (66, 69, 70, 81, 82). Just as the extremely low prevalence of *EGFR* mutations in squamous cell tumors (62) has shifted testing efforts towards adenocarcinoma tumors, it may be appropriate to focus future efforts on patients with a low or absent smoking history. However, reports of *EGFR* mutations in patients without typical clinical characteristics advise against strict testing limitations (83). When examining the other clinical characteristics thought to be associated with mutations, we found Asian race and BAC tumor histology to have non-significant trends towards predicting *EGFR* mutation status. The lack of statistical significance in these associations may be due to small sample size.

[0292] The test was feasible and fit into the time constraints of clinical cancer care. Nearly all of the tumors submitted for analysis produced interpretable results. The six specimens that failed PCR amplification were all paraffin-embedded, while none of the frozen specimens failed PCR amplification. When available, fresh frozen tissue is the preferable substrate for *EGFR* mutation testing.

[0293] There have been close to 2,500 NSCLC samples reported thus far that have undergone partial or complete *EGFR* sequence analysis. Our patients demonstrated mutations similar to previous reports, with overlapping exon 19 deletions of 9-23 base pairs and point mutations leading to single amino acid substitutions in exons 18 and 21. Five of the point mutations we found have been described above (E709V, G719S, G719A, L858R, and L861Q). One of the point mutations we found causes an amino acid substitution at a codon where a different amino acid substitution has been previously described (R776H). The E709V and R776H variants were each found in combination with a known gefitinib-sensitizing mutation involving codon 719. The P848L mutation in exon 21 was found in both the somatic and buccal samples, suggesting it may be a germline variant of uncertain significance. The patient was a never-smoking female with adenocarcinoma who had stable disease for 15 months on gefitinib treatment, prior to the *EGFR* mutation testing. When the P848L mutation was revealed, she had recently been found to have progressive disease and was started on erlotinib therapy. No information about response to erlotinib is available at this time.

[0294] The (2253_2276 del) deletion overlaps previously described exon 19 deletions. The deletions in our patients can be categorized into one of two groups: those spanning codons 747-749 at a minimum (amino acid sequence LRE), and those spanning codons 752-759 (Figure 11). Analysis of all exon 19 deletions reported to date suggests that a wide variety of amino acids can be deleted from the TK region spanning codons 747-759. There does not appear to be a required common codon deleted; however, all of the deletions we detected maintained a lysine residue at position 745.

[0295] One of our two exon 20 mutations are in a never-smoking female with recurrent adenocarcinoma who was treated with erlotinib after *EGFR* testing was performed and has had stable disease for two months at this time. The other is a former-smoking male with metastatic adenocarcinoma who was treated with an *EGFR*-TKI, but could not tolerate it due to severe rash. The identification of clinically relevant *EGFR* mutations in exon 20 underscores the importance of comprehensive sequencing of the TK region of *EGFR*.

[0296] In conclusion, this study demonstrates the feasibility and utility of comprehensive screening of the TK region of the *EGFR* gene for somatic mutations in NSCLC patients as part of clinical cancer care. The result of the test provides useful information regarding clinical predictors of *EGFR*-TKI response. Current smokers are less likely to harbor a mutation, as are former smokers with a high number of pack-years of smoking history.

EXAMPLE 5

[0297] *EGFR* GENE TEST FOR NON-SMALL CELL LUNG CANCER, a Standard Operating Procedure.

Clinical Indications:

[0298] This test is indicated for individuals with Non-Small Cell Lung Cancer.

Analytical Principle

[0299] The *EGFR* Gene Test is a genetic test that detects mutations in the kinase domain of *EGFR*. DNA is first obtained from a tumor biopsy. The DNA sequence of 7 exons (18, 19, 20, 21, 22, 23, 24) of *EGFR* is then determined by direct bi-directional gene sequencing. The sequence obtained is then compared to known *EGFR* sequence to identify DNA sequence changes. If a DNA sequence change is detected in tumor tissue, the test will be repeated on the original tissue sample. If the change has not previously been reported in a gefitinib- or erlotinib- responder, the test will also be conducted with a sample of the individual's blood to determine whether the mutation is constitutive (and therefore likely a normally occurring polymorphism) or occurred somatically in the tumor tissue.

Specimen Requirements:

[0300] A minimum of 100 ng of DNA is required from tissue sample. Note: Extremely small quantities of DNA may be extracted from tissue samples. The concentration of this DNA may not be accurately quantitated.

Quality Control:

Controls used

[0301] Two negative controls (water) and a positive control (human DNA) for each exon are included in the PCR reactions. The negative control should proceed through the entire procedure to ensure that the sequence obtained is not the result of contamination. A pGEM positive control and an ABI array control are included in the sequencing step.

Control Preparation and Storage:

[0302] The positive control for PCR is either Clontech human DNA or human DNA from an anonymous blood sample and is stored at 4°C. The negative control for the PCR reaction is HyPure Molecular Biology Grade water stored at room temperature. The pGEM positive sequencing reaction control and the ABI array control are stored at -20°C.

Tolerance Limits and Steps to Take if Individual Control Fails:

[0303] If the positive PCR control fails but the negative controls and samples pass, the PCR results will be designated as pass and sequencing will be performed. If a negative control shows evidence of DNA amplification, the whole reaction will be repeated with a new aliquot of patient's DNA. If the pGEM control fails and the test reactions fail, the sequencing run will be repeated with a second aliquot of the PCR product. If the sequencing controls fail but the test reactions pass, the sequencing does not need to be repeated. NOTE: Due to the low yield of DNA extraction from paraffin embedded tissue samples, external PCR reactions often do not yield visible products. Internal PCR reactions should yield visible products. The size of the product detected on the gel should be compared to the anticipated sizes (see below) to ensure that the appropriate PCR product has been obtained. If an internal PCR product is not visible on the gel, exon-specific PCR failures should be repeated.

[0304] If PCR amplification for an individual sample fails, a new round of PCR should be attempted with a two-fold increase in input DNA template. If PCR amplification fails again, a new DNA sample for that patient should be acquired if available. If the sample was a paraffin-embedded tissue sample, additional slides should be scraped. If available, more slides than used to generate the original sample should be scraped and digestion in Proteinase K should be allowed to occur for three nights.

[0305] Equipment and Reagents (All reagents stable for one year unless otherwise noted.)

PCR and Sequencing (in general, PCR and sequencing equipment and reagents are known to those of skill in the art and may be used herein, also noted above).

[0306] Primers: (see Table 6 and 7 below)

Table 6: External PCR Primers:

Exon	Forward Primer Sequence, (5'→3')	SEQ ID NOS	Reverse Primer Sequence, (5'→3')	SEQ ID NOS
18	TCAGAGCCTGTGTTTCTACCAA	653	TGGTCTCACAGGACCACTGATT	646
19	AAATAATCAGTGTGATTTCGTGGAG	654	GAGGCCAGTGCTGTCTCTAAGG	647
20	ACTTCACAGCCCTGCGTAAAC	655	ATGGGACAGGCACTGATTTGT	648
21	GCAGCGGGTTACATCTTCTTTC	656	CAGCTCTGGCTCACACTACCAG	649
22	CCTGAACTCCGTCAGACTGAAA	657	GCAGCTGGACTCGATTTCCT	650
23	CCTTACAGCAATCCTGTGAAACA	658	TGCCCAATGAGTCAAGAAGTGT	651
24	ATGTACAGTGCTGGCATGGTCT	659	CACTCACGGATGCTGCTTAGTT	652

TABLE 7: Internal PCR Primers:

Exon	Forward Primer Sequence, (5'→3')	Reverse Primer Sequence, (5'→3')	Product Length (bp)
18	TCCAAATGAGCTGGCAAGTG (SEQ ID NO 660)	TCCCAAACACTCAGTGAAACAAA (SEQ ID NO 667)	397
19	GTGCATCGCTGGTAACATCC (SEQ ID NO 661)	TGTGGAGATGAGCAGGGTCT (SEQ ID NO 668)	297
20	ATCGCATTCATGCGTCTTCA (SEQ ID NO 662)	ATCCCCATGGCAAACCTCTTG (SEQ ID NO 669)	378
21	GCTCAGAGCCTGGCATGAA (SEQ ID NO 663)	CATCCTCCCCTGCATGTGT (SEQ ID NO 670)	348
22	TGGCTUGTCTGTGTGTGTCA (SEQ ID NO 664)	CGAAAGAAAATACTTGCATGTCAGA (SEQ ID NO 671)	287
23	TGAAGCAAATTGCCCAAGAC (SEQ ID NO 665)	TGACATTTCTCCAGGGATGC (SEQ ID NO 672)	383
24	AAGTGTGCGCATCACCAATGC (SEQ ID NO 666)	ATGCGATCTGGGACACAGG (SEQ ID NO 673)	302
F primer linker	tgtaaacgacgcccagt (SEQ ID NO 645)	5' end of all forward primers	18
R primer linker	aacagctatgacctg (SEQ ID NO 674)	5' end of all reverse primers	16

Precautions

[0307]

Table 8

Task	Instruction(s)	Risk
1. PCR Setup	Use PCR Hood Use dedicated pipets and filtered tips Only open reagents in the hood	Contamination of PCR reaction
2. Use of PCR Hood	Do not use any post-PCR samples or reagents in the hood	Contamination of PCR reaction

Preparing PCR Reaction Mix for External PCR

[0308] All procedures performed in PCR hood for genomic DNA, not the clean hood.

1. Thaw out Taq Gold and dNTP on ice.
2. Prepare the master mix in a tube (eppendorf or 15mL tubes) using the table below. Water, Betaine, 10X Buffer, MgCl₂, DMSO, Taq Gold and dNTP should be added in the order listed. It is very important to mix the reagents by pipetting up-and-down gently while adding each reagent.
3. DNA should be added to the master mix before aliquoting. After making the large volume of master mix, aliquot 96 ul (enough for 8 rxns) to a separate tube for each patient or control. Add 8 ul of DNA at 5 ng/ul to the 96 ul of mastermix. 13ul can then be added to the individual wells of the plate or put in strip tubes and pipetted with a multi-channel pipettor.
4. For a full 384-well plate of reactions, make enough master mix for about 415 reactions.
5. Spin the plate of master mix to get rid of air bubbles.
6. If using a large set of primers, it would help to have them in 96-well plates with forward primers and reverse

primers in separate plates.

7. Add the primers using a multi-channel pipette. Make sure to mix by pipetting up-and-down gently.

8. Spin the plate to get rid of any air bubbles.

9. Use the cycle below to amplify.

Note: PCR is done in 384-well plates.

TABLE 9

Reagent	Volume per reaction (μL)
Autoclaved ddH ₂ O	4.90
5M Betaine	3.00
10X Buffer	1.50
Magnesium Chloride	1.50
DMSO	0.75
Taq	0.20
dNTP	0.15
PCR Forward Primer1 (conc. 20pmol/uL)	1.00
PCR Reverse Primer2 (conc. 20pmol/uL)	1.00
DNA (conc. 5ng/uL)	1.00
Total volume of PCR reaction	15.00

TABLE 10: PCR Amplification Cycle

Activate Taq Gold	10 minutes	95 °C	
Denature	30 seconds	95 °C	30 cycles
Anneal	30 seconds	60 °C	
	1 minutes	72 °C	
Extend	10 minutes	72 °C	
Hold	∞	4 °C	
Note: A cleanup is not necessary after performing the external PCR.			

Preparing PCR Reaction Mix for Internal PCR

[0309] The internal PCR set up is almost the same as the external PCR with a few exceptions.

1. Make the large volume of master mix as described for external PCR in the PCR hood.
2. Aliquot MM to 7 strip tubes and multichannel pipette 12ul into the 384-well plate.
3. Add 1ul each of forward and reverse internal primers. Temporarily seal plate.
4. Remove from hood, spin down plate and proceed to post PCR set-up area.
5. Use dedicated pipettes to aliquot 1ul of external PCR product into each reaction.
6. Heat seal and spin again.
7. Run same amplification cycle as external.

[0310] Run PCR products on a 1% gel before clean-up. Determine Pass/Failed exons for repeat PCR.

[0311] Clean-up Internal PCR Using Ampure Magnetic Bead Clean-up

Cleanup

[0312]

1. Vortex the plate of Ampure magnetic beads till there is no deposit of beads.
2. It is very important that the temperature of the Ampure beads is at room temperature.
3. Use the 384-well Ampure protocol on the Biomek and change the volume of reaction to 12uL to accommodate reagents used for cleanup. If this is not done, an error will be generated.
4. After the program is complete, hydrate plate with 20 uL of autoclaved ddH₂O per well. While adding water, make sure to mix by pipetting up-and-down gently.
5. Spin the plate to get rid of any air bubbles.
6. Place the plate on a magnet to separate out the beads. Now you should be able to take up 1 uL of the DNA to setup sequencing reactions. Save the rest at -20°C for future use.

Sequencing Protocol

Preparing Sequencing Reaction Mix

[0313]

1. Thaw out BigDye 3.1 in a dark place, on ice.
2. Prepare the master mix in a tube (eppendorf or 15mL tubes) using the table below. Water, buffer, DMSO and BigDye should be added in the order listed.
3. It is very important to mix the reagents by pipetting up-and-down gently while adding each reagent.
4. When using a universal primer for sequencing, the primer can also be added to the master mix at this time. If the primer is unique it should be added individually after the master mix is in the 384-well plate.
5. Usually for a full 384-well plate of reactions, make enough master mix for about 415 reactions.
6. Once the master mix is setup divide the mix into 8 wells of strip tubes. (Do not use reservoirs to aliquot master mix. That would be a waste of reagents.)
7. Now a multi-channel pipette can be used to aliquot the master mix into the 384-well plate
8. Spin the plate of master mix to get rid of air bubbles.
9. Add the PCR product to be sequenced, using a multi-channel pipette. Make sure to mix by pipetting up-and-down.
10. Spin the plate to get rid of any air bubbles.
11. Use the cycle below to amplify.

TABLE 11

<i>Reagent</i>	<i>Volume per reaction (μL)</i>
Autoclaved ddH ₂ O	4.38
5X ABI Buffer	3.65
DMSO	0.50
ABI BigDye 3.1	0.35
Sequencing Primer concentration	0.12
DNA from Internal PCR reaction	1.00
Total Volume of reaction	10.00

TABLE 12: Amplification Cycle for Sequencing

Denature	10 seconds	96 °C	25 cycles
Anneal	5 seconds	50 °C	
Extend	4 minutes	60 °C	
Hold	∞	4 °C	

Clean-up via Cleanseq Magnetic Bead Clean-up

[0314]

1. Vortex the plate of Cleanseq magnetic beads till there is no deposit of beads.
2. Use the Cleanseq 384-well plate program on the Biomek to clean-up the samples.
3. Once the program is done, save the original plate at -20 °C. The new plate with the clean samples is ready to go on the ABI 3730.

[0315] (Note: If the PCR products are shorter than 300 bps you might have to dilute the sample before putting it on the 3730)

Create Mutation Surveyor templates for the EGFR test and save them on LMM/Sequencing/ Sequences-MS Review/EGFR.

Repeat Result Criteria

[0316] All positive results are repeated by amplifying and sequencing the specific exon(s) in which a DNA sequence change has been detected from a second aliquot of patient DNA derived from the original tissue sample. In addition, DNA extracted from a sample of the patient's blood should be run in parallel to compare with tumor tissue if the sequence change detected has not previously been detected in a gefitinib- or erlotinib-responder.

[0317] Any exon that did not produce clear sequence will be repeated either from extraction, PCR or sequencing, based on the specific technical problems.

Assay Parameters

[0318] Sensitivity of the Test - Somatic *EGFR* kinase domain mutations have been found in approximately 13% of individuals with NSCLC (Paez JG et al., 2004). In addition, somatic *EGFR* kinase domain mutations have been found in 13/14 (92.8%) individuals with NSCLC that were gefitinib-responsive (Paez JG et al., 2004, Lynch, et al., 2004). Validation of the technical sensitivity of the test demonstrated 100% sensitivity to known mutations and validation of the sequencing platform in our lab shows 100% sensitivity (see "Accuracy of the Technique" below). The sensitivity for mutation detection of mosaic samples has been determined to be 25% (ie, heterozygous mutations can be detected when present at 50% of a cell mixture). We have found that up to 20% of paraffin-embedded tissue do not yield high quality DNA. We are unable to obtain sequence information from these samples.

[0319] Specificity of the Test - To date, published literature indicates that no individuals with a somatic mutation in *EGFR* were not responsive to gefitinib (11/11). The chance of finding a mutation due to an artifact of bi-directional sequencing is close to 0% (see "Accuracy of the Technique" below). As such, the specificity of the test is approximately 100%.

[0320] Accuracy of the Technique - The technique of DNA sequencing is the gold standard in molecular diagnostics. This lab uses the ABI 3730 DNA Analyzer that has a reported accuracy of 98.5%. Combining this with bi-directional sequencing, automated chromatogram analysis with Mutation Surveyor, and manual analysis of false positives, we have achieved an accuracy rate of 100%. This is based upon an analysis of over 100,000 bases of raw sequence. For details of this assessment, see our Quality Assurance Program manual.

[0321] Note: We do not assume that these results guarantee 100% accuracy of this platform. It is known that sequencing errors can occur and, as such, we report our accuracy to be 99.99% that has been found by large scale sequencing projects (Hill et al. 2000).

[0322] Reproducibility of the Test - Due to the accuracy of the test, when results are achieved, they are reproducible equal to the accuracy of the test (99.99%). However, on occasion, the test can fail due to factors listed below (see Limitations of Method) or because of PCR or sequencing failure due to unexplained technical reasons. In these cases, no results are achieved and the assay is repeated until a result is achieved or the patient specimen is deemed unacceptable. Specific rates of failure of each assay step and of specimens can be found in the validation reports in our Quality Assurance Program manual.

[0323] Normal Range of the Results - The normal sequence of the *EGFR* gene can be found online using GenBank accessions: NT_033968.5 (genomic sequence) and NM_005228.3 (mRNA sequence).

Limitations of Method:

[0324] Large deletions spanning one or more exons will not be detected by the sequencing method, particularly if present in heterozygosity- Mutations in the *EGFR* gene outside of the kinase domain will not be detected by this assay. Inhibitors may be present in the DNA sample preventing amplification by PCR. Degraded DNA may not produce analyzable data and re-submission of the specimen may be required. Rare sequence variations or secondary structures of the targeted primer sequences could affect PCR amplification and therefore mutation(s) could be missed in that region of one allele.

EXAMPLE 6

[0325] Gefitinib (Iressa) is a tyrosine kinase inhibitor that targets the epidermal growth factor receptor (EGFR), and induces dramatic clinical responses in non-small cell lung cancers (NSCLCs) with activating mutations within the EGFR kinase domain. We report that these mutant EGFRs selectively activate Akt and STAT signaling pathways, which promote cell survival, but have no effect on Erk/MAPK signaling, which induces proliferation. NSCLCs expressing mutant EGFRs underwent extensive apoptosis following siRNA-mediated knockdown of the mutant EGFR or treatment with pharmacological inhibitors of Akt and STAT signaling, and were relatively resistant to apoptosis induced by conventional chemotherapeutic drugs. Thus, mutant EGFRs selectively transduce survival signals on which NSCLCs become dependent; consequently, inhibition of those signals by Gefitinib may underlie striking clinical responses.

[0326] Receptor tyrosine kinases of the EGFR family regulate essential cellular functions including proliferation, survival, migration, and differentiation, and appear to play a central role in the etiology and progression of solid tumors (R. N. Jorissen et al., *Exp. Cell Res.* 284, 31 (2003), H. S. Earp, T. L. Dawson, X. Li, H. Yu, *Breast Cancer Res. Treat.* 35, 115 (1995)). EGFR is frequently overexpressed in breast, lung, colon, ovarian, and brain tumors, prompting the development of specific pharmacological inhibitors, such as Gefitinib, which disrupts EGFR kinase activity by binding the ATP pocket within the catalytic domain (A- E. Wakeling et al., *Cancer Res.* 62, 5749 (2002)). Gefitinib has induced dramatic clinical responses in approximately 10% of patients with chemotherapy-refractory NSCLC (J. Baselga et al., *J. Clin. Oncol.* 20, 4292 (2002), M. Fukuoka et al., *J. Clin. Oncol.* 21, 2237 (2003), G. Giaccone et al., *J Clin Oncol.* 22, 777 (2004), M. G. Kris et al., *JAMA* 290, 2149 (2003)). Virtually all Gefitinib-responsive lung cancers harbor somatic mutations within the EGFR kinase domain, whereas no mutations have been seen in non-responsive cases (T. J. Lynch et al., *N. Engl. J. Med.* 350, 2129 (2004), J. G. Paez et al., *Science* 304, 1497 (2004).) These heterozygous mutations include small in-frame deletions and missense substitutions clustered within the ATP-binding pocket.

[0327] Using transient transfections of mutant EGFRs, we showed previously that both types of mutations lead to increased EGF-dependent receptor activation, as measured by autophosphorylation of Y1068, one of the prominent C-terminal phosphorylation sites of EGFR. (T. J. Lynch et al., *N. Engl. J. Med.* 350, 2129 (2004)).

[0328] To enable studies of qualitative differences in signaling by mutant EGFRs, we generated stable lines of non-transformed mouse mammary epithelial cells (NMuMg) expressing wild-type or mutant EGFRs, and analyzed EGF-mediated autophosphorylation of multiple tyrosine residues linked to activation of distinct downstream effectors (R. N. Jorissen et al., *Exp. Cell Res.* 284, 31 (2003)). Cell lines were generated that expressed either wild-type EGFR or one of two recurrent mutations detected in tumors from Gefitinib-responsive patients: the missense mutation L858R and the 18bp in-frame deletion, de1L747-P753insS. Significantly different tyrosine phosphorylation patterns were observed between wild-type and the two mutant EGFRs at several C-terminal sites. EGF-induced phosphorylation of Y1045 and Y1173 was virtually indistinguishable between wild-type and mutant EGFRs, whereas phosphorylation of Y992 and Y1068 was substantially increased in both mutants. Interestingly, Y845 was highly phosphorylated in the L858R missense mutant, but not in the wild-type or the deletion mutant, and hence appears to be unique in distinguishing between the two types of EGFR mutations. The differential EGF-induced tyrosine phosphorylation pattern seen with wild-type and mutant receptors was reproducible in transiently transfected COS7 cells, ensuring against potential cell type specific effects.

[0329] Thus, Gefitinib-sensitive mutant EGFRs transduce signals that are qualitatively distinct from those mediated by wild-type EGFR. These differences may result directly from structural alterations within the catalytic pocket affecting substrate specificity, or from altered interactions with accessory proteins that modulate EGFR signaling.

[0330] The establishment of cell lines stably transfected with mutant EGFRs made it possible to compare the phosphorylation status of the major downstream targets of EGFR in a shared cellular background. EGF-induced activation of Erk1 and Erk2, via Ras, of Akt via PLC γ /PI3K, and of STAT3 and STAT5 via JAK2, are essential downstream pathways mediating oncogenic effects of EGFR (R. N. Jorissen et al., *Exp. Cell Res.* 284, 31 (2003)). EGF-induced Erk activation was essentially indistinguishable among cells expressing wild-type EGFR or either of the two activating EGFR mutants. In contrast, phosphorylation of both Akt and STAT5 was substantially elevated in cells expressing either of the mutant EGFRs. Increased phosphorylation of STAT3 was similarly observed in cells expressing mutant EGFRs. The unaltered Erk activation by the mutant EGFRs is consistent with the absence of increased phosphorylation of Y1173, an important

docking site for the Shc and Grb-2 adaptors that leads to Ras activation and subsequent Erk phosphorylation (R. N. Jorissen et al., Exp. Cell Res. 284, 31 (2003)). The increased Akt and STAT phosphorylation following activation of the mutant EGFRs is consistent with the increase in Y992 and Y1068 phosphorylation, both of which have been previously linked to Akt and STAT activation (R. N. Jorissen et al., Exp. Cell Res. 284, 31 (2003)). Thus, the selective EGF-induced autophosphorylation of C-terminal tyrosine residues within EGFR mutants is well correlated with the selective activation of downstream signaling pathways.

[0331] To extend these observations to lung cancer cells in which *EGFR* mutations appear to drive tumorigenesis, we studied lines derived from five NSCL tumors. NCI-H1975 carries the recurrent heterozygous missense mutation L858R and NCI-H1650 has the in-frame deletion delE746-A750, whereas NCI-358, NCI-H1666, and NCI-H1734 express wild-type *EGFR*. As in transfected cells, EGF-induced autophosphorylation of Y992 and Y1068 was markedly elevated in the two lines with endogenous *EGFR* mutations, as was phosphorylation of Akt and STAT5, but not Erk.

[0332] The oncogenic activity of *EGFR* reflects the activation of signals that promote both cell proliferation and cell survival (S. Grant, L. Qiao, P. Dent, Front. Biosci. 7, d376 (2002)). While these pathways exhibit overlap, Ras-mediated activation of the Erk kinases contributes substantially to the proliferative activity of *EGFR*, whereas activation of Akt and STATs is largely linked to an anti-apoptotic function (S. Grant, L. Qiao, P. Dent, Front. Biosci. 7, d376 (2002), F. Chang et al., Leukemia 17, 1263 (2003), F. Chang et al., Leukemia 17, 590 (2003), F. Chang et al., Int. J. Oncol. 22, 469 (2003), V. Calo et al., J. Cell Physiol. 197, 157 (2003), T. J. Ahonen et al., J. Biol. Chem. 278, 27287 (2003)). The two lung cancer cell lines harboring *EGFR* mutations exhibited a proliferative response to EGF at low serum concentrations that was not observed in cells with wild-type receptors. However, their proliferation rate and cell density at confluence were comparable at normal serum concentrations.

SiRNA

[0333] In contrast, apoptotic pathways were markedly different in lung cancer cells with mutant EGFRs: siRNA-mediated specific inactivation of mutant *EGFR* in these cell lines resulted in rapid and massive apoptosis. About 90% of NCI-H1975 cells transfected with L858R-specific siRNA died within 96 hours, as did NCI-H1650 cells transfected with delE746-A750-specific siRNA. SiRNA specific for either *EGFR* mutation had no effect on cells expressing the alternative mutation, and siRNA that targets both wild-type and mutant *EGFR* had minimal effect on the viability of cells expressing only wild-type receptor, but induced rapid cell death in lines expressing *EGFR* mutants. The ability of siRNAs to specifically target the corresponding *EGFR* alleles was confirmed in transfected COS7 cells by immunoblotting. Thus, expression of mutant EGFRs appears essential for suppression of pro-apoptotic signals in lung cancers harboring these mutations. The fact that lung cancer cells expressing only wild-type receptors do not display a similar dependence on *EGFR* expression may also account for the relative Gefitinib-insensitivity of human tumors that overexpress wild-type *EGFR*.

[0334] The effectiveness of Gefitinib in lung cancers harboring mutant EGFRs may reflect both its inhibition of critical anti-apoptotic pathways on which these cells have become strictly dependent, as well as altered biochemical properties of the mutant receptors. We previously reported that mutant EGFRs are more sensitive to Gefitinib inhibition of EGF-dependent autophosphorylation than wild-type receptors (T. J. Lynch et al., N. Engl. J. Med. 350, 2129 (2004)). This increased drug sensitivity by mutant receptors was also observed for both Erk and STAT5 activation. Thus, while EGF-induced signaling by mutant receptors demonstrates selective activation of downstream effectors via differential autophosphorylation events, their enhanced inhibition by Gefitinib is uniform, and may reflect altered drug binding to the mutant ATP pocket.

[0335] To establish the relevance of increased Akt and STAT signaling in *EGFR*-mediated NSCLC survival, we targeted these pathways with specific pharmacological inhibitors. Lung cancer cells harboring *EGFR* mutations were 100-fold more sensitive to Gefitinib than cells with wild-type receptor. Cells expressing mutant EGFRs were also more sensitive to pharmacological inhibition of Akt or STAT signaling than cells expressing only wild-type *EGFR*. While *EGFR*-mutant lung cancer cells exhibited increased sensitivity to disruption of Akt/STAT-mediated anti-apoptotic signals, they demonstrated markedly increased resistance to cell death signals induced by the commonly used chemotherapeutic agents doxorubicin and cisplatin, and the pro-apoptotic Fas-ligand.

[0336] Enhanced Akt/STAT signaling in cells with mutant *EGFR* might therefore provide an additional therapeutic target, while raising the possibility that conventional chemotherapy may be less effective against these tumors.

[0337] "Oncogene addiction" has been proposed to explain the apoptosis of cancer cells following suppression of a proliferative signal on which they have become dependent (I. B. Weinstein, Science 297, 63 (2002)). Interestingly, Imatinib (Gleevec) efficiently triggers cell death in chronic myeloid leukemias expressing the BCR-ABL translocation product and in gastrointestinal stromal tumors expressing activating c-Kit mutations, both of which exhibit frequently constitutive STAT activation that is effectively inhibited by the drug (T. Kindler et al., Leukemia 17, 999 (2003), G. P. Paner et al., Anticancer Res. 23, 2253 (2003)). Similarly, in lung cancer cells with *EGFR* kinase mutations, Gefitinib-responsiveness may result in large part from its effective inhibition of essential anti-apoptotic signals transduced by the mutant receptor.

Materials and Methods

Immunoblotting

5 **[0338]** Lysates from cultured cells were prepared in ice-cold RIPA lysis solution (1% Triton X-100, 0.1% SDS, 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 10 mM β -glycerol-phosphate, 10 mM NaF, 1 mM Naorthovanadate, containing protease inhibitors. Debris was removed by centrifugation in a microfuge at 12,000 x g for 10 min at 4°C. Clarified lysates were boiled in gel loading buffer and separated by 10% SDS-PAGE. Proteins were electrotransferred to nitrocellulose and detected with specific antibodies followed by incubation with horseradish peroxidase-conjugated
10 secondary goat antibody (Cell signaling (Beverly, MA; 1:2000) and development with enhanced chemiluminescence (DuPont NEN) followed by autoradiography. The phospho-EGFR Y845, Y992, Y1045, Y1068, phospho-STAT5 (tyr694), phospho-AKT(Ser473), phospho-ERK1/2(Thr202/Tyr204), AKT, STAT5, and ERK1/2 antibodies were obtained from New England Biolabs (Beverly, MA). The total EGFR Ab-20 antibody was obtained from NeoMarkers (Fremont, CA). The phospho-EGFR Y1173 antibody was from Upstate Biotechnology (Lake Placid, NY) and the total phosphotyrosine
15 antibody PY-20 was from Transduction Laboratories (Lexington, KY). All antibodies were used at a 1:1000 dilution.

EGFR expression vectors

20 **[0339]** Full-length EGFR expression constructs encoding the wild type, L858R or del L747-P753insS mutations were sub-cloned using standard methods into plasmid pUSEamp. All constructs were confirmed by DNA sequence analysis.

Cell lines and transfections

25 **[0340]** COS7 cells and NMuMg (normal mouse mammary epithelial) cells were grown in DMEM (Dulbecco's modified Eagle's media) with 10% fetal calf serum in the presence of 2mM L-glutamine and 50 U/ml penicillin/streptomycin. The NCI-H358, NCI-H1650, NCI-H1734, NCI-H1666, and NCI-H1975 human lung cancer cell lines were obtained from the American Type Culture Collection collection and were grown in RPMI1640 with 10% fetal bovine serum, 2mM L-glutamine, 50 U/ml penicillin/streptomycin and 1mM sodium pyruvate. They are referred to in the text, in an abbreviated manner,
30 as H358, H1650, H1734, H1666, and H1975, respectively. Transient transfection of COS7 cells was performed using Lipofectamine 2000 (Invitrogen; Carlsbad, CA). Plasmid (1 μ g) was transfected into cells at 80% confluence on a 10 cm dish. After 12 hours, the cells were harvested and reseeded in 12-well plates in the absence of serum. The following day, cells were stimulated with 30ng/ml of EGF. Stable NMuMg cell lines were prepared by co-transfecting the EGFR expression constructs with the drug-selectable plasmid pBABE puro, followed by selection in 3 μ g/ml puromycin. Pools of drug-resistant cells were used for analysis. Expression of EGFR in stably transfected cells was confirmed by immu-
35 noblotting.

SiRNA-mediated "knockdown" of EGFR expression

40 **[0341]** SiRNA for EGFR L858R was designed to target the nucleotide sequence CACAGATTTTGGCGGGCCAA (SEQ ID NO.: 688), while the GCTATCAAAACATCTCCGAAA (SEQ ID NO.: 689) sequence was used for the delE745-A750 (Qiagen; Valencia, CA). To target all forms of EGFR, commercially prepared siRNA corresponding to human wild-type EGFR was obtained from Dharmacon (Lafayette, CO). Transfection of siRNAs was performed with Lipofectamine 2000 (Invitrogen) as per the manufacturer's instructions. Cells were assayed for viability after 96 hours using the MTT
45 assay.

Apoptosis assay

50 **[0342]** 10,000 cells were seeded into individual wells of a 96-well plate. After 6 hours, the medium was changed and the cells were maintained in the presence of increasing concentrations of doxorubicin (Sigma; St. Louis, MO), cisplatin (Sigma), Fas-ligand (human activating, clone CH11; Upstate Biotechnology), Ly294002 (Sigma), or AG490 (Calbiochem; La Jolla, CA). After 96 hours, the viability of cells was determined using the MTT assay. For caspase immunostaining, 10,000 cells were seeded onto 10 mm coverslips. The next day they were transfected with siRNA (see previous section for details). After 72 hours the cells were fixed in 4% paraformaldehyde at room temperature for 10 min. They were subsequently permeabilized for 5 min in 0.5% Triton X-100 and blocked for 1hr in 5% normal goat serum (NGS). The
55 coverslips were then incubated overnight at 4°C in primary antibody (cleaved caspase-3 Asp175 5A1 from Cell Signaling) at a 1:100 dilution. The next day the coverslips were washed 3 times in PBS and incubated for 1 hour with secondary antibody (goat anti-rabbit Texas-red conjugated; from Jackson ImmunoResearch; West Grove, PA) at a 1:250 dilution in 5% normal goat serum and 0.5 μ g/ml of DAPI (4',6-Diamidino-2-phenylindole). After 3 washes in PBS the coverslips

were mounted with ProLong Gold anti-fade reagent from Molecular Probes (Eugene, OR).

Cell viability assay

- 5 **[0343]** 10 μ l of 5mg/ml MTT (Thiazolyl blue; Sigma) solution was added to each well of a 96-well plate. After 2 hours of incubation at 37°C, the medium was removed and the MTT was solubilized by the addition of 100 μ l of acidic isopropanol (0.1N HCL) to each well. The absorbance was determined spectrophotometrically at 570nm.

Growth curve

- 10 **[0344]** Growth curves for H-358, H-1650, H-1734, and H-1975 cells were obtained by seeding 1000 cells in individual wells of 96-well plates. Each cell line was plated in 8 separate wells. On consecutive days, the cells were fixed in 4% formaldehyde and stained with 0.1 % (w/v) crystal violet solution. The crystal violet was then solubilized in 100 μ l of 10% acetic acid, and the absorbance was measured at 570nm using a plate reader to determine the relative cell number.

Mutation identification

- 15 **[0345]** To identify sporadic NSCLC cell lines harboring mutations within EGFR, we sequenced exons 19 and 21 within a panel of 15 NSCLC cell-lines, as described above. Cell lines were selected for analysis based on their derivation from tumors of bronchoalveolar histology irrespective of smoking history (NCI-H358, NCI-H650, NCI-H1650), or from adenocarcinomas arising within non-smokers (NCI-H1435, NCI-H1563, NCI-H1651, NCI-H1734, NCI-H1793, NCI-H1975, NCI-H2291, NCI-H2342, NCI-H2030, NCI-H1838, NCI-H2347, NCI-H2023). NCI-H1666 has been reported to harbor only wild type EGFR (see examples above). All cell lines are available from the American Type Culture Collection.

REFERENCES

[0346]

- 30 1. Schiller JH, Harrington D, Belani CP, et al. Comparison of four chemotherapy regimens for advanced non-small cell lung cancer. *N Engl J Med* 2002; 346:92-98.
2. Druker BJ, Talpaz M, Resta DJ et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in Chronic Myeloid Leukemia. *N Engl J Med* 2001;344:1031-1037.
3. Arteaga CL. ErbB-targeted therapeutic approaches in human cancer. *Exp Cell Res*. 2003; 284:122-30.
- 35 4. Jorissen RN, Walker F, Pouliot N, Garrett TP, Ward CW, Burgess AW. Epidermal growth factor receptor: mechanisms of activation and signaling. *Exp Cell Res* 2003;284:31-53
5. Luetkeke NC, Phillips HK, Qui TH, Copeland NG, Earp HS, Jenkins NA, Lee DC. The mouse waved-2 phenotype results from a point mutation in the EGF receptor tyrosine kinase. *Genes Dev* 1994;8:399-413.
6. Nicholson RI, Gee JMW, Harper ME. EGFR and cancer prognosis. *Eur J Cancer*. 2001;37:S9-15
7. Wong AJ, Ruppert JM, Bigner SH, et al. Structural alterations of the epidermal growth factor receptor gene in human gliomas. *Proc Natl Acad Sci*. 1992;89:2965-2969.
- 40 8. Ciesielski MJ, Genstermaker RA. Oncogenic epidermal growth factor receptor mutants with tandem duplication: gene structure and effects on receptor function. *Oncogene* 2000; 19:810-820.
9. Frederick L, Wang W-Y, Eley G, James CD. Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas. *Cancer Res* 2000; 60:1383-1387.
- 45 10. Huang H-JS, Nagane M, Klingbeil CK, et al. The enhanced tumorigenic activity of a mutant epidermal growth factor receptor common in human cancers is mediated by threshold levels of constitutive tyrosine phosphorylation and unattenuated signaling. *J Biol Chem* 1997;272:2927-2935
11. Pegram MD, Konecny G, Slamon DJ. The molecular and cellular biology of HER2/neu gene amplification/over-expression and the clinical development of herceptin (trastuzumab) therapy for breast cancer. *Cancer Treat Res* 2000;103:57-75.
- 50 12. Ciardiello F, Tortora G. A novel approach in the treatment of cancer targeting the epidermal growth factor receptor. *Clin Cancer Res*. 2001;7:2958-2970
13. Wakeling AE, Guy SP, Woodburn JR et al. ZD1839 (Iressa): An orally active inhibitor of Epidermal Growth Factor signaling with potential for cancer therapy. *Cancer Res* 2002;62:5749-5754.
- 55 14. Moulder SL, Yakes FM, Muthuswamy SK, Bianco R, Simpson JF, Arteaga CL. Epidermal growth factor receptor (HER1) tyrosine kinase inhibitor ZD1839 (Iressa) inhibits HER2/neu (erbB2)-overexpressing breast cancer cells in vitro and in vivo. *Cancer Res* 2001;61:8887-8895.
15. Moasser MM, Basso A, Averbuch SD, Rosen N. The tyrosine kinase inhibitor ZD1839 ("Iressa") inhibits HER2-

driven signaling and suppresses the growth of HER-2 overexpressing tumor cells. *Cancer Res* 2001;61:7184-7188.

16. Ranson M, Hammond LA, Ferry D, et al. ZD1839, a selective oral epidermal growth factor receptor-tyrosine kinase inhibitor, is well tolerated and active in patients with solid, malignant tumors: results of a phase I trial. *J Clin Oncol*. 2002; 20: 2240-2250.

17. Herbst RS, Maddox A-M, Rothernberg ML, et al. Selective oral epidermal growth factor receptor tyrosine kinase inhibitor ZD1839 is generally well tolerated and has activity in non-small cell lung cancer and other solid tumors: results of a phase I trial. *J Clin Oncol*. 2002;20:3815-3825.

18. Baselga J, Rischin JB, Ranson M, et al. Phase I safety, pharmacokinetic and pharmacodynamic trial of ZD1839, a selective oral Epidermal Growth Factor Receptor tyrosine kinase inhibitor, in patients with five selected solid tumor types. *J Clin Oncol* 2002;20:4292-4302.

19. Albanell J, Rojo F, Averbuch S, et al. Pharmacodynamic studies of the epidermal growth factor receptor inhibitor ZD1839 in skin from cancer patients: histopathologic and molecular consequences of receptor inhibition. *J Clin Oncol*. 2001; 20:110-124.

20. Kris MG, Natale RB, Herbst RS, et al. Efficacy of Gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: A randomized trial. *JAMA* 2003;290:2149-2158.

21. Fukuoka M, Yano S, Giaccone G, et al. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer. *J Clin Oncol* 2003;21:2237-2246.

22. Giaccone G, Herbst RS, Manegold C, et al. Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: A phase III trial-INTACT 1. *J Clin Oncol* 2004;22:777-784.

23. Herbst RS, Giaccone G, Schiller JH, et al. Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: A phase III trial - INTACT 2. *J Clin Oncol* 2004;22:785-794.

24. Rich JN, Reardon DA, Peery T, et al. Phase II Trial of Gefitinib in recurrent glioblastoma. *J Clin Oncol* 2004;22:133-142

25. Cohen MH, Williams GA, Sridhara R, et al. United States Food and Drug Administration Drug Approval Summary: Gefitinib (ZD1839;Iressa) Tablets. *Clin Cancer Res*. 2004;10:1212-1218.

26. Cappuzzo F, Gregorc V, Rossi E, et al. Gefitinib in pretreated non-small-cell lung cancer (NSCLC): Analysis of efficacy and correlation with HER2 and epidermal growth factor receptor expression in locally advanced or Metastatic NSCLC. *J Clin Oncol*. 2003;21:2658-2663.

27. Fitch KR, McGowan KA, van Raamsdonk CD, et al. Genetics of Dark Skin in mice. *Genes & Dev* 2003;17:214-228.

28. Nielsen UB, Cardone MH, Sinskey AJ, MacBeath G, Sorger PK. Profiling receptor tyrosine kinase activation by using Ab microarrays. *Proc Natl Acad Sci USA* 2003;100:9330-5.

29. Burgess AW, Cho H, Eigenbrot C, et al. An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol Cell* 2003;12:541-552.

30. Stamos J, Sliwkowski MX, Eigenbrot C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J Biol Chem*. 2002;277:46265-46272.

31. Lorenzato A, Olivero M, Patrane S, et al. Novel somatic mutations of the MET oncogene in human carcinoma metastases activating cell motility and invasion. *Cancer Res* 2002; 62:7025-30.

32. Davies H, Bignell GR, Cox C, et al. Mutations of the BRAF gene in human cancer. *Nature* 2002;417:906-7.

33. Bardelli A, Parsons DW, Silliman N, et al. Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* 2003;300:949.

34. Daley GQ, Van Etten RA, Baltimore D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science* 1990;247:824-30.

35. Heinrich MC, Corless CL, Demetri GD, et al. Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J Clin Oncol* 2003;21:4342-4349.

36. Li B, Chang C, Yuan M, McKenna WG, Shu HG. Resistance to small molecule inhibitors of epidermal growth factor receptor in malignant gliomas. *Cancer Res* 2003;63:7443-7450.

37. C. L. Sawyers, *Genes Dev* 17, 2998-3010 (2003).

38. G. D. Demetri et al., *N Engl J Med* 347, 472-80 (2002).

39. B. J. Druker et al., *N Engl J Med* 344, 1038-42. (2001).

40. D. J. Slamon et al., *N Engl J Med* 344, 783-92 (2001).

41. H. Davies et al., *Nature* 417, 949-54 (2002).

42. Bardelli et al., *Science* 300, 949 (2003).

43. Y. Samuels et al., *Science* (2004).

44. Jemal et al., *CA Cancer J Clin* 54, 8-29 (2004).

45. S. Breathnach et al., *J Clin Oncol* 19, 1734-1742 (2001).

46. V. Rusch et al., *Cancer Res* 53, 2379-85 (1993).

47. R. Bailey et al., *Lung Cancer* 41 S2, 71 (2003).

48. M. Fukuoka et al., *J Clin Oncol* 21, 2237-46 (2003).
49. P. A. Janne et al., *Lung Cancer* 44, 221-230 (2004).
50. M. G. Kris et al., *Jama* 290, 2149-58 (2003).
51. V. A. Miller et al., *J Clin Oncol* 22, 1103-9 (2004).
52. M. Huse, J. Kuriyan, *Cell* 109, 275-82 (2002).
53. K. Naoki, T. H. Chen, W. G. Richards, D. J. Sugarbaker, M. Meyerson, *Cancer Res* 62, 7001-3 (2002).
54. J. Stamos, M. X. Sliwkowski, C. Eigenbrot, *J Biol Chem* 277, 46265-72 (2002).
55. T. Fujishita et al., *Oncology* 64, 399-406 (2003).
56. M. Ono et al., *Mol Cancer Ther* 3, 465-472 (2004).
57. M. C. Heinrich et al., *J Clin Oncol* 21, 4342-9 (2003).
58. G. Giaccone et al., *J Clin Oncol* 22, 777-84 (2004).
59. R. S. Herbst et al., *J Clin Oncol* 22, 785-94 (2004).
60. H. Yamazaki et al., *Mol Cell Biol* 8, 1816-20 (1988).
61. M. E. Gorre et al., *Science* 293, 876-80 (2001).
62. Marchetti A, Martella C, Felicioni L, et al: EGFR mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. *J Clin Oncol* 23:857-65, 2005.
63. Franklin WA: Diagnosis of lung cancer: pathology of invasive and preinvasive neoplasia. *Chest* 117:80S-89S, 2000.
64. Paez JG, Janne PA, Lee JC, et al: EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 304:1497-500, 2004.
65. Lynch TJ, Bell DW, Sordella R, et al: Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 350:2129-39, 2004.
66. Pao W, Miller V, Zakowski M, et al: EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci U S A* 101:13306-11, 2004
67. Huang SF, Liu HP, Li LH, et al: High frequency of epidermal growth factor receptor mutations with complex patterns in non-small cell lung cancers related to gefitinib responsiveness in Taiwan. *Clin Cancer Res* 10:8195-203, 2004.
68. Han SW, Kim TY, Hwang PG, et al: Predictive and Prognostic Impact of Epidermal Growth Factor Receptor Mutation in Non-Small-Cell Lung Cancer Patients Treated With Gefitinib. *J Clin Oncol*, 2005.
69. Tokumo M, Toyooka S, Kiura K, et al: The relationship between epidermal growth factor receptor mutations and clinicopathologic features in non-small cell lung cancers. *Clin Cancer Res* 11:1167-73, 2005.
70. Mitsudomi T, Kosaka T, Endoh H, et al: Mutations of the Epidermal Growth Factor Receptor Gene Predict Prolonged Survival After Gefitinib Treatment in Patients with Non-Small-Cell Lung Cancer With Postoperative Recurrence. *J Clin Oncol*, 2005.
71. Pao W, Wang TY, Riely GJ, et al: KRAS Mutations and Primary Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib. *PLoS Med* 2:e17, 2005
72. Read WL, Page NC, Tierney RM, et al: The epidemiology of bronchioloalveolar carcinoma over the past two decades: analysis of the SEER database. *Lung Cancer* 45:137-42, 2004.
73. Sanderson Cox L, Sloan JA, Patten CA, et al: Smoking behavior of 226 patients with diagnosis of stage IIIA/IIIB non-small cell lung cancer. *Psychooncology* 11:472-8, 2002.
74. Radzikowska E, Glaz P, Roszkowski K: Lung cancer in women: age, smoking, histology, performance status, stage, initial treatment and survival. Population-based study of 20 561 cases. *Ann Oncol* 13:1087-93, 2002.
75. Tong L, Spitz MR, Fueger JJ, et al: Lung carcinoma in former smokers. *Cancer* 78:1004-10, 1996.
76. de Perrot M, Licker M, Bouchardy C, et al: Sex differences in presentation, management, and prognosis of patients with non-small cell lung carcinoma. *J Thorac Cardiovasc Surg* 119:21-6, 2000
77. Capewell S, Sankaran R, Lamb D, et al: Lung cancer in lifelong non-smokers. Edinburgh Lung Cancer Group. *Thorax* 46:565-8, 1991
78. Gritz ER, Nisenbaum R, Elashoff RE, et al: Smoking behavior following diagnosis in patients with stage I non-small cell lung cancer. *Cancer Causes Control* 2:105-12, 1991
79. Sridhar KS, Raub W A, Jr.: Present and past smoking history and other predisposing factors in 100 lung cancer patients. *Chest* 101:19-25, 1992
80. Barbone F, Bovenzi M, Cavallieri F, et al: Cigarette smoking and histologic type of lung cancer in men. *Chest* 112:1474-9, 1997
81. Shigematsu H, Lin L, Takahashi T, et al: Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst* 97:339-46, 2005
82. Kosaka T, Yatabe Y, Endoh H, et al: Mutations of the epidermal growth factor receptor gene in lung cancer:

biological and

clinical implications. Cancer Res 64:8919-23, 2004

83. Cho D, Kocher O, Tenen DG, et al: Unusual cases in multiple myeloma and a dramatic response in metastatic lung cancer: case 4. Mutation of the epidermal growth factor receptor in an elderly man with advanced, gefitinib-responsive, non-small-cell lung cancer. J Clin Oncol 23:235-7, 2005

Table J. Characteristics of Nine Patients with Non-Small-Cell Lung Cancer and a Response to Gefitinib

Patient No.	Sex	Age at Beginning of Gefitinib Therapy yr	Pathological Type [*]	No. of Prior Regimens	Smoking-Status [†]	Duration of Therapy mo	Overall Survival [‡] mo	EGFR Mutations [§]	Response [¶]
1	F	70	BAC	3	Never	15.6	18.8	Yes	Major; improved lung lesions
2	M	66	BAC	0	Never	>14.0	>14.0	Yes	Major; improved bilateral lung lesions
3	M	64	Adeno	2	Never	9.6	12.9	Yes	Partial; improved lung lesions and soft-tissue mass
4	F	82	Adeno	1	Former	>13.3	>21.4	Yes	Minor; improved pleural disease
5	F	45	Adeno	2	Never	>14.7	>14.7	Yes	Partial; improved liver lesions
6	M	32	BAC	3	Never	>7.8	>7.8	Yes	Major; improved lung lesions
7	F	62	Adeno	1	Former	>4.3	>4.3	Yes	Partial; improved liver and lung lesions
8	F	58	Adeno	1	Former	11.7	17.9	Yes	Partial; improved liver lesions
9	F	42	BAC	2	Never	>33.5	>33.5	No	Partial; improved lung nodules

* Adenocarcinoma (Adeno) with any element of bronchoalveolar carcinoma (BAC) is listed as BAC.

† Smoking status was defined as former if the patient had not smoked any cigarettes within 12 months before entry and never if the patient had smoked less than 100 cigarettes in his or her lifetime.

‡ Overall survival was measured from the beginning of gefitinib treatment to death.

§ EGFR denotes the epidermal growth factor receptor gene.

¶ A partial response was evaluated with the use of response evaluation criteria in solid tumors; major and minor responses were evaluated by two physicians in patients in whom the response could not be measured with the use of these criteria.

Table 2. Somatic Mutations in the Tyrosine Kinase Domain of EGFR in Patients with Non-Small-Cell Lung Cancer

Patient	Mutation	Effect of Mutation
Patients with a response to gefitinib		
Patient 1	Deletion of 15 nucleotides (2235–2249)	In-frame deletion (746–750)
Patient 2	Deletion of 12 nucleotides (2240–2251)	In-frame deletion (747–751) and insertion of a serine residue
Patient 3	Deletion of 18 nucleotides (2240–2257)	In-frame deletion (747–753) and insertion of a serine residue
Patient 4	Deletion of 18 nucleotides (2240–2257)	In-frame deletion (747–753) and insertion of a serine residue
Patient 5	Substitution of G for T at nucleotide 2573	Amino acid substitution (L858R)
Patient 6	Substitution of G for T at nucleotide 2573	Amino acid substitution (L858R)
Patient 7	Substitution of A for T at nucleotide 2582	Amino acid substitution (L861Q)
Patient 8	Substitution of T for G at nucleotide 2155	Amino acid substitution (G719C)
Patients with no exposure to gefitinib*		
Patient A	Deletion of 18 nucleotides (2240–2257)	In-frame deletion (747–753) and insertion of a serine residue
Patient B	Deletion of 15 nucleotides (2235–2249)	In-frame deletion (746–750)

* Among the 25 patients with no exposure to gefitinib (15 with bronchoalveolar cancer, 7 with adenocarcinoma, and 3 with large-cell carcinoma), 2 (Patients A and B) — both of whom had bronchoalveolar cancer — had EGFR mutations. No mutations were found in 14 lung-cancer cell lines representing diverse histologic types: non-small-cell lung cancer (6 specimens), small-cell lung cancer (6 specimens), bronchus carcinoid (1 specimen), and an unknown type (1 specimen). Polymorphic variants identified within EGFR included the following: the substitution of A for G at nucleotide 1562, the substitution of A for T at nucleotide 1887, and a germ-line variant of unknown functional significance, the substitution of A for G at nucleotide 2885, within the tyrosine kinase domain.

Table 4: Population Characteristics Among 100 Patients Tested for *EGFR* Mutations as Part of NSCLC Care

Characteristic	Frequency
Mean age, years (standard deviation)	60.7 (11.0)
Female	63
Race	
White	76
Asian	7
Other	5
Unknown	12
Stage at Time of Test	
I	15
II	4
III	10
IV	67
Unknown	4
Histology	
Pure BAC	1
Adenocarcinoma with BAC Features	24
Adenocarcinoma	69
NSCLC, all other subtypes	6
Smoking Status	
Current	17
Former	48
Never	29
Unknown	6
Mean amount smoked by current and former smokers, pack-years (standard deviation)	39.0 (32.3)
Mean time from diagnosis to <i>EGFR</i> test, months (standard deviation)	18.7 (78.4)
Prior Chemotherapy Treatment	47
Prior <i>EGFR</i> Targeted Treatment	11

BAC = Bronchioloalveolar Carcinoma, EGFR = Epidermal Growth Factor Receptor

Table 5: Epidermal Growth Factor Receptor Somatic Gene Mutations Identified

Patient	Gender	Histology	Pack-Years Smoked	Exon	Nucleotide Change	Amino Acid Change
1	F	Adeno	0	18	2126A>T	E709V
2	F	A+BAC	60	18	2155G>A	G719S
3	F	A+BAC	0	19	2156G>C	G719A
4	M	A+BAC	0	20	2327G>A	R776H
5	F	Adeno	5	19	2235 2249 del	K745 A750 del ins K
6	M	Adeno	Unknown	19	2235 2249 del	K745 A750 del ins K
7	F	Adeno	0	19	2236 2250 del	E746 A750 del
8	M	Adeno	45	19	2236 2250 del	E746 A750 del
9	F	Adeno	Unknown	19	2236 2250 del	E746 A750 del
10	M	A+BAC	12	19	2237 2255 del ins T	E746 S752 del ins V
11	M	Adeno	1	19	2239 2248 del ins C	L747 A750 del ins P
12	M	A+BAC	0	19	2239 2251 del ins C	L747 T751 del ins P
13	F	Adeno	30	19	2253 2276 del	T751 I759 del ins T
14	F	Adeno	0	19	2254 2277 del	S752 I759 del
15	F	Adeno	0	20	2303 2311 dup	D770 N771 ins SVD
16	M	Adeno	5	20	2313 2318 dup CCCCCA	P772 H773 dup
17	F	Adeno	0	21	2543C>T	P848L*
18	M	BAC	0	21	2573T>G	L858R
19	F	A+BAC	0	21	2573T>G	L858R
20	M	A+BAC	1	21	2573T>G	L858R
21	F	Adeno	0	21	2573T>G	L858R
22	F	Adeno	15	21	2573T>G	L858R
23	F	Adeno	0	21	2582T>A	L861Q

Adeno = Adenocarcinoma, Adeno + BAC = Adenocarcinoma with Bronchioloalveolar Carcinoma Features, BAC = Pure Bronchioloalveolar Carcinoma

* This mutation was identified as a germline variant

Table S1A: Primers for amplification of selected EGFR and receptor tyrosine kinase exons (SEQ ID NOS: 1-212)

Gene	RefSeq	Exon	SEQ ID NO	F Nested	R Nested
ALK	NM_004304	24	1,2	GGAAATATAGGGAAGGGAAGGAA	TGACAGGGTACCAGGAGATGA
ALK	NM_004304	25	3,4	CTGAACCGCCAAGGACTCAT	TTTCCCTCCCTACTAAACACACG
AXL	NM_021913	19	5,6	ACTGATGCCCTGACCCTGTT	CCCATGGTCCCCACTCTT
CSF1R	NM_005211	18	7,8	AGGGACTCCAAAGCCATGTG	CTCTCTGGGGCCATCCACT
CSF1R	NM_005211	19	9,10	CATTGTCAAGGGCAATGTAAGTG	CTCTACCAACCTCGCTGT
DDR1	NM_013994	15	11,12	ACATGGGGAGCCAGAGTGAC	TGCAACCCAGAGAAAAGTGTG
DDR2	NM_006182	16	13,14	TGAGCTTCAACCTAGTTGTGTG	GTTTGCTCTCTGCTGTCTCA
DKFZp761P1010	NM_018423	8	15,16	TGTCCTTGTGTTTTGAAGATTCC	TGCAGACAGATGACAAACATGAA
EGFR	NM_005228	2	17,18	TGGGTGAGTCTCTGTGTGGAG	CATTGCCATAGCAAAAATAAACACA
EGFR	NM_005228	3	19,20	GGTTCAACTGGGCGTCTTA	CCTTCTCCGAGGTGGAATTG
EGFR	NM_005228	4	21,22	CGACCATGGCATCTCTTTA	AAAACGATCTCTATGTCCGTGGT
EGFR	NM_005228	5	23,24	CAGCCAGCCAAACAATCAGA	TCTTTGGAGTCTTCAGAGGGAAA
EGFR	NM_005228	6	25,26	TGTGGTTTCGTTGGAAGCAA	AATTGACAGCTCCCCACAG
EGFR	NM_005228	7	27,28	GGCTTCTGACGGGAGTCAA	CCACCCAAAGACTCTCAAGA
EGFR	NM_005228	8	29,30	CCTTCCATCACCCTCAAG	AGTGCCCTCCCATTTGCCATA
EGFR	NM_005228	9	31,32	ACCGGAATTCCTTCTGCTT	CACTGAAACAAACAACAGGGTGA
EGFR	NM_005228	10	33,34	AGGGGGTGAGTCACAGGTTT	TCAGAAGAAATGTTTTATTCCAAGG
EGFR	NM_005228	11	35,36	GCAAAATCCAATTTTCCACTT	GCAGGAGCTCTGTGCCCTAT
EGFR	NM_005228	12	37,38	TCCCACAGCATGACCTACCA	TTTGCTTCTTAAGGAACGTGAAAA
EGFR	NM_005228	13	39,40	TGTACCCCAAGGTCATGGAG	CAAAAGCCCAAGGGCAAGAA
EGFR	NM_005228	14	41,42	GGAGTCCCAACTCCTTGACC	GTCCTGCCACACACAGGATG
EGFR	NM_005228	15	43,44	GCTTTCCTCACTCACACACA	CAAACTCGGCAATTTCTTG
EGFR	NM_005228	16	45,46	CCACCAATCCAACATCCAGA	TGGCCAGAGCCATAGAAAC
EGFR	NM_005228	17	47,48	TTCCAAGATCATCTACAAGATGTCA	GCACATTGAGAGATCTTTCTGTC
EGFR	NM_005228	18	49,50	TCCAAATGAGCTGGCAAGTG	TCCCAACACTCAGTGAAACAAA
EGFR	NM_005228	19	51,52	GTGCATCGCTGGTAACATCC	TGTGGAGATGAGCAGGCTCT
EGFR	NM_005228	20	53,54	ATCGCATTCATGCGTCTTCA	ATCCCCATGGCAAACCTCTTG
EGFR	NM_005228	21	55,56	GCTCAGAGCCTGGCATGAA	CATCCTCCCTGCATGTTGT
EGFR	NM_005228	22	57,58	TGGCTCGTCTGTGTGTGTCA	CGAAAGAAAATACTTGATCATGCAGA
EGFR	NM_005228	23	59,60	TGAAGCAAATGCCCCAAGAC	TGACATTTCTCCAGGGATGTC
EGFR	NM_005228	24	61,62	AAGTGTGCGATACCAATGC	ATGCGATCTGGGACACAGG
EGFR	NM_005228	25	63,64	GGCACCTGCTGGCAATAGAC	TGACTTCATATCCATGTGAGTTTCACT
EGFR	NM_005228	26	65,66	TATACCTCCATGAGGCACA	GGGAAAAACCCACACAGGAA
EGFR	NM_005228	27	67,68	TCAGAACCAGCATCTCAAGGA	GATGCTGGAGGGAGCACTT
EGFR	NM_005228	28	69,70	CCTTGTGAGGACATTCACAGG	ATGTGCCCGAGGTGGAAGTA
EPHA1	NM_005232	14	71,72	GGAGGGCAGAGGACTAGCTG	GTGCTGGCCAAGTCTTTGT
EPHA1	NM_005232	15	73,74	CTGCAGCTAGCAACAGAGC	AAGAACCAGAGGAGCCAGGA
EPHA2	NM_004431	13	75,76	CGGGTAAGGATGTGGTTGT	CAGGTGTTCTGCTCTCTGAA
EPHA2	NM_004431	14	77,78	GCCTCAGGAGGCAACACACC	GGAGCAAGCCTAAGAAAGTTCA
EPHA3	NM_005233	10	79,80	GCCTTGTATCCATTGCCACA	TGACAACACGTTTTGGGTCAT
EPHA3	NM_005233	11	81,82	TGCATATTCCATTTCAGAACAGA	AAACAGTTTCATTGCTGCTAAAT
EPHA4	NM_004438	13	83,84	CCGGATACAGATACCCAAAAAGA	GGAGGCTTCAAGGGATGAGA
EPHA4	NM_004438	14	85,86	GCTGTGTCTCTGCTGGCTA	TGGTTGTAATGTGGAACCTAGCTTGC
EPHA7	NM_004440	13	87,88	TGGCTGTACGCTAAATAAGCATGT	TCAATTGCTTCATTCTCTCTGTT
EPHA7	NM_004440	14	89,90	TGCTGCTGAACCTACCAACCAA	TGTGGTAGTAATGTGCGAAAACCTG
EPHA8	NM_020526	13	91,92	CAAAGCACCGTCTCAACTCG	CCCGAAACTGCCAATCTTCAT
EPHA8	NM_020526	14	93,94	GGAAAACAGGACCCAGTGT	CCCTCTCCACAGAGCTGAT
EPHB1	NM_004441	7	95,96	GACAGAAGCTGACAAGCAGCA	AGGTTCATTCCCTCCCAGT
EPHB1	NM_004441	8	97,98	TGGGAGTGAGAGTTTGAAGAA	TATGAGGCCGTGAGCTGAAA
EPHB2	NM_017449	11	99,100	AGGGCCCTGCTCTGGTTT	CCAATTGGGCGTTAGTGAAA
EPHB2	NM_017449	12	101,	CTCATGAGATTGGGGCATCA	AGGCCCATGATCTCAGAACGC
EPHB3	NM_004443	11	103,	GGTTGCAGGAGAGACGAGGT	AGGCCCTTACCCCTGTGAC
EPHB3	NM_004443	12	105	ATGACCCCTCCGATCCTACC	TAATCCTGCTCCAGGCCATT

Table S1A: Continued

5	EPHB4	NM_004444	14	107	GGAAAAAGCAGAGGCAGGTG	TGGTCTCAAGAACCCAGCAG
	EPHB6	NM_004445	16	109	GACACCCTCCCCCTCTCAT	ACTATGACACCCCGGCTCAG
	EPHB6	NM_004445	17	111	TGCTTGATGTAAACCCCTGG	GCAATCCAACAGCCATGAGA
	ERBB2	NM_004448	21	113	GGAGCAAACCCCTATGTCCA	TCTCCAACGTGTGTCTGTGG
	ERBB3	NM_001982	21	115	TGGGGACCACTGCTGAGAG	TGCAGCCTTCTCTCTGAA
10	FGFR1	NM_000604	14	117	GCAGAGCAGTGTGGCAGAAG	ACAGGTGGGAAGGGACTGG
	FGFR1	NM_000604	15	119	AGTGGGGTGGGCTGAGAAC	TCTCTGGGGCAGAAAGAGGA
	FGFR2	NM_000141	14	121	ACCCGGCCACACTGTATTIC	CATCCCAACCCAGCTCTCAAC
	FGFR2	NM_000141	15	123	AGGGCATAGCCCTATTGAGC	CCCAGGAAAAAGCCAGAGAA
	FGFR3	NM_000142	13	125	CAGGTGTGGGTGGAGTAGGC	CTCAGGCGCCATCCACTT
15	FGFR3	NM_000142	14	127	AAGAAGACGACCAACGTGAGC	AGGAGCTCCAGGGCACAAG
	FGFR4	NM_002011	14	129	CCTCTCTGTAAAGTGGGTGGA	AGAGGGCCTCAGTGCAGAGT
	FGFR4	NM_002011	15	131	AGATGGGGCAGAACTGGATG	GGGTCCCAGACCAAATCTGA
	FLT1	NM_002019	23	133	AGGTGCTCCCTTCACAGCAT	TTCAGGGACTACAGCTGAGGAA
	FLT1	NM_002019	24	135	GCCGTATGTTATCTGGGAGGT	TGGGCCCATTACACTTTAAGA
20	FLT3	NM_004119	20	137	TTCCATCACCGGTACCTCCT	CCATAAATCAAAATGCACCACA
	FLT3	NM_004119	21	139	GAGTGGTCTTAGGAAGATGATGC	AAAGTCATGGGCTGCAAATACAA
	FLT4	NM_002020	23	141	ATGGTCCCCACTGCTTGG	AGGAGCTCACCTCACCCCTGT
	IGF1R	NM_000875	18	143	CCTTGCGTCTCTCCACACAT	TGGCAACGGGTAACAATGAA
	INSR	NM_000208	18	145	GGCTGAGGTAAGCTGCTTCG	AAAAAGAAGTATCTTGCACCTTT
25	INSR	NM_000208	19	147	AACCCCTCTTAGGGCTCTGTG	CAGGAGGATGGCAGGCTTTC
	KDR	NM_002253	24	149	CGTAGAGAGCTTCAGGACCTGTG	TTCCGAGAAGTTTGGCTTGA
	KIT	NM_000222	17	151	TGTGAACATCATTCAAGGCGTA	AAAATGTGTGATATCCCTAGACAGG
	KIT	NM_000222	18	153	TCCACATTTACGAAACAGCA	GGCTGCTTCTGAGACACAGT
	LTK	NM_002344	16	155	TATCTACCGGTGCGGGACTT	AGGTGTAGCCTCCCTCACA
30	MERTK	NM_006343	17	157	AGGCTGGTGGTGTCTCTGTG	CAAGCTGCCAACCCCTCAGTT
	MET	NM_000245	19	159	TGGATTTCAAATACTGAAGCCACT	TGGAATTGGTGGTGTGAAATTT
	MUSK	NM_005592	15	161	GGGCTTCCATATGTTCTGACATGG	CAGAGGACACGCCATAAGG
	MUSK	NM_005592	15	163	CCGAGATTTAGCCACCAAGGA	CCTGGGAAGCAAACAACACA
	NTRK1	NM_002529	15	165	AGGTCCCCAGTCTCTCTCC	AGACCCATGCAGCCATCTCA
35	NTRK1	NM_002529	16	167	CGTGAACCAACCGAGCTTGT	AGAGGGGCAGAAAGGGAAC
	NTRK2	NM_006180	15	169	GGTGGGGGTGAGGAGCTTAG	TCGTTTAAGCCACCCAGTCA
	NTRK2	NM_006180	16	171	TGCAAAATAAGGAAAGCAAACA	TCCTGACATGGTCTTCCAAACC
	NTRK3	NM_002530	17	173	CAGCATCTTCACACCTCTGA	GCTGGCTCTAAATCCCACT
	NTRK3	NM_002530	18	175	CTAATCCGGGAAGTTGTTGC	TTCTGTATCAGCAGCTCTCTGTG
40	PDGFRA	NM_006206	18	177	CAAGTGCCACCATGGATCA	GGCAGTGTACTGACCCCTTGA
	PDGFRA	NM_006206	19	179	GCACAAGTTATTAAGAGCCCAAGG	AGCATACTGGCCTCACACCA
	PDGFRB	NM_002609	18	181	GCACATGGGCAGTGTGTATTT	GAGCCCCACACAGATTTCT
	PDGFRB	NM_002609	19	183	ATGGGACGGAGAAGTGGTTG	TCCCTGTATCAGGGCTCGTC
	PTK7	NM_002821	18	185	TTCTACGCAGCACACCAAT	GCAGGCACTAAACCTTTCC
45	PTK7	NM_002821	19	187	GCACGATGTGACCAATTIC	AGCCCTGAGAGGGAGGTAGG
	RET	NM_000323	15	189	CACACACCACCCCTCTGCT	AAAGATTTGGGGTGAGGCTA
	RET	NM_000323	16	191	CTGAAAGCTCAGGGATAGGG	CTGGCCAAGCTGCACAGA
	ROR1	NM_005012	09	193	TGCAGCCAACGATTTGAAAG	GGAAAGCCCCAAGTCTGAAA
	ROR1	NM_005012	09	195	TCATCATGAGATCCCCACACT	GCATTTCCCTGGAAGGAGT
	ROR1	NM_005012	09	197	TGGATTCACTAACCAGGAAGTGA	CCCATTCACCAAGGATGATT
50	ROR1	NM_005012	09	199	GTTTCCAGCTGCCCACTACC	GCTCGAAACCACATGTTCCA
	RYK	NM_002958	13	201	CTGGATTJGGGGTCTCTGC	CGGGAACAGCTAGCAGATTJTT
	TEK	NM_000459	18	203	GGGAATTTTGGAGGGGAAC	GCITCAGTCACCACAGAGCA
	TEK	NM_000459	19	205	TGAGTCTACCCAGCAATCATTTG	TTCCCGAGAGCTACAGGACA
	TIE	NM_005424	18	207	GGTAACAAGGTTACCCACGAA	GTTTGAGGGCTGAGTGTGG
55	TIE	NM_005424	19	209	CCTCACCTTAGGGCTTGTG	AGCCAGGCTATGCCTTAGA
	TYRO3	NM_006293	18	211,212	GGGTAGCTTGGGAGCAAAGA	CCAAACCCAGAGAGCAAGAC

Table S1B: Primers for amplification of selected EGFR and receptor tyrosine kinase exons (SEQ ID NOS: 213 - 424)

Gene	RefSeq	Exon	SEQ ID NO	F External	R External
ALK	NM 004304	24	213,214	CATTTCCCTAATCCTTTTCCA	GTGATCCCAGATTTAGGCCTTC
ALK	NM 004304	25	215	GCCTCTCGTGGTTTGTITGTC	CCCAGGGTAGGGTCCAATAATC
AXL	NM 021913	19	217	CTTCTGCTGGAGGTGACTGAT	CAGGCATAGTGTGTGATGGTCA
CSF1R	NM 005211	18	219	TCACGATACACATTCTCAGATCC	GAAGATCTCCCAGAGGAGGATG
CSF1R	NM 005211	19	221	CGTAACGTGCTGTTGACCAAT	AAACGAGGGAAGAGCCAGAAAG
DDR1	NM 013994	15	223	TGGGGAGCACAATAAAGAAGA	ACTCTTGGCTCCTGGATTCTTG
DDR2	NM 006182	16	225	GGAAGTCAGTGTGCAGGGAATA	TTTATGAGAAATAGGCAAGCA
DKFZp761P1010	NM 018423	8	227	TGGTAATCCTAAACACAATGCAGA	CTGGGCAACACAGTGAGATCCT
EGFR	NM 005228	2	229	TCACAAATTCTTTGCTGTGTCC	CATGGAATCCAGATTAGCCTGT
EGFR	NM 005228	3	231	GATTGTGTCAGATCGTGGACAT	CGCTTAAATCTTCCCATTCAG
EGFR	NM 005228	4	233	CTCCATGGCACCATCAATTAACA	CTCAGGACACAAGTGTCTGTCT
EGFR	NM 005228	5	235	GCAGTTCATGGTTCATCTTCTTT	CAAAATAGCCCACTCTGGATTA
EGFR	NM 005228	6	237	CTTCTGTCATTGCCCAAGATG	CAAGGTCTCAGTGAGTGGTGA
EGFR	NM 005228	7	239	GAGAAGGCTCTTCTGACTCTGC	CAGGTGTTTCTCTGTGAGGTG
EGFR	NM 005228	8	241	CACATTGCGGCTAGAATGTTA	ACCCCGTCACAACCTTCAGT
EGFR	NM 005228	9	243	GCCGTAGCCCAAAAGTGTACTA	TCAGCTCAAACCTGTGATTTCC
EGFR	NM 005228	10	245	CTCACTCTCCATAAATGCTACGAA	GACTTAACGTGTCCCCTTTTGC
EGFR	NM 005228	11	247	GCCTCTTCGGGTAAATCAGATA	GAAGTCTGTGGTTTJAGCGGACA
EGFR	NM 005228	12	249	ATCTTTTGCTGGAGGAACCTTT	CAGGGTAAATTCATCCCATTGA
EGFR	NM 005228	13	251	CAGCAGCCAGCACAACACTACTT	TTGGCTAGATGAACCATTTGATGA
EGFR	NM 005228	14	253	TGAATGAAGCTCTGTGTTTACTC	ATGTTTCATCGCAGGCTAATGTG
EGFR	NM 005228	15	255	AAAACAGGGAGAAGTCTAAGCAA	CATGGCAGAGTCAATCCCACT
EGFR	NM 005228	16	257	CAATGCTAGAACACGCTGTCT	TCCCTCCATGAGGACAAAGTT
EGFR	NM 005228	17	259	GAGGAGCTGTGAGAAAGTTGGA	ATTTCTCTGGATGGATGTACCA
EGFR	NM 005228	18	261	TCAGAGCCTGTGTTTCTACCAA	TGGTCTCACAGGACCACTGATT
EGFR	NM 005228	19	263	AAATAATCAGTGTGATTCGTGGAG	GAGGCCAGTCTGTCTCTAAGG
EGFR	NM 005228	20	265	ACTTCACAGCCCTGCGTAAAC	ATGGGACAGGCACTGATTGT
EGFR	NM 005228	21	267	GCAGCGGGTTACATCTTCTTTC	CAGCTCTGGCTCACACTACCAG
EGFR	NM 005228	22	269	CCTGAACCTCCGTCAGACTGAAA	GCAGCTGGACTCGATTTCTCT
EGFR	NM 005228	23	271	CCTTACAGCAATCCTGTGAAACA	TGCCCAATGAGTCAAGAAAGTGT
EGFR	NM 005228	24	273	ATGTACAGTGTGGCATGGTCT	CACTCACGGATGCTGCTTAGTT
EGFR	NM 005228	25	275	TAAGGCACCCACATCATGTCA	TGGACCTAAAAGGCTTACAATCA
EGFR	NM 005228	26	277	GCCTTTTAGGTCCACTATGGAATG	CCAGGCGATGCTACTACTGGTC
EGFR	NM 005228	27	279	TCATAGCACACCTCCCTCACTG	ACACAACAAAGAGCTTGTGCAG
EGFR	NM 005228	28	281	CCATTACTTTGAGAAGGACAGGAA	TATTCTTGCTGGATGCGTTTCT
EPHA1	NM 005232	14	283	AGGAGGGCAGAGGACTAGCTG	GGCAATGTGAATGTGCACTG
EPHA1	NM 005232	15	285	CTTGAACCTGGGAGGTGGAG	ATCAGGGTGGGAGGAGTAAAGA
EPHA2	NM 004431	13	287	CCCACCTTACCTCTCACCTGTGC	GTGAACTTCCGGTAGGAAATGG
EPHA2	NM 004431	14	289	AGGGGACCTCAAGGGAGAAG	AGATCATGCCAGTGAACCTCAG
EPHA3	NM 005233	10	291	GGACCAGGAAAGTCTTGCTTT	GGTGGGGAACATTAACCTGAGG
EPHA3	NM 005233	11	293	GCTTCAGCTGTGTTTGTGTCAG	ACCTTCTGTTGAGGGAAATATG
EPHA4	NM 004438	13	295	CCCAGCTCCTAGGGTACAGTCT	CAGTCAGCTTCAAAATCCCTCTT
EPHA4	NM 004438	14	297	TCACTTCCCTGTGAGTAAAGAAAA	GGCCATTTAATCTTGTCTTGA
EPHA7	NM 004440	13	299	TGGACTTGTGCAAACTCAAACCTG	TCCCAATATAGGGCAGTCAATGT
EPHA7	NM 004440	14	301	TCTCAATCAGTTGAGTTGCCTTG	AGCTGTGCAAGTGTGGAACAT
EPHA8	NM 020526	13	303	GCTGTGAGGGTAAATGAGACCA	GTCTCCTGGTGAGTGACTGTGG
EPHA8	NM 020526	14	305	CCTTCTCTCTCTCCACAGC	GTCTTGTGCCAACAGTCGAG
EPHB1	NM 004441	7	307	GCTTGGCAAGGAGAAGAGAACA	GCTTGCTTTCTGCTTGAACAAC
EPHB1	NM 004441	8	309	GCTGGTCACCTTGAGCTTCTCT	CCATGCTGGGCTCTTTGATTA
EPHB2	NM 017449	11	311	CACCACTCTGAAGTTGGCTCT	ATGGCTCTGCACATTTGTTC
EPHB2	NM 017449	12	313	CAGAGTGGGAAAAGGCACTTCA	CCAGAGTCTGTGCAGACATTC
EPHB3	NM 034443	11	315	ATGGGGATTAAGTGGGATGTTG	CGTAGCTCCAGACATCACTAGCA
EPHB3	NM 004443	12	317	GCAACCTGGTCTGCAAGTCTC	ACCCAGCAGTCCAGCATGAG

TABLE S1B CONT.

	EPHB4	NM 004444	14	319	GAGTTTCAGTGAGCCAAGATCG	TTACAGGCTTGAGCCACTAGGC
	EPHB6	NM 004445	16	321	AAGCTTCCAGGAGACGAGGTC	GTCCCTGAAATCCCTCAAACC
5	EPHB6	NM 004445	17	323	TGCTCCATAAACGTGACTATTGC	GTAAGAGGGTGGGCTGGAATCT
	ERBB2	NM 004448	21	325	CTTAGACCATGTCCGGGAAAAC	CACATCACTCTGGTGGGTGAAC
	ERBB3	NM 001982	21	327	AAATTTTCATCCCAAAACCAACC	CCAGTCCCAAGTCTTGATCATT
	FGFR1	NM 000604	14	329	ACAAGTCGGCTAGTIGCATGG	TCTCAGATGAAACCAACAGCAC
	FGFR1	NM 000604	15	331	TTCTCTGAGAAAGCAAGGAGTGG	CCAGGGAGAAAGCAGGACTCTA
10	FGFR2	NM 000141	14	333	TTCTGGCGGTGTTTTGAAATTA	CTCAACATTGACGGCCTTTCTT
	FGFR2	NM 000141	15	335	TCAGCTCTTAAACAGGGCATAGC	GAAATGCAGCAGCCACTAAAGA
	FGFR3	NM 000142	13	337	CTCACCTTCAAGGACCTGGTGT	CAGGGAGGGGTAGAAACACA
	FGFR3	NM 000142	14	339	GGAGAGGTGGAGAGGCTTCAG	GAGACTCCCAGGACAGACACCT
	FGFR4	NM 002011	14	341	CACTCGTCTCTCAACCTTCC	AGGACTCACACGTCACTCTGGT
15	FGFR4	NM 002011	15	343	GGACAATGTGATGAAGATTGCTG	ATAGCAGGATCCCAAAAGACCA
	FLT1	NM 002019	23	345	GGCTTGGGGACCTGTATTGTGA	CAGTGGCCTTTCTGAGCCTTAC
	FLT1	NM 002019	24	347	GCACCTAGCTCCCTCTTTTATG	TTTTACAGTAGAGGGCAGACATGC
	FLT3	NM 004119	20	349	GCCACCATAGCTGCAGAAATTAG	CCCAAGGACAGATGTGATGCTA
	FLT3	NM 004119	21	351	GCCTTTGTTTCGAGAGGAGTTGT	GTTACGCTCTCAAGCAGGTTA
20	FLT4	NM 002020	23	353	ATTCACACAAGCTCTCTCCATGA	CTTGCCCCAAGATGCCTAAG
	IGF1R	NM 000875	18	355	TGCTTGGTATTGCTCATCATGT	CCCTTAGCTAGCCCACTGACAA
	INSR	NM 000208	18	357	CTCCTGGGAGTGGTGTCCTAA	CCTGGGCAACAGACAGAGTAAG
	INSR	NM 000208	19	359	CTTCACCTCCCATGCGTACC	GGGTTCAATGCCTACAGGA
	KDR	NM 002253	24	361	AAAATCTGTGACTTTGGCTTGG	GGGAGGAGACATTCTTTGATTTG
25	KIT	NM 000222	17	363	GCAGTCTGAGAAGAAACAGC	CTTCACATGCCCAAAATTACA
	KIT	NM 000222	18	365	TGAGCCATGTATTTCAGAGGTGA	TACATTTCAGCAGGTGCGTGTT
	LTK	NM 002344	16	367	TTGCCTACTCTGTAGGGATAATTGC	ATAGGGCATGTAGCCAGTGA
	MERTK	NM 006343	17	369	GCTCTGCTGTTGGTCTCACT	TTGCAAAGCACACATCTTCTGA
	MET	NM 000245	19	371	TGGCAATGTCAATGTCAAGCAT	GTATGTTGCCCACTCAACAAA
30	MUSK	NM 005592	15 1	373	TGCATTTCCTAGCTGAGACTCC	TGCCATCTCGACGTAGTAAAT
	MUSK	NM 005592	15 2	375	CTCTCCTGTGCTGAGCAGCTTT	TGTTTCCAATCACTGGCTTTCA
	NTRK1	NM 002529	15	377	GAACCATGGGCTGTCTCTGG	ATCTGGGATAGCGAAGGAGACA
	NTRK1	NM 002529	16	379	ATTACAGGCCACACGCCATC	AAGGCAAGAATAAGGGAGGAAGA
	NTRK2	NM 006180	15	381	GCTCTCAGGACTGCAGAAGTACA	GAGGAACCAATCCCACTCACAC
35	NTRK2	NM 006180	16	383	TCACCTTTTGCCTTCTGTCTCTG	GCACTGTGCTTTGCTTTCTCAG
	NTRK3	NM 002530	17	385	TGTCCTCTTTATCGTAGGTCTCCA	CACCACATTTCTTACAGTTCCA
	NTRK3	NM 002530	18	387	CACGTGTGACCAAGACAGACAAA	TGTGGTTTTCTGTATCAGCAGCTT
	PDGFRA	NM 006206	18	389	CAGGGAGTCTGAAATCATCAGG	TCAAGTATCTAGCCCCAAATCCA
	PDGFRA	NM 006206	19	391	GGCAATATTGACCATTCATCATT	AGGCCAGGAGTAAGACGCAAC
40	PDGFRB	NM 002609	18	393	AAGAACGTACGTGTGGTGTGG	CGCTATACTTGCTCCATGCACT
	PDGFRB	NM 002609	19	395	AGGAAACAGCCTCTGGTCTCTC	GTCAATGCTCAGACAGGGAGAT
	PTK7	NM 002821	18	397	CCCAGGAAGGCAGGTACTGTTA	TTTTACAACCACCAAGGGTGTG
	PTK7	NM 002821	19	399	TCGTGTGGTTACCTCCAGATTTT	AAATTAGCCAGGGAGTGGAGGT
	RET	NM 000323	15	401	CATGCCATGCTATGGCTCAC	AGGCTGAGCGGAGTTCTAATTG
45	RET	NM 000323	16	403	ATCTCAGCAATCCACAGGAGGT	ATTTGCCCTACGAACACATCAT
	ROR1	NM 005012	09 1	405	TGGAAAGTTGTCTATGGCACCTC	ATGGGCAGCAAGGACTTACTCT
	ROR1	NM 005012	09 2	407	CACCCCAATATTGTCTGCCTTC	GGCTCGGGAACATGTAATTAGG
	ROR1	NM 005012	09 3	409	CCATCATGTATGGCAAAATCTCTT	TGGCGTCTCCTAGTAAAGATGCT
	ROR1	NM 005012	09 4	411	GCCAGATTGCTGGTTTCATTG	GGCTAAAACACAAAGCACCATT
50	RYK	NM 002958	13	413	GGGAAGTCATCCACAAAGACCT	GGTCTGGGTCACAGCTCCTC
	TEK	NM 000459	18	415	TTCTTCTGCCAAGATGTGGTGT	TGCAGATGCTGCAATCATGTITA
	TEK	NM 000459	19	417	TGGACCCCGAAAGATAAATAGG	TTCTGCACTCCTCTGGAACCTG
	TIE	NM 005424	18	419	GGGTGAGAGCCAACTGATCT	CTGTGCCCTCTCATCTCACACT
	TIE	NM 005424	19	421	AGAACCTAGCCTCCAAGATTGC	ACACCTTCCAAGACTCCTTCCA
55	TYRO3	NM 006293	18	423,424	GACTCGAGGGTGGGAGACAG	GCTGTCACTAGGTGTCTGAGC

Table S2: EGFR mutation status in untreated lung cancer

Sample	Histology	Source	Gender	Exon	Sequence alteration	SEQ ID NO	Nucleotide	Amino acid
S0514	adenocarcinoma	U.S.	F	18	Substitution	425	2155G>A	G719S
S0377	adenocarcinoma	Japan	F	18	Substitution	426	2155G>A	G719S
S0418	adenocarcinoma	Japan	F	19	Del-1a	427	2235_2249delGGAAATTAAGAGAAAGC	E746_A750del
S0363	large cell ca.	Japan	F	19	Del-1a	428	2235_2249delGGAAATTAAGAGAAAGC	E746_A750del
S0380	adenocarcinoma	Japan	M	19	Del-1a	429	2235_2249delGGAAATTAAGAGAAAGC	E746_A750del
S0399	adenocarcinoma	Japan	F	19	Del-1a	430	2235_2249delGGAAATTAAGAGAAAGC	E746_A750del
S0353	adenocarcinoma	Japan	F	19	Del-1a	431	2235_2249delGGAAATTAAGAGAAAGC	E746_A750del
S0385	adenocarcinoma	Japan	M	19	Del-1a	432	2235_2249delGGAAATTAAGAGAAAGC	E746_A750del
S0301	adenocarcinoma	Japan	M	19	Del-1a	433	2235_2249delGGAAATTAAGAGAAAGC	E746_A750del
S0412	adenocarcinoma	Japan	M	19	Del-1b	434	2236_2250delGAAATTAAGAGAAAGCA	E746_A750del
S0335	adenocarcinoma	Japan	M	19	Del-1b	435	2236_2250delGAAATTAAGAGAAAGCA	E746_A750del
S0405	adenocarcinoma	Japan	F	19	Del-1b	436	2236_2250delGAAATTAAGAGAAAGCA	E746_A750del
S0439	adenocarcinoma	Japan	M	19	Del-2	437	2254_2277delTCCGAAAGCCCAACGGAATC	S752_I759del
S0361	adenocarcinoma	Japan	F	21	Substitution	438	2573T>G	L858R
S0388	adenocarcinoma	Japan	F	21	Substitution	439	2573T>G	L858R
S0389	adenocarcinoma	Japan	F	21	Substitution	440	2573T>G	L858R

Table S3.A: EGFR mutation status in gefitinib-treated lung cancer

Gefitinib sensitivity	Sample	Histology	Source	Gender	Exons	Sequence alteration	SEQ ID NO	Nucleotide	Amino acid
Y	IR1T	adenocarcinoma	U.S.	M	19	Del-3	441	2239_2247delTTAAGAGAA, 2248G>C	L747_E749del, A750P
Y	P003	adenocarcinoma	U.S.	M	19	Del-3	442	2239_2247delTTAAGAGAA, 2248G>C	L747_E749del, A750P
Y	IR4T	bronchioloalveolar carcinoma	U.S.	F	19	Del-4	443	2240_2257delTTAAGAGAAAGCAACATCTC	L747_S752del, P753S
Y	IR2T	adenocarcinoma	U.S.	F	19	Del-5	444	2238_2255delATTAGAGAGAAAGCAACATC, 2237A>T	L747_S752del, E746V
Y	IR3T	adenocarcinoma	U.S.	F	21	Substitution	445	2573T>G	L858R
Y in vitro	IRG H3255	adenocarcinoma	U.S.	F	21	Substitution	446	2573T>G	L858R
N	IR5	adenocarcinoma	U.S.	F	18-24	None detected		n/a	n/a
N	IR6	adenocarcinoma	U.S.	M	18-24	None detected		n/a	n/a
N	IR8	adenocarcinoma	U.S.	F	18-24	None detected		n/a	n/a
N	IR9	NSCLC	U.S.	F	18-24	None detected		n/a	n/a

Table S3B: *EGFR* mutations not shown in Table 2, Table S2, or Table S3A.

Sample	Tissue	Exon	Sequence alteration	Nucleotide	Amino acid
Tar4T	Lung adenocarcinoma	19	Deletion	2239-2250delTTAAGAGAAAGCA; 2251A>C	L747_A750del; T751T
AD355	Lung adenocarcinoma	19	Deletion	2240-2254delTAAGAGAAAGCA	L747_T751del
IRTT	Lung adenocarcinoma	19	Deletion	2257-2271delCCGAAAGCCCAACAAG	P753_K757del
AD240	Lung adenocarcinoma	20	Insertion	2309-2310insCAACCCCGG	D770_N771ins NPG
AD261	Lung adenocarcinoma	20	Insertion	2311-2312insGCGTGGACA	D770_N771ins SVD
	Lung adenocarcinoma	20	Insertion	2316-2317insGGT	P772_H773ins V
AD356	Lung adenocarcinoma	20	Substitution	2334-2335GG>AA	G779S
SP02-23	Acute myeloid leukemia	21	Substitution	2570G>T	G857V
SP08-94	Glioma	21	Substitution	2582T>A	L861Q
SP06-45	Sarcoma	21	Substitution	2648T>C	L883S
AD241	Colon adenocarcinoma	22	Substitution	2686G>T	D896Y

Table S3C Position of BCR-ABL mutants resistant to imatinib and analogous positions in EGFR

Abl1 residue subject to resistance mutation	Analogous EGFR residue	Identical/similar/non-conserved
Met-244	Lys-714	Non-conserved
Leu-248	Leu-718	Identical
Gly-250	Ser-720	Non-conserved
Gln-252	Ala-722	Non-conserved
Tyr-253	Phe-723	Similar
Glu-255	Thr-725	Non-conserved
Asp-276	Ala-750	Non-conserved
Thr-315	Thr-790	Identical
Phe-317	Leu-792	Similar
Met-351	Met-825	Identical
Glu-355	Glu-829	Identical
Phe-359	Leu-833	Similar
His-396	His-870	Identical
Ser-417	Thr-892	Similar
Phe-486	Phe-961	Identical

Table S4: Primers used for cDNA sequencing

Primer name	SEQ ID NO	Primer sequence 5' to 3'
cDNA_EGFR_aF	447	TGTAACGACGGCCAGTCGCCAGACCGGACGACA
cDNA_EGFR_aR	448	CAGGAAACAGCTATGACCAGGGCAATGAGGACATAACCA
cDNA_EGFR_bF	449	TGTAACGACGGCCAGTGGTGGTCCTTGGGAATTTGG
cDNA_EGFR_bR	450	CAGGAAACAGCTATGACCCCATCGACATGTTGCTGAGAAA
cDNA_EGFR_cF	451	TGTAACGACGGCCAGTGAAGGAGCTGCCCATGAGAA
cDNA_EGFR_cR	452	CAGGAAACAGCTATGACCCGTGGCTTCGTCTCGGAATT
cDNA_EGFR_dF	453	TGTAACGACGGCCAGTGAACTGACCAAATCATCTGT
cDNA_EGFR_dR	454	CAGGAAACAGCTATGACCTACCTATTCCGTTACACACTTT
cDNA_EGFR_eF	455	TGTAACGACGGCCAGTCCGTAATTATGTGGTGACAGAT
cDNA_EGFR_eR	456	CAGGAAACAGCTATGACCGCGTATGATTTCTAGGTTCTCA
rDNA_EGFR_fF	457	TGTAACGACGGCCAGTCTGAAACCGTAAAGGAAATCAC
cDNA_EGFR_fR	458	CAGGAAACAGCTATGACCCCTGCCTCGGCTGACATTC
cDNA_EGFR_gF	459	TGTAACGACGGCCAGTTAAGCAACAGAGGTGAAAACAG
cDNA_EGFR_gR	460	CAGGAAACAGCTATGACCGGTGTTGTTTTCTCCCATGACT
cDNA_EGFR_hF	461	TGTAACGACGGCCAGTGGACCAGACAACTGTATCCA
cDNA_EGFR_hR	462	CAGGAAACAGCTATGACCTTCCTTCAAGATCCTCAAGAGA
cDNA_EGFR_iF	463	TGTAACGACGGCCAGTGATCGGCCTCTTCATGCGAA
cDNA_EGFR_iR	464	CAGGAAACAGCTATGACCACGGTGGAGGTGAGGCAGAT
cDNA_EGFR_jF	465	TGTAACGACGGCCAGTCGAAAGCCAACAAGGAAATCC

(continued)

Primer name	SEQ ID NO	Primer sequence 5' to 3'
cDNA_EGFR_iR	466	CAGGAAACAGCTATGACCATTCCAATGCCATCCACTTGAT
cDNA_EGFR_kF	467	TGTAACACGACGGCCAGTAACACCGCAGCATGTCAAGAT
cDNA_EGFR_kR	468	CAGGAAACAGCTATGACCCTCGGGCCATTTTGGAGAATT
cDNA_EGFR_IF	469	TGTAACACGACGGCCAGTTCAGCCACCCATATGTACCAT
cDNA_EGFR_IR	470	CAGGAAACAGCTATGACCGCTTTGCAGCCCATTCTATC
cDNA_EGFR_mF	471	TGTAACACGACGGCCAGTACAGCAGGGCTTCTTCAGCA
cDNA_EGFR_mR	472	CAGGAAACAGCTATGACCTGACACAGGTGGGCTGGACA
cDNA_EGFR_nF	473	TGTAACACGACGGCCAGTGAATCCTGTCTATCACAAATCAG
cDNA_EGFR_nR	474	CAGGAAACAGCTATGACCGGTATCGAAAGAGTCTGGATTT
cDNA_EGFR_oF	475	TGTAACACGACGGCCAGTGCTCCACAGCTGAAAATGCA
cDNA_EGFR_oR	476	CAGGAAACAGCTATGACCACGTTGCAAAACCAGTCTGTG

[0347] The following numbered paragraphs define particular embodiments of the present invention.

1. A method for determining the likelihood of effectiveness of an epidermal growth factor receptor (EGFR) targeting treatment in a human patient affected with or at risk for developing cancer comprising: detecting the presence or absence of at least one nucleic acid variance in the kinase domain of the erbB1 gene of said patient relative to the wildtype erbB1 gene, wherein the presence of the at least one nucleic acid variance indicates that the EGFR targeting treatment is likely to be effective.

2. The method of paragraph 1, wherein the nucleic acid variance increases kinase activity.

3. The method of paragraph 1, wherein the erbB1 gene is obtained from a biological sample from said patient.

4. The method of paragraph 1, wherein the variance in the kinase domain of the erbB1 gene effects the conformational structure of the ATP-binding pocket.

5. The method of paragraph 1, wherein the variance in the kinase domain of erbB1 is in an exon of the erbB1 gene selected from the group consisting of exon 18, 19, 20 or 21.

6. The method of paragraph 5, wherein the variance is in exon 18, 19 or 21.

7. The method of paragraph 1, wherein the variance in the kinase domain of the erbB1 gene is an in frame deletion, substitution, or insertion.

8. The method of paragraph 7, wherein the in frame deletion is in exon 19 of the erbB1 gene.

9. The method of paragraph 8, wherein the in frame deletion in exon 19 of the erbB1 gene comprises a deletion of at least amino acids leucine, arginine, glutamic acid and alanine, at codons 747, 748, 749, and 750 of SEQ. ID. NO.512.

10. The method of paragraph 8, wherein the in frame deletion in exon 19 of the erbB1 gene comprises a deletion of at least amino acids leucine, arginine, and glutamic acid at codon 747, 748 and 749 of SEQ. ID. NO. 512.

11. The method of paragraph 8, wherein the in frame deletion comprises nucleotides selected from the group consisting of 2235 to 2249, 2240 to 2251, and 2240 to 2257 of SEQ ID NO: 511.

12. The method of paragraph 7, wherein the substitution is in exon 21 of the erbB1 gene.

13. The method of paragraph 12, wherein the substitution in exon 21 comprises at least one amino acid.

14. The method of paragraph 12, wherein the substitution in exon 21 comprises a substitution from the group consisting of a guanine for a thymine at nucleotide 2573 of SEQ ID NO: 511, and an adenine for a thymine at nucleotide 2582 of SEQ ID NO: 511.

15. The method of paragraph 7, wherein the substitution is in exon 18 of the erbB1 gene.

16. The method of paragraph 15, wherein the substitution in exon 18 is a thymine for a guanine or a serine for a guanine at nucleotide 2155 of SEQ ID NO: 511.

17. The method of paragraph 1, wherein the detection of the presence or absence of said at least one variance comprises amplifying a segment of nucleic acid.

18. The method of paragraph 17, wherein the segment to be amplified is 1000 nucleotides in length or less.

19. The method of paragraph 17, wherein the segment to be amplified includes a plurality of variances.

20. The method of paragraph 1, wherein the detection of the presence or absence of said at least one variance comprises contacting the erbB1 nucleic acid with at least one nucleic acid probe, wherein said at least one probe preferentially hybridizes with a nucleic acid sequence comprising said variance under selective hybridization conditions.

21. The method of paragraph 1, wherein the detection of the presence or absence of said at least one variance comprises sequencing at least one nucleic acid sequence.

22. The method of paragraph 1, wherein the detection of the presence or absence of said at least one variance comprises mass spectrometric determination of at least one nucleic acid sequence.

23. The method of paragraph 1, wherein the detection of the presence or absence of said at least one variance comprises performing a polymerase chain reaction (PCR) to amplify nucleic acid comprising the erbB1 coding sequence, and determining nucleotide sequence of the amplified nucleic acid.

24. The method of paragraph 23, wherein determining the nucleotide sequence of the amplified nucleic acid comprises sequencing at least one nucleic acid segment.

25. The method of paragraph 23, wherein determining the nucleotide sequence of the amplified nucleic acid comprises running the amplified nucleic acid segment on a gel and determining the segments size.

26. The method of paragraph 1, wherein the detection of the presence or absence of said at least one variance comprises determining the haplotype of a plurality of variances in a gene.

27. A probe which specifically binds under selective binding conditions to a nucleic acid sequence comprising at least one variance in the erbB1 gene, wherein the variance is a mutation in the kinase domain of erbB1 that confers a structural change in the ATP-binding pocket.

28. The probe of paragraph 27, wherein said variance is in an exon of the erbB1 gene selected from the group consisting of exon 18, 19, 20 or 21.

29. The probe of paragraph 27, wherein said probe comprises a nucleic acid sequence 500 nucleotide bases or fewer in length.

30. The probe of paragraph 27, wherein said probe comprises DNA.

31. The probe of paragraph 27, wherein said probe comprises DNA and at least one nucleic acid analog.

32. The probe of paragraph 27, wherein said probe comprises peptide nucleic acid (PNA).

33. The probe of paragraph 27, further comprising a detectable label.

34. The probe of paragraph 27, wherein the detectable label is a fluorescent label.

35. The probe of paragraph 27, wherein said probe comprises at least 10 consecutive nucleic acids consisting of at least nucleic acids 15-25 of SEQ ID NO 495, or compliments thereof.

36. The probe of paragraph 27, wherein said probe comprises at least 10 consecutive nucleic acids consisting of at least nucleic acids 20-30 of SEQ ID NO 497, or compliments thereof.

37. The probe of paragraph 27, wherein said probe comprises at least 10 consecutive nucleic acids consisting of at least nucleic acids 20-30 of SEQ ID NO 499, or compliments thereof.

38. A method for determining the likelihood of effectiveness of an EGFR targeting treatment in a patient comprising: determining the kinase activity of the erbB1 gene in a biological sample from said patient, wherein an increase in kinase activity following stimulation with an EGFR ligand, compared to a control, indicates that the EGFR targeting treatment is likely to be effective.

39. The method of paragraphs 1 and 38, wherein the EGFR targeting treatment is a tyrosine kinase inhibitor.

40. The method of paragraph 39, wherein the tyrosine kinase inhibitor is an anilinoquinazoline.

41. The method of paragraph 40, wherein the anilinoquinazoline is a synthetic anilinoquinazoline.

42. The method of paragraph 41, wherein the synthetic anilinoquinazoline is selected from the group consisting of gefitinib and erlotinib.

43. A method for determining the likelihood of effectiveness of an epidermal growth factor receptor (EGFR) targeting treatment in a human patient affected with or at risk for developing cancer comprising:

- a. detecting the presence or absence of at least one nucleic acid variance in exon 18, 19, 20, or 21 by performing a polymerase chain reaction (PCR) to amplify a portion of exon 18, 19, 20, or 21; and
- b. determining the nucleotide sequence of the amplified nucleic acid by sequencing at least one portion of the amplified exon 18, 19, 20, or 21, wherein the presence of at least one nucleotide variance in exon 18, 19, 20 or 21 compared to a wildtype erbB1 control indicates that the EGFR targeting treatment is likely to be effective.

44. A method of treating a patient affected with or at risk for developing cancer, comprising detecting the presence or absence of at least one nucleic acid variance in the kinase domain of the erbB1 gene of the patient, wherein the patient is administered an EGFR targeting treatment if the presence of the said at least one nucleic acid variance is detected.

45. A method of treating a patient affected with or at risk for developing cancer, comprising:

- a. detecting the presence or absence of at least one nucleic acid variance in exon 18, 19, 20, or 21 by performing a polymerase chain reaction (PCR) to amplify a portion of exon 18, 19, 20, or 21,
- b. determining the nucleotide sequence of the amplified nucleic acid by sequencing at least one portion of the amplified exon 18, 19, 20 or 21; and
- c. and administering an EGFR targeting treatment to the patient if the presence of the said at least one nucleic acid variance is detected.

46. The method of paragraphs 44 and 45, wherein the EGFR targeting treatment is a tyrosine kinase inhibitor.

47. The method of paragraph 46, wherein the tyrosine kinase inhibitor is an anilinoquinazoline.

48. The method of paragraph 47, wherein the anilinoquinazoline is a synthetic anilinoquinazoline.

49. The method of paragraph 48, wherein the synthetic anilinoquinazoline is selected from the group consisting of gefitinib and erlotinib.

50. The method of paragraphs 44 and 45, wherein the erbB1 gene is obtained from a biological sample from said patient.

51. The method of paragraphs 44 and 45, wherein said cancer is selected from the group consisting of gastrointestinal cancer, prostate cancer, ovarian cancer, breast cancer, head and neck cancer, lung cancer, non-small cell lung cancer, cancer of the nervous system, kidney cancer, retina cancer, skin cancer, liver cancer, pancreatic cancer, genital-urinary cancer and bladder cancer.

52. The method of paragraph 51, wherein said cancer is non-small cell lung cancer.

53. The method of paragraphs 44 and 45, wherein the nucleic acid variance increases kinase activity.

54. The method of paragraphs 44 and 45, wherein the variance in the kinase domain of the erbB1 gene effects the conformation of the ATP -binding pocket.

55. The method of paragraphs 44 and 45, wherein the variance in the kinase domain of erbB1 is in an exon of the erbB1 gene selected from the group consisting of exon 18, 19, 20 or 21.

56. The method of paragraph 55, wherein the variance is in exon 18, 19 or 21.

57. The method of paragraph 44 and 45, wherein the variance in the kinase domain of the erbB1 gene is an in frame deletion or a substitution.

58. The method of paragraph 57, wherein the in frame deletion is in exon 19 of the erbB1 gene.

59. The method of paragraph 58, wherein the in frame deletion in exon 19 of the erbB1 gene comprises a deletion of at least amino acids leucine, arginine, glutamic acid and alanine, at codons 747, 748, 749, and 750.

60. The method of paragraph 58, wherein the in frame deletion in exon 19 of the erbB1 gene comprises a deletion of at least amino acids leucine, arginine, and glutamine acid at codon 747, 748 and 749 of SEQ. ID. NO. 512.

61. The method of paragraph 59, wherein the in frame deletion comprises nucleotides selected from the group consisting of 2235 to 2249, 2240 to 2251, and 2240 to 2257 of SEQ ID NO: 511.

62. The method of paragraph 57, wherein the substitution is in exon 21 of the erbB1 gene.

63. The method of paragraph 62, wherein the substitution in exon 21 comprises a substitution from the group consisting of a guanine for a thymine at nucleotide 2573 of SEQ ID NO: 511, and an adenine for a thymine at nucleotide 2582 of SEQ ID NO: 511.

64. The method of paragraph 62, wherein the substitution in exon 21 comprises at least one amino acid.

65. The method of paragraph 57, wherein the substitution is in exon 18 of the erbB1 gene.

66. The method of paragraph 65, wherein the substitution in exon 18 is a thymine for a guanine at nucleotide 2155 of SEQ ID NO: 511.

67. A kit comprising:

- a. at least one degenerate primer pair designed to anneal to nucleic acid regions bordering or within Exon 18, 19, 20 or 21 of the EGFR kinase domain;
- b. products and reagents required to carry out PCR amplification; and
- c. instructions.

68. A kit according to paragraph 67, wherein the primer comprises the sequence primers selected from the group consisting of SEQ ID NOS: 505-508, and SEQ ID NOS. 646-673 with SEQ ID NO. 645 on the 5' end of all forward primers and SEQ ID NO. 674 on the 5' end of all reverse primers.

69. A kit comprising:

- a. at least one probe designed to anneal to nucleic acid regions within exons 18, 19, 20 or 21 of the EGFR kinase domain;
- b. products and reagents required to carry out the annealing reaction; and
- c. instructions.

70. The kit of paragraph 69 wherein at least one probe is bound to a solid support.

71. A kit comprising:

- a. at least one probe designed to bind to the ATP-binding pocket of the EGFR kinase domain protein;
- b. the products and reagents required to carry out the binding reaction; and
- c. instructions.

72. The kit according to paragraph 71, wherein the probe is an antibody, antibody fragment, or chimeric antibody.

73. The kit according to paragraph 72, wherein the probe further comprises a detectable label.

74. A method for selecting a compound that inhibits the catalytic kinase activity of a variant epidermal growth factor receptor (EGFR), comprising:

- a. contacting the compound with a variant EGFR; and
- b. detecting the resultant kinase activity of the variant EGFR, wherein a compound is selected that inhibits the kinase activity of the variant EGFR.

75. The method of paragraph 74, wherein the variant EGFR is labeled.

76. The method of paragraph 74, wherein the variant EGFR is bound to a solid support.

77. The method of paragraph 74, wherein said solid support is a protein chip.

78. A compound that inhibits the catalytic kinase activity of a variant EGFR as identified in paragraph 74, wherein the compound is selected from the group consisting of an antibody, antibody fragment, small molecule, peptide, protein, antisense nucleic acid, ribozyme, PNA, siRNA, oligonucleotide aptamer, and peptide aptamer.

79. The method of paragraph 74, wherein the variant EGFR comprises a secondary mutation in the kinase domain of the erbB1 gene.

80. A pharmaceutical composition comprising the inhibitor of EGFR kinase activity identified in paragraphs 74-79.

81. A method of treating a patient having an EGFR mediated disease, comprising administering to said patient the pharmaceutical composition of paragraph 80.

82. The method of paragraph 81, wherein the EGFR mediated disease is cancer.

83. The method of paragraph 82, wherein the cancer is selected from the group consisting of gastrointestinal cancer, prostate cancer, ovarian cancer, breast cancer, head and neck cancer, lung cancer, non-small cell lung cancer, cancer of the nervous system, kidney cancer, retina cancer, skin cancer, liver cancer, pancreatic cancer, genital-urinary cancer and bladder cancer.

84. The method of paragraph 83 wherein the cancer is non-small cell lung cancer.

85. A method for predicting the acquisition of secondary mutations in the kinase domain of the erbB1 gene comprising:

- a. contacting a cell having a variant form of the erbB1 gene with a sub-lethal dose of a tyrosine kinase inhibitor,
- b. selecting cells that are resistant to a growth arrest effect of the tyrosine kinase inhibitor; and
- c. analyzing the erbB1 nucleic acid from said resistant cells for the presence of secondary mutations in the

erbB1 kinase domain.

86. The method of paragraph 85, wherein the cell is in vitro.

87. The method of paragraph 85, wherein the cell is obtained from a transgenic animal.

88. The method of paragraph 87, wherein the transgenic animal is a mouse.

89. The method of paragraph 87, wherein the cell is obtained from a tumor biopsy.

90. The method of paragraph 85, further comprising first contacting the cells with an effective amount of a mutagenizing agent.

91. The method of paragraph 90, wherein the mutagenizing agent is selected from the group consisting of ethyl methanesulfonate (EMS), N-ethyl-N-nitrosourea (ENU), N-methyl-N-nitrosourea (MNU), phorbaxine hydrochloride (Prc), methyl methanesulfonate (MeMS), chlorambucil (Chl), melphalan, porcarbazine hydrochloride, cyclophosphamide (Cp), diethyl sulfate (Et_2SO_4), acrylamide monomer (AA), triethylene melamin (TEM), nitrogen mustard, vincristine, dimethylnitrosamine, N-methyl-N'-nitro-Nitrosoguanidine (MNNG), 7,12 dimethylbenz(a)anthracene (DMBA), ethylene oxide, hexamethylphosphoramide, bisulfan, and ethyl methanesulfonate (EtMs).

92. The method of paragraph 85, further comprising propagating a variant form of the EGFR gene in a DNA repair-deficient bacterial strain before introducing it into a cell.

93. A method for selecting a compound comprising contacting the compound with a variant epidermal growth factor receptor (EGFR) having a secondary mutation in the kinase domain and detecting the resulted kinase activity, wherein a compound is selected that inhibits the kinase activity of the variant EGFR.

94. The method of paragraph 93, wherein the secondary mutation results in a resistance to gefitinib or erlotinib.

95. An isolated nucleic acid comprising SEQ ID NO: 495.

96. An isolated nucleic acid comprising SEQ ID NO: 511, wherein nucleotides 2235 to 2249 are deleted.

97. An isolated nucleic acid comprising SEQ ID NO:497.

98. An isolated nucleic acid comprising SEQ ID NO: 511, wherein nucleotides 2240 to 2251 are deleted.

99. An isolated nucleic acid comprising SEQ ID NO:499.

100. An isolated nucleic acid comprising SEQ ID NO: 511, wherein nucleotides 2240 to 2257 are deleted.

101. An isolated nucleic acid comprising SEQ ID NO:502.

102. An isolated nucleic acid comprising SEQ ID NO: 511, wherein the guanine at nucleotide 2573 is substituted for a thymine.

103. An isolated nucleic acid comprising SEQ ID NO: 504.

104. An isolated nucleic acid comprising SEQ ID NO: 511, wherein the adenine at nucleotide 2582 is substituted for a thymine.

105. An isolated nucleic acid of SEQ ID NO: 511, wherein a thymine at nucleotide 2155 is substituted for a guanine.

106. An isolated protein having amino acid sequence of SEQ ID NO: 512, wherein amino acids 746 to 750 are deleted.

107. An isolated protein having amino acid sequence of SEQ ID NO: 512, where in amino acids 747 to 751 are deleted.

108. An isolated protein having amino acid sequence of SEQ ID NO: 512, wherein amino acids 747 to 753 are deleted.

109. An isolated protein having amino acid sequence of SEQ ID NO: 512, wherein the Leucine at amino acid 858 is substituted with an Arginine.

110. An isolated protein having amino acid sequence of SEQ ID NO: 512, wherein the Leucine at amino acid 861 is substituted with a Glutamine.

111. An isolated protein having amino acid sequence of SEQ ID NO: 512, wherein the Glycine at amino acid 719 is substituted with a Cysteine.

112. The method of paragraph 1, wherein the nucleic acid variance of the erbB1 gene is selected from the group consisting of a substitution of a thymine for a guanine or a serine for a guanine at nucleotide 2155 of SEQ ID NO 511, a deletion of nucleotides 2235 to 2249, 2240 to 2251, 2240 to 2257, 2236 to 2250, 2254 to 2277, or 2236 to 2244 of SEQ ID NO 511, and a substitution of a guanine for a thymine at nucleotide 2573 or an adenine for a thymine at nucleotide 2582 of SEQ ID NO 511.

113. A method for determining the likelihood of effectiveness of an epidermal growth factor receptor (EGFR) targeting treatment in a patient affected with or at risk for developing cancer comprising:

- a. obtaining a biological sample from said patient; and
- b. determining whether Akt, STAT5, or STAT3 are activated in said patient, wherein activated Akt, STAT5, or STAT3 indicates that said EGFR targeting treatment is likely to be effective.

114. The method, of paragraph 113, wherein the biological sample is a biopsy or an aspirate.

115. The method of paragraph 113, wherein activated Akt, STAT3, or STAT5 is phosphorylated.

116. The method of paragraph 113, wherein the activated Akt, STAT5, or STAT3 is determined immunologically.

117. The method of paragraph 116, wherein the immunological detection methods are selected from the group consisting of immunohistochemistry, immunocytochemistry, FACS scanning, immunoblotting, radioimmunoassays, western blotting, immunoprecipitation, or enzyme-linked immunosorbant assays (ELISA).

118. The method of paragraph 116, wherein the immunological detection method is immunohistochemistry or immunocytochemistry using anti-phospho Akt, anti-phospho STAT3 or anti-phospho STAT5 antibodies.

SEQUENCE LISTING

[0348]

<110> THE GENERAL HOSPITAL CORPORATION DANA-FARBER CANCER INSTITUTE, INC.

<120> METHOD TO DETERMINE RESPONSIVENESS OF CANCER TO EPIDERMAL GROWTH FACTOR RECEPTOR TARGETING TREATMENTS

<130> 030258-55145

<140> PCT/US05/10645

<141> 2005-03-31

<150> 60/558,218

<151> 2004-03-31

<150> 60/561,095

<151> 2004-04-09

<150> 60/565,753

<151> 2004-04-27

<150> 60/565,985
 <151> 2004-04-27

5 <150> 60/574,035
 <151> 2004-05-25

 <150> 60/577,916
 <151> 2004-06-07

10 <150> 60/592,287
 <151> 2004-07-29

 <160> 762

15 <170> PatentIn Ver. 3.3

 <210> 1
 <211> 23
 <212> DNA
 20 <213> Homo sapiens

 <400> 1
 ggaaatatag ggaagggaag gaa 23

25 <210> 2
 <211> 22
 <212> DNA
 <213> Homo sapiens

30 <400> 2
 ttgacagggt accaggagat ga 22

 <210> 3
 <211> 20
 35 <212> DNA
 <213> Homo sapiens

 <400> 3
 ctgaaccgcc aaggactcat 20

40 <210> 4
 <211> 23
 <212> DNA
 <213> Homo sapiens

45 <400> 4
 tttccctcc ctactaacac acg 23

 <210> 5
 50 <211> 20
 <212> DNA
 <213> Homo sapiens

 <400> 5
 55 actgatgccc tgaccctgtt 20

 <210> 6
 <211> 19

<212> DNA
 <213> Homo sapiens

5 <400> 6
 cccatgggtc cccactctt 19

10 <210> 7
 <211> 20
 <212> DNA
 <213> Homo sapiens

15 <400> 7
 agggactcca aagccatgtg 20

20 <210> 8
 <211> 19
 <212> DNA
 <213> Homo sapiens

25 <400> 8
 ctctctgggg ccatccact 19

30 <210> 9
 <211> 23
 <212> DNA
 <213> Homo sapiens

35 <400> 9
 cattgtcaag ggcaatgtaa gtg 23

40 <210> 10
 <211> 20
 <212> DNA
 <213> Homo sapiens

45 <400> 10
 ctctcaccaa ccctcgctgt 20

50 <210> 11
 <211> 20
 <212> DNA
 <213> Homo sapiens

55 <400> 11
 acatggggag ccagagtgac 20

60 <210> 12
 <211> 20
 <212> DNA
 <213> Homo sapiens

65 <400> 12
 tgcaaccag agaaagtgtg 20

70 <210> 13
 <211> 24
 <212> DNA
 <213> Homo sapiens

<400> 13
 tgagctttca accctagttt gttg 24

5
 <210> 14
 <211> 20
 <212> DNA
 <213> Homo sapiens

10
 <400> 14
 gtttgctctc tgctgtctca 20

15
 <210> 15
 <211> 24
 <212> DNA
 <213> Homo sapiens

20
 <400> 15
 tgccttggtg ttttgaaga ttcc 24

25
 <210> 16
 <211> 23
 <212> DNA
 <213> Homo sapiens

30
 <400> 16
 tgcagacaga tgacaaacat gaa 23

35
 <210> 17
 <211> 21
 <212> DNA
 <213> Homo sapiens

40
 <400> 17
 tgggtgagtc tctgttgga g 21

45
 <210> 18
 <211> 25
 <212> DNA
 <213> Homo sapiens

50
 <400> 18
 cattgccata gcaaaaataa acaca 25

55
 <210> 19
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 19
 ggttcaactg ggcgtccta 19

<210> 20
 <211> 20
 <212> DNA
 <213> Homo sapiens

<400> 20
 ccttctccga ggtggaattg 20

<210> 21
 <211> 20
 <212> DNA
 <213> Homo sapiens
 5
 <400> 21
 cgccacatgg catctcttta 20
 10
 <210> 22
 <211> 23
 <212> DNA
 <213> Homo sapiens
 15
 <400> 22
 aaaacgatct ctatgtccgt ggt 23
 20
 <210> 23
 <211> 20
 <212> DNA
 <213> Homo sapiens
 25
 <400> 23
 cagccagcca aacaatcaga 20
 30
 <210> 24
 <211> 23
 <212> DNA
 <213> Homo sapiens
 35
 <400> 24
 tctttggagt cttcagaggg aaa 23
 40
 <210> 25
 <211> 20
 <212> DNA
 <213> Homo sapiens
 45
 <400> 25
 tgtggttcg ttgaagcaa 20
 50
 <210> 26
 <211> 20
 <212> DNA
 <213> Homo sapiens
 55
 <400> 26
 aattgacagc tccccacag 20
 60
 <210> 27
 <211> 20
 <212> DNA
 <213> Homo sapiens
 65
 <400> 27
 ggctttctga cgggagtaa 20
 70
 <210> 28
 <211> 21

<212> DNA
 <213> Homo sapiens

5 <400> 28
 ccacccaaag actctccaag a 21

10 <210> 29
 <211> 20
 <212> DNA
 <213> Homo sapiens

15 <400> 29
 cctttccatc acccctcaag 20

20 <210> 30
 <211> 20
 <212> DNA
 <213> Homo sapiens

25 <400> 30
 agtgccttc cattgcctaa 20

30 <210> 31
 <211> 20
 <212> DNA
 <213> Homo sapiens

35 <400> 31
 accggaattc cttcctgctt 20

40 <210> 32
 <211> 23
 <212> DNA
 <213> Homo sapiens

45 <400> 32
 cactgaaaca aacaacagg tga 23

50 <210> 33
 <211> 20
 <212> DNA
 <213> Homo sapiens

55 <400> 33
 agggggtgag tcacaggttc 20

60 <210> 34
 <211> 26
 <212> DNA
 <213> Homo sapiens

65 <400> 34
 tcagaagaaa tgttttatt ccaagg 26

70 <210> 35
 <211> 21
 <212> DNA
 <213> Homo sapiens

	<400> 35	
	gcaaatccaa tttccact t	21
5	<210> 36	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 36	
	gcaggagctc tgtgccctat	20
	<210> 37	
	<211> 20	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 37	
	tccacagca tgacctacca	20
20	<210> 38	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 38	
	ttgcttctt aaggaactga aaa	23
	<210> 39	
30	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 39	
35	tgtcacccaa ggtcatggag	20
	<210> 40	
	<211> 20	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 40	
	caaaagccaa gggcaaagaa	20
45	<210> 41	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 41	
	ggagtccaa ctcctgacc	20
	<210> 42	
	<211> 19	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 42	
	gtcctgccca cacaggatg	19

<210> 43
 <211> 20
 <212> DNA
 <213> Homo sapiens
 5
 <400> 43
 gctttcccca ctcacacaca 20
 10
 <210> 44
 <211> 20
 <212> DNA
 <213> Homo sapiens
 15
 <400> 44
 caaacctcgg caattgttg 20
 20
 <210> 45
 <211> 20
 <212> DNA
 <213> Homo sapiens
 25
 <400> 45
 ccaccaatcc aacatccaga 20
 30
 <210> 46
 <211> 20
 <212> DNA
 <213> Homo sapiens
 35
 <400> 46
 tggcccagag ccatagaaac 20
 40
 <210> 47
 <211> 26
 <212> DNA
 <213> Homo sapiens
 45
 <400> 47
 ttccaagatc attctacaag atgtca 26
 50
 <210> 48
 <211> 23
 <212> DNA
 <213> Homo sapiens
 55
 <400> 48
 gcacattcag agattcttc tgc 23
 60
 <210> 49
 <211> 20
 <212> DNA
 <213> Homo sapiens
 65
 <400> 49
 tccaaatgag ctggcaagtg 20
 70
 <210> 50
 <211> 23

<212> DNA
 <213> Homo sapiens

5 <400> 50
 tcccaaacac tcagtgaac aaa 23

10 <210> 51
 <211> 20
 <212> DNA
 <213> Homo sapiens

15 <400> 51
 gtgcatcgct ggtaacatcc 20

20 <210> 52
 <211> 20
 <212> DNA
 <213> Homo sapiens

25 <400> 52
 tgtggagatg agcagggtct 20

30 <210> 53
 <211> 20
 <212> DNA
 <213> Homo sapiens

35 <400> 53
 atcgcatcca tgcgtcttca 20

40 <210> 54
 <211> 20
 <212> DNA
 <213> Homo sapiens

45 <400> 54
 atcccatgg caaactcttg 20

50 <210> 55
 <211> 19
 <212> DNA
 <213> Homo sapiens

55 <400> 55
 gctcagagcc tggcatgaa 19

60 <210> 56
 <211> 19
 <212> DNA
 <213> Homo sapiens

65 <400> 56
 catcctcccc tgcattgt 19

70 <210> 57
 <211> 20
 <212> DNA
 <213> Homo sapiens

	<400> 57		
	tggtctgtct gtgtgtgtca	20	
5	<210> 58		
	<211> 25		
	<212> DNA		
	<213> Homo sapiens		
10	<400> 58		
	cgaaagaaaa tacttgcattg tcaga	25	
	<210> 59		
	<211> 20		
	<212> DNA		
15	<213> Homo sapiens		
	<400> 59		
	tgaagcaaatt tgccaagac	20	
20	<210> 60		
	<211> 20		
	<212> DNA		
	<213> Homo sapiens		
25	<400> 60		
	tgacatttct ccagggatgc	20	
	<210> 61		
	<211> 20		
30	<212> DNA		
	<213> Homo sapiens		
	<400> 61		
	aagtgtcgca tcaccaatgc	20	
35	<210> 62		
	<211> 19		
	<212> DNA		
	<213> Homo sapiens		
40	<400> 62		
	atgcgatctg ggacacagg	19	
	<210> 63		
45	<211> 20		
	<212> DNA		
	<213> Homo sapiens		
	<400> 63		
50	ggcacctgct ggcaatagac	20	
	<210> 64		
	<211> 27		
	<212> DNA		
55	<213> Homo sapiens		
	<400> 64		
	tgacttcata tccatgtgag ttctact	27	

	<210> 65	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
5		
	<400> 65	
	tataacctcc atgaggcaca	20
10		
	<210> 66	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
15		
	<400> 66	
	gggaaaaacc cacacaggaa	20
20		
	<210> 67	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 67	
	tcagaaccag catctcaagg a	21
25		
	<210> 68	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
30		
	<400> 68	
	gatgctggag ggagcacct	19
35		
	<210> 69	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
40		
	<400> 69	
	ccttggtgag gacattcaca gg	22
45		
	<210> 70	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 70	
	atgtgccga ggtggaagta	20
50		
	<210> 71	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
55		
	<400> 71	
	ggagggcaga ggactagctg	20
	<210> 72	
	<211> 20	

<212> DNA
 <213> Homo sapiens

5 <400> 72
 gtgcctggcc aagtctttgt 20

10 <210> 73
 <211> 20
 <212> DNA
 <213> Homo sapiens

15 <400> 73
 ctgcagccta gcaacagagc 20

20 <210> 74
 <211> 20
 <212> DNA
 <213> Homo sapiens

25 <400> 74
 aagaaccaga ggagccagga 20

30 <210> 75
 <211> 20
 <212> DNA
 <213> Homo sapiens

35 <400> 75
 cgggtaagga tgtgggtgt 20

40 <210> 76
 <211> 20
 <212> DNA
 <213> Homo sapiens

45 <400> 76
 caggtgttct gcctcctgaa 20

50 <210> 77
 <211> 20
 <212> DNA
 <213> Homo sapiens

55 <400> 77
 gcttcaggag gcagaacacc 20

60 <210> 78
 <211> 22
 <212> DNA
 <213> Homo sapiens

65 <400> 78
 ggagcaagcc taagaaggtt ca 22

70 <210> 79
 <211> 21
 <212> DNA
 <213> Homo sapiens

	<400> 79	
	gccttgatc cattgccac a	21
5	<210> 80	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 80	
	tgacaacacg tttgggtca t	21
	<210> 81	
	<211> 23	
	<212> DNA	
15	<213> Homo sapiens	
	<400> 81	
	tgcatattcc attcagaac aga	23
20	<210> 82	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 82	
	aaacagtttc attgctgcta aat	23
	<210> 83	
	<211> 23	
30	<212> DNA	
	<213> Homo sapiens	
	<400> 83	
	ccggatacag atacccaaaa aga	23
35	<210> 84	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
40	<400> 84	
	ggaggcttca agggatgaga	20
	<210> 85	
45	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 85	
50	gctgtgtcc tgcttgcta	20
	<210> 86	
	<211> 25	
	<212> DNA	
55	<213> Homo sapiens	
	<400> 86	
	tggtgtaat gttgaactag cttgc	25

<210> 87
 <211> 24
 <212> DNA
 <213> Homo sapiens
 5
 <400> 87
 tggctgtcag ctaaataagc atgt 24
 10
 <210> 88
 <211> 24
 <212> DNA
 <213> Homo sapiens
 15
 <400> 88
 tcaatttgct tcatttctcc tggt 24
 20
 <210> 89
 <211> 21
 <212> DNA
 <213> Homo sapiens
 25
 <400> 89
 tgctgctgaa ctaccaacca a 21
 30
 <210> 90
 <211> 24
 <212> DNA
 <213> Homo sapiens
 35
 <400> 90
 tgtggtagta attgtggaaa actg 24
 40
 <210> 91
 <211> 20
 <212> DNA
 <213> Homo sapiens
 45
 <400> 91
 caaagcaccg tctcaactcg 20
 50
 <210> 92
 <211> 20
 <212> DNA
 <213> Homo sapiens
 55
 <400> 92
 cccgaaactg ccaacttcat 20
 60
 <210> 93
 <211> 20
 <212> DNA
 <213> Homo sapiens
 65
 <400> 93
 ggaaaacagg accccagtgt 20
 70
 <210> 94
 <211> 20

<212> DNA
 <213> Homo sapiens

5 <400> 94
 ccctcctcca cagagctgat 20

10 <210> 95
 <211> 21
 <212> DNA
 <213> Homo sapiens

15 <400> 95
 gacagaagct gacaagcagc a 21

20 <210> 96
 <211> 20
 <212> DNA
 <213> Homo sapiens

25 <400> 96
 aggttcatt ccctcccagt 20

30 <210> 97
 <211> 22
 <212> DNA
 <213> Homo sapiens

35 <400> 97
 tgggagtgag agtttgaag aa 22

40 <210> 98
 <211> 20
 <212> DNA
 <213> Homo sapiens

45 <400> 98
 tatgaggccg tgagctgaaa 20

50 <210> 99
 <211> 18
 <212> DNA
 <213> Homo sapiens

55 <400> 99
 agggccctgc tctggtt 18

60 <210> 100
 <211> 20
 <212> DNA
 <213> Homo sapiens

65 <400> 100
 ccaattgggc gttagtgaag 20

70 <210> 101
 <211> 20
 <212> DNA
 <213> Homo sapiens

	<400> 101	
	ctcatgagat tggggcatca	20
5	<210> 102	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 102	
	aggcccatga tctcagaagc	20
	<210> 103	
	<211> 20	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 103	
	ggttgagga gagacgaggt	20
20	<210> 104	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 104	
	aggcccttca ccctgtgac	19
	<210> 105	
30	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 105	
35	atgaccctc cgatcctacc	20
	<210> 106	
	<211> 20	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 106	
	taatcctgct ccacggcatt	20
	<210> 107	
45	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 107	
50	ggaaaaagca gaggcaggtg	20
	<210> 108	
	<211> 20	
	<212> DNA	
55	<213> Homo sapiens	
	<400> 108	
	tggtctcaag aaccagcag	20

	<210> 109	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
5		
	<400> 109	
	gacaccctcc ccctctcat	19
10		
	<210> 110	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
15		
	<400> 110	
	actatgacac cccggctgag	20
20		
	<210> 111	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 111	
	tgcttgatgt aaaacccttg g	21
25		
	<210> 112	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
30		
	<400> 112	
	gcaatccaac agccatgaga	20
35		
	<210> 113	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
40		
	<400> 113	
	ggagcaaacc cctatgtcca	20
45		
	<210> 114	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 114	
	tcctccaact gtgtgtgtg g	21
50		
	<210> 115	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
55		
	<400> 115	
	tggggaccac tgctgagag	19
	<210> 116	
	<211> 20	

	<212> DNA	
	<213> Homo sapiens	
5	<400> 116 tgcagccttc tctccttgaa	20
	<210> 117	
	<211> 20	
10	<212> DNA	
	<213> Homo sapiens	
	<400> 117 gcagagcagt gtggcagaag	20
15	<210> 118	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
20	<400> 118 acaggtggga agggactgg	19
	<210> 119	
	<211> 19	
25	<212> DNA	
	<213> Homo sapiens	
	<400> 119 agtggggtgg gctgagaac	19
30	<210> 120	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
35	<400> 120 tctctggggc agaaagagga	20
	<210> 121	
40	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 121 acccggccac actgtatttc	20
45	<210> 122	
	<211> 20	
	<212> DNA	
50	<213> Homo sapiens	
	<400> 122 catcccaccc agctctcaac	20
55	<210> 123	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	

	<400> 123	
	agggcatagc cctattgagc	20
5	<210> 124	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 124	
	cccaggaaaa agccagagaa	20
	<210> 125	
	<211> 20	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 125	
	caggtgtggg tggagtaggc	20
20	<210> 126	
	<211> 18	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 126	
	ctcaggcgcc atccactt	18
	<210> 127	
30	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 127	
35	aagaagacga ccaacgtgag c	21
	<210> 128	
	<211> 18	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 128	
	aggagctcca gggcacag	18
45	<210> 129	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 129	
	cctcctctgt aaagtgggtg ga	22
	<210> 130	
	<211> 20	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 130	
	agagggcctc agtgcagagt	20

<210> 131
 <211> 20
 <212> DNA
 <213> Homo sapiens
 5
 <400> 131
 agatggggca gaactggatg 20
 10
 <210> 132
 <211> 20
 <212> DNA
 <213> Homo sapiens
 15
 <400> 132
 ggggtcccaga ccaaattctga 20
 20
 <210> 133
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 133
 aggtgctccc ttcacagcat 20
 25
 <210> 134
 <211> 22
 <212> DNA
 <213> Homo sapiens
 30
 <400> 134
 ttcagggact acagctgagg aa 22
 35
 <210> 135
 <211> 21
 <212> DNA
 <213> Homo sapiens
 <400> 135
 gccgtatgtt atctgggagg t 21
 40
 <210> 136
 <211> 21
 <212> DNA
 <213> Homo sapiens
 45
 <400> 136
 tgggccatt acactttaag a 21
 50
 <210> 137
 <211> 20
 <212> DNA
 <213> Homo sapiens
 55
 <400> 137
 ttccatcacc ggtacctcct 20
 <210> 138
 <211> 23

<212> DNA
 <213> Homo sapiens

5 <400> 138
 ccataaatca aaaatgcacc aca 23

10 <210> 139
 <211> 23
 <212> DNA
 <213> Homo sapiens

15 <400> 139
 gagtggctt aggaagatga tgc 23

20 <210> 140
 <211> 22
 <212> DNA
 <213> Homo sapiens

25 <400> 140
 aaagtcattg gctgcaatac aa 22

30 <210> 141
 <211> 18
 <212> DNA
 <213> Homo sapiens

35 <400> 141
 atgggtcccca ctgcttgg 18

40 <210> 142
 <211> 20
 <212> DNA
 <213> Homo sapiens

45 <400> 142
 aggagctcac ctcaccctgt 20

50 <210> 143
 <211> 20
 <212> DNA
 <213> Homo sapiens

55 <400> 143
 ccttgctgt ctccacacat 20

60 <210> 144
 <211> 20
 <212> DNA
 <213> Homo sapiens

65 <400> 144
 tggcaacggg taacaatgaa 20

70 <210> 145
 <211> 20
 <212> DNA
 <213> Homo sapiens

	<400> 145	
	ggctgaggta agctgcttcg	20
5	<210> 146	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 146	
	aaaaagaagt atctgcccc ttt	23
	<210> 147	
	<211> 21	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 147	
	aaccctctt agggctctgt g	21
20	<210> 148	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 148	
	caggaggatg gcaggcttc	19
	<210> 149	
30	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
	<400> 149	
35	cgtagagagc ttcaggacct gtg	23
	<210> 150	
	<211> 20	
40	<212> DNA	
	<213> Homo sapiens	
	<400> 150	
	ttccgagaag tttgcctga	20
45	<210> 151	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 151	
	tgtgaacatc attcaaggcg ta	22
	<210> 152	
55	<211> 25	
	<212> DNA	
	<213> Homo sapiens	
	<400> 152	
	aaaatgtgtg atatccctag acagg	25

	<210> 153	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
5		
	<400> 153	
	tccacatttc agcaacagca	20
10		
	<210> 154	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
15		
	<400> 154	
	ggctgcttc tgagacacag t	21
20		
	<210> 155	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
25		
	<400> 155	
	tatctaccgg tgcgggactt	20
30		
	<210> 156	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
35		
	<400> 156	
	aggtgtagcc tccctcaca	20
40		
	<210> 157	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
45		
	<400> 157	
	aggctggtgg tgtctctgtg	20
50		
	<210> 158	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
55		
	<400> 158	
	caagctgcca accctcagtt	20
60		
	<210> 159	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
65		
	<400> 159	
	tggatttcaa atactgaagc cact	24
70		
	<210> 160	
	<211> 22	

	<212> DNA	
	<213> Homo sapiens	
5	<400> 160 tgggaattggt ggtgttgaat tt	22
	<210> 161	
	<211> 23	
10	<212> DNA	
	<213> Homo sapiens	
	<400> 161 gggcttcata tgttctgaca tgg	23
15	<210> 162	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
20	<400> 162 cagaggacca cgccatagg	19
	<210> 163	
	<211> 20	
25	<212> DNA	
	<213> Homo sapiens	
	<400> 163 ccgagattta gccaccagga	20
30	<210> 164	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
35	<400> 164 cctgggaagc aaacaacaca	20
	<210> 165	
40	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 165 aggtcccag tctctctcc	20
45	<210> 166	
	<211> 20	
	<212> DNA	
50	<213> Homo sapiens	
	<400> 166 agacccatgc agccatccta	20
55	<210> 167	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	

	<400> 167	
	cgatgaaccac cgagcttgt	19
5	<210> 168	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 168	
	agaggggcag aaggggaac	19
	<210> 169	
	<211> 20	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 169	
	ggtgggggtg aggagcttag	20
20	<210> 170	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 170	
	tcgtttaagc caccagctca	20
	<210> 171	
30	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 171	
35	tgcaaataag gaaagcaaac a	21
	<210> 172	
	<211> 21	
40	<212> DNA	
	<213> Homo sapiens	
	<400> 172	
	tcttgacatg gtctccaac c	21
	<210> 173	
45	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 173	
50	cagcatcttc acacacctct ga	22
	<210> 174	
	<211> 20	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 174	
	gctggctcta aatccacct	20

<210> 175
 <211> 20
 <212> DNA
 <213> Homo sapiens
 5
 <400> 175
 ctaatccggg aagttgtgc 20
 10
 <210> 176
 <211> 24
 <212> DNA
 <213> Homo sapiens
 15
 <400> 176
 ttctgtatca gcagcttctc tgtg 24
 20
 <210> 177
 <211> 19
 <212> DNA
 <213> Homo sapiens
 25
 <400> 177
 caagtgccac catggatca 19
 30
 <210> 178
 <211> 21
 <212> DNA
 <213> Homo sapiens
 35
 <400> 178
 ggcagtgtag tgaccccttg a 21
 40
 <210> 179
 <211> 24
 <212> DNA
 <213> Homo sapiens
 45
 <400> 179
 gcacaagtta ttaagagccc aagg 24
 50
 <210> 180
 <211> 20
 <212> DNA
 <213> Homo sapiens
 55
 <400> 180
 agcatactgg cctcacacca 20
 60
 <210> 181
 <211> 22
 <212> DNA
 <213> Homo sapiens
 65
 <400> 181
 gcacatgggc agtgttgtat tt 22
 70
 <210> 182
 <211> 20

	<212> DNA	
	<213> Homo sapiens	
5	<400> 182 gagccccaca cagatttcct	20
	<210> 183	
	<211> 20	
10	<212> DNA	
	<213> Homo sapiens	
	<400> 183 atgggacgga gaagtgggtg	20
15	<210> 184	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
20	<400> 184 tccctgtatc agggctcgtc	20
	<210> 185	
	<211> 20	
25	<212> DNA	
	<213> Homo sapiens	
	<400> 185 ttcctacgca gcacaccaat	20
30	<210> 186	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
35	<400> 186 gcaggcacta aaccctttcc	20
	<210> 187	
40	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 187 gcacgcatgt gaccaatttc	20
45	<210> 188	
	<211> 20	
	<212> DNA	
50	<213> Homo sapiens	
	<400> 188 agccctgaga gggaggtagg	20
55	<210> 189	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	

	<400> 189	
	cacacaccac ccctctgct	19
5	<210> 190	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 190	
	aaagatttgg ggtgagggct a	21
	<210> 191	
	<211> 20	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 191	
	ctgaaagctc agggataggg	20
20	<210> 192	
	<211> 18	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 192	
	ctggccaagc tgcacaga	18
	<210> 193	
30	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 193	
35	tgcagccaac gatttgaaag	20
	<210> 194	
	<211> 20	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 194	
	ggaaagcccc aagtctgaaa	20
45	<210> 195	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 195	
	tcatcatgag atccccacac t	21
	<210> 196	
	<211> 20	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 196	
	gcatttcccc ctgaaggagt	20

<210> 197
 <211> 23
 <212> DNA
 <213> Homo sapiens
 5
 <400> 197
 tggattcagt aaccaggaag tga 23
 10
 <210> 198
 <211> 20
 <212> DNA
 <213> Homo sapiens
 15
 <400> 198
 cccattccac caggatgatt 20
 20
 <210> 199
 <211> 20
 <212> DNA
 <213> Homo sapiens
 25
 <400> 199
 gttccagct gccactacc 20
 30
 <210> 200
 <211> 20
 <212> DNA
 <213> Homo sapiens
 35
 <400> 200
 gctcgaaacc acatgttcca 20
 40
 <210> 201
 <211> 20
 <212> DNA
 <213> Homo sapiens
 45
 <400> 201
 ctggattgg ggttctctgc 20
 50
 <210> 202
 <211> 22
 <212> DNA
 <213> Homo sapiens
 55
 <400> 202
 cggaacagc tagcagattt tt 22
 60
 <210> 203
 <211> 20
 <212> DNA
 <213> Homo sapiens
 65
 <400> 203
 gggaatttg gaggggaact 20
 70
 <210> 204
 <211> 20

	<212> DNA	
	<213> Homo sapiens	
5	<400> 204 gcttcagtca ccacagagca	20
	<210> 205	
	<211> 23	
10	<212> DNA	
	<213> Homo sapiens	
	<400> 205 tgagtctacc cagcaatcat ttg	23
15	<210> 206	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
20	<400> 206 ttcccagagag ctacaggaca	20
	<210> 207	
	<211> 21	
25	<212> DNA	
	<213> Homo sapiens	
	<400> 207 ggtaacaagg gtaccacga a	21
30	<210> 208	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
35	<400> 208 gtttgagggg ctgagtgtg	20
	<210> 209	
40	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 209 cctcaccctt agggcttg	20
45	<210> 210	
	<211> 20	
	<212> DNA	
50	<213> Homo sapiens	
	<400> 210 agcccagggtc atgccttaga	20
55	<210> 211	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	

	<400> 211	
	gggtagcttg ggagcaaaga	20
5	<210> 212	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 212	
	ccaaacccca gagagcagac	20
	<210> 213	
	<211> 22	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 213	
	catttccct aatcctttc ca	22
20	<210> 214	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 214	
	gtgatcccag atttaggcct tc	22
	<210> 215	
30	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 215	
35	gcctctcgtg gttgttttg tc	22
	<210> 216	
	<211> 22	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 216	
	cccagggtag ggtccaataa tc	22
45	<210> 217	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 217	
	cttctggtg gaggtgactg at	22
	<210> 218	
	<211> 22	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 218	
	caggcatagt gtgtgatggt ca	22

<210> 219
 <211> 23
 <212> DNA
 <213> Homo sapiens
 5
 <400> 219
 tcacgataca cattctcaga tcc 23
 10
 <210> 220
 <211> 22
 <212> DNA
 <213> Homo sapiens
 15
 <400> 220
 gaagatctcc cagaggagga tg 22
 20
 <210> 221
 <211> 21
 <212> DNA
 <213> Homo sapiens
 <400> 221
 cgtaacgtgc tgttgaccaa t 21
 25
 <210> 222
 <211> 22
 <212> DNA
 <213> Homo sapiens
 30
 <400> 222
 aaacgagggga agagccagaa ag 22
 35
 <210> 223
 <211> 22
 <212> DNA
 <213> Homo sapiens
 <400> 223
 tggggagcac aataaaagaa ga 22
 40
 <210> 224
 <211> 22
 <212> DNA
 <213> Homo sapiens
 45
 <400> 224
 actcttgct cctggattct tg 22
 50
 <210> 225
 <211> 22
 <212> DNA
 <213> Homo sapiens
 55
 <400> 225
 ggaagtcagt gtcagggaa ta 22
 <210> 226
 <211> 22

<212> DNA
 <213> Homo sapiens

5 <400> 226
 ttttagcaga aataggcaag ca 22

10 <210> 227
 <211> 24
 <212> DNA
 <213> Homo sapiens

15 <400> 227
 tggtaatcct aaacacaatg caga 24

20 <210> 228
 <211> 22
 <212> DNA
 <213> Homo sapiens

25 <400> 228
 ctgggcaaca cagtgagatc ct 22

30 <210> 229
 <211> 23
 <212> DNA
 <213> Homo sapiens

35 <400> 229
 tcacaaattt ctttgctgtg tcc 23

40 <210> 230
 <211> 23
 <212> DNA
 <213> Homo sapiens

45 <400> 230
 catggaactc cagattagcc tgt 23

50 <210> 231
 <211> 22
 <212> DNA
 <213> Homo sapiens

55 <400> 231
 gattgttgca gatcgtggac at 22

60 <210> 232
 <211> 22
 <212> DNA
 <213> Homo sapiens

65 <400> 232
 cgcttaaadc ttccattcc ag 22

70 <210> 233
 <211> 22
 <212> DNA
 <213> Homo sapiens

	<400> 233	
	ctccatggca ccatcattaa ca	22
5	<210> 234	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 234	
	ctcaggacac aagtgctctg ct	22
	<210> 235	
	<211> 24	
	<212> DNA	
15	<213> Homo sapiens	
	<400> 235	
	gcagttcatg gttcatcttc tttt	24
20	<210> 236	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 236	
	caaaatagcc caccctggat ta	22
	<210> 237	
	<211> 21	
30	<212> DNA	
	<213> Homo sapiens	
	<400> 237	
	ctttctgcat tgcccaagat g	21
35	<210> 238	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
40	<400> 238	
	caaggtctca gtgagtggg ga	22
	<210> 239	
45	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
	<400> 239	
50	gagaagggtc ttctgactc tgc	23
	<210> 240	
	<211> 22	
	<212> DNA	
55	<213> Homo sapiens	
	<400> 240	
	caggtgttc tcctgtgagg tg	22

<210> 241
 <211> 22
 <212> DNA
 <213> Homo sapiens
 5
 <400> 241
 cacattgcgg cctagaatgt ta 22
 10
 <210> 242
 <211> 20
 <212> DNA
 <213> Homo sapiens
 15
 <400> 242
 accccgtcac aaccttcagt 20
 20
 <210> 243
 <211> 22
 <212> DNA
 <213> Homo sapiens
 <400> 243
 gccgtagccc caaaggtac ta 22
 25
 <210> 244
 <211> 22
 <212> DNA
 <213> Homo sapiens
 30
 <400> 244
 tcagctcaaa cctgtgattt cc 22
 35
 <210> 245
 <211> 24
 <212> DNA
 <213> Homo sapiens
 <400> 245
 ctactctcc ataaatgcta cgaa 24
 40
 <210> 246
 <211> 22
 <212> DNA
 <213> Homo sapiens
 45
 <400> 246
 gacttaacgt gtcccctttt gc 22
 50
 <210> 247
 <211> 22
 <212> DNA
 <213> Homo sapiens
 55
 <400> 247
 gcctcttcgg ggtaatcaga ta 22
 <210> 248
 <211> 22

	<212> DNA	
	<213> Homo sapiens	
5	<400> 248 gaagtctgtg gtttagcgga ca	22
	<210> 249	
	<211> 22	
10	<212> DNA	
	<213> Homo sapiens	
	<400> 249 atcttttgcc tggaggaact tt	22
15	<210> 250	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
20	<400> 250 cagggtaaat tcatccatt ga	22
	<210> 251	
	<211> 22	
25	<212> DNA	
	<213> Homo sapiens	
	<400> 251 cagcagccag cacaactact tt	22
30	<210> 252	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
35	<400> 252 ttggctagat gaaccattga tga	23
	<210> 253	
40	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 253 tgaatgaagc tcctgtgtt actc	24
	<210> 254	
	<211> 22	
	<212> DNA	
50	<213> Homo sapiens	
	<400> 254 atgttcatcg caggctaattg tg	22
55	<210> 255	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	

	<400> 255	
	aaaacaggga gaacttctaa gcaa	24
5	<210> 256	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 256	
	catggcagag tcattcccac t	21
	<210> 257	
	<211> 22	
	<212> DNA	
15	<213> Homo sapiens	
	<400> 257	
	caatgctaga acaacgcctg tc	22
20	<210> 258	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 258	
	tccctccact gaggacaaag tt	22
	<210> 259	
	<211> 22	
30	<212> DNA	
	<213> Homo sapiens	
	<400> 259	
	gggagagctt gagaaagttg ga	22
35	<210> 260	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
40	<400> 260	
	atttctcgg atggatgtac ca	22
	<210> 261	
45	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 261	
50	tcagagcctg tgtttctacc aa	22
	<210> 262	
	<211> 22	
	<212> DNA	
55	<213> Homo sapiens	
	<400> 262	
	tggctcaca ggaccactga tt	22

<210> 263
 <211> 24
 <212> DNA
 <213> Homo sapiens
 5
 <400> 263
 aaataatcag tgtgattcgt ggag 24
 10
 <210> 264
 <211> 22
 <212> DNA
 <213> Homo sapiens
 15
 <400> 264
 gagggcagtg ctgtctctaa gg 22
 20
 <210> 265
 <211> 21
 <212> DNA
 <213> Homo sapiens
 <400> 265
 acttcacagc cctgcgtaaa c 21
 25
 <210> 266
 <211> 21
 <212> DNA
 <213> Homo sapiens
 30
 <400> 266
 atgggacagg cactgatttg t 21
 35
 <210> 267
 <211> 22
 <212> DNA
 <213> Homo sapiens
 <400> 267
 gcagcgggtt acatcttctt tc 22
 40
 <210> 268
 <211> 22
 <212> DNA
 <213> Homo sapiens
 45
 <400> 268
 cagctctggc tcacactacc ag 22
 50
 <210> 269
 <211> 22
 <212> DNA
 <213> Homo sapiens
 55
 <400> 269
 cctgaactcc gtcagactga aa 22
 <210> 270
 <211> 20

<212> DNA
 <213> Homo sapiens

5 <400> 270
 gcagctggac tcgatttcct 20

10 <210> 271
 <211> 23
 <212> DNA
 <213> Homo sapiens

15 <400> 271
 ccttacagca atcctgtgaa aca 23

20 <210> 272
 <211> 22
 <212> DNA
 <213> Homo sapiens

25 <400> 272
 tgcccaatga gtcaagaagt gt 22

30 <210> 273
 <211> 22
 <212> DNA
 <213> Homo sapiens

35 <400> 273
 atgtacagtg ctggcatggt ct 22

40 <210> 274
 <211> 22
 <212> DNA
 <213> Homo sapiens

45 <400> 274
 cactcacgga tgctgcttag tt 22

50 <210> 275
 <211> 21
 <212> DNA
 <213> Homo sapiens

55 <400> 275
 taaggcaccc acatcatgtc a 21

60 <210> 276
 <211> 24
 <212> DNA
 <213> Homo sapiens

65 <400> 276
 tggacctaaa aggcttaca tcaa 24

70 <210> 277
 <211> 24
 <212> DNA
 <213> Homo sapiens

	<400> 277	
	gccttttagg tccactatgg aatg	24
5	<210> 278	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 278	
	ccaggcgatg ctactactgg tc	22
	<210> 279	
	<211> 22	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 279	
	tcatagcaca cctccctcac tg	22
20	<210> 280	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 280	
	acacaacaaa gagcttgtgc ag	22
	<210> 281	
30	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 281	
35	ccattacttt gagaaggaca ggaa	24
	<210> 282	
	<211> 22	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 282	
	tattcttgct ggatgcgttt ct	22
45	<210> 283	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 283	
	aggagggcag aggactagct g	21
	<210> 284	
	<211> 20	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 284	
	ggcaatgtga atgtgcactg	20

<210> 285
 <211> 20
 <212> DNA
 <213> Homo sapiens
 5
 <400> 285
 cttgaacctg ggaggtggag 20
 10
 <210> 286
 <211> 22
 <212> DNA
 <213> Homo sapiens
 15
 <400> 286
 atcagggtgg gaggagtaaa ga 22
 20
 <210> 287
 <211> 22
 <212> DNA
 <213> Homo sapiens
 25
 <400> 287
 cccacttacc ttcacctgt gc 22
 30
 <210> 288
 <211> 22
 <212> DNA
 <213> Homo sapiens
 35
 <400> 288
 gtgaactcc ggtaggaaat gg 22
 40
 <210> 289
 <211> 20
 <212> DNA
 <213> Homo sapiens
 45
 <400> 289
 aggggacctc aagggagaag 20
 50
 <210> 290
 <211> 22
 <212> DNA
 <213> Homo sapiens
 55
 <400> 290
 agatcatgcc agtgaactcc ag 22
 60
 <210> 291
 <211> 22
 <212> DNA
 <213> Homo sapiens
 65
 <400> 291
 ggaccaggaa agtccttgct tt 22
 70
 <210> 292
 <211> 22

<212> DNA
 <213> Homo sapiens

5 <400> 292
 ggtggggaac attaaactga gg 22

10 <210> 293
 <211> 22
 <212> DNA
 <213> Homo sapiens

15 <400> 293
 gcttcaggtt gttttgtgc ag 22

20 <210> 294
 <211> 22
 <212> DNA
 <213> Homo sapiens

25 <400> 294
 acccttgctt gagggaaata tg 22

30 <210> 295
 <211> 22
 <212> DNA
 <213> Homo sapiens

35 <400> 295
 cccagctcct aggggtacagt ct 22

40 <210> 296
 <211> 23
 <212> DNA
 <213> Homo sapiens

45 <400> 296
 cagtcagctt caaaatccct ctt 23

50 <210> 297
 <211> 24
 <212> DNA
 <213> Homo sapiens

55 <400> 297
 tcacttcctt gtgagtaaag aaaa 24

60 <210> 298
 <211> 23
 <212> DNA
 <213> Homo sapiens

65 <400> 298
 ggccatttaa ttctgtcct tga 23

70 <210> 299
 <211> 23
 <212> DNA
 <213> Homo sapiens

	<400> 299	
	tggactgtg caaactcaaa ctg	23
5	<210> 300	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 300	
	tccaatata gggcagtcac gtt	23
	<210> 301	
	<211> 23	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 301	
	tctcaatcag ttgagttgcc ttg	23
20	<210> 302	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 302	
	agctgtgcaa gttgggaaac at	22
	<210> 303	
30	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 303	
35	gctgtgaggg taaatgagac ca	22
	<210> 304	
	<211> 22	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 304	
	gtctcctggt gagtgactgt gg	22
	<210> 305	
45	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
	<400> 305	
50	ccttccttcg tctccacagc	20
	<210> 306	
	<211> 21	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 306	
	gtccttgatc caacagtcga g	21

<210> 307
 <211> 22
 <212> DNA
 <213> Homo sapiens
 5
 <400> 307
 gcttggcaag gagaagagaa ca 22
 10
 <210> 308
 <211> 23
 <212> DNA
 <213> Homo sapiens
 15
 <400> 308
 gcttgcttc ttgctgaac aac 23
 20
 <210> 309
 <211> 22
 <212> DNA
 <213> Homo sapiens
 25
 <400> 309
 gctggtcacc ttgagcttct ct 22
 30
 <210> 310
 <211> 21
 <212> DNA
 <213> Homo sapiens
 35
 <400> 310
 ccatgctggg ctcttgatt a 21
 40
 <210> 311
 <211> 22
 <212> DNA
 <213> Homo sapiens
 45
 <400> 311
 caccactctg aagttgcct ct 22
 50
 <210> 312
 <211> 21
 <212> DNA
 <213> Homo sapiens
 55
 <400> 312
 atggctctgc acattgttc c 21
 60
 <210> 313
 <211> 22
 <212> DNA
 <213> Homo sapiens
 65
 <400> 313
 cagagtggga aaaggcactt ca 22
 70
 <210> 314
 <211> 22

<212> DNA
 <213> Homo sapiens

5 <400> 314
 ccagagtcct gtgcagacat tc 22

10 <210> 315
 <211> 22
 <212> DNA
 <213> Homo sapiens

15 <400> 315
 atggggatta actgggatgt tg 22

20 <210> 316
 <211> 23
 <212> DNA
 <213> Homo sapiens

25 <400> 316
 cgtagctcca gacatcacta gca 23

30 <210> 317
 <211> 22
 <212> DNA
 <213> Homo sapiens

35 <400> 317
 gcaacctggt ctgcaaagtc tc 22

40 <210> 318
 <211> 20
 <212> DNA
 <213> Homo sapiens

45 <400> 318
 acccagcagt ccagcatgag 20

50 <210> 319
 <211> 22
 <212> DNA
 <213> Homo sapiens

55 <400> 319
 gagtttcagt gagccaagat cg 22

60 <210> 320
 <211> 22
 <212> DNA
 <213> Homo sapiens

65 <400> 320
 ttacaggctt gagccactag gc 22

70 <210> 321
 <211> 21
 <212> DNA
 <213> Homo sapiens

	<400> 321 aagcttcag gagacgaggt c	21
5	<210> 322 <211> 21 <212> DNA <213> Homo sapiens	
10	<400> 322 gtccctgaaa tccctcaaac c	21
15	<210> 323 <211> 23 <212> DNA <213> Homo sapiens	
20	<400> 323 tgctccataa acgtgactat tgc	23
25	<210> 324 <211> 22 <212> DNA <213> Homo sapiens	
30	<400> 324 gtaagagggt gggctggaat ct	22
35	<210> 325 <211> 22 <212> DNA <213> Homo sapiens	
40	<400> 325 cttagaccat gtccgggaaa ac	22
45	<210> 326 <211> 22 <212> DNA <213> Homo sapiens	
50	<400> 326 cacatcactc tggtaggtga ac	22
55	<210> 327 <211> 22 <212> DNA <213> Homo sapiens	
	<400> 327 aaatttcac ccaaaaccaa cc	22
	<210> 328 <211> 23 <212> DNA <213> Homo sapiens	
	<400> 328 ccagtcccaa gttcttgatc att	23

<210> 329
 <211> 21
 <212> DNA
 <213> Homo sapiens
 5
 <400> 329
 acaagtcggc tagttgcatg g 21
 10
 <210> 330
 <211> 22
 <212> DNA
 <213> Homo sapiens
 15
 <400> 330
 tctcagatga aaccaccagc ac 22
 20
 <210> 331
 <211> 23
 <212> DNA
 <213> Homo sapiens
 <400> 331
 ttcatctgag aagcaaggag tgg 23
 25
 <210> 332
 <211> 22
 <212> DNA
 <213> Homo sapiens
 30
 <400> 332
 ccagggagaa agcaggactc ta 22
 35
 <210> 333
 <211> 22
 <212> DNA
 <213> Homo sapiens
 <400> 333
 ttctggcggg gtttgaaat ta 22
 40
 <210> 334
 <211> 22
 <212> DNA
 <213> Homo sapiens
 45
 <400> 334
 ctcaacattg acggcctttc tt 22
 50
 <210> 335
 <211> 23
 <212> DNA
 <213> Homo sapiens
 55
 <400> 335
 tcagctctta aacaggcat agc 23
 <210> 336
 <211> 22

<212> DNA
 <213> Homo sapiens

5 <400> 336
 gaaatgcagc agccactaaa ga 22

10 <210> 337
 <211> 22
 <212> DNA
 <213> Homo sapiens

15 <400> 337
 ctcaccttca aggacctggt gt 22

20 <210> 338
 <211> 21
 <212> DNA
 <213> Homo sapiens

25 <400> 338
 caggagaggg tagaaaccac a 21

30 <210> 339
 <211> 21
 <212> DNA
 <213> Homo sapiens

35 <400> 339
 ggagaggtgg agaggcttca g 21

40 <210> 340
 <211> 22
 <212> DNA
 <213> Homo sapiens

45 <400> 340
 gagactccca ggacagacac ct 22

50 <210> 341
 <211> 20
 <212> DNA
 <213> Homo sapiens

55 <400> 341
 cactcgttcc tcacccttcc 20

60 <210> 342
 <211> 22
 <212> DNA
 <213> Homo sapiens

65 <400> 342
 aggactcaca cgctactctg gt 22

70 <210> 343
 <211> 23
 <212> DNA
 <213> Homo sapiens

	<400> 343	
	ggacaatgtg atgaagattg ctg	23
5	<210> 344	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 344	
	atagcaggat cccaaaagac ca	22
	<210> 345	
	<211> 22	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 345	
	ggcttgggga cctgtatttg ta	22
20	<210> 346	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 346	
	cagtggcctt tctgagcctt ac	22
	<210> 347	
30	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
	<400> 347	
35	gcactctagc tccctctttt agc	23
	<210> 348	
	<211> 24	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 348	
	ttttacagta gagggcagac atgc	24
45	<210> 349	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 349	
	gccaccatag ctgcagaatt ag	22
	<210> 350	
	<211> 22	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 350	
	cccaaggaca gatgtgatgc ta	22

	<210> 351	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
5		
	<400> 351	
	gcctttgttc gagaggagt gt	22
10		
	<210> 352	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
15		
	<400> 352	
	gttcacgctc tcaagcaggt ta	22
20		
	<210> 353	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 353	
	attccacaag ctctctcat ga	22
25		
	<210> 354	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
30		
	<400> 354	
	cttgccccaa gatgcctaag	20
35		
	<210> 355	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
40		
	<400> 355	
	tgcttggtat ttgctcatca tgt	23
45		
	<210> 356	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 356	
	cccttagcta gccactgac aa	22
50		
	<210> 357	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
55		
	<400> 357	
	ctcctgggag tgggtgcca	20
	<210> 358	
	<211> 22	

<212> DNA
 <213> Homo sapiens

5 <400> 358
 cctgggcaac agacagagta ag 22

10 <210> 359
 <211> 21
 <212> DNA
 <213> Homo sapiens

15 <400> 359
 cttcacttcc ccatgctac c 21

20 <210> 360
 <211> 21
 <212> DNA
 <213> Homo sapiens

25 <400> 360
 gggttcacaa tgcctacagg a 21

30 <210> 361
 <211> 22
 <212> DNA
 <213> Homo sapiens

35 <400> 361
 aaaatctgtg actttggctt gg 22

40 <210> 362
 <211> 23
 <212> DNA
 <213> Homo sapiens

45 <400> 362
 gggaggagac attctttgat ttg 23

50 <210> 363
 <211> 22
 <212> DNA
 <213> Homo sapiens

55 <400> 363
 gcagtcctga gaagaaaaca gc 22

60 <210> 364
 <211> 22
 <212> DNA
 <213> Homo sapiens

65 <400> 364
 cttcacatgc cccaaaatta ca 22

70 <210> 365
 <211> 23
 <212> DNA
 <213> Homo sapiens

	<400> 365	
	tgagccatgt attcagagg tga	23
5	<210> 366	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 366	
	tacatttcag caggtgcgtg tt	22
	<210> 367	
	<211> 24	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 367	
	ttgcctactc tgtagggata ttgc	24
20	<210> 368	
	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 368	
	atagggcatg tagcccagtg a	21
	<210> 369	
30	<211> 21	
	<212> DNA	
	<213> Homo sapiens	
	<400> 369	
35	gctctgctgt tggctctcac t	21
	<210> 370	
	<211> 22	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 370	
	ttgcaaagca cacatcttct ga	22
45	<210> 371	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 371	
	tggcaatgtc aatgtcaagc at	22
	<210> 372	
	<211> 22	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 372	
	gtatgttgcc ccactcaaca aa	22

<210> 373
 <211> 22
 <212> DNA
 <213> Homo sapiens
 5
 <400> 373
 tgcatttcct agctgagact cc 22
 10
 <210> 374
 <211> 22
 <212> DNA
 <213> Homo sapiens
 15
 <400> 374
 tgccatctcg cacgtagtaa at 22
 20
 <210> 375
 <211> 22
 <212> DNA
 <213> Homo sapiens
 25
 <400> 375
 ctctcctgtg ctgagcagct tt 22
 30
 <210> 376
 <211> 22
 <212> DNA
 <213> Homo sapiens
 35
 <400> 376
 tgtttccaat cactggcttt ca 22
 40
 <210> 377
 <211> 20
 <212> DNA
 <213> Homo sapiens
 45
 <400> 377
 gaaccatggg ctgtctctgg 20
 50
 <210> 378
 <211> 22
 <212> DNA
 <213> Homo sapiens
 55
 <400> 378
 atctgggata gcgaaggaga ca 22
 60
 <210> 379
 <211> 20
 <212> DNA
 <213> Homo sapiens
 65
 <400> 379
 attacaggcc acacgccatc 20
 70
 <210> 380
 <211> 23

<212> DNA
 <213> Homo sapiens

5 <400> 380
 aaggcaagaa taagggagga aga 23

10 <210> 381
 <211> 23
 <212> DNA
 <213> Homo sapiens

15 <400> 381
 gctctcagga ctgcagaagt aca 23

20 <210> 382
 <211> 22
 <212> DNA
 <213> Homo sapiens

25 <400> 382
 gaggaaccaa tccactcac ac 22

30 <210> 383
 <211> 23
 <212> DNA
 <213> Homo sapiens

35 <400> 383
 tcactctttg ccttctgtct ctg 23

40 <210> 384
 <211> 22
 <212> DNA
 <213> Homo sapiens

45 <400> 384
 gcactgtgct ttgctttctc ag 22

50 <210> 385
 <211> 24
 <212> DNA
 <213> Homo sapiens

55 <400> 385
 tgtctccttt atcgtaggtc tcca 24

60 <210> 386
 <211> 22
 <212> DNA
 <213> Homo sapiens

65 <400> 386
 caccacattt cctacagttc ca 22

70 <210> 387
 <211> 22
 <212> DNA
 <213> Homo sapiens

	<400> 387	
	cactgtgcac cagacagaca aa	22
5	<210> 388	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 388	
	tgtggttttc tgtatcagca gctt	24
	<210> 389	
	<211> 22	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 389	
	cagggagtct gaaatcatca gg	22
20	<210> 390	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 390	
	tcaagtatct agccccaaat cca	23
	<210> 391	
30	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 391	
35	ggcaatattg accattcatc attc	24
	<210> 392	
	<211> 21	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 392	
	aggccaggag taagacgcaa c	21
45	<210> 393	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 393	
	aagaacgtac gtgtggtgtt gg	22
	<210> 394	
	<211> 22	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 394	
	cgctatactt gctccatgca ct	22

<210> 395
 <211> 21
 <212> DNA
 <213> Homo sapiens
 5
 <400> 395
 aggaaacagc ctctgtcct c 21
 10
 <210> 396
 <211> 22
 <212> DNA
 <213> Homo sapiens
 15
 <400> 396
 gtcaatgctc agacagggag at 22
 20
 <210> 397
 <211> 22
 <212> DNA
 <213> Homo sapiens
 <400> 397
 cccaggaagg caggtactgt ta 22
 25
 <210> 398
 <211> 22
 <212> DNA
 <213> Homo sapiens
 30
 <400> 398
 tttacaacc accaaggggtg tg 22
 35
 <210> 399
 <211> 23
 <212> DNA
 <213> Homo sapiens
 <400> 399
 tcgtgtggtt acctccagat ttt 23
 40
 <210> 400
 <211> 22
 <212> DNA
 <213> Homo sapiens
 45
 <400> 400
 aaattagcca gggagtggag gt 22
 50
 <210> 401
 <211> 20
 <212> DNA
 <213> Homo sapiens
 55
 <400> 401
 catgccatgc tatggctcac 20
 <210> 402
 <211> 22

<212> DNA
 <213> Homo sapiens

5 <400> 402
 aggctgagcg gagttctaattg 22

10 <210> 403
 <211> 22
 <212> DNA
 <213> Homo sapiens

15 <400> 403
 atctcagcaa tccacaggaggt 22

20 <210> 404
 <211> 22
 <212> DNA
 <213> Homo sapiens

25 <400> 404
 atttgctca cgaacacatcatt 22

30 <210> 405
 <211> 23
 <212> DNA
 <213> Homo sapiens

35 <400> 405
 tggaaagttg tctatggcacctc 23

40 <210> 406
 <211> 22
 <212> DNA
 <213> Homo sapiens

45 <400> 406
 atgggcagca aggacttactct 22

50 <210> 407
 <211> 22
 <212> DNA
 <213> Homo sapiens

55 <400> 407
 caccccaata ttgtctgccttc 22

60 <210> 408
 <211> 22
 <212> DNA
 <213> Homo sapiens

65 <400> 408
 ggctcgggaa catgtaattagg 22

70 <210> 409
 <211> 24
 <212> DNA
 <213> Homo sapiens

	<400> 409	
	ccatcatgta tggcaaattc tctt	24
5	<210> 410	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 410	
	tggcgtctcc tagtaaagat gct	23
	<210> 411	
	<211> 21	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 411	
	gccagattgc tggtttcatt g	21
20	<210> 412	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 412	
	ggctaaaaca caaagcacca tt	22
	<210> 413	
30	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 413	
35	gggaagtcac ccacaaagac ct	22
	<210> 414	
	<211> 20	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 414	
	ggtctgggtc acagctctc	20
	<210> 415	
45	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 415	
50	ttctctgcc aagatgtggt gt	22
	<210> 416	
	<211> 22	
	<212> DNA	
55	<213> Homo sapiens	
	<400> 416	
	tgcatgatgt gcaatcatgt ta	22

<210> 417
 <211> 22
 <212> DNA
 <213> Homo sapiens
 5
 <400> 417
 tggaccccga aagataaata gg 22
 10
 <210> 418
 <211> 22
 <212> DNA
 <213> Homo sapiens
 15
 <400> 418
 ttctgcactc ctctggaac tg 22
 20
 <210> 419
 <211> 22
 <212> DNA
 <213> Homo sapiens
 25
 <400> 419
 gggtagagc caacactgat ct 22
 30
 <210> 420
 <211> 22
 <212> DNA
 <213> Homo sapiens
 35
 <400> 420
 ctgtgccctc tcattcaca ct 22
 40
 <210> 421
 <211> 22
 <212> DNA
 <213> Homo sapiens
 45
 <400> 421
 agaactagc ctccaagatt gc 22
 50
 <210> 422
 <211> 22
 <212> DNA
 <213> Homo sapiens
 55
 <400> 422
 acacctcca agactcctc ca 22
 60
 <210> 423
 <211> 20
 <212> DNA
 <213> Homo sapiens
 65
 <400> 423
 gactcgaggg tgggagacag 20
 70
 <210> 424
 <211> 22

<212> DNA
<213> Homo sapiens

5 <400> 424
gctgtcacta ggtgtcctga gc 22

10 <210> 425
 <211> 3878
 <212> DNA
 <213> Homo sapiens

 <400> 425

15

20

25

30

35

40

45

50

55

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgaggcggc	cggaggtccc	agctagcccc	ggcgggccgc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctcc	cccaggtccc	cgctcgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
5	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggtatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
10	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgatagtg	1200
	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgctgtctc	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaagggaat	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cttgccctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
30	tcatgggaga	aaacaacacc	ctggcttggg	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcccata	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctctct	tatgctcttg	2220
	tgggtggccct	ggggatcgcc	ctcttcctgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacacccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtcta	2400
35	gctccgggtg	gttcgggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	acaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tgacaaccc	ccacgtgtgc	agcctgtctg	2580
	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtc	aactgggtgtg	2700
	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
40	cagccaggaa	cgtactgggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggtctg	2820
	ccaacttctg	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgctctatca	2880
	agtggatggc	atttgaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacggggt	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccacccatat	3060
	gtaccatcga	tgtctacatg	atcatgggtca	agtgtctggat	gatagacgca	gatagtgcgc	3120
45	caaagtctcc	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgttacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtgga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agcccccca	cgtcacggac	tccccctctg	agctctctga	3360
	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
50							
	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgcttgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggctctgtg	cagaatcctg	tctatcacia	tcagcctctg	aaccccgcg	3600
	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccagtatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	caactgggcc	3720
55	agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gtccacagc	tgaatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaatttatgt	gagcatga			3878

<210> 426
 <211> 3878
 <212> DNA
 <213> Homo sapiens

5

<400> 426

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccg	gccgcccaga	ccggacgaca	120
10	ggccacctcg	tcggcgctcc	cccaggtccc	cgctcgccg	ccaacgccac	aaccaccg	180
	cacggccccc	tgactccgct	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
15	taaagacct	ccaggaggtg	gctgggttat	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagctct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
20	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgagg	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggt	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtg	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
25	aggaagacgg	cgtccgcaag	tgtaaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggttttt	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
30	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	tgcgctctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaaggaggt	tgttggagaa	ctctgagtgc	atagatgcc	1920
35	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggacctaa	gatccccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgctgctgg	2220
	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgacgc	2280
40	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacacccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtgta	2400
	gctccgggtg	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cgcctgctgg	2580
45	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtc	aactgggtgtg	2700
	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgc	cgcgacctgg	2760
	cagccaggaa	cgtactggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggctgg	2820
	ccaaactgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	atttgaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacgggg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
50	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccacccatat	3060
	gtaccatcga	tgtctacatg	atcatggtca	agtgtgtgat	gatagacgca	gatagtcgcc	3120

55

```

      caaagttccg tgagttgatc atcgaattct ccaaaatggc ccgagacccc cagcgctacc 3180
      ttgtcattca gggggatgaa agaatgcatt tgccaagtcc tacagactcc aacttctacc 3240
      gtgccctgat ggatgaagaa gacatggacg acgtgggtgga tgccgacgag tacctcatcc 3300
      cacagcaggg cttcttcagc agccccctcca cgtcacggac tcccctcctg agctctctga 3360
5      gtgcaaccag caacaattcc accgtggcct gcattgatag aaatgggctg caaagctgtc 3420
      ccatcaagga agacagcttc ttgcagcgat acagctcaga cccacaggc gccttgactg 3480
      aggacagcat agacgacacc ttcttcccag tgcctgaata cataaaccag tccgttccca 3540
      aaaggccgcg tggctctgtg cagaatcctg tctatcacia tcagcctctg aaccccgcg 3600
      ccagcagaga cccacactac caggaccccc acagcactgc agtgggcaac cccgagtatc 3660
      tcaacactgt ccagcccacc tgtgtcaaca gcacattcga cagccctgcc cactgggccc 3720
10      agaaaggcag ccaccaaatt agcctggaca accctgacta ccagcaggac ttctttccca 3780
      aggaagccaa gccaaatggc atctttaagg gctccacagc tgaaaatgca gaatacctaa 3840
      gggtcgcgcc acaaagcagt gaattttatt gagcatga 3878

```

```

15      <210> 427
      <211> 3863
      <212> DNA
      <213> Homo sapiens

```

```

20      <400> 427

```

```

25

```

```

30

```

```

35

```

```

40

```

```

45

```

```

50

```

```

55

```

	cccggcgag	cgcgccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcg	60
	ccgagggcg	cggagtcgg	agctagcccc	ggcgccgccc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctccg	cccaggtccc	cgctctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggagggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccitttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
10	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccagtagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtc	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccgggag	agcgactgcc	tgttctggccg	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatat	agctttggtg	ccacctgctg	gaagaagtgt	ccccgtaatt	1140
	atgtggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggat	tggtagaatt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgggca	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtagtg	1620
	gagatgtgat	aattttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccggg	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgctgtctt	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	acccagagtg	ctgcctcag	gcatggaaca	tcactgcac	aggacgggga	ccagacaact	1980
30	gtatccagtg	tgcccaactac	attgacggcc	cccactgctg	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggatggtggg	ggccctcctc	ttgtgctggt	2220
	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgcttcg	aagcgcacgc	2280
	tgcgagggct	gctgcaggag	agggagcttg	tggagcctct	tacaccagtg	ggagaagctc	2340
35	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtgtg	2400
	gctccgggtg	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaaa	acatctccga	aagccaacaa	ggaaatcctc	gatgaagcct	2520
	acgtgatggc	cagcgtggac	aacccccacg	tgtgccgcct	gctgggcac	tgcttcacct	2580
	ccacgtgca	gctcatcacg	cagctcatgc	cctctggctg	cctcctggac	tacgtccggg	2640
	aacacaagaa	caatattggc	tcccagttacc	tgctcaactg	gtgtgtgcag	atcgcaagg	2700
	gcatgaacta	cttggaggac	cgctcgcttg	tgaccgcga	cctggcagcc	aggaacgtac	2760
40	tgggtgaaac	accgcagcat	gtcaagatca	cagattttgg	gctggccaaa	ctgctgggtg	2820
	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcattgg	2880
45	aatcaatttt	acacagaatc	tatacccacc	agagtgatgt	ctggagctac	ggggtgactg	2940
	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	ggtcaagtgc	tggatgatag	acgcagatag	tcgccc aaag	ttccgtgagt	3120
	tgatcatcga	atttctcaaa	atggccccgag	acccccagcg	ctacctgttc	attcaggggg	3180
	atgaaagaat	gcattttgca	agtcctacag	actccaactt	ctaccgtgcc	ctgatggatg	3240
	aagaagacat	ggacgacgtg	gtggatggcg	acgagtacct	catcccacag	cagggcttct	3300
50	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgc	accagcaaca	3360
	attccaccgt	ggcttgcatt	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
	gcttcttgca	gcgatacagc	tcagacccca	caggcgctt	gactgaggac	agcatagacg	3480
	acaccttctt	cccagtgcct	gaatacataa	accagttccgt	tcccaaaagg	cccgttggct	3540
	ctgtgcagaa	tccgttctat	cacaatcagc	ctctgaaccc	cgcgcccgac	agagacccac	3600
	actaccagga	ccccacagc	actgcagtgg	gcaacccgga	gtatctcaac	actgtccagc	3660
55	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaaagccaa	3780
	atggcatctt	taagggctcc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
	gcagtgaatt	tattggagca	tga				3863

<210> 428
 <211> 3863
 <212> DNA
 <213> Homo sapiens

5

<400> 428

10	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgagggcggc	cggagtcccg	agctagcccc	ggcggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgctctgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcagagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgtgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcc	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
15	aggtggtcct	tggaattttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggagggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccattgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
20	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	ccccgggag	agcgactgcc	tggtctgccg	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
25	atgtggtgac	agatcacggc	tcgtgctgcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaa	gtgtgtaacg	1260
	gaataggat	tggtgaattt	aaagactcac	tcctcataaa	tgctacgaat	attaaacat	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tcacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
30	ccttttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgttttg	gacctccggg	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgctgtctt	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
35	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	acccagagtg	cctgcctcag	gcatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggctctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggacagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgctgctgg	2220
40	tggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggaggct	gctgcaggag	agggagcttg	tgagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgctgg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaaa	acatctccga	aagccaacaa	ggaaatcctc	gatgaagcct	2520
	acgtgatggc	cagcgtggac	aacccccacg	tgtgccgcct	gctgggcctc	tgctcacct	2580
45	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcggctg	cctcctggac	tatgtccggg	2640

50

55

```

aacacaaaga caatattggc tcccagtacc tgctcaactg gtgtgtgcag atcgcaaagg 2700
gcatgaacta cttggaggac cgtcgccttg tgcaccgcga cctggcagcc aggaacgtac 2760
tggtgaaaac accgcagcat gtcaagatca cagattttgg gctggccaaa ctgctgggtg 2820
cggaagagaa agaataccat gcagaaggag gcaaagtgcc tatcaagtgg atggcattgg 2880
aatcaatttt acacagaatc tatacccacc agagtgatgt ctggagctac ggggtgactg 2940
5   tttgggagtt gatgaccttt ggatccaagc catatgacgg aatccctgcc agcgagatct 3000
cctccatcct ggagaaagga gaacgcctcc ctcagccacc catatgtacc atcgatgtct 3060
acatgatcat ggtcaagtgc tggatgatag acgcagatag tcgcccaaag ttccgtgagt 3120
tgatcatcga attctccaaa atggcccag agccccagcg ctaccttgct attcaggggg 3180
atgaaagaat gcatttgcca agtcctacag actccaactt ctaccgtgcc ctgatggatg 3240
10  aagaagacat ggacgacgtg gtggatgccg acgagtacct catcccacag cagggcttct 3300
tcagcagccc ctccacgtca cggactcccc tcctgagctc tctgagtgca accagcaaca 3360
attccaccgt ggcttgcatg gatagaaatg ggctgcaaag ctgtcccatc aaggaagaca 3420
gcttcttgca gcgatacagc tcagacccca caggcgcctt gactgaggac agcatagacg 3480
acaccttcct cccagtgcct gaatacataa accagtccgt tcccaaaagg cccgctggct 3540
ctgtgcagaa tcctgtctat cacaatcagc ctctgaacct cgcgcccagc agagaccac 3600
15  actaccagga cccccacagc actgcagtgg gcaaccccga gtatctcaac actgtccagc 3660
ccacctgtgt caacagcaca ttcgacagcc ctgcccactg ggcccagaaa ggcagccacc 3720
aaattagcct ggacaaccct gactaccagc aggacttctt tcccaaggaa gccaaagcaa 3780
atggcatctt taagggctcc acagctgaaa atgcagaata cctaagggtc gcgccacaaa 3840
gcagtgaatt tattggagca tga                                     3863

```

```

20  <210> 429
    <211> 3863
    <212> DNA
    <213> Homo sapiens

```

```

25  <400> 429

```

```

30

```

```

35

```

```

40

```

```

45

```

```

50

```

```

55

```

[illegible]

<210> 430
 <211> 3863
 <212> DNA
 <213> Homo sapiens

5

<400> 430

10
 15
 20
 25
 30
 35
 40
 45
 50
 55

```

ccccggcgag cgcgggccgca gcagcctccg cccccgcac ggtgtgagcg cccgacgcgg 60
ccgagggcggc cggagtcccc agctagcccc ggcgggccgcc gccgcccaga cccgacgaca 120
ggccacctcg tcggcgctccg cccgagtccc cgctctcgccg ccaacgccac aaccaccgcg 180
cacggccccc tgactccgtc cagtattgat cgggagagcc ggagcgagct cttcggggag 240
cagcgatgcg accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgctct 300
gcccggcgag tcgggctctg gaggaaaaaga aagtttgcca aggcacgagt aacaagctca 360
cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
aggtggtcct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
taaagaccat ccaggagggtg gctggttatg tcctcattgc cctcaacaca gtggagcgaa 540
ttcctttgga aaacctgcag atcatcagag gaaatatgta ctacgaaaat tcctatgcct 600
tagcagtcct atctaactat gatgcaaaata aaaccggact gaaggagctg cccatgagaa 660
atttacagga aatcctgcat ggcgccgtgc ggttcagcaa caacctgcc ctgtgcaacg 720
tgagagcat ccagtggcgg gacatagtca gcagtgactt tctcagcaac atgtcgatgg 780
acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
gctgggggtgc aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagt 900
gctccgggcg ctgccgtggc aagtccccc gtgactgctg ccacaaccag tgtgctgcag 960
gctgcacagg ccccggggag agcgactgcc tggctgccc caaattccga gacgaagcca 1020
cgtgcaagga cacctgcccc ccactcatgc tctacaacct caccacgtac cagatggatg 1080
tgaaccccga gggcaaatac agctttgggtg ccacctgctg gaagaagtgt ccccgtaatt 1140
atgtggtgac agatcacggc tcgtgctcc gagcctgtgg ggccgacagc tatgagatgg 1200
aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgcaa gtgtgtaacg 1260
gaataggtat tggngaattt aaagactcac tctccataaa tgctacgaat attaaacact 1320
tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
aggaaatcac aggggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
cctttgagaa cctagaaatc atacgcggca ggaccaagca acatggtcag ttttctcttg 1560
cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
aactgtttgg gacctccggt cagaaaacca aaattataag caacagaggt gaaaacagct 1740
gcaaggccac aggccagggtc tgccatgcct tgtgctcccc cgagggctgc tggggcccgg 1800
agcccaggga ctgctctct tgcgggaatg tcagccgagg cagggaatgc gtggacaagt 1860
gcaaccttct ggaggggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
acccagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
gtatccagtg tgcccactac attgacggcc cccactgctg caagacctgc ccggcaggag 2040
tcatgggaga aaacaacacc ctggtctgga agtacgcaga cgccggccat gtgtgccacc 2100
tgtgccatcc aaactgcacc tacggatgca ctgggcccag tcttgaaggc tgtccaacga 2160

```

atgggcctaa gatccccgtcc atcgccactg ggatgggtggg ggccctcctc ttgctgctgg 2220
 tggtagccct ggggatcggc ctcttcacgc gaaggcgcca catcgttcgg aagcgacgc 2280
 tgcggaggct gctgcaggag agggagcttg tggagcctct tacaccagt ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgctgg 2400
 gctccggtgc gttcggcacg gtgtataagg gactctggat ccagaagggt gagaaagtta 2460
 aaattcccgt cgctatcaaa acatctccga aagccaacaa ggaaatcctc gatgaagcct 2520
 acgtgatggc cagcgtggac aacccccacg tgtgccgcct gctgggcac tgcctcacct 2580
 ccaccgtgca gctcatcacg cagctcatgc ccttcggctg cctcctggac tatgtccggg 2640
 aacacaaaga caatattggc tccagttacc tgctcaactg gtgtgtgcag atcgcaaagg 2700
 gcatgaacta cttggaggac cgtcgcttgg tgcaccgca cctggcagcc aggaacgtac 2760
 tggtagaaaac accgcagcat gtcaagatca cagattttgg gctggccaaa ctgctgggtg 2820
 cggaagagaa agaataccat gcagaaggag gcaaagtgcc tatcaagtgg atggcattgg 2880
 aatcaatttt acacagaatc tatacccacc agagtgatgt ctggagctac ggggtgactg 2940
 tttgggagtt gatgaccttt ggatccaagc catatgacgg aatccctgcc agcgagatct 3000
 cctccatcct ggagaaagga gaacgcctcc ctacagccacc catatgtacc atcgatgtct 3060
 acatgatcat ggtcaagtgc tggatgatag acgcagatag tgcgccaaag ttccgtgagt 3120
 tgatcatcga atttctcaaa atggcccagag acccccagcg ctaccttgct attcaggggg 3180
 atgaaagaat gcatttgcca agtccctacag actccaactt ctaccgtgcc ctgatggatg 3240
 aagaagacat ggacgacgtg gtggatgccg acgagtacct catcccacag cagggcttct 3300
 tcagcagccc ctccacgtca cggactcccc tctgagctc tctgagtgc accagcaaca 3360
 attccaccgt ggcttgcat gatagaaatg ggctgcaaag ctgtcccatc aaggaagaca 3420
 gcttcttgca gcgatacagc tcagacccca caggcgctt gactgaggac agcatagacg 3480
 acaccttcct ccagtgctc gaatacataa accagtccgt tcccaaaagg cccgctggct 3540
 ctgtgcagaa tcctgtctat cacaatcagc ctctgaacct cgcgccagc agagacccac 3600
 actaccagga ccccccacagc actgcagtgg gcaaccccg gtatctcaac actgtccagc 3660
 ccacctgtgt caacagcaca ttcgacagcc ctgcccactg ggcccagaaa ggcagccacc 3720
 aaattagcct ggacaaccct gactaccagc aggacttct tcccaaggaa gccaaagcaa 3780
 atggcatctt taagggctcc acagctgaaa atgcagaata cctaagggtc gcgccacaaa 3840
 gcagtgaatt tattggagca tga 3863

<210> 431

<211> 3863

<212> DNA

<213> Homo sapiens

<400> 431

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	gggtgtgagcg	cccgcgcg	60
	ccgaggcggc	cggaggtccc	agctagcccc	ggcgggccgc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctcc	cccaggtccc	cgccctcgcc	ccaacgccac	aaccaccg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggttat	tcctcattgc	cctcaacaca	gtggagcgaa	540
10	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagctct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagagggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
15	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgggag	agcgactggc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtg	agatcacggc	tcgtgctgcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacctcct	gccgggtggc	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaatc	atacgcggca	ggaccaagca	acatgggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
25	gagatgtgat	aatttcagga	aacaaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgctgtctt	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
30	accagaggtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccaactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtggg	ggccctcctc	ttgctgctgg	2220
35	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcggttcg	aagcgcacgc	2280
	tgcgaggagct	gctgcaggag	agggagcttg	tggagcctct	tacaccagtg	ggagaagctc	2340
	ccaacaagagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaaa	acatctccga	aagccaacaa	ggaaatccct	gatgaagcct	2520
	acgtgatggc	cagcgtggag	aacccccacg	tgtgccgcct	gctgggcac	tgctccacct	2580
40	ccacctgtga	gctcatcacg	cagctcatgc	cctctggctg	cctcctggag	tatcgccggg	2640
	aacacaaaga	caatatggcg	tcccagtagc	tgctcaactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttggaggac	cgctcgcttg	tgaccgcga	cctggcagcc	aggaacgtac	2760
	tggtgaaaac	accgcagcat	gtcaagatca	cagattttgg	gctggccaaa	ctgctgggtg	2820
	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcatttg	2880
	aatcaatttt	acacagaatc	tatacccacc	agatgtatgt	ctggagctac	gggggtgactg	2940
45	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	ggtcaagtgc	tggatgatag	acgcagatag	tcgcccagg	ttccgtgagt	3120
	tgatcatcga	attctccaaa	atggcccag	accccagcg	ctacctgtgc	attcaggggg	3180
	atgaaagaat	gcatttgcca	agtcttacag	actccaactt	ctaccgtgac	ctgatggatg	3240
	aagaagacat	ggacgacgtg	gtggatgccg	acgagtacct	catcccacag	cagggtctct	3300
50	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgc	accagcaaca	3360
	attccaccgt	ggcttgcat	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
	gcttcttgca	gcgatacagc	tcagaccca	caggcgctt	gactgaggac	agcatagacg	3480
	acaccttcct	cccagtgctt	gaatacataa	accagtccgt	tcccaaaagg	cccgtggct	3540
	ctgtgcagaa	tcctgtctat	cacaatcagc	ctctgaacct	cgcgcccagc	agagaccac	3600
	actacagga	ccccacagc	actgcagtgg	gcaaccccga	gtatctcagc	actgtccagc	3660
55	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaagccaa	3780
	atggcatctt	taagggtctc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
	gcagtgaatt	tattggagca	tga				3863

<210> 432
 <211> 3863
 <212> DNA
 <213> Homo sapiens

5

<400> 432

```

10      cccggcgcgag cgcgggccgca gcagcctccg ccccccgcac ggtgtgagcg cccgacgcgg 60
      ccgaggcgggc cggaggtccc agctagcccc ggcgggccgcc gccgcccaga ccggacgaca 120
      ggccacctcg tcggcggtccg cccgaggtccc cgctcgcccg ccaacgccac aaccaccgcg 180
      cacggccccc tgactccgtc cagtattgat cgggagagcc ggagcgagct cttcggggag 240
      cagcgatgcy accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgctct 300
      gcccggcgag tcgggtctcg gaggaagaaga aagtttgcca aggcacgagt aacaagctca 360
      cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
      aggtggctct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
15      taaagacccat ccaggaggtg gctggttatg tcctcattgc cctcaacaca gtggagcgaa 540
      ttccctttgga aaacctgcag atcatcagag gaaatatgta ctacgaaaat tcctatgcct 600
      tagcagtcct atctaactat gatgcaaata aaaccggact gaaggagctg cccatgagaa 660
      atttacagga aatcctgcat ggcgccgtgc gggtcagcaa caaccctgcc ctgtgcaacg 720
      tggagagcat ccagtggcgg gacatagtca gcagtgaact tctcagcaac atgtcgatgg 780
      acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
20      gctgggggtgc aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagt 900
      gctccggggcg ctgccgtggc aagtcccca gtgactgctg ccacaaccag tgtgctgcag 960
      gctgcacagg ccccggggag agcgactgcc tggctgccc caaattccga gacgaagcca 1020
      cgtgcaagga cacctgcccc cactcatgc tctacaaccc caccacgtac cagatggatg 1080
      tgaaccccga gggcaaatac agctttggtg ccacctgctg gaagaagtgt ccccgtaatt 1140
      atgtggtgac agatcacggc tcgtgcgtcc gagcctgtgg ggccgacagc tatgatgg 1200
25      aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgcaaa gtgtgtaacg 1260
      gaataggtat tgggtgaattt aaagactcac tctccataaa tgctacgaat attaaacact 1320
      tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
      actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
      aggaaatcac agggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
      cctttgagaa cctagaaatc atacgcggca ggaccaagca acatggtcag ttttctcttg 1560
30      cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
      gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680

```

35

40

45

50

55

5 aactgtttgg gacctccggt cagaaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaaggccac aggccagggtc tgccatgcct tgtgtcctccc cgaggggtgc tggggcccgg 1800
 agcccaggga ctgctgtctt tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
 gcaacctttt ggaggggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
 acccagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
 gtatccagtg tgcccactac attgacggcc cccactgcgt caagacctgc ccggcaggag 2040
 tcatgggaga aaacaacacc ctggcttggg agtacgcaga cgccggccat gtgtgccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctggggcagg tcttgaaggc tgtccaacga 2160
 atgggcctaa gatcccgtcc atcgccactg ggatgggtgg ggccctcctc ttgctgtctg 2220
 10 tgggtggccct ggggatcggc ctcttcatgc gaaggcgcca catcgttcgg aagcgacgc 2280
 tgcggagggt gctgcaggag agggagcttg tggagcctct tacaccaggt ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtctg 2400
 gctccggtgc gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
 aaattcccgt cgctatcaaa acatctccga aagccaacaa ggaaatcctc gatgaagcct 2520
 acgtgatggc cagcgtggac aacccccacg tgtgccgcct gctgggcata tgcctcacct 2580
 ccaccgtgca gctcatcacg cagctcatgc ccttcggctg cctcctggac tatgtccggg 2640
 15 aacacaaaga caatattggc tcccagtacc tgctcaactg gtgtgtgcag atcgcaaagg 2700
 gcatgaacta cttggaggag cgctcgcttg tgcaccgcga cctggcagcc aggaacgtac 2760
 tggtgaaaac accgcagcat gtcaagatca cagattttgg gctggccaaa ctgctgggtg 2820
 cggaagagaa agaataccat gcagaaggag gcaaagtgcc tatcaagtgg atggcattgg 2880
 aatcaatttt acacagaatc tatacccacc agagtgatgt ctggagctac ggggtgactg 2940
 20 tttggaggtt gatgaccttt ggatccaagc catatgacgg aatccctgcc agcgagatct 3000
 cctccatcct ggagaaagga gaacgcctcc ctacgccacc catatgtacc atcgatgtct 3060
 acatgatcat ggtcaagtgc tggatgatag acgcagatag tcgccc aaag ttccgtgagt 3120
 tgatcatcga attctccaaa atggcccag agccccagcg ctaccttgct attcaggggg 3180
 atgaaagaat gcatttgcca agtccctacg actccaactt ctaccgtgcc ctgatggatg 3240
 aagaagacat ggacgacgtg gtggatgccg acgagtacct catcccacag cagggtctct 3300
 25 tcagcagccc ctccacgtca cggactcccc tctgagctc tctgagtgca accagcaaca 3360
 attccaccgt ggcttgcatg gatagaaatg ggctgcaaag ctgtcccatc aaggaagaca 3420
 gcttcttgca gcgatacagc tcagacccca caggcgctct gactgaggac agcatagacg 3480
 acaccttcct cccagtgcct gaatacataa accagtccgt tcccaaaagg cccgctggct 3540
 ctgtgcagaa tcctgtctat cacaatcagc ctctgaacct cgcgcccagc agagaccac 3600
 actaccagga cccccacagc actgcagtgg gcaacccga gtatctcaac actgtccagc 3660
 30 ccacctgtgt caacagcaca ttcgacagcc ctgcccactg ggcccagaaa ggcagccacc 3720
 aaattagcct ggacaaccct gactaccagc aggacttctt tcccaaggaa gccaaagccaa 3780
 atggcatctt taagggctcc acagctgaaa atgcagaata cctaagggtc gcgccacaaa 3840
 cgagtgaatt tattggagca tga 3863

<210> 433
 <211> 3863
 <212> DNA
 <213> Homo sapiens

<400> 433

	cccggcgag	cgcgccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcg	60
	ccgagggcg	cggagtcgg	agctagcccc	ggcgccgccc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctcg	cccaggtccc	cgcttcgccc	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgtctc	300
	gcccggcgag	tcgggctctg	gaggaaaaa	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tggaattttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggagggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
10	tagcagctct	atctaactat	gatgcaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tgagagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggg	ctgcccgtgg	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccgggag	agcgactgcc	tggtctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatat	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggat	tggtagaatt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
25	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgttttg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgctgtctc	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaaggaggt	ttgtggagaa	ctctgagtg	atacagtgcc	1920
30	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagt	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgagggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgtgtcttg	2220
	tggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	atagcgacgc	2280
35	tgcgagggt	gctgcaggag	agggagcttg	tgaggcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtg	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaaa	acatctccga	aagccaacaa	ggaaatcctc	gatgaagcct	2520
	acgtgatggc	cagcgtggac	aacccccacg	tgtgcccgtc	gctgggcact	tgcttcacct	2580
40	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcggctg	cctcctggac	tatgtccggg	2640
	aacacaaaga	caatattggc	tcccagttac	tgctcaactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttgaggagc	cgctcgcttg	tgcaccgcga	cctggcagcc	aggaacgtac	2760
	tggtgaaaac	accgcagcat	gtcaagatca	cagatttttg	gctggccaaa	ctgctgggtg	2820
	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcatttg	2880
	aatcaatttt	acacagaatc	tataccacc	agagtgatgt	ctggagctac	ggggtgactg	2940
45	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	ggtcaagtgc	tggatgatag	acgcagatag	tcgccc aaag	ttccgtgagt	3120
	tgatcatcga	atttctcaaa	atggccccgag	acccccagcg	ctacctgtgc	attcaggggg	3180
	atgaaagaat	gcattttgcca	agtcctacag	actccaactt	ctaccgtgcc	ctgatggatg	3240
	aagaagacat	ggacgacgtg	gtggatgccg	acgagtacct	catcccacag	cagggcttct	3300
50	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgc	accagcaaca	3360
	attccaccgt	ggcttgcatt	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
	gcttcttgca	gcgatacagc	tcagacccca	caggcgctt	gactgaggac	agcatagacg	3480
	acaccttctc	cccagtgctc	gaatacataa	accagtcctg	tcccaaaagg	cccgtggctg	3540
	ctgtgcagaa	tctgtctat	cacaatcagc	ctctgaacc	cgcgcccgag	agagaccac	3600
	actaccagga	ccccacagc	actgcagtgg	gcaacccgga	gtatctcaac	actgtccagc	3660
	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
55	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaaagcaa	3780
	atggcatctt	taagggctcc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
	gcagtgaatt	tattggagca	tga				3863

<210> 434
 <211> 3863
 <212> DNA
 <213> Homo sapiens

5

<400> 434

10	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgctctgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccg	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gccccggcgag	tcggggtctg	gaggaaaaga	aagtgttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
15	aggtgggtcct	tggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggagggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtctt	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
20	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggg	ctgccgtggc	aagtcccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	ccccgggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
25	atgtggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200

30

35

40

45

50

55

5 aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgcaaa gtgtgtaacg 1260
 gaataggtat tgggtgaattt aaagactcac tctccataaa tgctacgaat attaaacact 1320
 tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 aggaaatcac aggggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
 cctttgagaa cctagaaatc atacgcggca ggaccaagca acatggtcag ttttctcttg 1560
 cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
 gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
 aactgtttgg gacctccggt cagaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaagggcac agggcaggtc tgccatgcct tgtgctcccc cgagggtgc tggggcccgg 1800
 10 agcccaggga ctgctctct tgcgggaatg tcagccgagg cagggaatgc gtggacaagt 1860
 gcaaccttct ggaggggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
 acccagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
 gtatccagtg tgccactac attgacggcc cccactgcgt caagacctgc ccggcaggag 2040
 tcatgggaga aaacaacacc ctggtctgga agtacgcaga cgccggccac gtgagccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctgggccagg tcttgaaggc tgtccaacga 2160
 15 atgggcctaa gatcccgctc atcgccactg ggatgggtgg ggccctctc ttgctgctgg 2220
 tgggtggccct ggggatcggc ctcttcctgc gaaggcgcca catcggtcgg aagcgcacgc 2280
 tgcggaggct gctgcaggag agggagcttg tggagcctct tacaccagtg ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtctg 2400
 gctccgggtg gtccggcacg gtgtataagg gactctggat ccagaagggt gagaaagtta 2460
 aaattcccgt cgctatcaag acatctccga aagccaacaa ggaaatctc gatgaagcct 2520
 20 acgtgatggc cagcgtggac aacccccacg tgtgccgctt gctgggcate tgcctcacct 2580
 ccaccgtgca gctcatcacg cagctcatgc ccttcggctg cctcctggac tatgtccggg 2640
 aacacaaaga caatattggc tccagttacc tgctcaactg gtgtgtgcag atcgcaaagg 2700
 gcatgaacta cttggaggag cgtcgcttgg tgaccgcga cctggcagcc aggaacgtac 2760
 tggtgaaaaac accgcagcat gtcaagatca cagattttgg gctggccaaa ctgctgggtg 2820
 cggaagagaa agaataccat gcagaaggag gcaaagtgcc tatcaagtgg atggcatttg 2880
 25 aatcaatttt acacagaatc tataccacc agagtgatgt ctggagctac ggggtgactg 2940
 tttgggagtt gatgacctt ggatccaagc catatgacgg aatccctgcc agcgagatct 3000
 cctccatcct ggagaaagga gaacgcctcc ctgagccacc catatgtacc atcgatgtct 3060
 acatgatcat ggtcaagtgc tggatgatag acgcagatag tcgccc aaag ttccgtgagt 3120
 tgatcatcga atttccaaa atggcccgag acccccagcg ctacctgtc attcaggggg 3180
 atgaaagaat gcatttgcca agtcctacag actccaactt ctaccgtgcc ctgatggatg 3240
 30 aagaagacat ggacgacgtg gtggatgccc acgagtacct catcccacag cagggtcttct 3300
 tcagcagccc ctccacgtca cggactcccc tctgagctc tctgagtgca accagcaaca 3360
 attccaccgt ggcttgcat gatagaaatg ggctgcaaag ctgtcccatc aaggaagaca 3420
 gcttcttgca gcgatacagc tcagaccca caggcgctt gactgaggac agcatagacg 3480
 acaccttcct cccagtgcct gaatacataa accagtccgt tcccaaaagg cccgtggct 3540
 ctgtgcagaa tcctgtctat cacaatcagc ctctgaacc cgcgcccagc agagaccac 3600
 35 actaccagga ccccccacagc actgcagtgg gcaacccga gtatctcaac actgtccagc 3660
 ccacctgtgt caacagcaca ttcgacagcc ctgcccactg ggcccagaaa ggcagccacc 3720
 aaattagcct ggacaaccct gactaccagc aggacttct tcccaaggaa gccaaagcaa 3780
 atggcatctt taagggctcc acagctgaaa atgcagaata cctaagggtc gcgccacaaa 3840
 gcagtgaatt tattggagca tga 3863

40 <210> 435
 <211> 3863
 <212> DNA
 <213> Homo sapiens

45 <400> 435

50

55

	cccgccgcag	cgccggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggagtgccc	agctagcccc	ggcggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctcc	cccaggtccc	cgctctcgcc	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgct	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
5	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gccccggcag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aaataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
10	tagcagtctt	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
15							
	gctgcacagg	ccccggggag	agcgactgcc	tggtctgccg	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
20	atgtggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	ttgatagatg	1200
	aggaagacgg	cgctccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
25	cctttgagaa	cctagaaatc	atacgcggca	ggaccaagca	acatggctac	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agccccagga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
30	gcaaccttct	ggaggggtgag	ccaaggaggt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccaactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctctct	ttgtcttggt	2220
35	tggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggaggct	gctgcaggag	aggagccttg	tggagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcgggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	acatctccga	aagccaacaa	ggaaatcctc	gatgagcct	2520
	acgtgatggc	cagcgtggac	aacccccacg	tgctggcctc	gctgggcctc	tgccctacct	2580
40	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcggctg	cctcctggac	tatgtccggg	2640
	aacacaaaga	caatattggc	tcccagtagc	tgctcaactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttgaggagc	cgctcgcttg	tgcaccgcga	cctggcagcc	aggaacgtac	2760
	tggtgaaaac	accgcagcat	gtcaagatca	cagatttttg	gctggccaaa	ctgctgggtg	2820
	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcatttg	2880
	aatcaatttt	acacagaatc	tatacccacc	agagtgtatg	ctggagctac	ggggtgactg	2940
45	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	ggtcaagtgc	tggatgatag	acgcagatag	tcgcccacaa	ttccgtgagt	3120
	tgatcatcga	attctccaaa	atggcccagc	acccccagcg	ctacctgtgc	attcaggggg	3180
	atgaaagaat	gcatttgcca	agtcctacag	actccaactt	ctaccgtgcc	ctgatggatg	3240
50	aagaagacat	ggacgacgtg	gtggatgccg	acgagtacct	catcccacag	cagggcttct	3300
	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgc	accagcaaca	3360
	attccaccgt	ggcttgcat	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
	gcttcttgca	gcgatacagc	tcagaccccc	caggcgctct	gactgaggac	agcatagagc	3480
	acaccttctc	cccagtgcct	gaatacataa	accagtcctg	tcccaaaagg	cccgtggct	3540
	ctgtgcagaa	tcctgtctat	cacaatcagc	ctctgaaccc	cgcgcccagc	agagaccac	3600
	actaccagga	ccccacagc	actgcagtgg	gcaaccccg	gtatctcaac	actgtccagc	3660
55	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaaagcaa	3780
	atggcatctt	taagggtctc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
	gcagtgaaatt	tattggagca	tga				3863

<210> 436
 <211> 3863
 <212> DNA
 <213> Homo sapiens

5

<400> 436

10	cccgggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgaacgag	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccgaagtccc	cgccctcgccg	ccaacgccac	aaccaccgag	180
	cacggccccc	tgactccgctc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgaggttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
15	aggtgggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtctt	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720

20

25

30

35

40

45

50

55

5 tggagagcat ccagtggcgg gacatagtc gacgtgactt tctcagcaac atgtcgatgg 780
 acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
 gctgggggtgc aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagct 900
 gctccgggcg ctgccgtggc aagtccccc gtgactgctg ccacaaccag tgtgctgcag 960
 gctgcacagg ccccgggag agcgactgcc tggctgccc caaattccga gacgaagcca 1020
 cgtgcaagga cacctgcccc ccactcatgc tctacaaccc caccacgtac cagatggatg 1080
 tgaacccga gggcaaatac agctttgggtg ccacctgcgt gaagaagtgt ccccgtaatt 1140
 atgtggtgac agatcacggc tctgctgctc gagcctgtgg ggccgacagc tatgagatgg 1200
 aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgcaa gtgtgtaacg 1260
 gaataggtat tgggtgaatt aaagactcac tctcataaaa tgctacgaat attaaacact 1320
 10 tcaaaaactg accctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 aggaaatcac agggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
 cttttgagaa cctagaaatc atacgaggca ggaccaagca acatggtcag ttttctcttg 1560
 cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
 gagatgtgat aatttcagga acaaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
 15 aactgtttgg gacctccggt cagaaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaaggccac aggccaggtc tgccatgcct tgtgtcccc cgagggctgc tggggcccg 1800
 agcccaggga ctgctctct tgcgggaatg tcagccgagg cagggaatgc gtggacaagt 1860
 gcaaccttct ggagggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
 acccagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
 gtatccagtg tgcccactac attgacggcc cccactgcgt caagacctgc cggcaggag 2040
 20 tcatgggaga aaacaacacc ctggtctgga agtacgcaga cgccggccat gtgtgccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctgggcccagg tcttgaaggc tgtccaacga 2160
 atgggcctaa gatcccgtcc atcgccactg ggatgggtgg ggccctctc ttgctgctgg 2220
 tgggtggccct ggggatcggc ctcttcatgc gaaggcgcca catcgttcgg aagcgacgc 2280
 tgcggaggct gctgcaggag agggagcttg tggagcctct tacaccagat ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtg 2400
 25 gctccggtgc gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
 aaattcccgt cgctatcaag acatctccga aagccaacaa ggaaatcctc gatgaagcct 2520
 acgtgatggc cagcgtggac aacccccacg tgtgccgcct gctgggcac tgcctcacct 2580
 ccaccgtgca gctcatcacg cagctcatgc ccttcggctg cctcctggac tatgtccggg 2640
 aacacaaaga caatattggc tcccagtacc tccactgact cctctgagc atcgagatct 2700
 gcatgaacta cttggaggac cgtcgttgg tgcaccgca cctggcagcc aggaacgtac 2760
 30 tggtgaaaac accgcagcat gtcaagatca cagattttgg gctggccaaa ctgctgggtg 2820
 cggaagagaa agaataccat gcagaaggag gcaaagtgcc tatcaagtgg atggcattgg 2880
 aatcaatttt acacagaatc tatacccacc agagtgatgt ctggagctac ggggtgactg 2940
 tttggaggtt gatgacctt ggatccaagc catatgacgg aatccctgcc agcgagatct 3000
 cctccatcct ggagaaagga gaacgcctcc ctcagccacc catatgtacc atcgatgtct 3060
 35 acatgatcat ggtcaagtgc tggatgatag acgcagatag tcgccc aaag ttccgtgagt 3120
 tgatcatcga attctccaaa atggcccag acccccagcg ctacctgtc attcaggggg 3180
 atgaaagaat gcatttgcca agtcttacag actccaact ctaccgtgcc ctgatggatg 3240
 aagaagacat ggacgacgtg gtggatgcc acgagtacct catcccacag cagggtctt 3300
 tcagcagccc ctccacgtca cggactcccc tctgagctc tctgagtga accagcaaca 3360
 attccaccgt ggcttgcat gatagaaatg ggctgcaaag ctgtcccatc aaggaagaca 3420
 gcttcttgca gcgatacagc tcagaccca caggcgctt gactgaggac agcatagacg 3480
 40 acaccttcct cccagtgcct gaatacataa accagtccgt tcccaaaagg cccgctggct 3540
 ctgtgcagaa tcctgtctat cacaatcagc ctctgaacct cgcgcccagc agagaccac 3600
 actaccagga cccccacagc actgcagtgg gcaaccccga gtatctcaac actgtccagc 3660
 ccacctgtgt caacagcaca ttcgacagcg ctgcccactg ggcccagaaa ggcagccacc 3720
 aaattagcct ggacaaccct gactaccagc aggacttct tcccaaggaa gccagccaa 3780
 atggcatctt taagggtcc acagctgaaa atgcagaata cctaagggtc gcgccacaaa 3840
 45 gcagtgaatt tattggagca tga 3863

<210> 437

<211> 3854

<212> DNA

50 <213> Homo sapiens

<400> 437

55

	cccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgaggcggc	cggagtccc	agctagcccc	ggcgggccgc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctccg	cccaggtccc	cgctctcgcc	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gccccggcag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
10	taaagaccat	ccaggagggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
	tagcagctct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
15	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	ccccgggag	agcgactgcc	tggctcggcc	caaattccga	gacgaagcca	1020
	cgtgcaaggga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
20	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaat	tgtagaagt	gcgaagggcc	ttgccgcaa	gtgtgtaact	1260
	gaataggat	tgggtgaatt	aaagactcac	tcctcataaa	tgctacgaat	aactggaata	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatatctctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
25	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgtcctcc	cgagggtgctg	tgggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	caggggaatgc	gtggacaagt	1860
30	gcaaccttct	ggagggtgag	ccaaggaggt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgccctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggctctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggacagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtggg	ggccctcctc	ttgctgctgg	2220
35	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcggttcgg	aagcgcacgc	2280
	tgcgaggagct	gctgcaggag	agggagcttg	tgagcctct	tacaccagct	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtggtg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacact	cgatgaagcc	tacgtgatgg	2520
	ccagcgtgga	caacccccac	gtgtgccgcc	tgctgggcat	ctgcctcacc	tccaccgtgc	2580
40	agctcatcac	gcagctcatg	cccttcgggt	gcctcttgga	ctatgtccgg	gaacacaaag	2640
	acaatatattg	ctcccagtac	ctgctcaact	ggtgtgtgca	gatcgcaaag	ggcatgaact	2700
	acttggagga	ccgtcgcttg	gtgcaccgcg	acctggcagc	caggaacgta	ctgggtgaaaa	2760
	caccgcagca	tgtcaagatc	acagattttg	ggctggccaa	actgctgggt	gcggaagaga	2820
	aagaatacca	tgcagaagga	ggcaaagtgc	ctatcaagtg	gatggcattg	gaatcaattt	2880
	tacacagaat	ctatacccac	cagagtgatg	tctggagcta	cgggtgact	gtttgggagt	2940
45	tgatgacctt	tggatccaag	ccatatgacg	gaatccctgc	cagcgagatc	tcctccatcc	3000
	tggagaaaagg	agaacgcctc	cctcagccac	ccatatgtac	catcgatgtc	tacatgatca	3060
	tggatcaagt	ctggatgata	gacgcagata	gtcgcccaaa	gttccgtgag	ttgatcatcg	3120
	aatttctcaa	aatggcccga	gacccccagc	gctaccttgt	cattcagggg	gatgaaagaa	3180
	tgcattttgcc	aagtccctaca	gactccaact	tctaccgtgc	cctgatggat	gaagaagaca	3240
	tggagcagct	gggtgatgcc	gacgagtacc	tcattccaca	gcagggtctc	ttcagagacc	3300
50	cctccacgtc	acggactccc	ctcctgagct	ctctgagtgc	aaccagcaac	aattccaccg	3360
	tggcttgc	tgatagaaat	gggctgcaaa	gctgtcccat	caaggaagac	agcttctttgc	3420
	agcgatacag	ctcagacccc	acaggcgctt	tgactgagga	cagcatagac	gacaccttcc	3480
	tcccagtgcc	tgaatacata	aaccagtccg	ttcccaaaag	gcccgtggc	tctgtgcaga	3540
	atcctgtcta	tcacaatcag	cctctgaacc	ccgcgcccag	cagagaccca	cactaccagc	3600
	acccccacag	cactgcagtg	ggcaaccctg	agtatctcaa	cactgtccag	cccactgtg	3660
	tcaacagcac	atttcgacag	cctgcccact	gggcccagaa	aggcagccac	caaattagcc	3720
55	tggacaaccc	tgactaccag	caggacttct	ttcccaagga	agccaagcca	aatggcatct	3780
	ttaagggctc	cacagctgaa	aatgcagaat	acctaagggt	cgcgccacaa	agcagtgaat	3840
	ttattggagc	atga					3854

<210> 438
 <211> 3878
 <212> DNA
 <213> Homo sapiens

5

<400> 438

10

cccggcgcag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
ccgaggcggc	cggagtcccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
ggccacctcg	tcggcggtccg	cccgagtccc	cgctctgccg	ccaacgccac	aaccaccgcg	180
cacggcccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcgggggag	240

15

20

25

30

35

40

45

50

55

5 cagcgatgcg accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgctct 300
 gcccggcgag tcgggctctg gaggaagaaga aagtttgcca aggcacgagt aacaagctca 360
 cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
 aggtggctct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
 taaagacat ccaggaggtg gctgggtatg tcctcattgc cctcaacaca gtggagcgaa 540
 ttccitttga aaacctgcag atcatcagag gaaatatgta ctacgaaaat tcctatgcct 600
 tagcagtcct atctaactat gatgcaaata aaaccggact gaaggagctg cccatgagaa 660
 atttacagga aatcctgcat ggcgccgtgc ggttcagcaa caaccctgcc ctgtgcaacg 720
 tggagagcat ccagtggcgg gacatagtcg gcagtgaatt tctcagcaac atgtcgatgg 780
 acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
 10 gctgggggtg aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagt 900
 gctccggggc ctgccgtggc aagtcccccga gtgactgctg ccacaaccag tgtgctgcag 960
 gctgcacagg ccccggggag agcgactgcc tggctgccc gaaattccga gacgaagcca 1020
 cgtgcaagga cacctgcccc cactcatgct tctacaacc caccacgtac cagatggatg 1080
 tgaaccccg gggcaaatat agctttgggt ccacctgcgt gaagaagtgt ccccgtaatt 1140
 atgtggtgac agatcacggc tcgtgcgtcc gagcctgtgg ggcgacagc tatgagatgg 1200
 15 aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgcccgaac gtgtgtaacg 1260
 gaatagggat tgggtgaatt aaagactcac tctccataaa tgctacgaat attaaacact 1320
 tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 aggaaatcac aggggttttt ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
 cctttgagaa cctagaaatc atacgcggca ggaccaagca acatggtcag ttttctcttg 1560
 20 tggtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtatg 1620
 gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
 aactgtttgg gacctccggt cagaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaaggccac aggccagggtc tgccatgcct tgtgctcccc cgagggctgc tggggcccg 1800
 agcccaggga ctgctctctc tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
 25 gcaaccttct ggagggtgag ccaaggaggt ttgtggagaa ctctgagtgc atacagtgc 1920
 acccagagtg cctgcctcag gccatgaaca tcacctgcac aggcgggga ccagacaact 1980
 gtatccagtg tgcccactac attgacggcc cccactgcgt caagacctgc ccggcaggag 2040
 tcatgggaga aaacaacacc ctggtctgga agtacgcaga cgccggccat gtgtgccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctgggcccag tcttgaaggc tgtccaacga 2160
 atgggcctaa gatcccgtcc atcgccactg ggatggtggg ggcctcctc ttgtgctgg 2220
 30 tgggtggccct ggggatcgcc ctcttcattg gaaggcgcca catcgttcgg aagcgcacgc 2280
 tgccggaggct gctgcaggag agggagcttg tggagcctct tacaccagat ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtgg 2400
 gctccgggtg gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
 aaattcccgt cgctatcaag gaattaagag aagcaacatc tccgaaagcc aacaaggaaa 2520
 35 tcctctgtag agcctacgtg atggccagcg tggcaaccc ccactgtgtg gcctgtgtg 2580
 gcatctgcct cacctccacc gtgcagctca tcacgcagct catgccctc ggctgcctc 2640
 tggactatgt ccgggaacac aaagacaata ttggctccca gtacctgtc aactggtgtg 2700
 tgcagatcgc aaagggcatg aactacttgg aggaccgtcg cttggtgcac cgcgacctg 2760
 cagccaggaa cgtaactgtg aaaacaccgc agcatgtcaa gatcacagat tttgggcgg 2820
 ccaaactgct ggggtgcgga gagaaagaat accatgcaga aggaggcaaa gtgcctatca 2880
 40 agtggatggc attggaatca attttacaca gaattctata ccaccagat gatgtctgga 2940
 gctacggggt gactgtttgg gagttgatga cctttggatc caagccatat gacggaatcc 3000
 ctgccagcga gatctcctcc atcctggaga aaggagaacg cctccctcag ccacccatat 3060
 gtaccatcga tgtctacatg atcatggtca agtgcctggat gatagacgca gatagtgcgc 3120
 caaagttccg tgagttgatc atcgaattct ccaaaatggc ccgagacccc cagcgctacc 3180
 45 ttgtcattca gggggatgaa agaatgcatt tgccaagtcc tacagactcc aacttctacc 3240
 gtgccctgat ggatgaagaa gacatggacg acgtggtgga tgccgacgag tacctcatcc 3300
 cacagcagg cttcttcagc agcccccca cgtcacggac tccccctctg agctctctga 3360
 gtgcaaccag caacaattcc accgtggctt gcattgatag aaatgggctg caaagctgtc 3420
 ccatcaagga agacagcttc ttgcagcgat acagctcaga ccccacaggc gccttgactg 3480
 aggacagcat agacgacacc ttcttcccag tgcctgaata cataaaccag tccgttccca 3540
 50 aaaggcccg tggtctctgt cagaatcctg tctatcacia tcagcctctg aaccccgcgc 3600
 ccagcagaga cccacactac caggaccccc acagcactgc agtgggcaac cccgagtatc 3660
 tcaacactgt ccagcccacc tgtgtcaaca gcacattcga cagccctgcc cactgggccc 3720
 agaaaggcag ccaccaaatt agcctggaca accctgacta ccagcaggac ttctttccca 3780
 aggaagccaa gccaaatggc atctttaagg gctccacagc tgaaaatgca gaatacctaa 3840
 gggtcgcgcc acaaagcagt gaatttatg gagcatga 3878

<210> 439

<211> 3878

<212> DNA

<213> Homo sapiens

<400> 439

	cccggcgcag	cgcggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccg	gccgcccaga	ccggacgaca	120
5	ggccacctcg	tccggcgccc	cccaggtccc	cgccctcgcc	ccaacgccac	aaccaccg	180
	cacggccccc	tgactccg	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatg	accctccgg	acggccggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tccgggctc	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
10	taaagaccat	ccaggagggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	ttaggtctt	atctaactat	gatgcataata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caacctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtc	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
15	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	gtgtctgcag	960
	gctgcacagg	ccccggggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatat	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
20	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tggtagaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccgggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccittgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
25	cagtcgctcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccctct	ggagggtgag	ccaaggaggt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
30	acccaggtg	cctgcctcag	cctatgaaca	tcactgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaagg	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggtatgggtgg	ggccctcctc	ttgtgtcttg	2220
	tgggtggccct	ggggatcggc	ctcttcattg	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
35	tgccggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtgc	gttcggcagc	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tgacaacccc	ccacgtgtgc	cgctgtcttg	2580
	gcattctgct	cacctccacc	gtgcagctca	tgcgcagct	catgcccttc	ggctgcctcc	2640
40	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtct	aactgggtgtg	2700
	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgaccttg	2760
	cagccaggga	cgtactggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggagg	2820
	ccaaactgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	attggaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
45	gctagggggg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccaccatat	3060
	gtaccatcga	tgtctacatg	atcatggtca	agtgtggat	gatagacgca	gatagtcgcc	3120
	caaagttccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtgga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agcccccca	cgtcacggac	tccccctctg	agctctctga	3360
50	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
	ccatcaaggga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgcctgaata	cataaaccag	tccgttccca	3540
	aaaggccgcg	tggctctgtg	cagaatcctg	tctatcacia	tcagcctctg	aaccccgcg	3600
	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccagtatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactggggcc	3720
55	agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gctccacagc	tgaaaatgca	gaatacctaa	3840
	gggtcgcg	acaaagcagt	gaatttattg	gagcatga			3878

<210> 440
<211> 3878
<212> DNA
<213> Homo sapiens

5

<400> 440

10

15

20

25

30

35

40

45

50

55

	cccgggcgag	cgcgggccga	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccg	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtcc	cccaggtccc	cgctctcgcc	ccaacgccac	aaccaccg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gagggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggttat	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
10	tgacagtctt	atctaactat	gatgcaaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	gtgtctgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatat	agctttgggt	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgctgcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaatagggtat	tggtagaatt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggttttt	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccgg	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	acccagagtgc	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtgc	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
30	tcatgggaga	aaacaacacc	ctgggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgagggc	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgtgtctgg	2220
	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcggttcg	aagcgcacgc	2280
	tgcgagggtc	gctgcaggag	agggagcttg	tggagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
35	gctccgggtg	gttcggcagc	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattccccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tgacaaccc	ccacgtgtgc	ggctgtctgg	2580
	gcactggcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	cgcctgcctc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtct	aactgggtgtg	2700
	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
40	cagccaggga	cgtactgggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggagg	2820
	ccaaactgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	attggaatca	atcttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctactgggtg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccaccatat	3060
	gtaccatcga	tgtctacatg	atcatggtca	agtgtggtat	gatagacgca	gatagtgcgc	3120
45	caaagtctcc	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtgga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agcccccca	cgtcacggac	tccccctctg	agctctctga	3360
	gtgcaaccag	caacaattcc	accgtggcct	gcattgatag	aaatgggctg	caaagctgtc	3420
	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
50	aggacagcat	agacgacacc	ttcctccag	tgctgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggtctgtg	cagaatcctg	tctatcaca	tcagcctctg	aaccccgcg	3600
	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccgagtatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactggggcc	3720
	agaaaggcag	ccaccaaat	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gtccacagc	tgaaaatgca	gaatacctaa	3840
55	gggtcgcg	acaaagcagt	gaatttattg	gagcatga			3878

<210> 441

<211> 3869

<212> DNA

<213> Homo sapiens

<400> 441

5

10

15

20

25

30

35

40

45

50

55

```

cccggcgcag cgcggccgca gcagcctccg cccccgcac ggtgtgagcg cccgacgcgg 60
ccgaggcggc cggagtcccg agctagcccc ggcgccgccc gccgcccaga cccgacgaca 120
ggccacctcg tcggcgctccg cccgagtccc cgctctcgccg ccaacgccac aaccaccgcg 180
cacggccccc tgactccgtc cagtattgat cgggagagcc ggagcgagct cttcggggag 240
cagcgatgcg accctccggg acggccgggg cagcgctcct ggcgtgctg gctgcgctct 300
gcccggcgag tcgggctctg gaggaaaaga aagtttgcca aggcacgagt aacaagctca 360
cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
aggtggtcct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
taaagaccat ccaggagggtg gctggttatg tcctcattgc cctcaacaca gtggagcgaa 540
ttcctttgga aaacctgcag atcatcagag gaaatatgta ctacgaaaaa tcctatgcct 600
tagcagtcct atctaactat gatgcaataa aaaccggact gaaggagctg cccatgagaa 660
atttacagga aatcctgcat ggcgccgtgc ggttcagcaa caacctgcc ctgtgcaacg 720
tggagagcat ccagtggcgg gacatagtca gcagtgaact tctcagcaac atgtcgatgg 780
acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
gctgggggtg aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagt 900
gctccggggc ctgccgtggc aagtccccc gtgactgctg ccacaaccag tgtgctgcag 960
gctgcacagg ccccgggag agcgactgcc tctcataaaa caaattccga gacgagacca 1020
cgtgcaagga cacctgcccc ccactcatgc tctacaaccc caccacgtac cagatggatg 1080
tgaaccccgga gggcaaacac agctttgggtg ccacctgctg gaagaagtgt ccccgtaatt 1140
atgtggtgac agatcacggc tcgtgcgtcc gagcctgtgg ggccgacagc tatgagatgg 1200
aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgcaaa gtgtgtaacg 1260
gaataggtag tggtagaatt aaagactcac tctccataaa tgctacgaat attaaacact 1320
tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
aggaaatcac aggggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
cctttgagaa cctagaaatc atacgaggca ggaccaagca acatggtcag ttttctcttg 1560
cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
aactgtttgg gacctccggt cagaaaacca aaattataag caacagaggt gaaaacagct 1740
gcaaggccac aggccagggtc tgccatgcct tgtgtcccc cgagggctgc tggggcccgg 1800
agcccaggga ctgctctctc tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
gcaaccttct ggagggtgag ccaaggaggt ttgtggagaa ctctgagtgc atacagtgcc 1920
acccagagtg cctgcctcag gccatgaaaca aggacgggga ccagacaact 1980
gtatccagtg tgcccactac attgacggcc cccactgctg caagacctgc ccggcaggag 2040
tcatgggaga aaacaacacc ctggtctgga agtacgcaga cgccggccat gtgtgccacc 2100
tgtgccatcc aaactgcacc tacggatgca ctgggcccagg tcttgagggc tgtccaacga 2160
atgggcctaa gatccccgtc atcgccactg ggatgggtggg ggccctcctc ttgctgctgg 2220
tggtgccctt ggggatcggc ctcttcattg gaagcgcca catcgctcgg aagcgaagtc 2280
tgccggaggc gctgcaggag agggagcttg tggagcctct tacaccagtg gagagaagtc 2340
ccaaccaagc tctcttgagg atcttgagg aaactgaatt caaaaagatc aaagtgtctg 2400
gctccggtgc gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
aaattcccgt cgctatcaag gaaccaacat ctccgaaagc caacaaggaa atcctcgatg 2520
aagcctacgt gatggccagc gtggacaacc cccacgtgtg ccgctgctg ggcactgtcc 2580
tcacctccac cgtgcagctc atcacgcagc tcattgccct cggctgcctc ctgactatg 2640
tccgggaaca caaagacaat attggctccc agtacctgct caactgggtg gtgcagatcg 2700
caaagggcat gaactacttg gaggaccgtc gcttgggtga ccgcgacctg gcagccagga 2760
acgtactggg gaaaacaccg cagcatgtca agatcacaga ttttgggctg gccaaactgc 2820
tgggtgcgga agagaaagaa taccatgcag aaggaggcaa agtgccatc agtgccatg 2880
cattggaatc aatitttacac agaattctata cccaccagag tgatgtctgg agctacgggg 2940
tgactgtttg ggagttgatg acctttggat ccaagccata tgacggaatc cctgccagcg 3000
agatctcctc catcctggag aaaggagaac gcctccctca gccaccata tgtaccatcg 3060
atgtctacat gatcatggtc aagtgttgga tgatagacgc agatagtcgc ccaaagttcc 3120
gtgagttgat catcgaattc tccaaaatgg cccgagaccc ccagcgctac cttgtcattc 3180
agggggatga aagaatgcat ttgccaaagt ctacagactc caacttctac cgtgccctga 3240
tggatgaaga agacatggag gacgtgggtg atgccgacga gtacctcatc ccacagcagg 3300
gcttcttcag agccccctcc acgtcacgga ctccccctct gagctctctg agtgcaacca 3360
gcaacaattc caccgtggct tgcattgata gaaatgggct gcaaagctgt cccatcaagg 3420
aagacagctt cttgcagcga tacagctcag accccacagg cgccttgact gaggacagca 3480
tagacgacac cttcctccca gtgcctgaat acataaacca gtccgttccc aaaaggcccc 3540
ctggctctgt gcagaatcct gtctatcaca atcagcctct gaaccccgcg cccagcagag 3600
accacaccta ccaggacccc cacagcactg cagtgggcaa ccccgagtat ctcaacactg 3660

```

```
tccagccac ctgtgtcaac agcacattcg acagccctgc ccactgggcc cagaaaggca 3720
gccaccaaata tagcctggac aaccctgact accagcagga cttctttccc aaggaagcca 3780
agccaaatgg catctttaag ggctccacag ctgaaaatgc agaataccta agggtcgcgc 3840
caciaagcag tgaatttatt ggagcatga 3869
```

5

```
<210> 442
<211> 3869
<212> DNA
<213> Homo sapiens
```

10

```
<400> 442
```

15

20

25

30

35

40

45

50

55

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccgc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctcc	cccaggtccc	cgctcgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
5	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggtatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
10	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccgga	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgatagtg	1200
	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaatagggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgctgtctc	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaagggaagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
30	tcatggggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatggtggg	ggccctctct	gtgtctcttg	2220
	tgggtggccct	gggggatcgcc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgacgcg	2280
	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacacccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
35	gctccgggtgc	gttcgggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaaccaacat	ctccgaaagc	caacaaggaa	atcctcgatg	2520
	aagcctacgt	gatggccagc	gtggacaacc	cccactgtgtg	cgccctgtgc	gtgactgtcc	2580
	tcacctccac	cgtgcagctc	atcacgcagc	tcattgccctt	cggctgcctc	ctggactatg	2640
	tccgggaaca	caaagacaat	attggctccc	agtacctgct	caactggtgt	gtgcagatcg	2700
	caaagggcat	gaactacttg	gaggaccgtc	gcttggtgca	ccgcgacctg	gcagccagga	2760
40	acgtactggt	gaaaacaccg	cagcatgtca	agatcacaga	ttttgggctg	gccaaactgc	2820
	tgggtgcgga	agagaaagaa	taccatgcag	aaggaggcaa	agtgccctac	aagtggatgg	2880
	catttgaatc	aattttacac	agaatctata	cccaccagag	tgatgtctgg	agctacgggg	2940
	tgactgtttg	ggagttgatg	acctttggat	ccaagccata	tgacggaatc	cctgccagcg	3000
	agatctcctc	catcctggag	aaaggagaac	gcctccctca	gccacccata	tgtaccatcg	3060
	atgtctacat	gatcatggtc	aagtgtctga	tgatagacgc	agatagtgcg	ccaaagttcc	3120
45	gtgagttgat	catcgaattc	tccaaaaatg	cccagagacc	ccagcgctac	cttgtcattc	3180
	agggcgatga	aagaatgcat	ttgccaagtc	ctacagactc	caacttctac	ggctgcctga	3240
	tggatgaaga	agacatggac	gacgtggtgg	atgccgacga	gtacctcatc	ccacagcagg	3300
	gcttcttcag	cagccccctc	acgtcacgga	ctccccctct	gagctctctg	agtgcaacca	3360
	gcaacaattc	caccgtggct	tgcattgata	gaaatgggct	gcaaagctgt	cccatcaagg	3420
50							
	aagacagctt	cttgacgcga	tacagctcag	acccacaggg	cgccttgact	gaggacagca	3480
	tagacgacac	cttcctccca	gtgccgtgaat	acataaacca	gtccgttccc	aaaaggcccc	3540
	ctggctctgt	gcagaatcct	gtctatcaca	atcagcctct	gaaccccgcg	cccagcagag	3600
	acccacacta	ccaggacccc	cacagcactg	cagtgggcaa	ccccgagtat	ctcaacactg	3660
	tccagcccac	ctgtgtcaac	agcacattcg	acagccctgc	ccactggggc	cagaaaggca	3720
55	gccaccaa	tagcctggac	aaccctgact	accagcagga	cttctttccc	aagggaagcca	3780
	agccaaatgg	catctttaag	ggctccacag	ctgaaaatgc	agaataccta	agggtcgcgc	3840
	cacaaagcag	tgaatttatt	ggagcatga				3869

<210> 443
 <211> 3860
 <212> DNA
 <213> Homo sapiens

5

<400> 443

10	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgctcgcgcg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgtct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
15	taaagaccat	ccaggaggtg	gctgggtatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccitttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaatt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
20	gctgggggtg	aggagagggg	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	tgtgtgtcag	960
	gctgcacagg	cccccgagg	agcgactgcc	tggctgtccg	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaacccccg	gggcaaatac	agctttgggt	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
25	aggaagacgg	cgtccgcaag	tgttaagaagt	gcgaagggcc	ttgcccgaac	gtgtgtaacg	1260
	gaataggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggttttt	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
30	ccttcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgtat	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgtctatg	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgtctccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgctgtctc	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
35	gcaaccttct	ggagggtgag	ccaaggaggt	ttgtggagaa	ctctgagtgc	ctacatggcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctctct	ttgtctcttg	2220
	tgggtggccct	ggggatcggc	ctcttcattg	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
40	tgcgagggct	gctgcaggag	agggagcttg	tggagcctct	tacaccagct	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtg	gttcgggcag	gtgtataaag	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaatcgaaag	ccaacaagga	aatcctcgat	gaagcctacg	2520
	tgatggccag	cgtaggacaac	ccccacgtgt	gccgcctgct	gggcatctgc	ctcacctcca	2580
45	ccgtgcagct	catcacgcag	ctcatgccct	tcggctgcct	cctggactat	gtccgggaac	2640
	acaaagacaa	tattggctcc	cagtacctgc	tcaactgggt	tgtgcagatc	gcaaagggca	2700
	tgaactactt	ggaggaccgt	cgcttgggtg	accgcgacct	ggcagccagg	aacgtactgg	2760
	tgaaaacacc	gcagcatgtc	aagatcacag	atcttgggct	ggccaaactg	ctgggtgcgg	2820
	aagagaaaga	ataccatgca	gaaggaggca	aagtgcctat	caagtggatg	gcatttgaat	2880
	caatttttaca	cagaatctat	acccaccaga	gtgatgtctg	gagctacggg	gtgactgttt	2940
	gggagttgat	gaccttttga	tccaagccat	atgacggaat	ccctgccagc	gagatctcct	3000
50	ccatcctgga	gaaaggagaa	cgctccctc	agccacccat	atgtaccatc	gatgtctaca	3060
	tgatcatggt	caagtgtctg	atgatagacg	cagatagtcg	cccaaagttc	cgtgagttga	3120
	tcatcgaatt	ctccaaaatg	gcccagagacc	cccagcgcta	ccttgtcatt	cagggggatg	3180

55

```

5  aaagaatgca tttgccaagt cctacagact ccaacttcta ccgtgccctg atggatgaag 3240
   aagacatgga cgacgtggtg gatgccgacg agtacctcat cccacagcag ggcttcttca 3300
   gcagccccctc cacgtcacgg actccccctcc tgagctctct gagtgcaacc agcaacaatt 3360
   ccaccgtggc ttgcattgat agaaatgggc tgcaaagctg tcccatcaag gaagacagct 3420
   tcttgacgagc atacagctca gaccccacag gcgccttgac tgaggacagc atagacgaca 3480
   ccttcctccc agtgcctgaa tacataaaacc agtccgttcc caaaaggccc gctggctctg 3540
   tgcagaatcc tgtctatcac aatcagcctc tgaaccccgcc gccacgcaga gacccacact 3600
   accaggacccc ccacagcact gcagtgggca accccgagta tctcaacact gtccagccca 3660
   cctgtgtcaa cagcacattc gacagccctg cccactgggc ccagaaaggc agccaccaa 3720
10 ttagcctgga caaccctgac taccagcagg acttctttcc caaggaagcc aagccaaatg 3780
   gcatctttaa gggctccaca gctgaaaatg cagaatacct aagggtcgcg ccacaaagca 3840
   gtgaatttat tggagcatga

```

<210> 444

<211> 3860

15 <212> DNA

<213> Homo sapiens

<400> 444

20

25

30

35

40

45

50

55

	cccggcgag	cgcgccgca	gcagcctccg	cccccgac	ggtgtgagcg	cccagcgcg	60
	ccgaggcggc	cggagtcccg	agctagcccc	ggcgccgccc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctcg	cccaggtccc	cgccctcgccc	ccaacgccac	aaccaccgag	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	atttacagga	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tggaattttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggttat	tcctcattgc	cctcaacaca	gtggagcgaa	540
10	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagctct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagagggg	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
15	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgggg	agcgactgcc	tgtctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	ccactgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataaggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	gaaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccgggtggca	ctgtgggtgtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccttttgagaa	cctagaaaatc	atacgcggca	ggaccaagca	acatgggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgttttg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	ccactgcgt	caagacctgc	ccggcaggag	2040
30	tcatggggaga	aaacaacacc	ctggcttggg	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgctgctgg	2220
	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggagggt	gctgcaggag	agggagcttg	tggagcctct	tacaccagct	ggagaagctc	2340
35	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctgg	2400
	gctccgggtg	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gttccgaaag	ccaacaagga	aatcctcgat	gaagcctacg	2520
	tgatggccag	cgtggacaac	ccccacgtgt	gccgcctgct	gggcatctgc	ctcacctcca	2580
	ccgtgcagct	catcacgcag	ctcatgccct	tcggctgcct	cctggactat	gtccgggaac	2640
	acaaagacaa	tattggctcc	cagtacgtgc	tgaactgggt	tgtgcagatg	gcaaaaggga	2700
40	tgaactactt	ggaggaccgt	cgcttgggtg	accgcgacct	ggcagccagg	aacgtactgg	2760
	tgaaaacacc	gcagcatgtc	aagatcacag	attttgggct	ggccaaactg	ctgggtgcgg	2820
	aagagaaaaga	ataccatgca	gaaggaggca	aagtgcctat	caagtggatg	gcattggaat	2880
	caatttttaca	cagaatctat	acccaccaga	gtgatgtctg	gagctacggg	gtgactgttt	2940
45	gggagttgat	gaccttttga	tccaagccat	atgacggaat	ccctgccagc	gagatctcct	3000
	ccatcctgga	gaaaggagaa	cgctccctc	agccacccat	atgtaccatc	gatgtctaca	3060
	tgatcatggt	caagtgtctg	atgatagacg	cagatagtgc	cccaaagttc	cgtgagttga	3120
	tcatcgaatt	ctccaaaatg	gcccgagacc	cccagcgcta	ccttgtcatt	cagggggatg	3180
	aaagaatgca	tttgccaagt	cctacagact	ccaacttcta	ccgtgccctg	atggtatgaag	3240
	aagacatgga	cgacgtgggtg	gatgccgacg	agtacctcat	cccacagcag	ggcttcttca	3300
50	gcagccccct	cacgtcacgg	actccccctc	tgagctctct	gagtgcaccc	agcaacaatt	3360
	ccaccgtggc	ttgcattgat	agaaaatggg	tgcaaagctg	tcccatcaag	gaagacagct	3420
	tcttgacgag	atacagctca	gacccccacg	gcgccttgac	tgaggacagc	atagacgaca	3480
	ccttcctccc	agtgcctgaa	tacataaacc	agtccgttcc	caaaaggccc	gctggctctg	3540
	tgcagaatcc	tgtctatcac	aatcagcctc	tgaaccccgc	gcccagcaga	gacccacact	3600
	accagagccc	ccacagcact	gcagtgggca	accccgagta	tctcaacact	gtccagccca	3660
	cctgtgtcaa	cagcacattc	gacagccctg	cccactgggc	ccagaaaggc	agccaccaa	3720
55	ttagcctgga	caaccctgac	taccagcagg	acttctttcc	caaggaagcc	aagccaaatg	3780
	gcatcttttaa	gggctccaca	gctgaaaatg	cagaatacct	aagggtcgcg	ccacaaagca	3840
	gtgaattttat	tggagcatga					3860

<210> 445
 <211> 3878
 <212> DNA
 <213> Homo sapiens

5

<400> 445

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcg	60
	ccgaggcggc	cggagtcctcg	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
10	ggccacctcg	tcggcgctccg	cccagatccc	cgctctcgccg	ccaacgccac	aaccaccgag	180
	cacggccccc	tgactccgctc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tggaattttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
15	taaagaccat	ccaggagggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tgagagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
20	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgagg	agcgactgcc	tggtctgccg	caaattccga	gacgaagcca	1020
	cgtgcaaggga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
25	aggaagcagg	cgtccgcaag	tgtaagaaat	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaaatc	atagcgggca	ggaccaagca	acatgggtcag	ttttctcttg	1560
30	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggtgc	tgggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
35	gcaaccttct	ggagggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	acccagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctcttc	ttgctgctgg	2220
	tggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
40	tgcgagggct	gctgcaggag	agggagcttg	tggagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaaag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cgctgtctgg	2580
45	gcactgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgctc	aactggtgtg	2700

50

55

	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttggtgcac	cgcgacctgg	2760
	cagccaggaa	cgtactgggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggcggg	2820
	ccaaactgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
5	agtggatggc	attggaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacggggg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccacccatat	3060
	gtaccatcga	tgtctacatg	atcatgggtca	agtgtggat	gatagacgca	gatagtcgcc	3120
	caaagtccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaagtcc	tacagactcc	aacttctacc	3240
10	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agccccctcca	cgtcacggac	tccccctctg	agctctctga	3360
	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgctgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggctctgtg	cagaatcctg	tctatcacia	tcagcctctg	aaccccgcg	3600
15	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccagtatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactgggccc	3720
	agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gctccacagc	tgaaaatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaattttattg	gagcatga			3878

20 <210> 446
 <211> 3878
 <212> DNA
 <213> Homo sapiens

25 <400> 446

30

35

40

45

50

55

	cccgccgcag	cgccggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgccgcgc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgccctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggttat	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccitttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
10	tagcagtcct	atctaactat	gatgcataata	aaaccggact	gaaggagctg	cccatgagag	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtc	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aggtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccga	gggcaaatat	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	ctcaaggag	ataatgtatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	acccaggtg	ctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
30	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggatgggtggg	ggccctcctc	ttgtgcttgg	2220
	tggtggccct	ggggatcggc	ctcttcattg	gaaggcgcca	catcgttcgg	aagcgacgc	2280
	tgccggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccagct	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
35	gctccgggtg	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460

	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cgcctgctgg	2580
40	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtct	aactgggtgtg	2700
	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
	cagccaggaa	cgtactgggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggctgg	2820
	ccaaactgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgacctatg	2880
	agtggatggc	attggaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
45	gctacgggg	gactgtttgg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccaccatat	3060
	gtaccatcga	tgtctacatg	atcatggtca	agtgtggat	gatagacgca	gatagtgcgc	3120
	caaagttccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtga	tgccgacgag	tacctcatcc	3300
	cacagcagg	cttcttcagc	agcccccca	cgtcacggac	tccccctctg	agctctctga	3360
50	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgcctgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggctctgtg	cagaatcctg	tctatcacaa	tcagcctctg	aaccccgcgc	3600
	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccagtatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	caactggccc	3720
55	agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gtcccacagc	tgaatatgca	gaatacctaa	3840
	gggtcgcgccc	acaaagcagt	gaatttattg	gagcatga			3878

<210> 447
 <211> 36
 <212> DNA
 <213> Homo sapiens
 5
 <400> 447
 tgtaaacga cggccagtgc cccagaccgg acgaca 36
 10
 <210> 448
 <211> 39
 <212> DNA
 <213> Homo sapiens
 15
 <400> 448
 caggaaacag ctatgaccag ggcaatgagg acataacca 39
 20
 <210> 449
 <211> 38
 <212> DNA
 <213> Homo sapiens
 <400> 449
 tgtaaacga cggccagtgc tggccttgc gaattgg 38
 25
 <210> 450
 <211> 40
 <212> DNA
 <213> Homo sapiens
 30
 <400> 450
 caggaaacag ctatgacccc atcgacatgt tgctgagaaa 40
 35
 <210> 451
 <211> 38
 <212> DNA
 <213> Homo sapiens
 <400> 451
 tgtaaacga cggccagtga aggagctgcc catgagaa 38
 40
 <210> 452
 <211> 38
 <212> DNA
 <213> Homo sapiens
 45
 <400> 452
 caggaaacag ctatgaccgc tggcttcgtc tcggaatt 38
 50
 <210> 453
 <211> 40
 <212> DNA
 <213> Homo sapiens
 55
 <400> 453
 tgtaaacga cggccagtga aactgaccaa aatcatctgt 40
 <210> 454
 <211> 40

<212> DNA
 <213> Homo sapiens

<400> 454
 5 caggaaacag ctatgaccta cctattccgt tacacacttt 40

<210> 455
 <211> 40
 <212> DNA
 10 <213> Homo sapiens

<400> 455
 tgtaaaacga cggccagtcc gtaattatgt ggtgacagat 40

15 <210> 456
 <211> 40
 <212> DNA
 <213> Homo sapiens

20 <400> 456
 caggaaacag ctatgaccgc gtatgatttc taggttctca 40

25 <210> 457
 <211> 41
 <212> DNA
 <213> Homo sapiens

30 <400> 457
 tgtaaaacga cggccagtct gaaaaccgta aaggaaatca c 41

35 <210> 458
 <211> 37
 <212> DNA
 <213> Homo sapiens

<400> 458
 caggaaacag ctatgacccc tgcctcggct gacattc 37

40 <210> 459
 <211> 40
 <212> DNA
 <213> Homo sapiens

45 <400> 459
 tgtaaaacga cggccagtta agcaacagag gtgaaaacag 40

50 <210> 460
 <211> 40
 <212> DNA
 <213> Homo sapiens

<400> 460
 caggaaacag ctatgaccgg tgtgttttc tcccatgact 40

55 <210> 461
 <211> 38
 <212> DNA
 <213> Homo sapiens

	<400> 461 tgtaaaacga cggccagtg accagacaac tgtatcca	38
5	<210> 462 <211> 40 <212> DNA <213> Homo sapiens	
10	<400> 462 caggaaacag ctatgacctt cctcaagat cctcaagaga	40
15	<210> 463 <211> 38 <212> DNA <213> Homo sapiens	
20	<400> 463 tgtaaaacga cggccagtga tcggcctctt catgcgaa	38
25	<210> 464 <211> 38 <212> DNA <213> Homo sapiens	
30	<400> 464 caggaaacag ctatgaccac ggtggagggtg aggcagat	38
35	<210> 465 <211> 39 <212> DNA <213> Homo sapiens	
40	<400> 465 tgtaaaacga cggccagtgc aaagccaaca aggaaatcc	39
45	<210> 466 <211> 40 <212> DNA <213> Homo sapiens	
50	<400> 466 caggaaacag ctatgaccat tccaatgcca tccacttgat	40
55	<210> 467 <211> 39 <212> DNA <213> Homo sapiens	
	<400> 467 tgtaaaacga cggccagtaa caccgcagca tgtcaagat	39
	<210> 468 <211> 39 <212> DNA <213> Homo sapiens	
	<400> 468 caggaaacag ctatgaccct cgggccattt tggagaatt	39

	<210> 469	
	<211> 39	
	<212> DNA	
	<213> Homo sapiens	
5		
	<400> 469	
	tgtaaaacga cggccagttc agccacccat atgtaccat	39
10		
	<210> 470	
	<211> 39	
	<212> DNA	
	<213> Homo sapiens	
15		
	<400> 470	
	caggaaacag ctatgaccgc ttgcagccc atttctatc	39
20		
	<210> 471	
	<211> 38	
	<212> DNA	
	<213> Homo sapiens	
25		
	<400> 471	
	tgtaaaacga cggccagtac agcagggctt cttcagca	38
30		
	<210> 472	
	<211> 38	
	<212> DNA	
	<213> Homo sapiens	
35		
	<400> 472	
	caggaaacag ctatgacctg acacaggtgg gctggaca	38
40		
	<210> 473	
	<211> 40	
	<212> DNA	
	<213> Homo sapiens	
45		
	<400> 473	
	tgtaaaacga cggccagtga atcctgtcta tcacaatcag	40
50		
	<210> 474	
	<211> 40	
	<212> DNA	
	<213> Homo sapiens	
55		
	<400> 474	
	caggaaacag ctatgaccgg tatcgaaaga gtctggattt	40
60		
	<210> 475	
	<211> 38	
	<212> DNA	
	<213> Homo sapiens	
65		
	<400> 475	
	tgtaaaacga cggccagtgc tccacagctg aaaatgca	38
70		
	<210> 476	
	<211> 39	

<212> DNA
 <213> Homo sapiens

<400> 476
 5 caggaaacag ctatgaccac gttgcaaaac cagtctgtg 39

<210> 477
 <211> 25
 <212> PRT
 10 <213> Homo sapiens

<400> 477

15 Lys Thr Pro Gln His Val Lys Ile Thr Asp Phe Gly Arg Ala Lys Leu
 1 5 10 15
 Leu Gly Ala Glu Glu Lys Glu Tyr His
 20 25

20 <210> 478
 <211> 25
 <212> PRT
 <213> Homo sapiens

25 <400> 478

Lys Thr Pro Gln His Val Lys Ile Thr Asp Phe Gly Leu Ala Lys Leu
 1 5 10 15
 30 Leu Gly Ala Glu Glu Lys Glu Tyr His
 20 25

<210> 479
 <211> 25
 35 <212> PRT
 <213> Homo sapiens

<400> 479

40 His Glu Asp Leu Thr Val Lys Ile Gly Asp Phe Gly Leu Ala Thr Val
 1 5 10 15
 Lys Ser Arg Trp Ser Gly Ser His Gln
 20 25

45 <210> 480
 <211> 25
 <212> PRT
 <213> Homo sapiens

50 <400> 480

Glu Thr Glu Phe Lys Lys Ile Lys Val Leu Ser Ser Gly Ala Phe Gly
 1 5 10 15
 55 Thr Val Tyr Lys Gly Leu Trp Ile Pro
 20 25

<210> 481
<211> 25
<212> PRT
<213> Homo sapiens

<400> 481

Glu Thr Glu Phe Lys Lys Ile Lys Val Leu Gly Ser Gly Ala Phe Gly
1 5 10 15
Thr Val Tyr Lys Gly Leu Trp Ile Pro
20 25

<210> 482
<211> 25
<212> PRT
<213> Homo sapiens

<400> 482

Asp₁ Gly Gln Ile Thr₅ Val Gly Gln Arg Ile₁₀ Gly Ser Gly Ser Phe₁₅ Gly
Thr Val Tyr Lys₂₀ Gly Lys Trp His Gly₂₅

<210> 483
<211> 19
<212> PRT
<213> Homo sapiens

<400> 483

Val Ala Ile Lys Thr Ser Pro Lys Ala Asn Lys Glu Ile Leu Asp Glu
1 5 10 15
Ala Tyr Val

<210> 484
<211> 16
<212> PRT
<213> Homo sapiens

<400> 484

Val Ala Ile Lys Glu Leu Arg Glu Ala Thr Leu Asp Glu Ala Tyr Val
1 5 10 15

<210> 485
<211> 21
<212> PRT
<213> Homo sapiens

<400> 485

Val Ala Ile Lys Glu Pro Thr Ser Pro Lys Ala Asn Lys Glu Ile Leu
1 5 10 15

Asp Glu Ala Tyr Val
5 20

<210> 486
<211> 18
<212> PRT
10 <213> Homo sapiens

<400> 486

Val Ala Ile Lys Glu Ser Lys Ala Asn Lys Glu Ile Leu Asp Glu Ala
15 1 5 10 15

Tyr Val

<210> 487
20 <211> 18
<212> PRT
<213> Homo sapiens

<400> 487

Val Ala Ile Lys Val Pro Lys Ala Asn Lys Glu Ile Leu Asp Glu Ala
25 1 5 10 15

Tyr Val

30 <210> 488
<211> 24
<212> PRT
<213> Homo sapiens

35 <400> 488

Val Ala Ile Lys Glu Leu Arg Glu Ala Thr Ser Pro Lys Ala Asn Lys
40 1 5 10 15

Glu Ile Leu Asp Glu Ala Tyr Val
20

<210> 489
45 <211> 25
<212> PRT
<213> Homo sapiens

<400> 489

50 Val Ala Val Lys Met Leu Asn Val Thr Ala Pro Thr Pro Gln Gln Leu
1 5 10 15

Gln Ala Phe Lys Asn Glu Val Gly Val
55 20 25

<210> 490
<211> 6

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: illustrative mutation motif

```
<220>  
<221> MOD_RES  
<222> (2)  
<223> variable residue
```

```
<220>  
<221> MOD_RES  
<222> (4)..(5)  
<223> variable residue
```

<400> 490

Gly Xaa Gly Xaa Xaa Gly
1 5

<210> 491
<211> 252
<212> PRT
<213> Homo sapiens

<400> 491

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

<210> 492

<211> 268

45 <212> PRT

<213> Homo sapiens

<400> 492

Phe Lys Lys Ile Lys Val Leu Gly Ser Gly Ala Phe Gly Thr Val Tyr
 1 5 10 15

55

5 Lys Gly Leu Trp₂₀ Ile Pro Glu Gly Glu₂₅ Lys Val Lys Ile Pro Val Ala
 Ile Lys Glu₃₅ Leu Arg Glu Ala Thr₄₀ Ser Pro Lys Ala Asn₄₅ Lys Glu Ile
 Leu Asp₅₀ Glu Ala Tyr Val Met₅₅ Ala Ser Val Asp Asn₆₀ Pro His Val Cys
 10 Arg Leu Leu Gly Ile Cys₇₀ Leu Thr Ser Thr Val₇₅ Gln Leu Ile Thr Gln₈₀
 Leu Met Pro Phe Gly₈₅ Cys Leu Leu Asp Tyr₉₀ Val Arg Glu His Lys₉₅ Asp
 15 Asn Ile Gly Ser₁₀₀ Gln Tyr Leu Leu Asn₁₀₅ Trp Cys Val Gln Ile Ala Lys
 Gly Met Asn₁₁₅ Tyr Leu Glu Asp Arg Arg Leu Val His Arg₁₂₅ Asp Leu Ala
 20 Ala Arg Asn Val Leu Val Lys₁₃₅ Thr Pro Gln His Val₁₄₀ Lys Ile Thr Asp
 Phe Gly Leu Ala Lys Leu₁₅₀ Leu Gly Ala Glu Glu₁₅₅ Lys Glu Tyr His Ala₁₆₀
 25 Glu Gly Gly Lys Val₁₆₅ Pro Ile Lys Trp Met₁₇₀ Ala Leu Glu Ser Ile₁₇₅ Leu
 His Arg Ile Tyr₁₈₀ Thr His Gln Ser Asp₁₈₅ Val Trp Ser Tyr Gly₁₉₀ Val Thr
 30 Val Trp Glu₁₉₅ Leu Met Thr Phe Gly₂₀₀ Ser Lys Pro Tyr Asp₂₀₅ Gly Ile Pro
 Ala Ser Glu Ile Ser Ser Ile₂₁₅ Leu Glu Lys Gly Glu₂₂₀ Arg Leu Pro Gln
 35 Pro Pro Ile Cys Thr Ile₂₃₀ Asp Val Tyr Met Ile₂₃₅ Met Val Lys Cys Trp₂₄₀
 Met Ile Asp Ala Asp₂₄₅ Ser Arg Pro Lys Phe₂₅₀ Arg Glu Leu Ile Ile₂₅₅ Glu
 40 Phe Ser Lys Met₂₆₀ Ala Arg Asp Pro Gln Arg Tyr Leu

<210> 493

<211> 54

<212> DNA

45 <213> Homo sapiens

<400> 493

aaaattcccg tcgctatcaa ggaattaaga gaagcaacat ctccgaaagc caac 54

50 <210> 494

<211> 24

<212> DNA

<213> Homo sapiens

55 <400> 494

tttgggctgg ccaaactgct gggt 24

<210> 495

<211> 39
 <212> DNA
 <213> Homo sapiens

5 <400> 495
 aaaattcccg tcgctatcaa aacatctccg aaagccaac 39

<210> 496
 <211> 24
 10 <212> DNA
 <213> Homo sapiens

<400> 496
 15 ttgggctgg ccaaactgct gggt 24

<210> 497
 <211> 42
 <212> DNA
 20 <213> Homo sapiens

<400> 497
 aaaattcccg tcgctatcaa ggaatcatct ccgaaagcca ac 42

25 <210> 498
 <211> 24
 <212> DNA
 <213> Homo sapiens

30 <400> 498
 ttgggctgg ccaaactgct gggt 24

<210> 499
 <211> 36
 <212> DNA
 35 <213> Homo sapiens

<400> 499
 aaaattcccg tcgctatcaa ggaatcgaaa gccaac 36

40 <210> 500
 <211> 24
 <212> DNA
 <213> Homo sapiens

45 <400> 500
 ttgggctgg ccaaactgct gggt 24

<210> 501
 <211> 54
 50 <212> DNA
 <213> Homo sapiens

<400> 501
 55 aaaattcccg tcgctatcaa ggaattaaga gaagcaacat ctccgaaagc caac 54

<210> 502
 <211> 24
 <212> DNA

<213> Homo sapiens
 <400> 502
 5 tttgggcggg ccaaactgct gggt 24
 <210> 503
 <211> 54
 <212> DNA
 <213> Homo sapiens
 10
 <400> 503
 aaaattcccg tcgctatcaa ggaattaaga gaagcaacat ctccgaaagc caac 54
 <210> 504
 15 <211> 24
 <212> DNA
 <213> Homo sapiens
 <400> 504
 20 tttgggctgg ccaaacagct gggt 24
 <210> 505
 <211> 25
 <212> DNA
 25 <213> Homo sapiens
 <400> 505
 gcaatatcag ccttaggtgc ggctc 25
 30 <210> 506
 <211> 26
 <212> DNA
 <213> Homo sapiens
 35 <400> 506
 catagaaagt gaacatttag gatgtg 26
 <210> 507
 <211> 24
 40 <212> DNA
 <213> Homo sapiens
 <400> 507
 45 ctaacgttcg ccagccataa gtcc 24
 <210> 508
 <211> 26
 <212> DNA
 <213> Homo sapiens
 50 <400> 508
 gctgcgagct caccagaat gtctgg 26
 <210> 509
 55 <211> 18
 <212> PRT
 <213> Homo sapiens

<400> 509

5 Lys Ile Pro Val Ala Ile Lys Glu Leu Arg Glu Ala Thr Ser Pro Lys
1 5 10 15
Ala Asn

<210> 510

<211> 8

10 <212> PRT

<213> Homo sapiens

<400> 510

15 Phe Gly Leu Ala Lys Leu Leu Gly
1 5

<210> 511

<211> 3878

20 <212> DNA

<213> Homo sapiens

<220>

<221> CDS

25 <222> (246)..(3875)

<400> 511

30

35

40

45

50

55

cccggcgag cgcgccgca gcagcctccg ccccccgcac ggtgtgagcg cccgacgcgg 60
 ccgaggcggc cggagtcccc agctagcccc ggcggccgcc gccgcccaga ccggacgaca 120
 5 ggccacctcg tcggcgctcg cccgagtccc cgctctgccg ccaacgccac aaccaccgcg 180
 cacggccccc tgactccgtc cagtattgat cgggagagcc ggagcgagct cttcgggggag 240
 cagcg atg cga ccc tcc ggg acg gcc ggg gca gcg ctc ctg gcg ctg ctg 290
 10 Met Arg Pro Ser Gly Thr Ala Gly Ala Ala Leu Leu Ala Leu Leu 15
 gct gcg ctc tgc ccg gcg agt cgg gct ctg gag gaa aag aaa gtt tgc 338
 Ala Ala Leu Cys Pro Ala Ser Arg Ala Leu Glu Glu Lys Lys Val Cys 30
 15 caa ggc acg agt aac aag ctc acg cag ttg ggc act ttt gaa gat cat 386
 Gln Gly Thr Ser Asn Lys Leu Thr Gln Leu Gly Thr Phe Glu Asp His 45
 ttt ctc agc ctc cag agg atg ttc aat aac tgt gag gtg gtc ctt ggg 434
 Phe Leu Ser Leu Gln Arg Met Phe Asn Asn Cys Glu Val Val Leu Gly 60
 20 aat ttg gaa att acc tat gtg cag agg aat tat gat ctt tcc ttc tta 482
 Asn Leu Glu Ile Thr Tyr Val Gln Arg Asn Tyr Asp Leu Ser Phe Leu 75
 25 aag acc atc cag gag gtg gct ggt tat gtc ctc att gcc ctc aac aca 530
 Lys Thr Ile Gln Glu Val Ala Gly Tyr Val Leu Ile Ala Leu Asn Thr 95
 gtg gag cga att cct ttg gaa aac ctg cag atc atc aga gga aat atg 578
 Val Glu Arg Ile Pro Leu Glu Asn Leu Gln Ile Ile Arg Gly Asn Met 110
 30 tac tac gaa aat tcc tat gcc tta gca gtc tta tct aac tat gat gca 626
 Tyr Tyr Glu Asn Ser Tyr Ala Leu Ala Val Leu Ser Asn Tyr Asp Ala 125
 35 aat aaa acc gga ctg aag gag ctg ccc atg aga aat tta cag gaa atc 674
 Asn Lys Thr Gly Leu Lys Glu Leu Pro Met Arg Asn Leu Gln Glu Ile 140
 ctg cat ggc gcc gtg cgg ttc agc aac aac cct gcc ctg tgc aac gtg 722
 40 Leu His Gly Ala Val Arg Phe Ser Asn Asn Pro Ala Leu Cys Asn Val 155
 gag agc atc cag tgg cgg gac ata gtc agc agt gac ttt ctc agc aac 770

	Glu 160	Ser	Ile	Gln	Trp	Arg 165	Asp	Ile	Val	Ser	Ser 170	Asp	Phe	Leu	Ser	Asn 175	
5	atg Met	tcg Ser	atg Met	gac Asp	ttc Phe 180	cag Gln	aac Asn	cac His	ctg Leu	ggc Gly 185	agc Ser	tgc Cys	caa Gln	aag Lys	tgt Cys 190	gat Asp	818
	cca Pro	agc Ser	tgt Cys	ccc Pro 195	aat Asn	ggg Gly	agc Ser	tgc Cys	tgg Trp 200	ggc Gly	gca Ala	gga Gly	gag Glu	gag Glu 205	aac Asn	tgc Cys	866
10	cag Gln	aaa Lys	ctg Leu 210	acc Thr	aaa Lys	atc Ile	atc Ile	tgt Cys 215	gcc Ala	cag Gln	cag Gln	tgc Cys	tcc Ser 220	ggg Gly	cgc Arg	tgc Cys	914
15	cgt Arg	ggc Gly	aag Lys	tcc Ser	ccc Pro	agt Ser	gac Asp	tgc Cys	tgc Cys	cac His	aac Asn	cag Gln	tgt Cys	gct Ala	gca Ala	ggc Gly	962
	225			230			235										
20	tgc Cys 240	aca Thr	ggc Gly	ccc Pro	cgg Arg	gag Glu 245	agc Ser	gac Asp	tgc Cys	ctg Leu	gtc Val 250	tgc Cys	cgc Arg	aaa Lys	ttc Phe	cga Arg 255	1010
	gac Asp	gaa Glu	gcc Ala	acg Thr	tgc Cys 260	aag Lys	gac Asp	acc Thr	tgc Cys	ccc Pro 265	cca Pro	ctc Leu	atg Met	ctc Leu	tac Tyr 270	aac Asn	1058
25	ccc Pro	acc Thr	acg Thr	tac Tyr 275	cag Gln	atg Met	gat Asp	gtg Val	aac Asn 280	ccc Pro	gag Glu	ggc Gly	aaa Lys	tac Tyr 285	agc Ser	ttt Phe	1106
	ggc Gly	gcc Ala	acc Thr 290	tgc Cys	gtg Val	aag Lys	aag Lys	tgt Cys 295	ccc Pro	cgt Arg	aat Asn	tat Tyr	gtg Val 300	gtg Val	aca Thr	gat Asp	1154
30	cac His	ggc Gly 305	tcg Ser	tgc Cys	gtc Val	cga Arg	gcc Ala 310	tgt Cys	ggg Gly	gcc Ala	gac Asp	agc Ser 315	tat Tyr	gag Glu	atg Met	gag Glu	1202
35	gaa Glu 320	gac Asp	ggc Gly	gtc Val	cgc Arg	aag Lys 325	tgt Cys	aag Lys	aag Lys	tgc Cys	gaa Glu 330	ggg Gly	cct Pro	tgc Cys	cgc Arg	aaa Lys 335	1250
	gtg Val	tgt Cys	aac Asn	gga Gly	ata Ile 340	ggt Gly	att Ile	ggt Gly	gaa Glu	ttt Phe 345	aaa Lys	gac Asp	tca Ser	ctc Leu	tcc Ser 350	ata Ile	1298
40	aat Asn	gct Ala	acg Thr	aat Asn 355	att Ile	aaa Lys	cac His	ttc Phe	aaa Lys 360	aac Asn	tgc Cys	acc Thr	tcc Ser	atc Ile 365	agt Ser	ggc Gly	1346
45	gat Asp	ctc Leu	cac His 370	atc Ile	ctg Leu	ccg Pro	gtg Val	gca Ala 375	ttt Phe	agg Arg	ggc Gly	gac Asp	tcc Ser 380	ttc Phe	aca Thr	cat His	1394
	act Thr	cct Pro 385	cct Pro	ctg Leu	gat Asp	cca Pro	cag Gln 390	gaa Glu	ctg Leu	gat Asp	att Ile	ctg Leu 395	aaa Lys	acc Thr	gta Val	aag Lys	1442
50	gaa Glu 400	atc Ile	aca Thr	ggg Gly	ttt Phe	ttg Leu 405	ctg Leu	att Ile	cag Gln	gct Ala	tgg Trp 410	cct Pro	gaa Glu	aac Asn	agg Arg	acg Thr 415	1490
55	gac Asp	ctc Leu	cat His	gcc Ala	ttt Phe 420	gag Glu	aac Asn	cta Leu	gaa Glu	atc Ile 425	ata Ile	cgc Arg	ggc Gly	agg Arg	acc Thr 430	aag Lys	1538
	caa	cat	ggc	cag	ttt	tct	ctt	gca	gtc	gtc	agc	ctg	aac	ata	aca	tcc	1586

	Gln	His	Gly	Gln	Phe	Ser	Leu	Ala	Val	Val	Ser	Leu	Asn	Ile	Thr	Ser	
				435					440					445			
5	ttg	gga	tta	cgc	tcc	ctc	aag	gag	ata	agt	gat	gga	gat	gtg	ata	att	1634
	Leu	Gly	Leu	Arg	Ser	Leu	Lys	Glu	Ile	Ser	Asp	Gly	Asp	Val	Ile	Ile	
			450					455					460				
	tca	gga	aac	aaa	aat	ttg	tgc	tat	gca	aat	aca	ata	aac	tgg	aaa	aaa	1682
	Ser	Gly	Asn	Lys	Asn	Leu	Cys	Tyr	Ala	Asn	Thr	Ile	Asn	Trp	Lys	Lys	
		465					470					475					
10	ctg	ttt	ggg	acc	tcc	ggg	cag	aaa	acc	aaa	att	ata	agc	aac	aga	ggg	1730
	Leu	Phe	Gly	Thr	Ser	Gly	Gln	Lys	Thr	Lys	Ile	Ile	Ser	Asn	Arg	Gly	
	480					485					490					495	
	gaa	aac	agc	tgc	aag	gcc	aca	ggc	cag	gtc	tgc	cat	gcc	ttg	tgc	tcc	1778
15	Glu	Asn	Ser	Cys	Lys	Ala	Thr	Gly	Gln	Val	Cys	His	Ala	Leu	Cys	Ser	
					500					505					510		
	ccc	gag	ggc	tgc	tgg	ggc	ccg	gag	ccc	agg	gac	tgc	gtc	tct	tgc	cgg	1826
	Pro	Glu	Gly	Cys	Trp	Gly	Pro	Glu	Pro	Arg	Asp	Cys	Val	Ser	Cys	Arg	
				515					520					525			
20	aat	gtc	agc	cga	ggc	agg	gaa	tgc	gtg	gac	aag	tgc	aac	ctt	ctg	gag	1874
	Asn	Val	Ser	Arg	Gly	Arg	Glu	Cys	Val	Asp	Lys	Cys	Asn	Leu	Leu	Glu	
			530					535					540				
	ggg	gag	cca	agg	gag	ttt	gtg	gag	aac	tct	gag	tgc	ata	cag	tgc	cac	1922
25	Gly	Glu	Pro	Arg	Glu	Phe	Val	Glu	Asn	Ser	Glu	Cys	Ile	Gln	Cys	His	
		545					550					555					
	cca	gag	tgc	ctg	cct	cag	gcc	atg	aac	atc	acc	tgc	aca	gga	cgg	gga	1970
	Pro	Glu	Cys	Leu	Pro	Gln	Ala	Met	Asn	Ile	Thr	Cys	Thr	Gly	Arg	Gly	
	560					565					570				575		
30	cca	gac	aac	tgt	atc	cag	tgt	gcc	cac	tac	att	gac	ggc	ccc	cac	tgc	2018
	Pro	Asp	Asn	Cys	Ile	Gln	Cys	Ala	His	Tyr	Ile	Asp	Gly	Pro	His	Cys	
					580					585					590		
	gtc	aag	acc	tgc	ccg	gca	gga	gtc	atg	gga	gaa	aac	aac	acc	ctg	gtc	2066
35	Val	Lys	Thr	Cys	Pro	Ala	Gly	Val	Met	Gly	Glu	Asn	Asn	Thr	Leu	Val	
				595					600					605			
	tgg	aag	tac	gca	gac	gcc	ggc	cat	gtg	tgc	cac	ctg	tgc	cat	cca	aac	2114
	Trp	Lys	Tyr	Ala	Asp	Ala	Gly	His	Val	Cys	His	Leu	Cys	His	Pro	Asn	
			610					615					620				
40	tgc	acc	tac	gga	tgc	act	ggg	cca	ggg	ctt	gaa	ggc	tgt	cca	acg	aat	2162
	Cys	Thr	Tyr	Gly	Cys	Thr	Gly	Pro	Gly	Leu	Glu	Gly	Cys	Pro	Thr	Asn	
		625					630					635					
	ggg	cct	aag	atc	ccg	tcc	atc	gcc	act	ggg	atg	gtg	ggg	gcc	ctc	ctc	2210
45	Gly	Pro	Lys	Ile	Pro	Ser	Ile	Ala	Thr	Gly	Met	Val	Gly	Ala	Leu	Leu	
	640					645					650				655		
	ttg	ctg	ctg	gtg	gtg	gcc	ctg	ggg	atc	ggc	ctc	ttc	atg	cga	agg	cgc	2258
	Leu	Leu	Leu	Val	Val	Ala	Leu	Gly	Ile	Gly	Leu	Phe	Met	Arg	Arg	Arg	
					660					665					670		
	cac	atc	gtt	cgg	aag	cgc	acg	ctg	cgg	agg	ctg	ctg	cag	gag	agg	gag	2306
50	His	Ile	Val	Arg	Lys	Arg	Thr	Leu	Arg	Arg	Leu	Leu	Gln	Glu	Arg	Glu	
				675					680					685			
	ctt	gtg	gag	cct	ctt	aca	ccc	agt	gga	gaa	gct	ccc	aac	caa	gct	ctc	2354
	Leu	Val	Glu	Pro	Leu	Thr	Pro	Ser	Gly	Glu	Ala	Pro	Asn	Gln	Ala	Leu	
			690					695					700				
55	ttg	agg	atc	ttg	aag	gaa	act	gaa	ttc	aaa	aag	atc	aaa	gtg	ctg	ggc	2402
	Leu	Arg	Ile	Leu	Lys	Glu	Thr	Glu	Phe	Lys	Lys	Ile	Lys	Val	Leu	Gly	

	705					710					715						
5	tcc Ser 720	ggt Gly	gcg Ala	ttc Phe	ggc Gly	acg Thr 725	gtg Val	tat Tyr	aag Lys	gga Gly	ctc Leu 730	tgg Trp	atc Ile	cca Pro	gaa Glu	ggt Gly 735	2450
	gag Glu	aaa Lys	gtt Val	aaa Lys	att Ile 740	ccc Pro	gtc Val	gct Ala	atc Ile	aag Lys 745	gaa Glu	tta Leu	aga Arg	gaa Glu	gca Ala 750	aca Thr	2498
10	tct Ser	ccg Pro	aaa Lys	gcc Ala 755	aac Asn	aag Lys	gaa Glu	atc Ile	ctc Leu 760	gat Asp	gaa Glu	gcc Ala	tac Tyr	gtg Val 765	atg Met	gcc Ala	2546
	agc Ser	gtg Val	gac Asp 770	aac Asn	ccc Pro	cac His	gtg Val	tgc Cys 775	cgc Arg	ctg Leu	ctg Leu	ggc Gly	atc Ile 780	tgc Cys	ctc Leu	acc Thr	2594
15	tcc Ser	acc Thr 785	gtg Val	cag Gln	ctc Leu	atc Ile	acg Thr 790	cag Gln	ctc Leu	atg Met	ccc Pro	ttc Phe 795	ggc Gly	tgc Cys	ctc Leu	ctg Leu	2642
	gac Asp 800	tat Tyr	gtc Val	cgg Arg	gaa Glu	cac His 805	aaa Lys	gac Asp	aat Asn	att Ile	ggc Gly 810	tcc Ser	cag Gln	tac Tyr	ctg Leu	ctc Leu 815	2690
20	aac Asn	tgg Trp	tgt Cys	gtg Val	cag Gln 820	atc Ile	gca Ala	aag Lys	ggc Gly	atg Met 825	aac Asn	tac Tyr	ttg Leu	gag Glu	gac Asp 830	cgt Arg	2738
	cgc Arg	ttg Leu	gtg Val	cac His 835	cgc Arg	gac Asp	ctg Leu	gca Ala	gcc Ala 840	agg Arg	aac Asn	gta Val	ctg Leu	gtg Val 845	aaa Lys	aca Thr	2786
30	ccg Pro	cag Gln	cat His 850	gtc Val	aag Lys	atc Ile	aca Thr	gat Asp 855	ttt Phe	ggg Gly	ctg Leu	gcc Ala	aaa Lys 860	ctg Leu	ctg Leu	ggt Gly	2834
	gcg Ala	gaa Glu 865	gag Glu	aaa Lys	gaa Glu	tac Tyr	cat His 870	gca Ala	gaa Glu	gga Gly	ggc Gly	aaa Lys 875	gtg Val	cct Pro	atc Ile	aag Lys	2882
35	tgg Trp 880	atg Met	gca Ala	ttg Leu	gaa Glu	tca Ser 885	att Ile	tta Leu	cac His	aga Arg	atc Ile 890	tat Tyr	acc Thr	cac His	cag Gln	agt Ser 895	2930
	gat Asp	gtc Val	tgg Trp	agc Ser	tac Tyr 900	ggg Gly	gtg Val	act Thr	gtt Val	tgg Trp 905	gag Glu	ttg Leu	atg Met	acc Thr	ttt Phe 910	gga Gly	2978
40	tcc Ser	aag Lys	cca Pro	tat Tyr 915	gac Asp	gga Gly	atc Ile	cct Pro	gcc Ala 920	agc Ser	gag Glu	atc Ile	tcc Ser	tcc Ser 925	atc Ile	ctg Leu	3026
	gag Glu	aaa Lys	gga Gly 930	gaa Glu	cgc Arg	ctc Leu	cct Pro	cag Gln 935	cca Pro	ccc Pro	ata Ile	tgt Cys	acc Thr 940	atc Ile	gat Asp	gtc Val	3074
50	tac Tyr	atg Met 945	atc Ile	atg Met	gtc Val	aag Lys	tgc Cys 950	tgg Trp	atg Met	ata Ile	gac Asp	gca Ala 955	gat Asp	agt Ser	cgc Arg	cca Pro	3122
	aag Lys 960	ttc Phe	cgt Arg	gag Glu	ttg Leu	atc Ile 965	atc Ile	gaa Glu	ttc Phe	tcc Ser	aaa Lys 970	atg Met	gcc Ala	cga Arg	gac Asp	ccc Pro 975	3170
55	cag Gln	cgc Arg	tac Tyr	ctt Leu	gtc Val 980	att Ile	cag Gln	ggg Gly	gat Asp	gaa Glu 985	aga Arg	atg Met	cat His	ttg Leu	cca Pro 990	agt Ser	3218

	cct	aca	gac	tcc	aac	ttc	tac	cgt	gcc	ctg	atg	gat	gaa	gaa	gac	atg	3266
	Pro	Thr	Asp	Ser	Asn	Phe	Tyr	Arg	Ala	Leu	Met	Asp	Glu	Glu	Asp	Met	
				995					1000					1005			
5	gac	gac	gtg	gtg	gat	gcc	gac	gag	tac	ctc	atc	cca	cag	cag	ggc	ttc	3314
	Asp	Asp	Val	Val	Asp	Ala	Asp	Glu	Tyr	Leu	Ile	Pro	Gln	Gln	Gly	Phe	
			1010					1015					1020				
10	ttc	agc	agc	ccc	tcc	acg	tca	cgg	act	ccc	ctc	ctg	agc	tct	ctg	agt	3362
	Phe	Ser	Ser	Pro	Ser	Thr	Ser	Arg	Thr	Pro	Leu	Leu	Ser	Ser	Leu	Ser	
		1025					1030					1035					
	gca	acc	agc	aac	aat	tcc	acc	gtg	gct	tgc	att	gat	aga	aat	ggg	ctg	3410
	Ala	Thr	Ser	Asn	Asn	Ser	Thr	Val	Ala	Cys	Ile	Asp	Arg	Asn	Gly	Leu	
	1040					1045					1050					1055	
15	caa	agc	tgt	ccc	atc	aag	gaa	gac	agc	ttc	ttg	cag	cga	tac	agc	tca	3458
	Gln	Ser	Cys	Pro	Ile	Lys	Glu	Asp	Ser	Phe	Leu	Gln	Arg	Tyr	Ser	Ser	
				1060						1065					1070		
20	gac	ccc	aca	ggc	gcc	ttg	act	gag	gac	agc	ata	gac	gac	acc	ttc	ctc	3506
	Asp	Pro	Thr	Gly	Ala	Leu	Thr	Glu	Asp	Ser	Ile	Asp	Asp	Thr	Phe	Leu	
				1075					1080					1085			
	cca	gtg	cct	gaa	tac	ata	aac	cag	tcc	ggt	ccc	aaa	agg	ccc	gct	ggc	3554
	Pro	Val	Pro	Glu	Tyr	Ile	Asn	Gln	Ser	Val	Pro	Lys	Arg	Pro	Ala	Gly	
			1090					1095					1100				
25	tct	gtg	cag	aat	cct	gtc	tat	cac	aat	cag	cct	ctg	aac	ccc	gcg	ccc	3602
	Ser	Val	Gln	Asn	Pro	Val	Tyr	His	Asn	Gln	Pro	Leu	Asn	Pro	Ala	Pro	
		1105					1110					1115					
30	agc	aga	gac	cca	cac	tac	cag	gac	ccc	cac	agc	act	gca	gtg	ggc	aac	3650
	Ser	Arg	Asp	Pro	His	Tyr	Gln	Asp	Pro	His	Ser	Thr	Ala	Val	Gly	Asn	
	1120					1125					1130					1135	
	ccc	gag	tat	ctc	aac	act	gtc	cag	ccc	acc	tgt	gtc	aac	agc	aca	ttc	3698
	Pro	Glu	Tyr	Leu	Asn	Thr	Val	Gln	Pro	Thr	Cys	Val	Asn	Ser	Thr	Phe	
				1140						1145					1150		
35	gac	agc	cct	gcc	cac	tgg	gcc	cag	aaa	ggc	agc	cac	caa	att	agc	ctg	3746
	Asp	Ser	Pro	Ala	His	Trp	Ala	Gln	Lys	Gly	Ser	His	Gln	Ile	Ser	Leu	
				1155					1160					1165			
40	gac	aac	cct	gac	tac	cag	cag	gac	ttc	ttt	ccc	aag	gaa	gcc	aag	cca	3794
	Asp	Asn	Pro	Asp	Tyr	Gln	Gln	Asp	Phe	Phe	Pro	Lys	Glu	Ala	Lys	Pro	
			1170					1175					1180				
	aat	ggc	atc	ttt	aag	ggc	tcc	aca	gct	gaa	aat	gca	gaa	tac	cta	agg	3842
	Asn	Gly	Ile	Phe	Lys	Gly	Ser	Thr	Ala	Glu	Asn	Ala	Glu	Tyr	Leu	Arg	
		1185					1190					1195					
45	gtc	gcg	cca	caa	agc	agt	gaa	ttt	att	gga	gca	tga					3878
	Val	Ala	Pro	Gln	Ser	Ser	Glu	Phe	Ile	Gly	Ala						
	1200					1205					1210						
50	<210> 512																
	<211> 1210																
	<212> PRT																
	<213> Homo sapiens																
55	<400> 512																

	Met	Arg	Pro	Ser	Gly	Thr	Ala	Gly	Ala	Ala	Leu	Leu	Ala	Leu	Leu	Ala
	1				5					10					15	
5	Ala	Leu	Cys	Pro	Ala	Ser	Arg	Ala	Leu	Glu	Glu	Lys	Lys	Val	Cys	Gln
				20					25					30		
10																
15																
20																
25																
30																
35																
40																
45																
50																
55																

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

	Ile	Thr	Gly	Phe	Leu 405	Leu	Ile	Gln	Ala	Trp 410	Pro	Glu	Asn	Arg	Thr 415	Asp
5	Leu	His	Ala	Phe 420	Glu	Asn	Leu	Glu	Ile 425	Ile	Arg	Gly	Arg	Thr 430	Lys	Gln
	His	Gly	Gln	Phe 435	Ser	Leu	Ala	Val 440	Val	Ser	Leu	Asn	Ile 445	Thr	Ser	Leu
10	Gly	Leu 450	Arg	Ser	Leu	Lys	Glu 455	Ile	Ser	Asp	Gly	Asp 460	Val	Ile	Ile	Ser
	Gly 465	Asn	Lys	Asn	Leu	Cys 470	Tyr	Ala	Asn	Thr	Ile 475	Asn	Trp	Lys	Lys	Leu 480
15	Phe	Gly	Thr	Ser	Gly 485	Gln	Lys	Thr	Lys	Ile 490	Ile	Ser	Asn	Arg	Gly 495	Glu
	Asn	Ser	Cys	Lys 500	Ala	Thr	Gly	Gln	Val 505	Cys	His	Ala	Leu	Cys 510	Ser	Pro
20	Glu	Gly	Cys 515	Trp	Gly	Pro	Glu	Pro 520	Arg	Asp	Cys	Val	Ser 525	Cys	Arg	Asn
	Val	Ser 530	Arg	Gly	Arg	Glu	Cys 535	Val	Asp	Lys	Cys	Asn 540	Leu	Leu	Glu	Gly
25	Glu 545	Pro	Arg	Glu	Phe 550	Val	Glu	Asn	Ser	Glu	Cys 555	Ile	Gln	Cys	His	Pro 560
	Glu	Cys	Leu	Pro	Gln 565	Ala	Met	Asn	Ile	Thr 570	Cys	Thr	Gly	Arg	Gly 575	Pro
30	Asp	Asn	Cys	Ile 580	Gln	Cys	Ala	His	Tyr 585	Ile	Asp	Gly	Pro	His 590	Cys	Val
	Lys	Thr	Cys 595	Pro	Ala	Gly	Val	Met 600	Gly	Glu	Asn	Asn	Thr 605	Leu	Val	Trp
35	Lys	Tyr 610	Ala	Asp	Ala	Gly	His 615	Val	Cys	His	Leu	Cys 620	His	Pro	Asn	Cys
	Thr 625	Tyr	Gly	Cys	Thr	Gly 630	Pro	Gly	Leu	Glu	Gly 635	Cys	Pro	Thr	Asn	Gly 640
40	Pro	Lys	Ile	Pro	Ser 645	Ile	Ala	Thr	Gly	Met 650	Val	Gly	Ala	Leu	Leu 655	Leu
	Leu	Leu	Val	Val 660	Ala	Leu	Gly	Ile	Gly 665	Leu	Phe	Met	Arg	Arg 670	Arg	His
45	Ile	Val	Arg 675	Lys	Arg	Thr	Leu	Arg 680	Arg	Leu	Leu	Gln	Glu 685	Arg	Glu	Leu
	Val	Glu 690	Pro	Leu	Thr	Pro	Ser 695	Gly	Glu	Ala	Pro	Asn 700	Gln	Ala	Leu	Leu
50	Arg 705	Ile	Leu	Lys	Glu	Thr 710	Glu	Phe	Lys	Lys	Ile 715	Lys	Val	Leu	Gly	Ser 720
	Gly	Ala	Phe	Gly	Thr 725	Val	Tyr	Lys	Gly	Leu 730	Trp	Ile	Pro	Glu	Gly 735	Glu
55	Lys	Val	Lys	Ile 740	Pro	Val	Ala	Ile	Lys 745	Glu	Leu	Arg	Glu	Ala 750	Thr	Ser
	Pro	Lys	Ala 755	Asn	Lys	Glu	Ile	Leu 760	Asp	Glu	Ala	Tyr	Val 765	Met	Ala	Ser

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

	Glu Tyr Leu Asn Thr Val Gln Pro Thr Cys Val Asn Ser Thr Phe Asp	
	1140 1145 1150	
5	Ser Pro Ala His Trp Ala Gln Lys Gly Ser His Gln Ile Ser Leu Asp	
	1155 1160 1165	
	Asn Pro Asp Tyr Gln Gln Asp Phe Phe Pro Lys Glu Ala Lys Pro Asn	
	1170 1175 1180	
10	Gly Ile Phe Lys Gly Ser Thr Ala Glu Asn Ala Glu Tyr Leu Arg Val	
	1185 1190 1195 1200	
	Ala Pro Gln Ser Ser Glu Phe Ile Gly Ala	
	1205 1210	
15	<210> 513	
	<211> 25	
	<212> DNA	
	<213> Homo sapiens	
20	<400> 513	
	cagatttggc tcgacctgga catag 25	
25	<210> 514	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
30	<400> 514	
	cagctgatct caaggaaaca gg 22	
35	<210> 515	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
40	<400> 515	
	gtattatcag tcactaaagc tcac 24	
45	<210> 516	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 516	
	cacactcaa gtggaattct gc 22	
55	<210> 517	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
60	<400> 517	
	tgcatagggt tcaactgg 19	
65	<210> 518	
	<211> 26	
	<212> DNA	
	<213> Homo sapiens	

	<400> 518	
	ccttctccga ggtggaattg agtgac	26
5	<210> 519	
	<211> 26	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 519	
	gctaattgcg ggactcttgt tcgcac	26
	<210> 520	
	<211> 22	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 520	
	tacatgcttt tctagtggtc ag	22
20	<210> 521	
	<211> 25	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 521	
	ggtctcaagt gattctacaa accag	25
	<210> 522	
30	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 522	
35	ccttcaccta ctggttcaca tctg	24
	<210> 523	
	<211> 26	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 523	
	catggtttga ctagtttga atgtgg	26
45	<210> 524	
	<211> 27	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 524	
	ggatactaaa gatactttgt caccagg	27
	<210> 525	
	<211> 26	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 525	
	gaacactagg ctgcaaagac agtaac	26

<210> 526
 <211> 24
 <212> DNA
 <213> Homo sapiens
 5
 <400> 526
 ccaagcaagg caaacacatc cacc 24
 10
 <210> 527
 <211> 23
 <212> DNA
 <213> Homo sapiens
 15
 <400> 527
 ggaggatgga gcctttccat cac 23
 20
 <210> 528
 <211> 24
 <212> DNA
 <213> Homo sapiens
 <400> 528
 gaagaggaag atgtgttcct ttgg 24
 25
 <210> 529
 <211> 24
 <212> DNA
 <213> Homo sapiens
 30
 <400> 529
 gaatgaagga tgatgtggca gtgg 24
 35
 <210> 530
 <211> 21
 <212> DNA
 <213> Homo sapiens
 <400> 530
 caaaacatca gccattaacg g 21
 40
 <210> 531
 <211> 26
 <212> DNA
 <213> Homo sapiens
 45
 <400> 531
 ccacttactg tcatataat acagag 26
 50
 <210> 532
 <211> 25
 <212> DNA
 <213> Homo sapiens
 55
 <400> 532
 catgtgagat agcatttggg aatgc 25
 <210> 533
 <211> 26

<212> DNA
 <213> Homo sapiens

5 <400> 533
 catgacctac catcattgga aagcag 26

10 <210> 534
 <211> 21
 <212> DNA
 <213> Homo sapiens

15 <400> 534
 gtaatttcac agttaggaat c 21

20 <210> 535
 <211> 25
 <212> DNA
 <213> Homo sapiens

25 <400> 535
 gtcaccaag gtcattggagc acagg 25

30 <210> 536
 <211> 23
 <212> DNA
 <213> Homo sapiens

35 <400> 536
 cagaatgcct gtaaagctat aac 23

40 <210> 537
 <211> 24
 <212> DNA
 <213> Homo sapiens

45 <400> 537
 gtcctggagt cccaactcct tgac 24

50 <210> 538
 <211> 25
 <212> DNA
 <213> Homo sapiens

55 <400> 538
 ggaagtggct ctgatggccg tcctg 25

60 <210> 539
 <211> 27
 <212> DNA
 <213> Homo sapiens

65 <400> 539
 ccactcacac aactaaata tttaag 27

70 <210> 540
 <211> 27
 <212> DNA
 <213> Homo sapiens

	<400> 540 gaccaaaca ccttaagtaa ctgactc	27
5	<210> 541 <211> 24 <212> DNA <213> Homo sapiens	
10	<400> 541 ccaatccaac atccagacac atag	24
15	<210> 542 <211> 24 <212> DNA <213> Homo sapiens	
20	<400> 542 ccagagccat agaaacttga tcag	24
25	<210> 543 <211> 29 <212> DNA <213> Homo sapiens	
30	<400> 543 gtatggacta tggcacttca attgcatgg	29
35	<210> 544 <211> 28 <212> DNA <213> Homo sapiens	
40	<400> 544 ccagagaaca tggcaaccag cacaggac	28
45	<210> 545 <211> 25 <212> DNA <213> Homo sapiens	
50	<400> 545 caaatgagct ggcaagtgcc gtgtc	25
55	<210> 546 <211> 25 <212> DNA <213> Homo sapiens	
	<400> 546 gagtttcca aacactcagt gaaac	25
	<210> 547 <211> 25 <212> DNA <213> Homo sapiens	
	<400> 547 gcaatatcag ccttaggtgc ggctc	25

<210> 548
 <211> 26
 <212> DNA
 <213> Homo sapiens
 5
 <400> 548
 catagaaagt gaacatttag gatgtg 26
 10
 <210> 549
 <211> 24
 <212> DNA
 <213> Homo sapiens
 15
 <400> 549
 ccatgagtac gtatittgaa actc 24
 20
 <210> 550
 <211> 24
 <212> DNA
 <213> Homo sapiens
 25
 <400> 550
 catatcccca tggcaaactc ttgc 24
 30
 <210> 551
 <211> 24
 <212> DNA
 <213> Homo sapiens
 35
 <400> 551
 ctaacgttcg ccagccataa gtcc 24
 40
 <210> 552
 <211> 26
 <212> DNA
 <213> Homo sapiens
 45
 <400> 552
 gctgcgagct caccagaat gtctgg 26
 50
 <210> 553
 <211> 25
 <212> DNA
 <213> Homo sapiens
 55
 <400> 553
 gacgggtcct ggggtgatct ggctc 25
 60
 <210> 554
 <211> 3866
 <212> DNA
 <213> Homo sapiens
 65
 <400> 554

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgaggcggc	cggaggtccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgccctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gccccggcag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tggaattttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
10	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagagggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tgtctgccc	caaattcccg	gacgagcgat	1020
	cgtgcaaggga	cccttgcccc	ccactcatgc	tgtacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
20	atgtggtgac	agatcacggc	tcgtgctgcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggccctga	aaacaggacg	gacctccatg	1500
25	ccitttgagaa	cctagaaatc	atacgggcca	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttccagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgtcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgctgtctt	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
30	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gcatgaaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggatggtggg	ggccctcctc	ttgtgtctgg	2220
35	tggttggcct	ggggatcggc	ctcttcattg	gaaggcgcca	catcggttcg	aagcgcacgc	2280
	tgctggaggc	gctgcaggag	agggagcttg	tggagcctct	tacaccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtg	gttcgggcag	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaaccatctc	cgaaagccaa	caaggaaatc	ctcgatgaag	2520
	cctacgtgat	ggccagcgtg	gacaaccccc	acgtgtgccg	cctgctgggc	atctgcctca	2580
40	cctccaccgt	gcagctcatc	acgcagctca	tgcccttcgg	ctgcctcctg	gactatgtcc	2640
	gggaacacaa	agacaatatt	ggctcccagt	acctgctcaa	ctgggtgtgtg	cagatcgcaa	2700
	agggcatgaa	ctacttgagg	gaccgtcgct	tggtgcaccg	cgacctggca	gccaggaacg	2760
	tactggtgaa	aacaccgcag	catgtcaaga	tcacagattt	tgggctggcc	aaactgctgg	2820
	gtgcggaaga	gaaagaatac	catgcagaag	gaggcaaagt	gcctatcaag	tggtatggcat	2880
	tggaaatcaat	tttacacaga	atctataccc	accagagtga	tgtctggagc	tacggggtga	2940
45	ctgtttggga	gttgatgacc	tttggatcca	agccatatga	cggaatccct	gccagcgaga	3000
	tctcctccat	cctggagaaa	ggagaacgcc	tccctcagcc	acccatatgt	accatcgatg	3060
	tctacatgat	catggtcaag	tgctggatga	tagacgcaga	tagtcgcccc	aagttccgtg	3120
	agttgatcat	cgaattctcc	aaaatggccc	gagaccccca	gcgctacctt	gtcattcagg	3180
	gggatgaaag	aatgcatttg	ccaagtccca	cagactccaa	cttctaccgt	gccctgatgg	3240
	atgaagaaga	catggacgac	gtgggtgatg	ccgacgagta	cctcatccca	cagcagggtc	3300
50	tcctcagcag	ccccctccacg	tcacggactc	ccctcctgag	ctctctgagt	gcaaccagca	3360
	acaattccac	cgtggcttgc	attgatagaa	atgggctgca	aagctgtccc	atcaaggaag	3420
	acagcttctt	gcagcgatac	agctcagacc	ccacaggcgc	cttgactgag	gacagcatag	3480
	acgacacctt	cctcccagtg	cctgaatata	taaaccagtc	cgttcccaaa	aggcccgtg	3540
	gcctctgtga	gaatcctgtc	tatcacaaat	agcctctgaa	ccccgcgccc	agcagagacc	3600
	cacactacca	ggacccccac	agcaactgcag	tgggcaaccc	cgagtatctc	aacactgtcc	3660
	agcccacctg	tgtcaacagc	acattcgaca	gccctgcccc	ctgggcccag	aaaggcagcc	3720
55	accaaattag	cctggacaac	cctgactacc	agcaggactt	ctttcccaag	gaagccaagc	3780
	caaattggcat	ctttaagggc	tccacagctg	aaaatgcaga	atacctaagg	gtcgcgccac	3840
	aaagcagtga	atttatttga	gcata				3866

<210> 555
 <211> 24
 <212> DNA
 <213> Homo sapiens
 5
 <400> 555
 caggactaca gaaatgtagg tttc 24
 10
 <210> 556
 <211> 26
 <212> DNA
 <213> Homo sapiens
 15
 <400> 556
 gtcctgcct taagtaatgt gatgac 26
 20
 <210> 557
 <211> 25
 <212> DNA
 <213> Homo sapiens
 25
 <400> 557
 gactggaagt gtcgcatcac caatg 25
 30
 <210> 558
 <211> 25
 <212> DNA
 <213> Homo sapiens
 35
 <400> 558
 ggtttaataa tgcatctgg gacac 25
 40
 <210> 559
 <211> 25
 <212> DNA
 <213> Homo sapiens
 45
 <400> 559
 gcagctataa ttagagaac caagg 25
 50
 <210> 560
 <211> 21
 <212> DNA
 <213> Homo sapiens
 55
 <400> 560
 aaaattgact tcatttccat g 21
 60
 <210> 561
 <211> 22
 <212> DNA
 <213> Homo sapiens
 65
 <400> 561
 cctagttgct ctaaaactaa cg 22
 70
 <210> 562
 <211> 24

<212> DNA
 <213> Homo sapiens

5 <400> 562
 ctgtgaggcg tgacagccgt gcag 24

10 <210> 563
 <211> 25
 <212> DNA
 <213> Homo sapiens

15 <400> 563
 caacctacta atcagaacca gcatc 25

20 <210> 564
 <211> 24
 <212> DNA
 <213> Homo sapiens

25 <400> 564
 ccttcactgt gtctgcaaat ctgc 24

30 <210> 565
 <211> 24
 <212> DNA
 <213> Homo sapiens

35 <400> 565
 cctgtcataa gtctccttgt tgag 24

40 <210> 566
 <211> 25
 <212> DNA
 <213> Homo sapiens

45 <400> 566
 cagtctgtgg gtctaagagc taatg 25

50 <210> 567
 <211> 24
 <212> DNA
 <213> Homo sapiens

55 <400> 567
 caggaatggg tgagtctctg tgtg 24

60 <210> 568
 <211> 23
 <212> DNA
 <213> Homo sapiens

65 <400> 568
 gtggaattct gccaggcct ttc 23

70 <210> 569
 <211> 25
 <212> DNA
 <213> Homo sapiens

	<400> 569	
	gattctacaa accagccagc caaac	25
5	<210> 570	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 570	
	cctactggtt cacatctgac cctg	24
	<210> 571	
	<211> 24	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 571	
	gtttgaatgt ggttcgtg gaag	24
20	<210> 572	
	<211> 27	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 572	
	cttgtcacc aggcagaggg caatatc	27
	<210> 573	
	<211> 23	
30	<212> DNA	
	<213> Homo sapiens	
	<400> 573	
	gacagtaact tgggctttct gac	23
35	<210> 574	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
40	<400> 574	
	catccacca aagactctcc aag	23
	<210> 575	
45	<211> 25	
	<212> DNA	
	<213> Homo sapiens	
	<400> 575	
50	ctgttcatat aatacagagt ccctg	25
	<210> 576	
	<211> 22	
	<212> DNA	
55	<213> Homo sapiens	
	<400> 576	
	gagagatgca ggagctctgt gc	22

<210> 577
 <211> 25
 <212> DNA
 <213> Homo sapiens
 5
 <400> 577
 gcagtttgta gtcaatcaaa ggtgg 25
 <210> 578
 <211> 22
 <212> DNA
 <213> Homo sapiens
 10
 <400> 578
 gtaatttaaa tgggaatagc cc 22
 <210> 579
 <211> 24
 <212> DNA
 <213> Homo sapiens
 20
 <400> 579
 caactccttg accattacct caag 24
 <210> 580
 <211> 23
 <212> DNA
 <213> Homo sapiens
 25
 <400> 580
 gatggccgtc ctgccacac agg 23
 <210> 581
 <211> 21
 <212> DNA
 <213> Homo sapiens
 30
 <400> 581
 gagtagttta gcatatattg c 21
 <210> 582
 <211> 23
 <212> DNA
 <213> Homo sapiens
 35
 <400> 582
 gacagtcaga aatgcaggaa age 23
 <210> 583
 <211> 26
 <212> DNA
 <213> Homo sapiens
 40
 <400> 583
 caagtgccgt gtcctggcac ccaagc 26
 <210> 584
 <211> 24

<212> DNA
 <213> Homo sapiens

5 <400> 584
 ccaaacactc agtgaacaa agag 24

10 <210> 585
 <211> 21
 <212> DNA
 <213> Homo sapiens

15 <400> 585
 ccttaggtgc ggctccacag c 21

20 <210> 586
 <211> 22
 <212> DNA
 <213> Homo sapiens

25 <400> 586
 catttaggat gtggagatga gc 22

30 <210> 587
 <211> 23
 <212> DNA
 <213> Homo sapiens

35 <400> 587
 gaaactcaag atcgattca tgc 23

40 <210> 588
 <211> 23
 <212> DNA
 <213> Homo sapiens

45 <400> 588
 gcaaactctt gctatcccag gag 23

50 <210> 589
 <211> 22
 <212> DNA
 <213> Homo sapiens

55 <400> 589
 cagccataag tcctcgacgt gg 22

60 <210> 590
 <211> 24
 <212> DNA
 <213> Homo sapiens

65 <400> 590
 catcctcccc tgcattgttt aaac 24

70 <210> 591
 <211> 24
 <212> DNA
 <213> Homo sapiens

	<400> 591	
	gtaggtttct aaacatcaag aaac	24
5	<210> 592	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 592	
	gtgatgacat ttctccaggg atgc	24
	<210> 593	
	<211> 23	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 593	
	catcaccaat gccttcttta age	23
20	<210> 594	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 594	
	gctggagggt ttaataatgc gate	24
	<210> 595	
30	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 595	
35	gcaaacacac aggcacctgc tggc	24
	<210> 596	
	<211> 21	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 596	
	catgtgagtt tcactagatg g	21
	<210> 597	
45	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
	<400> 597	
50	gaccggacga caggccacct cgtc	24
	<210> 598	
	<211> 25	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 598	
	gaagaacgaa acgtcccgtt cctcc	25

<210> 599
 <211> 23
 <212> DNA
 <213> Homo sapiens
 5
 <400> 599
 gttgagcact cgtgtgcatt agg 23
 <210> 600
 10 <211> 22
 <212> DNA
 <213> Homo sapiens
 <400> 600
 15 ctcagtgcac gtgtactggg ta 22
 <210> 601
 <211> 35
 <212> DNA
 20 <213> Homo sapiens
 <400> 601
 gttcactggg ctaattgcgg gactcttgtt cgcac 35
 25 <210> 602
 <211> 28
 <212> DNA
 <213> Homo sapiens
 30 <400> 602
 ggtaaataca tgcttttcta gtggtcag 28
 <210> 603
 <211> 23
 35 <212> DNA
 <213> Homo sapiens
 <400> 603
 ggaggatgga gcctttccat cac 23
 40 <210> 604
 <211> 24
 <212> DNA
 <213> Homo sapiens
 45 <400> 604
 gaagaggaag atgtgttcct ttgg 24
 <210> 605
 50 <211> 24
 <212> DNA
 <213> Homo sapiens
 <400> 605
 55 gaatgaagga tgatgtggca gtgg 24
 <210> 606
 <211> 25

<212> DNA
 <213> Homo sapiens

5 <400> 606
 gtatgtgtga aggagtcact gaaac 25

10 <210> 607
 <211> 22
 <212> DNA
 <213> Homo sapiens

15 <400> 607
 ggtgagtcac aggttcagtt gc 22

20 <210> 608
 <211> 21
 <212> DNA
 <213> Homo sapiens

25 <400> 608
 caaaacatca gccattaacg g 21

30 <210> 609
 <211> 22
 <212> DNA
 <213> Homo sapiens

35 <400> 609
 gtagccagca tgtctgtgtc ac 22

40 <210> 610
 <211> 23
 <212> DNA
 <213> Homo sapiens

45 <400> 610
 cagaatgcct gtaaagctat aac 23

50 <210> 611
 <211> 23
 <212> DNA
 <213> Homo sapiens

55 <400> 611
 catttgctt tccccactca cac 23

60 <210> 612
 <211> 27
 <212> DNA
 <213> Homo sapiens

65 <400> 612
 gacaaaaaca ccttaagtaa ctgactc 27

70 <210> 613
 <211> 25
 <212> DNA
 <213> Homo sapiens

	<400> 613	
	gaagctacat agtgtctcac ttcc	25
5	<210> 614	
	<211> 25	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 614	
	cacaactgct aatggcccgt tctcg	25
	<210> 615	
	<211> 23	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 615	
	gctcctgctc cctgtcataa gtc	23
20	<210> 616	
	<211> 25	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 616	
	gaagtctctgc tggtagtcag ggttg	25
	<210> 617	
	<211> 24	
30	<212> DNA	
	<213> Homo sapiens	
	<400> 617	
35	ctgcagtggg caaccccgag tadc	24
	<210> 618	
	<211> 20	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 618	
	tgtgggtcta agagctaag	20
	<210> 619	
45	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 619	
50	gacaggccac ctcgtcggcg tc	22
	<210> 620	
	<211> 22	
	<212> DNA	
55	<213> Homo sapiens	
	<400> 620	
	cagctgatct caaggaaaca gg	22

<210> 621
 <211> 25
 <212> DNA
 <213> Homo sapiens
 5
 <400> 621
 ctcggtgca ttaggttca actgg 25
 10
 <210> 622
 <211> 26
 <212> DNA
 <213> Homo sapiens
 15
 <400> 622
 ccttctccga ggtggaattg agtgac 26
 20
 <210> 623
 <211> 26
 <212> DNA
 <213> Homo sapiens
 <400> 623
 gctaattgcg ggactctgt tcgcac 26
 25
 <210> 624
 <211> 22
 <212> DNA
 <213> Homo sapiens
 30
 <400> 624
 tacatgcttt tctagtggc ag 22
 35
 <210> 625
 <211> 23
 <212> DNA
 <213> Homo sapiens
 <400> 625
 cctttccatc acccctcaag agg 23
 40
 <210> 626
 <211> 25
 <212> DNA
 <213> Homo sapiens
 45
 <400> 626
 gatgtgtcc ttggaggtg gcatg 25
 50
 <210> 627
 <211> 24
 <212> DNA
 <213> Homo sapiens
 55
 <400> 627
 gatgtggcag tggcggtcc ggtg 24
 <210> 628
 <211> 24

<212> DNA
 <213> Homo sapiens

5 <400> 628
 ggagtcactg aaacaacaa cagg 24

10 <210> 629
 <211> 21
 <212> DNA
 <213> Homo sapiens

15 <400> 629
 ggttcagttg ctgtataaa g 21

20 <210> 630
 <211> 24
 <212> DNA
 <213> Homo sapiens

25 <400> 630
 ccattaacgg taaaattca gaag 24

30 <210> 631
 <211> 20
 <212> DNA
 <213> Homo sapiens

35 <400> 631
 ccaaggcat ggagcacagg 20

40 <210> 632
 <211> 25
 <212> DNA
 <213> Homo sapiens

45 <400> 632
 ctgtaaagct ataacaacaa cctgg 25

50 <210> 633
 <211> 27
 <212> DNA
 <213> Homo sapiens

55 <400> 633
 ccactcacac acactaaata tttaag 27

60 <210> 634
 <211> 24
 <212> DNA
 <213> Homo sapiens

65 <400> 634
 gtaactgact caaatacaaa ccac 24

70 <210> 635
 <211> 25
 <212> DNA
 <213> Homo sapiens

```

<400> 635
gaagctacat agtgtctcac ttcc      25

<210> 636
<211> 20
<212> DNA
<213> Homo sapiens

<400> 636
ctgctaattgg cccgttctcg      20

<210> 637
<211> 24
<212> DNA
<213> Homo sapiens

<400> 637
cctgtcataa gtctccttgt tgag      24

<210> 638
<211> 20
<212> DNA
<213> Homo sapiens

<400> 638
ggtagtcagg gttgtccagg      20

<210> 639
<211> 23
<212> DNA
<213> Homo sapiens

<400> 639
cgagtatctc aacactgtcc age      23

<210> 640
<211> 25
<212> DNA
<213> Homo sapiens

<400> 640
ctaagagcta atgcgggcat ggctg      25

<210> 641
<211> 372
<212> DNA
<213> Homo sapiens

<400> 641

gcaatatcag ccttaggtgc ggctccacag cccagtgct cctcaccttc ggggtgcatc 60
gctggtaaca tccaccaga tcactgggca gcatgtggca ccatctcaca attgccagtt 120
aacgtcttcc ttctctctct gtcataaggga ctctggatcc cagaaggtga gaaagttaaa 180
attcccgtcg ctatcaagga attaagagaa gcaacatctc cgaaagccaa caaggaaatc 240
ctcgatgtga gtttctgctt tgctgtgtgg ggggtccatgg ctctgaacct caggcccacc 300
ttttctcatg tctggcagct gctctgctct agaccctgct catctccaca tcctaaatgt 360
tcactttcta tg      372

```

<210> 642
 <211> 415
 <212> DNA
 <213> Homo sapiens
 5
 <400> 642
 10
 ctaacgttcg ccagccataa gtcctcgacg tggagaggct cagagcctgg catgaacatg 60
 accctgaatt cggatgcaga gcttcttccc atgatgatct gtccctcaca gcaggggtctt 120
 ctctgtttca gggcatgaac tacttggagg accgtcgctt ggtgcaccgc gacctggcag 180
 ccaggaacgt actggtgaaa acaccgcagc atgtcaagat cacagatttt gggctggcca 240
 aactgctggg tgcggaagag aaagaatacc atgcagaagg aggcaaagta aggagggtggc 300
 tttagggtcag ccagcatttt cctgacacca gggaccaggc tgccttccca ctagctgtat 360
 tgtttaacac atgcagggga ggatgctctc cagacattct gggtgagctc gcagc 415
 15
 <210> 643
 <211> 35
 <212> DNA
 <213> Homo sapiens
 20
 <400> 643
 tatcaaggaa ttaagagaag caacatctcc gaaag 35
 25
 <210> 644
 <211> 24
 <212> DNA
 <213> Homo sapiens
 30
 <400> 644
 cgaaagccaa caaggaaatc ctgc 24
 35
 <210> 645
 <211> 18
 <212> DNA
 <213> Homo sapiens
 40
 <400> 645
 tgtaaacga cgccagt 18
 45
 <210> 646
 <211> 22
 <212> DNA
 <213> Homo sapiens
 50
 <400> 646
 tggctcaca ggaccactga tt 22
 55
 <210> 647
 <211> 22
 <212> DNA
 <213> Homo sapiens
 <400> 647
 gaggccagtg ctgtctctaa gg 22
 <210> 648
 <211> 21
 <212> DNA

<213> Homo sapiens
 <400> 648
 atgggacagg cactgatttg t 21
 5
 <210> 649
 <211> 22
 <212> DNA
 <213> Homo sapiens
 10
 <400> 649
 cagctctggc tcacactacc ag 22
 15
 <210> 650
 <211> 20
 <212> DNA
 <213> Homo sapiens
 20
 <400> 650
 gcagctggac tcgatttcct 20
 25
 <210> 651
 <211> 22
 <212> DNA
 <213> Homo sapiens
 30
 <400> 651
 tgcccaatga gtcaagaagt gt 22
 35
 <210> 652
 <211> 22
 <212> DNA
 <213> Homo sapiens
 40
 <400> 652
 cactcacgga tgctgcttag tt 22
 45
 <210> 653
 <211> 22
 <212> DNA
 <213> Homo sapiens
 50
 <400> 653
 tcagagcctg tgtttctacc aa 22
 55
 <210> 654
 <211> 24
 <212> DNA
 <213> Homo sapiens
 60
 <400> 654
 aaataatcag tgtgattcgt ggag 24
 65
 <210> 655
 <211> 21
 <212> DNA
 <213> Homo sapiens

	<400> 655	
	acttcacagc cctgcgtaaa c	21
5	<210> 656	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 656	
	gcagcggggt acatcttctt tc	22
	<210> 657	
	<211> 22	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 657	
	cctgaactcc gtcagactga aa	22
20	<210> 658	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 658	
	ccttacagca atcctgtgaa aca	23
	<210> 659	
30	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
	<400> 659	
35	atgtacagtg ctggcatggt ct	22
	<210> 660	
	<211> 20	
	<212> DNA	
40	<213> Homo sapiens	
	<400> 660	
	tccaaatgag ctggcaagtg	20
45	<210> 661	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
50	<400> 661	
	gtgcatcgct ggtaacatcc	20
	<210> 662	
	<211> 20	
55	<212> DNA	
	<213> Homo sapiens	
	<400> 662	
	atcgcatcca tgcgtcttca	20

	<210> 663	
	<211> 19	
	<212> DNA	
	<213> Homo sapiens	
5		
	<400> 663	
	gctcagagcc tggcatgaa	19
10		
	<210> 664	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
15		
	<400> 664	
	tggctcgtct gtgtgtgtca	20
20		
	<210> 665	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
25		
	<400> 665	
	tgaagcaaatt tgccaagac	20
30		
	<210> 666	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
35		
	<400> 666	
	aagtgtcgca tcaccaatgc	20
40		
	<210> 667	
	<211> 23	
	<212> DNA	
	<213> Homo sapiens	
45		
	<400> 667	
	tcccaaacac tcagtgaac aaa	23
50		
	<210> 668	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
55		
	<400> 668	
	tgtggagatg agcagggtct	20
60		
	<210> 669	
	<211> 20	
	<212> DNA	
	<213> Homo sapiens	
65		
	<400> 669	
	atccccatgg caaactcttg	20
70		
	<210> 670	
	<211> 19	

<212> DNA
 <213> Homo sapiens

5 <400> 670
 catcctcccc tgcattgt 19

10 <210> 671
 <211> 25
 <212> DNA
 <213> Homo sapiens

15 <400> 671
 cgaaagaaaa tacttgcattg tcaga 25

20 <210> 672
 <211> 20
 <212> DNA
 <213> Homo sapiens

25 <400> 672
 tgacatttct ccagggatgc 20

30 <210> 673
 <211> 19
 <212> DNA
 <213> Homo sapiens

35 <400> 673
 atgcgatctg ggacacagg 19

40 <210> 674
 <211> 16
 <212> DNA
 <213> Homo sapiens

45 <400> 674
 aacagctatg accatg 16

50 <210> 675
 <211> 26
 <212> DNA
 <213> Homo sapiens

55 <400> 675
 caagtgccgt gtctggcac ccaagc 26

60 <210> 676
 <211> 24
 <212> DNA
 <213> Homo sapiens

65 <400> 676
 ccaaactc agtgaacaa agag 24

70 <210> 677
 <211> 25
 <212> DNA
 <213> Homo sapiens

	<400> 677	
	gcaccaagc ccatgccgtg gctgc	25
5	<210> 678	
	<211> 26	
	<212> DNA	
	<213> Homo sapiens	
10	<400> 678	
	gaaacaaaga gtaaagtaga tgatgg	26
	<210> 679	
	<211> 24	
15	<212> DNA	
	<213> Homo sapiens	
	<400> 679	
	caccttcaca atataccctc catg	24
20	<210> 680	
	<211> 22	
	<212> DNA	
	<213> Homo sapiens	
25	<400> 680	
	gacagccgtg cagggaaaaa cc	22
	<210> 681	
	<211> 24	
30	<212> DNA	
	<213> Homo sapiens	
	<400> 681	
	gaaccagcat ctcaaggaga tctc	24
35	<210> 682	
	<211> 24	
	<212> DNA	
	<213> Homo sapiens	
40	<400> 682	
	gagcacctgg ctggacact ggag	24
	<210> 683	
45	<211> 26	
	<212> DNA	
	<213> Homo sapiens	
	<400> 683	
50	gagcagccct gaactccgtc agactg	26
	<210> 684	
	<211> 26	
	<212> DNA	
55	<213> Homo sapiens	
	<400> 684	
	ctcagtacaa tagatagaca gcaatg	26

<210> 685
 <211> 25
 <212> DNA
 <213> Homo sapiens
 5
 <400> 685
 gacgggtcct ggggtgatct ggctc 25
 10
 <210> 686
 <211> 26
 <212> DNA
 <213> Homo sapiens
 15
 <400> 686
 ctcagtacaa tagatagaca gcaatg 26
 20
 <210> 687
 <211> 26
 <212> DNA
 <213> Homo sapiens
 <400> 687
 cgtggagagg ctcagagcct ggcatg 26
 25
 <210> 688
 <211> 21
 <212> DNA
 <213> Homo sapiens
 30
 <400> 688
 cacagatttt gggcggggcca a 21
 35
 <210> 689
 <211> 21
 <212> DNA
 <213> Homo sapiens
 <400> 689
 gctatcaaaa catctccgaa a 21
 40
 <210> 690
 <211> 35
 <212> DNA
 <213> Homo sapiens
 45
 <400> 690
 tatcaaggaa ttaagagaag caacatctcc gaaag 35
 50
 <210> 691
 <211> 29
 <212> DNA
 <213> Homo sapiens
 55
 <400> 691
 agttaaaatt cccgtcgcta tcaaggaat 29
 <210> 692
 <211> 35

<212> DNA

<213> Homo sapiens

<400> 692

5 tatcaaggaa ttaagagaag caacatctcc gaaag 35

<210> 693

<211> 29

<212> DNA

10 <213> Homo sapiens

<400> 693

aacatctccg aaagccaaca aggaaatcc 29

15 <210> 694

<211> 35

<212> DNA

<213> Homo sapiens

20 <400> 694

tatcaaggaa ttaagagaag caacatctcc gaaag 35

<210> 695

<211> 21

25 <212> DNA

<213> Homo sapiens

<400> 695

taaaattccc gtcgctatca a 21

30

<210> 696

<211> 35

<212> DNA

<213> Homo sapiens

35

<400> 696

tatcaaggaa ttaagagaag caacatctcc gaaag 35

<210> 697

<211> 25

40 <212> DNA

<213> Homo sapiens

<400> 697

45 ccatctccga aagccaacaa ggaaa 25

<210> 698

<211> 35

<212> DNA

50 <213> Homo sapiens

<400> 698

tatcaaggaa ttaagagaag caacatctcc gaaag 35

55

<210> 699

<211> 23

<212> DNA

<213> Homo sapiens

<400> 699
 aattcccgtc gctatcaagg aac 23

5
 <210> 700
 <211> 15
 <212> DNA
 <213> Homo sapiens

10
 <400> 700
 gggctggcca aactg 15

15
 <210> 701
 <211> 15
 <212> DNA
 <213> Homo sapiens

20
 <400> 701
 gggcgggcca aactg 15

25
 <210> 702
 <211> 15
 <212> DNA
 <213> Homo sapiens

30
 <400> 702
 gggctggcca aactg 15

35
 <210> 703
 <211> 15
 <212> DNA
 <213> Homo sapiens

40
 <400> 703
 gggctggcca aacag 15

45
 <210> 704
 <211> 27
 <212> DNA
 <213> Homo sapiens

50
 <400> 704
 gatcaaagtg ctgggctccg gtgcgtt 27

55
 <210> 705
 <211> 27
 <212> DNA
 <213> Homo sapiens

60
 <220>
 <221> modified_base
 <222> (14)
 <223> a, t, c, g, other or unknown

65
 <400> 705
 gatcaaagtg ctgngctccg gtgcgtt 27

70
 <210> 706
 <211> 31

<212> DNA
 <213> Homo sapiens

5 <400> 706
 tcacagattt tgggctggcc aaactgctgg g 31

10 <210> 707
 <211> 31
 <212> DNA
 <213> Homo sapiens

15 <220>
 <221> modified_base
 <222> (16)
 <223> a, t, c, g, other or unknown

20 <400> 707
 tcacagattt tgggcnggcc aaactgctgg g 31

25 <210> 708
 <211> 21
 <212> DNA
 <213> Homo sapiens

30 <400> 708
 tatcaaggaa ttaagagaag c 21

35 <210> 709
 <211> 21
 <212> DNA
 <213> Homo sapiens

40 <220>
 <221> modified_base
 <222> (17)
 <223> a, t, c, g, other or unknown

45 <400> 709
 tatcaaggaa ttaaganaag c 21

50 <210> 710
 <211> 22
 <212> DNA
 <213> Homo sapiens

55 <400> 710
 tatcaaggaa ttaagagaag ca 22

60 <210> 711
 <211> 27
 <212> DNA
 <213> Homo sapiens

65 <400> 711
 cgtcgctatc aaggaattaa gagaagc 27

<210> 712
 <211> 20

<212> DNA
 <213> Homo sapiens

5 <400> 712
 cgctatcaag gaacaacatc 20

10 <210> 713
 <211> 19
 <212> DNA
 <213> Homo sapiens

15 <400> 713
 sgatcaygga ascaacatc 19

20 <210> 714
 <211> 28
 <212> DNA
 <213> Homo sapiens

25 <400> 714
 aggaattaag agaagcaaca tctccgaa 28

30 <210> 715
 <211> 28
 <212> DNA
 <213> Homo sapiens

35 <400> 715
 aggaattaaa agaaaccaca tcttcat 28

40 <210> 716
 <211> 28
 <212> DNA
 <213> Homo sapiens

45 <400> 716
 aggaaccaag agaagccacg tattcgaa 28

50 <210> 717
 <211> 37
 <212> DNA
 <213> Homo sapiens

55 <400> 717
 tatcaaggaa ttaagagaag caacatctcc gaaagcc 37

60 <210> 718
 <211> 19
 <212> DNA
 <213> Homo sapiens

65 <400> 718
 tatcaagggt ccgaaagcc 19

70 <210> 719
 <211> 19
 <212> DNA
 <213> Homo sapiens

<400> 719

tatcaagggtt ccgaaagcc 19

<210> 720

<211> 3867

<212> DNA

<213> Homo sapiens

<400> 720

```

ccccggcgag cgcgggccga gcagcctccg ccccccgcac ggtgtgagcg cccgacgcgg 60
ccgaggcggc cggagtcctc agctagcccc ggcgggccgc gccgcccaga ccggacgaca 120
ggccacctcg tcggcgctcc cccgagtcct cgctcgccg ccaacgccac aaccaccgcg 180
cacggcccc tgaactccgtc cagtattgat cgggagagcc ggagcgagct cttcggggag 240
cagcgatgcg accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgctct 300
gccccggcag tcgggctctg gaggaagaaga aagtttgcca aggcacgagt aacaagctca 360
cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
aggtggtcct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
taaagaccat ccaggaggtg gctgggttatg tcctcattgc cctcaacaca gtggagcgaa 540
ttcctttgga aaacctgcag atcatcagag gaaatatgta ctacgaaaat tcctatgcct 600
tagcagcttt atctaactat gatgcaaata aaaccggact gaaggagctg cccatgagaa 660
atttacagga aatcctgcat ggcgccgtgc gggttcagcaa caaccctgcc ctgtgcaacg 720
tgagagcat ccagtggcgg gacatagtca gcagtgactt tctcagcaac atgtcgatgg 780
acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
gctgggggtgc aggagaggag aactgccaga aactgaccaa aatcatctgt gcccagcagt 900

```

5 gctccgggcg ctgccgtggc aagtccccc gtgactgctg ccacaaccag tgtgctgcag 960
 gctgcacagg ccccccgggag agcgactgcc tggctgccc caaattccga gacgaagcca 1020
 cgtgcaagga cacctgcccc ccactcatgc tctacaaccc caccacgtac cagatggatg 1080
 tgaaccccgga gggcaaatat agcttttggtg ccacctgctg gaagaagtgt ccccgtaatt 1140
 atgtggtgac agatcacggc tcgtgctgctc gagcctgtgg ggccgacagc tatgagatgg 1200
 aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgcaaa gtgtgtaacg 1260
 gaatagggtat tgggtgaattt aaagactcac tctccataaa tgctacgaat attaaacact 1320
 tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 10 aggaaatcac aggggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
 cctttgagaa cctagaaatc atacgcggca ggaccaagca acatggtcag ttttctcttg 1560
 cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
 gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
 aactgtttgg gacctccggt cagaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaaggccac aggccaggctc tgccatgcct tgtgctcccc cgagggctgc tggggcccg 1800
 15 agcccaggga ctgctctctc tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
 gcaaccttct ggagggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
 acccagagtg cctgcctcag gccatgaaca taccctgcac aggacgggga ccagacaact 1980
 gtatccagtg tgcccactac attgacggcc cccactgctg caagacctgc ccggcaggag 2040
 tcatgggaga aaacaacacc ctggtctgga agtacgcaga cgccggccat gtgtgccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctgggcccagg tcttgaaggc tgtccaacga 2160
 atgggcctaa gatccccgtcc atcgccactg ggatggtggg ggccctcctc ttgctgctgg 2220
 20 tggtgccctt ggggatcggc ctcttcctgc gaaggcgcca catcgctcgg catcgctcgg 2280
 tgcggaggct gctgcaggag agggagcttg tggagcctct tacaccaggt ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtctg 2400
 gctccggtgc gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
 aaattcccggt cgctatcaag gaatacatct ccgaaagcca acaaggaaat cctcgatgaa 2520
 25 gcctacgtga tggccagcgt ggacaacccc cacgtgtgcc gcctgctggg catctgcctc 2580
 acctccaccg tgcagctcat caccagctc gctgcctcct gctgcctcct ggactatgtc 2640
 cgggaacaca aagacaatat tggctcccag tacctgctca actggtgtgt gcagatcgca 2700
 aagggcatga actacttggg ggaccgtcgc ttggtgcacc gcgacctggc agccaggaac 2760
 gtactggtga aaacaccgca gcatgtcaag atcacagatt ttgggctggc caaactgctg 2820
 ggtgcggaag agaaagaata ccatgcagaa ggaggcaaag tgcctatcaa gtggatggca 2880
 30 ttggaatcaa ttttacacag aatctatacc caccagagt atgtctggag ctacggggtg 2940
 actgtttggg agttgatgac ctttggatcc aagccatatg acggaatccc tgccagcgag 3000
 atctcctcca tcctggagaa aggagaacgc ctccctcagc caccatattg taccatcgat 3060
 gtctacatga tcatggtcaa gtgctggatg atagacgcag atagtgcgcc aaagttccgt 3120
 gagttgatca tcgaattctc caaaatggcc agcgtacct agcgctacct tgtcattcag 3180
 ggggatgaaa gaatgcattt gccaagtcct acagactcca acttctaccg tgccctgatg 3240
 35 gatgaagaag acatggacga cgtggtggat gccgacgagt acctcatccc acagcagggc 3300
 ttcttcagca gccccccac gtcacggact cccctcctga gctctctgag tgcaaccagc 3360
 aacaattcca ccgtggcttg cattgataga aatgggctgc aaagctgtcc catcaaggaa 3420
 gacagcttct tgcagcgata cagctcagac cccacaggcg ccttgactga ggacagcata 3480
 gacgacacct tcctcccagt gcctgaatac ataaaccagt ccgttcccaa aaggcccgtc 3540
 ggctctgtgc agaatcctgt ctatcacaa cagcctctga accccgcgcc cagcagagac 3600
 40 ccacactacc aggacccccca cagcactgca gtgggcaacc ccgagtatct caacactgtc 3660
 cagcccacct gtgtcaacag cacattcgac agccctgccc actgggcca gaaaggcagc 3720
 caccaaatta gcctggacaa ccctgactac cagcaggact tctttcccaa ggaagccaag 3780
 ccaaattggca tctttaaggg ctccacagct gaaaatgcag aatacctaag ggtcgcgcca 3840
 caaagcagtg aatttatgg agcatga 3867

45 <210> 721
 <211> 3863
 <212> DNA
 <213> Homo sapiens

50 <400> 721

55

	cccgggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggagtgccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctccg	cccaggtccc	cgctctcgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgctc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
5	cagcgatcg	accctccggg	acggccgggg	cagcgctcct	aatcatctgt	gctgcgctct	300
	gccccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaaat	tcctatgcct	600
10							
	tagcagtcctt	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
15	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgagg	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagg	tgtgatagtg	1200
20	aggaagacgg	cgctccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaatc	atacgcggca	ggaccaagca	acatggctcag	ttttctcttg	1560
25	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agccccagga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaacccttct	ggaggggtgag	ccaaggaggt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
30	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatggtggg	ggccctctct	tgtctgtgg	2220
	tgggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
35	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tgaaatcctc	gatgaagcct	2520
	acgtgatggc	cagcgtggac	aacccccacg	tgtgcgcctc	gctgggcatac	tgccctacct	2580
	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcggctg	cctcctggac	tatgtccggg	2640
40	aacacaaaga	caatattggc	tcccagtacc	tgctcaactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttgaggagc	cgctcgcttg	tgcaccgcga	cctggcagcc	aggaacgtac	2760
	tggtgaaaac	accgcagcat	gtcaagatca	cagatttttg	gctggccaaa	ctgctgggtg	2820
	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcattgg	2880
	aatcaatttt	acacagaatc	tatacccacc	agagtgtatg	ctggagctac	ggggtgactg	2940
	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
45	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	ggtcaagtgc	tggatgatag	acgcagatag	tcgcccacaa	ttccgtgagt	3120
	tgatcatcga	attctccaaa	atggcccag	acccccagcg	ctaccttgct	attcaggggg	3180
	atgaaagaat	gcatttgcca	agtcctacag	actccaactt	ctaccgtgcc	ctgatggatg	3240
	aagaagacat	ggacgacgtg	gtggatgccg	acgagtacct	catcccacag	cagggcttct	3300
	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgca	accagcaaca	3360
50	attccaccgt	ggcttgcat	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
	gcttcttgca	gcgatacagc	tcagacccca	caggcgctt	gactgaggac	agcatagacg	3480
	acaccttctc	cccagtgcct	gaatacataa	accagtcctg	tcccacaaag	cccgtggct	3540
	ctgtgcagaa	tcctgtctat	cacaatcagc	ctctgaaccc	cgcgcccagc	agagaccac	3600
	actaccagga	ccccacagc	actgcagtgg	gcaaccccg	gtatctcaac	actgtccagc	3660
	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
55	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccacaaag	gccaagccaa	3780
	atggcatctt	taagggtctc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
	gcagtgaaatt	tattggagca	tga				3863

<210> 722
 <211> 3886
 <212> DNA
 <213> Homo sapiens

5

<400> 722

10

```

cccgggcgcag cgcgggccgca gcagcctccg ccccccgcac ggtgtgagcg cccgacgcgg 60
ccgagggcggc cggagtcccc agctagcccc ggcggccgcc gccgcccaga ccggacgaca 120
ggccacctcg tcggcgctccg cccgagtccc cgcctcgccg ccaacgccac aaccaccgcg 180
cacggcccc tgactccgtc cagtattgat cgggagagcc ggagcgagct cttcggggag 240
cagcgatgcg accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgctct 300
gcccggcgag tcgggctctg gaggaaaaga aagtttgcca aggcacgagt aacaagctca 360

```

15

20

25

30

35

40

45

50

55

	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
5	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
	tagcagtcit	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccattgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagagggg	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
10	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgagg	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaaggga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccgga	gggcaaatac	agcttttggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
15	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaatagggtat	tggtagaattt	aaagactcac	tctccataaa	tgctacgaat	atataaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccittgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
20	cagtgcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgctctct	tgccggaatg	tcagccgagg	caggggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
25	acccagagtg	cctgcctcag	gcatgaaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggatggtggg	ggccctcctc	ttgtgctg	2220
	tgggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgacgc	2280
30	tgcgagggtc	gctgcaggag	agggagcttg	tgagcctct	tacaccag	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtg	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggaacccc	ggcaaccccc	acgtgtgccg	2580
	cctgctgggc	atctgcctca	cctccaccgt	gcactcatc	acgcagctca	tgcccttcgg	2640
35	ctgcctcctg	gactatgtcc	gggaacacaa	agacaatatt	ggctcccagt	acctgtctca	2700
	ctggtgtgtg	cagatcgcaa	agggcatgaa	ctacttgag	gaccgtcgct	tgggtgcccg	2760
	cgacctggca	gccagggaacg	tactggtgaa	aacaccgcag	catgtcaaga	tcacagattt	2820
	tgggctggcc	aaactgctgg	gtgcggaaga	gaaagaatac	catgcagaag	gaggcaaagt	2880
	gcctatcaag	tggatggcat	tggaaatcaat	tttacacaga	atctataccc	accagagtga	2940
40	tgtctggagc	tacgggggtga	ctgtttggga	gttgatgacc	tttggatcca	agccatatga	3000
	cggaatccct	gccagcgaga	tctcctccat	cctggagaaa	ggagaacgcc	tccctcagcc	3060
	acccatatgt	accatcgatg	tctacatgat	catggtcaag	tgctggatga	tagacgcaga	3120
	tagtcgcccc	aagttccgtg	agttgatcat	cgaattctcc	aaaatggccc	gagaccccc	3180
	cgctacacct	gtcattcagg	gggatgaaag	aatgcatttg	ccaagtccta	cagactccaa	3240
	cttctaccgt	gccctgatgg	atgaagaaga	catggacgac	gtggtggatg	ccgacgagta	3300
	cctcatccca	cagcagggct	tcttcagcag	ccccctccag	tcacggactc	ccctcctgag	3360
45	ctctctgagt	gcaaccagca	acaattccac	cgtaggcttg	attgatagaa	atgggctgca	3420
	aagctgtccc	atcaagggaag	acagcttctt	gcagcgatac	agctcagacc	ccacaggcgc	3480
	cttgactgag	gacagcatag	acgacacctt	cctcccagtg	cctgaatata	taaaccagtc	3540
	cgttcccaaa	aggcccgtg	gtctgtgca	gaatcctgtc	tatcacaatc	agcctctgaa	3600
	ccccgcgccc	agcagagacc	cacactacca	ggacccccac	agcactgcag	tgggcaaccc	3660
	cgagtatctc	aacactgtcc	agccccactg	tgtcaacagc	acattcgaca	gccctgcccc	3720
	ctgggcccag	aaaggcagcc	accaaattag	cctggacaac	cctgactacc	agcaggactt	3780
50	ctttcccaag	gaagccaagc	caaattggcat	ctttaagggc	tccacagctg	aaaatgcaga	3840
	atacctaagg	gtcgcgccac	aaagcagtga	atttatttga	gcatga		3886

<210> 723

<211> 3887

55 <212> DNA

<213> Homo sapiens

<400> 723

ccccggcgag cgcgggccgca gcagcctccg ccccccgcac ggtgtgagcg cccgacgcgg 60

5	ccgaggcggc	cggagtcccg	agctagcccc	ggcgggccgc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tggggtccg	cccaggtccc	cgctctcgcc	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgaggagagc	ggagcgaagt	cttcggggag	240
	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tggggtctg	gaggaaaaa	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtgggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
10	taaagaccat	ccaggagggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaat	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
15	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	ccccggggag	agcgactgcc	tggtctgccc	caaattccga	gacgagacca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccgga	gggcaaatat	agctttgggtg	ccacctgctg	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
20	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgggcca	ggaccaagca	acatggtcag	ttttctcttg	1560
25	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgtcccc	cgagggtgctg	tgggggcccg	1800
	agcccaggga	ctgctgtctt	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaagggtg	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
30	accaagagtg	cctgcctcag	gcatgaaaca	tcactgacac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgctg	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgaaagg	tgtccaacga	2160
	atgggcctaa	gatccccgtc	atcgccactg	ggatggtggg	ggccctcctc	ttgtgctggt	2220
	tgggtggcct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgctcgg	agcgcacgc	2280
35	tgccggagct	ctgcaggag	agggagcttg	tgagcctct	tacaccagtc	ggagagcctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaaag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacagcgt	ggacaacccc	cacgtgtgcc	2580
	gctgtgtggg	catctgcctc	acctccaccg	tgcagctcat	cacgcagctc	atgcccttcg	2640
40	gctgcctcct	ggactatgtc	cgggaaacaca	agacaatat	tggctcccag	tacctgtctc	2700
	actggtgtgt	gcagatcgca	aagggtcatga	actacttgga	ggaccgtcgc	ttggtgcacc	2760
	gcgacctggc	agccagggaac	gtactgggtga	aaacaccgca	gcatgtcaag	atcacagatt	2820
	ttgggctggc	caaactgctg	ggtgcggaag	agaaagaata	ccatgcagaa	ggaggcaaa	2880
	tgccatcaaa	gtggatggca	ttggaatcaa	ttttacacag	aatctatacc	caccagagtg	2940
45	atgtctggag	ctacgggggtg	actgtttggg	agttgatgac	ctttggatcc	aagccatattg	3000
	acggaatccc	tgccagcgag	atctcctcca	tcctggagaa	aggagaacgc	ctccctcagc	3060
	cacccatatg	taccatcgat	gtctacatga	tcattggtcaa	gtgctggatg	atagacgcag	3120
	atagtcgccc	aaagtccctg	gagttgatca	tcgaattctc	caaatggcc	cgagaccccc	3180
	agcgctacct	tgtcattcag	ggggatgaaa	gaatgcattt	gccaagtcc	acagactcca	3240
	acttctaccg	tgccctgatg	gatgaagaag	acatggacga	cgtggtggat	gccgacgagt	3300
50	acctcatccc	acagcagggc	ttcttcagca	gccccccac	gtcacggact	cccctcctga	3360
	gctctctgag	tgcaaccagc	aacaattcca	ccgtggcttg	cattgataga	aatgggctgc	3420
	aaagctgtcc	catcaaggaa	gacagcttct	tgacgcgata	cagctcagac	cccacaggcg	3480
	ccttgactga	ggacagcata	gacgacacct	tcctcccagt	gcctgaatac	ataaaccagt	3540
	ccgttcccaa	aaggcccgct	ggctctgtgc	agaatcctgt	ctatcacaat	cagcctctga	3600
	accccgcgcc	cagcagagac	ccacactacc	aggaccccca	cagcactgca	gtgggcaacc	3660
	ccgagtatct	caacactgtc	cagcccacct	gtgtcaacag	cacattcgac	agccctgccc	3720
55	actgggcccc	gaaaggcagc	caccaaatta	gcctggacaa	ccctgactac	cagcaggact	3780
	tctttcccaa	ggaagccaag	ccaaatggca	tctttaaggg	ctccacagct	gaaaatgcag	3840
	aatacctaag	ggtcgcgcca	caaagcagtg	aatttattgg	agcatga		3887

<210> 724
<211> 3881
<212> DNA
<213> Homo sapiens

5

<400> 724

10

15

20

25

30

35

40

45

50

55

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgctctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gagggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tggaattttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
10	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgcccgtgg	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tggtctggcg	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaaggt	gcgaagggcc	ttgccgcaa	gtgtgtaacg	1260
20	gaataggtat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
25	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgtat	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggg	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgtctccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgctctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaaggagtg	ttgtggagaa	ctctgagtgc	ataagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcactgacac	aggacgggga	ccagacaact	1980
30	gtatccagt	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctctct	ttgtgtctgg	2220
	tggtggccct	ggggatcggc	ctcttcattg	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
35	tgcgagggct	gctgcaggag	agggagcttg	tgagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtg	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggaacaacc	cggtcacgtg	tgccgctctg	2580
	tgggcatctg	cctcacctcc	accgtgcagc	tcatacgcga	gctcatgccc	ttcggctgcc	2640
	tcctggacta	tgtccgggaa	cacaaagaca	atattggctc	ccagtacctg	ctcaactggt	2700
40	gtgtgcagat	cgcaaaaggc	atgaactact	tggaggaccg	tcgcttggtg	caccgcgacc	2760
	tggcagccag	gaacgtactg	gtgaaaacac	cgagcatgt	caagatcaca	gattttgggc	2820
	tggccaaact	gctgggtgcg	gaagagaaa	aataccatgc	agaaggaggc	aaagtgccta	2880
	tcaattggat	ggcattggaa	tcaattttac	acagaatcta	taccaccag	agtgtgtct	2940
	ggagctacgg	ggtgactgtt	tgggagttga	tgaccttttg	atccaagcca	tatgacggaa	3000
45	tccctgccag	cgagatctcc	tccatcctgg	agaaaggaga	acgcctccct	cagccaccca	3060
	tatgtaccat	cgatgtctac	atgatcatgg	tcaagtgtctg	gatgatagac	gcagatagtc	3120
	gccc aaagt	ccgtgagttg	atcatcgaat	tctccaaaat	ggcccagagc	ccccagcgct	3180
	acctgtgtcat	tcagggggat	gaaagaatgc	atttgccaag	tcctacagac	tccaacttct	3240
	accgtgccct	gatggatgaa	gaagacatgg	acgacgtgg	ggatgccgac	gagtacctca	3300
	tcccacagca	gggttctctc	agcagccct	ccacgtcacg	gactccccct	ctgagctctc	3360
50	tgagtgcac	cagcaacaat	tccaccgtgg	cttgcatgga	tagaaatggg	ctgcaaagct	3420
	gtcccatcaa	ggaagacagc	ttcttgcagc	gatacagctc	agacccaca	ggcgcttga	3480
	ctgaggacag	catagacgac	accttctctc	cagtgcctga	atacataaac	cagtcctgtc	3540
	ccaaaaggcc	cgtggctctc	gtgcagaatc	ctgtctatca	caatcagcct	ctgaaccccg	3600
	cgcccagcag	agacccacac	taccaggacc	cccacagcac	tgcatgtggc	aaccccgagt	3660
	atctcaacac	tgtccagccc	acctgtgtca	acagcacatt	cgacagccct	gcccactggg	3720
	cccagaaagg	cagccacca	attagcctgg	acaaccctga	ctaccagcag	gacttctttc	3780
55	ccaaggaagc	caagccaaat	ggcatcttta	agggctccac	agctgaaaat	gcagaatacc	3840
	taagggctcg	gccacaaagc	agtgaattta	ttggagcatg	a		3881

<210> 725

<211> 3878
 <212> DNA
 <213> Homo sapiens

5 <400> 725

```

cccggcgag cgcggccgca gcagcctccg ccccccgcac ggtgtgagcg cccgacgagg 60
ccgaggcggc cggagtcccg agctagcccc ggcgcccgcc gccgcccaga ccggacgaca 120
ggccacctcg tcggcgctcg cccgagctccc cgcttcgccc ccaacgccac aaccaccgag 180
cacggccccc tgactccgtc cagtagtgat cgggagagcc ggagcgagct cttcggggag 240
cagcgatgag accctccggg acggccgggg cagcgctcct ggcgtgctg gctgcgctct 300
gccggcgag tcgggctctg gaggaagaaga aagtttgcca aggcacgagt aacaagctca 360
cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
aggtggctct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
taaagaccat ccaggagggtg gctgggttatg tcctcattgc cctcaacaca gtggagcgaa 540
ttcctttgga aaacctgcag atcatcagag gaaatatgta ctacgaaaaa tcctatgcct 600
tagcagctct atctaactat gatgcaataa aaaccggact gaaggagctg cccatgagaa 660
atttacagga aatcctgcat ggcgccgtgc ggttcagcaa caacctgcc ctgtgcaacg 720
tgagagagcat ccagtggcgg gacatagtcg gcagtgactt tctcagcaac atgtcgatgg 780
acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
gctgggggtg aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagct 900
gctccggggc ctgcccgtggc aagtccccc gtgactgctg ccacaaccag tgtgctgcag 960
gctgcacagg ccccgggag agcgactgcc tggctgccc gaaattccga gacgaagcca 1020
cgtgcaagga cacctgcccc ccactcatgc tctacaacct caccacgtac cagatggatg 1080
tgaacccga gggcaaatag agctttgggt ccacctgctg gaagaagtgt ccccgtaatt 1140
atgtgggtgac agatcacggc tcgtgcgtcc gagcctgtgg ggcgacagc tatgagatgg 1200
aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgaaa gtgtgtaacg 1260
gaataggatg tggtagaatt aaagactcac tccataaaa tgctacgaat attaaacact 1320
tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
aggaaatcac aggggtttttg ctgattcagg cttggcctga aaacaggagc gacctccatg 1500
cctttgagaa cctagaatac atacgcccga ggaccaagca acatggtcag ttttctcttg 1560
cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
aactgtttgg gacctccggt cagaaaacca aaattataag caacagaggt gaaaacagct 1740
gcaaggccac aggccagggtc tgccatgcct tgtgctcccc cgagggctgc tggggcccg 1800
agcccaggga ctgctctctc tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
gcaaccttct ggagggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
cccagagtgc cctgcctcag gcatgcaaca aggcaggga aggcaggga ccagacact 1980
gtatccagtg tgccactac attgacggcc cccactgctg caagacctgc ccggcaggag 2040
tcatgggaga aaacaacacc ctggtctgga agtacgcaga cgccggccat gtgtgccacc 2100
tgtgccatcc aaactgcacc tacggatgca ctgggcccag tcttgaggc tgtccaacga 2160
atgggcctaa gatcccgtcc atcgccactg ggatgggtgg ggccctctc ttgctgctgg 2220
tgggtggcct ggggatcggc ctcttcctgc gaaggcgcca catcgctcgg aagcgacgc 2280
tgccggaggc gctgcaggag agggagcttg tggagcctct tacaccagt caccagagctc 2340
ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtctg 2400
gctccgggtg gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
aaattcccgt cgctatcaag gaattaaagag aagcaacatc tccgaaagcc aacaaggaaa 2520
tcctcgatga agcctacgtg atggccagcg tggacaacct ccacgtgtgc cgcctgctaa 2580
gcatctgcct cacctccacc gtgcagctca tcacgcagct catgcccttc ggctgcctcc 2640
tggtactatg ccgggaacac aaagacaata ttggctccca gtacctgtc aactgggtgtg 2700
tgcagatcgc aaaggcatg aactacttgg aggaccgtcg cttggtgcac cgcgacctgg 2760
cagccaggaa cgtactgggtg aaaacaccgc agcatgtcaa gatcacagat tttgggctgg 2820
ccaaactgct ggggtgcggaa gagaaagaat accatgcaga aggaggcaaa gtgcctatca 2880
agtggatggc attggaatca attttacaca gaactatac ccaccagat gatgtctgga 2940
gctacggggg gactgttttg gagttgatga cctttggatc caagccatat gacggaatcc 3000
ctgccagcga gatctcctcc atcctggaga aaggagaacg cctccctcag ccacccatat 3060
gtaccatcga tgtctacatg atcatgggtc agtgctggat gatagacgca gatagtgcgc 3120
caaagttccg tgagttgatc atcgaattct ccaaaatggc cagagacccc cagcgctacc 3180
ttgtcattca gggggatgaa agaatgcatt tgccaagtcc tacagactcc aacttctacc 3240
gtgccttgat ggtgaagaa gacatggacg acgtgggtga tgccgacgag tacttctacc 3300
ccagcagggt cttcttcagc agcccccca cgtcacggac tcccctctg agctctctga 3360
gtgcaaccag caacaattcc accgtggctt gcattgatag aaatgggctg caaagctgtc 3420
ccatcaaggga agacagcttc ttgcagcgat acagctcaga cccacaggc gccttgactg 3480
aggacagcat agacgacacc ttctcccag tgctgaata cataaaccag tccgttccca 3540
aaaggcccg tcgtctgtg cagaatcctg tctatcacia tcagcctctg aacccgcgc 3600
ccagcagaga cccacactac caggaccccc acagcactgc agtgggcaac cccgagtatc 3660

```

tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactggggccc	3720
agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
aggaagccaa	gccaaatggc	atctttaagg	gctccacagc	tgaaaatgca	gaatacctaa	3840
gggtcgcgcc	acaaagcagt	gaatttattg	gagcatga			3878

5

<210> 726
 <211> 3878
 <212> DNA
 <213> Homo sapiens

10

<400> 726

15

20

25

30

35

40

45

50

55

	cccggcgag	cgcgccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgagggcggc	cggagtcctcg	agctagcccc	ggcgcccgcc	gccgcccaga	ccggacgaca	120
	ggccaccttcg	tcggcgctccg	cccagatccc	cgctcgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
5	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaaattg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	ttaaagaccat	ccaggaggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
10	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tggtctgccg	caaattccga	gacgaagcca	1020
	cgtgcaaggga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgatagtg	1200
	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgccgca	ggaccaagca	acatgggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
25	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgttttg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggtcgc	tggggcccgg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
30	gtatccagtg	tgcccaactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtggg	ggccctctc	ttgtgctg	2220
	tggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcgagggct	gctgcaggag	agggagcttg	tgagccctct	tacacccagt	ggagaagctc	2340
35	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataaag	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgatatcaag	gaattaaag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggaacacc	ccacgtgtgc	cgctgctg	2580
	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgctc	aactgggtgtg	2700
40	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgaccttg	2760
	cagccaggaa	cgtactggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgtgtg	2820
	ccaactgct	gggtgaggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	attggaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacggggg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccacccatat	3060
45	gaaccaatcga	tgtctacatg	atcatggtca	agtgctggat	gatagacgca	gatagctgcc	3120
	caaagtcccg	tgagttagatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agccccctca	cgtcacggac	tcccctcctg	agctctctga	3360
50	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
	ccatcaaggga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgctgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggctctgtg	cagaatcctg	tctatcacia	tcagcctctg	aaccccgcg	3600
	ccagcagaga	cccacactac	caggaccccc	acagactgc	agtgggcaac	cccgagtac	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactgggccc	3720
55	agaaaggcag	ccaccaaat	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gctccacagc	tgaatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaatttattg	gagcatga			3878

<210> 727
 <211> 3878
 <212> DNA
 <213> Homo sapiens

5

<400> 727

	cccggcgag	cgcgccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgaggcggc	cggagtcctc	agctagcccc	ggcgccgccc	gccgcccaga	ccggacgaca	120
10	ggccacctcg	tcggcgctcc	cccagatccc	cgccctcgcc	ccaacgccac	aaccaccgag	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tggaattttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
15	taaagaccat	ccaggagggt	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagctct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tgagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	actttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
20	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgagg	agcgactgcc	tggtctgccg	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaacccga	gggcaaatac	agctttgggt	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
25	aggagagcgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggagc	gacctccatg	1500
	cctttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatgggtcag	ttttctcttg	1560
30	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggtgct	tgggggcccg	1800
	agcccaggga	tgcggtctct	tgccggaatg	tcagccagg	cagggaatgc	tggtgacaat	1860
	gcaaccttct	ggagggtgag	ccaaggagg	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
35	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggccagg	tcttgaaggc	tgccaacga	2160
	atggggcctaa	gatcccgctc	atcgccactg	ggatgggtgg	ggccctcttc	ttgctgctgg	2220
	tggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcggttcg	aagcgacgc	2280
40	tgccggaggct	gctgcaggag	agggagcttg	tgagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaaag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggaacaacc	ccacgtgtgc	cgctgctg	2580
	gcactgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
45	tgactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtc	aactgggtgtg	2700
	tgagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgc	cgagacctgg	2760
	cagccaggaa	cgtactgtgtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggctgg	2820
	ccaaacagct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	atttgaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacggggt	gactgtttgg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
50	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccaccatat	3060

55

```

5      gtaccatcga tgtctacatg atcatgggtca agtgctggat gatagacgca gatagtcgcc 3120
      caaagttccg tgagttgatc atcgaattct ccaaaatggc ccgagacccc cagcgctacc 3180
      ttgtcattca gggggatgaa agaatgcatt tgccaagtcc tacagactcc aacttctacc 3240
      gtgccctgat ggatgaagaa gacatggacg acgtgggtga tgccgacgag tacctcatcc 3300
      cacagcaggg cttcttcagc agcccccca cgtcacggac tccccctctg agctctctga 3360
      gtgcaaccag caacaattcc accgtggctt gcattgatag aaatgggctg caaagctgtc 3420
      ccatcaagga agacagcttc ttgcagcgat acagctcaga cccacagggc gccttgactg 3480
      aggacagcat agacgacacc ttcttcccag tgcctgaata cataaaccag tccgttccca 3540
      aaaggcccg c tggctctgtg cagaatcctg tctatcaca tcagcctctg aaccccgcg 3600
      ccagcagaga cccacactac caggaccccc acagactgc agtgggcaac cccgagtatc 3660
10     tcaacactgt ccagcccacc tgtgtcaaca gcacattcga cagccctgcc cactgggccc 3720
      agaaaggcag ccaccaaatt agcctggaca accctgacta ccagcaggac ttctttccca 3780
      aggaagccaa gccaaatggc atctttaagg gctccacagc tgaaaatgca gaatacctaa 3840
      gggtcgcgcc acaaagcagt gaattttatt gagcatga 3878

```

```

15     <210> 728
      <211> 3878
      <212> DNA
      <213> Homo sapiens

```

```

20     <400> 728

```

```

25

```

```

30

```

```

35

```

```

40

```

```

45

```

```

50

```

```

55

```

	cccgggcgag	cgcgggccga	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcgg	cggaggtccc	agctagcccc	ggcgggccgc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtcc	cccaggtccc	cgctctcgcc	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gtgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtgggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggttat	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
10	tgcaggtctt	atctaactat	gatgcaataa	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtc	gcagtgaatt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatat	agctttgggt	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attnaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggttttt	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgggcca	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
25							
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagaggtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
30	gtatccagtg	tgccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgg	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctctct	ttgtgctgg	2220
	tgggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccaggt	ggagaagctc	2340
35	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtg	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tgacaacccc	ccacgtgtgc	cgcctgtctg	2580
	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtct	aactgggtgtg	2700
40	tgcagatcgc	aaagggcagt	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
45							
	cagccaggaa	cgtactgggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggctgg	2820
	ccaaactgct	gggtgcggaa	gagaaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	atcgggaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacggggg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccaccatat	3060
	gtaccatcga	tgtctacatg	atcatggtca	agtgtggat	gatagacgca	gatagtcgcc	3120
	caaagtctcc	tgagtgtgat	atcgaattct	ccaaaatggc	ccgagacccc	cagcgtacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	cgaagaatgc	tacagactcc	aacttctacc	3240
50	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agcccccca	cgtcacggac	tccccctctg	agctctctga	3360
	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aatgggctg	caaagctgtc	3420
	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgcctgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggctctgtg	cagaatcctg	tctatcacaa	tcagcctctg	aaccccgcg	3600
	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccagtatc	3660
55	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	actggggccc	3720
	agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gtccacagc	tgaatatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaatttatgt	gagcatga			3878

<210> 729
 <211> 3878
 <212> DNA
 <213> Homo sapiens

5

<400> 729

10

15

20

25

30

35

40

45

50

55

```

ccccggcgag cgcgggccga gcagcctccg cccccgcac ggtgtgagcg cccgacgcgg 60
ccgaggcggc cggagtcctcg agctagcccc ggcgggccgc gccgcccaga ccggacgaca 120
ggccacctcg tcggcgctccg cccgagtcctc cgctctcgccg ccaacgccac aaccaccgcg 180
cacggccccc tgactccgtc cagtattgat cgggagagcc ggagcgagct cttcggggag 240
cagcgatgcg accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgctct 300
gcccggcgag tcgggctctg gaggaagaaga aagtttgcca aggcacgagt aacaagctca 360
cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
aggtggctct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
taaagaccat ccaggaggtg gctgggtatg tcctcattgc cctcaacaca gtggagcgaa 540
ttccttttga aaacctgcag atcatcagag gaaatatgta ctacgaaaat tcctatgcct 600
tagcagtcct atctaactat gatgcaaata aaaccggact gaaggagctg cccatgagaa 660
atttacagga aatcctgcat ggcccgctgc ggttcagcaa caacctgcc ctgtgcaacg 720
tgagagcat ccagtggcgg gacatagtca gcagtgactt tctcagcaac atgtcgatgg 780
acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
gctggggtgc aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagct 900
gctccgggcg ctgccgtggc aagtccccca gtgactgctg ccacaaccag tgtgctgcag 960
gctgcacagg ccccggggag agcgactgcc tggctgccc ccaattccga gacgaagcca 1020
cgtgcaagga cacctgcccc ccactcatgc tctacaacc caccacgtac cagatggatg 1080
tgaaccccg gggcaaatat agctttgggt ccacctgcgt gaagaagtgt ccccgtaatt 1140
atgtggtgac agatcacggc tcgtgcgtcc gagcctgtgg ggccgacagc tatgagatgg 1200
aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgcaaa gtgtgtaacg 1260
gaataggtag tggtagaatt aaagactcac tctccataaa tgctacgaat attaaacact 1320
tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
aggaaatcac aggggttttt ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
cctttgagaa cctagaaatc atacgcggca ggaccaagca acatggtcag ttttctcttg 1560
cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
aactgtttgg gacctccggt cagaaaacca aaattataag caacagaggt gaaaacagct 1740
gcaaggccac aggccagggtc tgccatgcct tgtgctcccc cgagggctgc tggggcccg 1800
agcccaggga ctgctctctc tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
gcaaccttct ggagggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
acccagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
gtatccagtg tgccactac attgacggcc cccactgcgt caagacctgc ccggcaggag 2040
tcatgggaga aaactgcacc ctggtctgga agtacgcaga cgccggccat gtgtgccacc 2100
tgtgccatcc aaactgcacc tacggatgca ctgggcccagg tcttgaaggc tgtccaacga 2160
atgggcctaa gatcccgtcc atcgccactg ggaatggggg ggccctctct ttgctgctgg 2220
tggtggccct ggggatcggc ctcttcacgc gaaggcgcca catcgcttcg aagcgacgc 2280
tgcgagggct gctgcaggag agggagcttg tggagcctct tacaccagtg ggagaagctc 2340
ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtggt 2400
gctccggtgc gttcggcacg gtgtataagg gactctggat ccagaagggt gagaaagtta 2460

```

	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cgctgtctgg	2580
	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
5	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtc	aactgggtgtg	2700
	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
	cagccaggaa	cgtactgggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggctgg	2820
	ccaaactgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	attggaatca	attttacaca	gaatctatac	ccaccagagt	tatgtctgga	2940
	gctacggggg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccacccatat	3060
10	gtaccatcga	tgtctacatg	atcatgggtca	agtgtggat	gatagacgca	gatagtcgcc	3120
	caaagttccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agccccicca	cgtcacggac	tccccctctg	agctctctga	3360
	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
15	ccatcaaggga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgctgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggtctctgtg	cagaatcctg	tctatcacia	tcagcctctg	aaccccgccg	3600
	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccaggtatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactggggcc	3720
	agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
20	aggaagccaa	gccaatggc	atctttaagg	gctccacagc	tgaaaatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaattttattg	gagcatga			3878

<210> 730

<211> 3863

25 <212> DNA

<213> Homo sapiens

<400> 730

30

35

40

45

50

55

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgaggcggc	cggaggtccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgctctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gccccggcag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
10	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ttgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagagggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	cccccgggag	agcgactgcc	tgtctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	ccactgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccga	gggcaaatac	agctttggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaaactg	cacctccatc	agtgccgtgc	tccacatcct	gccgggtggca	tttggtgggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaaatc	atacgcggca	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgtatagc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgtcccc	cgagggtgc	tggggcccg	1800
	agcccaggga	ctgctgtctt	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaagggagt	ttgtggagaa	ctctgagtg	atacagtgcc	1920
30	acccagagtg	cctgcctcag	gcatggaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgctc	atcgccactg	ggatggtggg	ggccctcctc	ttgctgctgg	2220
35	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacacccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattccccgt	cgctatcaaa	acatctccga	aagccaacaa	ggaaatcctc	gatgaagcct	2520
40	acgtgatggc	cagcgtggag	aaacccacg	tgtgcgcct	gctgggcact	gctcctcct	2580
	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcggctg	cctcctggac	tatgtccggg	2640
	aacacaaaga	caatattggc	tcccagtacc	tgctcaactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttgaggagc	cgctcgcttg	tgcaccgcga	cctggcagcc	aggaacgtac	2760
	tggtgaaaac	accgcagcat	gtcaagatca	cagatttttg	gctggccaaa	ctgctgggtg	2820
	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcattgg	2880
45	aatcaatttt	acacagaatc	tatacccacc	agagtgatgt	ctggagctac	ggggtgactg	2940
	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	ggtcaagtgc	tggatgatag	acgcagatag	tcgcccaaa	ttccgtgagt	3120
	tgatcatcga	attctccaaa	atggccccgag	acccccagcg	ctaccttgct	attcaggggg	3180
	atgaaagaat	gcatttgcca	agtcctacag	actccaaact	ctaccgtgcc	ctatggatg	3240
50	aagaagacat	ggacgacgtg	gtggatgccg	acgagtacct	catcccacag	cagggcttct	3300
	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgca	accagcaaca	3360
	attccaccgt	ggcttgcat	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
	gcttcttgca	gcgatacagc	tcagacccca	caggcgctt	gactgaggac	agcatagacg	3480
	acaccttctt	cccagtgcct	gaatacataa	accagttccg	tcccaaaaag	cccgtgggt	3540
	ctgtgcagaa	tccgtgtctat	cacaatcagc	ctctgaaccc	cgcgcccgag	agagaccac	3600
55	actaccagga	cccccacagc	actgcagtgg	gcaaccccg	gtatctcaac	actgtccagc	3660
	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaaagcaa	3780
	atggcatctt	taagggctcc	acagctgaaa	atgcagaata	cctaaggggtc	gcgccacaaa	3840
	gcagtgaatt	tattggagca	tga				3863

<210> 731
 <211> 3866
 <212> DNA
 <213> Homo sapiens

5

<400> 731

10	ccccggcgag	cgcgggccga	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcgg	cggaggtccc	agctagcccc	ggcggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtcc	cccaggtccc	cgctctgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgtgctg	gctgcgtctt	300
	gcccggcgag	tcgggctctg	gaggaaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
15	aggtggtcct	tggaattttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggagggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tgagagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
20	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	ccccggggag	agcgactgcc	tggtctgccg	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
25	atgtggtgac	agatcacggc	tcgtgctgcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
30	ccttttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccgg	1800
	agcccaggga	ctgctgtctt	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
35	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920

40

45

50

55

5 acccagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
 gtatccagtg tgcccactac attgacggcc cccactgcgt caagacctgc ccggcaggag 2040
 tcatgggaga aaacaacacc ctggctctgga agtacgcaga cgccggccat gtgtgccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctgggccagg tcttgaaggc tgtccaacga 2160
 atggggcctaa gatcccgtcc atcgccactg ggatgggtggg ggccctcctc ttgctgctgg 2220
 tgggtggccct ggggatcggc ctcttcatgc gaaggcgcca catcgttcgg aagcgcacgc 2280
 tgcggaggct gctgcaggag agggagcttg tggagcctct tacacccagt ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtctg 2400
 gctccggtgc gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
 aaattcccgt cgctatcaag gaatcatctc cgaaagccaa caaggaaatc ctcgatgaag 2520
 10 cctacgtgat ggccagcgtg gacaaccccc acgtgtgccg cctgctgggc atctgcctca 2580
 cctccaccgt gcagctcatc acgcagctca tgcccttcgg ctgcctcctg gactatgtcc 2640
 gggaacacaa agacaatatt ggctcccagt acctgctcaa ctggtgtgtg cagatcgcaa 2700
 agggcatgaa ctacttgagg gaccgtcgtt tgggtgcaccg cgacctggca gccaggaacg 2760
 tactgggtgaa aacaccgcag catgtcaaga tcacagattt tgggctggcc aaactgctgg 2820
 gtgcggaaga gaaagaatac catgcagaag gaggcaaagt gcctatcaag tggatggcat 2880
 15 tggaaatcaat tttaacacaga atctataccc accagagtga tgtctggagc tacggggtga 2940
 ctgtttggga gttgatgacc tttggatcca agccatatga cggaatccct gccagcgaga 3000
 tctcctccat cctggagaaa ggagaacgcc tccctcagcc acctatatgt accatcgatg 3060
 tctacatgat catggtcaag tgctggatga tagacgcaga tagtcgcca aagttccgtg 3120
 agttgatcat cgaattctcc aaaaatggccc gagaccccca gcgctacctt gtcattcagg 3180
 gggatgaaag aatgcatttg ccaagtccta cagactccaa cttctaccgt gccctgatgg 3240
 20 atgaagaaga catggacgac gtgggtggatg ccgacgagta cctcatccca cagcagggct 3300
 tcttcagcag cccctccacg tcacggactc ccctcctgag ctctctgagt gcaaccagca 3360
 acaattccac cgtggcttgc attgatagaa atgggctgca aagctgtccc atcaaggaag 3420
 acagcttctt gcagcgatac agctcagacc ccacaggcgc cttgactgag gacagcatag 3480
 acgacacctt cctcccagtg cctgaatata taaaccagtc cgttcccaaa aggcccgctg 3540
 25 gctctgtgca gaatcctgtc tatcacaatc agcctctgaa ccccgcgccc agcagagacc 3600
 cacactacca ggacccccac agcactgcag tgggcaaccg cgagtatctc aacactgtcc 3660
 agccccacctg tgtcaacagc acattcgaca gccctgcccc ctgggcccag aaaggcagcc 3720
 accaaattag cctggacaac cctgactacc agcaggactt ctttcccaag gaagccaagc 3780
 caaatggcat ctttaagggtc tccacagctg aaaatgcaga atacctaagg gtcgcgccac 3840
 aaagcagtgaa atttattgga gcatga 3866

30 <210> 732
 <211> 3860
 <212> DNA
 <213> Homo sapiens

35 <400> 732

40

45

50

55

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgaggcggc	cggagtgccc	agctagcccc	ggcgggcgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgccctcgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcagact	cttcggggag	240
5	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
10	tagcagtcct	atctaactat	gatgcaaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgcccgtgg	aagtccccca	gtgactgctg	ccacaaccag	tgtgtgcag	960
15	gctgcacagg	cccccgggag	agcgactgcc	tggcttgccg	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatat	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgctgcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgcccgaac	gtgtgtaacg	1260
	gaataggtat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcggtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25							
	aactgttttg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgctgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
30	gcaaccttct	ggagggtgag	ccaaggaggt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccaactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccaatc	aaactgcacc	tacggatgca	ctgggcccagg	tcttgaaggc	tgtccaacga	2160
35	atgggacctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgtctgtgg	2220
	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgacgc	2280
	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtg	gttcggcacg	gtgtataaag	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaatcgaaag	cacacaagga	aatcctcag	gaagcctacg	2520
40	tgatggccag	cgtggacaac	ccccacgtgt	gccgcctgct	gggcatctgc	ctcacctcca	2580
	ccgtgcagct	catcacgcag	ctcatgccct	tcggctgcct	cctggactat	gtccgggaac	2640
	acaaagacaa	tattggctcc	cagtacctgc	tcaactgggtg	tgtgcagatc	gcaaagggca	2700
	tgaactactt	ggaggaccgt	cgcttggtgc	accgcgacct	ggcagccagg	aacgtactgg	2760
	tgaaaaacacc	gcagcatgtc	aagatcacag	atcttgggct	ggccaaactg	ctgggtgcgg	2820
	aagagaaaaga	ataccatgca	gaaggaggca	aagtgcctat	caagtggatg	gcattggaat	2880
45	caatttttaca	cagaatctat	acccaccaga	gtgatgtctg	gagctacggg	gtgactgttt	2940
	gggagttgat	gaccttttga	tccaagccat	atgacggaat	ccctgccagc	gagatctcct	3000
	ccatcctgga	gaaaggagaa	cgcttccctc	agccacccat	atgtaccatc	gatgtctaca	3060
	tgatcatggt	caagtgtctg	atgatagacg	cagatagtgc	cccaaagttc	cgtgagttga	3120
	tcacgcaatt	ctccaaaatg	gcccagacc	cccagcgcta	ccttgtcatt	caggggatg	3180
	aaagaatgca	tttgccaagt	cctacagact	ccaacttcta	ccgtgccttg	atggatgaag	3240
50	aagacatgga	cgacgtggtg	gatgccgacg	agtacctcat	cccacagcag	ggcttcttca	3300
	gcagccccctc	cacgtcacgg	actccccctc	tgagctctct	gagtgcaccc	agcaacaatt	3360
	ccaccgtggc	ttgcattgat	agaaatgggc	tgcaaagctg	tcccatcaag	gaagacagct	3420
	tcttgacagc	atacagctca	gacccacag	gcgccttgac	tgaggacagc	atagacgaca	3480
	ctttcctccc	agtgccctgaa	tacataaaacc	agtcctgttc	caaaaggccc	gctggctctg	3540
	tgcagaatcc	tgtctatcac	aatcagcctc	tgaacccgcg	gcccagcaga	gacccacact	3600
	accaggaccc	ccacagcact	gcagtgggca	accccgagta	tctcaacact	gtccagccca	3660
55	cctgtgtcaa	cagcacattc	gacagccctg	cccactgggc	ccagaaaggc	agccaccaa	3720
	ttagcctgga	caaccctgac	taccagcagg	acttctttcc	caaggaagcc	aagccaaatg	3780
	gcatctttaa	gggctccaca	gctgaaaatg	cagaatacct	aagggtcgcg	ccacaaagca	3840
	gtgaatttat	tggagcatga					3860

<210> 733
 <211> 3860
 <212> DNA
 <213> Homo sapiens

5

<400> 733

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgagggcggc	cggagtcccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
10	ggccacctcg	tcggcggtccg	cccgagtccc	cgctcgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttaggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtgggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
15	taaagacccat	ccaggagggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttggg	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtcctt	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
20	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgagg	agcgactgcc	tggtctgccg	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccga	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
25	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380

30

35

40

45

50

55

5 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 aggaaatcac aggggtttttg ctgattcagg ctgggcctga aaacaggacg gacctccatg 1500
 cctttgagaa cctagaaatc atacgcggca ggaccaagca acatgggtcag ttttctcttg 1560
 cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
 gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
 aactgtttgg gacctccggt cagaaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaaggccac aggccaggtc tgccatgcct tgtgctcccc cgagggctgc tggggcccg 1800
 agcccaggga ctgctgtctt tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
 gcaaccttct ggaggggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
 10 acccagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
 gtatccagtg tgcccactac attgacggcc ccactgctg caagacctgc ccggcaggag 2040
 tcatgggaga aaacaacacc ctggcttggga agtacgcaga cgccggccat gtgtgccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctggggcagg tcttgaaggc tgtccaacga 2160
 atgggcctaa gatcccgtcc atcgccactg ggatgggtgg ggccctctc ttgctgctgg 2220
 tgggtggccct ggggatcggc ctcttcatgc gaaggcgcca catcgttcgg aagcgcacgc 2280
 15 tgcggagggt gctgcaggag agggagcttg tggagcctct tacaccaggt ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgctgg 2400
 gctccggtgc gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
 aaattcccgt cgctatcaag gaatcgaaag ccaacaagga aatcctcgat gaagcctacg 2520
 tgatggccag cgtggacaac cccacgtgt ctcatgccct tcggctgcct cctggactat gtccgggaac 2640
 ccgtgcagct catcacgcag cagtacctgc tcaactgggt tgtgcagatg tgtgcagatg 2700
 20 acaaagacaa tattggctcc cgcttgggtg accgcgacct ggagccagg ggccaaactg aacgtactgg 2760
 tgaactactt gcagcatgtc aagatcacag attttgggt ggccaaactg ctgggtgcgg 2820
 aagagaaaga ataccatgca gaaggaggca aagtgcctat caagtggatg gcattggaat 2880
 caattttaca cagaatctat acccaccaga gtgatgtctg gagctacggg gtgactgttt 2940
 gggagttagt gacctttgga tccaagccat atgacggaat ccctgccagc gagatctcct 3000
 25 ccattcctgga gaaaggagaa cgcttccctc agccacccat atgtaccatc gatgtctaca 3060
 tgatcatggt caagtgtctg atgatagacg cagatagtcg cccaaagttc cgtgagttga 3120
 tcatcgaatt ctccaaaatg gcccgagacc cccagcgcta ccttgtcatt caggggggatg 3180
 aaagaatgca ttgtccaagt cctacagact ccaacttcta ccgtgccctg atggatgaag 3240
 aagacatgga cgacgtggtg gatgccgacg agtacctcat cccacagcag ggcttcttca 3300
 cgagccccctc cacgtcacgg actccccctc tgagctctct gactgcaacc agcaacaatt 3360
 ccaccgtggc ttgcattgat agaaatgggc tgcaaagctg tcccatcaag gaagacagct 3420
 30 tcttgacgagc atacagctca gacccacag gcgccttgac tgaggacagc atagacgaca 3480
 ccttcctccc agtgctgaa tacataaacc agtccgttcc caaaaggccc gctggctctg 3540
 tgcagaatcc tgtctatcac aatcagcctc tgaacccgc gccacgcaga gacccacact 3600
 accaggaccc ccacagcact gcagtgggca accccagta tctcaacact gtccagccca 3660
 cctgtgtcaa cagcacattc gacagccctg cccactgggc ccagaaaggc agccacaaa 3720
 35 ttagcctgga caaccctgac taccagcagg acttctttcc caaggaagcc aagccaaatg 3780
 gcatctttta gggctccaca gctgaaaatg cagaatacct aagggtcgcg ccacaaagca 3840
 gtgaatttat tggagcatga 3860

<210> 734

<211> 3878

40 <212> DNA

<213> Homo sapiens

<400> 734

45

50

55

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgaggcggc	cggagtcccg	agctagcccc	ggcgggccg	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgctctcgccg	ccaacgccac	aaccaccg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gccccgcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tggaattttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagacccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
10	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccattgaaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagagggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccgggag	agcgactgcc	tgttctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaacc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
20	atgtggtgac	agatcacggc	tcgtgctgcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaa	gtgtgtaacg	1260
	gaataggtat	tgggtgaatt	aaagactcac	tcctcataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
25	ccitttgagaa	cctagaaatc	atacgggcga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgtatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgtcctcc	cgagggtg	tgggggcccg	1800
	agcccaggga	ctgctgtctc	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
30	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	agccagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggctctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggacagg	tcttgaaggc	tgtccaacga	2160
35	atgggacctaa	gatccccgtcc	atcgccactg	ggatggtggg	ggccctcctc	ttgctgctgg	2220
	tgggtggccct	ggggatcgcc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgacgc	2280
	tgccggaggct	gctgcaggag	agggagcttg	tggagcctct	tacacccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
40	tcctcgatga	agcctacgtg	atggccagcg	tggaacaacc	ccacgtgtgc	gcgtctgtgg	2580
	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtc	aactggtgtg	2700
	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
	cagccaggaa	cgtactggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggcggg	2820
	ccaaactgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
45	agtggatggc	attggaatca	atittacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacggggg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccacccatat	3060
	gtaccatcga	tgtctacatg	atcatggtca	agtgtgtgat	gatagacgca	gatagtgcgc	3120
	caaagtcccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaaagtcc	tacagactcc	aacttctacc	3240
50	gtgccctgat	ggatgaagaa	gacatggacg	acgtggtgga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agcccccca	cgtcacggac	tccccctctg	agctctctga	3360
	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgcctgaata	cataaaccag	tccgttccca	3540
	aaaggccccg	tggctctgtg	cagaatcctg	tctatcacia	tcagcctctg	aacccccgcg	3600
55	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccgagatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactgggccc	3720
	agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gtcccacagc	tgaaaatgca	gaatacctaa	3840
	gggtcgcg	acaaagcagt	gaatttattg	gagcatga			3878

<210> 735
 <211> 3878
 <212> DNA
 <213> Homo sapiens

5

<400> 735

	cccggcgcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgaggcggc	cggagtcccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
10	ggccacctcg	tcggcggtccg	cccgagttccc	cgccctcgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
15	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtctt	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tgagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
20	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840

25

30

35

40

45

50

55

5 gctgggggtgc aggagaggag aactgccaga aactgaccaa aatcatctgt gcccagcagt 900
 gctccggggcg ctgccgtggc aagtcaccca gtgactgctg ccacaaccag tgtgctgcag 960
 gctgcacagg ccccccgggag agcgactgcc tggctgcccg caaatccga gacgaagcca 1020
 cgtgcaagga cactgcccc cactcatgc tctacaaccc caccacgtac cagatggatg 1080
 tgaacccccga gggcaaatatc agctttgggtg ccacctgcgt gaagaagtgt ccccgtaatt 1140
 atgtggtgac agatcacggc tctgtgctcc gagcctgtgg ggccgacagc tatgagatgg 1200
 aggaagacgg cgtccgcaag tgttaagaagt gcgaagggcc ttgccgcaaa gtgtgtaacg 1260
 gaataggat tgggtgaattt aaagactcac tctccataaa tgctacgaat attaaacact 1320
 tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 10 aggaaatcac aggggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
 cctttgagaa cctagaaatc atacgaggca ggaccaagca acatggtcag ttttctcttg 1560
 cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
 gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
 aactgtttgg gacctccgtg cagaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaagggcac agggcaggac tgccatgcct tgtgctcccc cgagggctgc tggggcccg 1800
 15 agcccaggga ctgctctct tggcggaaatg tcagccgagg cagggaatgc gtggacaagt 1860
 gcaaccttct ggaggggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgc 1920
 acccagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
 gtatccagt tgcacctac attgacggcg cccactgcgt caagacctgc ccggcaggag 2040
 tcatgggaga aaacaacacc ctggtctgga aggacgcaga cgccggccat gtgtgccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctggggcagg tcttgaaggc tgtccaacga 2160
 20 atgggcctaa gatcccgctc atcgccactg ggatgggtgg ggccctctc ttgctgtgtg 2220
 tgggtggccct ggggatcggc ctcttcacgc gaaggcgcca catcgttcgg aagcgacgc 2280
 tgcggaggct gctgcaggag agggagcttg tggagcctc tacaccaggt ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatg aaagtgtgtg 2400
 gctccgggtg gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
 aaattcccg ctctatcaag gaattaagag aagcaacatc tccgaaagcc aacaaggaaa 2520
 25 tctcgcgatga agcctacgtg atggccagcg tggacaaccc ccacgtgtgc cgcctgctgg 2580
 gcatctgcct cacctccacc gtgcagctca tcacgcagct catgccctc ggctgcctcc 2640
 tggactatgt ccgggaacac aaagacaata ttggctccca gtacctgtc aactgggtgtg 2700
 tgccagatgc aaagggcatg aactacttgg aggacgtcg cgtggtgcac aactgggtgtg 2760
 cagccaggaa cgtactgggtg aaaacaccgc agcatgtcaa gatcacagat tttgggcggg 2820
 ccaactgct ggggtgcggaa gagaaagaat accatgcaga aggaggcaaa gtgcctatca 2880
 30 agtggatggc attggaatca attttacaca gaatctatac ccaccagagt gatgtctgga 2940
 gctacggggg gactgttttg gagttgatga cctttggatc caagccatat gacggaatcc 3000
 ctgccagcga gatctcctc atcctggaga aaggagaacg cctccctcag ccacccatat 3060
 gtaccatcga tgtctacatg atcatggatc agtgctggat gatagacgca gatagtgcgc 3120
 caaagttccg tgagttgatc atcgaattct ccaaaatggc ccgagacccc cagcgctacc 3180
 ttgtcattca gggggatgaa agaatgcatt tgccaagtcc tacagactcc aacttctacc 3240
 35 gtgccctgat ggatgaagaa gacatggacg acgtgggtga tgccgacgag tacctcatcc 3300
 cacagcaggg cttcttcagc agcccccca cgtcacggac tccccctctg agctctctga 3360
 gtgcaaccag caacaattcc accgtggctt gcattgatag aaatgggctg caaagctgtc 3420
 ccatcaagga agacagcttc ttgcagcgat acagctcaga cccacaggc gccttgactg 3480
 aggacagcat agacgacacc ttctcccag tgcctgaata cataaaccag tccgttccca 3540
 aaagggccgc tggctctgtg cagaatcctg tctatcacia tcagcctctg aaccccgcg 3600
 40 ccagcagaga cccacactac caggaccccc acagactgc agtgggcaac cccgagtatc 3660
 tcaacactgt ccagcccacc tgtgtcaaca gcacattcga cagccctgcc aactgggccc 3720
 agaaaggcag ccaccaaatt agcctggaca accctgacta ccagcaggac ttctttccca 3780
 aggaagccaa gccaaatggc atctttaagg gctccacagc tgaaaatgca gaatacctaa 3840
 gggtcgcgcc acaaagcagt gaatttattg gagcatga 3878

45 <210> 736
 <211> 3878
 <212> DNA
 <213> Homo sapiens

50 <400> 736

55

	cccgggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggagtgccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctcc	cccaggtccc	cgctctcgcc	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgct	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gccccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaaat	tcctatgcct	600
10							
	tagcagtctt	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
15	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	tgtgtgcag	960
	gctgcacagg	cccccgggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaacccccga	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
20	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtag	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccittgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
25	ccttcgtcag	cctgaacata	acatccttgg	gtgtggagaa	cctcaaggag	ataagtgtg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgctctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaaggaggt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
30	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagt	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatggtggg	ggccctctct	ttgtgtgtgg	2220
	tgttgccct	ggggatcggc	ctcttcattgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
35	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacacccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	agcgaacatc	tccgaaagcc	aacaaggaaa	2520
40	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cgctgtctgg	2580
	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtct	aactggtgtg	2700
	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
	cagccaggaa	cgtactggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggctgg	2820
	ccaaacagct	gggtgcgga	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
45	agtggatggc	atttgaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacggggg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccacccatat	3060
	gtaccatcga	tgtctacatg	atcatggtca	agtgtggat	gatagacgca	gatagtgcgc	3120
	caaagttccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtggtgga	tgccgacgag	tacctcatcc	3300
50	cacagcaggg	cttcttcagc	agcccccca	cgtcacggac	tccccctctg	agctctctga	3360
	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aatgggctg	caaagctgtc	3420
	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gcttgcactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgcctgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggctctgtg	cagaatcctg	tctatcacia	tcagcctctg	aaccccgcg	3600
	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccagtatc	3660
55	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactgggccc	3720
	agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gtccacagc	tgaaaatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaatttatgt	gagcatga			3878

<210> 737
 <211> 3878
 <212> DNA
 <213> Homo sapiens

5

<400> 737

10

```

ccccggcgag cgcgggccga gcagcctccg ccccccgcac ggtgtgagcg cccgacgcgg 60
ccgagggcggc cggagtcccc agctagcccc ggcggccgcc gccgcccaga ccggacgaca 120
ggccacctcg tcggcgctcc cccgagtccc cgcctcgccg ccaacgccac aaccaccgcg 180
cacggcccc tgactccgtc cagtattgat cgggagagcc ggagcgagct cttcggggag 240
cagcgatgcg accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgctct 300

```

15

20

25

30

35

40

45

50

55

	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcagcagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
5	taaagaccat	ccaggagggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccitttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	ctcatgcct	600
	tagcagtcct	atctaactat	gatgcaaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
10	gctccgggag	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgagg	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatat	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgctcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagat	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
15	gaatagggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
20	cagtcggtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgctctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
25	acccagagtgc	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtgc	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggatggtggg	ggccctcctc	ttgtgtctgg	2220
	tgggtggccct	ggggatcggc	ctcttcctgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
30	tgccggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtgtg	2400
	gctccgggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgtatga	agcctacgtg	atggccagcg	tggaacaacc	ccacgtgtgc	ggcctgcttg	2580
35	gcactctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	cgcctgctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtct	aactggtgtg	2700
	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
	cagccaggga	cgtactggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggctgg	2820
	ccaaactgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	attggaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
40	gctacgggggt	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccacccatat	3060
	gtaccatcga	tgtctacatg	atcatggtca	agtgtggtg	gatagacgca	gatagtgcgc	3120
	caaagttccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaagtcc	tacagactcc	aacttctacc	3240
45	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtgga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agcccccca	cgtcacggac	tccccctctg	agctctctga	3360
	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccc	tgctgaata	cataaaccag	tccgttccca	3540
	aaaggccccg	tggctctgtg	cagaatcctg	tctatcacaa	tcagcctctg	aaccccgcgc	3600
50	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccgagtatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactgggccc	3720
	agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gctccacagc	tgaatatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaatttattg	gagcatga			3878

<210> 738

<211> 3860

55 <212> DNA

<213> Homo sapiens

<400> 738

ccccggcgag cgcgggccgca gcagcctccg cccccgcac ggtgtgagcg cccgacgcgg 60

5	ccgagggcgc	cggagtcctc	agctagcccc	ggcgggccgc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctcg	cccaggtccc	cgctctcgcc	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgaggagacc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tggaattttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
10	taaagaccat	ccaggagggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaaa	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccattgaaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
15	acttccagaa	ccacctgggc	agctgcaaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgcccgtgg	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	ccccgggag	agcgactgcc	tgttctgccg	caaattccga	gacgaagcca	1020
	cgtgcaagg	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
20	atgtggtgac	agatcacggc	tcgtgctgct	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaact	1260
	gaataggat	tgggtgaattt	aaagactcac	tcctcataaa	tgctacgaat	attaacact	1320
	tcaaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatatctctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
25	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgctgctct	tgccggaatg	tcagccgagg	caggggaatgc	gtggacaagt	1860
30	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatggtggg	ggccctcctc	ttgctgctgg	2220
35	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggaagct	gctgcaggag	agggagcttg	tggagcctct	tacaccagtc	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaatcgaagg	ccaacaagga	aatcctcgat	gaagcctacg	2520
40	tgatggccag	cgtggacaac	ccccacgtgt	gccgcctgct	gggcatctgc	ctcacctcca	2580
	ccgtgcagct	catcacgcag	ctcatgccct	tcggctgcct	cctggactat	gtccgggaac	2640
	acaaagacaa	tattggctcc	cagtacctgc	tcaactgggtg	tgtgcagatc	gcaaagggca	2700
	tgaactactt	ggaggaccgt	cgcttggtgc	accgcgacct	ggcagccagg	aacgtactgg	2760
	tgaaaacacc	gcagcatgtc	aagatcacag	attttgggct	ggccaaactg	ctgggtgcgg	2820
	aagagaaaga	ataccatgca	gaaggaggca	aagtgcctat	caagtggatg	gcattggaat	2880
45	caattttaca	cagaatctat	acccaccaga	gtgatgtctg	gagctacggg	gtgactgttt	2940
	gggagttgat	gaccttttga	tccaagccat	atgacggaat	ccctgccagc	gagatctcct	3000
	ccatcctgga	gaaaggagaa	cgctccctc	agccacccat	atgtaccatc	gatgtctaca	3060
	tgatcatggt	caagtgtctg	atgatagacg	cagatagtcg	cccaaagttc	cgtgagttga	3120
	tcatcgaatt	ctccaaaatg	gcccagagacc	cccagcgcta	ccttgtcatt	cagggggatg	3180
	aaagaatgca	tttgccaagt	cctacagact	ccaacttcta	ccgtgccctg	atggatgaag	3240
50	aagacatgga	cgacgtgggtg	gatgccgacg	agtacctcat	cccacagcag	ggcttcttca	3300
	gcagccccct	cacgtcacgg	actccccctc	tgagctctct	gagtgaacc	agcaacaatt	3360
	ccaccgtggc	ttgcattgat	agaaatgggc	tgcaaagctg	tcccatcaag	gaagacagct	3420
	tcttgcaagc	atacagctca	gacccccacg	gcgccttgac	tgaggacagc	atagacgaca	3480
	cttctctccc	agtgcctgaa	tacataaacc	agtccgttcc	caaaaggccc	gctggctctg	3540
	tgcagaatcc	tgtctatcac	aatcagcctc	tgaaccccg	gcccagcaga	gacctccact	3600
55	accaggaccc	ccacagcact	gcagtgggca	accccgagta	tctcaacact	gtccagccca	3660
	cctgtgtcaa	cagcacattc	gacagccctg	cccactgggc	ccagaaaggc	agccaccaa	3720
	ttagcctgga	caaccctgac	taccagcagg	acttctttcc	caaggaagcc	aagccaaatg	3780
	gcatctttaa	gggctccaca	gctgaaaatg	cagaatacct	aagggtcgcg	ccacaaagca	3840
	gtgaatttat	tggagcatga					3860

<210> 739
<211> 3863
<212> DNA
<213> Homo sapiens

5

<400> 739

10

15

20

25

30

35

40

45

50

55

	cccgggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgctctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggagggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
10	tagcagtcct	atctaactat	gatgcaataa	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgcccgtgg	aagtccccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	cccccgggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatat	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgctgcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgcccgaac	gtgtgtaacg	1260
	gaataggtat	tggtagaatt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcggtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgctctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
30	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgagggc	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgtctgttg	2220
	tgggtggccct	ggggatcggc	ctcttcactg	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
35	gctccgggtg	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattccccgt	cgctatcaaa	acatctccga	aagccaacaa	ggaaatcctc	gatgaagcct	2520
	acgtgatggc	cagcgtggac	aacccccacg	tgtgcccgtc	gctgggcact	tgcctcactt	2580
	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcggctg	cctcctggac	tatgtccggg	2640
	aacacaaaga	caatattggc	tcccagtacc	tgtcactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttgaggagc	cgctcgcttg	tgcaccgcga	cctggcagcc	aggaacgtac	2760
40	tggtgaaaac	accgcagcat	gtcaagatca	cagatttttg	gctggccaaa	ctgctgggtg	2820
	cggagagaaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcatttg	2880
	aatcaatttt	acacagaatc	tatacccacc	agagtgatgt	ctggagctac	gggggtgactg	2940
	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	ggtcaagtgc	tggatgatag	acgcagatag	tcgccccaa	ttccgtgagt	3120
45	tgatcatcga	attctccaaa	atggccccgag	acccccagcg	ctaccttgct	attcaggggg	3180
	atgaaagaat	gcattttgcca	agtcttacag	actccaactt	ctaccgtgcc	ctgatggatg	3240
	aagaagacat	ggacgacgtg	gtggatgccg	acgagtacct	catcccacag	cagggcttct	3300
	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgc	accagcaaca	3360
	attccaccgt	ggcttgcat	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
	gcttcttgca	gcgatacagc	tcagacccca	caggcgctt	gactgaggac	agcatagacg	3480
50	acaccttctt	cccagtgcct	gaatacataa	accagtccgt	tcccaaaaag	cccgtgggt	3540
	ctgtgcagaa	tccgtcttat	cacaatcagc	ctctgaacc	cgcgccagcc	agagaccac	3600
	actaccagga	ccccacagc	actgcagtgg	gcaacccga	gtatctcaac	actgtccagc	3660
	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaaagcaa	3780
	atggcatctt	taagggctcc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
55	gcagtgaatt	tattggagca	tga				3863

<210> 740

<211> 3878

<212> DNA

<213> Homo sapiens

<400> 740

5

10

15

20

25

30

35

40

45

50

55

```

ccccggcgag cgcggccgca gcagcctccg cccccgcac ggtgtgagcg cccgacgagg 60
ccgaggcggc cggagtcccg agctagcccc ggcgggcgcc gccgcccaga cccgacgaca 120
ggccacctcg tcggcggtccg cccgagtccc cgctctcgcc ccaacgccac aaccaccgag 180
cacggccccc tgactccgtc cagtattgat cgggagagcc ggagcgagct cttcggggag 240
cagcgtatgag accctccggg acggccgggg cagcgctcct ggcgtgctg gctgcgctct 300
gcccggcgag tcggggtctg gaggaaaaga aagtttgcca aggcacgagt aacaagctca 360
cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
aggtggtcct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
taaagaccat ccaggagggtg gctgggttatg tcctcattgc cctcaacaca gtggagcgaa 540
ttcctttgga aaacctgcag atcatcagag gaaatatgta ctacgaaaaa tcctatgcct 600
tagcagtcct atctaactat atcgcaataa aaaccggact gaaggagctg cccatgagaa 660
atttacagga aatcctgcat ggcggcgtgc ggttcagcaa caacctgcc ctgtgcaacg 720
tggagagcat ccagtggcgg gacatagtca gcagtgaact tctcagcaac atgtcgatgg 780
acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
gctgggggtg aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagt 900
gctccggggc ctgccgtggc aagtccccc gtgactgctg ccacaaccag tgcgtgcag 960
gctgcacagg ccccgggag agcgactgcc tctctgccc caaattccga gtagtagaaa 1020
cgtgcaagga cacctgcccc ccactcatgc tctacaaccc caccacgtac cagatggatg 1080
tgaaccccg gggcaaatat agctttgggtg ccacctgctg gaagaagtgt ccccgtaatt 1140
atgtggtgac agatcacggc tcgtgcgtcc gagcctgtgg ggccgacagc tatgagatgg 1200
aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgaaa gtgtgtaacg 1260
gaataggtag tggtagaatt aaagactcac tctcataaaa tgctacgaat attaaacact 1320
tcaaaaactg cacttccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
aggaaatcac aggggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
cctttgagaa cctagaaatc atacgaggca ggaccaagca acatggtcag ttttctcttg 1560
cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
aactgtttgg gacctccggt cagaaaacca aaattataag caacagaggt gaaaacagct 1740
gcaaggccac aggccagggtc tgccatgcct tgtgtcccc cgagggctgc tggggcccg 1800
agcccaggga ctgctgtctt tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
gcaaccttct ggagggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
accaagtggt cctgcctcag gccatgaaca taccctgcac aggacgggga ccagacaact 1980
gtatccagtg tgcccactac attgacggcc cccactgctg caagacctgc ccggcaggag 2040
tcatgggaga aaacaacacc ctggcttggg agtacgcaga cgccggccat gtgtgccacc 2100
tgtgccatcc aaactgcacc tacggatgca ctgggcccagg tcttgaaggc tgtccaacga 2160
atgggcctaa gatccccgtc atcgccactg ggatgggtggg ggccctcctc ttgctgctgg 2220
tggtgccctt ggggatcggc ctcttcattg gaaggcgcca catcgcttcg gaggcagcgc 2280
tgccggaggc gctgcaggag agggagcttg tggagcctct tacaccagtg ggagacgctc 2340
ccaaccaagc tctcttgagg atcttgaagg taactgaatt caaaaagatc aaagtgtcta 2400
gctccggtgc gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
aaattcccgt cgctatcaag gaattaagag aagcaacatc tccgaaagcc aacaaggaaa 2520
tcctcgatga agcctacgtg atggccagcg tggacaaccc ccacgtgtgc cgcctgcttg 2580
gcatctgcct cacctccacc gtgcagctca catgcccttc ggctgcctcc 2640
tggactatgt ccgggaacac aaagacaata ttggctccca gtacctgtct aactgggtgtg 2700

tgcagatcgc aaagggcatg aactacttgg aggaccgtcg cttggtgcac cgcgacctgg 2760
cagccaggaa cgtactggtg aaaacaccgc agcatgtcaa gatcacagat tttgggctgg 2820
ccaaactgct ggggtgcggaa gagaaagaat accatgcaga aggaggcaaa gtgcctatca 2880
agtggatggc attggaatca attttacaca gaatctatac ccaccagagt gatgtctgga 2940
gctacggggg gactgttttg gagttgatga cctttggatc caagccatat gacggaatcc 3000
ctgccagcga gatctcctcc atcctggaga aaggagaacg cctccctcag ccacccatat 3060
gtaccatcga tgtctacatg atcatggtca agtgctggat gatagacgca gatagtgcgc 3120
caaagtcccg tgagttgatc atcgaattct ccaaattggc cagagacccc cagcgctacc 3180
ttgtcattca gggggatgaa agaatgcatt tgccaagtcc tacagactcc aacttctacc 3240
gtgccctgat ggtatgaaga gacatggacg acgtgggtga tgccgacgag tacctcatcc 3300
cacagcaggg cttcttcagc agcccccca cgtcacggac tccccctctg agctctctga 3360
gtgcaaccag caacaattcc accgtggctt gcattgatag aaatgggctg caaagctgtc 3420
ccatcaagga agacagcttc ttgcagcgat acagctcaga cccacaggc gccttgactg 3480
aggacagcat agacgacacc ttctcccgag tgctgaata cataaacag tccgttccca 3540
aaaggccccg tggctctgtg cagaatcctg tctatcacia tcagcctctg aaccccgcg 3600

```

```

      ccagcagaga cccacactac caggaccccc acagcactgc agtgggcaac cccgagtatc 3660
      tcaacactgt ccagcccacc tgtgtcaaca gcacattcga cagccctgcc cactggggccc 3720
      agaaaggcag ccaccaaatt agcctggaca accctgacta ccagcaggac ttctttccca 3780
      aggaagccaa gccaaatggc atctttaagg gctccacagc tgaaaatgca gaatacctaa 3840
5      gggtcgcgcc acaaagcagt gaatttattg gagcatga          3878

```

<210> 741

<211> 3878

<212> DNA

10 <213> Homo sapiens

<400> 741

15

20

25

30

35

40

45

50

55

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	gggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctcc	cccaggtccc	cgccctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataaactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggttat	tcctcattgc	cctcaacaca	gtggagcgaa	540
10	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagctct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagagggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
15	gctccggggcg	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgggag	agcgactggc	tgtctgccc	caaattccga	gacgaagcca	1020
	cgtagcaagg	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataaggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccgggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaaatc	atacgcggca	ggaccaagca	acatgggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccggt	cagaaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
30	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatggggaga	aaacaacacc	ctgggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgctgctgg	2220
	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggagggt	gctgcaggag	agggagcttg	tggagcctct	tacaccagct	ggagaagctc	2340
35	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagagt	aaagtgtctg	2400
	cctccgggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cacctgctgg	2580
	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtc	actgggtgtg	2700
40	tgcatatcgc	aaagggcatg	aaactacttg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
	cagccaggaa	cgtactggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggctgg	2820
	caaactgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	atttgaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacgggg	gactgtttgg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccaccatat	3060
45	gtaccatcga	tgtctacatg	atcatgggtca	agtgtctggat	gatagacgca	gatagtgcgc	3120
	caaagtcccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agccccctca	cgtcacggac	tccccctctg	agctctctga	3360
50	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
	ccatcaaggga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgcttgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggctctgtg	cagaatcctg	tctatcacia	tcagcctctg	aacccgcgc	3600
	ccagcagaga	ccacactac	caggaccccc	acagactgc	agtgggcaac	cccgagatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactgggccc	3720
55	agaaaggcag	ccaccaaat	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gctccacagc	tgaaaatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaatttattg	gagcatga			3878

<210> 742
 <211> 3863
 <212> DNA
 <213> Homo sapiens

5

<400> 742

	cccggcgag	cgcgccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgaggcggc	cggagtcctc	agctagcccc	ggcgccgccc	gccgcccaga	ccggacgaca	120
10	ggccacctcg	tcggcgctcc	cccagatccc	cgctcgccc	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgctctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattaccct	atgtgcagag	gaattatgat	ctttccttct	480
15	taaagaccat	ccaggaggtg	gctgggttat	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagctct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
20	gctggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	ccccggggag	agcgactgcc	tggtctgccg	caaattccga	gacgaagcca	1020
	cgtgcaaggga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
25	aggaagaccg	cgtccgcaag	tgtaaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaatc	ataccgcca	ggaccaagca	acatggtcag	ttttctcttg	1560
30	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggg	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agccccagga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaaatg	gtgggaagct	1860
	gcaacccttct	ggagggtgag	ccaagggaagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
35	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccaactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggatgggtgg	ggccctcttc	ttgctgctgg	2220
	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgacgc	2280
40	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccagct	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaaa	acatctccga	aagccaacaa	ggaaatcctc	gatgaagcct	2520
	acgtgatggc	cagcgtggac	aacccccacg	tgtgccgcct	gctgggcac	tgccctacct	2580
	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcggctg	cctcctggac	tatgtccggg	2640
45	aacacaaaga	caatattggc	tcccagtacc	tgctcaactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttggaggac	cgtcgcttgg	tgaccgcga	cctggcagcc	aggaacgtac	2760
	tggtgaaaac	accgcagcat	gtcaagatca	cagatttttg	gctggccaaa	ctgctgggtg	2820
	cgaagagaa	agaataccat	gcagaaggag	gcaaagtgc	tatcaagtgg	atggcattgg	2880
50	aatcaatttt	acacagaatc	tatacccacc	agagtgtatg	ctggagctac	ggggtgactg	2940
	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060

55

```

      acatgatcat ggtcaagtgc tggatgatag acgcagatag tcgccc aaag ttccgtgagt 3120
      tgatcatcga atttctccaaa atggcccagag acccccagcg ctaccttggtc attcaggggg 3180
      atgaaagaat gcatttgcca agtcctacag actccaactt ctaccgtgcc ctgatggatg 3240
      aagaagacat ggacgacgtg gtggatgccg acgagtacct catcccacag cagggcttct 3300
5      tcagcagccc ctccacgtca cggactcccc tcctgagctc tctgagtgca accagcaaca 3360
      attccaccgt ggcttgcatg gatagaaatg ggctgcaaag ctgtcccatc aaggaagaca 3420
      gcttcttgca gcgatacagc tcagacccca caggcgctt gactgaggac agcatagacg 3480
      acaccttcct ccagtgctc gaatacataa accagtccgt tcccaaaagg cccgctggct 3540
      ctgtgcagaa tcctgtctat cacaatcagc ctctgaacc cgcgcccagc agagaccac 3600
10      actaccagga ccccccacagc actgcagtgg gcaaccccga gtatctcaac actgtccagc 3660
      ccacctgtgt caacagcaca ttcgacagcc ctgcccactg ggcccagaaa ggcagccacc 3720
      aaattagcct ggacaaccct gactaccagc aggacttct tcccaaggaa gccaaagcaa 3780
      atggcatctt taagggctcc acagctgaaa atgcagaata cctaagggtc gcgccacaaa 3840
      gcagtgaatt tattggagca tga                                     3863

```

```

15      <210> 743
      <211> 3863
      <212> DNA
      <213> Homo sapiens

```

```

20      <400> 743

```

```

25

```

```

30

```

```

35

```

```

40

```

```

45

```

```

50

```

```

55

```

	cccgggcgag	cgcgggccga	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccg	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtcc	cccaggtccc	cgcttcgccc	ccaacgccac	aaccaccg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
5	cagcgatg	accctccgg	acggccggg	cagcgctcct	ggcgctgctg	gctgcgtct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
10	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgcccgtgg	aagtccccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	cccccgggag	agcgactgcc	tggcttgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggt	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgctgcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	tgcccgcaag	tgtaagaagt	gcgaagggcc	ttgcccgaac	gtgtgtaacg	1260
	gaataggtat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggttttt	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcggtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgtatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgctctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaaggagtg	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagatg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
30	tcatgggaga	aaacaacacc	ctgggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgaaagg	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgtgtcttg	2220
	tggtggccct	ggggatcggc	ctcttcacgt	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcgagggtc	gctgcaggag	agggagcttg	tggagcctct	tacacccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaaag	aaactgaatt	caaaaagatc	aaagtgtctg	2400
35	gctccgggtg	gttcggcacg	gtgtataaag	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaaa	acatctccga	aagccaacaa	ggaaatcctc	gatgaagcct	2520
	acgtgatggc	cagcgtggac	aacccccacg	tgtgcgcctc	gctgggcact	tgcctacact	2580
	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcggctg	cctcctggac	tatgtccggg	2640
	aacacaaaga	caatattggc	tcccagtagc	tgtcactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttgaggagc	cgctcgcttg	tgcaccgcga	cctggcagcc	aggaacgtac	2760
40	tggtgaaaac	accgcagcat	gtcaagatca	cagatttttg	gctggccaaa	ctgctgggtg	2820
	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcatttg	2880
	aatcaatttt	acacagaatc	tataccacc	agagtgatgt	ctggagctac	ggggtgactg	2940
45	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	ggtcaagtgc	tggatgatag	acgcagatag	tcgccaaaag	ttccgtgagt	3120
	tgatcatcga	attctccaaa	atggccccg	acccccagcg	ctaccttgct	attcaggggg	3180
	atgaaagaat	gcatttgcca	agtcctacag	actccaactt	ctaccgtgcc	ctgatggatg	3240
	aagaagacat	ggacgcagtg	gtggatgccc	acgagtacct	catcccacag	cagggcttct	3300
	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgca	accagcaaca	3360
50	attccaccgt	ggcttgcat	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
	gcttcttgca	gcgatacagc	tcagacccca	caggcgctt	gactgaggac	agcatagacg	3480
	acacgttctc	cccagtgcct	gaatacataa	accagctcgt	tcccaaaagg	cccgtggct	3540
	ctgtgcagaa	tctgtctat	cacaatcagc	ctctgaaccc	cgcgcccgag	agagaccac	3600
	actaccagga	cccccacagc	actgcagtgg	gcaacccgga	gtatctcaac	actgtccagc	3660
	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
55	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaaagcaa	3780
	atggcatctt	taagggtctc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
	gcagtgaaat	tattggagca	tga				3863

<210> 744
 <211> 3863
 <212> DNA
 <213> Homo sapiens

5

<400> 744

10 cccggcgag cgcggccgca gcagcctccg ccccccgcac ggtgtgagcg cccgacgagg 60
 cccgagcgcc cggaggtccc agctagcccc ggccggccgc gccgcccaga cccgacgaca 120
 ggccacctcg tcggcggtcc cccgaggtccc cgcttcgccc ccaacgccac aaccaccgag 180
 cacggccccc tgactccgtc cagttattgat cgggagagcc ggagcgagct cttcggggag 240
 cagcgatgag accctccggg acggccgggg cagcgctcct ggcgtgctg gctgcgctct 300
 gcccggcgag tcgggctctg gaggaagaaga aagtttgcca aggcacgagt aacaagctca 360
 cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
 15 aggtggctct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
 taaagaccat ccaggagggt gctgggttatg tcctcattgc cctcaacaca gtggagcgaa 540
 ttccitttga aaacctgcag atcatcagag gaaatatgta ctacgaaaat tcctatgcct 600
 tagcagtcct atctaactat gatgcaaata aaaccggact gaaggagctg cccatgagaa 660
 atttacagga aatcctgcat ggcgccgtgc ggttcagcaa caacctgcc ctgtgcaacg 720
 tggagagcat ccagtggcgg gacatagtcg gcagtgactt tctcagcaac atgtcgatgg 780
 20 acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
 gctgggggtg aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagt 900
 gctccgggag ctgcccgtgg aagtccccc gtgactgctg ccacaaccag tgtgtgcag 960
 gctgcacagg ccccggggag agcgactgcc tggctgccc ccaattccga gacgaagcca 1020
 cgtgcaagga cacctgcccc ccactcatgc tctacaaccc caccacgtac cagatggatg 1080
 tgaaccccga gggcaaatat agctttgggt ccacctgctg gaagaagtgt ccccgtaatt 1140
 25 atgtgggtgac agatcacggc tcgtgcgtcc gagcctgtgg ggccgacagc tatgagatgg 1200
 aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgaaa gtgtgtaacg 1260
 gaatagggtat tgggtgaattt aaagactcac tctccataaa tgctacgaat attaaacact 1320
 tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 aggaaatcac aggggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
 30 cctttgagaa cctagaaatc atacgcccga ggaccaagca acatggtcag ttttctcttg 1560
 cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
 gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
 aactgtttgg gacctccggt cagaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaaggccac aggccagggtc tgccatgcct tgtgctcccc cgagggctgc tggggcccgg 1800
 agcccaggga ctgctgtctt tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
 35 gcaaccttct ggaggggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
 acccagagtg cctgcctcag gccatgaaca tcacctgcac aggcagggga ccagacaact 1980
 gtatccagtg tgccactac attgacggcc cccactgctg caagacctgc ccggcaggag 2040
 tcatgggaga aaacaacacc ctggtctgga agtacgcaga cgccggccat gtgtgccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctgggcccag tcttgaaggc tgtccaacga 2160
 atgggcctaa gatcccgtcc atcgccactg ggatgggtgg ggccctcctc ttgctgctgg 2220
 40 tggtgccctt ggggatcggc ctcttcatgc gaaggcgcca catcgttcgg aagcgacgc 2280
 tgccgaggct gctgcaggag agggagcttg tggagcctct tacaccagt ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtctg 2400
 gctccgggtg gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
 aaattcccgt cgctatcaaa acatctccga aagccaacaa ggaaatcctc gatgaagcct 2520
 acgtgatggc cagcgtggac aacccccacg tgtgccgcct gctgggcatc tgcctcacct 2580

45

50

55

	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcgggctg	cctcctggac	tatgtccggg	2640
	aacacaaaga	caatattggc	tcccagtagc	tgctcaactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttggaggac	cgtcgcttgg	tgacccgcga	cctggcagcc	aggaacgtac	2760
5	tggtgaaaac	accgcagcat	gtcaagatca	cagatttttg	gctggccaaa	ctgctgggtg	2820
	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcattgg	2880
	aatcaatttt	acacagaatc	tatacccacc	agagtgatgt	ctggagctac	ggggtgactg	2940
	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
10	acatgatcat	ggatcaagtgc	tggatgatag	acgcagatag	tcgccccaaag	ttccgtgagt	3120
	tgatcatcga	attcttccaaa	atggcccagag	acccccagcg	ctaccttgtc	attcaggggg	3180
	atgaaagaat	gcattttgcca	agtcctacag	actccaactt	ctaccgtgcc	ctgatggatg	3240
	aagaagacat	ggacgacgtg	gtggatgccg	acgagtacct	catcccacag	cagggcttct	3300
	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgcg	accagcaaca	3360
	attccaccgt	ggcttgcatc	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
15	gcttcttgca	gcgatacagc	tcagaccca	caggcgctt	gactgaggac	agcatagacg	3480
	acaccttcct	cccagtgctt	gaatacataa	accagtccgt	tcccaaaagg	cccgtggct	3540
	ctgtgcagaa	tcctgtctat	cacaatcagc	ctctgaaccc	cgcgcccagc	agagaccac	3600
	actaccagga	ccccacagc	actgcagtgg	gcaaccccga	gtatctcaac	actgtccagc	3660
	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaagccaa	3780
20	atggcatctt	taagggctcc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
	gcagtgaatt	tattggagca	tga				3863

<210> 745

<211> 3863

<212> DNA

25 <213> Homo sapiens

<400> 745

30

35

40

45

50

55

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgaggcggc	cggagtcctcg	agctagcccc	ggcgggccg	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctccg	cccaggtccc	cgccctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgctc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
10	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	cttggaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagagggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
15	gctccggggcg	ctgccgtggc	aagtcaccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgggag	agcgactgcc	tgtctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tgctacaacc	caccacgtac	cagatggatg	1080
	tgaaccccga	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtaaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataaggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccgggtggca	cttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaaatc	atacgcgcca	ggaccaagca	acatgggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgttttg	tgccctccgt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
30	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggctctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgctgctgg	2220
	tgggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
35	tgccggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccagct	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtggt	2400
	gctccgggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaaa	acatctccga	aagccaacaa	ggaaatcctc	gatgaagcct	2520
40	acgtgatggc	cagcggtggac	aacccccacg	tgtgccgcct	gctgggcatc	tgctctacct	2580
	ccacctgtga	gctcatcacg	cagctcatgc	cctctggctg	cctcctggag	tgcctccggg	2640
	aacacaaaga	caatatggcg	tcccagttacc	tgctcaactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttggaggac	cgctcgcttg	tgaccgcga	cctggcagcc	aggaacgtac	2760
	tgggtgaaaac	accgcagcat	gtcaagatca	cagattttgg	gctggccaaa	ctgctgggtg	2820
	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcatttg	2880
	aatcaatttt	acacagaatc	tatacccacc	agagtgatgt	ctggagctac	ggggtgactg	2940
45	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	ggtcaagtgc	tggatgatag	acgcagatag	tcgcccaga	ttccgtgagt	3120
	tgatcatcga	attctccaaa	atggcccag	acccccagcg	ctacctgtgc	attcaggggg	3180
	atgaaagaat	gcattttgcca	agtcttacag	actccaactt	ctaccgtgac	ctgatggatg	3240
	aagaagacat	ggacgacgtg	gtggatgccg	acgagtacct	catcccacag	cagggtctct	3300
50	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgc	accagcaaca	3360
	attccaccgt	ggcttgcat	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
	gcttcttgca	gcgatacagc	tcagaccca	caggcgctt	gactgaggac	agcatagacg	3480
	acaccttct	cccagtgctt	gaatacataa	accagtccgt	tcccaaaagg	cccgtggtct	3540
	ctgtgcagaa	tcctgtctat	cacaatcagc	ctctgaacct	cgcgcccagc	agagaccac	3600
	actacagga	ccccacagc	actgcagtgg	gcaaccccga	gtatctcagc	actgtccagc	3660
55	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaagccaa	3780
	atggcatctt	taagggtctc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
	gcagtgaatt	tattggagca	tga				3863

<210> 746
 <211> 3863
 <212> DNA
 <213> Homo sapiens

5

<400> 746

10 cccggcgag cgcggccgca gcagcctccg ccccccgcac ggtgtgagcg cccgacgcgg 60
 cccgagggcgc cggaggtccc agctagcccc ggcgcccgcc gccgcccaga ccggacgaca 120
 ggccacctcg tcggcggtccg cccgaggtccc cgccctcgccg ccaacgccac aaccaccgcg 180
 cacggccccc tgactccgtc cagtattgat cgggagagcc ggagcgagct cttcggggag 240
 cagcgatgag accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgtctt 300
 gcccggcgag tcgggctctg gaggaagaaga aagtttgcca aggcacgagt aacaagctca 360
 cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
 aggtggctct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
 15 taaagaccat ccaggaggtg gctgggtatg tcctcattgc cctcaacaca gtggagcgaa 540
 ttccctttgga aaacctgcag atcatcagag gaaatatgta ctacgaaaat tcctatgcct 600
 tagcagctct atctaactat gatgcaaata aaaccggact gaaggagctg cccatgagaa 660
 atttacagga aatcctgcat ggcgccgtgc ggttcagcaa caaccctgcc ctgtgcaacg 720
 tggagagcat ccagtgccgg gacatagtca gcagtgaact tctcagcaac atgtcgatgg 780
 acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
 20 gctgggggtgc aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagt 900
 gctccggggcg ctgccgtggc aagtccccca gtgactgctg ccacaaccag tgtgctgcag 960
 gctgcacagg ccccggggag agcgactgcc tggctgccc caaattccga gacgaagcca 1020
 cgtgcaagga cacctgcccc ccaactcatgc tctacaaccc caccacgtac cagatggatg 1080
 tgaaccccgga gggcaaatac agctttgggtg ccacctgcgt gaagaagtgt ccccgtaatt 1140
 atgtgggtgac agatcacggc tcgtgcgtcc gagcctgtgg ggccgacagc tatgatggatg 1200
 25 aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgcaaa gtgtgtaacg 1260
 gaataggtat tgggtgaattt aaagactcac tctccataaa tgctacgaat attaaacact 1320
 tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 aggaaatcac agggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
 cctttgagaa cctagaaatc atacgcggca ggaccaagca acatggtcag ttttctcttg 1560
 30 cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
 gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
 aactgtttgg gacctccggg cagaaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaaggccac aggccagggtc tgccatgcct tgtgctcccc cgagggctgc tggggcccgg 1800
 agcccaggga ctgctctctc tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
 35 gcaaccttct ggaggggtgag ccaagggtg ttgtggagaa ctctgagtgc atacagtgcc 1920
 acccagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
 gtatccagtg tgcccactac attgacggcc cccactgcgt caagacctgc ccggcaggag 2040

40

45

50

55

	tcatgggaga	aaacaacacc	ctgggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtggg	ggccctcctc	ttgctgctgg	2220
5	tgggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacacccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgctgg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	acatctccga	aagccaacaa	ggaaatcctc	gatgaagcct	2520
	acgtgatggc	cagcgtggac	aacccccacg	tgtgccgcct	gctgggcatac	tgcctcacct	2580
10	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcggctg	cctcctggac	tatgtccggg	2640
	aacacaaaga	caatattggc	tcccagtagc	tgtcaactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttggaggag	cgtcgcttgg	tgcaccgcga	cctggcagcg	aggaacgtac	2760
	tgggtgaaaac	accgcagcat	gtcaagatca	cagatttttg	gctggccaaa	ctgctgggtg	2820
	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcatttg	2880
	aatcaatttt	acacagaatc	tatacccacc	agagtgatgt	ctggagctac	ggggtgactg	2940
15	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	gggtcaagtgc	tggatgatag	acgcagatag	tgcgccaaag	ttccgtgagt	3120
	tgatcatcga	attctccaaa	atggccccgag	acccccagcg	ctaccttgtc	attcagggggg	3180
	atgaaagaat	gcatttgcca	agtcctacag	actccaactt	ctaccgtgcc	ctgatggatg	3240
20	aagaagacat	ggacgacgtg	gtggatgccg	acgagtacct	catcccacag	cagggcttct	3300
	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgca	accagcaaca	3360
	attccaccgt	ggcttgcatt	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
	gcttcttgca	gcgatacagc	tcagacccca	caggcgcctt	gactgaggac	agcatagacg	3480
	acaccttcct	cccagtgctt	gaatacataa	accagtccgt	tcccaaaagg	cccgtggctt	3540
	ctgtgcagaa	tcctgtctat	cacaatcagc	ctctgaaccc	cgcgcccagc	agagacccac	3600
25	actaccagga	ccccacacgc	actgcagtgg	gcaaccccg	gtatctcaac	actgtccagc	3660
	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaaagcaa	3780
	atggcatctt	taagggctcc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
	gcagtgaatt	tattggagca	tga				3863
30	<210> 747						
	<211> 3863						
	<212> DNA						
	<213> Homo sapiens						
35	<400> 747						
40							
45							
50							
55							

	cccgccgcag	cgccggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggagtcctcg	agctagcccc	ggcgccgcgc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctccg	cccaggtccc	cgctctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaa	aaagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccitttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
10	tagcagctct	atctaactat	gatgcgaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtc	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtg	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aaagtcctcc	gtgactgctg	ccacaaccag	tgtgtgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tggtctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatat	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagtt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaatagggtat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	ctcaaggag	ataatgagtg	1620
	gagatgtgat	aattttcagga	aacaaaaatt	tgtgtctatgc	aaatacaata	aactggaaaa	1680
25	aactgttttg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740

	gcaaggccac	aggccaggtc	tgccatgcct	tgtgtctccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgctctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
30	gcaaccttct	ggagggtgag	ccaaggaggt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctctct	ttgtgtctgg	2220
	tggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
35	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccagtg	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtg	gttcgggcag	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	acatctccga	aagccaacaa	ggaaatcctc	gatgaagcct	2520
	acgtgatggc	cagcgtggac	aacccccacg	tgtgccgctc	gctgggcact	tgcttcacct	2580
	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcggctg	cctcctggac	tatgtccggg	2640
40	aacacaaaga	caatattggc	tcccagtagc	tgctcaactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttggaggac	cgctcgcttg	tgcaccgcga	cctggcagcc	aggaacgtac	2760
	tggtgaaaac	accgcagcat	gtcaagatca	cagatttttg	gctggccaaa	ctgctgggtg	2820
	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcatttg	2880
	aatcaatttt	acacagaatc	tatacccacc	agagtgatgt	ctggagctac	ggggtgactg	2940
	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
45	cctccatcct	ggagaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	ggtcaagtgc	tggatgatag	acgcagatag	tcgccc aaag	ttccgtgagt	3120
	tgatcatcga	attctccaaa	atggccccag	acccccagcg	ctaccttgct	attcaggggg	3180
	atgaaagaat	gcatttgcca	agtcctacag	actccaactt	ctaccgtgcc	ctgatggatg	3240
	aagaagacat	ggacgcagtg	gtggatgccc	acgagtacct	catcccacag	cagggcttct	3300
	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtga	accagcaaca	3360
50	attccaccgt	ggcttgcat	gatagaaatg	ggctgcaaa	ctgtcccatc	aaggaagaca	3420
	gcttcttgca	gcgatacagc	tcagacccca	caggcgctct	gactgaggac	agcatagacg	3480
	acaccttctc	cccagtgcct	gaatacataa	accagtcctg	tcccaaaagg	cccgtggctg	3540
	ctgtgcagaa	tctgtcttat	cacaatcagc	ctctgaaccc	cgcgcccgag	acgagaccac	3600
	actaccagga	cccccacagc	actgcagtgg	gcaaccccca	gtatctcaac	actgtccagc	3660
	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
55	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaaagcaa	3780
	atggcatctt	taagggtctc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
	gcagtgaatt	tattggagca	tga				3863

<210> 748
 <211> 3863
 <212> DNA
 <213> Homo sapiens

5

<400> 748

10 cccggcgag cgcggccgca gcagcctccg ccccccgcac ggtgtgagcg cccgacgcgg 60
 cccgagggcggc cggaggtccc agctagcccc ggcgggccgc gccgcccaga ccggacgaca 120
 ggccacctcg tcggcggtccg cccgaggtccc cgcctcgccg ccaacgccac aaccaccgcg 180
 cagggccccc tgactccgtc cagtattgat cgggagagcc ggagcgagct cttcggggag 240
 cagcgatgcg accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgctct 300
 gcccggcgag tcgggctctg gaggaaaaga aagtttgcca aggcacgagt aacaagctca 360
 cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
 15 aggtgggtcct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
 taaagaccat ccaggagggtg gctgggttatg tcctcattgc cctcaacaca gtggagcgaa 540
 ttccitttga aaacctgcag atcatcagag gaaatatgta ctacgaaaat tcctatgcct 600
 tagcagtcct atctaactat gatgcaaata aaaccggact gaaggagctg cccatgagaa 660
 atttacagga aatcctgcat ggcgccgtgc ggttcagcaa caacctgcc ctgtgcaacg 720
 tggagagcat ccagtggcgg gacatagtcg gcagtgactt tctcagcaac atgtcgatgg 780
 20 acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
 gctggggtgc aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagt 900
 gctccgggcg ctgccgtggc aagtcccca gtgactgctg ccacaaccag tgtgctgcag 960
 gctgcacagg ccccggggag agcgactgcc tgggtctgccg caaattccga gacgaagcca 1020
 cgtgcaagga cacctgcccc ccactcatgc tctacaaccc caccacgtac cagatggatg 1080
 tgaaccccga gggcaaatac agctttgggtg ccacctgctg gaagaagtgt ccccgtaatt 1140
 25 atgtgggtgac agatcacggc tcgtgctgctc gagcctgtgg ggccgacagc tatgagatgg 1200
 aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgcaaa gtgtgtaacg 1260
 gaataggtat tgggtgaattt aaagactcac tctccataaa tgctacgaat attaaacact 1320
 tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 aggaaatcac aggggtttttg ctgattcagg cttggcctga aacaggacg gacctccatg 1500

30

35

40

45

50

55

	cctttgagaa	cctagaaaatc	atacgcggca	ggaccaagca	acatgggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aattttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
5	aactgtttgg	gacctccggt	cagaaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgtctccc	cgagggtgc	tggggcccgg	1800
	agcccaggga	ctgcgctctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaacctttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	acccagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tgtgaggaga	aaacaacacc	ctggctctga	agtacgcaga	cgccggccat	gtgtgccacc	2100
10	tgtgccatcc	aaactgcacc	tacggatgca	ctgggccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggatgggtggg	ggccctcctc	ttgtctgttg	2220
	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggagggt	gctgcaggag	agggagcttg	tggagcctct	tacaccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaagg	gagaaagtta	2460
15	aaattcccggt	cgctatcaag	acatctccga	aagccaacaa	ggaaatcctc	gatgaagcct	2520
	acgtgatggc	cagcgtggac	aacccccacg	tgtgccgcct	gctgggcac	tcctcacct	2580
	ccaccgtgca	gctcatcacg	cagctcatgc	ccttcggctg	cctcctggac	tatgtccggg	2640
	aacacaaaga	caatattggc	tccagttacc	tgctcaactg	gtgtgtgcag	atcgcaaagg	2700
	gcatgaacta	cttgaggagc	cgctcgcttg	tgaccgcga	cctggcagcc	aggaacgtac	2760
	tggtgaaaac	accgcagcat	gtcaagatca	cagattttgg	gctggccaaa	ctgtgggtg	2820
20	cggaagagaa	agaataccat	gcagaaggag	gcaaagtgcc	tatcaagtgg	atggcatttg	2880
	aatcaatttt	acacagaatc	tatacccacc	agagtgatgt	ctggagctac	ggggtgactg	2940
	tttgggagtt	gatgaccttt	ggatccaagc	catatgacgg	aatccctgcc	agcgagatct	3000
	cctccatcct	ggagaaaagga	gaacgcctcc	ctcagccacc	catatgtacc	atcgatgtct	3060
	acatgatcat	ggtcaagtgc	tggatgatag	acgcagatag	tcgccc aaag	ttccgtgagt	3120
	tgatcatcga	atttctccaaa	atggcccagag	acccccagcg	ctaccttgtc	attcaggggg	3180
25	atgaaagaat	gcatttgcca	agtcctacag	actccaactt	ctaccgtgcc	ctgatggatg	3240
	aagaagacat	ggacgacgtg	gtggatgccg	acgagtacct	catcccacag	cagggcttct	3300
	tcagcagccc	ctccacgtca	cggactcccc	tcctgagctc	tctgagtgc	accagcaaca	3360
	attccaccgt	ggcttgcat	gatagaaatg	ggctgcaaag	ctgtcccatc	aaggaagaca	3420
30	gcttcttgca	gcgatacagc	tcagacccca	caggcgcctt	gactgaggac	agcatagacg	3480
	acaccttcct	cccagtgctt	gaatacataa	accagtccgt	tcccaaaagg	cccgtggct	3540
	ctgtgcagaa	tcctgtctat	cacaatcagc	ctctgaacct	cgcgcccagc	agagaccac	3600
	actaccagga	ccccacagc	actgcagtgg	gcaacccga	gtatctcaac	actgtccagc	3660
	ccacctgtgt	caacagcaca	ttcgacagcc	ctgcccactg	ggcccagaaa	ggcagccacc	3720
	aaattagcct	ggacaaccct	gactaccagc	aggacttctt	tcccaaggaa	gccaaagcaa	3780
	atggcatctt	taagggtctc	acagctgaaa	atgcagaata	cctaagggtc	gcgccacaaa	3840
35	gcagtgaatt	tattggagca	tga				3863

<210> 749

<211> 3860

<212> DNA

40 <213> Homo sapiens

<400> 749

45

50

55

	cccggcgag	cgcgggcgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccagcgcg	60
	ccgaggcggc	cggagtcctcg	agctagcccc	ggcgggcgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctccg	cccaggtccc	cgccctcgccg	ccaacgccac	aaccaccgag	180
5	cacggccccc	tgactccgctc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggagggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
10	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
15	gctccggggcg	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgagg	agcgactggc	tgtctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaacct	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatat	agctttggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
20	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaaggggc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatatctctg	aaaaccgtaa	1440
25	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatgggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgtcctcc	cgagggtgctg	tgggggcccg	1800
	agcccaggga	ctgctgtctt	tgccggaatg	tcagccgagg	caggggaatgc	gtggacaagt	1860
30	gcaaccttct	ggagggtgag	ccaaggaggt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accaagatgtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggacagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatggtggg	ggccctcctc	ttgctgctgg	2220
35	tggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcggttcg	aagcgcacgc	2280
	tgcggaagct	gctgcaggag	agggagcttg	tggagcctct	tacaccagtc	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gttccgaaaag	ccaacaagga	aatcctcgat	gaagcctacg	2520
	tgatggccag	cgtggacaac	ccccacgtgt	gccgctgctg	gggcatctgc	ctcacctcca	2580
40	ccgtgcagct	catcacgcag	ctcatgccct	tcggctgcct	cctggactac	gtccgggaac	2640
	acaaagacaa	tattggctcc	cagtaccctg	tcaactgggtg	tgtgcagatc	gcaaagggca	2700
	tgaactactt	ggaggaccgt	cgcttggtgc	accgcgacct	ggcagccagg	aacgtactgg	2760
	tgaaaacacc	gcagcatgtc	aagatcacag	atcttgggct	ggccaaactg	ctgggtgcgg	2820
	aagagaaaaga	ataccatgca	gaaggaggca	aagtgcctat	caagtggatg	gcattggaat	2880
45	caattttaca	cagaatctat	acccaccaga	gtgatgtctg	gagctacggg	gtgactgttt	2940
	gggagttgat	gacctttgga	tccaagccat	atgacggaat	ccctgccagc	gagatctcct	3000
	ccatcctgga	gaaaggagaa	cgctccctc	agccacccat	atgtaccatc	gatgtctaca	3060
	tgatcatggt	caagtgtctg	atgatagacg	cagatagtcg	cccaaagttc	cgtgagttga	3120
	tcatcgaatt	ctccaaaatg	gcccgagacc	cccagcgcta	ccttgctcatt	cagggggatg	3180
	aaagaatgca	tttgccaagt	cctacagact	ccaacttcta	ccgtgccctg	atggatgaag	3240
50	aagacatgga	cgacgtggtg	gatgccgacg	agtacctcat	cccacagcag	ggcttcttca	3300
	gcagccccctc	cacgtcacgg	actccccctc	tgagctctct	gagtgcacac	agcaacaatt	3360
	ccaccgtggc	ttgcattgat	agaaaatggc	tgcaaagctg	tcccatcaag	gaagacagct	3420
	tcttgcagcg	atacagctca	gacccacag	gcgccttgac	tgaggacagc	atagacgaca	3480
	ccttcctccc	agtgcctgaa	tacataaacc	agtccgttcc	caaaaggccc	gctggctctg	3540
	tgcagaatcc	tgtctatcac	aatcagcctc	tgaaccccg	gcccagcaga	gtccacacat	3600
55	accagagccc	ccacagcact	gcagtgggca	accccgagta	tctcaacact	gaccagccca	3660
	cctgtgtcaa	cagcacattc	gacagccctg	cccactgggc	ccagaaaggc	agccaccaa	3720
	ttagcctgga	caacctgac	taccagcagg	acttctttcc	caaggaagcc	aagccaaatg	3780
	gcatctttta	gggtccaca	gctgaaaatg	cagaatacct	aagggtcgcg	ccacaaagca	3840
	gtgaatttat	tggagcatga					3860

<210> 750
 <211> 3869
 <212> DNA
 <213> Homo sapiens

5

<400> 750

```

10      cccggcgcgag  cgcgggccgca  gcagcctccg  ccccccgcac  ggtgtgagcg  cccgacgcgg  60
      ccgaggcggc  cggagtcccg  agctagcccc  ggcggccgcc  gccgcccaga  cccggacgaca  120
      ggccacctcg  tcggcgtccg  cccgagtcct  cgcctcgccg  ccaacgccac  aaccaccgcg  180
      cacggcccc  tgactccgtc  cagtattgat  cgggagagcc  ggagcgagct  cttcggggag  240
      cagcgatgcg  accctccggg  acggccgggg  cagcgctcct  ggcgctgctg  gctgcgctct  300
      gcccggcgag  tcgggctctg  gaggaaaaga  aagtttgcca  aggcacgagt  aacaagctca  360
      cgcagttggg  cacttttgaa  gatcattttc  tcagcctcca  gaggatgttc  aataactgtg  420
      aggtggtcct  tgggaatttg  gaaattacct  atgtgcagag  gaattatgat  ctttccttct  480
15      taaagaccat  ccaggagggtg  gctggttatg  tcctcattgc  cctcaacaca  gtggagcgaa  540
      ttcctttgga  aaacctgcag  atcatcagag  gaaatatgta  ctacgaaaat  tcctatgcct  600
      tagcagtctt  atctaactat  gatgcaaata  aaaccggact  gaaggagctg  cccatgagaa  660
      atttacagga  aatcctgcat  ggcgccgtgc  ggttcagcaa  caaccctgcc  ctgtgcaacg  720
      tggagagcat  ccagtggcgg  gacatagtca  gcagtgactt  tctcagcaac  atgtcgatgg  780
20      acttccagaa  ccacctgggc  agctgccaaa  agtgtgatcc  aagctgtccc  aatgggagct  840
      gctgggggtgc  aggagaggag  aactgccaga  aactgaccaa  aatcatctgt  gcccgagcgt  900
      gctccgggcg  ctgccgtggc  aagtcccca  gtgactgctg  ccacaaccag  tgtgctgcag  960

```

25

30

35

40

45

50

55

5 gctgcacagg cccccgggag agcgactgcc tggctctgcc caaattccga gacgaagcca 1020
 cgtgcaaggga cacctgcccc ccactcatgc tctacaaccc caccacgtac cagatggatg 1080
 tgaaccccgga gggcaaaatac agctttgggtg ccacctgcgt gaagaagtgt ccccgtaatt 1140
 atgtgggtgac agatcacggc tctgtcgtcc gagcctgtgg ggccgacagc tatgagatgg 1200
 aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgcaaa gtgtgtaacg 1260
 gaatagggtat tgggtgaattt aaagactcac tctccataaa tgctacgaat attaaacact 1320
 tcaaaaactg caccctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 aggaaatcac aggggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
 ccttttgagaa cctagaaatc atacgcgga ggaccaagca acatggtcag ttttctcttg 1560
 10 cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
 gagatgtgat aatttcaggga acaaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
 aactgtttgg gacctccggg cagaaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaaggccac aggccaggtc tgccatgcct tgtgtcctcc cgagggctgc tggggcccg 1800
 agcccaggga ctgctgtctc tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
 gcaaccttct ggagggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
 15 acccagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
 gtatccagtg tgcccactac attgacggcc cccactgcgt caagacctgc ccggcaggag 2040
 tcatgggaga aaacaacacc ctggtctgga agtacgcaga cgccggccat gtgtgccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctgggacagg tcttgaaggc tgtccaacga 2160
 atgggcctaa gatcccgtcc atcgccactg ggatgggtgg ggccctctc ttgctgctgg 2220
 20 tgggtggccct ggggatcggc ctcttcacgc gaaggcgcca catcgttcgg aagcgacgc 2280
 tgcgagggtc gctgcaggag agggagcttg tggagcctct tacaccagtg ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtg 2400
 gctccggtgc gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
 aaattcccgt cgctatcaag gaaccaacat ctccgaaagc caacaaggaa atcctcgatg 2520
 aagcctacgt gatggccagc gtggacaacc cccacgtgtg ccgcctgctg ggcactctgc 2580
 25 tcacctccac cgtgcagctc atcacgcagc tcatgccctt cggetgcctc ttggactatg 2640
 tccgggaaca caaagacaat attggctccc agtacctgct caactgggtg gtgcagatcg 2700
 caaagggcat gaactacttg gaggaccgtc gcttgggtga ccgcgacctg gcagccagga 2760
 acgtactggt gaaaacaccg cagcatgtca agatcacaga ttttgggctg gccaaactgc 2820
 tgggtgcgga agagaaaagaa taccatgcag aaggaggcaa agtgccctatc aagtggatgg 2880
 30 cattggaatc aattttacac agaactctata cccaccagag tgatgtctgg agctacgggg 2940
 tgactgtttg ggagtgtgat acctttggat ccaagccata tgacggaatc cctgccagcg 3000
 agatctcctc catcctggag aaaggagaac gcctccctca gccaccata tgtaccatcg 3060
 atgtctacat gatcatggtc aagtgtctgga tgatagacgc agatagtcgc ccaaagttcc 3120
 gtgagttgat catcgaattc tccaaaatgg cccgagaccc ccagcgctac cttgtcattc 3180
 aggggggatga aagaatgcat ttgccaaagtc ctacagactc caacttctac cgtgccctga 3240
 35 tggatgaaga agacatggac gacgtggtgg atgccgacga gtacctcatc ccacagcagg 3300
 gcttcttcag cagccccctc acgtcacgga ctccccctc gagctctctg agtgcaacca 3360
 gcaacaattc caccgtggct tgcattgata gaaatgggct gcaaagctgt cccatcaagg 3420
 aagacagctt cttgcagcga tacagctcag accccacagg cgccttgact gaggacagca 3480
 tagacgacac cttcctccca gtgcctgaat acataaacca gtccgttccc aaaaggccc 3540
 ctggctctgt gcagaatcct gtctatcaca atcagcctct gaaccccgcg ccagcagag 3600

 40 acccacacta ccaggacccc cacagcactg cagtgggcaa ccccgagtat ctcaacactg 3660
 tccagcccac ctgtgtcaac agcacattcg acagccctgc ccactgggcc cagaaaggca 3720
 gccaccaaata tagcctggac aacctgact accagcagga cttctttccc aaggaagcca 3780
 agccaaatgg catctttaag ggctccacag ctgaaaatgc agaataccta agggtcgcgc 3840
 cacaaagcag tgaatttatt ggagcatga 3869

45 <210> 751
 <211> 3866
 <212> DNA
 <213> Homo sapiens

50 <400> 751

55

	cccgccgcag	cgccggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcg	60
	ccgaggccgc	cggagtcccc	agctagcccc	ggcgccgcgc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcgctccg	cccaggtccc	cgccctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgctc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcy	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gccccggcag	tcgggctctg	gaggaaaaga	aagtttgccca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aaataactgtg	420
	aggtgggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
10	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagctctt	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
15	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgggag	agcgactgcc	tggctgtccg	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccgga	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
20	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcgaag	tgtaaagaagt	gcgaaggggc	ttgccgcaaa	gtgtgtaacg	1260
	gaatagggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
25	ccittgagaa	cctagaaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	ccttcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgtctccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
30	gcaaccttct	ggagggtgag	ccaaggaggt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctctct	ttgtgtctgg	2220
35	tgggtggccct	ggggatcgcc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtgc	gttcgggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaaccatctc	cgaaagccaa	caaggaaatc	ctcgatgaag	2520
	ctatcgatgat	ggccagcggt	gacaaccccc	acgtgtgccg	cctgtctggc	atctgcctca	2580
40	cctccaccgt	gcagctcatc	acgcagctca	tgcccttcgg	ctgcctcctg	gactatgtcc	2640
	gggaacacaa	agacaatatt	ggctcccagt	acctgtctaa	ctggtgtgtg	cagatcgcaa	2700
	agggcatgaa	ctacttgag	gaccgtcgct	tgggtgcaccg	cgacctggca	gccagggaacg	2760
	tactgggtgaa	aacaccgcag	catgtcaaga	tcacagattt	tgggctggcc	aaactgtctg	2820
	gtgcggaaga	gaaagaatac	catgcagaag	gaggcaaagt	gcctatcaag	tggatggcat	2880
	tggaatcaat	tttacacaga	atctataccc	accagagtga	tgtctggagc	tacggggtga	2940
45	ctgtttggga	gttgatgacc	tttggatcca	agccatatga	cggaatccct	gccagcgaga	3000
	tctcctccat	cctggagaaa	ggagaacgcc	tccctcagcc	acccatatgt	accatcgatg	3060
	tctacatgat	catggtcaag	tgctggatga	tagacgcaga	tagtcgccca	aagttccgtg	3120
	agttgatcat	cgaattctcc	aaaatggccc	gagaccccca	gcgtacctt	gtcattcagg	3180
	gggatgaaag	aatgcatttg	ccaagtccca	cagactccaa	cttctaccgt	gccctgatgg	3240
	atgaagaaga	catggacgac	gtgggtggatg	ccgacgagta	cctcatccca	cagcagggtc	3300
50	tcctcagcag	cccctccacg	tcacggactc	ccctcctgag	ctctctgagt	gcaaccagca	3360
	acaattccac	cgtggcttgc	attgatagaa	atgggctgca	aagctgtccc	atcaaggaag	3420
	acagcttctt	gcagcgatac	agctcagacc	ccacaggcgc	cttgactgag	gacagcatag	3480
	acgacacctt	cctcccagtg	cctgaatata	taaaccagtc	cgttcccaaa	aggccccgtg	3540
	gctctgtgca	gaatcctgtc	tatcacaaatc	agcctctgaa	ccccgcgccc	agcagagacc	3600
	cacactacca	ggacccccac	agcactgcag	tgggcaaccc	cgagtatctc	aacactgtcc	3660
	agcccacctg	tgtcaacagc	acattcgaca	gccctgccca	ctgggcccag	aaaggcagcc	3720
55	accaaattag	cctggacaac	cctgactacc	agcaggactt	ctttcccaag	gaagccaagc	3780
	caaatggcat	ctttaagggc	tccacagctg	aaaatgcaga	atacctaagg	gtcgcgccac	3840
	aaagcagtga	atttattgga	gcatga				3866

<210> 752
 <211> 3854
 <212> DNA
 <213> Homo sapiens

5

<400> 752

10	cccg	ggcg	cag	cg	ggccg	ca	gc	agcct	ccg	cccc	cg	ca	ggtgtgagcg	cccg	acgcgg	60
	ccg	aggcg	ggc	cg	gagtc	ccc	ag	ctagcccc	ggc	ggccg	cc	gccgccc	aga	ccgg	acgaca	120
	ggcc	acctcg	tc	ggcgt	ccg	ccc	gagtc	ccc	cg	cctcg	ccg	cca	acgcc	ac	aacc	accg
	cac	ggcccc	tg	actcc	gtc	cag	tatt	gat	cg	ggag	agcc	gg	agcgag	ct	cttc	ggggag
	cag	cgatgcg	acc	ctcc	ggg	ac	ggcc	gggg	cag	cgctc	ct	gg	cgctg	ctg	gctg	cgctct
	gccc	ggcgag	tc	gggct	ctg	gag	gaaa	aaga	aag	tttg	cca	agg	cacgag	t	aaca	agctca
	cg	cagttggg	ca	ctttt	gaa	gat	cat	ttttc	tc	agcct	cca	gag	gatgttc	aata	actgtg	420

15

20

25

30

35

40

45

50

55

	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggagggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
5	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggcg	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
10	gctgcacagg	cccccgggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtaaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggat	tgggtgaattt	aaagactcac	tctccataaa	tgtacgaat	attaaacact	1320
	tcaaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccgggtggca	tttaggggtg	1380
15	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccttttgagaa	cctagaaaatc	atacgcgcca	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
20	aactgttttg	gacctccggt	cagaaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	acccagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
25	tcatgggaga	aaacaacacc	ctggcttggga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtggg	ggccctcctc	ttgctgctgg	2220
	tgggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcggagggt	gctgcaggag	agggagcttg	tggagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
30	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaaccct	cgatgaagcc	tacgtgatgg	2520
	ccagcgtgga	caacccccac	gtgtgccgcc	tgctgggcat	ctgcctcacc	tccaccgtgc	2580
	agctcatcac	gcagctcatg	cccttcgggt	gcctcctgga	ctatgtccgg	gaacacaaag	2640
	acaatatttg	ctcccagtac	ctgtctcaact	gggtgtgtga	gatcgcaaag	ggcatgaact	2700
	acttggaggga	ccgtcgcttg	gtgcaccgcy	acctggcagc	caggaacgta	ctggtgaaaa	2760
35	caccgcagca	tgtcaagatc	acagattttg	ggctggccaa	actgctgggt	gcggaagaga	2820
	aagaatacca	tgcagaagga	ggcaaaagtgc	ctatcaagtg	gatggcattg	gaatcaattt	2880
	tacacagaat	ctatacccac	cagagtgatg	tctggagcta	cggggtgact	gtttgggagt	2940
	tgatgacctt	tggatccaag	ccatatgacg	gaatccctgc	cagcgagatc	tcctccatcc	3000
	tggagaaaagg	tgaacgcctc	cctcagccac	ccatatgtac	catcgatgtc	tacatgatca	3060
	tgggtcaagtg	ctggatgata	gacgcagata	gtcgcccaaa	gttccgtgag	ttgatcatcg	3120
40	aattctccaa	aatggcccga	gacccccagc	gctaccttgt	cattcagggg	gatgaaagaa	3180
	tgcattttgcc	aagtccctaca	gactccaact	tctaccgtgc	cctgatggat	gaagaagaca	3240
	tggagcacgt	ggtggatgcc	gacgagtacc	tcacccca	gcagggcttc	ttcagcagcc	3300
	cctccacgtc	acggactccc	ctcctgagct	ctctgagtgc	aaccagcaac	aattccaccg	3360
	tggcttgcac	tgatagaaat	gggctgcaaa	gctgtcccat	caaggaagac	agcttcttgc	3420
	agcgatacag	ctcagacccc	acaggcgctt	tgactgagga	cagcatagac	gacaccttcc	3480
45	tcacagtgcc	tgaatacata	aaccagtccg	ttcccaaaag	gcccgcgtggc	tctgtgcaga	3540
	atcctgtcta	tcacaatcag	cctctgaacc	ccgcgcccag	cagagaccca	cactaccagg	3600
	acccccacag	cactgcagtg	ggcaaccccc	agtatctcaa	cactgtccag	cccacctgtg	3660
	tcaacagcac	attcgacagc	cctgcccact	gggcccagaa	aggcagccac	caaattagcc	3720
	tggacaaccc	tgactaccag	caggacttct	ttcccaagga	agccaagcca	aatggcatct	3780
50	ttaagggctc	cacagctgaa	aatgcagaat	acctaagggt	cgcgccacaa	agcagtgaat	3840
	ttattggagc	atga					3854

<210> 753

<211> 3854

<212> DNA

55 <213> Homo sapiens

<400> 753

cccggcgag cgcgggcgca gcagcctccg ccccccgcac ggtgtgagcg cccgacgcgg 60
 ccgagggcggc cggagtcgcc agctagcccc ggcggccgcc gccgcccaga ccggacgaca 120

5 ggccacctcg tcggcgctccg cccgagtcct cgctctcgcc ccaacgccac aaccaccgcg 180
 cacggccccc tgactccgtc cagtattgat cgggagagcc ggagcgagct ctctggggag 240
 cagcgatgag accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgctct 300
 gccggcgag tcgggctctg gaggaaaaga aagtttgcca aggcacgagt aacaagctca 360
 cgaggttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
 10 aggtggtcct tgggaatttg gaaattacct atgtgcagag gaattatgat ctctccttct 480
 taaagaccat ccaggaggtg gctggttatg tcctcattgc cctcaacaca gtggagcgaa 540
 ttccitttga aaacctgcag atcatcagag gaaatatgta ctacgaaaat tcctatgcct 600
 tagcagctct atctaactat gatgcaaata aaaccggact gaaggagctg cccatgagaa 660
 atttacagga aatcctgcat ggcgccgtgc ggttcagcaa caaccctgcc ctgtgcaacg 720
 tggagagcat ccagtggcgg gacatagtca gcagtgaact tctcagcaac atgtcgatgg 780
 15 acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
 gctgggggtg aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagt 900
 gctccggggc ctgcccgtggc aagtccccc gtgactgtcg ccacaaccag tgtgctgcag 960
 gctgcacagg ccccgggag agcgactgcc tggctcgccg caaattccga gacgaagcca 1020
 cgtgcaagga cacctgcccc ccactcatgc tctacaaccc caccacgtac cagatggatg 1080
 tgaacccgga gggcaaatac agctttgggt ccacctgcgt gaagaagtgt ccccgtaatt 1140
 atgtggtgac agatcacggc tcgtgcgtcc gagcctgtgg ggccgacagc tatgagatgg 1200
 20 aggaagacgg cgtccgcaat tgtaagaagt gcgaagggcc ttgccgcaa gtgtgtaacg 1260
 gaataggtat tgggtgaatt aaagactcac tctcataaaa tgctacgaat aactggaaaa 1320
 tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 aggaaatcac agggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
 cctttgagaa cctagaaatc atacgcggca ggaccaagca acatggtcag ttttctcttg 1560
 25 cagtcgtcag cctgaacata acatccttgg gattacgtc cctcaaggag ataagtgatg 1620
 gagatgtgat aatttcagga acaaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
 aactgtttgg gacctccggt cagaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaaggccac aggccaggtc tgccatgcct tgtgtcccc cgagggctgc tggggcccg 1800
 agcccaggga ctgctctct tgcgggaatg tcagccgagg cagggaatgc gtggacaagt 1860
 30 gcaaccttct ggaggggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
 accagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
 gtatccagtg tgcctactac attgacggcc cccactgcgt caagacctgc ccgagaggag 2040
 tcatgggaga aaacaacacc ctggcttgga agtacgcaga cgccggccat gtgtgccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctgggcccag tcttgaaggc tgtccaacga 2160
 atgggcctaa gatcccgtcc atcgccactg ggatggtggg ggccctctc ttgctgctgg 2220
 35 tgggtggccct ggggatcggc ctcttcattg gaaggcgcca catcgttcgg aagcgacgc 2280
 tgcggagcgt agtgacgttg agggagcttg tggagcctct tacaccagt ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtg 2400
 gctccggtgc gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
 aaattcccgt cgctatcaag gaattaagag aagcaacact cgatgaagcc tacgtgatgg 2520
 ccagcgtgga caacccccac gtgtgccgcc tgtgggcat ctgcctcacc tccaccgtgc 2580
 40 agctcatcac gcagctcatg cccttcgggt gcctctgga ctatgtccgg gaacacaaag 2640
 acaatatttg ctcccagtac ctgctcaact ggtgtgtgca gatcgcaaag ggcataaact 2700
 acttgaggga ccgtcgcttg gtgcaccgag acctggcagc caggaacgta ctggtgaaaa 2760
 caccgcagca tgtcaagatc acagattttg ggctggccaa actgctgggt gcggaagaga 2820
 aagaatacca tgcagaagga ggcaaatgct ctatcaagtg gatggcattg gaatcaattt 2880
 45 tacacagaat ctataccac cagagtgatg tctggagcta cggggtgact gtttgggagt 2940
 tgatgacctt tggatccaag ccatatgacg gaatccctgc cagcgagatc tcctccatcc 3000
 tggagaaaagg agaacgcctc cctcagccac ccatatgtac catcgatgtc tacatgatca 3060
 tggatcaagt ctggatgata gacgcagata gtcgcccata gttccgtgag ttgatcatcg 3120
 aatttcccaa aatggcccga gacccccagc gctaccttgt cattcagggg gatgaaagaa 3180
 tgcattttgcc aagtccata gactccaact tctaccgtgc cctgatggat gaagaagaca 3240
 tggagcagct ggtggatgcc gacgagtacc tcattccaca gcagggttc ttcagagcc 3300
 50 cctccacgtc acggactccc ctctgagct ctctgagtc aaccagcaac aattccaccg 3360
 tggcttgcat tgatagaaat gggctgcaaa gctgtcccat caaggaagac agcttcttgc 3420
 agcgatacag ctgagacccc acaggcgctt tgactgagga cagcatagac gacaccttcc 3480
 tcccagtgcc tgaatacata aaccagtccg ttcccaaaag gccgctggc tctgtgcaga 3540
 atcctgtcta tcacaatcag cctctgaacc cgcgcccag cagagacca cactaccagg 3600
 55 acccccacag cactgcagtg ggcaacccc agtatctcaa cactgtccag cccactgtg 3660
 tcaacagcac attcgacag cctgcccact gggcccagaa aggcagccac caaattagcc 3720
 tggacaaccc tgactaccag caggacttct ttcccaagga agccaagcca aatggcatct 3780
 ttaagggctc cacagctgaa aatgcagaat acctaagggt cgcgccacaa agcagtgaat 3840
 ttattggagc atga 3854

<210> 754
<211> 3887
<212> DNA
<213> Homo sapiens

5

<400> 754

10

15

20

25

30

35

40

45

50

55

	cccgccgcag	cgccggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggagtcgcc	agctagcccc	ggcgcccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgccctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aaagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggagggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccitttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
10	tagatgcttt	atctaactat	gatgcataata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtc	gcagtgaatt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aaagtcctcca	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaatat	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagat	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgggca	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttccagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	acccagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
30	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggatgggtggg	ggccctcctc	ttgtgtctgg	2220
	tgggtggcct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgcttcg	cacgcacgc	2280
	tgccggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccagct	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
35	gctccgggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tgacagcgt	ggacaacccc	cagtggtgcc	2580
	gcctgtcggg	cactgcctc	acctccaccg	tgcactcat	cacgcagctc	catgcccttcg	2640
	gctgcctcct	ggactatgtc	cgggaacaca	aagacaatat	tggctcccag	tacctgtctca	2700
	actggtgtgt	gcagatcgca	aagggcagatg	actacttgga	ggaccgtcgc	ttggtgcacc	2760
40	gcgacctggc	agccagggaac	gtactgggtga	aaacaccgca	gcatgtcaag	atcacagatt	2820
	ttgggctggc	caaactgctg	ggtgcggaag	agaaagaata	ccatgcagaa	ggaggcaag	2880
	tgccatcatga	gtggatggca	ttggaatcaa	ttttacacag	aatctatacc	caccagagt	2940
	atgtctggag	ctacgggggtg	actgtttggg	agttgatgac	ctttggatcc	aagccatag	3000
	acggaatccc	tgccagcgag	atctcctcca	tcctggagaa	aggagaacgc	ctccctcagc	3060
	cacccatatg	taccatcgat	gtctacatga	tcagtgtcaa	gtgctggatg	atagacgcag	3120
45	atagtcgccc	aaagtccctg	gagttgatca	tcgaattctc	caaatggcc	cgagaccccc	3180
	agcgctacct	tgtcattcag	ggggatgaaa	gaatgcattt	gccaagtctt	acagactcca	3240
	acttctaccg	tgccctgatg	gatgaagaag	acatggacga	cgtgggtggat	gccgacgagt	3300
	acctcatccc	acagcagggc	ttcttcagca	gccccctccac	gtcacggact	cccctcctga	3360
	gctctctgag	tgcaaccagc	aacaattcca	ccgtggcttg	cattgataga	aatgggctgc	3420
	aaagctgtcc	catcaaggaa	gacagcttct	tgacgcgata	cagctcagac	cccacaggcg	3480
50	ccttgactga	ggacagcata	gacgacacct	tcctcccagt	gcctgaatac	ataaaccagt	3540
	ccgttcccaa	aaggcccgct	ggctctgtgc	agaattcctgt	ctatcacaat	cagcctctga	3600
	accccgcggc	cagcagagac	ccacactacc	aggaccccca	cagcactgca	gtgggcaacc	3660
	ccgagtatct	caacactgtc	cagcccacct	gtgtcaacag	cacattcgac	agccctgccc	3720
	actgggcccc	gaaaggcagc	caccaaatta	gcctggacaa	ccctgactac	cagcaggact	3780
	tctttcccaa	ggaagccaag	ccaaatggca	tctttaaggg	ctccacagct	gaaaatgcag	3840
55	aatacctaag	ggtcgcgcca	caaagcagtg	aatttattgg	agcatga		3887

<210> 755

<211> 3884

<212> DNA

<213> Homo sapiens

<400> 755

5

10

15

20

25

30

35

40

45

50

55

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	gggtgtgagcg	cccgcgcgg	60
	ccgaggcggc	cggaggtccc	agctagcccc	ggcgggccg	gcccgcagca	ccggacgaca	120
	ggccacctcg	tcggcgctcg	cccaggtccc	cgccctcgcc	ccaacgccac	aaccaccg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gccccggcag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tggaattttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggagggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
10	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tgagagagcat	ccagtgccgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagagggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
15	gctccggggc	ctgcccgtgg	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	cccccgagg	agcgactg	tgtctgccc	caaatttcca	gacgaagcca	1020
	cgtgcaagga	ccctgcccc	ccactcatgc	tgtaacaacc	caccacgtac	catatggatg	1080
	tgaaccccga	gggcaaatat	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	tgaataggtat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	gaaaaaactg	cacctccatc	agtgccgtatc	tccacatcct	gcccgtggca	cttggtggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaaatc	atacgccgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgtatg	1620
25	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgttttg	gacctccgg	cagaaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggagggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	acccagagt	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
30	gtatcccagt	tgcccactac	attgacggcc	ccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggctctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgctgctgg	2220
	tggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgctcgg	aagcgcacgc	2280
	tgccggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccagct	ggagagagctc	2340
35	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacccccac	gtgtgccgcc	2580
	tgctgggcat	ctgcctcacc	tccaccgtgc	agctcatcac	gcagctcatg	cccttcggct	2640
40	gctccttggg	ctatgtccgg	gaacacaaag	acaattattg	ctcccagtag	ctgctcaact	2700
	gggtgtgtgca	gatcgcaaa	ggcatgaact	acttggagga	ccgtcgcttg	gtgcaccg	2760
	acctggcagc	caggaacgta	ctggtgaaaa	caccgcagca	tgtcaagatc	acagattttg	2820
	ggctggccaa	actgctgggt	gcggaagaga	agaataacca	tgcagaagga	ggcaaagtgc	2880
	ctatcaagt	gatggcattg	gaatcaattt	tacacagaat	ctatacccac	cagagtgtatg	2940
	tctggagcta	cgggggtgact	gtttgggagt	tgatgacctt	tggatccaa	ccatagacg	3000
45	gaatccctgc	cagcgagatc	tcctccatcc	tggagaaagg	agaacgcctc	cctcagccac	3060
	ccatatgtac	catcgatgtc	tacatgatca	tggtcaagt	ctggatgata	gacgcagata	3120
	gtcgcacaaa	gttccgtgag	ttgatcatcg	aattctccaa	aatggcccga	gacccccagc	3180
	gctaccttgt	cattcagggg	gatgaaagaa	tgcatttgcc	aagtcctaca	gactccaact	3240
	tctaccgtgc	cctgatggat	gaagaagaca	tggagcagct	ggatggatgcc	gacgagtacc	3300
50	tcaccccaca	gcagggcctt	ttcagcagcc	ctccacgtc	acggactccc	ctcctgagct	3360
	ctctgagtgc	aaccagcaac	aattccaccg	tggcttgcat	tgatagaaat	gggctgcaaa	3420
	gctgtcccat	caaggaagac	agcttcttgc	agcgatacag	ctcagacccc	acaggcgctt	3480
	tgactgagga	cagcatagac	gacaccttcc	tcccagtgcc	tgaatacata	aaccagtccg	3540
	ttcccaaaaag	gccccgtggc	tctgtgcaga	atcctgtcta	tcacaatcag	cctctgaacc	3600
	cccgccccag	cagagaccga	cactaccagg	acccccacag	cactgcagtg	ggcaaccctg	3660
	agtatctcaa	cactgtccag	cccacctgtg	tcaacagcac	attcgacagc	cctgcccact	3720
55	gggcccagaa	aggcagccac	caaattagcc	tggacaaccc	tgactaccag	caggacttct	3780
	ttcccaagga	agccaagcca	aatggcatct	ttaagggctc	cacagctgaa	aatgcagaat	3840
	acctaaggg	cgcgccacaa	agcagtgaat	ttattggagc	atga		3884

<210> 756
 <211> 3878
 <212> DNA
 <213> Homo sapiens

5

<400> 756

	ccccggcgag	cgcgggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
10	ggccacctcg	tcggcggtccg	cccaggtccc	cgctcgcgcg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgtctt	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
15	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccttttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
20	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gccacgacgt	900
	gctccggggcg	ctgccgtggc	aagtccccca	gtgactgctg	ccacaaccag	tgtgtgtcag	960
	gctgcacagg	cccccgggag	agcgactgcc	tggctgtccg	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaacccccga	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
25	aggaagacgg	cgtccgcaag	tgttaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
30	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgtatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgtctatg	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgtctccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgctgtctt	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
35	gcaaccttct	ggagggtgag	ccaaggaggt	ttgtggagaa	ctctgagtgc	atcacgtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctggggcagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctcttc	tgtgtctgtg	2220
	ttgttggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
40	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacacccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtgtg	2400
	gctccgggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cgcctgtgtg	2580
45	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtct	aactggtgtg	2700
	tgcagatcgc	aaagggcag	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
	cagccaggaa	cgtactggtg	aaaacactgc	agcatgtcaa	gatcacagat	tttgggctgg	2820
	ccaaactgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	attggaatca	attttacaca	gaatctatac	ccaccagagt	gatgtcttga	2940
50	gctacggggg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccacccatat	3060
	gtaccatcga	tgtctacatg	atcatgggtca	agtgtgtgat	gatagacgca	gatagtgcgc	3120
	caaagttccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agccccctca	cgtcacggac	tccccctctg	agctctctga	3360
55	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctccag	tgcctgaata	cataaaccag	tccgttccca	3540

5 aaaggcccg tggctctgtg cagaatcctg tctatcaca tcagcctctg aaccccgcg 3600
 ccagcagaga cccacactac caggaccccc acagcactgc agtgggcaac cccgagtatc 3660
 tcaacactgt ccagcccacc tgtgtcaaca gcacattcga cagccctgcc cactgggccc 3720
 agaaaggcag ccaccaaatt agcctggaca accctgacta ccagcaggac ttctttccca 3780
 aggaagccaa gccaaatggc atctttaagg gtccacagc tgaaaatgca gaatacctaa 3840
 gggtcgcgcc acaaagcagt gaatttattg gagcatga 3878

10 <210> 757
 <211> 3878
 <212> DNA
 <213> Homo sapiens

15 <400> 757

20

25

30

35

40

45

50

55

	ccccggcgag	cgcgggccgca	gcagcctccc	ccccccgcac	ggtgtgagcg	cccgacgchg	60
	ccgagggcggc	cggagtcctcg	agctagcccc	ggcgggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcgc	tcggcgctccg	cccaggtccc	cgctctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgctc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggtcct	tggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggaggtg	gctggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
10	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
	tagcagctct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccagtgaaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagagggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
15	gctccggggcg	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	ccccggggag	agcgactgcc	tggcttgccg	caaattccga	gacgaagcca	1020
	cgtgcaaggga	cacctggccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccgga	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagacgg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgaaa	gtgtgtaacg	1260
20	tgataggtat	tggtgaattt	aaagactcac	tccataaaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	cctggcgatc	tcccatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
25	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgctgtctt	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	acccagagtgc	cttgccctcag	gcatgaaca	tcacctgcac	aggacgggga	ccagacactt	1980
30	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctcctc	ttgctgctgg	2220
	tggtggccct	ggggatcggc	ctcttcctgc	gaaggcgcca	catcggtcgg	aagcgcacgc	2280
	tgcggaagct	ctgcagggag	agggagcttg	tggagcctct	tacaccagct	ggagaagctc	2340
35	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtggt	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cgctgtctgg	2580
	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcttcc	2640
	tggaactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtct	aactgggtgtg	2700
40	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
	cagccaggaa	cgtactggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggcggg	2820
	ccaaactgct	gggtgcgga	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	attggaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacgggggt	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccaccatat	3060
45	gtaccaatcga	tgtctacatg	atcatggtca	agtgtggtg	gatagacgca	gatagtcgcc	3120
	caaagttccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtggtgga	tgccgacgag	tacctcatcc	3300
50	cacagcaggg	cttcttcagc	agcccccca	cgtcacggac	tccccctctg	agctctctga	3360
	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
	ccatcaaggga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgctgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggctctgtg	cagaatcctg	tctatcacia	tcagcctctg	aaccccgcg	3600
	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccgagtatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactgggccc	3720
55	agaaaggcag	ccaccaaat	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gctccacagc	tgaaaatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaatttattg	gagcatga			3878

<210> 758
 <211> 3878
 <212> DNA
 <213> Homo sapiens

5

<400> 758

	cccggcgag	cgcgccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgaggcggc	cggagtcgcc	agctagcccc	ggcgcccgcc	gccgcccaga	ccggacgaca	120
10	ggccacctcg	tcggcgctcc	cccagatccc	cgccctcgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgct	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
15	taaagaccat	ccaggaggtg	gctgggttat	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
	tagcagctct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
20	gctggggtgc	aggagagggg	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
	gctgcacagg	ccccggggag	agcgactgcc	tggtctgccg	caaattccga	gacgaagcca	1020
	cgtgcaaggga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccca	gggcaaaata	agctttgggt	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
25	aggaagacgg	cgtccgcaag	tgtaaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatatctctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaaatc	atagcgggca	ggaccaagca	acatggtcag	ttttctcttg	1560
30	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
	aactgtttgg	gacctccggg	cagaaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tgggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaaatg	tcaggcagg	cagggaaatg	gtggagcagt	1860
	gcaacctttct	ggagggtgag	ccaagggaat	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
35	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
	tcatgggaga	aaacaacacc	ctggtctgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgctc	atcgccactg	ggatgggtgg	ggccctcttc	ttgctgctgg	2220
	tgggtggccct	ggggatcggc	ctcttcacgc	gaaggcgcca	catcgttcgg	aagcgacgc	2280
40	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacaccagct	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccgggtg	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cgctgctg	2580
	gcactgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
45	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtc	aactgggtgtg	2700
	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
	cagccaggaa	cgtactggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggcccgg	2820
	ccaaactgct	gggtgcggaa	gagaaaagaat	accatgcaga	aggaggcaaa	gtgccatatca	2880
	agtggatggc	atttgaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacggggt	gactgtttgg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
50	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccaccatat	3060

55

```

5      gtaccatcga tgtctacatg atcatgggtca agtgctggat gatagacgca gatagtcgcc 3120
      caaagttccg tgagttgatc atcgaattct ccaaaatggc ccgagacccc cagcgctacc 3180
      ttgtcattca gggggatgaa agaatgcatt tgccaagtcc tacagactcc aacttctacc 3240
      gtgccctgat ggatgaagaa gacatggacg acgtgggtga tgccgacgag tacctcatcc 3300
      cacagcaggg cttcttcagc agcccccca cgtcacggac tccccctctg agctctctga 3360
      gtgcaaccag caacaattcc accgtggctt gcattgatag aaatgggctg caaagctgtc 3420
      ccatcaagga agacagcttc ttgcagcgat acagctcaga cccacagggc gccttgactg 3480
      aggacagcat agacgacacc ttcttcccag tgcctgaata cataaaccag tccgttccca 3540
      aaaggcccg c tggctctgtg cagaatcctg tctatcaca tcagcctctg aaccccgcg 3600
      ccagcagaga cccacactac caggaccccc acagactgc agtgggcaac cccgagtatc 3660
10     tcaacactgt ccagcccacc tgtgtcaaca gcacattcga cagccctgcc cactgggccc 3720
      agaaaggcag ccaccaaatt agcctggaca accctgacta ccagcaggac ttctttccca 3780
      aggaagccaa gccaaatggc atctttaagg gctccacagc tgaaaatgca gaatacctaa 3840
      gggtcgcgcc acaaagcagt gaattttatt gagcatga 3878

```

```

15     <210> 759
      <211> 3878
      <212> DNA
      <213> Homo sapiens

```

```

20     <400> 759

```

```

25

```

```

30

```

```

35

```

```

40

```

```

45

```

```

50

```

```

55

```

	cccggcgag	cgcggccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgacgcgg	60
	ccgagggcggc	cggaggtccc	agctagcccc	ggcggccgcc	gccgcccaga	ccggacgaca	120
	ggccacctcg	tcggcggtccg	cccaggtccc	cgccctcgccg	ccaacgccac	aaccaccgcg	180
5	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
	cagcgatgcg	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaa	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	taaagaccat	ccaggagggtg	gctgggttatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttccitttga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaa	tcctatgcct	600
10	taagcagctt	atctaactat	gatgcaaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgactt	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gcccagcagt	900
	gctccggggc	ctgccgtggc	agtcctccca	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tgttctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatat	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgagatgg	1200
	aggaagagcg	cgtccgcaag	tgtagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaatagagtg	tggtagaatt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaaatcac	aggggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	ccitttgagaa	cctagaaatc	atacgcgga	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	ctcaaggagc	atattctatg	1620
	gagatgtgat	aattttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccaggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
30	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatccccgtcc	atcgccactg	ggatgggtggg	ggccctcctc	ttgtgtctgg	2220
	tgggtggcct	ggggatcggc	ctcttcattg	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
	tgcgagggtc	gctgcaggag	agggagcttg	tggagcctct	tacaccaggt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
35	gctccgggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccgt	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cgctgtcttg	2580
	gcactctgct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtct	aactgggtgtg	2700
	tgcagatcgc	aaagggcag	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
40	cagccaggaa	cgtactggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggcggg	2820
	ccaaactgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	attggaatca	attttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
45	gctacggggg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccaccatat	3060
	gtaccatcga	tgtctacatg	atcatggtca	agtgtggtat	gatagacgca	gatagtgcgc	3120
	caaagtcccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtggtgga	tgccgacgag	tacctcatcc	3300
	cacagcaggg	cttcttcagc	agcccccca	cgtcacggac	tccccctctg	agctctctga	3360
50	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctccag	tgcctgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggctctgtg	cagaatcctg	tctatcacaa	tcagcctctg	aaccctcgcg	3600
	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccagtatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	caactgggcc	3720
55	agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gtcccacagc	tgaatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaatttattg	gagcatga			3878

<210> 760
 <211> 3878
 <212> DNA
 <213> Homo sapiens

5

<400> 760

10 cccggcgag cgcggccgca gcagcctccg ccccccgcac ggtgtgagcg cccgacgagg 60
 ccgaggcggc cggagtcccg agctagcccc ggcgcccgcc gccgcccaga ccggacgaca 120
 ggccacctcg tcggcgctcg cccgagtccc cgctctcgccg ccaacgccac aaccaccgag 180
 cacggccccc tgactccgct cagtattgat cgggagagcc ggagcgagct cttcggggag 240
 cagcgatgag accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgctct 300
 gcccggcgag tcgggctctg gaggaaaaga aagtttgcca aggcacgagt aacaagctca 360
 cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
 15 aggtggctct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
 taaagaccat ccaggagggt gctgggttatg tcctcattgc cctcaacaca gtggagcgaa 540
 ttccitttga aaacctgcag atcatcagag gaaatatgta ctacgaaaat tcctatgcct 600
 tagcagtcct atctaactat gatgcaaata aaaccggact gaaggagctg cccatgagaa 660
 atttacagga aatcctgcat ggcgccgtgc ggttcagcaa caacctgcc ctgtgcaacg 720
 tggagagcat ccagtggcgg gacatagtcg gcagtgactt tctcagcaac atgtcgatgg 780
 20 acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
 gctgggggtg aggagaggag aactgccaga aactgaccaa aatcatctgt gccagcagt 900
 gctccggggc ctgcccgtgg aagtccccc gtgactgctg ccacaaccag tgtgtgcag 960
 gctgcacagg ccccggggag agcgactgcc tggctgccc ccaattccga gacgaagcca 1020
 cgtgcaagga cacctgcccc ccactcatgc tctacaaccc caccacgtac cagatggatg 1080
 tgaaccccga gggcaaatac agctttgggt ccacctgctg gaagaagtgt ccccgtaatt 1140
 25 atgtgggtgac agatcacggc tcgtgcgtcc gagcctgtgg ggccgacagc tatgagatgg 1200
 aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgaaa gtgtgtaacg 1260
 gaatagggtat tgggtaattt aaagactcac tctccataaa tgctacgaat attaaacact 1320
 tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
 actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
 aggaaatcac aggggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
 30 cctttgagaa cctagaaatc atacgcccga ggaccaagca acatggtcag ttttctcttg 1560
 cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
 gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
 aactgtttgg gacctccggt cagaaaacca aaattataag caacagaggt gaaaacagct 1740
 gcaaggccac aggccagggtc tgccatgcct tgtgtcccc cgagggctgc tggggcccgg 1800
 agcccaggga ctgctgtctt tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
 35 gcaaccttct ggagggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
 acccagagtg cctgcctcag gccatgaaca tcacctgcac aggcagggga ccagacaact 1980
 gtatccagtg tgccactac attgacggcc cccactgctg caagacctgc ccggcaggag 2040
 tcatgggaga aaacaacacc ctggctctgga agtacgcaga cgccggccat gtgtgccacc 2100
 tgtgccatcc aaactgcacc tacggatgca ctgggcccagg tcttgaaggc tgtccaacga 2160
 atgggcctaa gatcccgtcc atcgccactg ggatgggtgg ggccctcctc ttgctgctgg 2220
 40 tggtagccct ggggatcggc ctcttcatgc gaaggcgcca catcggtcgg aagcgcacgc 2280
 tgccgaggct gctgcaggag agggagcttg tggagcctct tacaccagt ggagaagctc 2340
 ccaaccaagc tctcttgagg atcttgaagg aaactgaatt caaaaagatc aaagtgtctg 2400
 gctccgggtg gttcggcacg gtgtataagg gactctggat cccagaaggt gagaaagtta 2460
 aaattcccgt cgctatcaag gaattaagag aagcaacatc tccgaaagcc aacaaggaaa 2520
 tcctcgatga agcctacgtg atggccagcg tggacaaccc ccacgtgtgc cgcctgctgg 2580

45

50

55

```

5      gcacatctgcct caccctccacc gtgcagctca tcacgcagct catgcccttc ggctgcctcc 2640
      tggactatgt cccgggaacac aaagacaata ttggctccca gtacctgctc aactgggtgtg 2700
      tgcagatcgc aaagggcatg aactacttgg aggaccgtcg cttgggtgcac cgcgacctgg 2760
      cagccaggaa cgtactggtg aaaacaccgc agcatgtcaa gatcacagat tttgggcggg 2820
      ccaaactgct ggggtgcggaa gagaaaagaat accatgcaga aggaggcaaa gtgcctatca 2880
      agtggatggc attggaatca attttacaca gaatctatac ccaccagagt gatgtctgga 2940
      gctacggggg gactgttttg gagttgatga cttttggatc caagccatat gacggaatcc 3000
      ctgccagcga gatctcctcc atcctggaga aaggagaacg cctccctcag ccacccatat 3060
      gtaccatcga tgtctacatg atcatggtca agtgctggat gatagacgca gatagtcgcc 3120
      caaagtccg tgagttgatc atcgaattct ccaaaatggc ccgagacccc cagcgctacc 3180
10     ttgtcattca ggggggatgaa agaatgcatt tgccaagtcc tacagactcc aacttctacc 3240
      gtgccctgat ggatgaagaa gacatggacg acgtgggtga tgccgacgag tacctcatcc 3300
      cacagcaggg cttcttcagc agcccccca cgtcacggac tccccctctg agctctctga 3360
      gtgcaaccag caacaattcc accgtggctt gcattgatag aaatgggctg caaagctgtc 3420
      ccatcaagga agacagcttc ttgcagcgat acagctcaga cccacagggc gccttgactg 3480
      aggacagcat agacgacacc ttcctcccag tgcctgaata cataaaccag tccgttccca 3540
15     aaaggccgc tggctctgtg cagaatcctg tctatcaca tcagcctctg aaccccgcg 3600
      ccagcagaga cccacactac caggaccccc acagactgc agtgggcaac cccgagtatc 3660
      tcaacactgt ccagcccacc tgtgtcaaca gcacattcga cagccctgcc cactgggccc 3720
      agaaaggcag ccaccaaatt agcctggaca accctgacta ccagcaggac ttctttccca 3780
      aggaagccaa gccaaatggc atctttaagg gctccacagc tgaaaatgca gaatacctaa 3840
      gggtcgcgcc acaaagcagt gaatttattg gagcatga 3878

```

20

```

<210> 761
<211> 3878
<212> DNA
<213> Homo sapiens

```

25

```

<400> 761

```

30

35

40

45

50

55

	cccggcgag	cgcgccgca	gcagcctccg	ccccccgcac	ggtgtgagcg	cccgcgcgg	60
	ccgagggcgc	cggagtcgcc	agctagcccc	ggcgccgcgc	gccgcccaga	ccggacgaca	120
	ggccaccttcg	tcggcgctccg	cccagatccc	cgctcgccg	ccaacgccac	aaccaccgcg	180
	cacggccccc	tgactccgtc	cagtattgat	cgggagagcc	ggagcgagct	cttcggggag	240
5	cagcgatgag	accctccggg	acggccgggg	cagcgctcct	ggcgctgctg	gctgcgctct	300
	gcccggcgag	tcgggctctg	gaggaaaaga	aagtttgcca	aggcacgagt	aacaagctca	360
	cgcagttggg	cacttttgaa	gatcattttc	tcagcctcca	gaggatgttc	aataactgtg	420
	aggtggctct	tgggaatttg	gaaattacct	atgtgcagag	gaattatgat	ctttccttct	480
	ttaaagaccat	ccaggaggtg	gctgggtatg	tcctcattgc	cctcaacaca	gtggagcgaa	540
	ttcctttgga	aaacctgcag	atcatcagag	gaaatatgta	ctacgaaaat	tcctatgcct	600
10	tagcagtcct	atctaactat	gatgcaaata	aaaccggact	gaaggagctg	cccatgagaa	660
	atttacagga	aatcctgcat	ggcgccgtgc	ggttcagcaa	caaccctgcc	ctgtgcaacg	720
	tggagagcat	ccagtggcgg	gacatagtca	gcagtgaact	tctcagcaac	atgtcgatgg	780
	acttccagaa	ccacctgggc	agctgccaaa	agtgtgatcc	aagctgtccc	aatgggagct	840
	gctgggggtgc	aggagaggag	aactgccaga	aactgaccaa	aatcatctgt	gccagcagct	900
	gctccggggcg	ctgccgtggc	aagtccccc	gtgactgctg	ccacaaccag	tgtgctgcag	960
15	gctgcacagg	ccccggggag	agcgactgcc	tggctgccc	caaattccga	gacgaagcca	1020
	cgtgcaagga	cacctgcccc	ccactcatgc	tctacaaccc	caccacgtac	cagatggatg	1080
	tgaaccccg	gggcaaatac	agctttgggtg	ccacctgcgt	gaagaagtgt	ccccgtaatt	1140
	atgtgggtgac	agatcacggc	tcgtgcgtcc	gagcctgtgg	ggccgacagc	tatgatagtg	1200
	aggaagacgg	cgtccgcaag	tgtaagaagt	gcgaagggcc	ttgccgcaaa	gtgtgtaacg	1260
	gaataggtat	tgggtgaattt	aaagactcac	tctccataaa	tgctacgaat	attaaacact	1320
20	tcaaaaactg	cacctccatc	agtggcgatc	tccacatcct	gccgggtggca	tttaggggtg	1380
	actccttcac	acatactcct	cctctggatc	cacaggaact	ggatattctg	aaaaccgtaa	1440
	aggaatcac	agggtttttg	ctgattcagg	cttggcctga	aaacaggacg	gacctccatg	1500
	cctttgagaa	cctagaaatc	atacgcggca	ggaccaagca	acatggtcag	ttttctcttg	1560
	cagtcgtcag	cctgaacata	acatccttgg	gattacgctc	cctcaaggag	ataagtgatg	1620
	gagatgtgat	aatttcagga	aacaaaaatt	tgtgctatgc	aaatacaata	aactggaaaa	1680
25	aactgtttgg	gacctccggt	cagaaaacca	aaattataag	caacagaggt	gaaaacagct	1740
	gcaaggccac	aggccagggtc	tgccatgcct	tgtgctcccc	cgagggctgc	tggggcccg	1800
	agcccaggga	ctgcgtctct	tgccggaatg	tcagccgagg	cagggaatgc	gtggacaagt	1860
	gcaaccttct	ggaggggtgag	ccaagggagt	ttgtggagaa	ctctgagtgc	atacagtgcc	1920
	accagagtg	cctgcctcag	gccatgaaca	tcacctgcac	aggacgggga	ccagacaact	1980
	gtatccagtg	tgcccaactac	attgacggcc	cccactgcgt	caagacctgc	ccggcaggag	2040
30	tcatgggaga	aaacaacacc	ctggcttgga	agtacgcaga	cgccggccat	gtgtgccacc	2100
	tgtgccatcc	aaactgcacc	tacggatgca	ctgggcccag	tcttgaaggc	tgtccaacga	2160
	atgggcccata	gatcccgtcc	atcgccactg	ggatgggtgg	ggccctctct	ttgctgtgg	2220
	tgggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgacgc	2280
	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacacccagt	ggagaagctc	2340
35	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataagg	gactctggat	cccagaaggt	gagaaagtta	2460
	aatctcccg	cgctatcaag	gaattaaag	aagcaacatc	tccgaaagct	aacaaggaaa	2520
40	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cgctgtctg	2580
	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtc	aactgggtgtg	2700
	tgcagatcgc	aaagggcatg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
	cagccaggaa	cgtactggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggcccgg	2820
	ccaaactgct	gggtgcggaa	gagaaaagaat	accatgcaga	aggaggcaaa	gtgccctatca	2880
	agtggatggc	atttgaatca	atitttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
45	gctacgggg	gactgtttgg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccacccatat	3060
	gtaccatcga	tgtctacatg	atcatggtca	agtgtggat	gatagacgca	gatagtgcgc	3120
	caaagtccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtgggtga	tgccgacgag	tacctcatcc	3300
	cacagcagg	cttcttcagc	agcccccca	cgtcacggac	tccccctctg	agctctctga	3360
50	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aatgggctg	caaagctgtc	3420
	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gcttgcactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgcttgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggctctgtg	cagaatcctg	tctatcacia	tcagcctctg	aaccccgcg	3600
	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccagtatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	caactggccc	3720
55	agaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
	aggaagccaa	gccaaatggc	atctttaagg	gtcccacagc	tgaaaatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaatttatgt	gagcatga			3878

<210> 762
 <211> 3878
 <212> DNA
 <213> Homo sapiens

5

<400> 762

```

    cccggcgcgag cgcggccgca gcagcctccg ccccccgcac ggtgtgagcg cccgacgcgg 60
    ccgaggcgggc cggaggtccc agctagcccc ggcgggccgc gccgcccaga ccggacgaca 120
  10  ggccacctcg tcggcggtccg cccgaggtccc cgctcgccg ccaacgccac aaccaccgcg 180
    cacggccccc tgactccgtc cagtattgat cgggagagcc ggagcgagct cttcggggag 240
    cagcgatgcy accctccggg acggccgggg cagcgctcct ggcgctgctg gctgcgtctt 300
    gcccggcgag tcgggtctcg gaggaagaaga aagtttgcca aggcacgagt aacaagctca 360
    cgcagttggg cacttttgaa gatcattttc tcagcctcca gaggatgttc aataactgtg 420
  15  aggtggctct tgggaatttg gaaattacct atgtgcagag gaattatgat ctttccttct 480
    taaagaccat ccaggaggtg gctgggtatg tcctcattgc cctcaacaca gtggagcgaa 540
    ttccctttgga aaacctgcag atcatcagag gaaatatgta ctacgaaaat tcctatgcct 600
    tagcagtcct atctaactat gatgcaaata aaaccggact gaaggagctg cccatgagaa 660
    atttacagga aatcctgcat ggcgccgtgc gggtcagcaa caaccctgcc ctgtgcaacg 720
    tggagagcat ccagtgccgg gacatagtca gcagtgactt tctcagcaac atgtcgatgg 780
    acttccagaa ccacctgggc agctgccaaa agtgtgatcc aagctgtccc aatgggagct 840
  20  gctgggggtgc aggagaggag aactgccaga aactgaccaa aatcatctgt gcccagcagt 900
    gctccggggcg ctgcccgtggc aagtcccca gtgactgctg ccacaaccag tgtgctgcag 960
    gctgcacagg ccccggggag agcgactgcc tggctgccc caaattccga gacgaagcca 1020
    cgtgcaagga cacctgcccc cactcatgc tctacaaccc caccacgtac cagatggatg 1080
    tgaaccccgga gggcaaatac agctttgggt ccacctgcgt gaagaagtgt ccccgtaatt 1140
    atgtgggtgac agatcacggc tcgtgcgtcc gagcctgtgg ggccgacagc tatgagatgg 1200
  25  aggaagacgg cgtccgcaag tgtaagaagt gcgaagggcc ttgccgcaaa gtgtgtaacg 1260
    gaataggtat tgggtgaattt aaagactcac tctccataaa tgctacgaat attaaacact 1320
    tcaaaaactg cacctccatc agtggcgatc tccacatcct gccggtggca tttaggggtg 1380
    actccttcac acatactcct cctctggatc cacaggaact ggatattctg aaaaccgtaa 1440
    aggaaatcac agggtttttg ctgattcagg cttggcctga aaacaggacg gacctccatg 1500
    cctttgagaa cctagaaatc atacgcggca ggaccaagca acatggtcag ttttctcttg 1560
  30  cagtcgtcag cctgaacata acatccttgg gattacgctc cctcaaggag ataagtgatg 1620
    gagatgtgat aatttcagga aacaaaaatt tgtgctatgc aaatacaata aactggaaaa 1680
    aactgtttgg gacctccggt cagaaaaacca aaattataag caacagaggt gaaaacagct 1740
    gcaaggccac aggccagggtc tgccatgcct tgtgctcccc cgagggctgc tggggcccgg 1800
    agcccaggga ctgctctctc tgccggaatg tcagccgagg cagggaatgc gtggacaagt 1860
  35  gcaaccttct ggaggggtgag ccaagggagt ttgtggagaa ctctgagtgc atacagtgcc 1920
    acccagagtg cctgcctcag gccatgaaca tcacctgcac aggacgggga ccagacaact 1980
    gtatccagtg tgccactac attgacggcc cccactgcgt caagacctgc ccggcaggag 2040
    tcatgggaga aaacaacacc ctggtctgga agtacgcaga cgccggccat gtgtgccacc 2100

```

40

45

50

55

	tgtgccatcc	aaactgcacc	tacggatgca	ctggggccagg	tcttgaaggc	tgtccaacga	2160
	atgggcctaa	gatcccgctc	atcgccactg	ggatgggtggg	ggccctcctc	ttgctgctgg	2220
	tgggtggccct	ggggatcggc	ctcttcatgc	gaaggcgcca	catcgttcgg	aagcgcacgc	2280
5	tgcggaggct	gctgcaggag	agggagcttg	tggagcctct	tacacccagt	ggagaagctc	2340
	ccaaccaagc	tctcttgagg	atcttgaagg	aaactgaatt	caaaaagatc	aaagtgtctg	2400
	gctccggtgc	gttcggcacg	gtgtataaag	gactctggat	cccagaaggt	gagaaagtta	2460
	aaattcccg	cgctatcaag	gaattaagag	aagcaacatc	tccgaaagcc	aacaaggaaa	2520
	tcctcgatga	agcctacgtg	atggccagcg	tggacaaccc	ccacgtgtgc	cgctgtctgg	2580
	gcatctgcct	cacctccacc	gtgcagctca	tcacgcagct	catgcccttc	ggctgcctcc	2640
	tggactatgt	ccgggaacac	aaagacaata	ttggctccca	gtacctgtc	aactgggtgtg	2700
10	tgcagatcgc	aaagggcacg	aactacttgg	aggaccgtcg	cttgggtgcac	cgcgacctgg	2760
	cagccaggaa	cgtactgggtg	aaaacaccgc	agcatgtcaa	gatcacagat	tttgggctgg	2820
	ccaacacgct	gggtgcggaa	gagaaagaat	accatgcaga	aggaggcaaa	gtgcctatca	2880
	agtggatggc	attggaatca	atitttacaca	gaatctatac	ccaccagagt	gatgtctgga	2940
	gctacgggg	gactgttttg	gagttgatga	cctttggatc	caagccatat	gacggaatcc	3000
	ctgccagcga	gatctcctcc	atcctggaga	aaggagaacg	cctccctcag	ccacccatat	3060
15	gtaccatcga	tgtctacatg	atcatggtca	agtgtctggat	gatagacgca	gatagtcgcc	3120
	caaagttccg	tgagttgatc	atcgaattct	ccaaaatggc	ccgagacccc	cagcgctacc	3180
	ttgtcattca	gggggatgaa	agaatgcatt	tgccaagtcc	tacagactcc	aacttctacc	3240
	gtgccctgat	ggatgaagaa	gacatggacg	acgtggtgga	tgccgacgag	tacctcatcc	3300
	cacagcagg	cttcttcagc	agccccctcca	cgtcacggac	tccccctctg	agctctctga	3360
	gtgcaaccag	caacaattcc	accgtggctt	gcattgatag	aaatgggctg	caaagctgtc	3420
20	ccatcaagga	agacagcttc	ttgcagcgat	acagctcaga	ccccacaggc	gccttgactg	3480
	aggacagcat	agacgacacc	ttcctcccag	tgcctgaata	cataaaccag	tccgttccca	3540
	aaaggcccg	tggctctgtg	cagaatcctg	tctatcacia	tcagcctctg	aaccccgcg	3600
	ccagcagaga	cccacactac	caggaccccc	acagcactgc	agtgggcaac	cccgagtatc	3660
	tcaacactgt	ccagcccacc	tgtgtcaaca	gcacattcga	cagccctgcc	cactggggccc	3720
	agaaaaggcag	ccaccaaatt	agcctggaca	accctgacta	ccagcaggac	ttctttccca	3780
25	aggaagccaa	gccaaatggc	atctttaagg	gctccacagc	tgaaaatgca	gaatacctaa	3840
	gggtcgcgcc	acaaagcagt	gaattttattg	gagcatga			3878

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO 2005018677 A [0014]
- WO 0004194 A [0066] [0070]
- US 5493531 A [0082]
- US 5459039 A [0084]
- US 5919523 A [0093] [0095]
- US 5837832 A [0095]
- US 5831070 A [0095]
- US 5770722 A [0095]
- US 5807522 A, Brown and Shalon [0095]
- EP 0566226 A [0129]
- EP 0520722 A [0129]
- US 4925648 A [0135]
- US 4932412 A [0135]
- US 4331647 A [0140]
- US 4348376 A [0140]
- US 4361544 A [0140]
- US 4468457 A [0140]
- US 4444744 A [0140]
- US 4460459 A [0140]
- US 4460561 A [0140]
- US 4624846 A [0140] [0141]
- US 4818709 A [0140]
- US 4046722 A [0140]
- US 4671958 A [0140]
- US 4046784 A [0140]
- US 5332567 A [0140]
- US 5443953 A [0140]
- US 5541297 A [0140]
- US 5601825 A [0140]
- US 5635603 A [0140]
- US 5637288 A [0140]
- US 5677427 A [0140]
- US 5686578 A [0140]
- US 5698178 A [0140]
- US 5789554 A [0140]
- US 5922302 A [0140]
- US 6187287 B [0140]
- US 6319500 B [0140]
- US 4789734 A [0182]
- US 4906474 A [0183]
- US 4925673 A [0183]
- US 3625214 A [0183]
- US 5593688 A, Baldeschweiler [0184]
- US 5410016 A, Hubbell [0186]
- US 6015670 A [0200]
- US 4736866 A, Leder [0204]
- US 5175383 A [0204]
- EP 186833 A [0208]
- WO 9216553 A [0208]
- US 6090923 A [0208]
- WO 9909016 A [0210]
- WO 9843960 A [0210]
- WO 9738983 A [0210]
- WO 9906378 A [0210]
- WO 9906396 A [0210]
- WO 9630347 A [0210]
- WO 9633978 A [0210]
- WO 9633977 A [0210]
- WO 9633980 A [0210]
- US 6699843 B [0212]
- US 5707796 A [0212]
- US 5763177 A [0212]
- US 6011577 A [0212]
- US 5580737 A [0212]
- US 5567588 A [0212]
- US 5660985 A [0212]
- US 5814500 A [0215]
- US 5811234 A [0215]
- US 5780607 A [0215]
- US 20030198627 A [0216]
- US 20040063906 A [0224]
- US 0510645 W [0348]
- US 60558218 B [0348]
- US 60561095 B [0348]
- US 60565753 B [0348]
- US 60565985 B [0348]
- US 60574035 B [0348]
- US 60577916 B [0348]
- US 60592287 B [0348]

Non-patent literature cited in the description

- SCHILLER JH et al. *N Engl J Med*, 2002, vol. 346, 92-98 [0002]
- W. J. GULLICK et al. *Cancer Res.*, 1986, vol. 46, 285-292 [0003]
- S. COHEN et al. *J. Biol. Chem.*, 1980, vol. 255, 4834-4842 [0003]
- A. B. SCHREIBER et al. *J. Biol. Chem.*, 1983, vol. 258, 846-853 [0003]

- **L. HARRIS et al.** *Int. J. Biol. Markers*, 1999, vol. 14, 8-15 [0004]
- **J. MENDELSON ; J. BASELGA.** *Oncogene*, 2000, vol. 19, 6550-6565 [0004]
- **A. L. ULLRICH et al.** *Nature*, 1984, vol. 307, 418-425 [0005]
- **J. DOWNWARD et al.** *Nature*, 1984, vol. 307, 521-527 [0005]
- **C. R. CARLIN et al.** *Mol. Cell. Biol.*, 1986, vol. 6, 257-264 [0005]
- **F. L. V. MAYES ; M. D. WATERFIELD.** *The EMBO J.*, 1984, vol. 3, 531-537 [0005]
- **G. CARPENTER ; S. COHEN.** *Ann. Rev. Biochem.*, 1979, vol. 48, 193-216 [0006]
- **M.-J. OH et al.** *Clin. Cancer Res.*, 2000, vol. 6, 4760-4763 [0006]
- **DE BONO, J. S. ; ROWINSKY, E. K.** The ErbB Receptor Family: A Therapeutic Target For Cancer. *Trends in Molecular Medicine*, 2002, vol. 8, 19-26 [0007]
- **TORRANCE et al.** *Nature Medicine*, September 2000, vol. 6 (9), 1024 [0014] [0129] [0210]
- **GREENBERGER et al.** Proc. 11th NCI EORTC-AACR Symposium on New Drugs in Cancer Therapy. *Clinical Cancer Res.*, November 2000, vol. 6, ISSN 1078-0432 [0014]
- **RABINDRAN et al.** *Cancer Res.*, 2004, vol. 64, 3958-3965 [0014]
- **HOLBRO ; HYNES.** *Ann. Rev. Pharm. Tox.*, 2004, vol. 44, 195-217 [0014]
- **TSOU et al.** *J. Med. Chem.*, 2005, vol. 48, 1107-1131 [0014]
- **TEJPAR et al.** *J. Clin. Oncol. ASCO Annual Meeting Proc.*, 2004, vol. 22 (14), 3579 [0014]
- **CARPENTER et al.** *Ann. Rev. Biochem.*, 1987, vol. 56, 881-914 [0044]
- **HUMPHREY et al.** *PNAS*, 1990, vol. 87, 4207-4211 [0044]
- **ROLFF, A et al.** PCR: Clinical Diagnostics and Research. Springer, 1994 [0064]
- **LANDEGRAN et al.** *Science*, 1988, vol. 241, 1077-1080 [0071]
- **NAKAZAWA et al.** *Proc. Natl. Acad. Sci. USA*, 1994, vol. 91, 360-364 [0071]
- **ABRAVAYA et al.** *Nucl. Acids Res.*, 1995, vol. 23, 675-682 [0071]
- **GUATELLI et al.** *Proc. Natl. Acad. Sci. USA*, 1990, vol. 87, 1874-1 878 [0072]
- **KWOH et al.** *Proc. Natl. Acad. Sci. USA*, 1989, vol. 86, 1173-1177 [0072]
- **LIZARDI et al.** *BioTechnology*, 1988, vol. 6, 1197 [0072]
- **Molecular Cloning: A Laboratory Manual.** Cold Spring Harbor Laboratory, 15 January 2001, vol. 3 [0080]
- **M. J. MCPHERSON ; S. G. MØLLER ; R. BEYNON ; C. HOWE.** PCR amplification is PCR (Basics: From Background to Bench). Springer Verlag, 15 October 2000 [0080]
- **MYERS et al.** *Science*, 1985, vol. 230, 1242 [0083]
- **COTTON et al.** *Proc. Natl. Acad. Sci. USA*, 1988, vol. 85, 4397 [0083]
- **SALEEBA et al.** *Methods Enzymol.*, 1992, vol. 2 (17), 286-295 [0083]
- **HSU et al.** *Carcinogenesis*, 1994, vol. 15, 1657-1662 [0084]
- **ORITA et al.** *Proc. Natl. Acad. Sci. USA*, 1989, vol. 86, 2766 [0085]
- **COTTON.** *Mutat. Res.*, 1993, vol. 285, 125-144 [0085]
- **HAYASHI.** *Genet. Anal. Tech. Appl.*, 1992, vol. 9, 73-79 [0085]
- **KEEN et al.** *Trends Genet.*, 1991, vol. 7, 5 [0085]
- **MYERS et al.** *Nature*, 1985, vol. 313, 495 [0086]
- **ROSENBAUM ; REISSNER.** *Biophys. Chem.*, 1987, vol. 265, 12753 [0086]
- **SAIKI et al.** *Nature*, 1986, vol. 324, 163 [0087]
- **SAIKI et al.** *Proc. Natl. Acad. Sci. USA*, 1989, vol. 86, 6230 [0087]
- **GIBBS et al.** *Nucl. Acids Res.*, 1989, vol. 17, 2437-24, 48 [0088]
- **PROSSNER.** *Tibtech.*, 1993, vol. 11, 238 [0088]
- **GASPARINI et al.** *Mol. Cell Probes*, 1992, vol. 6, 1 [0088]
- **BARANY.** *Proc. Natl. Acad. Sci. USA*, 1991, vol. 88, 189 [0088]
- **Current Protocols in Molecular Biology.** John Wiley & Sons [0089]
- **Molecular Cloning: A Laboratory Manual.** Cold Spring Harbor Laboratory, 2001, vol. 1-3 [0092]
- **SAMBROOK et al.** *Molecular Cloning.* Cold Spring Harbor Laboratory Press, 2001 [0092]
- **EMBRETSON et al.** *Nature*, 1993, vol. 362, 359-362 [0098]
- **GOSDEN et al.** *BioTechniques*, 1993, vol. 15 (1), 78-80 [0098]
- **HENIFORD et al.** *Nuc. Acid Res.*, 1993, vol. 21 (14), 3159-3166 [0098]
- **LONG et al.** *Histochemistry*, 1993, vol. 99, 151-162 [0098]
- **NUOVO et al.** *PCR Methods and Applications*, 1993, vol. 2 (4), 305-312 [0098]
- **PATTERSON et al.** *Science*, 1993, vol. 260, 976-979 [0098]
- **SAMBROOK et al.** *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, 2001 [0100]
- **C. R. NEWTON ; A. GRAHAM.** PCR. BIOS Publishers, 1994 [0100]
- **WIEDMANN, M. et al.** *PCR Methods Appl.*, 1994, vol. 3, 57-64 [0100]
- **F. BARNAY.** *Proc. Natl. Acad. Sci. USA*, 1991, vol. 88, 189-93 [0100]
- **G. TERRANCE WALKER et al.** *Nucleic Acids Res.*, 1994, vol. 22, 2670-77 [0100]
- **HIGUCHI et al.** *Biol Technology*, 1993, vol. 11, 1026-1030 [0100]

- **NIELSEN et al.** *Science*, 1991, vol. 254, 1497 [0102]
- Mass Spectrometry. Methods of Enzymology. Academic Press, 1990, vol. 193 [0105]
- **R. N. JORISSEN et al.** *Exp. Cell Res.*, 2003, vol. 284, 31 [0108] [0326] [0328] [0330]
- **JALKANEN et al.** *J. Cell. Biol.*, 1985, vol. 101, 976-985 [0115]
- **JALKANEN et al.** *J. Cell. Biol.*, 1987, vol. 105, 3087-3096 [0115]
- **HARLOW et al.** Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory Press, 1988 [0116]
- Antibodies A Laboratory Manual. Cold Spring Harbor Press, 1988 [0117]
- Current Protocols In Molecular Biology. John Wiley and Sons, 1987 [0117]
- **BRAUER et al.** *FASEB J*, 2001, vol. 15, 2689-2701 [0119]
- **DENNY.** *Farmaco*, January 2001, vol. 56 (1-2), 51-6 [0129]
- **G. D. DEMETRI et al.** *N. Engl. J. Med.*, 2002, vol. 347, 472 [0130]
- **B. J. DRUKER et al.** *N. Engl. J. Med.*, 2001, vol. 344, 1038 [0130]
- **D.J. SLAMON et al.** *N. Engl. J. Med.*, 2001, vol. 344, 783 [0130]
- **STACHI-FAINARO et al.** *Cancer Cell*, 2005, vol. 7 (3), 251 [0131]
- **RIBATTI et al.** *J Hematother Stem Cell Res.*, 2003, vol. 12 (1), 11-22 [0132]
- REMINGTON'S PHARMACEUTICAL SCIENCES. Mack Publishing Co, 1995 [0135]
- Goodman and Gilman's The Pharmacological Basis of Therapeutics. Macmillan Publishing Co, 1980 [0135]
- **WAWRZYNCZAK ; THORPE.** Introduction to the Cellular and Molecular Biology of Cancer. Oxford University Press, 1986, 378-410 [0140]
- Immunoconjugates: Antibody Conjugates in Radioimaging and Therapy of Cancer. Oxford University Press, 1987, 3-300 [0140]
- **DILLMAN, R. O.** CRC Critical Reviews in Oncology/Hematology. CRC Press, Inc, 1984, vol. 1, 357 [0140]
- **PASTAN et al.** *Cell*, 1986, vol. 47, 641 [0140]
- **VITETTA et al.** *Science*, 1987, vol. 238, 1098-1104 [0140]
- **BRADY et al.** *Int. J. Rad. Oncol. Biol. Phys.*, 1987, vol. 13, 1535-1544 [0140]
- **HARBERT.** Nuclear Medicine Therapy. Thieme Medical Publishers, vol. 1087, 1-340 [0144]
- Current Protocols in Immunology. John Wiley & Sons, 2001 [0153]
- **DHANABAL et al.** *Cancer Res.*, 1999, vol. 59, 189-197 [0164] [0167]
- **XIN et al.** *J. Biol. Chem.*, 1999, vol. 274, 9116-9121 [0164] [0167]
- **SHEU et al.** *Anticancer Res.*, vol. 18, 4435-4441 [0164] [0167]
- **AUSPRUNK et al.** *Dev. Biol.*, 1974, vol. 38, 237-248 [0164] [0167]
- **GIMBRONE et al.** *J. Natl. Cancer Inst.*, vol. 52, 413-427 [0164] [0167]
- **NICOSIA et al.** *In vitro*, vol. 18, 538-549 [0164] [0167]
- Liposomes. **G. GREGORIADIS.** Drug Carriers in Biology and Medicine. Academic Press, 1979, 287-341 [0182]
- **JEIN.** *TIPS*, 1998, vol. 19, 155-157 [0183]
- Current Protocols in Molecular Biology. John Wiley & Sons, 2004 [0198]
- **SCAPPINI et al.** *Cancer*, 01 April 2004, vol. 100, 1459 [0201]
- **OHASHI et al.** *Cell*, 1991, vol. 65, 305-317 [0204]
- **ADAMS et al.** *Nature*, 1987, vol. 325, 223-228 [0204]
- **ROMAN et al.** *Cell*, 1990, vol. 61, 383-396 [0204]
- **TRAXLER et al.** *J. Med Chem*, 1996, vol. 39, 2285-2292 [0210]
- **LAXMIN ARAYANA et al.** *Carcinogen*, 1995, vol. 16, 1741-1745 [0210]
- **BUCHDUNGER et al.** *Clin. Cancer Res.*, 1995, vol. 1, 813-821 [0210]
- **DINNEY et al.** *Clin. Cancer Res.*, 1997, vol. 3, 161-168 [0210]
- **BRUNTON et al.** *Anti Cancer Drug Design*, 1996, vol. 11, 265-295 [0210]
- Antisense DNA and RNA. Cold Spring Harbor Laboratory, 1988 [0211]
- **MCMANUS et al.** *Nature Reviews Genetics*, 2002, vol. 3 (10), 737 [0212]
- **CECH.** *Science*, 1989, vol. 245 (4915), 276 [0214]
- **BRUMMELKAMP et al.** *Science*, 2002, vol. 296, 550-553 [0216]
- **JAQUE et al.** *Nature*, 2002, vol. 418, 435-438 [0216]
- **ELBASHIR S. M. et al.** *Nature*, 2001, vol. 411, 494-498 [0216]
- **MCCAFFREY et al.** *Nature*, 2002, vol. 418, 38-39 [0216]
- **XIA H. et al.** *Nat. Biotech.*, 2002, vol. 20, 1006-1010 [0216]
- **NOVINA et al.** *Nat. Med.*, 2002, vol. 8, 681-686 [0216]
- **NYCE et al.** *Nature*, 1997, vol. 385, 720 [0219]
- **NIELSEN et al.** *Science*, 1991, vol. 254, 1457-1500 [0224]
- **EGHOLM et al.** *Nature*, 1993, vol. 365, 566-568 [0224]
- **DEMIDOV et al.** *Biochem. Pharmacol.*, 1994, vol. 48, 1310-1313 [0224]
- **NIELSEN ; HAAIMA.** *Chemical Society Reviews*, 1997, 73-78 [0224]
- **BRADFORD et al.** *Blood.*, 01 July 2003, vol. 102 (1), 276-83 [0226]
- **HOCHHAUS et al.** *Leukemia*, November 2002, vol. 16 (11), 2190-6 [0226]
- **AL-ALI et al.** *Hematol J.*, 2004, vol. 5 (1), 55-60 [0226]

- **A. B. MILLER ; B. HOOGSTRATEN ; M. STAQUET ; A. WINKLER.** *Cancer*, 1981, vol. 47, 207-14 [0326]
- **H. S. EARP ; T. L. DAWSON ; X. LI ; H. YU.** *Breast Cancer Res. Treat.*, 1995, vol. 35, 115 [0326]
- **A- E. WAKELING et al.** *Cancer Res.*, 2002, vol. 62, 5749 [0326]
- **J. BASELGA et al.** *J. Clin. Oncol.*, 2002, vol. 20, 4292 [0326]
- **M. FUKUOKA et al.** *J. Clin. Oncol.*, 2003, vol. 21, 2237 [0326]
- **G. GIACCONE et al.** *J Clin Oncol.*, 2004, vol. 22, 777 [0326]
- **M. G. KRIS et al.** *JAMA*, 2003, vol. 290, 2149 [0326]
- **T. J. LYNCH et al.** *N. Engl. J. Med.*, 2004, vol. 350, 2129 [0326] [0327] [0334]
- **J. G. PAEZ et al.** *Science*, 2004, vol. 304, 1497 [0326]
- **S. GRANT ; L. QIAO ; P. DENT.** *Front. Biosci.*, 2002, vol. 7, d376 [0332]
- **F. CHANG et al.** *Leukemia*, 2003, vol. 17, 1263 [0332]
- **F. CHANG et al.** *Leukemia*, 2003, vol. 17, 590 [0332]
- **F. CHANG et al.** *Int. J. Oncol.*, 2003, vol. 22, 469 [0332]
- **V. CALO et al.** *J. Cell Physiol.*, 2003, vol. 197, 157 [0332]
- **T. J. AHONEN et al.** *J. Biol. Chem.*, 2003, vol. 278, 27287 [0332]
- **I. B. WEINSTEIN.** *Science*, 2002, vol. 297, 63 [0337]
- **T. KINDLER et al.** *Leukemia*, 2003, vol. 17, 999 [0337]
- **G. P. PANER et al.** *Anticancer Res.*, 2003, vol. 23, 2253 [0337]
- **SCHILLER JH ; HARRINGTON D ; BELANI CP et al.** Comparison of four chemotherapy regimens for advanced non-small cell lung cancer. *N Engl J Med*, 2002, vol. 346, 92-98 [0346]
- **DRUKER BJ ; TALPAZ M ; RESTADJ et al.** Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in Chronic Myeloid Leukemia. *N Engl J Med*, 2001, vol. 344, 1031-1037 [0346]
- **ARTEAGA CL.** ErbB-targeted therapeutic approaches in human cancer. *Exp Cell Res.*, 2003, vol. 284, 122-30 [0346]
- **JORISSEN RN ; WALKER F ; POULIOT N ; GARRETT TP ; WARD CW ; BURGESS AW.** Epidermal growth factor receptor: mechanisms of activation and signaling. *Exp Cell Res*, 2003, vol. 284, 31-53 [0346]
- **LUETTEKE NC ; PHILLIPS HK ; QUI TH ; COPELAND NG ; EARP HS ; JENKINS NA ; LEE DC.** The mouse waved-2 phenotype results from a point mutation in the EGF receptor tyrosine kinase. *Genes Dev*, 1994, vol. 8, 399-413 [0346]
- **NICHOLSON RI ; GEE JMW ; HARPER ME.** EGFR and cancer prognosis. *Eur J Cancer.*, 2001, vol. 37, 9-15 [0346]
- **WONG AJ ; RUPPERT JM ; BIGNER SH et al.** Structural alterations of the epidermal growth factor receptor gene in human gliomas. *Proc Natl Acad Sci.*, 1992, vol. 89, 2965-2969 [0346]
- **CIESIELSKI MJ ; GENSTERMAKER RA.** Oncogenic epidermal growth factor receptor mutants with tandem duplication: gene structure and effects on receptor function. *Oncogene*, 2000, vol. 19, 810-820 [0346]
- **FREDERICK L ; WANG W-Y ; ELEY G ; JAMES CD.** Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas. *Cancer Res*, 2000, vol. 60, 1383-1387 [0346]
- **HUANG H-JS ; NAGANE M. KLINGBEIL CK et al.** The enhanced tumorigenic activity of a mutant epidermal growth factor receptor common in human cancers is mediated by threshold levels of constitutive tyrosine phosphorylation and unattenuated signaling. *J Biol Chem*, 1997, vol. 272, 2927-2935 [0346]
- **PEGRAM MD ; KONECNY G ; SLAMON DJ.** The molecular and cellular biology of HER2/neu gene amplification/overexpression and the clinical development of herceptin (trastuzumab) therapy for breast cancer. *Cancer Treat Res*, 2000, vol. 103, 57-75 [0346]
- **CIARDIELLO F ; TORTORA G.** A novel approach in the treatment of cancer targeting the epidermal growth factor receptor. *Clin Cancer Res.*, 2001, vol. 7, 2958-2970 [0346]
- **WAKELING AE ; GUY SP ; WOODBURN JR et al.** ZD1839 (Iressa): An orally active inhibitor of Epidermal Growth Factor signaling with potential for cancer therapy. *Cancer Res*, 2002, vol. 62, 5749-5754 [0346]
- **MOULDER SL ; YAKES FM ; MUTHUSWAMY SK ; BIANCO R ; SIMPSON JF ; ARTEAGA CL.** Epidermal growth factor receptor (HER1) tyrosine kinase inhibitor ZD1839 (Iressa) inhibits HER2/neu (erbB2)-overexpressing breast cancer cells in vitro and in vivo. *Cancer Res*, 2001, vol. 61, 8887-8895 [0346]
- **MOASSER MM ; BASSO A ; AVERBUCH SD ; ROSEN N.** The tyrosine kinase inhibitor ZD1839 ("Iressa") inhibits HER2-driven signaling and suppresses the growth of HER-2 overexpressing tumor cells. *Cancer Res*, 2001, vol. 61, 7184-7188 [0346]
- **RANSON M ; HAMMOND LA ; FERRY D et al.** ZD1839, a selective oral epidermal growth factor receptor-tyrosine kinase inhibitor, is well tolerated and active in patients with solid, malignant tumors: results of a phase I trial. *J Clin Oncol.*, 2002, vol. 20, 2240-2250 [0346]
- **HERBST RS ; MADDOX A-M ; ROTHERNBERG ML et al.** Selective oral epidermal growth factor receptor tyrosine kinase inhibitor ZD1839 is generally well tolerated and has activity in non-small cell lung cancer and other solid tumors: results of a phase I trial. *J Clin Oncol.*, 2002, vol. 20, 3815-3825 [0346]

- **BASELGA J ; RISCHIN JB ; RANSON M et al.** Phase I safety, pharmacokinetic and pharmacodynamic trial of ZD1839, a selective oral Epidermal Growth Factor Receptor tyrosine kinase inhibitor, in patients with five selected solid tumor types. *J Clin Oncol*, 2002, vol. 20, 4292-4302 [0346]
- **ALBANELL J ; ROJO F ; AVERBUCH S et al.** Pharmacodynamic studies of the epidermal growth factor receptor inhibitor ZD1839 in skin from cancer patients: histopathologic and molecular consequences of receptor inhibition. *J Clin Oncol.*, 2001, vol. 20, 110-124 [0346]
- **KRIS MG ; NATALE RB ; HERBST RS et al.** Efficacy of Gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: A randomized trial. *JAMA*, 2003, vol. 290, 2149-2158 [0346]
- **FUKUOKA M ; YANO S ; GIACCONE G et al.** Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer. *J Clin Oncol*, 2003, vol. 21, 2237-2246 [0346]
- **GIACCONE G ; HERBST RS ; MANEGOLD C et al.** Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: A phase III trial-INTACT 1. *J Clin Oncol*, 2004, vol. 22, 777-784 [0346]
- **HERBST RS ; GIACCONE G ; SCHILLER JH et al.** Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: A phase III trial - INTACT 2. *J Clin Oncol*, 2004, vol. 22, 785-794 [0346]
- **RICH JN ; REARDON DA ; PEERY T et al.** Phase II Trial of Gefitinib in recurrent glioblastoma. *J Clin Oncol*, 2004, vol. 22, 133-142 [0346]
- **COHEN MH ; WILLIAMS GA ; SRIDHARA R et al.** United States Food and Drug Administration Drug Approval Summary: Gefitinib (ZD1839;Iressa) Tablets. *Clin Cancer Res.*, 2004, vol. 10, 1212-1218 [0346]
- **CAPPUZZO F ; GREGORC V ; ROSSI E et al.** Gefitinib in pretreated non-small-cell lung cancer (NSCLC): Analysis of efficacy and correlation with HER2 and epidermal growth factor receptor expression in locally advanced or Metastatic NSCLC. *J Clin Oncol.*, 2003, vol. 21, 2658-2663 [0346]
- **FITCH KR ; MCGOWAN KA ; VAN RAAMSDONK CD et al.** Genetics of Dark Skin in mice. *Genes & Dev*, 2003, vol. 17, 214-228 [0346]
- **NIELSEN UB ; CARDONE MH ; SINSKEY AJ ; MACBEATH G ; SORGER PK.** Profiling receptor tyrosine kinase activation by using Ab microarrays. *Proc Natl Acad Sci USA*, 2003, vol. 100, 9330-5 [0346]
- **BURGESS AW ; CHO H ; EIGENBROT C et al.** An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol Cell*, 2003, vol. 12, 541-552 [0346]
- **STAMOS J ; SLIWKOWSKI MX ; EIGENBROT C.** Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J Biol Chem.*, 2002, vol. 277, 46265-46272 [0346]
- **LORENZATO A ; OLIVERO M ; PATRANE S et al.** Novel somatic mutations of the MET oncogene in human carcinoma metastases activating cell motility and invasion. *Cancer Res*, 2002, vol. 62, 7025-30 [0346]
- **DAVIES H ; BIGNELL GR ; COX C et al.** Mutations of the BRAF gene in human cancer. *Nature*, 2002, vol. 417, 906-7 [0346]
- **BARDELLI A ; PARSONS DW ; SILLIMAN N et al.** Mutational analysis of the tyrosine kinase in colorectal cancers. *Science*, 2003, vol. 300, 949 [0346]
- **DALEY GQ ; VAN ETEN RA ; BALTIMORE D.** Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science*, 1990, vol. 247, 824-30 [0346]
- **HEINRICH, MC ; CORLESS CL ; DEMETRI GD et al.** Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J Clin Oncol*, 2003, vol. 21, 4342-4349 [0346]
- **LI B ; CHANG C ; YUAN M ; MCKENNA WG ; SHU HG.** Resistance to small molecule inhibitors of epidermal growth factor receptor in malignant gliomas. *Cancer Res*, 2003, vol. 63, 7443-7450 [0346]
- **C. L. SAWYERS.** *Genes Dev*, 2003, vol. 17, 2998-3010 [0346]
- **G. D. DEMETRI et al.** *N Engl J Med*, 2002, vol. 347, 472-80 [0346]
- **B. J. DRUKER et al.** *N Engl J Med*, 2001, vol. 344, 1038-42 [0346]
- **D. J. SLAMON et al.** *N Engl J Med*, 2001, vol. 344, 783-92 [0346]
- **H. DAVIES et al.** *Nature*, 2002, vol. 417, 949-54 [0346]
- **BARDELLI et al.** *Science*, 2003, vol. 300, 949 [0346]
- **Y. SAMUELS et al.** *Science*, 2004 [0346]
- **JEMAL et al.** *CA Cancer J Clin*, 2004, vol. 54, 8-29 [0346]
- **S. BREATHNACH et al.** *J Clin Oncol*, 2001, vol. 19, 1734-1742 [0346]
- **V. RUSCH et al.** *Cancer Res*, 1993, vol. 53, 2379-85 [0346]
- **R. BAILEY et al.** *Lung Cancer*, 2003, vol. 41 (S2), 71 [0346]
- **M. FUKUOKA et al.** *J Clin Oncol*, 2003, vol. 21, 2237-46 [0346]
- **P. A. JANNE et al.** *Lung Cancer*, 2004, vol. 44, 221-230 [0346]
- **M. G. KRIS et al.** *Jama*, 2003, vol. 290, 2149-58 [0346]
- **V. A. MILLER et al.** *J Clin Oncol*, 2004, vol. 22, 1103-9 [0346]
- **M. HUSE ; J. KURIYAN.** *Cell*, 2002, vol. 109, 275-82 [0346]

- **K. NAOKI ; T. H. CHEN ; W. G. RICHARDS ; D. J. SUGARBAKER ; M. MEYERSON.** *Cancer Res*, 2002, vol. 62, 7001-3 [0346]
- **J. STAMOS ; M. X. SLIWKOWSKI ; C. EIGENBROT.** *J Biol Chem*, 2002, vol. 277, 46265-72 [0346]
- **T. FUJISHITA et al.** *Oncology*, 2003, vol. 64, 399-406 [0346]
- **M. ONO et al.** *Mol Cancer Ther*, 2004, vol. 3, 465-472 [0346]
- **M. C. HEINRICH et al.** *J Clin Oncol*, 2003, vol. 21, 4342-9 [0346]
- **G. GIACCONE et al.** *J Clin Oncol*, 2004, vol. 22, 777-84 [0346]
- **R. S. HERBST et al.** *J Clin Oncol*, 2004, vol. 22, 785-94 [0346]
- **H. YAMAZAKI et al.** *Mol Cell Biol*, 1988, vol. 8, 1816-20 [0346]
- **M. E. GORRE et al.** *Science*, 2001, vol. 293, 876-80 [0346]
- **MARCHETTIA ; MARTELLA C ; FELICIONI L et al.** EGFR mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. *J Clin Oncol*, 2005, vol. 23, 857-65 [0346]
- **FRANKLIN WA.** Diagnosis of lung cancer: pathology of invasive and preinvasive neoplasia. *Chest*, 2000, vol. 117, 80S-89S [0346]
- **PAEZ JG ; JANNE PA ; LEE JC et al.** EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*, 2004, vol. 304, 1497-500 [0346]
- **LYNCH TJ ; BELL DW ; SORDELLA R et al.** Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med*, 2004, vol. 350, 2129-39 [0346]
- **PAO W ; MILLER V ; ZAKOWSKI M et al.** EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci U S A*, 2004, vol. 101, 13306-11 [0346]
- **HUANG SF ; LIU HP ; LI LH et al.** High frequency of epidermal growth factor receptor mutations with complex patterns in non-small cell lung cancers related to gefitinib responsiveness in Taiwan. *Clin Cancer Res*, 2004, vol. 10, 8195-203 [0346]
- **HAN SW ; KIM TY ; HWANG PG et al.** Predictive and Prognostic Impact of Epidermal Growth Factor Receptor Mutation in Non-Small-Cell Lung Cancer Patients Treated With Gefitinib. *J Clin Oncol*, 2005 [0346]
- **TOKUMO M ; TOYOOKA S ; KIURA K et al.** The relationship between epidermal growth factor receptor mutations and clinicopathologic features in non-small cell lung cancers. *Clin Cancer Res*, 2005, vol. 11, 1167-73 [0346]
- **MITSUDOMI T ; KOSAKA T ; ENDOH H et al.** Mutations of the Epidermal Growth Factor Receptor Gene Predict Prolonged Survival After Gefitinib Treatment in Patients with Non-Small-Cell Lung Cancer With Postoperative Recurrence. *J Clin Oncol*, 2005 [0346]
- **PAO W ; WANG TY ; RIELY GJ et al.** KRAS Mutations and Primary Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib. *PLoS Med*, 2005, vol. 2, e17 [0346]
- **READ WL ; PAGE NC ; TIERNEY RM et al.** The epidemiology of bronchioloalveolar carcinoma over the past two decades: analysis of the SEER database. *Lung Cancer*, 2004, vol. 45, 137-42 [0346]
- **SANDERSON COX L ; SLOAN JA ; PATTEN CA et al.** Smoking behavior of 226 patients with diagnosis of stage IIIA/IIIB non-small cell lung cancer. *Psychooncology*, 2002, vol. 11, 472-8 [0346]
- **RADZIKOWSKA E ; GLAZ P ; ROSZKOWSKI K.** Lung cancer in women: age, smoking, histology, performance status, stage, initial treatment and survival. Population-based study of 20 561 cases. *Ann Oncol*, 2002, vol. 13, 1087-93 [0346]
- **TONG L ; SPITZ MR ; FUEGER JJ et al.** Lung carcinoma in former smokers. *Cancer*, 1996, vol. 78, 1004-10 [0346]
- **DE PERROT M ; LICKER M ; BOUCHARDY C et al.** Sex differences in presentation, management, and prognosis of patients with non-small cell lung carcinoma. *J Thorac Cardiovasc Surg*, 2000, vol. 119, 21-6 [0346]
- **CAPEWELL S ; SANKARAN R ; LAMB D et al.** Lung cancer in lifelong non-smokers. Edinburgh Lung Cancer Group. *Thorax*, 1991, vol. 46, 565-8 [0346]
- **GRITZ ER ; NISENBAUM R ; ELASHOFF RE et al.** Smoking behavior following diagnosis in patients with stage I non-small cell lung cancer. *Cancer Causes Control*, 1991, vol. 2, 105-12 [0346]
- **SRIDHAR KS ; RAUB W A, JR.** Present and past smoking history and other predisposing factors in 100 lung cancer patients. *Chest*, 1992, vol. 101, 19-25 [0346]
- **BARBONE F ; BOVENZI M ; CAVALLIERI F et al.** Cigarette smoking and histologic type of lung cancer in men. *Chest*, 1997, vol. 112, 1474-9 [0346]
- **SHIGEMATSU H ; LIN L ; TAKAHASHI T et al.** Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst*, 2005, vol. 97, 339-46 [0346]
- **KOSAKA T ; YATABE Y ; ENDOH H et al.** Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. *Cancer Res*, 2004, vol. 64, 8919-23 [0346]

- **CHO D ; KOCHER O ; TENEN DG et al.** Unusual cases in multiple myeloma and a dramatic response in metastatic lung cancer: case 4. Mutation of the epidermal growth factor receptor in an elderly man with advanced, gefitinib-responsive, non-small-cell lung cancer. *J Clin Oncol*, 2005, vol. 23, 235-7 **[0346]**

Patentkrav

1. Fremgangsmåde til bestemmelse af sandsynligheden for effektivitet af en EGFR-tyrosinkinaseinhibitor til at behandle cancer i en patient, der er ramt af cancer, hvilken fremgangsmåde omfatter at bestemme, om erbB1-genet fra patienten omfatter mindst en nucleinsyreafvigelse, hvilken afvigelse er valgt blandt:

- a. en mutation i exon 18 der resulterer i en substitution af cystein for glycine ved kodon 719 i SEQ. ID. NO.: 511 (G719C) eller i en substitution af serin for glycine ved kodon 719 i SEQ. ID. NO.: 511 (G719S); eller
- 10 b. en in-frame-deletion i exon 19 der resulterer i en deletion af mindst aminosyrerne: leucin, arginin, glutaminsyre og alanin ved kodonerne 747, 748, 749 og 750 i SEQ. ID. NO.: 511; eller
- c. en mutation i exon 21 der resulterer i en aminosyresubstitution af arginin for leucin ved kodon 858 i SEQ. ID. NO.: 511 (L858R) eller af
- 15 glutamin for leucin i position 861 i SEQ. ID. NO.: 511 (L861Q).

2. Fremgangsmåden ifølge krav 1, hvor bestemmelsen af om erbB1-genet fra patienten omfatter mindst en af nucleinsyreafvigelserne, omfatter:

- A. at bestemme kinaseaktiviteten af erbB1-genet i en biologisk prøve fra patienten, hvor en forøgelse i kinaseaktivitet efter stimulering med en EGFR-ligand i sammenligning med en kontrol indikerer, at EGFR-tyrosinkinaseinhibitoren sandsynligvis er effektiv; eller
- 20 B. (a) at detektere tilstedeværelsen eller fraværet af mindst en nucleinsyreafvigelse i exon 18, 19 eller 21 ved at udføre en polymerasekædereaktion (PCR) for at amplificere en del af exon 18, 19 eller 21; og
- (b) at bestemme nucleotidsekvensen af den amplificerede nucleinsyre ved sekventering af mindst en del af den amplificerede exon 18, 19 eller
- 25 21; eller

C. (a) at opnå en biologisk prøve fra patienten; og

(b) at bestemme om Akt, STAT5 eller STAT3 er aktiveret i patienten, hvor aktiveret Akt, STAT5 eller STAT3 indikerer, at den EGFR-målrettede behandling sandsynligvis vil være effektiv.

5 **3.** Fremgangsmåden ifølge krav 1, hvor tilstedeværelsen af den mindst ene nucleinsyreafvigelse indikerer, at den EGFR-målrettede behandling sandsynligvis vil være effektiv.

10 **4.** Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 3, hvor canceren vælges blandt gastrointestinal cancer, prostatacancer, ovariecancer, brystcancer, hoved- og halscancer, lungecancer, ikke-småcellet lungecancer (NSCLC), cancer i nervesystemet, nyrecancer, retinacancer, hudcancer, levercancer, pancreascancer, genital-urinalcancer og blærecancer.

15 **5.** Fremgangsmåden ifølge krav 4, hvor canceren er NSCLC.

6. Fremgangsmåde ifølge et hvilket som helst af de foregående krav, hvor erbB1-genet opnås fra en biologisk prøve, der er opnået fra patienten.

20 **7.** Fremgangsmåden ifølge krav 6, hvor den biologiske prøve er en prøve af væv eller væske, der er isoleret fra individet, og vælges blandt en tumorbiopsi-prøve, tumorprøve, resektion, bronkoskopisk biopsi og celleblok fra pleuralvæske.

25 **8.** Fremgangsmåde ifølge et hvilket som helst af de foregående krav, hvor bestemmelsen af om erbB1-genet omfatter mindst en nucleinsyreafvigelse, anvender DNA eller RNA der er opsamlet fra biologiske prøver og udføres ved polymerasekædereaktion (PCR), hybridisering med allelspecifikke prober, enzymatisk mutationsdetektion, kemisk spaltning af mismatches, massespektrometri
30 eller DNA-sekventering.

9. Fremgangsmåden ifølge krav 8, hvor amplifikationsprodukterne i PCR separeres ved anvendelse af sekvensforskelle ved at anvende: enkelstrengskonformationspolymorfi (SSCP), denatureringsgradientgelelektroforese (DGGE), temperaturfølsom gelelektroforese (TSGE), kemisk spaltning eller restriktionsfragmentpolymorfier, eller hybridisering til et nucleinsyre-array.

10. EGFR-tyrosinkinaseinhibitor til anvendelse ved behandling af cancer i en patient, der er ramt af cancer, ved at administrere en effektiv mængde af tyrosinkinaseinhibitoren til patienten, hvor kinasedomænet af erbB1-genet i patienten er blevet bestemt til at indeholde en form af erbB1-genet, der omfatter mindst en nucleinsyreafvigelse, hvilket indikerer, at en EGFR-målet behandling vil være effektiv i patienten, hvilken nucleinsyreafvigelse er valgt blandt:

15 a. en mutation i exon 18 der resulterer i en substitution af cystein for glycine ved kodon 719 i SEQ. ID. NO.: 511 (G719C) eller i en substitution af serin for glycine ved kodon 719 i SEQ. ID. NO.: 511 (G719S); eller

b. en in-frame-deletion i exon 19 der resulterer i en deletion af mindst aminosyrerne: leucin, arginin, glutaminsyre og alanin ved kodonerne 747, 748, 749 og 750 i SEQ. ID. NO.: 511; eller

20 c. en mutation i exon 21 der resulterer i en aminosyresubstitution af arginin for leucin ved kodon 858 i SEQ. ID. NO.: 511 (L858R) eller af glutamin for leucin ved kodon 861 i SEQ. ID. NO.: 511 (L861Q);

hvor bestemmelsen af sandsynligheden for effektiviteten af en EGFR-målet behandling i den berørte patient udføres ved fremgangsmåden ifølge et hvilket som helst af kravene 1 til 9.

25

11. Anvendelse af primere i en fremgangsmåde ifølge et hvilket som helst af kravene 1 til 9, hvilke primere er udvalgt blandt:

A. primere til amplificering af en targetnucleinsyre i kinasedomænet i erbB1-genet, hvilken targetnucleinsyre omfatter en nucleinsyreafvigelse,

som er valgt blandt:

- a. en mutation i exon 18 der resulterer i en substitution af cystein for glycine ved kodon 719 i SEQ. ID. NO.: 511 (G719C) eller i en substitution af serin for glycine ved kodon 719 i SEQ. ID. NO.: 511 (G719S); eller
- 5 b. en in-frame-deletion i exon 19 der resulterer i en deletion af mindst aminosyrerne: leucin, arginin, glutaminsyre og alanin ved kodonerne 747, 748, 749 og 750 i SEQ. ID. NO.: 511; eller
- c. en mutation i exon 21 der resulterer i en aminosyresubstitution af arginin for leucin ved kodon 858 i SEQ. ID. NO.: 511 (L858R) eller af
- 10 glutamin for leucin ved kodon 861 i SEQ. ID. NO.: 511 (L861Q); og

B. en primer som er:

B (1) et primer-par der er valgt blandt følgende:

- SEQ ID NO.: 49 og 50 (exon 18);
- SEQ ID NO.: 51 og 52 (exon 19);
- 15 SEQ ID NO.: 55 og 56 (exon 21);
- SEQ ID NO.: 113 og 114 (exon 18);
- SEQ ID NO.: 261 og 262 (exon 18);
- SEQ ID NO.: 263 og 264 (exon 19);
- SEQ ID NO.: 267 og 268 (exon 21);
- 20 SEQ ID NO.: 325 og 326 (exon 21);
- SEQ ID NO.: 505 og 506 (exon 19);
- SEQ ID NO.: 507 og 508 (exon 21);

SEQ ID NO.: 687 og 508 (exon 19);

SEQ ID NO.: 653 og 646 (exon 18);

SEQ ID NO.: 654 og 647 (exon 19);

SEQ ID NO.: 656 og 649 (exon 21);

5 SEQ ID NO.: 660 og 667 (exon 18);

SEQ ID NO.: 661 og 668 (exon 19);

SEQ ID NO.: 663 og 670 (exon 21); eller

B (2) enhver af de primere der er omtalt i B(1); eller

B (3) enhver anden kombination af de ovennævnte primere.

10 **12.** Anvendelse af en probe til at skelne tilstedeværelsen af en bestemt afvigelse eller afvigelser i erbB1-genet, hvilken afvigelse eller afvigelser er valgt blandt:

15 a. en mutation i exon 18 der resulterer i en substitution af cystein for glycine ved kodon 719 i SEQ. ID. NO.: 511 (G719C) eller i en substitution af serin for glycine ved kodon 719 i SEQ. ID. NO.: 511 (G719S); eller

b. en in-frame-deletion i exon 19 der resulterer i en deletion af mindst aminosyrerne: leucin, arginin, glutaminsyre og alanin ved kodonerne 747, 748, 749 og 750 i SEQ. ID. NO.: 511; eller

20 c. en mutation i exon 21 der resulterer i en aminosyresubstitution af arginin for leucin ved kodon 858 i SEQ. ID. NO.: 511 (L858R) eller af glutamin for leucin ved kodon 861 i SEQ. ID. NO.: 511 (L861Q).

13. Anvendelsen ifølge krav 12, hvilken anvendelse er ifølge en fremgangsmåde ifølge et hvilket som helst af kravene 1 til 9.

14. Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 9, eller EGFR-tyrosinkinaseinhibitoren til anvendelse ved behandling af cancer ifølge krav 10, eller anvendelsen ifølge krav 11 eller krav 12 eller krav 13, ved hvilken afvigelsen eller afvigelserne er valgt blandt:

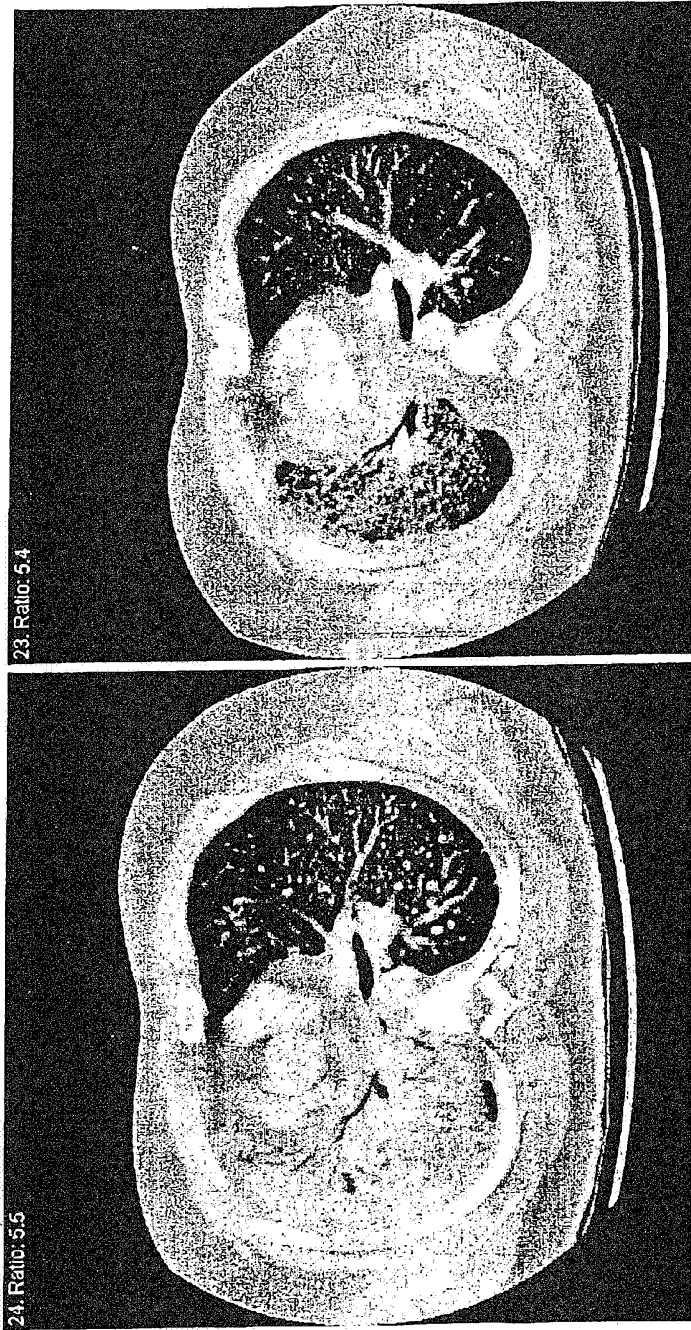
- 5 (i) en substitution i exon 18 af EGFR der omfatter en substitution af et thymin for en guanin ved nucleotid 2155 i SEQ ID NO.: 511 ifølge nummereringen der er anvendt i figur 5, hvilket resulterer i en aminosyresubstitution, hvor vildtype-glycin er substitueret med en cystein ved kodon 719 i SEQ ID NO.: 511;
- 10 (ii) en substitution i exon 18 af EGFR der omfatter en substitution af en adenin for en guanin ved nucleotid 2155 i SEQ ID NO.: 511 i overensstemmelse med nummereringen der er anvendt i figur 5, hvilket resulterer i en aminosyresubstitution, hvor vildtype-glycin er substitueret med en serin ved kodon 719 i SEQ ID NO.: 511;
- 15 (iii) en in-frame-deletion i exon 19 af EGFR (erbB1) der omfatter deletion af mindst aminosyrerne: leucin, arginin, glutaminsyre og alanin ved kodonerne 747, 748, 749 og 750 i SEQ ID NO.: 511;

 (iv) en in-frame-deletion i exon 19 der omfatter en deletion af aminosyrerne 746 til 750 i SEQ ID NO.: 511;
- 20 (v) en in-frame-deletion der omfatter nucleotiderne 2235 til 2249 i SEQ ID NO.: 511 ifølge nummereringen der er anvendt i figur 5, som deleterer aminosyrerne 746 til 750 i SEQ ID NO.: 511 som er sekvensen: glutaminsyre, leucin, arginin, glutaminsyre og alanin;
- 25 (vi) en in-frame-deletion der omfatter nucleotiderne 2240 til 2251 i SEQ ID NO.: 511 ifølge nummereringen der er anvendt i figur 5;
- (vii) en in-frame-deletion der omfatter nucleotiderne 2240 til 2257 i SEQ ID NO.: 511 ifølge nummereringen der er anvendt i figur 5;

- (viii) en in-frame-deletion der omfatter nucleotiderne 2239 til 2247 sammen med en substitution af cytosin for guanin ved nucleotid 2248 i SEQ ID NO.: 511 ifølge nummereringen der er anvendt i figur 5;
- 5 (ix) en in-frame-deletion i exon 19 der omfatter en deletion af nucleotiderne 2239 til 2250 sammen med en substitution af cytosin for adenin ved nucleotid 2251 i SEQ ID NO.: 511 ifølge nummereringen der er anvendt i figur 5;
- 10 (x) en in-frame-deletion der deleterer nucleotiderne 2238 til 2255 sammen med en substitution af thymin for adenin ved nucleotid 2237 i SEQ ID NO.: 511 ifølge nummereringen der er anvendt i figur 5;
- (xi) en in-frame-deletion der omfatter nucleotiderne 2239 til 2250 delT-TAAGAGAAGCA i SEQ ID NO.: 511 ifølge nummereringen der er anvendt i figur 5;
- 15 (xii) en in-frame-deletion der omfatter nucleotiderne 2240-2254 delTAA-GAGAAGCA i SEQ ID NO.: 511 ifølge nummereringen der er anvendt i figur 5;
- (xiii) en mutation i exon 21 der omfatter en substitution af guanin for thymin ved nucleotid 2573 i SEQ ID NO.: 511 ifølge nummereringen der er anvendt i figur 5;
- 20 (xiv) en mutation i exon 21 der omfatter en substitution af adenin for thymin ved nucleotid 2582 i SEQ ID NO.: 511 ifølge nummereringen der er anvendt i figur 5.

DRAWINGS

Figure 1 A and B



A B

Figures 2A - 2E

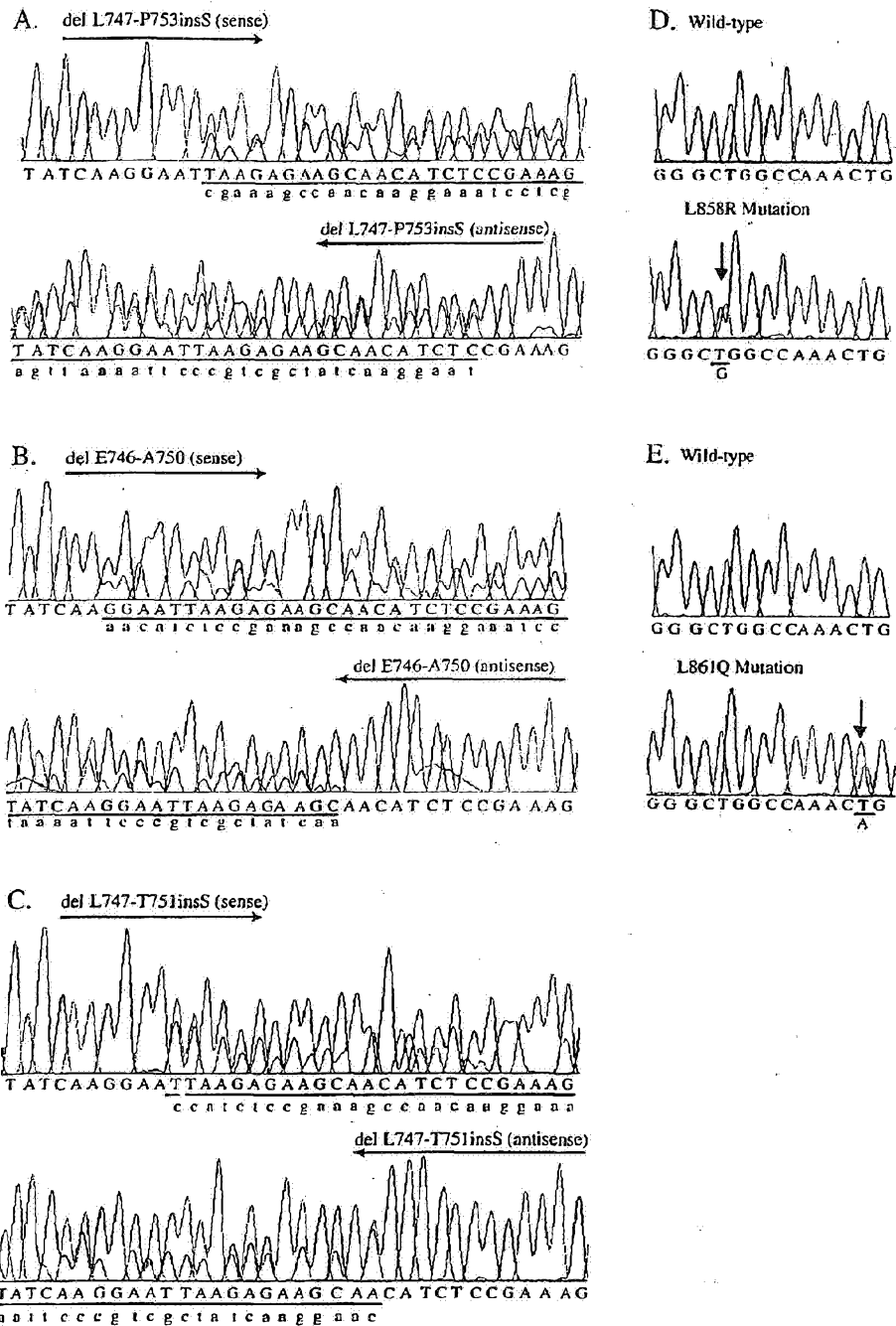
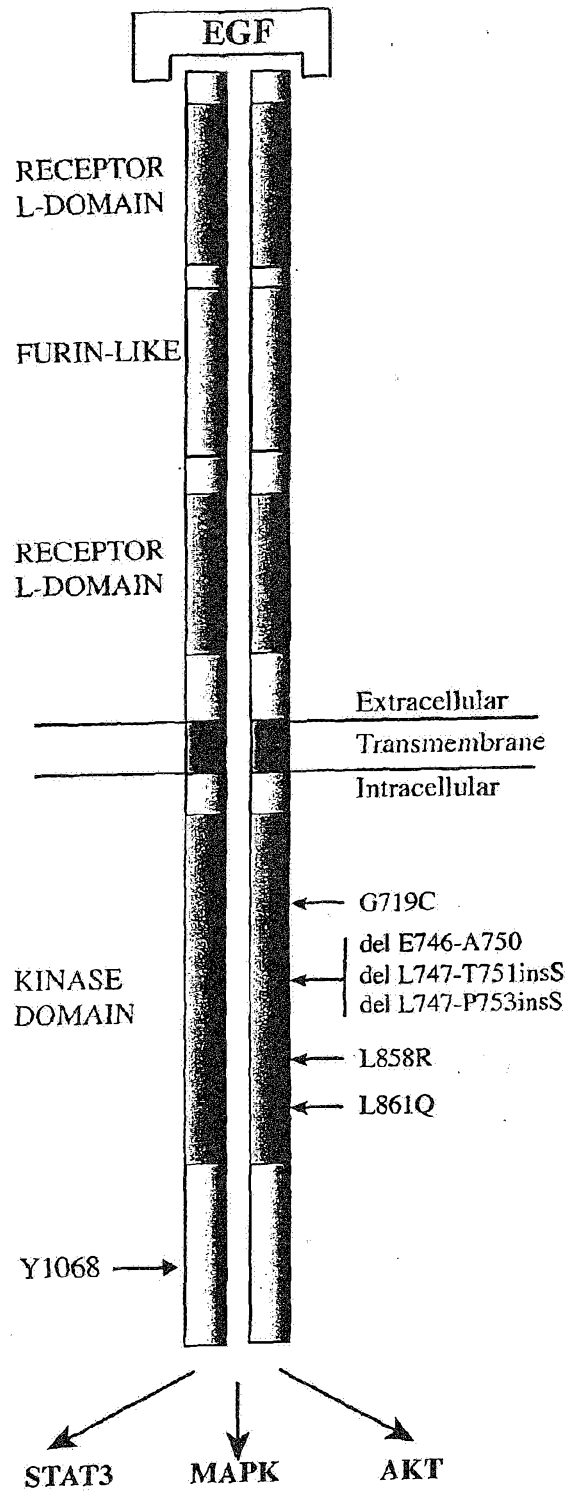
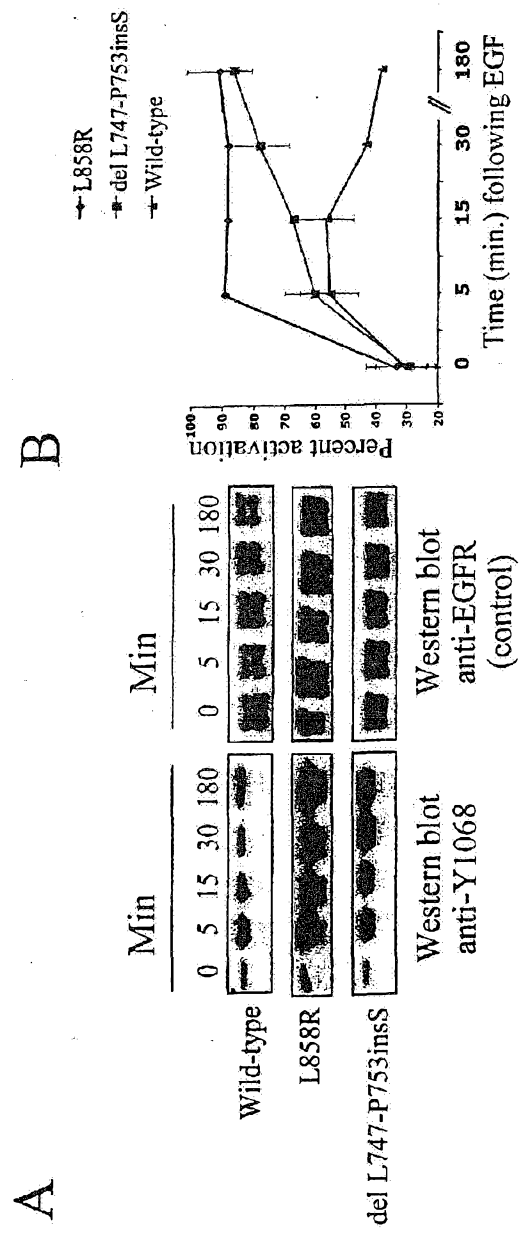


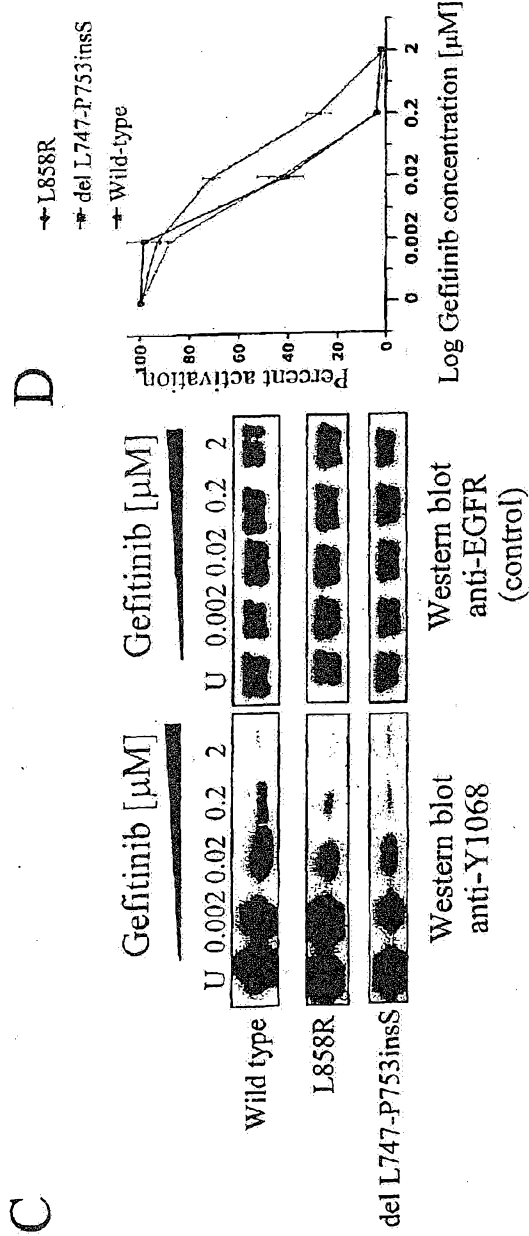
Figure 2F



Figures 3A and 3B



Figures 3C and 3D



Figures 4B-C

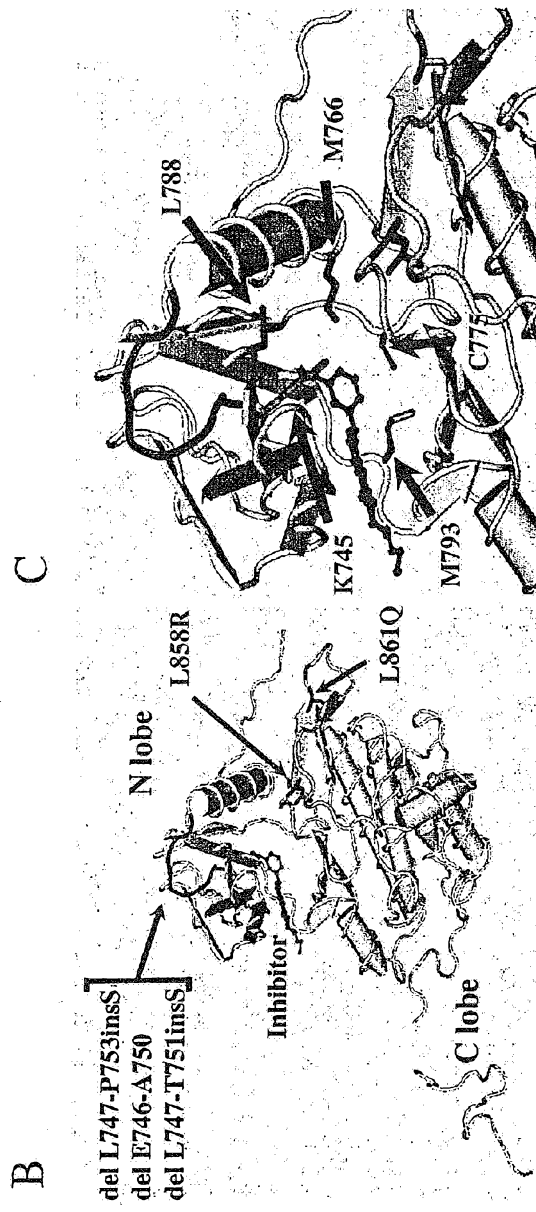


Figure 5

CCGGGCGCAGCGCGGCGCGCAGCAGCCTCCGCCCCCGCACGGTGTGAGCGCCCGACGCGG -185

 CCGAGGCGGCGCGAGTCCCAGCTAGCCCCGGCGCGCGCGCGCCAGACCGGACGACA -125

 GGCCACCTCGTCGGCGTCCGCCCCGAGTCCCCGCCTCGCCGCCAACGCCACAACCACCGCG -65

 CACGGCCCCCTGACTCCGTCCAGTATTGATCGGGAGAGCCGGAGCGAGCTCTTCGGGGAG -5

 CAGCGATCGCACCTCCGGGACGGCGGGGCGAGCGCTCCTGGCGCTGCTGGCTGCGCTCT 55
-M--R--P--S--G--T--A--G--A--A--L--L--A--L--L--A--A--L-- 18
 GCGGCGGAGTCCGGGCTCTGGAGGAAAAGAAAGTTTGCCAAGGCACGAGTAACAAGCTCA 115
 C--P--A--S--R--A--L--E--E--K--K--V--C--Q--G--T--S--N--K--L-- 38
 CGCAGTTGGGCACTTTTGAAGATCATTTTCTCAGCCTCCAGAGGATGTTCATAAAGTGTG 175
 T--Q--L--G--T--F--E--D--H--F--L--S--L--Q--R--M--F--N--N--C-- 58
 AGGTGGTCTTGGGAATTTGGAATTTACCTATGTGCAGAGGAATTATGATCTTTCTTCT 235
 E--V--V--L--G--N--L--E--I--T--Y--V--Q--R--N--Y--D--L--S--F-- 78
 TAAAGACCATCCAGGAGGTGGCTGCTTATGTCTCATTCGCTCAACACAGTGGAGCGAA 295
 L--K--T--I--Q--E--V--A--G--Y--V--L--I--A--L--N--T--V--E--R-- 98
 TTCCTTTGGAAAACCTGCAGATCATCAGAGGAAATATGTACTACGAAAATTCCTATGCCT 355
 I--P--L--E--N--L--Q--I--I--R--G--N--M--Y--Y--E--N--S--Y--A-- 118
 TAGCAGTCTTATCTAACTATGATGCAAATAAAACCGACTGAAGGAGCTGCCCATGAGAA 415
 L--A--V--L--S--N--Y--D--A--N--K--T--G--L--K--E--L--P--M--R-- 138
 ATTTACAGGAAATCCTGCATGGCGCGTTCGGTTTTCAGCAACAACCTGCCCTGTGCAACG 475
 N--L--Q--E--I--L--H--G--A--V--R--F--S--N--N--P--A--L--C--N-- 158
 TGGAGAGCATCCAGTGGCGGGACATAGTCAGCAGTGAAGTCTTCTCAGCAACATGTCGATGG 535
 V--E--S--I--Q--W--R--D--I--V--S--S--D--F--L--S--N--M--S--M-- 178
 ACTTCAGAACCACTGGGCGAGCTGCCAAAAGTGTGATCCAAGCTGTCCCAATGGGAGCT 595
 D--F--Q--N--H--L--G--S--C--Q--K--C--D--P--S--C--P--N--G--S-- 198
 GCTGGGGTGCAGGAGAGGAGAACTGCCAGAAACTGACCAAAATCATCTGTGCCCGAGCAGT 655
 C--W--G--A--G--E--E--N--C--Q--K--L--T--K--I--I--C--A--Q--Q-- 218
 GCTCCGGGCGCTGCCGTGGCAAGTCCCCAGTGAAGTCTGCCACAACCAAGTGTGCTGCAG 715
 C--S--G--R--C--R--G--K--S--P--S--D--C--C--H--N--Q--C--A--A-- 238
 GCTGCACAGGCCCCCGGGAGAGCGACTGCCTGGTCTGCCGCAAATTCGAGACGAAGCCA 775
 G--C--T--G--P--R--E--S--D--C--L--V--C--R--K--F--R--D--E--A-- 258
 CGTGCAAGGACACCTGCCCCCACTCATGCTCTACAACCCCAACACGTACCAGATGGATG 835
 T--C--K--D--T--C--P--P--L--M--L--Y--N--P--T--T--Y--Q--M--D-- 278
 TGAACCCCGAGGGCAAATACAGCTTTGGTGCCACCTGCGTGAAGAAGTGTCCCGTAATT 895
 V--N--P--E--G--K--Y--S--F--G--A--T--C--V--K--K--C--P--R--N-- 298

Figure 5 (cont.)

ATGTGGTGACAGATCACGGCTCGTGCCTCCGAGCCTGTGGGGCCGACAGCTATGAGATGG 955
 Y--V--V--T--D--H--G--S--C--V--R--A--C--G--A--D--S--Y--E--M-- 318
 AGCAAGACGGCGTCCGCAAGTGTAAAGAAGTGCAGAGGGCCTTGCCGCAAAGTGTGTAACG 1015
 E--E--D--G--V--R--K--C--K--K--C--E--G--P--C--R--K--V--C--N-- 338
 GAATAGGTATTGGTGAATTTAAAGACTCACTCTCCATAAATGCTACGAATATTAAACACT 1075
 G--I--G--I--G--E--F--K--D--S--L--S--I--N--A--T--N--I--K--H-- 358
 TCAAAAACCTGCACCTCCATCAGTGGCGATCTCCACATCCTGCCGGTGGCATTAGGGGTG 1135
 F--K--N--C--T--S--I--S--G--D--L--H--I--L--P--V--A--F--R--G-- 378
 ACTCCTTCACACATACTCCTCCTCTGGATCCACAGGAAGTGGATATTCTGAAAACCGTAA 1195
 D--S--F--T--H--T--P--P--L--D--P--Q--E--L--D--I--L--K--T--V-- 398
 AGGAAATCACAGGGTITTTGCTGATTTCAGGCTTGGCCTGAAAACAGGACCGACCTCCATG 1255
 K--E--I--T--G--F--L--L--I--Q--A--W--P--E--N--R--T--D--L--H-- 418
 CCTTTGAGAACCTAGAAATCATACGCGGCAGGACCAAGCAACATGGTCAGTTTTCTCTTG 1315
 A--F--E--N--L--E--I--I--R--G--R--T--K--Q--H--G--Q--F--S--L-- 438
 CAGTCGTGAGCCTGAACATAACATCCTTGGGATTACGCTCCCTCAAGGAGATAAGTGATG 1375
 A--V--V--S--L--N--I--T--S--L--G--L--R--S--L--K--E--I--S--D-- 458
 GAGATGTGATAATTTTCAGGAAACAAAAATTTGTGCTATGCAAATACAATAAACTGGAAAA 1435
 G--D--V--I--I--S--G--N--K--N--L--C--Y--A--N--T--I--N--W--K-- 478
 AACTGTTTGGGACCTCCGGTCAGAAAACCAAATTTATAAGCAACAGAGGTGAAAACAGCT 1495
 K--L--F--G--T--S--Q--K--T--K--I--I--S--N--R--G--E--N--S-- 498
 GCAAGGCCACAGGCCAGGTCTGCCATGCCTTGTGCTCCCCGAGGGCTGCTGGGGCCCGG 1555
 C--K--A--T--G--Q--V--C--H--A--L--C--S--P--E--G--C--W--G--P-- 518
 AGCCCAGGGAAGTGCCTCTCTTGCCTGGAATGTGAGCCGAGGAGGGAATGCGTGGACAAGT 1615
 E--F--R--D--C--V--S--C--R--N--V--S--R--G--R--E--C--V--D--K-- 538
 GCAACCTTCTGGAGGGTGAGCCAAGGAGTTTGTGGAGAAGTCTGAGTGCATACAGTGCC 1675
 C--N--L--L--E--G--E--P--R--E--F--V--E--N--S--E--C--I--Q--C-- 558
 ACCCAGAGTGCCTGCCTCAGGCCATGAACATCACCTGCACAGGACGGGGACCAGACAACT 1735
 H--P--E--C--L--P--Q--A--M--N--I--T--C--T--G--R--G--P--D--N-- 578
 GTATCCAGTGTGCCCACTACATTGACGGCCCCCACTGCGTCAAGACCTGCCCGGCAGGAG 1795
 C--I--Q--C--A--H--Y--I--D--G--P--H--C--V--K--T--C--P--A--G-- 598
 TCATGGGAGAAAAACAACCCCTGGTCTGGAAGTACGCAGACGCCGGCCATGTGTGCCACC 1855
 V--M--G--E--N--N--T--L--V--W--K--Y--A--D--A--G--H--V--C--H-- 618
 TGTGCCATCCAACTGCACCTACGGATGCACCTGGGCCAGGTCTTGAAGGCTGTCCAACGA 1915
 L--C--H--P--N--C--T--Y--G--C--T--G--P--G--L--E--G--C--P--T-- 638
 ATGGGCCTAAGATCCCGTCCATCGCCACTGGGATGGTGGGGGCCCTCCTCTTGCTGCTGG 1975
 N--G--P--K--I--P--S--I--A--T--G--M--V--G--A--L--L--L--L--L-- 658
 TGGTGGCCCTGGGGATCGGCCTCTTCATGCGAAGGCGCCACATCGTTCCGAAGCGCACGC 2035
 V--V--A--L--G--I--G--L--F--M--R--R--R--H--I--V--R--K--R--T-- 678

Figure 5 (cont.)

TGC GGAGGCTGCTGCAGGAGAGGGAGCTTGTGGAGCCTCTTACACCCAGTGGAGAAGCTC 2095
 L--R--R--L--L--Q--E--R--E--L--V--E--P--L--T--P--S--G--E--A-- 698

 CCAACCAAGCTCTCTTGAGGATCTTGAAGGAACTGAATTCAAAAAGATCAAAGTGCTGG 2155
 P--N--Q--A--L--L--R--I--L--K--E--T--E--F--K--K--I--K--V--L-- 718

 GCTCCGGTGCCTTCGGCACGGTGTATAAGGGACTCTGGATCCCAGAAGGTGAGAAAGTTA 2215
 G--S--G--A--F--G--T--V--Y--K--G--L--W--I--P--E--G--E--K--V-- 738

 AAATTCCCCTGCTATCAAGGAATTAAAGAGAAGCAACATCTCCAAAGCCAACAAGGAAA 2275
 K--I--P--V--A--I--K--E--L--R--E--A--T--S--P--K--A--N--K--E-- 758

 TCCTCGATGAAGCCTACGTGATGGCCAGCGTGGACAACCCCCACGTGTGCCGCTGCTGG 2335
 I--L--D--E--A--Y--V--M--A--S--V--D--N--P--H--V--C--R--L--L-- 778

 GCATCTGCCTCACCTCCACCGTGCAGCTCATCACGCAGCTCATGCCCTTCGGCTGCCTCC 2395
 G--I--C--L--T--S--T--V--Q--L--I--T--Q--L--M--P--F--G--C--L-- 798

 TGGACTATGTCCGGGAACACAAAGACAATATTGGCTCCCAGTACCTGCTCAACTGGTGTG 2455
 L--D--Y--V--R--E--H--K--D--N--I--G--S--Q--Y--L--L--N--W--C-- 818

 TGCAGATCGCAAAGGGCATGAACCTACTTGGAGGACCGTGCCTTGGTGCACCGCGACCTGG 2515
 V--Q--I--A--K--G--M--N--Y--L--E--D--R--R--L--V--H--R--D--L-- 838

 CAGCCAGGAACGTACTGCTGAAAACACCGCAGCATGTCAAGATCACAGATTTTGGGCTGG 2575
 A--A--R--N--V--L--V--K--T--P--Q--H--V--K--I--T--D--F--G--L-- 858

 CCAAAGTCTGGGTGCGGAAGAGAAAGAATACCATGCAGAAGGAGGCAAAGTGCCTATCA 2635
 A--K--L--L--G--A--E--E--K--E--Y--H--A--E--G--G--K--V--P--I-- 878

 AGTGGATGGCATTTGGAATCAATTTTACACAGAATCTATACCCACCAGAGTGATGTCTGGA 2695
 K--W--M--A--L--E--S--I--L--H--R--I--Y--T--H--Q--S--D--V--W-- 898

 GCTACGGGGTGACTGTTTGGGAGTTGATGACCTTTGGATCCAAGCCATATGACGGAATCC 2755
 S--Y--G--V--T--V--W--E--L--M--T--F--G--S--K--P--Y--D--G--I-- 918

 CTGCCAGCGAGATCTCCTCCATCCTGGAGAAAGGAGAACGCCTCCCTCAGCCACCCATAT 2815
 P--A--S--E--I--S--S--I--L--E--K--G--E--R--L--P--Q--P--P--I-- 938

 GTACCATCGATGTCTACATGATCATGGTCAAGTGCTGGATGATAGACGCAGATAGTCGCC 2875
 C--T--I--D--V--Y--M--I--M--V--K--C--W--M--I--D--A--D--S--R-- 958

 CAAAGTTCCGTGAGTTGATCATCGAATTTCTCCAAATGGCCCGAGACCCCGAGCGCTACC 2935
 P--K--F--R--E--L--I--I--E--F--S--K--M--A--R--D--P--Q--R--Y-- 978

 TTGTCAATTCAGGGGGATGAAAGAATGCATTTGCCAAGTCCTACAGACTCCAACCTCTACC 2995
 L--V--I--Q--G--D--E--R--M--H--L--P--S--P--T--D--S--N--F--Y-- 998

 GTGCCCTGATGGATGAAGAAGACATGGACGACGTGGTGGATGCCGACGAGTACCTCATCC 3055
 R--A--L--M--D--E--E--D--M--D--D--V--V--D--A--D--E--Y--L--I-- 1018

 CACAGCAGGGCTTCTTCAGCAGCCCCCTCCACGTACCGGACTCCCCTCCTGAGCTCTCTGA 3115
 P--Q--Q--G--F--F--S--S--P--S--T--S--R--T--P--L--L--S--S--L-- 1038

 GTGCAACCAGCAACAATTCACCGTGGCTTGCAATTGATAGAAATGGGCTGCAAAGCTGTC 3175
 S--A--T--S--N--N--S--T--V--A--C--I--D--R--N--G--L--Q--S--C-- 1058

Figure 5 (cont.)

CCATCAAGGAAGACAGCTTCTTGCGAGGATACAGCTCAGACCCACAGGCGCCTTGACTG 3235
 P--I--K--E--D--S--F--L--Q--R--Y--S--S--D--P--T--G--A--L--T-- 1078

 AGGACAGCATAGACGACACCTTCCTCCAGTGCCTGAATACATAAACAGTCCGTTCCCA 3295
 E--D--S--I--D--D--T--F--L--P--V--P--E--Y--I--N--Q--S--V--P-- 1098

 AAAGGCCCGCTGGCTCTGTGCAGAAATCCTGTCTATCACAATCAGCCTCTGAACCCCGCGC 3355
 K--R--P--A--G--S--V--Q--N--P--V--Y--H--N--Q--P--L--N--P--A-- 1118

 CCAGCAGAGACCCACACTACCAGGACCCCCACAGCACTGCAGTGGGCAACCCCGAGTATC 3415
 P--S--R--D--P--H--Y--Q--D--P--H--S--T--A--V--G--N--P--E--Y-- 1138

 TCAACACTGTCCAGCCCACCTGTGTCAACAGCACATTCGACAGCCCTGCCCACTGGGCCC 3475
 L--N--T--V--Q--P--T--C--V--N--S--T--F--D--S--P--A--H--W--A-- 1158

 AGAAAGGCAGCCACCAAATTAGCCTGGACAACCCTGACTACCAGCAGGACTTCTTTCCCA 3535
 Q--K--G--S--H--Q--I--S--L--D--N--P--D--Y--Q--Q--D--F--F--P-- 1178

 AGGAAGCCAAGCCAAATGGCATCTTTAAGGGCTCCACAGCTGAAAATGCAGAATACCTAA 3595
 K--E--A--K--P--N--G--I--F--K--G--S--T--A--E--N--A--R--Y--L-- 1198

 GGGTCGCGCCACAAAGCAGTGAATTTATTGGAGCATGA 3633 (SEQ ID NO 511)
 R--V--A--P--Q--S--S--E--F--I--G--A--* 1210 (SEQ ID NO 512)

FIG. 6

A

			SEQ ID NO
		activation loop	
L858R	KTPQHVKITDFGR	AKLLGAEEKEYH	870 477
EGFR	KTPQHVKITDFGL	AKLLGAEEKEYH	870 478
BRAF	HEDLTVKIGD	ECIATVKS	608 479
	***	*****	

B

		P-loop	
G719S	ETEFKKIKVL	SSGAFGT	733 480
EGFR	ETEFKKIKVL	SSGAFGT	733 481
BRAF	DGQITVGQRI	CSGSGT	477 482
	***	*****	*

C

	742	750 752	
Del-1	VAIK	ET-SPKANKEILDEAYV	765 483
Del-2	VAIKELREAT-	ET-SPKANKEILDEAYV	765 484
Del-3	VAIKE	ET-SPKANKEILDEAYV	765 485
Del-4	VAIKE	ET-SPKANKEILDEAYV	765 486
Del-5	VAIKV	ET-SPKANKEILDEAYV	765 487
EGFR	VAIKELREAT-	SPKANKEILDEAYV	765 488
BRAF	VAVKMLNVTAPT	PQQLQAFKNEVG	503 489
	** *	* *	

FIG. 7

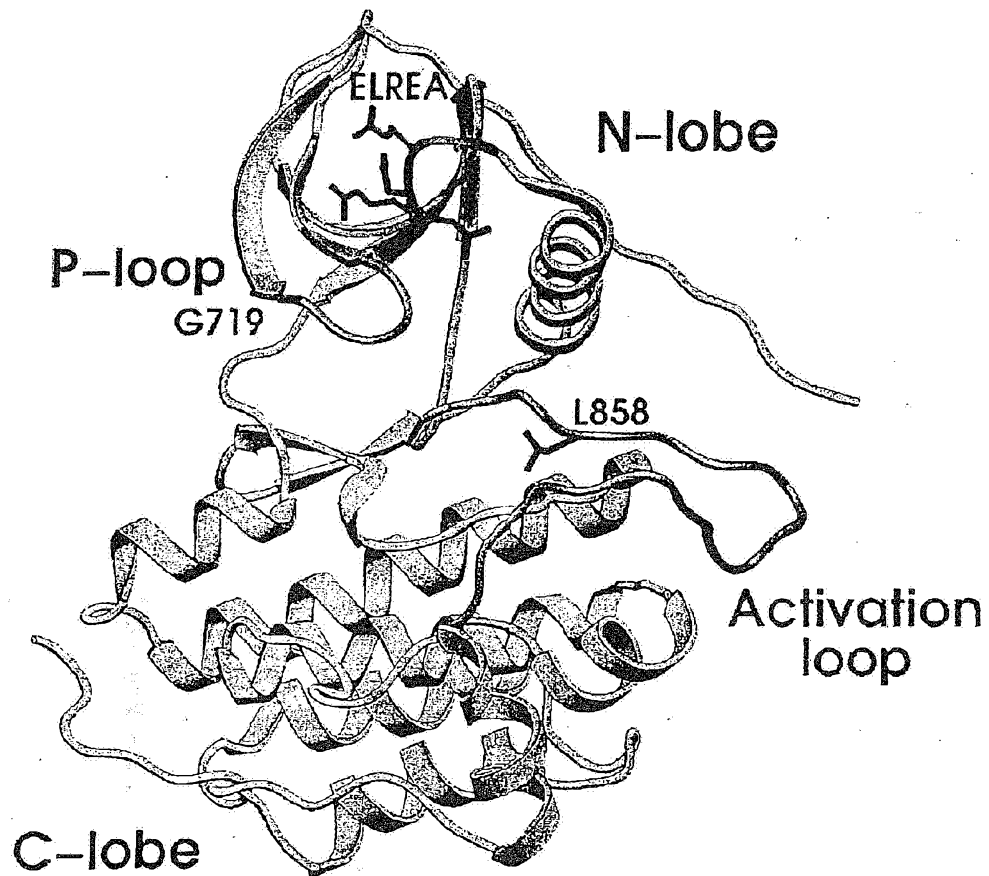


FIG. 8

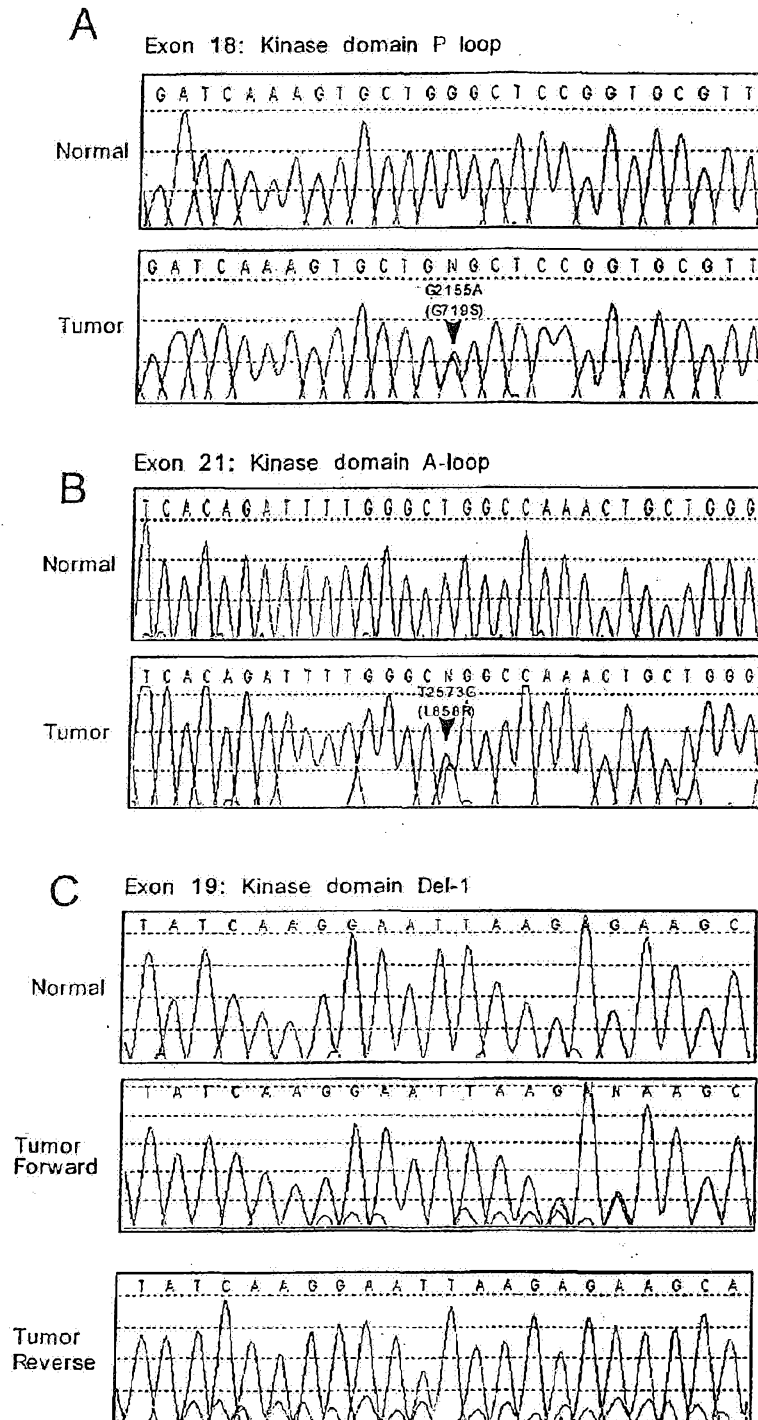


FIG. 8 (cont.)

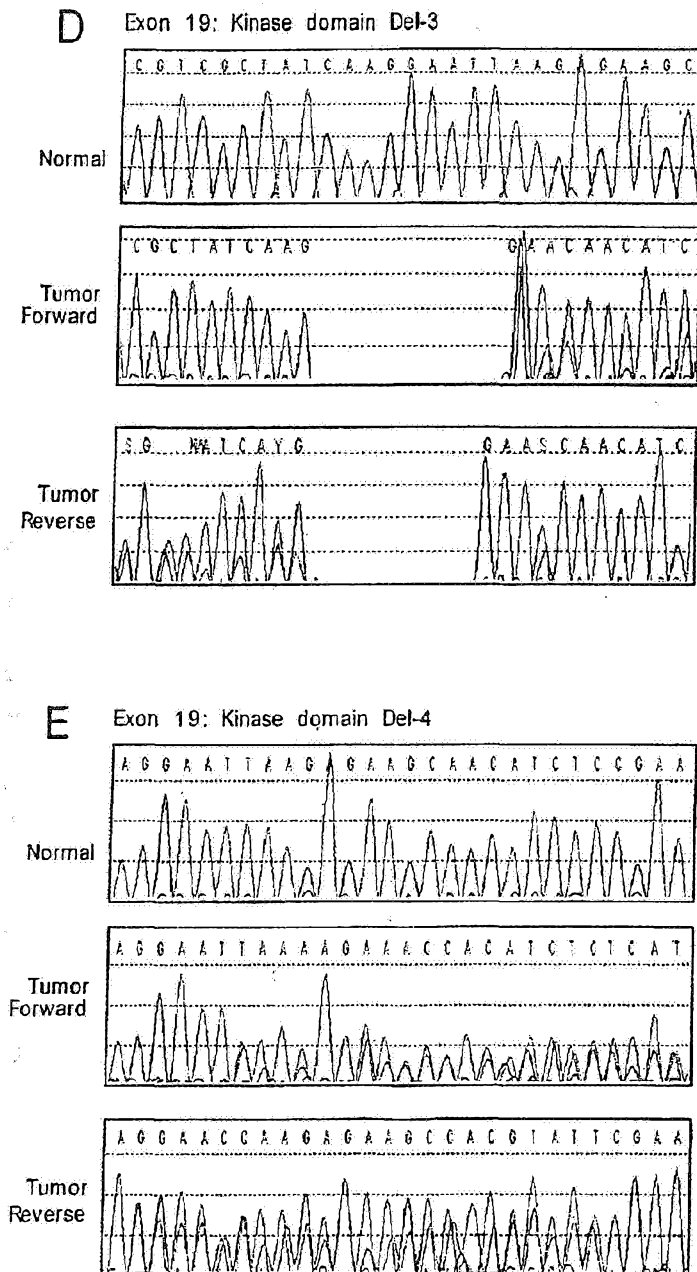


FIG. 8 (cont.)

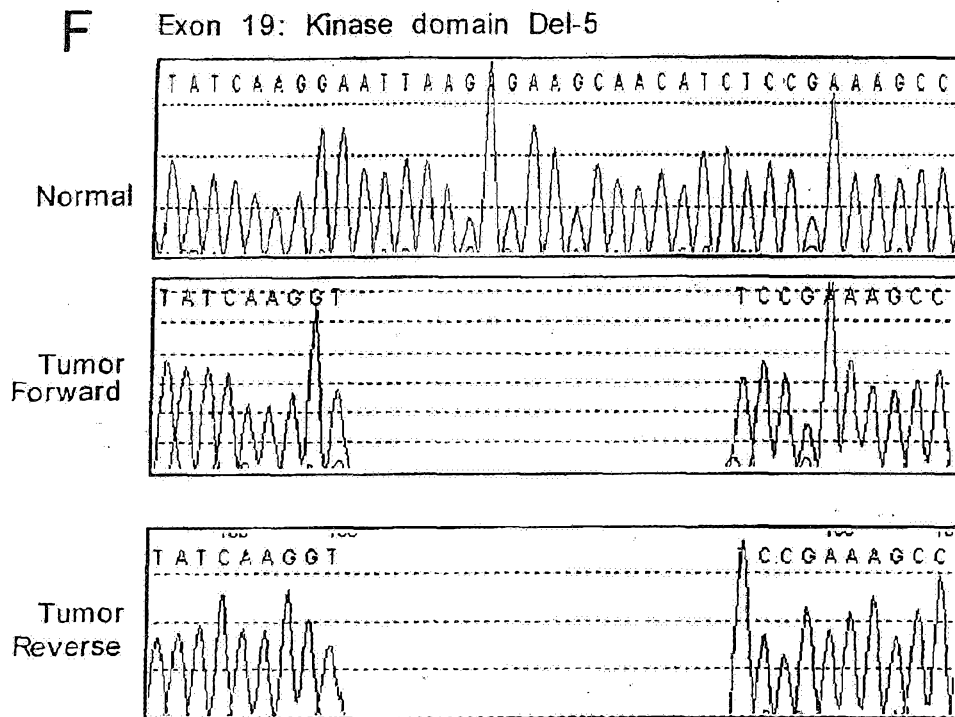


FIG 9

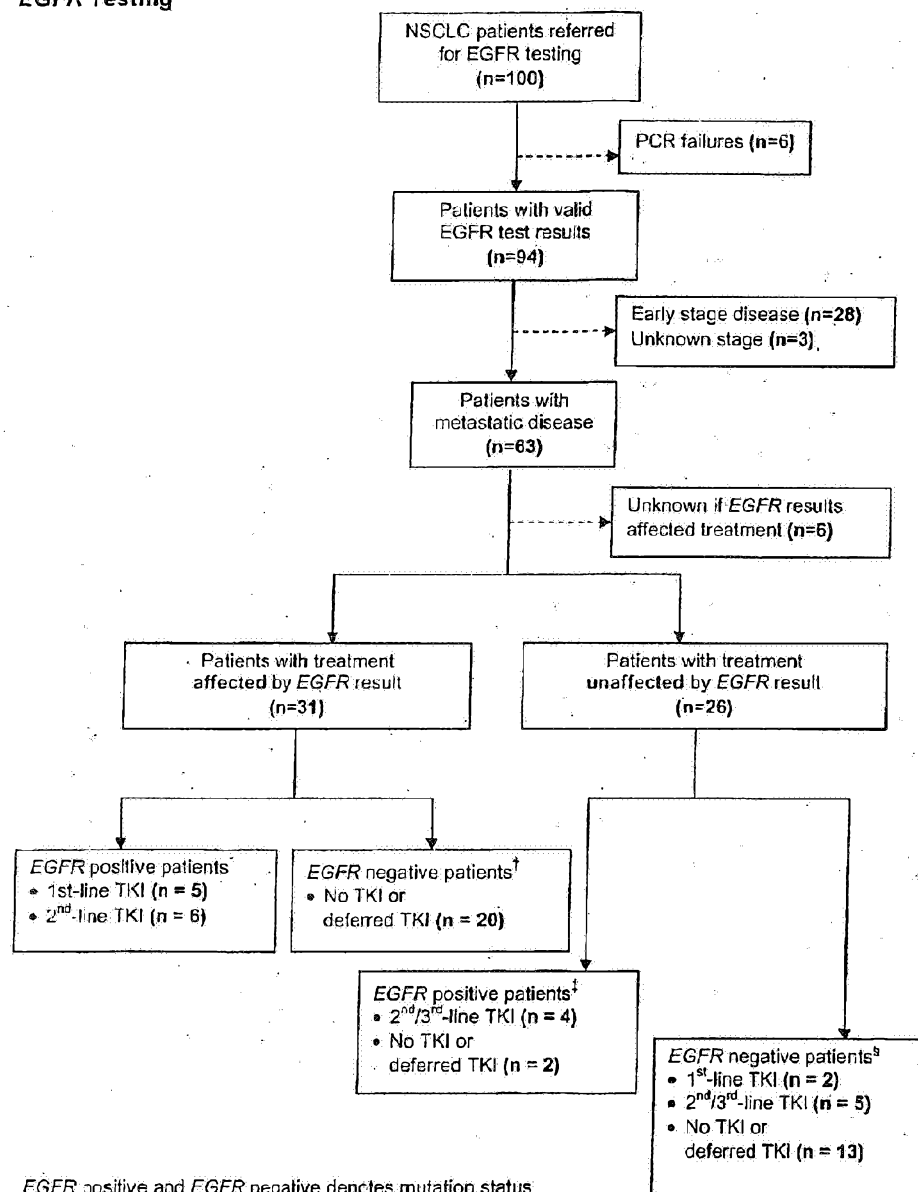
Alignment of Abl1 to EGFR

```

712                               790
Abl1 (242) ITMKHKLGGQYCEVYEGVWK-----KSLTVAVKTLKEDT--MEVEEFLKEAAVNMKEIKHPNLVQLLGVCTREPPFFYIIT
EGFR (712) FKIKIVLGSAGFTVYKGLWPEGEKVKIPVAIKELREATSPKANKEILDEAYVNASVDNPHVCRLLGICLTS-TVQLIIT
Consensus   LG G PG VY GLW   K I VAIK LKE T   E L EA VM I P L LLGIC   IIT
791
Abl1 (316) EFMTYGNLLDYLRRCNRQEVNAVVLVLYMATQISSAMEYLEKKNFIFHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTA
EGFR (791) QLMPPFGCLLDYVREHKDN-IGSQYLLNWCVCQIAKGMNYLEDRLVHRDLAARNVLVKTPOHVKITDFGLAKLLGAEEKEY
Consensus   M FG LLDYLR   N I A LL   QIA AM YLE K IHRDLAARN LV   VKI DFGLAKLL AD
870
Abl1 (396) HA-GAKFPIKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDL SQVVELLEKDYRMERPEGCEKVVYELMRA
EGFR (870) HAEGGKVPKWMVALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGPASEISILEKGERLPQDPICITIDVYMTIMVK
Consensus   HA GAK PIKW A ESI H FS SDVWAFGV LWEL TFG PY GI S I ILEK RL P C VY IM
950
Abl1 (475) CWQNPSPDRPSEAIHQAF----- (SEQ ID NO 491)
EGFR (950) CWMIADSRPKFRELIIEFSKWARDPQRYL (SEQ ID NO 492)
Consensus   CW   RP F EI   F
979

```

Figure 1C The Decision Making Process for Patients with Metastatic NSCLC Undergoing EGFR Testing



EGFR positive and EGFR negative denotes mutation status

* 2 patients enrolled in a 1st line TKI clinical trial

† 2 patients receiving TKI prior to testing

‡ 3 patients receiving TKI prior to testing

§ 1 patient enrolled in 1st line TKI clinical trial and 4 patients receiving TKI prior to testing

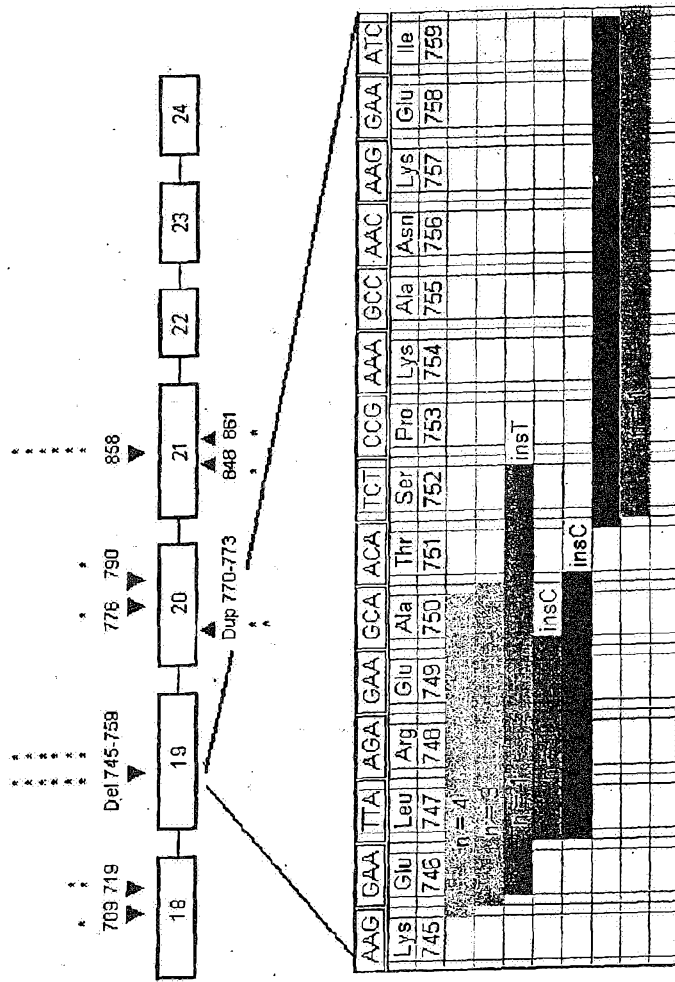


Figure 11 Diagram of EGFR exons 18-24, which encode the TK region (not to scale). Arrows depict the location of identified mutations. Asterisks denote the number of patients with mutations at each location. Blow-up diagram depicts the overlap of the exon 19 deletions, and the number of patients (n) with each deletion.