

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2007310412 B2**

(54) Title
Cancer susceptibility variants on Chr8q24.21

(51) International Patent Classification(s)
C12Q 1/68 (2006.01)

(21) Application No: **2007310412**

(22) Date of Filing: **2007.10.26**

(87) WIPO No: **WO08/050356**

(30) Priority Data

(31) Number
8560

(32) Date
2006.10.27

(33) Country
IS

(43) Publication Date: **2008.05.02**

(44) Accepted Journal Date: **2013.02.14**

(71) Applicant(s)
deCODE Genetics ehf.

(72) Inventor(s)
Amundadottir, Laufey;Kong, Augustine;Gudmundsson, Julius;Manolescu, Andrei;Sulem, Patrick

(74) Agent / Attorney
Freehills Patent Attorneys, Level 43 101 Collins Street, Melbourne, VIC, 3000

(56) Related Art
Freedman et al (2006) PNAS USA vol. 103, pps 14068-14073
Amundadottir et al (2006) Nature Genetics, vol. 38, pps 652-658

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
2 May 2008 (02.05.2008)

PCT

(10) International Publication Number
WO 2008/050356 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/IS2007/000019

(22) International Filing Date: 26 October 2007 (26.10.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
8560 27 October 2006 (27.10.2006) IS

(71) Applicant (for all designated States except US): **DECODE GENETICS** [IS/IS]; Sturlugata 8, IS-101 Reykjavik (IS).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GUDMUNDSSON, Julius** [IS/IS]; Kvistaland 17, IS-108 Reykjavik (IS). **SULEM, Patrick** [FR/IS]; Eskihlid 22, IS-105 Reykjavik (IS). **KONG, Augustine** [—/IS]; Maragata 2, IS-101 Reykjavik (IS). **MANOLESCU, Andrei** [RO/IS]; Eskihlid 22a, IS-105 Reykjavik (IS). **AMUNDADOTTIR, Laufey** [IS/US]; 461 Phelps Street, Gaithersburg, MD 20878 (US).

(74) Agent: **JONSSON, Thorlakur**; deCODE Genetics ehf., Sturlugata 8, IS-101 Reykjavik (IS).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— with international search report

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

— with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

(54) Title: **CANCER SUSCEPTIBILITY VARIANTS ON CHR8Q24.21**

(57) **Abstract:** A region on chromosome 8q24.21 has been demonstrated to play a major role in particular forms of cancer. It has been discovered that certain markers and haplotypes are indicative of a susceptibility to particular cancers, including prostate cancer. Diagnostic applications for identifying a susceptibility to cancer using these markers and haplotypes are described.



WO 2008/050356 A1

CANCER SUSCEPTIBILITY VARIANTS ON CHR8Q24.21

BACKGROUND OF THE INVENTION

Cancer, the uncontrolled growth of malignant cells, is a major health problem of the modern
5 medical era and is one of the leading causes of death in developed countries. In the United States, one
in four deaths is caused by cancer (Jemal, A. *et al.*, *CA Cancer J. Clin.* 52:23-47 (2002)).

The incidence of prostate cancer has dramatically increased over the last decades and prostate
cancer is now a leading cause of death in the United States and Western Europe (Peschel, R.E. and
J.W. Colberg, *Lancet* 4:233-41 (2003); Nelson, W.G. *et al.*, *N. Engl. J. Med.* 349(4):366-81 (2003)).
10 Prostate cancer is the most frequently diagnosed noncutaneous malignancy among men in
industrialized countries, and in the United States, 1 in 8 men will develop prostate cancer during his life
(Simard, J. *et al.*, *Endocrinology* 143(6):2029-40 (2002)). Although environmental factors, such as
dietary factors and lifestyle-related factors, contribute to the risk of prostate cancer, genetic factors have
also been shown to play an important role. Indeed, a positive family history is among the strongest
15 epidemiological risk factors for prostate cancer, and twin studies comparing the concordant occurrence
of prostate cancer in monozygotic twins have consistently revealed a stronger hereditary component in
the risk of prostate cancer than in any other type of cancer (Nelson, W.G. *et al.*, *N. Engl. J. Med.*
349(4):366-81 (2003); Lichtenstein P. *et al.*, *N. Engl. J. Med.* 343(2):78-85 (2000)). In addition, an
increased risk of prostate cancer is seen in 1st to 5th degree relatives of prostate cancer cases in a
20 nation wide study on the familiarity of all cancer cases diagnosed in Iceland from 1955-2003
(Amundadottir *et al.*, *PLoS Medicine* 1(3):e65 (2004)). The genetic basis for this disease, emphasized
by the increased risk among relatives, is further supported by studies of prostate cancer among
particular populations: for example, African Americans have among the highest incidence of prostate
cancer and mortality rate attributable to this disease: they are 1.6 times as likely to develop prostate
25 cancer and 2.4 times as likely to die from this disease than European Americans (Ries, L.A.G. *et al.*,
NIH Pub. No. 99-4649 (1999)).

An average 40% reduction in life expectancy affects males with prostate cancer. If detected
early, prior to metastasis and local spread beyond the capsule, prostate cancer can be cured (e.g.,
using surgery). However, if diagnosed after spread and metastasis from the prostate, prostate cancer is
30 typically a fatal disease with low cure rates. While prostate-specific antigen (PSA)-based screening has
aided early diagnosis of prostate cancer, it is neither highly sensitive nor specific (Punglia *et al.*, *N Engl
J Med.* 349(4):335-42 (2003)). This means that a high percentage of false negative and false positive
diagnoses are associated with the test. The consequences are both many instances of missed cancers
and unnecessary follow-up biopsies for those without cancer. As many as 65 to 85% of individuals
35 (depending on age) with prostate cancer have a PSA value less than or equal to 4.0 ng/mL, which has
traditionally been used as the upper limit for a normal PSA level (Punglia *et al.*, *N Engl J Med.*
349(4):335-42 (2003); Cookston, M.S., *Cancer Control* 8(2):133-40 (2001); Thompson, I.M. *et al.*, *N*

Engl J Med. 350:2239-46 (2004)). A significant fraction of those cancers with low PSA levels are scored as Gleason grade 7 or higher, which is a measure of an aggressive prostate cancer. *Id.*

In addition to the sensitivity problem outlined above, PSA testing also has difficulty with specificity and predicting prognosis. PSA levels can be abnormal in those without prostate cancer. For example, benign prostatic hyperplasia (BPH) is one common cause of a false-positive PSA test. In addition, a variety of noncancer conditions may elevate serum PSA levels, including urinary retention, prostatitis, vigorous prostate massage and ejaculation. *Id.*

Subsequent confirmation of prostate cancer using needle biopsy in patients with positive PSA levels is difficult if the tumor is too small to see by ultrasound. Multiple random samples are typically taken but diagnosis of prostate cancer may be missed because of the sampling of only small amounts of tissue. Digital rectal examination (DRE) also misses many cancers because only the posterior lobe of the prostate is examined. As early cancers are nonpalpable, cancers detected by DRE may already have spread outside the prostate (Mistry K.J., *Am. Board Fam. Pract.* 16(2):95-101 (2003)).

Thus, there is clearly a great need for improved diagnostic procedures that would facilitate early-stage prostate cancer detection and prognosis, as well as aid in preventive and curative treatments of the disease. In addition, there is a need to develop tools to better identify those patients who are more likely to have aggressive forms of prostate cancer from those patients that are more likely to have more benign forms of prostate cancer that remain localized within the prostate and do not contribute significantly to morbidity or mortality. This would help to avoid invasive and costly procedures for patients not at significant risk.

Loci Associated with Various Forms of Prostate Cancer

The incidence of prostate cancer has dramatically increased over the last decades. Prostate cancer is a multifactorial disease with genetic and environmental components involved in its etiology. It is characterized by heterogeneous growth patterns that range from slow growing tumors to very rapid highly metastatic lesions.

Although genetic factors are among the strongest epidemiological risk factors for prostate cancer, the search for genetic determinants involved in the disease has been challenging. Studies have revealed that linking candidate genetic markers to prostate cancer has been more difficult than identifying susceptibility genes for other cancers, such as breast, ovary and colon cancer. Several reasons have been proposed for this increased difficulty including: the fact that prostate cancer is often diagnosed at a late age thereby often making it difficult to obtain DNA samples from living affected individuals for more than one generation; the presence within high-risk pedigrees of phenocopies that are associated with a lack of distinguishing features between hereditary and sporadic forms; and the genetic heterogeneity of prostate cancer and the accompanying difficulty of developing appropriate statistical transmission models for this complex disease (Simard, J. *et al.*, *Endocrinology* 143(6):2029-40 (2002)).

Various genome scans for prostate cancer-susceptibility genes have been conducted and several prostate cancer susceptibility loci have been reported. For example, HPC1 (1q24-q25), PCAP (1q42-q43), HCPX (Xq27-q28), CAPB (1p36), HPC20 (20q13), HPC2/ELAC2 (17p11) and 16q23 have been proposed as prostate cancer susceptibility loci (Simard, J. *et al.*, *Endocrinology* 143(6):2029-40 (2002); Nwosu, V. *et al.*, *Hum. Mol. Genet.* 10(20):2313-18 (2001)). In a genome scan conducted by Smith *et al.*, the strongest evidence for linkage was at HPC1, although two-point analysis also revealed a LOD score of ≥ 1.5 at D4S430 and LOD scores ≥ 1.0 at several loci, including markers at Xq27-28 (Ostrander E.A. and J.L. Stanford, *Am. J. Hum. Genet.* 67:1367-75 (2000)). Another genome scan reported two-point LOD scores of ≥ 1.5 for chromosomes 10q, 12q and 14q using an autosomal dominant model of inheritance, and chromosomes 1q, 8q, 10q and 16p using a recessive model of inheritance. *Id.* Still another genome scan identified regions with nominal evidence for linkage on 2q, 12p, 15q, 16q and 16p. *Id.* A genome scan for prostate cancer predisposition loci using a small set of Utah high risk prostate cancer pedigrees and a set of 300 polymorphic markers provided evidence for linkage to a locus on chromosome 17p (Simard, J. *et al.*, *Endocrinology* 143(6):2029-40 (2002)). Eight new linkage analyses were published in late 2003, which depicted remarkable heterogeneity. Eleven peaks with LOD scores higher than 2.0 were reported, none of which overlapped (see Actane consortium, Schleutker *et al.*, Wiklund *et al.*, Witte *et al.*, Janer *et al.*, Xu *et al.*, Lange *et al.*, Cunningham *et al.*; all of which appear in *Prostate*, vol. 57 (2003)).

As described above, identification of particular genes involved in prostate cancer has been challenging. One gene that has been implicated is RNASEL, which encodes a widely expressed latent endoribonuclease that participates in an interferon-inducible RNA-decay pathway believed to degrade viral and cellular RNA, and has been linked to the HPC locus (Carpten, J. *et al.*, *Nat. Genet.* 30:181-84 (2002); Casey, G. *et al.*, *Nat. Genet.* 32(4):581-83 (2002)). Mutations in RNASEL have been associated with increased susceptibility to prostate cancer. For example, in one family, four brothers with prostate cancer carried a disabling mutation in RNASEL, while in another family, four of six brothers with prostate cancer carried a base substitution affecting the initiator methionine codon of RNASEL. *Id.* Other studies have revealed mutant RNASEL alleles associated with an increased risk of prostate cancer in Finnish men with familial prostate cancer and an Ashkenazi Jewish population (Rokman, A. *et al.*, *Am J. Hum. Genet.* 70:1299-1304 (2002); Rennert, H. *et al.*, *Am J. Hum. Genet.* 71:981-84 (2002)). In addition, the Ser217Leu genotype has been proposed to account for approximately 9% of all sporadic cases in Caucasian Americans younger than 65 years (Stanford, J.L., *Cancer Epidemiol. Biomarkers Prev.* 12(9):876-81 (2003)). In contrast to these positive reports, however, some studies have failed to detect any association between RNASEL alleles with inactivating mutations and prostate cancer (Wang, L. *et al.*, *Am. J. Hum. Genet.* 71:116-23 (2002); Wiklund, F. *et al.*, *Clin. Cancer Res.* 10(21):7150-56 (2004); Maier, C. *et al.*, *Br. J. Cancer* 92(6):1159-64(2005)).

The macrophage-scavenger receptor 1 (MSR1) gene, which is located at 8p22, has also been identified as a candidate prostate cancer-susceptibility gene (Xu, J. *et al.*, *Nat. Genet.* 32:321-25 (2002)). A mutant MSR1 allele was detected in approximately 3% of men with nonhereditary prostate cancer but only 0.4% of unaffected men. *Id.* However, not all subsequent reports have confirmed these initial findings (see, e.g., Lindmark, F. *et al.*, *Prostate* 59(2):132-40 (2004); Seppala, E.H. *et al.*, *Clin. Cancer Res.* 9(14):5252-56 (2003); Wang, L. *et al.*, *Nat Genet.* 35(2):128-29 (2003); Miller, D.C. *et al.*,

Cancer Res. 63(13):3486-89 (2003)). MSR1 encodes subunits of a macrophage-scavenger receptor that is capable of binding a variety of ligands, including bacterial lipopolysaccharide and lipoteichoic acid, and oxidized high-density lipoprotein and low-density lipoprotein in serum (Nelson, W.G. *et al.*, *N. Engl. J. Med.* 349(4):366-81 (2003)).

5 The *ELAC2* gene on Chr17 was the first prostate cancer susceptibility gene to be cloned in high risk prostate cancer families from Utah (Tavtigian, S.V., *et al.*, *Nat. Genet.* 27(2):172-80 (2001)). A frameshift mutation (1641InsG) was found in one pedigree. Three additional missense changes: Ser217Leu; Ala541Thr; and Arg781His, were also found to associate with an increased risk of prostate cancer. The relative risk of prostate cancer in men carrying both Ser217Leu and Ala541Thr was found
10 to be 2.37 in a cohort not selected on the basis of family history of prostate cancer (Rebbeck, T.R., *et al.*, *Am. J. Hum. Genet.* 67(4):1014-19 (2000)). Another study described a new termination mutation (Glu216X) in one high incidence prostate cancer family (Wang, L., *et al.*, *Cancer Res.* 61(17):6494-99 (2001)). Other reports have not demonstrated strong association with the three missense mutations, and a recent metaanalysis suggests that the familial risk associated with these mutations is more
15 moderate than was indicated in initial reports (Vesprini, D., *et al.*, *Am. J. Hum. Genet.* 68(4):912-17 (2001); Shea, P.R., *et al.*, *Hum. Genet.* 111(4-5):398-400 (2002); Suarez, B.K., *et al.*, *Cancer Res.* 61(13):4982-84 (2001); Severi, G., *et al.*, *J. Natl. Cancer Inst.* 95(11):818-24 (2003); Fujiwara, H., *et al.*, *J. Hum. Genet.* 47(12):641-48 (2002); Camp, N.J., *et al.*, *Am. J. Hum. Genet.* 71(6):1475-78 (2002)).

Polymorphic variants of genes involved in androgen action (e.g., the androgen receptor (AR)
20 gene, the cytochrome P-450c17 (CYP17) gene, and the steroid-5- α -reductase type II (SRD5A2) gene), have also been implicated in increased risk of prostate cancer (Nelson, W.G. *et al.*, *N. Engl. J. Med.* 349(4):366-81 (2003)). With respect to AR, which encodes the androgen receptor, several genetic epidemiological studies have shown a correlation between an increased risk of prostate cancer and the presence of short androgen-receptor polyglutamine repeats, while other studies have failed to detect
25 such a correlation. *Id.* Linkage data has also implicated an allelic form of CYP17, an enzyme that catalyzes key reactions in sex-steroid biosynthesis, with prostate cancer (Chang, B. *et al.*, *Int. J. Cancer* 95:354-59 (2001)). Allelic variants of SRD5A2, which encodes the predominant isozyme of 5- α -
reductase in the prostate and functions to convert testosterone to the more potent dihydrotestosterone, have been associated with an increased risk of prostate cancer and with a poor prognosis for men with
30 prostate cancer (Makridakis, N.M. *et al.*, *Lancet* 354:975-78 (1999); Nam, R.K. *et al.*, *Urology* 57:199-204 (2001)).

In short, despite the effort of many groups around the world, the genes that account for a substantial fraction of prostate cancer risk have not been identified. Although twin studies have implied that genetic factors are likely to be prominent in prostate cancer, only a handful of genes have been
35 identified as being associated with an increased risk for prostate cancer, and these genes account for only a low percentage of cases. Thus, it is clear that the majority of genetic risk factors for prostate cancer remain to be found. It is likely that these genetic risk factors will include a relatively high number of low-to-medium risk genetic variants. These low-to-medium risk genetic variants may, however, be responsible for a substantial fraction of prostate cancer, and their identification, therefore, a great benefit

for public health. Furthermore, none of the published prostate cancer genes have been reported to predict a greater risk for aggressive prostate cancer than for less aggressive prostate cancer.

Extensive genealogical information for a population containing cancer patients has in a recent study been combined with powerful gene sharing methods to map a locus on chromosome 8q24.21, which has been demonstrated to play a major role in cancer (e.g., breast cancer, prostate cancer, lung cancer, melanoma). Various cancer patients and their relatives were genotyped with a genome-wide marker set including 1100 microsatellite markers, with an average marker density of 3-4 cM. (Amundadottir L.T., *Nature Genet.* 38(6):652-658 (2006)). Association was detected to a single LD block within the locus between positions 128.414 and 128.506 Mb (NCBI build 34) in Utah CEPH HapMap samples.

Breast cancer is a significant health problem for women in the United States and throughout the world. Although advances have been made in detection and treatment of the disease, breast cancer remains the second leading cause of cancer-related deaths in women, affecting more than 180,000 women in the United States each year. For women in North America, the life-time odds of getting breast cancer are now one in eight.

No universally successful method for the treatment or prevention of breast cancer is currently available. Management of breast cancer currently relies on a combination of early diagnosis (e.g., through routine breast screening procedures) and aggressive treatment, which may include one or more of a variety of treatments, such as surgery, radiotherapy, chemotherapy and hormone therapy. The course of treatment for a particular breast cancer is often selected based on a variety of prognostic parameters including an analysis of specific tumor markers. See, e.g., Porter-Jordan and Lippman, *Breast Cancer* 8:73-100 (1994).

Although the discovery of BRCA1 and BRCA2 were important steps in identifying key genetic factors involved in breast cancer, it has become clear that mutations in BRCA1 and BRCA2 account for only a fraction of inherited susceptibility to breast cancer (Nathanson, K.L. *et al.*, *Human Mol. Gen.* 10(7):715-720 (2001); Anglican Breast Cancer Study Group. *Br. J. Cancer* 83(10):1301-08 (2000); and Syrjäkoski K. *et al.*, *J. Natl. Cancer Inst.* 92:1529-31 (2000)). In spite of considerable research into therapies for breast cancer, breast cancer remains difficult to diagnose and treat effectively, and the high mortality observed in breast cancer patients indicates that improvements are needed in the diagnosis, treatment and prevention of the disease.

deCODE has demonstrated an increased risk of breast cancer in 1st to 5th degree relatives of breast cancer cases in a nation wide study of the familiarity of all cancers diagnosed in Iceland from 1955-2003 (Amundadottir *et al.*, *PLoS Med.* 1(3):e65 (2004); Lichtenstein P. *et al.*, *N. Engl. J. Med.* 343(2):78-85 (2000)), where the authors show that breast cancer has one of the highest heritability of all cancers tested in a cohort of close to 45,000 twins.

It is estimated that only 5-10% of all breast cancers in women are associated with hereditary susceptibility due to mutations in autosomal dominant genes, such as BRCA1, BRCA2, p53, pTEN and STK11/LKB1 (Mincey, B.A. *Oncologist* 8:466-73 (2003)). One genetic locus, on Chromosome 8p, has

been proposed as a locus for a breast cancer-susceptibility gene based on studies documenting allelic loss in this region in sporadic breast cancer (Seitz, S. *et al.*, *Br. J. Cancer* 76:983-91 (1997); Kerangueven, F. *et al.*, *Oncogene* 10:1023 (1995)). Studies have also suggested that a breast cancer-susceptibility gene may be located on 13q21 (Kainu, T. *et al.*, *Proc. Natl. Acad. Sci. USA* 97:9603-08 (2000)). However, as with prostate cancer, identification of additional breast cancer-susceptibility genes has been difficult.

Lung cancer causes more deaths from cancer worldwide than any other form of cancer (Goodman, G.E., *Thorax* 57:994-999 (2002)). In the United States, lung cancer is the primary cause of cancer death among both men and women. In 2002, the death rate from lung cancer was an estimated 134,900 deaths, exceeding the combined total for breast, prostate and colon cancer. Lung cancer is also the leading cause of cancer death in all European countries and is rapidly increasing in developing countries. While environmental factors, such as lifestyle factors (e.g., smoking) and dietary factors, play an important role in lung cancer, genetic factors also contribute to the disease. For example, a family of enzymes responsible for carcinogen activation, degradation and subsequent DNA repair have been implicated in susceptibility to lung cancer. Studies have revealed that defects in both the p53 and RB/p16 pathway are essential for the malignant transformation of lung epithelial cells (Yokota, J. and T. Kohno, *Cancer Sci.* 95(3):197-204 (2004)). Other genes, such as K-ras, PTEN and MYO18B, are genetically altered less frequently than p53 and RB/p16 in lung cancer cells, suggesting that alterations in these genes are associated with further malignant progression or unique phenotypes in a subset of lung cancer cells. Molecular footprint studies that have been conducted at the sites of p53 mutations and RB/p16 deletions have further demonstrated that DNA repair activities and non-homologous end-joining of DNA double-strand breaks are important in the accumulation of genetic alterations in lung cancer cells. In addition, studies have identified candidate lung adenocarcinoma susceptibility genes, for example, drug carcinogen metabolism genes, such as NQO1 (NAD(P)H:quinone oxidoreductase) and GSTT1 (glutathione S-transferase T1), and DNA repair genes, such as XRCC1 (X-ray cross-complementary group 1) (Yanagitani, N. *et al.*, *Cancer Epidemiol. Biomarkers Prev.* 12:366-71 (2003); Lin, P. *et al.*, *J. Toxicol. Environ. Health A.* 58:187-97 (1999); Divine, K.K. *et al.*, *Mutat. Res.* 461:273-78 (2001); Sunaga, N. *et al.*, *Cancer Epidemiol. Biomarkers Prev.* 11:730-38 (2002)). A region of chromosome 19q13.3, which encompasses locus D19S246, has also been suggested as containing a gene(s) associated with lung adenocarcinoma (Yanagitani, N. *et al.*, *Cancer Epidemiol. Biomarkers Prev.* 12:366-71 (2003)). In addition, an increased risk to familial members outside of the nuclear family has been shown by deCODE geneticists by analysing all lung cancer cases diagnosed in Iceland over 48 years. This increased risk could not be entirely accounted for by smoking indicating that genetic variants may predispose certain individuals to lung cancer (Jonson *et al.*, *JAMA* 292(24):2977-83 (2004); Amundadottir *et al.*, *PLoS Med.* 1(3):e65 (2004)).

The five-year survival rate among all lung cancer patients, regardless of the stage of disease at diagnosis, is only 13%. This contrasts with a five-year survival rate of 46% among cases detected while the disease is still localized. However, only 16% of lung cancers are discovered before the disease has spread. Early detection is difficult as clinical symptoms are often not observed until the disease has reached an advanced stage. Currently, diagnosis is aided by the use of chest x-rays, analysis of the type of cells contained in sputum and fiberoptic examination of the bronchial passages. Treatment

regimens are determined by the type and stage of the cancer, and include surgery, radiation therapy and/or chemotherapy. In spite of considerable research into therapies for this and other cancers, lung cancer remains difficult to diagnose and treat effectively. Accordingly, there is a great need in the art for improved methods for detecting and treating such cancers.

5 The incidence of malignant melanoma is increasing more rapidly than any other type of human cancer in North America (Armstrong *et al.*, *Cancer Surv.* 19-20:219-240 (1994)). Although melanoma is curable when identified at an early stage, it requires detection and removal of the primary tumor before it has spread to distant sites. Malignant melanomas have great propensity to metastasize and are notoriously resistant to conventional cancer treatments, such as chemotherapy and -irradiation. Once
10 metastases have occurred the prognosis is very poor. Thus, early detection of melanoma is of vital importance in melanoma treatment and control.

 Studies have demonstrated that genetic factors play an important role in the stepwise progression of normal pigment cells to atypical nevi to invasive primary melanoma and finally to cells with aggressive metastatic potential (Kim, C.J., *et al.*, *Cancer Control* 9(1):49-53 (2002)). For example,
15 genetic aberrations, such as rearrangements on chromosome 1, which harbors a tumor-suppressor gene, have been implicated in malignant melanomas. However, the molecular and biological mechanisms of how a normal melanocyte of adult skin transforms into a melanoma cell remains unclear.

 Various studies have implicated genetic factors in melanoma. For example, elevated familial risk for early onset melanoma was noted by examination of a Utah population database (Cannon-
20 Albright, L.A., *et al.*, *Cancer Res.*, 54(9):2378-85 (1994)). In addition, the Swedish Family-Cancer Database reported a familial standardized incidence ratios (SIR) of 2.54 and 2.98 for cutaneous malignant melanoma (CMM) in a individual with an affected parent or sib, respectively. For an offspring whose parent had multiple primary melanomas, the SIR rose to 61.78 (Hemminki, K., *et al.*, *J. Invest. Dermatol.* 120(2):217-23 (2003)). In the Icelandic population-based study by Amundadottir *et al.* (*PLoS*
25 *Med.* 1(3):e65 (2004)), comparable SIR-values were found. Although figures vary, it has been reported that about 10% of CMM cases are familial (Hansen, C.B., *et al.*, *Lancet Oncol.* 5(5):314-19 (2004)). Given the known environmental risk factors for melanoma, shared environment in addition to genetics is likely to factor into these estimates. However, familial cases tend to have earlier ages of onset and a higher risk of multiple primary tumors, suggesting a genetic component (see, e.g., Tucker M., *Oncogene*
30 22(20):3042-52 (2003)). However, the molecular and biological mechanisms of how a normal melanocyte transforms into a melanoma cell remains unclear.

 A series of linkage-based studies have implicated CDKN2a on Chr9p21 as a major CMM-susceptibility gene (Bataille, V., *Eur. J. Cancer* 39(10):1341-47 (2003)). CDK4 was identified as a pathway candidate shortly thereafter, however, mutations in CDK4 have only been observed in a few
35 families worldwide (Zuo, L., *et al.*, *Nat. Genet.* 12(1):97-99 (1996)). CDKN2a encodes the cyclin dependent kinase inhibitor p16, which inhibits CDK4 and CDK6, thereby preventing G1 to S cell cycle transit. An alternate transcript of CKDN2a produces p14ARF, which encodes a cell cycle inhibitor that acts through the MDM2-p53 pathway. It is likely that CDKN2a mutant melanocytes are deficient in cell cycle control or the establishment of senescence, either as a developmental state or in response to DNA
40 damage (Ohtani, N., *et al.*, *J. Med. Invest.* 51(3-4):146-53 (2004)). Overall penetrance of CDKN2a

mutations in familial CMM cases is 67% by age 80. However, penetrance is increased in areas of high melanoma prevalence (Bishop, D.T., *et al.*, *J. Natl. Cancer Inst.* 94(12):894-903 (2002)).

The Melanoma Genetics Consortium recently completed a genome-wide scan for CMM, using a set of predominantly Australian, high-risk families unlinked to 9p21 or CDK4 (Gillanders, E., *et al.*, *Am. J. Hum. Genet.* 73(2):301-13 (2003)). The 10 cM resolution scan gave a non-parametric multipoint LOD score of 2.06 in the 1p22 region. Other locations on chromosomes 4, 7, 14, and 18 gave LODs in excess of 1.0. With additional markers to 1p22 and the application of an age-of-onset restriction, non-parametric LOD scores in excess of 5.0 were observed. Evidence suggests that a high-penetrance mutation of a tumor suppressor gene exists at this location, however the pattern of LOH is complex (Walker, G.J., *et al.*, *Genes Chromosomes Cancer*, 41(1):56-64 (2004)).

Another genetic locus that has been implicated in CMM is that which encodes the Melanocortin 1 Receptor (MC1R). MC1R is a G-protein coupled receptor that is involved in promoting the switch from pheomelanin to eumelanin synthesis. Numerous well-characterized variants of the MC1R gene have been implicated in red-haired, pale-skinned and freckle-prone phenotypes. More than half of red-haired individuals carry at least one of these MC1R variants (Valverde, P., *et al.*, *Nat. Genet.* 11(3):328-30 (1995); Palmer, J.S., *et al.*, *Am. J. Hum. Genet.* 66(1):176-86 (2000)). Subsequently, it was shown that the same variants conferred risk for CMM with odds ratios of about 2.0 for a single variant and about 4.0 for compound heterozygotes. Recent studies have shown that the stronger variants of MC1R increase the penetrance of CDKN2a mutations and lower the age of onset (Box, N.F., *et al.*, *Am. J. Hum. Genet.* 69(4):765-73 (2001); van der Velden, P.A., *et al.*, *Am. J. Hum. Genet.*, 69(4):774-79 (2001)).

A number of other candidate genes have been implicated in CMM. For example, a landmark study in cancer genomics identified somatic mutations in BRAF (the human B1 homolog of the v-raf murine sarcoma virus oncogene) in 60% of melanomas (Davies, H., *et al.*, *Nature* 417(6892):949-54 (2002)). Mutations are also common in nevi, both typical and atypical, suggesting that mutation is an early event. *Id.* Germline mutations have not been reported, however, a germline SNP variant of BRAF has been implicated in CMM risk (Meyer, P., *et al.*, *J. Carcinog.* 2(1):7 (2003)). Other candidate genes, which were identified through association studies and have been implicated in CMM risk include, e.g., XRCC3, XPD, EGF, VDR, NBS1, CYP2D6, and GSTM1 (Hayward, N.K., *Oncogene*, 22(20):3053-62 (2003)). However, such association studies frequently suffer from small sample sizes, reliance on single SNPs and potential population stratification.

Clearly, identification of markers and genes that are responsible for susceptibility to particular forms of cancer (e.g., prostate cancer, breast cancer, lung cancer, melanoma, colon cancer, testicular cancer) is one of the major challenges facing oncology today. Some of the pathways underlying cancer are shared in different forms of cancer. As a consequence, genetic risk factors identified for one particular form of cancer may also represent a risk factor for other cancer types. Diagnostic and therapeutic methods utilizing these risk factors may therefore have a common utility. Accordingly, therapeutic measures developed to target such risk factors may have implications for cancer in general, and not necessarily only the cancer for which the risk factor is originally identified. There is a need to identify means for the early detection of individuals that have a genetic susceptibility to cancer so that more aggressive screening and intervention regimens may be instituted for the early detection and treatment of cancer. Cancer genes may also reveal key molecular pathways that may be manipulated

(e.g., using small or large molecule weight drugs) and may lead to more effective treatments regardless of the cancer stage when a particular cancer is first diagnosed.

SUMMARY OF THE INVENTION

5 As described herein, it has been discovered that particular markers haplotypes in a specific DNA segment within chromosome 8q24.21 are indicative of susceptibility to particular cancers.

In a first aspect, the invention relates to a method for diagnosing a susceptibility to cancer in a human individual, comprising determining the presence or absence of at least one allele of at least one polymorphic marker in a nucleic acid sample obtained from the individual, wherein the at least one
10 polymorphic marker is associated with SEQ ID NO:2, and wherein the presence of the at least one allele is indicative of a susceptibility to the cancer. In one embodiment, the at least one marker is associated with SEQ ID NO:1. In another embodiment, the at least one marker is located within a genomic region whose nucleotide sequence is set forth in SEQ ID NO:2. In an alternative embodiment, the at least one marker is located within a genomic region whose nucleotide sequence is set forth in SEQ ID NO:1. In
15 one preferred embodiment, the at least one polymorphic marker comprises at least one marker selected from the group of markers set forth in Table 5A, 5B and 5C.

Due to the nature of linkage disequilibrium, the present invention may be practiced using a variety of polymorphic markers that are in linkage disequilibrium. Thus, in another embodiment, the at least one marker comprises at least one marker within Chr8q24.21 in strong linkage disequilibrium, as
20 defined by $|D'| > 0.8$ and/or $r^2 > 0.2$, with one or more markers selected from the group consisting of the markers in Table 4A and 4B. In another embodiment, the at least one polymorphic marker is in linkage disequilibrium with HapC. In one preferred embodiment, the at least one marker is marker rs16901979 (SEQ ID NO: 73) and markers in linkage disequilibrium therewith. In another preferred embodiment, the at least one marker is selected from the markers set forth in Table 4A and 4B.

25 In some embodiments of the invention, the method of diagnosing a susceptibility to cancer further comprises assessing the frequency of at least one haplotype in the individual. The haplotype comprises in one such embodiment the markers rs1456314 allele G, rs17831626 allele T, rs7825414 allele G, rs6993569 allele G, rs6994316 allele A, rs6470494 allele T, rs1016342 allele C, rs1031588 allele G, rs1016343 allele T, rs1551510 allele G, rs1456306 allele C, rs1378897 allele G, rs1456305
30 allele T, rs7816535 allele G

In certain embodiments of the invention, susceptibility is represented by values for relative risk (RR). In other embodiments, the susceptibility is represented by an odds ratio (OR). In certain
embodiments of the method diagnosing a susceptibility to cancer, the susceptibility is increased susceptibility, characterized by values for RR or OR of greater than one. In other embodiments, the
35 susceptibility is decreased susceptibility, characterized by values for RR or OR of less than one. In particular embodiments of the invention, the increased susceptibility is characterized by a relative risk of at least 1.5, including a relative risk of at least 1.7, a relative risk of at least 2.0, a relative risk of at least 2.5, a

relative risk of at least 3.0, a relative risk of at least 3.5, and a relative risk of at least 4.0. Other embodiments are characterized by relative risk of at least 1.75, 2.25, 2.75, 3.25, 3.75, and so on. Other values for the relative risk are however also within the scope of the present invention.

5 In certain other embodiments of the invention, certain alleles or haplotypes are found in decreased frequency in patients than in the population. Thus, certain alleles or haplotypes are found in decreased frequency in individuals diagnosed with, or at risk for, particular cancer (e.g., prostate cancer) than in the general population. Such markers are indicative of a protection against the cancer, or a decreased susceptibility of developing these disorders. Decreased susceptibility is in particular embodiments characterized by a relative risk of less than 0.7, including a relative risk of less than 0.6, a relative risk of less than 0.5, a relative risk of less than 0.4, a relative risk of less than 0.35, a relative risk of less than 0.3, and a relative risk of less than 0.25. Other values of relative risk characterizing the decreased susceptibility or decreased risk are however also possible and within scope of the invention, including, but not limited to, less than 0.8, less than 0.75, less than 0.65, less than 0.55, less than 0.45, less than 0.20, etc.

15 In particular embodiments of the methods of the invention, the at least one marker or haplotype comprises rs16901979 allele 1, with the at least one marker or haplotype conferring an increased susceptibility of the cancer. In another such embodiment, the at least one marker or haplotype is the marker rs16901979 allele 1. In other particular embodiments of the methods of the invention, the at least one marker or haplotype comprises rs16901979 allele 2, with the at least one marker or haplotype conferring an increased susceptibility of the cancer. In another such embodiment, the at least one marker or haplotype is the marker rs16901979 allele 2.

In certain embodiments of the methods of the invention, the cancer is selected from the group consisting of prostate cancer, colon cancer, breast cancer, testicular cancer, lung cancer and melanoma cancer. In a preferred embodiment, the cancer is prostate cancer.

25 The prostate cancer is in one embodiment an aggressive prostate cancer as defined by a combined Gleason score of 7(4+3)-10. The prostate cancer is in another embodiment a less aggressive prostate cancer as defined by a combined Gleason score of 2-7(3+4). In one embodiment, the at least one marker or haplotype is indicative of a more aggressive prostate cancer and/or a worse prognosis.

30 Another embodiment of the methods of the invention relates to the presence of the marker or haplotype being indicative of a different response rate of the subject to a particular treatment modality. In another embodiment, the presence of the at least one marker or haplotype is indicative of a predisposition to a somatic rearrangement of Chr8q24.21 in a tumor or its precursor. In one such embodiment, the somatic rearrangement is selected from the group consisting of an amplification, a translocation, an insertion and a deletion.

35 The methods, uses and kits of invention can in certain embodiments relate to individuals with a particular ancestry. Thus, in one embodiment of the invention, the individual is of a specific ancestry. In another embodiment, the ancestry is black African ancestry. Other ancestry of the individuals to be assessed by the methods of the invention are also possible, as described in further detail herein, and are also within scope of the present invention. In one embodiment, the ancestry is self-reported. In another embodiment, the ancestry is determined by detecting at least one allele of at least one

polymorphic marker in a sample from the individual, wherein the presence or absence of the allele is indicative of the ancestry of the individual.

Another aspect of the present invention relates to a method of identification of a marker for use in assessing susceptibility to cancer, the method comprising

- 5 a. identifying at least one polymorphic marker within SEQ ID NO:2, or at least one polymorphic marker in linkage disequilibrium therewith;
 - b. determining the genotype status of a sample of individuals diagnosed with, or having a susceptibility to, prostate cancer; and
 - c. determining the genotype status of a sample of control individuals;
- 10 wherein a significant difference in frequency of at least one allele in at least one polymorphism in individuals diagnosed with, or having a susceptibility to, prostate cancer, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one polymorphism being useful for assessing susceptibility to cancer.

Linkage disequilibrium is in one particular embodiment characterized by numerical values of r^2 of greater than 0.2 and/or $|D'|$ of greater than 0.8. In another embodiment, the at least one polymorphic marker is in linkage disequilibrium, as characterized by numerical values of r^2 of greater than 0.2 and/or $|D'|$ of greater than 0.8 with HapC and/or marker rs16901979. In another embodiment, an increase in frequency of the at least one allele in the at least one polymorphism in individuals diagnosed with, or having a susceptibility to, cancer, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one polymorphism being useful for assessing increased susceptibility to cancer. In yet another embodiment, a decrease in frequency of the at least one allele in the at least one polymorphism in individuals diagnosed with, or having a susceptibility to, exfoliation syndrome, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one polymorphism being useful for assessing decreased susceptibility to, or protection against, cancer.

25 The present invention, in another aspect, relates to a method of genotyping a nucleic acid sample obtained from a human individual at risk for, or diagnosed with, cancer, comprising determining the presence or absence of at least one allele of at least one polymorphic marker in the sample, wherein the at least one marker is selected from the group consisting of the markers set forth in Table 4A and 4B, and markers in linkage disequilibrium therewith, and wherein the presence or absence of the at least one allele of the at least one polymorphic marker is indicative of a susceptibility of cancer. In one embodiment the at least one marker is rs16901979 (SEQ ID NO:73), and markers in linkage disequilibrium therewith. In another embodiment, linkage disequilibrium is determined by numerical values for r^2 of at least 0.2 and/or numerical values of $|D'|$ of at least 0.8. In another embodiment, the genotyping comprises amplifying a segment of a nucleic acid that comprises the at least one polymorphic marker by Polymerase Chain Reaction (PCR), using a nucleotide primer pair flanking the at least one polymorphic marker. In yet another embodiment, genotyping is performed using a process selected from allele-specific probe hybridization, allele-specific primer extension, allele-specific

amplification, nucleic acid sequencing, 5'-exonuclease digestion, molecular beacon assay, oligonucleotide ligation assay, size analysis, and single-stranded conformation analysis. In one such embodiment, the process comprises allele-specific probe hybridization. In another embodiment, the process comprises nucleic acid sequencing. In another embodiment, the nucleic acid sequencing is

5 DNA sequencing.

One embodiment of a method of genotyping according to the invention, comprises the steps of:

- 1) contacting copies of the nucleic acid with a detection oligonucleotide probe and an enhancer oligonucleotide probe under conditions for specific hybridization of the oligonucleotide probe with the nucleic acid; wherein
 - 10 a) the detection oligonucleotide probe is from 5-100 nucleotides in length and specifically hybridizes to a first segment of the nucleic acid whose nucleotide sequence is given by SEQ ID NO:2 that comprises at least one polymorphic site;
 - b) the detection oligonucleotide probe comprises a detectable label at its 3' terminus and a quenching moiety at its 5' terminus;
 - 15 c) the enhancer oligonucleotide is from 5-100 nucleotides in length and is complementary to a second segment of the nucleotide sequence that is 5' relative to the oligonucleotide probe, such that the enhancer oligonucleotide is located 3' relative to the detection oligonucleotide probe when both oligonucleotides are hybridized to the nucleic acid; and
 - 20 d) a single base gap exists between the first segment and the second segment, such that when the oligonucleotide probe and the enhancer oligonucleotide probe are both hybridized to the nucleic acid, a single base gap exists between the oligonucleotides;
- 2) treating the nucleic acid with an endonuclease that will cleave the detectable label from the 3' terminus of the detection probe to release free detectable label when the detection probe
- 25 is hybridized to the nucleic acid; and
- 3) measuring free detectable label, wherein the presence of the free detectable label indicates that the detection probe specifically hybridizes to the first segment of the nucleic acid, and indicates the sequence of the polymorphic site as the complement of the detection probe.

In a particular embodiment, the copies of the nucleic acid are provided by amplification by

30 Polymerase Chain Reaction (PCR). In some embodiments the susceptibility to be detected is increased susceptibility. In other embodiment, the susceptibility is decreased susceptibility. In particular embodiments, the cancer is selected from prostate cancer, colon cancer, breast cancer, lung cancer, testicular cancer and melanoma. In a preferred embodiment, the cancer is prostate cancer. In one such embodiment, the prostate cancer is an aggressive prostate cancer as defined by a combined

35 Gleason score of 7(4+3)-10. In another such embodiment, the prostate cancer is a less aggressive prostate cancer as defined by a combined Gleason score of 2-7(3+4).

The methods of identification of markers for use in assessing a susceptibility to cancer, and the methods of genotyping, are in some embodiments practised on individual is of a specific ancestry. In one such embodiment, the ancestry is black African ancestry. Other ancestry is also within the scope of the invention, as described in detail in the herein. the ancestry is in one embodiment self-reported. In
5 another embodiment, the ancestry is determined by detecting at least one allele of at least one polymorphic marker in a sample from the individual, wherein the presence or absence of the allele is indicative of the ancestry of the individual.

Another aspect of the present invention relates to a method of assessing an individual for probability of response to a therapeutic agent for preventing and/or ameliorating symptoms associated
10 with cancer, comprising: determining the presence or absence of at least one allele of at least one polymorphic marker in a nucleic acid sample obtained from the individual, wherein the at least one polymorphic marker is selected from the group consisting of the polymorphic markers listed in Table 5A, 5B and 5C, and markers in linkage disequilibrium therewith, wherein the presence of the at least one allele of the at least one marker is indicative of a probability of a positive response to a symptoms
15 associated with exfoliation syndrome and/or glaucoma therapeutic agent. Another aspect of the present invention relates to a method of predicting prognosis of an individual diagnosed with, cancer, the method comprising determining the presence or absence of at least one allele of at least one polymorphic marker in a nucleic acid sample obtained from the individual, wherein the at least one polymorphic marker is selected from the group consisting of the polymorphic markers listed in Table 5A,
20 5B and 5C, and markers in linkage disequilibrium therewith, wherein the presence of the at least one allele is indicative of a worse prognosis of the cancer in the individual. Yet another aspect of the invention relates to a method of monitoring progress of a treatment of an individual undergoing treatment for cancer, the method comprising determining the presence or absence of at least one allele of at least one polymorphic marker in a nucleic acid sample obtained from the individual, wherein the at
25 least one polymorphic marker is selected from the group consisting of the polymorphic markers listed in Table 5A, 5B and 5C, and markers in linkage disequilibrium therewith, wherein the presence of the at least one allele is indicative of the treatment outcome of the individual. In any of these aspects, the at least one polymorphic marker is in one embodiment rs16901979 (SEQ ID NO:73) and markers in linkage disequilibrium therewith. In another embodiment, linkage disequilibrium is defined by numerical
30 values of r^2 of at least 0.2 and/or values of $|D'|$ of at least 0.8. In one preferred embodiment, the cancer is prostate cancer. In one such embodiment, the prostate cancer is an aggressive prostate cancer as defined by a combined Gleason score of 7(4+3)-10. In another embodiment, the prostate cancer is a less aggressive prostate cancer as defined by a combined Gleason score of 2-7(3+4).

The methods of the present invention may be utilized as such. In some embodiments, the
35 methods may also be utilized in combination with other useful in the methods described herein. In one such embodiment, the method further comprises assessing at least one biomarker in a sample from the individual. The biomarker may be any biological marker useful for aiding in any decision-making based on the methods as described herein. In one embodiment, the biomarker is PSA. In another embodiment, the sample is a blood sample or a cancer biopsy sample. Other sample types useful for
40 practicing the invention are however also contemplated, and within scope of the invention, such as other body fluids or tissue samples from any human tissue type.

Other embodiments of the methods of the invention, further comprise analyzing non-genetic information to make risk assessment, diagnosis, or prognosis of the individual. The non-genetic information are in one embodiment selected from age, gender, ethnicity, socioeconomic status, previous disease diagnosis, medical history of subject, family history of cancer, biochemical measurements, and
5 clinical measurements. In a preferred embodiment, the method further comprises calculating overall risk based on genetic and non-genetic information.

Another aspect of the present invention relates to a kit for assessing susceptibility to cancer in a human individual, the kit comprising reagents for selectively detecting at least one allele of at least one polymorphic marker in the genome of the individual, wherein the polymorphic marker is selected from
10 the group consisting of the polymorphic markers within the segment whose sequence is set forth in SEQ ID NO:2, and markers in linkage disequilibrium therewith, and wherein the presence of the at least one allele is indicative of a susceptibility to cancer. In one embodiment, the kit comprises at least one polymorphic marker selected from the group of markers set forth in Table 5A, 5B and 5C, and markers in linkage disequilibrium therewith. In another embodiment, the at least one polymorphic markers is
15 selected from the group of markers set forth in Table 4A and 4B. In one preferred embodiment, the at least one polymorphic marker is selected from rs16901979 (SEQ ID NO: 73), and markers in linkage disequilibrium therewith. In another embodiment, the at least one polymorphic marker is rs16901979. In one embodiment, linkage disequilibrium is defined by numerical values of r^2 of at least 0.2 and/or values of $|D'|$ of at least 0.8.. In another embodiment, the cancer is selected from prostate cancer, colon
20 cancer, breast cancer, testicular cancer, lung cancer and melanoma cancer. In a preferred embodiment, the cancer is prostate cancer. In one such embodiment, the prostate cancer is an aggressive prostate cancer as defined by a combined Gleason score of 7(4+3)-10. In another embodiment, the prostate cancer is a less aggressive prostate cancer as defined by a combined Gleason score of 2-7(3+4).

25 The kits of the invention may be used in any of the methods of the invention, as described herein. Thus, kits comprising reagents for specifically detecting at least one allele of at least one polymorphic marker, as described herein, may be utilized to practice any of the methods described herein, as will be apparent to the skilled person.

In one embodiment of the kit of the invention, the reagents comprise at least one contiguous
30 oligonucleotide that hybridizes to a fragment of the genome of the individual comprising the at least one polymorphic marker, a buffer and a detectable label. In one embodiment, the reagents comprise at least one pair of oligonucleotides that hybridize to opposite strands of a genomic nucleic acid segment obtained from the subject, wherein each oligonucleotide primer pair is designed to selectively amplify a fragment of the genome of the individual that includes one polymorphic marker, and wherein the
35 fragment is at least 30 base pairs in size. In a preferred embodiment, the at least one oligonucleotide is completely complementary to the genome of the individual. In another embodiment, the oligonucleotide is about 18 to about 50 nucleotides in length. In yet another embodiment, the oligonucleotide is 20-30 nucleotides in length.

In one preferred embodiment of the kit of the invention, the kit comprises:

- a. a detection oligonucleotide probe that is from 5-100 nucleotides in length;
- b. an enhancer oligonucleotide probe that is from 5-100 nucleotides in length; and
- c. an endonuclease enzyme;

5 wherein the detection oligonucleotide probe specifically hybridizes to a first segment of the nucleic acid whose nucleotide sequence is given by SEQ ID NO: 2 that comprises at least one polymorphic site; and

wherein the detection oligonucleotide probe comprises a detectable label at its 3' terminus and a quenching moiety at its 5' terminus;

wherein the enhancer oligonucleotide is from 5-100 nucleotides in length and is complementary to a
10 second segment of the nucleotide sequence that is 5' relative to the oligonucleotide probe, such that the enhancer oligonucleotide is located 3' relative to the detection oligonucleotide probe when both oligonucleotides are hybridized to the nucleic acid;

wherein a single base gap exists between the first segment and the second segment, such that when the oligonucleotide probe and the enhancer oligonucleotide probe are both hybridized to the nucleic
15 acid, a single base gap exists between the oligonucleotides; and

wherein treating the nucleic acid with the endonuclease will cleave the detectable label from the 3' terminus of the detection probe to release free detectable label when the detection probe is hybridized to the nucleic acid.

Another aspect of the invention relates to the use of an oligonucleotide probe in the
20 manufacture of a diagnostic reagent for diagnosing and/or assessing susceptibility to cancer in a human individual, wherein the probe hybridizes to a segment of a nucleic acid whose nucleotide sequence is given by SEQ ID NO: 2 that comprises at least one polymorphic site, wherein the fragment is 15-500 nucleotides in length. In one embodiment, the polymorphic site is selected from the polymorphic markers set forth in Table 5A, 5B and 5C, and polymorphisms in linkage disequilibrium therewith. In
25 another embodiment, the polymorphic site is rs16901979 (SEQ ID NO: 73). In one embodiment, the cancer is selected from prostate cancer, colon cancer, breast cancer, testicular cancer, lung cancer and melanoma cancer. In a preferred embodiment, the cancer is prostate cancer. In one such embodiment, the prostate cancer is an aggressive prostate cancer as defined by a combined Gleason score of 7(4+3)-10. In another embodiment, the prostate cancer is a less aggressive prostate cancer as defined
30 by a combined Gleason score of 2-7(3+4).

In yet another aspect, the present invention relates to a computer-readable medium on which is stored:

- a. an identifier for at least one polymorphic marker;

- b. an indicator of the frequency of at least one allele of said at least one polymorphic marker in a plurality of individuals diagnosed with cancer; and
- c. an indicator of the frequency of the least one allele of said at least one polymorphic markers in a plurality of reference individuals;

5 wherein the at least one polymorphic marker is selected from the polymorphic markers set forth in Table 5A, 5B and 5C, and polymorphisms in linkage disequilibrium therewith.

In one embodiment, the polymorphic site is marker rs16901979 (SEQ ID NO:73), and markers in linkage disequilibrium therewith, as defined by numerical values of r^2 of at least 0.2 and/or values of $|D'|$ of at least 0.8. In another embodiment, the cancer is selected from prostate cancer, colon cancer,
10 breast cancer, testicular cancer, lung cancer and melanoma cancer. In a preferred embodiment, the cancer is prostate cancer. In one such embodiment, the prostate cancer is an aggressive prostate cancer as defined by a combined Gleason score of 7(4+3)-10. In another embodiment, the prostate cancer is a less aggressive prostate cancer as defined by a combined Gleason score of 2-7(3+4).

The computer-readable medium according to the invention may, in some embodiments,
15 comprise information about the ancestry of the plurality of individuals. In another embodiment, the plurality of individuals diagnosed with cancer and the plurality of reference individuals is of a specific ancestry. In one embodiment, the ancestry is black African ancestry. In another embodiment, the ancestry is self-reported. In yet another embodiment, the ancestry is determined genetically, by genotyping a plurality of polymorphic markers to assess ancestry, as further described herein.

20 The invention furthermore relates to an apparatus for determining a genetic indicator for cancer in a human individual, comprising:

a computer readable memory; and

a routine stored on the computer readable memory;

wherein the routine is adapted to be executed on a processor to analyze marker and/or haplotype
25 information for at least one human individual with respect to at least one polymorphic marker selected from the markers set forth in Table 5A, 5B and 5C, and markers in linkage disequilibrium therewith, and generate an output based on the marker or haplotype information, wherein the output comprises a risk measure of the at least one marker or haplotype as a genetic indicator of cancer for the human individual. In one embodiment, the routine further comprises an indicator of the frequency of at least
30 one allele of at least one polymorphic marker or at least one haplotype in a plurality of individuals diagnosed with cancer, and an indicator of the frequency of at the least one allele of at least one polymorphic marker or at least one haplotype in a plurality of reference individuals, and wherein a risk measure is based on a comparison of the at least one marker and/or haplotype status for the human individual to the indicator of the frequency of the at least one marker and/or haplotype information for
35 the plurality of individuals diagnosed with cancer.

In one embodiment, the at least one polymorphic marker is rs16901979 (SEQ ID NO:73), and markers in linkage disequilibrium therewith, as defined by numerical values of r^2 of at least 0.2 and/or values of $|D'|$ of at least 0.8. In another embodiment, the risk measure is characterized by an Odds Ratio (OR) or a Relative Risk (RR).

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other aspects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawings.

FIG. 1 depicts the LD structure (HAPMAP) in the Chr8q24.21 LD Block C area. The LD structure for Caucasians (CEU) is shown in (A), while the LD structure for Africans from Yoruba (YRI) is shown in (B). The thick diagonal line indicates the location of LD block C (SEQ ID NO:1). Each marker is shown in a sequential order with equal distances between two consecutive markers.

FIG. 2 depicts the LD structure of the LD Block C' (SEQ ID NO:2) in the Chr8q24.21 region. The LD block as defined overlaps LD Block C, representing a refined analysis of the region within which variants associated with prostate cancer as described herein are located. The LD structure for Caucasians (CEU) is shown in (A), while the LD structure for Africans from Yoruba (YRI) is shown in (B). The thick diagonal line indicates the location of LD block C. Each marker is shown in a sequential order with equal distances between two consecutive markers.

DETAILED DESCRIPTION OF THE INVENTION

Reference to any prior art in the specification is not, and should not be taken as, an acknowledgment, or any form of suggestion, that this prior art forms part of the common general knowledge in Australia or any other jurisdiction or that this prior art could reasonably be expected to be ascertained, understood and regarded as relevant by a person skilled in the art.

Definitions

The following terms shall, in the present context, have the meaning as indicated:

As used herein, except where the context requires otherwise, the term "comprise" and variations of the term, such as "comprising", "comprises" and "comprised", are not intended to exclude other additives, components, integers or steps.

A "polymorphic marker", sometimes referred to as a "marker", as described herein, refers to a genomic polymorphic site. Each polymorphic marker has at least two sequence variations characteristic of particular alleles at the polymorphic site. Thus, genetic association to a polymorphic marker implies that there is association to at least one specific allele of that particular polymorphic marker. The marker
5 can comprise any allele of any variant type found in the genome, including single nucleotide polymorphisms (SNPs), microsatellites, insertions, deletions, duplications and translocations.

An "allele" refers to the nucleotide sequence of a given locus (position) on a chromosome. A polymorphic marker allele thus refers to the composition (i.e., sequence) of the marker on a chromosome. Genomic DNA from an individual contains two alleles for any given polymorphic marker,
10 representative of each copy of the marker on each chromosome.

2007310412 29 Jan 2013

A nucleotide position at which more than one sequence is possible in a population (either a natural population or a synthetic population, e.g., a library of synthetic molecules) is referred to herein as a "polymorphic site".

5 A "Single Nucleotide Polymorphism" or "SNP" is a DNA sequence variation occurring when a single nucleotide at a specific location in the genome differs between members of a species or between paired chromosomes in an individual. Most SNP polymorphisms have two alleles. Each individual is in this instance either homozygous for one allele of the polymorphism (i.e. both chromosomal copies of the individual have the same nucleotide at the SNP location), or the individual is heterozygous (i.e. the two sister chromosomes of the individual contain different nucleotides). The SNP nomenclature as reported
10 herein refers to the official Reference SNP (rs) ID identification tag as assigned to each unique SNP by the National Center for Biotechnological Information (NCBI).

A "variant", as described herein, refers to a segment of DNA that differs from the reference DNA. A "marker" or a "polymorphic marker", as defined herein, is a variant. Alleles that differ from the reference are referred to as "variant" alleles.

15 A "fragment" of a nucleotide or a protein, as described herein, comprises all or a part of the nucleotide or the protein.

An "animal", as described herein, refers to any domestic animal (e.g., cats, dogs, etc.), agricultural animal (e.g., cows, horses, sheep, chicken, etc.), or test species (e.g., rabbit, mouse, rat, etc.), and also includes humans.

20 A "microsatellite", as described herein, is a polymorphic marker that has multiple small repeats of bases that are 2-8 nucleotides in length (such as CA repeats) at a particular site, in which the number of repeat lengths varies in the general population.

An "indel", as described herein, is a common form of polymorphism comprising a small insertion or deletion that is typically only a few nucleotides long.

25 A "haplotype," as described herein, refers to a segment of genomic DNA within one strand of DNA that is characterized by a specific combination of alleles arranged along the segment. For diploid organisms such as humans, a haplotype comprises one member of the pair of alleles for each polymorphic marker or locus. In a certain embodiment, the haplotype can comprise two or more alleles, three or more alleles, four or more alleles, or five or more alleles.

30 The term "susceptibility", as described herein, encompasses both increased susceptibility and decreased susceptibility. Thus, particular polymorphic markers and/or haplotypes of the invention may be characteristic of increased susceptibility (i.e., increased risk) of glaucoma, as characterized by a relative risk (RR) of greater than one. Alternatively, the markers and/or haplotypes of the invention are characteristic of decreased susceptibility (i.e., decreased risk) of glaucoma, as characterized by a
35 relative risk of less than one.

A "nucleic acid sample", as described herein, is a sample obtained from an individual that contains nucleic acid (DNA or RNA). In certain embodiments, i.e. the detection of specific polymorphic markers and/or haplotypes, the nucleic acid sample comprises genomic DNA. Such a nucleic acid sample can be obtained from any source that contains genomic DNA, including as a blood sample, sample of amniotic fluid, sample of cerebrospinal fluid, or tissue sample from skin, muscle, buccal or conjunctival mucosa (buccal swab), placenta, gastrointestinal tract or other organs.

As used herein, "Chr8q24.21" and "8q24.21" refer to chromosomal band 8q24.21 which corresponds roughly to position 127,200,001-131,400,000 bp in UCSC Build 34 (from the UCSC Genome browser Build 34 at www.genome.ucsc.edu).

As used herein, "LD Block C" refers to the LD block on Chr8q24.21 wherein association of variants to cancer, i.e. prostate cancer, breast cancer, lung cancer and melanoma is observed. NCBI Build 34 position of this LD block is from 128,032,278 to 128,094,256 bp (SEQ ID NO:1).

As used herein, "LD Block C'" refers to the LD block on Chr8q24.21 wherein association of variants associated with cancer may preferably be detected. The NCBI Build 34 position of this LD block is from 128,029,113 to 128,126,447, and its sequence is set forth in SEQ ID NO:2

Bp. In NCBI Builds 35 and 36 the location of the region is from position 128,141,706 to 128,239,040. The sequence of the LD Block C' region in Builds 34, 35 and 36 is identical over the entire span of 97,335 bp.

The term "African ancestry", as described herein, refers to self-reported African ancestry of individuals.

The term "cancer therapeutic agent" refers to an agent that can be used to ameliorate or prevent symptoms associated with cancer (i.e., prostate cancer, lung cancer, breast cancer and/or melanoma).

The terms "associated with SEQ ID NO:2", "associated with SEQ ID NO:1" "associated with LD Block C", and "associated with LD Block C'", refer to those DNA segments (e.g. polymorphic markers) that are in linkage disequilibrium (LD) with the genomic segments represented by SEQ ID NO:2, SEQ ID NO:1, LD Block C and LD Block C'. In certain embodiments, such DNA segments are in LD with one or more markers within SEQ ID NO:2, SEQ ID NO:1, LD Block C or LD Block C' as measured by values for $|D'|$ greater than 0.8 and/or r^2 greater than 0.2.

Association to Chr8q24.21

As discussed above, linkage to chromosome 8q24.21 and association to a linkage disequilibrium (LD) block within the linkage region has recently been reported. As described herein, it has now been surprisingly found that variants (markers and/or haplotypes) residing within another DNA segment of extensive LD (i.e., another LD block) in the chromosome 8q24.21 region are also associated

with cancer. The association detected is independent of previously detected association in the region, a surprising result for the present inventors. In one embodiment of the invention, the association is detected by the haplotype HapC, comprising the markers rs1456314 allele G, rs17831626 allele T, rs7825414 allele G, rs6993569 allele G, rs6994316 allele A, rs6470494 allele T, rs1016342 allele C, rs1031588 allele G, rs1016343 allele T, rs1551510 allele G, rs1456306 allele C, rs1378897 allele G, rs1456305 allele T and rs7816535 allele G. As with other variants associated with human traits, a large number of surrogate variants (markers and/or haplotypes) can be described. One such surrogate marker for HapC is marker rs16901979. The region most likely to harbor surrogate variants is commonly defined as a region of extensive linkage disequilibrium, so called linkage disequilibrium blocks (LD block), as further described herein. In one embodiment, an LD block encompassing the variants associating with cancer is LD Block C characterized by the sequence set forth in SEQ ID NO:1. A further refinement of the signal originally detected by the inventors defines the LD Block C' as the region between two recombination hotspots on chromosome 8q24.21. The hotspots are located at approximately position 128,029,113 and 128,126,447 on chromosome 8, and the region thus defined is as set forth in SEQ ID NO:2. Surrogate markers and/or haplotypes for HapC, rs16901979 can be found within either LD Block as defined (i.e., SEQ ID NO:1, and SEQ ID NO:2) and further described in detail herein.

In various embodiments of the invention, certain markers and/or SNPs, identified using the methods described herein, can be used for a diagnosis of increased susceptibility to cancer (e.g., prostate cancer), and also for a diagnosis of a decreased susceptibility to cancer (e.g., prostate cancer), i.e. for identification of variants that are protective against cancer (e.g., prostate cancer). The diagnostic assays presented below can be used to identify the presence or absence of these particular variants.

The Gleason score is the most frequently used grading system for prostate cancer (DeMarzo, A.M. *et al.*, *Lancet* 361:955-64 (2003)). The system is based on the discovery that prognosis of prostate cancer is intermediate between that of the most predominant pattern of cancer and that of the second most predominate pattern. These predominant and second most prevalent patterns are identified in histological samples from prostate tumors and each is graded from 1 (most differentiated) to 5 (least differentiated) and the two scores are added. The combined Gleason grade, also known as the Gleason sum or score, thus ranges from 2 (for tumors uniformly of pattern 1) to 10 (for undifferentiated tumors). Most cases with divergent patterns, especially on needle biopsy, do not differ by more than one pattern.

The Gleason score is a prognostic indicator, with the major prognostic shift being between 6 and 7, as Gleason score 7 tumors behave much worse leading to more morbidity and higher mortality than tumors scoring 5 or 6. Score 7 tumors can further be subclassified into 3+4 or 4+3 (the first number is the predominant histologic subtype in the biopsied tumor sample and the second number is the next predominant histologic subtype), with the 4+3 score being associated with worse prognosis. A patient's Gleason score can also influence treatment options. For example, younger men with limited amounts of a Gleason score 5-6 on needle biopsy and low PSA concentrations may simply be monitored while men with Gleason scores of 7 or higher usually receive active management. In Table 1, the frequency of haplotype and the associated risk of aggressive prostate cancer (i.e., as indicated by a

combined Gleason score of 7(4+3 only) to 10) and less aggressive prostate cancer (i.e., as indicated by a combined Gleason score of 2 to 7 (3+4 only)) are indicated. However, the Gleason score is not a perfect predictor of prognosis. Thus, patients with tumors with low Gleason scores may still have more aggressive prostate cancer (defined as tumors extending beyond the prostate locally or through distant metastasis).

In certain methods described herein, an individual who is at risk (increased susceptibility) for cancer (e.g., prostate cancer (aggressive or high Gleason grade prostate cancer, less aggressive or low Gleason grade prostate cancer) is an individual in whom an at-risk marker or haplotype is identified. In one embodiment, the strength of the association of a marker or haplotype is measured by relative risk (RR). RR is the ratio of the incidence of the condition among subjects who carry one copy of the marker or haplotype to the incidence of the condition among subjects who do not carry the marker or haplotype. This ratio is equivalent to the ratio of the incidence of the condition among subjects who carry two copies of the marker or haplotype to the incidence of the condition among subjects who carry one copy of the marker or haplotype.

In one embodiment, the invention is a method of diagnosing a susceptibility to prostate cancer (e.g., aggressive or high Gleason grade prostate cancer, less aggressive or low Gleason grade prostate cancer), comprising detecting a marker or haplotype associated with LD Block C (e.g., a marker or haplotype as set forth in Table 5, having a value of relative risk (RR) greater than one, indicating the marker is associated with increased susceptibility to disease/increased risk of disease and thus is an "at-risk" variant; a marker or haplotype with values of RR less than one indicate the marker is associated with decreased susceptibility to disease/decreased risk of disease and thus is a "protective" variant), wherein the presence of the marker or haplotype is indicative of a susceptibility to prostate cancer.

In another embodiment, the invention is a method of diagnosing a susceptibility to prostate cancer (e.g., aggressive or high Gleason grade prostate cancer, less aggressive or low Gleason grade prostate cancer), comprising detecting marker rs16901979. In one embodiment, the susceptibility is increased susceptibility, wherein the presence of the 1 allele at marker rs16901979 is indicative of an increased susceptibility to prostate cancer. In a further embodiment, the invention is a method of diagnosing increased susceptibility to prostate cancer in an individual whose ancestry comprises African ancestry, comprising detecting marker rs16901979, wherein the presence of the 1 allele at marker rs16901979 is indicative of an increased susceptibility to prostate cancer or an increased risk of prostate cancer. In particular embodiments, the marker or haplotype that is associated with a susceptibility to prostate cancer has a relative risk of at least 1.3, such as at least 1.5 or at least 1.7 or at least 2.0. In another embodiment, the prostate cancer is an aggressive prostate cancer, as defined by a combined Gleason score of 7(4+3) to 10 and/or an advanced stage of prostate cancer (e.g., Stages 2 to 4). In yet another embodiment, the prostate cancer is a less aggressive prostate cancer, as defined by a combined Gleason score of 2 to 7(3+4) and/or an early stage of prostate cancer (e.g., Stage 1). In another embodiment, the presence of a marker or haplotype associated with LD Block C, in conjunction with the subject having a PSA level greater than 4 ng/ml, is indicative of a more aggressive prostate cancer and/or a worse prognosis. In yet another embodiment, in patients who have a normal PSA level

(e.g., less than 4 ng/ml), the presence of a marker or haplotype is indicative of a more aggressive prostate cancer and/or a worse prognosis.

In other embodiments, the invention is a method of diagnosing a decreased susceptibility to prostate cancer, comprising detecting a marker or haplotype associated with LD Block C, wherein the presence of that marker or haplotype is indicative of a decreased susceptibility to prostate cancer or of a protective marker or haplotype against prostate cancer. Thus, in one embodiment, the susceptibility is decreased susceptibility, wherein the presence of the 2 allele at marker rs16901979 is indicative of a decreased susceptibility to prostate cancer. In a further embodiment, the invention is a method of diagnosing decreased susceptibility to prostate cancer in an individual whose ancestry comprises African ancestry, comprising detecting marker rs16901979, wherein the presence of the s allele at marker rs16901979 is indicative of a decreased susceptibility to prostate cancer or a decreased risk of prostate cancer.

The segment on chromosome 8q24.21 of the present invention has also been found to play a role in other forms of cancer, e.g. breast cancer, colon cancer, lung cancer and melanoma. It has been discovered that particular markers and/or haplotypes in a specific DNA segment within the region are present at a higher than expected frequency in breast cancer subjects. Thus, in one embodiment, the invention is a method of diagnosing an increased susceptibility to cancer selected from breast cancer, lung cancer, colon cancer and melanoma, comprising detecting a marker or haplotype associated with the genomic segments whose sequence is set forth in SEQ ID NO:1 or SEQ ID NO:2, wherein the presence of the marker or haplotype is indicative of an increased susceptibility to the cancer (e.g., breast cancer, colon cancer, lung cancer and melanoma). In a particular embodiment, the marker or haplotype that is associated with a susceptibility to cancer (i.e., breast cancer, lung cancer and melanoma) has a relative risk of at least 1.3, such as at least 1.5, at least 1.7 or at least 2.0. In other embodiments, the invention is drawn to a method of diagnosing a decreased susceptibility to cancer (i.e., breast cancer, lung cancer and melanoma) comprising detecting a marker or haplotype associated with the genomic segments whose sequence is set forth in SEQ ID NO:1 or SEQ ID NO:2, wherein the presence of that marker or haplotype is indicative of a decreased susceptibility to cancer or of a protective marker or haplotype against breast cancer (protective against cancer (i.e., breast cancer, lung cancer and melanoma)). In a particular embodiment, the marker or haplotype that is associated with a decreased susceptibility to cancer (i.e., breast cancer, lung cancer and melanoma) has a relative risk of less than 0.9, such as less than 0.8, less than 0.7, less than 0.6 and less than 0.5. In another embodiment, the melanoma is malignant cutaneous melanoma.

Assessment for markers and haplotypes

The genomic sequence within populations is not identical when individuals are compared. Rather, the genome exhibits sequence variability between individuals at many locations in the genome. Such variations in sequence are commonly referred to as polymorphisms, and there are many such sites within each genome. For example, the human genome exhibits sequence variations which occur on average every 500 base pairs. The most common sequence variant consists of base variations at a

single base position in the genome, and such sequence variants, or polymorphisms, are commonly called Single Nucleotide Polymorphisms ("SNPs"). These SNPs are believed to have occurred in a single mutational event, and therefore there are usually two possible alleles possible at each SNPsite; the original allele and the mutated allele. Due to natural genetic drift and possibly also selective
5 pressure, the original mutation has resulted in a polymorphism characterized by a particular frequency of its alleles in any given population. Many other types of sequence variants are found in the human genome, including microsatellites, insertions, deletions, inversions and copy number variations. A polymorphic microsatellite has multiple small repeats of bases (such as CA repeats, TG on the complimentary strand) at a particular site in which the number of repeat lengths varies in the general
10 population. In general terms, each version of the sequence with respect to the polymorphic site represents a specific allele of the polymorphic site. These sequence variants can all be referred to as polymorphisms, occurring at specific polymorphic sites characteristic of the sequence variant in question. In general terms, polymorphisms can comprise any number of specific alleles. Thus in one embodiment of the invention, the polymorphism is characterized by the presence of two or more alleles
15 in any given population. In another embodiment, the polymorphism is characterized by the presence of three or more alleles. In other embodiments, the polymorphism is characterized by four or more alleles, five or more alleles, six or more alleles, seven or more alleles, nine or more alleles, or ten or more alleles. All such polymorphisms can be utilized in the methods and kits of the present invention, and are thus within the scope of the invention.

20 In some instances, reference is made to different alleles at a polymorphic site without choosing a reference allele. Alternatively, a reference sequence can be referred to for a particular polymorphic site. The reference allele is sometimes referred to as the "wild-type" allele and it usually is chosen as either the first sequenced allele or as the allele from a "non-affected" individual (e.g., an individual that does not display a trait or disease phenotype).

25 Alleles for SNP markers as referred to herein refer to the bases A, C, G or T as they occur at the polymorphic site in the SNP assay employed. The allele codes for SNPs used herein are as follows: 1=A, 2=C, 3=G, 4=T. The person skilled in the art will however realise that by assaying or reading the opposite DNA strand, the complementary allele can in each case be measured. Thus, for a polymorphic site (polymorphic marker) characterized by an A/G polymorphism, the assay employed may be designed
30 to specifically detect the presence of one or both of the two bases possible, i.e. A and G. Alternatively, by designing an assay that is designed to detect the opposite strand on the DNA template, the presence of the complementary bases T and C can be measured. Quantitatively (for example, in terms of relative risk), identical results would be obtained from measurement of either DNA strand (+ strand or – strand).

Typically, a reference sequence is referred to for a particular sequence. Alleles that differ from
35 the reference are sometimes referred to as "variant" alleles. A variant sequence, as used herein, refers to a sequence that differs from the reference sequence but is otherwise substantially similar. Alleles at the polymorphic genetic markers described herein are variants. Additional variants can include changes that affect a polypeptide. Sequence differences, when compared to a reference nucleotide sequence, can include the insertion or deletion of a single nucleotide, or of more than one nucleotide, resulting in a
40 frame shift; the change of at least one nucleotide, resulting in a change in the encoded amino acid; the

change of at least one nucleotide, resulting in the generation of a premature stop codon; the deletion of several nucleotides, resulting in a deletion of one or more amino acids encoded by the nucleotides; the insertion of one or several nucleotides, such as by unequal recombination or gene conversion, resulting in an interruption of the coding sequence of a reading frame; duplication of all or a part of a sequence; 5 transposition; or a rearrangement of a nucleotide sequence,. Such sequence changes can alter the polypeptide encoded by the nucleic acid. For example, if the change in the nucleic acid sequence causes a frame shift, the frame shift can result in a change in the encoded amino acids, and/or can result in the generation of a premature stop codon, causing generation of a truncated polypeptide. Alternatively, a polymorphism associated with a disease or trait can be a synonymous change in one or 10 more nucleotides (*i.e.*, a change that does not result in a change in the amino acid sequence). Such a polymorphism can, for example, alter splice sites, affect the stability or transport of mRNA, or otherwise affect the transcription or translation of an encoded polypeptide. It can also alter DNA to increase the possibility that structural changes, such as amplifications or deletions, occur at the somatic level. The polypeptide encoded by the reference nucleotide sequence is the "reference" polypeptide with a 15 particular reference amino acid sequence, and polypeptides encoded by variant alleles are referred to as "variant" polypeptides with variant amino acid sequences.

A haplotype refers to a segment of DNA that is characterized by a specific combination of alleles arranged along the segment. For diploid organisms such as humans, a haplotype comprises one member of the pair of alleles for each polymorphic marker or locus. In a certain embodiment, the 20 haplotype can comprise two or more alleles, three or more alleles, four or more alleles, or five or more alleles, each allele corresponding to a specific polymorphic marker along the segment. Haplotypes can comprise a combination of various polymorphic markers, *e.g.*, SNPs and microsatellites, having particular alleles at the polymorphic sites. The haplotypes thus comprise a combination of alleles at various genetic markers.

25 Detecting specific polymorphic markers and/or haplotypes can be accomplished by methods known in the art for detecting sequences at polymorphic sites. For example, standard techniques for genotyping for the presence of SNPs and/or microsatellite markers can be used, such as fluorescence-based techniques (Chen, X. *et al.*, *Genome Res.* 9(5): 492-98 (1999)), utilizing PCR, LCR, Nested PCR and other techniques for nucleic acid amplification. Specific methodologies available for SNP 30 genotyping include, but are not limited to, TaqMan genotyping assays and SNPlex platforms (Applied Biosystems), mass spectrometry (*e.g.*, MassARRAY system from Sequenom), minisequencing methods, real-time PCR, Bio-Plex system (BioRad), CEQ and SNPstream systems (Beckman), Molecular Inversion Probe array technology (*e.g.*, Affymetrix GeneChip), and BeadArray Technologies (*e.g.*, Illumina GoldenGate and Infinium assays). By these or other methods available to the person 35 skilled in the art, one or more alleles at polymorphic markers, including microsatellites, SNPs or other types of polymorphic markers, can be identified.

In certain methods described herein, an individual who is at an increased susceptibility (*i.e.*, increased risk) for any specific disease or trait under study, is an individual in whom at least one specific allele at one or more polymorphic marker or haplotype conferring increased susceptibility for the disease 40 or trait is identified (*i.e.*, at-risk marker alleles or haplotypes). In one aspect, the at-risk marker or

haplotype is one that confers a significant increased risk (or susceptibility) of the disease or trait. In one embodiment, significance associated with a marker or haplotype is measured by a relative risk (RR). In another embodiment, significance associated with a marker or haplotype is measured by an odds ratio (OR). In a further embodiment, the significance is measured by a percentage. In one embodiment, a significant increased risk is measured as a risk (relative risk and/or odds ratio) of at least 1.2, including but not limited to: at least 1.2, at least 1.3, at least 1.4, at least 1.5, at least 1.6, at least 1.7, 1.8, at least 1.9, at least 2.0, at least 2.5, at least 3.0, at least 4.0, and at least 5.0. In a particular embodiment, a risk (relative risk and/or odds ratio) of at least 1.2 is significant. In another particular embodiment, a risk of at least 1.3 is significant. In yet another embodiment, a risk of at least 1.4 is significant. In a further embodiment, a relative risk of at least about 1.5 is significant. In another further embodiment, a significant increase in risk is at least about 1.7 is significant. However, other cutoffs are also contemplated, e.g. at least 1.15, 1.25, 1.35, and so on, and such cutoffs are also within scope of the present invention. In other embodiments, a significant increase in risk is at least about 20%, including but not limited to about 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 150%, 200%, 300%, and 500%. In one particular embodiment, a significant increase in risk is at least 20%. In other embodiments, a significant increase in risk is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% and at least 100%. Other cutoffs or ranges as deemed suitable by the person skilled in the art to characterize the invention are however also contemplated, and those are also within scope of the present invention.

20 An at-risk polymorphic marker or haplotype of the present invention is one where at least one allele of at least one marker or haplotype is more frequently present in an individual at risk for the disease or trait (affected), compared to the frequency of its presence in a comparison group (control), and wherein the presence of the marker or haplotype is indicative of susceptibility to the disease or trait. The control group may in one embodiment be a population sample, i.e. a random sample from the general population. In another embodiment, the control group is represented by a group of individuals who are disease-free. Such disease-free control may in one embodiment be characterized by the absence of one or more specific disease-associated symptoms. In another embodiment, the disease-free control group is characterized by the absence of one or more disease-specific risk factors. Such risk factors are in one embodiment at least one environmental risk factor. Representative environmental factors are natural products, minerals or other chemicals which are known to affect, or contemplated to affect, the risk of developing the specific disease or trait. Other environmental risk factors are risk factors related to lifestyle, including but not limited to food and drink habits, geographical location of main habitat, and occupational risk factors. In another embodiment, the risk factors are at least one genetic risk factor.

35 As an example of a simple test for correlation would be a Fisher-exact test on a two by two table. Given a cohort of chromosomes, the two by two table is constructed out of the number of chromosomes that include both of the markers or haplotypes, one of the markers or haplotypes but not the other and neither of the markers or haplotypes.

In other embodiments of the invention, an individual who is at a decreased susceptibility (i.e., at a decreased risk) for a disease or trait is an individual in whom at least one specific allele at one or more

polymorphic marker or haplotype conferring decreased susceptibility for the disease or trait is identified. The marker alleles and/or haplotypes conferring decreased risk are also said to be protective. In one aspect, the protective marker or haplotype is one that confers a significant decreased risk (or susceptibility) of the disease or trait. In one embodiment, significant decreased risk is measured as a relative risk of less than 0.9, including but not limited to less than 0.9, less than 0.8, less than 0.7, less than 0.6, less than 0.5, less than 0.4, less than 0.3, less than 0.2 and less than 0.1. In one particular embodiment, significant decreased risk is less than 0.7. In another embodiment, significant decreased risk is less than 0.5. In yet another embodiment, significant decreased risk is less than 0.3. In another embodiment, the decrease in risk (or susceptibility) is at least 20%, including but not limited to at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95% and at least 98%. In one particular embodiment, a significant decrease in risk is at least about 30%. In another embodiment, a significant decrease in risk is at least about 50%. In another embodiment, the decrease in risk is at least about 70%. Other cutoffs or ranges as deemed suitable by the person skilled in the art to characterize the invention are however also contemplated, and those are also within scope of the present invention.

The person skilled in the art will appreciate that for markers with two alleles present in the population being studied (such as SNPs), and wherein one allele is found in increased frequency in a group of individuals with a trait or disease in the population, compared with controls, the other allele of the marker will be found in decreased frequency in the group of individuals with the trait or disease, compared with controls. In such a case, one allele of the marker (the one found in increased frequency in individuals with the trait or disease) will be the at-risk allele, while the other allele will be a protective allele.

Linkage Disequilibrium

The natural phenomenon of recombination, which occurs on average once for each chromosomal pair during each meiotic event, represents one way in which nature provides variations in sequence (and biological function by consequence). It has been discovered that recombination does not occur randomly in the genome; rather, there are large variations in the frequency of recombination rates, resulting in small regions of high recombination frequency (also called recombination hotspots) and larger regions of low recombination frequency, which are commonly referred to as Linkage Disequilibrium (LD) blocks (Myers, S. *et al.*, *Biochem Soc Trans* 34:526-530 (2006); Jeffreys, A.J., *et al.*, *Nature Genet* 29:217-222 (2001); May, C.A., *et al.*, *Nature Genet* 31:272-275(2002)).

Linkage Disequilibrium (LD) refers to a non-random assortment of two genetic elements. For example, if a particular genetic element (e.g., an allele of a polymorphic marker, or a haplotype) occurs in a population at a frequency of 0.50 (50%) and another element occurs at a frequency of 0.50 (50%), then the predicted occurrence of a person's having both elements is 0.25 (25%), assuming a random distribution of the elements. However, if it is discovered that the two elements occur together at a frequency higher than 0.25, then the elements are said to be in linkage disequilibrium, since they tend to be inherited together at a higher rate than what their independent frequencies of occurrence (e.g., allele or haplotype frequencies) would predict. Roughly speaking, LD is generally correlated with the

frequency of recombination events between the two elements. Allele or haplotype frequencies can be determined in a population by genotyping individuals in a population and determining the frequency of the occurrence of each allele or haplotype in the population. For populations of diploids, e.g., human populations, individuals will typically have two alleles for each genetic element (e.g., a marker, haplotype or gene).

Many different measures have been proposed for assessing the strength of linkage disequilibrium (LD). Most capture the strength of association between pairs of biallelic sites. Two important pairwise measures of LD are r^2 (sometimes denoted Δ^2) and $|D'|$. Both measures range from 0 (no disequilibrium) to 1 ('complete' disequilibrium), but their interpretation is slightly different. $|D'|$ is defined in such a way that it is equal to 1 if just two or three of the possible haplotypes are present, and it is <1 if all four possible haplotypes are present. Therefore, a value of $|D'|$ that is <1 indicates that historical recombination may have occurred between two sites (recurrent mutation can also cause $|D'|$ to be <1 , but for single nucleotide polymorphisms (SNPs) this is usually regarded as being less likely than recombination). The measure r^2 represents the statistical correlation between two sites, and takes the value of 1 if only two haplotypes are present.

The r^2 measure is arguably the most relevant measure for association mapping, because there is a simple inverse relationship between r^2 and the sample size required to detect association between susceptibility loci and SNPs. These measures are defined for pairs of sites, but for some applications a determination of how strong LD is across an entire region that contains many polymorphic sites might be desirable (e.g., testing whether the strength of LD differs significantly among loci or across populations, or whether there is more or less LD in a region than predicted under a particular model). Measuring LD across a region is not straightforward, but one approach is to use the measure r , which was developed in population genetics. Roughly speaking, r measures how much recombination would be required under a particular population model to generate the LD that is seen in the data. This type of method can potentially also provide a statistically rigorous approach to the problem of determining whether LD data provide evidence for the presence of recombination hotspots. For the methods described herein, a significant r^2 value can be at least 0.1 such as at least 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99 or 1.0. In one preferred embodiment, the significant r^2 value can be at least 0.2. Alternatively, linkage disequilibrium as described herein, refers to linkage disequilibrium characterized by values of $|D'|$ of at least 0.2, such as 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99. Thus, linkage disequilibrium represents a correlation between alleles of distinct markers. It is measured by correlation coefficient or $|D'|$ (r^2 up to 1.0 and $|D'|$ up to 1.0). In certain embodiments, linkage disequilibrium is defined in terms of values for both the r^2 and $|D'|$ measures. In one such embodiment, a significant linkage disequilibrium is defined as $r^2 > 0.1$ and $|D'| > 0.8$. In another embodiment, a significant linkage disequilibrium is defined as $r^2 > 0.2$ and $|D'| > 0.9$. Other combinations and permutations of values of r^2 and $|D'|$ for determining linkage disequilibrium are also possible, and within the scope of the invention. Linkage disequilibrium can be determined in a single human population, as defined herein, or it can be determined in a collection of samples comprising individuals from more than one human population. In one embodiment of the invention, LD is determined in a sample from one or more of the HapMap populations (caucasian (CEU), african (YRI), japanese (JPT), chinese (CHB)), as defined (<http://www.hapmap.org>). In one such

embodiment, LD is determined in the CEU population of the HapMap samples. In another embodiment, LD is determined in the YRI population. In yet another embodiment, LD is determined in samples from the Icelandic population.

If all polymorphisms in the genome were identical at the population level, then every single one of them would need to be investigated in association studies. However, due to linkage disequilibrium between polymorphisms, tightly linked polymorphisms are strongly correlated, which reduces the number of polymorphisms that need to be investigated in an association study to observe a significant association. Another consequence of LD is that many polymorphisms may give an association signal due to the fact that these polymorphisms are strongly correlated.

Genomic LD maps have been generated across the genome, and such LD maps have been proposed to serve as framework for mapping disease-genes (Risch, N. & Merkiangas, K, *Science* 273:1516-1517 (1996); Maniatis, N., *et al.*, *Proc Natl Acad Sci USA* 99:2228-2233 (2002); Reich, DE *et al.*, *Nature* 411:199-204 (2001)).

It is now established that many portions of the human genome can be broken into series of discrete haplotype blocks containing a few common haplotypes; for these blocks, linkage disequilibrium data provides little evidence indicating recombination (see, e.g., Wall, J.D. and Pritchard, J.K., *Nature Reviews Genetics* 4:587-597 (2003); Daly, M. *et al.*, *Nature Genet.* 29:229-232 (2001); Gabriel, S.B. *et al.*, *Science* 296:2225-2229 (2002); Patil, N. *et al.*, *Science* 294:1719-1723 (2001); Dawson, E. *et al.*, *Nature* 418:544-548 (2002); Phillips, M.S. *et al.*, *Nature Genet.* 33:382-387 (2003)).

There are two main methods for defining these haplotype blocks: blocks can be defined as regions of DNA that have limited haplotype diversity (see, e.g., Daly, M. *et al.*, *Nature Genet.* 29:229-232 (2001); Patil, N. *et al.*, *Science* 294:1719-1723 (2001); Dawson, E. *et al.*, *Nature* 418:544-548 (2002); Zhang, K. *et al.*, *Proc. Natl. Acad. Sci. USA* 99:7335-7339 (2002)), or as regions between transition zones having extensive historical recombination, identified using linkage disequilibrium (see, e.g., Gabriel, S.B. *et al.*, *Science* 296:2225-2229 (2002); Phillips, M.S. *et al.*, *Nature Genet.* 33:382-387 (2003); Wang, N. *et al.*, *Am. J. Hum. Genet.* 71:1227-1234 (2002); Stumpf, M.P., and Goldstein, D.B., *Curr. Biol.* 13:1-8 (2003)). More recently, a fine-scale map of recombination rates and corresponding hotspots across the human genome has been generated (Myers, S., *et al.*, *Science* 310:321-32324 (2005); Myers, S. *et al.*, *Biochem Soc Trans* 34:526530 (2006)). The map reveals the enormous variation in recombination across the genome, with recombination rates as high as 10-60 cM/Mb in hotspots, while closer to 0 in intervening regions, which thus represent regions of limited haplotype diversity and high LD. The map can therefore be used to define haplotype blocks/LD blocks as regions flanked by recombination hotspots. As used herein, the terms "haplotype block" or "LD block" includes blocks defined by any of the above described characteristics, or other alternative methods used by the person skilled in the art to define such regions.

Some representative methods for identification of haplotype blocks are set forth, for example, in U.S. Published Patent Application Nos. 20030099964, 20030170665, 20040023237 and 20040146870. Haplotype blocks can be used to map associations between phenotype and haplotype status, using single markers or haplotypes comprising a plurality of markers. The main haplotypes can be identified

in each haplotype block, and then a set of "tagging" SNPs or markers (the smallest set of SNPs or markers needed to distinguish among the haplotypes) can then be identified. These tagging SNPs or markers can then be used in assessment of samples from groups of individuals, in order to identify association between phenotype and haplotype. If desired, neighboring haplotype blocks can be
5 assessed concurrently, as there may also exist linkage disequilibrium among the haplotype blocks.

It has thus become apparent that for any given observed association to a polymorphic marker in the genome, it is likely that additional markers in the genome also show association. This is a natural consequence of the uneven distribution of LD across the genome, as observed by the large variation in recombination rates. The markers used to detect association thus in a sense represent "tags" for a
10 genomic region (i.e., a haplotype block or LD block) that is associating with a given disease or trait, and as such are useful for use in the methods and kits of the present invention. One or more causative (functional) variants or mutations may reside within the region found to be associating to the disease or trait. Such variants may confer a higher relative risk (RR) or odds ratio (OR) than observed for the tagging markers used to detect the association. The present invention thus refers to the markers used
15 for detecting association to the disease, as described herein, as well as markers in linkage disequilibrium with the markers. Thus, in certain embodiments of the invention, markers that are in LD with the markers and/or haplotypes of the invention, as described herein, may be used as surrogate markers. The surrogate markers have in one embodiment relative risk (RR) and/or odds ratio (OR) values smaller than for the markers or haplotypes initially found to be associating with the disease, as
20 described herein. In other embodiments, the surrogate markers have RR or OR values greater than those initially determined for the markers initially found to be associating with the disease, as described herein. An example of such an embodiment would be a rare, or relatively rare (< 10% allelic population frequency) variant in LD with a more common variant (> 10% population frequency) initially found to be associating with the disease, such as the variants described herein. Identifying and using such markers
25 for detecting the association discovered by the inventors as described herein can be performed by routine methods well known to the person skilled in the art, and are therefore within the scope of the present invention.

Determination of haplotype frequency

30 The frequencies of haplotypes in patient and control groups can be estimated using an expectation-maximization algorithm (Dempster A. *et al.*, *J. R. Stat. Soc. B*, 39:1-38 (1977)). An implementation of this algorithm that can handle missing genotypes and uncertainty with the phase can be used. Under the null hypothesis, the patients and the controls are assumed to have identical frequencies. Using a likelihood approach, an alternative hypothesis is tested, where a candidate at-risk-
35 haplotype, which can include the markers described herein, is allowed to have a higher frequency in patients than controls, while the ratios of the frequencies of other haplotypes are assumed to be the same in both groups. Likelihoods are maximized separately under both hypotheses and a corresponding 1-df likelihood ratio statistic is used to evaluate the statistical significance.

To look for at-risk and protective markers and haplotypes within a linkage region, for example, association of all possible combinations of genotyped markers is studied, provided those markers span a practical region. The combined patient and control groups can be randomly divided into two sets, equal in size to the original group of patients and controls. The marker and haplotype analysis is then repeated and the most significant p-value registered is determined. This randomization scheme can be repeated, for example, over 100 times to construct an empirical distribution of p-values. In a preferred embodiment, a p-value of <0.05 is indicative of a significant marker and/or haplotype association.

Haplotype Analysis

One general approach to haplotype analysis involves using likelihood-based inference applied to NESTED MODELS (Gretarsdottir S., et al., *Nat. Genet.* 35:131-38 (2003)). The method is implemented in the program NEMO, which allows for many polymorphic markers, SNPs and microsatellites. The method and software are specifically designed for case-control studies where the purpose is to identify haplotype groups that confer different risks. It is also a tool for studying LD structures. In NEMO, maximum likelihood estimates, likelihood ratios and p-values are calculated directly, with the aid of the EM algorithm, for the observed data treating it as a missing-data problem.

Even though likelihood ratio tests based on likelihoods computed directly for the observed data, which have captured the information loss due to uncertainty in phase and missing genotypes, can be relied on to give valid p-values, it would still be of interest to know how much information had been lost due to the information being incomplete. The information measure for haplotype analysis is described in Nicolae and Kong (Technical Report 537, Department of Statistics, University of Statistics, University of Chicago; *Biometrics*, 60(2):368-75 (2004)) as a natural extension of information measures defined for linkage analysis, and is implemented in NEMO.

For single marker association to a disease, the Fisher exact test can be used to calculate two-sided p-values for each individual allele. Usually, all p-values are presented unadjusted for multiple comparisons unless specifically indicated. The presented frequencies (for microsatellites, SNPs and haplotypes) are allelic frequencies as opposed to carrier frequencies. To minimize any bias due the relatedness of the patients who were recruited as families for the linkage analysis, first and second-degree relatives can be eliminated from the patient list. Furthermore, the test can be repeated for association correcting for any remaining relatedness among the patients, by extending a variance adjustment procedure described in Risch, N. & Teng, J. (*Genome Res.*, 8:1273-1288 (1998)), DNA pooling (*ibid*) for sibships so that it can be applied to general familial relationships, and present both adjusted and unadjusted p-values for comparison. The differences are in general very small as expected. To assess the significance of single-marker association corrected for multiple testing we can carry out a randomization test using the same genotype data. Cohorts of patients and controls can be randomized and the association analysis redone multiple times (e.g., up to 500,000 times) and the p-value is the fraction of replications that produced a p-value for some marker allele that is lower than or equal to the p-value we observed using the original patient and control cohorts.

For both single-marker and haplotype analyses, relative risk (RR) and the population attributable risk (PAR) can be calculated assuming a multiplicative model (haplotype relative risk model) (Terwilliger, J.D. & Ott, J., *Hum. Hered.* 42:337-46 (1992) and Falk, C.T. & Rubinstein, P, *Ann. Hum. Genet.* 51 (Pt 3):227-33 (1987)), i.e., that the risks of the two alleles/haplotypes a person carries multiply. For example, if RR is the risk of A relative to a, then the risk of a person homozygote AA will be RR times that of a heterozygote Aa and RR^2 times that of a homozygote aa. The multiplicative model has a nice property that simplifies analysis and computations — haplotypes are independent, i.e., in Hardy-Weinberg equilibrium, within the affected population as well as within the control population. As a consequence, haplotype counts of the affecteds and controls each have multinomial distributions, but with different haplotype frequencies under the alternative hypothesis. Specifically, for two haplotypes, h_i and h_j , $\text{risk}(h_i)/\text{risk}(h_j) = (f/p_i)/(f/p_j)$, where f and p denote, respectively, frequencies in the affected population and in the control population. While there is some power loss if the true model is not multiplicative, the loss tends to be mild except for extreme cases. Most importantly, p-values are always valid since they are computed with respect to null hypothesis.

15

Linkage Disequilibrium Using NEMO

LD between pairs of markers can be calculated using the standard definition of D' and r^2 (Lewontin, R., *Genetics* 49:49-67 (1964); Hill, W.G. & Robertson, A. *Theor. Appl. Genet.* 22:226-231 (1968)). Using NEMO, frequencies of the two marker allele combinations are estimated by maximum likelihood and deviation from linkage equilibrium is evaluated by a likelihood ratio test. The definitions of D' and r^2 are extended to include microsatellites by averaging over the values for all possible allele combination of the two markers weighted by the marginal allele probabilities. When plotting all marker combination to elucidate the LD structure in a particular region, we plot D' in the upper left corner and the p-value in the lower right corner. In the LD plots the markers can be plotted equidistant rather than according to their physical location, if desired.

25

Risk assessment and Diagnostics

As described herein, certain polymorphic markers and haplotypes comprising such markers are found to be useful for risk assessment of cancer (e.g. prostate cancer (e.g., aggressive prostate cancer, lung cancer, colon cancer, breast cancer, melanoma). Risk assessment can involve the use of the markers for diagnosing a susceptibility to cancer. Particular alleles of polymorphic markers are found more frequently in individuals with cancer, than in individuals without diagnosis of cancer. Therefore, these marker alleles have predictive value for detecting cancer, or a susceptibility to cancer, in an individual. Tagging markers within haplotype blocks or LD blocks comprising at-risk markers, such as the markers of the present invention, can be used as surrogates for other markers and/or haplotypes within the haplotype block or LD block. Markers with values of r^2 equal to 1 are perfect surrogates for the at-risk variants, i.e. genotypes for one marker perfectly predicts genotypes for the other. Markers with smaller values of r^2 than 1, can also be surrogates for the at-risk variant, or alternatively represent

35

variants with relative risk values as high as or possibly even higher than the at-risk variant. The at-risk variant identified may not be the functional variant itself, but is in this instance in linkage disequilibrium with the true functional variant. The present invention encompasses the assessment of such surrogate markers for the markers as disclosed herein. Such markers are annotated, mapped and listed in public
5 databases, as well known to the skilled person, or can alternatively be readily identified by sequencing the region or a part of the region identified by the markers of the present invention in a group of individuals, and identify polymorphisms in the resulting group of sequences. As a consequence, the person skilled in the art can readily and without undue experimentation genotype surrogate markers in linkage disequilibrium with the markers and/or haplotypes as described herein. The tagging or
10 surrogate markers in LD with the at-risk variants detected, also have predictive value for detecting association to the cancer, or a susceptibility to the cancer, in an individual. These tagging or surrogate markers that are in LD with the markers of the present invention can also include other markers that distinguish among haplotypes, as these similarly have predictive value for detecting susceptibility to cancer.

15 The markers and haplotypes of the invention, e.g., the markers presented in Table 4A, 4B, 5A, 5B, 5C, may be useful for risk assessment and diagnostic purposes for, either alone or in combination. Thus, even in cases where the increase in risk by individual markers is relatively modest, i.e. on the order of 10-30%, the association may have significant implications. Thus, relatively common variants may have significant contribution to the overall risk (Population Attributable Risk is high), or combination
20 of markers can be used to define groups of individual who, based on the combined risk of the markers, is at significant combined risk of developing the disease.

Thus, in one embodiment of the invention, a plurality of variants (genetic markers, biomarkers and/or haplotypes) is used for overall risk assessment. These variants are in one embodiment selected from the variants as disclosed herein. Other embodiments include the use of the variants of the present
25 invention in combination with other variants known to be useful for diagnosing a susceptibility to cancer. In such embodiments, the genotype status of a plurality of markers and/or haplotypes is determined in an individual, and the status of the individual compared with the population frequency of the associated variants, or the frequency of the variants in clinically healthy subjects, such as age-matched and sex-matched subjects. Methods known in the art, such as multivariate analyses or joint risk analyses, may
30 subsequently be used to determine the overall risk conferred based on the genotype status at the multiple loci. Assessment of risk based on such analysis may subsequently be used in the methods and kits of the invention, as described herein.

As described in the above, the haplotype block structure of the human genome has the effect that a large number of variants (markers and/or haplotypes) in linkage disequilibrium with the variant
35 originally associated with a disease or trait may be used as surrogate markers for assessing association to the disease or trait. The number of such surrogate markers will depend on factors such as the historical recombination rate in the region, the mutational frequency in the region (i.e., the number of polymorphic sites or markers in the region), and the extent of LD (size of the LD block) in the region. These markers are usually located within the physical boundaries of the LD block or haplotype block in
40 question as defined using the methods described herein, or by other methods known to the person

skilled in the art. However, sometimes marker and haplotype association is found to extend beyond the physical boundaries of the haplotype block as defined. Such markers and/or haplotypes may in those cases be also used as surrogate markers and/or haplotypes for the markers and/or haplotypes physically residing within the haplotype block as defined. As a consequence, markers and haplotypes in LD (typically characterized by r^2 greater than 0.1, such as r^2 greater than 0.2, including r^2 greater than 0.3, also including r^2 greater than 0.4) with the markers and haplotypes described herein are also within the scope of the invention, even if they are physically located beyond the boundaries of the haplotype block as defined. The invention thus relates to markers that are described herein (e.g., Table 4A, 4B, 5A, 5B, 5C), but may also include other markers that are in strong LD (e.g., characterized by r^2 greater than 0.1 or 0.2 and/or $|D'| > 0.8$) with one or more of the markers listed herein.

For the SNP markers described herein, the opposite allele to the allele found to be in excess in patients (at-risk allele) with a particular cancer (e.g., prostate cancer) is found in decreased frequency in patients with the cancer. These markers and haplotypes in LD and/or comprising such markers, are thus protective for the cancer, i.e. they confer a decreased risk or susceptibility of individuals carrying these markers and/or haplotypes developing the cancer. In other embodiments, haplotypes comprising at least two polymorphic markers may be found in decreased frequency in individuals with the particular cancer, and are thus protective for the cancer. Such markers and haplotypes are useful for diagnosing a decreased susceptibility to the cancer in an individual.

Certain variants of the present invention, including certain haplotypes comprise, in some cases, a combination of various genetic markers, e.g., SNPs and microsatellites. Detecting haplotypes can be accomplished by methods known in the art and/or described herein for detecting sequences at polymorphic sites. Furthermore, correlation between certain haplotypes or sets of markers and disease phenotype can be verified using standard techniques. A representative example of a simple test for correlation would be a Fisher-exact test on a two by two table.

In specific embodiments, a marker allele or haplotype found to be associated with cancer, (e.g., marker alleles as listed in Tables 1, 2, 3, 4A and 4B) is one in which the marker allele or haplotype is more frequently present in an individual at risk for cancer (affected), compared to the frequency of its presence in a healthy individual (control), wherein the presence of the marker allele or haplotype is indicative of cancer or a susceptibility to cancer. In other embodiments, at-risk markers in linkage disequilibrium with one or more markers found to be associated with cancer are tagging markers that are more frequently present in an individual at risk for cancer (affected), compared to the frequency of their presence in a healthy individual (control), wherein the presence of the tagging markers is indicative of increased susceptibility to cancer. In a further embodiment, at-risk markers alleles (i.e. conferring increased susceptibility) in linkage disequilibrium with one or more markers found to be associated with cancer (e.g., marker alleles as listed in Table 4A, 4B, 5A, 5B and 5C), are markers comprising one or more allele that is more frequently present in an individual at risk for cancer, compared to the frequency of their presence in a healthy individual (control), wherein the presence of the markers is indicative of increased susceptibility to .

Study population

In a general sense, the methods and kits of the invention can be utilized from samples containing genomic DNA from any source, i.e. any individual. In preferred embodiments, the individual is a human individual. The individual can be an adult, child, or fetus. The present invention also provides for assessing markers and/or haplotypes in individuals who are members of a target population. Such a target population is in one embodiment a population or group of individuals at risk of developing the disease, based on other genetic factors, biomarkers, biophysical parameters (e.g., weight, BMD, blood pressure), or general health and/or lifestyle parameters (e.g., history of disease or related diseases, previous diagnosis of disease, family history of disease).

The invention provides for embodiments that include individuals from specific age subgroups, such as those over the age of 40, over age of 45, or over age of 50, 55, 60, 65, 70, 75, 80, or 85. Other embodiments of the invention pertain to other age groups, such as individuals aged less than 85, such as less than age 80, less than age 75, or less than age 70, 65, 60, 55, 50, 45, 40, 35, or age 30. Other embodiments relate to individuals with age at onset of the disease in any of the age ranges described in the above. It is also contemplated that a range of ages may be relevant in certain embodiments, such as age at onset at more than age 45 but less than age 60. Other age ranges are however also contemplated, including all age ranges bracketed by the age values listed in the above. The invention furthermore relates to individuals of either gender, males or females.

The Icelandic population is a Caucasian population of Northern European ancestry. A large number of studies reporting results of genetic linkage and association in the Icelandic population have been published in the last few years. Many of those studies show replication of variants, originally identified in the Icelandic population as being associating with a particular disease, in other populations (Stacey, S.N., *et al.*, *Nat Genet.* May 27 2007 (Epub ahead of print; Helgadóttir, A., *et al.*, *Science* **316**:1491-93 (2007); Steinthorsdóttir, V., *et al.*, *Nat Genet.* **39**:770-75 (2007); Gudmundsson, J., *et al.*, *Nat Genet.* **39**:631-37 (2007); Amundadóttir, L.T., *et al.*, *Nat Genet.* **38**:652-58 (2006); Grant, S.F., *et al.*, *Nat Genet.* **38**:320-23 (2006)). Thus, genetic findings in the Icelandic population have in general been replicated in other populations, including populations from Africa and Asia.

The markers of the present invention found to be associated with cancer (e.g., prostate cancer) are believed to show similar association in other human populations. Particular embodiments comprising individual human populations are thus also contemplated and within the scope of the invention. Such embodiments relate to human subjects that are from one or more human population including, but not limited to, Caucasian populations, European populations, American populations, Eurasian populations, Asian populations, Central/South Asian populations, East Asian populations, Middle Eastern populations, African populations, Hispanic populations, and Oceanian populations. European populations include, but are not limited to, Swedish, Norwegian, Finnish, Russian, Danish, Icelandic, Irish, Kelt, English, Scottish, Dutch, Belgian, French, German, Spanish, Portuguese, Italian, Polish, Bulgarian, Slavic, Serbian, Bosnian, Czech, Greek and Turkish populations. The invention furthermore in other embodiments can be practiced in specific human populations that include Bantu, Mandenk, Yoruba, San, Mbuti Pygmy, Orcadian, Adygel, Russian, Sardinian, Tuscan, Mozabite,

Bedouin, Druze, Palestinian, Balochi, Brahui, Makrani, Sindhi, Pathan, Burusho, Hazara, Uygur, Kalash, Han, Dai, Daur, Hezhen, Lahu, Miao, Oroqen, She, Tujia, Tu, Xibo, Yi, Mongolian, Naxi, Cambodian, Japanese, Yakut, Melanesian, Papuan, Karitinan, Surui, Colombian, Maya and Pima.

In one preferred embodiment, the invention relates to populations that include black African ancestry such as populations comprising persons of African descent or lineage. Black African ancestry may be determined by self reporting as African-Americans, Afro-Americans, Black Americans, being a member of the black race or being a member of the negro race. For example, African Americans or Black Americans are those persons living in North America and having origins in any of the black racial groups of Africa. In another example, self-reported persons of black African ancestry may have at least one parent of black African ancestry or at least one grandparent of black African ancestry.

The racial contribution in individual subjects may also be determined by genetic analysis. Genetic analysis of ancestry may be carried out using unlinked microsatellite markers such as those set out in Smith *et al.* (*Am J Hum Genet* **74**, 1001-13 (2004)). In one embodiment, genetic ancestry is estimated using a set of microsatellite markers selected from the about 2000 microsatellites genotyped in a previously described study (Pritchard, J.K. *et al.*, *Genetics* **115**:945-59 (2000)), using a multi-ethnic cohort. In one such embodiment, a cohort of 35 European Americans, 88 African Americans, 34 Chinese, and 29 Mexican Americans is used, as described herein. One particular embodiment useful for estimating genetic ancestry from this set comprises 30 unlinked microsatellite markers. The selected set shows the most significant differences between European Americans, African Americans, and Asians of the 2000 markers described by Pritchard *et al.*, and also has good quality and yield. Thus, in one embodiment, genetic ancestry is determined by genotyping the set of microsatellite markers consisting of D1S2630, D1S2847, D1S466, D1S493, D2S166, D3S1583, D3S4011, D3S4559, D4S2460, D4S3014, D5S1967, DG5S802, D6S1037, D8S1719, D8S1746, D9S1777, D9S1839, D9S2168, D10S1698, D11S1321, D11S4206, D12S1723, D13S152, D14S588, D17S1799, D17S745, D18S464, D19S113, D20S878 and D22S1172. Appropriate primer pairs for amplifying a segment comprising marker DG5S802 is set forth in SEQ ID NO:4 and SEQ ID NO:5. The skilled person will realize that other combinations of microsatellite markers, or other types of polymorphic markers (e.g., SNPs) may also be used to estimate genetic ancestry.

In certain embodiments, the invention relates to markers and/or haplotypes identified in specific populations, as described in the above. The person skilled in the art will appreciate that measures of linkage disequilibrium (LD) may give different results when applied to different populations. This is due to different population history of different human populations as well as differential selective pressures that may have led to differences in LD in specific genomic regions. It is also well known to the person skilled in the art that certain markers, e.g. SNP markers, have different population frequency in different populations, or are polymorphic in one population but not in another. The person skilled in the art will however apply the methods available and as thought herein to practice the present invention in any given human population. This may include assessment of polymorphic markers in the LD region of the present invention, so as to identify those markers that give strongest association within the specific population. Thus, the at-risk variants of the present invention may reside on different haplotype background and in different frequencies in various human populations. However, utilizing methods

known in the art and the markers of the present invention, the invention can be practiced in any given human population.

Utility of Genetic Testing

5 The person skilled in the art will appreciate and understand that the variants described herein in general do not, by themselves, provide an absolute identification of individuals who will develop a particular cancer, e.g., prostate cancer. The variants described herein do however indicate increased and/or decreased likelihood that individuals carrying the at-risk or protective variants of the invention will develop a particular form of cancer, accompanied by symptoms associated with the cancer. This
10 information is however extremely valuable in itself, as outlined in more detail in the below, as it can be used to, for example, initiate preventive measures at an early stage, perform regular physical and/or mental exams to monitor the progress and/or appearance of symptoms, or to schedule exams at a regular interval to identify early signs of the cancer, so as to be able to apply treatment at an early stage.

 The knowledge about a genetic variant that confers a risk of developing cancer offers the
15 opportunity to apply a genetic-test to distinguish between individuals with increased risk of developing the cancer (i.e. carriers of the at-risk variant) and those with decreased risk of developing the cancer (i.e. carriers of the protective variant). The core values of genetic testing, for individuals belonging to both of the above mentioned groups, are the possibilities of being able to diagnose the cancer at an early stage and provide information to the clinician about prognosis/aggressiveness of the cancer in
20 order to be able to apply the most appropriate treatment. For example, the application of a genetic test for cancer (e.g., prostate cancer (aggressive or high Gleason grade prostate cancer, less aggressive or low Gleason grade prostate cancer)) can provide an opportunity for the detection of the disease at an earlier stage which may lead to the application of therapeutic measures at an earlier stage, and thus can minimize the deleterious effects of the symptoms and serious health consequences conferred by
25 cancer. Some advantages of genetic tests for cancer include:

1. To aid early detection

 The application of a genetic test for prostate cancer can provide an opportunity for the detection of the disease at an earlier stage which leads to higher cure rates, if found locally, and increases survival rates by minimizing regional and distant spread of the tumor. For prostate cancer, a genetic
30 test will most likely increase the sensitivity and specificity of the already generally applied Prostate Specific Antigen (PSA) test and Digital Rectal Examination (DRE). This can lead to lower rates of false positives (thus minimize unnecessary procedures such as needle biopsies) and false negatives (thus increasing detection of occult disease and minimizing morbidity and mortality due to PCA).

2. To determine aggressiveness

35 Genetic testing can provide information about pre-diagnostic prognostic indicators and enable the identification of individuals at high or low risk for aggressive tumor types that can lead to modification in screening strategies. For example, an individual determined to be a carrier of a high risk allele for the

development of aggressive prostate cancer will likely undergo more frequent PSA testing, examination and have a lower threshold for needle biopsy in the presence of an abnormal PSA value.

Furthermore, identifying individuals that are carriers of high or low risk alleles for aggressive tumor types will lead to modification in treatment strategies. For example, if prostate cancer is diagnosed in an individual that is a carrier of an allele that confers increased risk of developing an aggressive form of prostate cancer, then the clinician would likely advise a more aggressive treatment strategy such as a prostatectomy instead of a less aggressive treatment strategy.

As is known in the art, Prostate Specific Antigen (PSA) is a protein that is secreted by the epithelial cells of the prostate gland, including cancer cells. An elevated level in the blood indicates an abnormal condition of the prostate, either benign or malignant. PSA is used to detect potential problems in the prostate gland and to follow the progress of prostate cancer therapy. PSA levels above 4 ng/ml are indicative of the presence of prostate cancer (although as known in the art and described herein, the test is neither very specific nor sensitive).

In one embodiment, the method of the invention is performed in combination with (either prior to, concurrently or after) a PSA assay. In a particular embodiment, the presence of a marker or haplotype, in conjunction with the subject having a PSA level greater than 4 ng/ml, is indicative of a more aggressive prostate cancer and/or a worse prognosis. As described herein, particular markers and haplotypes are associated with high Gleason (i.e., more aggressive) prostate cancer. In another embodiment, the presence of a marker or haplotype, in a patient who has a normal PSA level (e.g., less than 4 ng/ml), is indicative of a high Gleason (i.e., more aggressive) prostate cancer and/or a worse prognosis. A "worse prognosis" or "bad prognosis" occurs when it is more likely that the cancer will grow beyond the boundaries of the prostate gland, metastasize, escape therapy and/or kill the host.

In one embodiment, the presence of a marker or haplotype is indicative of a predisposition to a somatic rearrangement of Chr8q24.21 (e.g., one or more of an amplification, a translocation, an insertion and/or deletion) in a tumor or its precursor. The somatic rearrangement itself may subsequently lead to a more aggressive form of prostate cancer (e.g., a higher histologic grade, as reflected by a higher Gleason score or higher stage at diagnosis, an increased progression of prostate cancer (e.g., to a higher stage), a worse outcome (e.g., in terms of morbidity, complications or death)). As is known in the art, the Gleason grade is a widely used method for classifying prostate cancer tissue for the degree of loss of the normal glandular architecture (size, shape and differentiation of glands). A grade from 1–5 is assigned successively to each of the two most predominant tissue patterns present in the examined tissue sample and are added together to produce the total or combined Gleason grade (scale of 2-10). High numbers indicate poor differentiation and therefore more aggressive cancer.

Aggressive prostate cancer is cancer that grows beyond the prostate, metastasizes and eventually kills the patient. As described herein, one surrogate measure of aggressivity is a high combined Gleason grade. The higher the grade on a scale of 2-10 the more likely it is that a patient has aggressive disease.

As used herein and unless noted differently, the term "stage" is used to define the size and physical extent of a cancer (e.g., prostate cancer). One method of staging various cancers is the TNM method, wherein in the TNM acronym, T stands for tumor size and invasiveness (e.g., the primary tumor in the prostate); N relates to nodal involvement (e.g., prostate cancer that has spread to lymph nodes); and M indicates the presence or absence of metastases (spread to a distant site).

The present invention furthermore relates to risk assessment for cancer (e.g., prostate cancer), including diagnosing whether an individual is at risk for developing the cancer. The polymorphic markers of the present invention can be used alone or in combination, as well as in combination with other factors, including other genetic or non-genetic risk factors or biomarkers (e.g., PSA), for risk assessment of an individual for a particular cancer (e.g., prostate cancer). Many factors may affect the predisposition of an individual towards developing risk of developing cancer are known to the person skilled in the art and can be utilized in such assessment. These include, but are not limited to, age, gender, smoking history and smoking status, physical activity, waist-to-hip circumference ratio, family history of cancer, previously diagnosed cancer, obesity, hypertriglyceridemia, low HDL cholesterol, hypertension, elevated blood pressure, cholesterol levels, HDL cholesterol, LDL cholesterol, triglycerides, apolipoprotein AI and B levels, fibrinogen, ferritin, C-reactive protein and leukotriene levels. Methods known in the art can be used for such overall risk assessment, including multivariate analyses or logistic regression.

METHODS OF THE INVENTION

Methods for the diagnosis of susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, melanoma) are described herein and are encompassed by the invention. Kits for assaying a sample from a subject to detect susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, melanoma) are also encompassed by the invention.

Diagnostic and screening assays

In certain embodiments, the present invention pertains to methods of diagnosing, or aiding in the diagnosis of, cancer or a susceptibility to cancer, by detecting particular alleles at genetic markers that appear more frequently in cancer subjects or subjects who are susceptible to cancer. In a particular embodiment, the invention is a method of diagnosing a susceptibility to prostate cancer (e.g., aggressive prostate cancer), breast cancer, colon cancer, lung cancer and/or melanoma by detecting one or more particular polymorphic markers (e.g., the markers or haplotypes described herein). The present invention describes methods whereby detection of particular markers or haplotypes is indicative of a susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, melanoma). Such prognostic or predictive assays can also be used to determine prophylactic treatment of a subject prior to the onset of symptoms associated with such cancers.

In addition, in certain other embodiments, the present invention pertains to methods of diagnosing, or aiding in the diagnosis of, a decreased susceptibility to cancer, by detecting particular genetic marker alleles or haplotypes that appear less frequently in cancer. In a particular embodiment, the invention is a method of diagnosing a decreased susceptibility to prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer and/or melanoma by detecting one or more particular genetic markers (e.g., the markers or haplotypes described herein). The present invention describes methods whereby detection of particular markers or haplotypes is indicative of a decreased susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, colon cancer, lung cancer, melanoma), or of a protective marker or haplotype against the cancer.

As described and exemplified herein, particular markers or haplotypes associated with the Chr8q24.21 LD Block C (SEQ ID NO:1) and LD Block C' (SEQ ID NO:2) are associated with cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, colon cancer, melanoma). In one embodiment, the marker or haplotype is one that confers a significant risk or susceptibility to prostate cancer, breast cancer, lung cancer, colon cancer and/or melanoma. In another embodiment, the invention pertains to methods of diagnosing a susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, melanoma) in a subject, by screening for a marker or haplotype associated with SEQ ID NO:2 (e.g., markers as set forth in 5A, 5B and 5C, and markers in linkage disequilibrium therewith) that is more frequently present in a subject having, or who is susceptible to, cancer (affected), as compared to the frequency of its presence in a healthy subject (control). In certain embodiments, the marker or haplotype has a p value < 0.05 . In other embodiments, the significance of association is characterized by smaller p-values, such as < 0.01 , < 0.001 , < 0.0001 , < 0.00001 , < 0.000001 , < 0.0000001 or < 0.00000001 .

In these embodiments, the presence of the marker or haplotype is indicative of a susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, colon cancer, lung cancer, melanoma). These diagnostic methods involve detecting the presence or absence of a marker or haplotype that is associated with cancer, as described herein. The haplotypes described herein include combinations of various genetic markers (e.g., SNPs, microsatellites). The detection of the particular genetic markers that make up the particular haplotypes can be performed by a variety of methods described herein and/or known in the art. For example, genetic markers can be detected at the nucleic acid level (e.g., by direct nucleotide sequencing) or at the amino acid level if the genetic marker affects the coding sequence of a protein encoded by a cancer-associated nucleic acid, e.g. a nucleic acid whose sequence is as set forth in SEQ ID NO:1 or SEQ ID NO:2 (e.g., by protein sequencing or by immunoassays using antibodies that recognize such a protein). The marker alleles or haplotypes of the present invention correspond to fragments of a genomic DNA sequence associated with cancer (e.g. prostate cancer). Such fragments encompass the DNA sequence of the polymorphic marker or haplotype in question, but may also include DNA segments in strong LD (linkage disequilibrium) with the marker or haplotype. In one embodiment, such segments comprises genomic segments in LD with the marker or haplotype as determined by a value of r^2 greater than 0.2 and/or $|D'| > 0.8$. Examples of such segments in LD with the variants of the invention are set forth in SEQ ID NO:1 and SEQ IN NO:2.

In one embodiment, diagnosis of a susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, melanoma) can be accomplished using hybridization methods, such as Southern analysis, Northern analysis, and/or *in situ* hybridizations (see Current Protocols in Molecular Biology, Ausubel, F. *et al.*, eds., John Wiley & Sons, including all
5 supplements). A biological sample from a test subject or individual (a "test sample") of genomic DNA, RNA, or cDNA is obtained from a subject suspected of having, being susceptible to, or predisposed for cancer (the "test subject"). The subject can be an adult, child, or fetus. The test sample can be from any source that contains genomic DNA, such as a blood sample, sample of amniotic fluid, sample of cerebrospinal fluid, or tissue sample from skin, muscle, buccal or conjunctival mucosa, placenta,
10 gastrointestinal tract or other organs. A test sample of DNA from fetal cells or tissue can be obtained by appropriate methods, such as by amniocentesis or chorionic villus sampling. The DNA, RNA, or cDNA sample is then examined. The presence of a specific marker allele can be indicated by sequence-specific hybridization of a nucleic acid probe specific for the particular allele. The presence of more than specific marker allele or a specific haplotype can be indicated by using several sequence-specific
15 nucleic acid probes, each being specific for a particular allele. In one embodiment, a haplotype can be indicated by a single nucleic acid probe that is specific for the specific haplotype (i.e., hybridizes specifically to a DNA strand comprising the specific marker alleles characteristic of the haplotype). A sequence-specific probe can be directed to hybridize to genomic DNA, RNA, or cDNA. A "nucleic acid probe", as used herein, can be a DNA probe or an RNA probe that hybridizes to a complementary
20 sequence. One of skill in the art would know how to design such a probe so that sequence specific hybridization will occur only if a particular allele is present in a genomic sequence from a test sample.

To diagnose a susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, melanoma), a hybridization sample is formed by contacting the test sample containing a cancer-associated nucleic acid, with at least one nucleic acid probe. A non-limiting
25 example of a probe for detecting mRNA or genomic DNA is a labeled nucleic acid probe that is capable of hybridizing to mRNA or genomic DNA sequences described herein. The nucleic acid probe can be, for example, a full-length nucleic acid molecule, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length that is sufficient to specifically hybridize under stringent conditions to appropriate mRNA or genomic DNA. For example, the nucleic acid probe can be
30 all or a portion of SEQ ID NO:1 or all or a portion of SEQ ID NO:2, optionally comprising at least one allele contained in the haplotypes described herein, or the probe can be the complementary sequence of such a sequence. In a particular embodiment, the nucleic acid probe is a portion of SEQ ID NO:1 or a portion of SEQ ID NO:2, optionally comprising at least one allele contained in the haplotypes described herein, or the probe can be the complementary sequence of such a sequence. Other suitable
35 probes for use in the diagnostic assays of the invention are described herein.

Hybridization can be performed by methods well known to the person skilled in the art (see, e.g., Current Protocols in Molecular Biology, Ausubel, F. *et al.*, eds., John Wiley & Sons, including all supplements). In one embodiment, hybridization refers to specific hybridization, i.e., hybridization with no mismatches (exact hybridization). In one embodiment, the hybridization conditions for specific
40 hybridization are high stringency.

Specific hybridization, if present, is detected using standard methods. If specific hybridization occurs between the nucleic acid probe and the nucleic acid in the test sample, then the sample contains the allele that is complementary to the nucleotide that is present in the nucleic acid probe. The process can be repeated for any markers of the present invention, or markers that make up a haplotype of the present invention, or multiple probes can be used concurrently to detect more than one marker alleles at a time. It is also possible to design a single probe containing more than one marker alleles of a particular haplotype (e.g., a probe containing alleles complementary to 2, 3, 4, 5 or all of the markers that make up a particular haplotype). Detection of the particular markers of the haplotype in the sample is indicative that the source of the sample has the particular haplotype (e.g., a haplotype) and therefore is susceptible to cancer (e.g., prostate cancer).

In one preferred embodiment, a method utilizing a detection oligonucleotide probe comprising a fluorescent moiety or group at its 3' terminus and a quencher at its 5' terminus, and an enhancer oligonucleotide, is employed, as described by Kutyavin *et al.* (*Nucleic Acid Res.* **34**:e128 (2006)). The fluorescent moiety can be Gig Harbor Green or Yakima Yellow, or other suitable fluorescent moieties. The detection probe is designed to hybridize to a short nucleotide sequence that includes the SNP polymorphism to be detected. Preferably, the SNP is anywhere from the terminal residue to -6 residues from the 3' end of the detection probe. The enhancer is a short oligonucleotide probe which hybridizes to the DNA template 3' relative to the detection probe. The probes are designed such that a single nucleotide gap exists between the detection probe and the enhancer nucleotide probe when both are bound to the template. The gap creates a synthetic abasic site that is recognized by an endonuclease, such as Endonuclease IV. The enzyme cleaves the dye off the fully complementary detection probe, but cannot cleave a detection probe containing a mismatch. Thus, by measuring the fluorescence of the released fluorescent moiety, assessment of the presence of a particular allele defined by nucleotide sequence of the detection probe can be performed.

The detection probe can be of any suitable size, although preferably the probe is relatively short. In one embodiment, the probe is from 5-100 nucleotides in length. In another embodiment, the probe is from 10-50 nucleotides in length, and in another embodiment, the probe is from 12-30 nucleotides in length. Other lengths of the probe are possible and within scope of the skill of the average person skilled in the art.

In a preferred embodiment, the DNA template containing the SNP polymorphism is amplified by Polymerase Chain Reaction (PCR) prior to detection. In such an embodiment, the amplified DNA serves as the template for the detection probe and the enhancer probe.

Certain embodiments of the detection probe, the enhancer probe, and/or the primers used for amplification of the template by PCR include the use of modified bases, including modified A and modified G. The use of modified bases can be useful for adjusting the melting temperature of the nucleotide molecule (probe and/or primer) to the template DNA, for example for increasing the melting temperature in regions containing a low percentage of G or C bases, in which modified A with the capability of forming three hydrogen bonds to its complementary T can be used, or for decreasing the melting temperature in regions containing a high percentage of G or C bases, for example by using modified G bases that form only two hydrogen bonds to their complementary C base in a double

stranded DNA molecule. In a preferred embodiment, modified bases are used in the design of the detection nucleotide probe. Any modified base known to the skilled person can be selected in these methods, and the selection of suitable bases is well within the scope of the skilled person based on the teachings herein and known bases available from commercial sources as known to the skilled person.

5 In another hybridization method, Northern analysis (see Current Protocols in Molecular Biology, Ausubel, F. et al., eds., John Wiley & Sons, *supra*) is used to identify the presence of a polymorphism associated with cancer or a susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, melanoma). For Northern analysis, a test sample of RNA is obtained from the subject by appropriate means. As described herein, specific hybridization of a nucleic acid probe to RNA from the subject is indicative of a particular allele complementary to the probe. For
10 representative examples of use of nucleic acid probes, see, for example, U.S. Patent Nos. 5,288,611 and 4,851,330.

Additionally, or alternatively, a peptide nucleic acid (PNA) probe can be used in addition to, or instead of, a nucleic acid probe in the hybridization methods described herein. A PNA is a DNA mimic
15 having a peptide-like, inorganic backbone, such as N-(2-aminoethyl)glycine units, with an organic base (A, G, C, T or U) attached to the glycine nitrogen via a methylene carbonyl linker (see, for example, Nielsen, P., et al., *Bioconjug. Chem.* 5:3-7 (1994)). The PNA probe can be designed to specifically hybridize to a molecule in a sample suspected of containing one or more of the genetic markers of a haplotype that is associated with cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast
20 cancer, lung cancer, melanoma). Hybridization of the PNA probe is diagnostic for cancer or a susceptibility to cancer.

In one embodiment of the invention, a test sample containing genomic DNA obtained from the subject is collected and the polymerase chain reaction (PCR) is used to amplify a fragment comprising one or more markers or haplotypes of the present invention. As described herein, identification of a
25 particular marker allele or haplotype associated with cancer can be accomplished using a variety of methods (e.g., sequence analysis, analysis by restriction digestion, specific hybridization, single stranded conformation polymorphism assays (SSCP), electrophoretic analysis, etc.). In another embodiment, diagnosis is accomplished by expression analysis using quantitative PCR (kinetic thermal cycling). This technique can, for example, utilize commercially available technologies, such as
30 TaqMan® (Applied Biosystems, Foster City, CA). The technique can assess the presence of an alteration in the expression or composition of a polypeptide or splicing variant(s) that is encoded by a nucleic acid associated with cancer (e.g., a nucleic acid whose sequence comprises all or a fragment of the sequence set forth in SEQ ID NO:1 or SEQ ID NO:2). Further, the expression of the variant(s) can be quantified as physically or functionally different.

35 In another method of the invention, analysis by restriction digestion can be used to detect a particular allele if the allele results in the creation or elimination of a restriction site relative to a reference sequence. A test sample containing genomic DNA is obtained from the subject. PCR can be used to amplify particular regions of SEQ ID NO:1 or SEQ ID NO:2 in the test sample from the test subject. Restriction fragment length polymorphism (RFLP) analysis can be conducted, e.g., as described in

Current Protocols in Molecular Biology, *supra*. The digestion pattern of the relevant DNA fragment indicates the presence or absence of the particular allele in the sample.

Sequence analysis can also be used to detect specific alleles at polymorphic sites associated with SEQ ID NO:1 or SEQ ID NO:2. Therefore, in one embodiment, determination of the presence or absence of a particular marker alleles or haplotypes comprises sequence analysis of a test sample of DNA or RNA obtained from a subject or individual. PCR or other appropriate methods can be used to amplify a portion of SEQ ID NO:1 or SEQ ID NO:2, and the presence of a specific allele can then be detected directly by sequencing the polymorphic site (or, alternatively, multiple polymorphic sites in a haplotype) of the genomic DNA in the sample.

Allele-specific oligonucleotides can also be used to detect the presence of a particular allele at a polymorphic site associated with cancer, through the use of dot-blot hybridization of amplified oligonucleotides with allele-specific oligonucleotide (ASO) probes (see, for example, Saiki, R. *et al.*, *Nature*, 324:163-166 (1986)). An "allele-specific oligonucleotide" (also referred to herein as an "allele-specific oligonucleotide probe") is an oligonucleotide of approximately 10-50 base pairs or approximately 15-30 base pairs, that specifically hybridizes to a region of SEQ ID NO:1 or SEQ ID NO:2, and which contains a specific allele at a polymorphic site (e.g., a polymorphism described herein). An allele-specific oligonucleotide probe that is specific for one or more particular polymorphisms associated with SEQ ID NO:1 or SEQ ID NO:2 can be prepared using standard methods (see, e.g., Current Protocols in Molecular Biology, *supra*). PCR can be used to amplify the desired region of SEQ ID NO:1 or SEQ ID NO:2. The DNA containing the amplified LD Block C region can be dot-blotted using standard methods (see, e.g., Current Protocols in Molecular Biology, *supra*), and the blot can be contacted with the oligonucleotide probe. The presence of specific hybridization of the probe to the amplified region can then be detected. Specific hybridization of an allele-specific oligonucleotide probe to DNA from the subject is indicative of a specific allele at a polymorphic site associated with cancer (e.g., prostate cancer) (see, e.g., Gibbs, R. *et al.*, *Nucleic Acids Res.*, 17:2437-2448 (1989) and WO 93/22456).

With the addition of such analogs as locked nucleic acids (LNAs), the size of primers and probes can be reduced to as few as 8 bases. LNAs are a novel class of bicyclic DNA analogs in which the 2' and 4' positions in the furanose ring are joined via an O-methylene (oxy-LNA), S-methylene (thio-LNA), or amino methylene (amino-LNA) moiety. Common to all of these LNA variants is an affinity toward complementary nucleic acids, which is by far the highest reported for a DNA analog. For example, particular all oxy-LNA nonamers have been shown to have melting temperatures (T_m) of 64°C and 74°C when in complex with complementary DNA or RNA, respectively, as opposed to 28°C for both DNA and RNA for the corresponding DNA nonamer. Substantial increases in T_m are also obtained when LNA monomers are used in combination with standard DNA or RNA monomers. For primers and probes, depending on where the LNA monomers are included (e.g., the 3' end, the 5' end, or in the middle), the T_m could be increased considerably.

In another embodiment, arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from a subject, can be used to identify polymorphisms in a cancer-associated nucleic acid. For example, an oligonucleotide array can be used. Oligonucleotide arrays

typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. These oligonucleotide arrays, also described as "Genechips™," have been generally described in the art (see, e.g., U.S. Patent No. 5,143,854, PCT Patent Publication Nos. WO 90/15070 and 92/10092). These arrays can generally be produced using mechanical

5 synthesis methods or light directed synthesis methods that incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis methods (Fodor, S. *et al.*, *Science*, 251:767-773 (1991); Pirrung *et al.*, U.S. Patent No. 5,143,854 (see also published PCT Application No. WO 90/15070); and Fodor, S. *et al.*, published PCT Application No. WO 92/10092 and U.S. Patent No. 5,424,186, the entire teachings of each of which are incorporated by reference herein). Techniques for

10 the synthesis of these arrays using mechanical synthesis methods are described in, e.g., U.S. Patent No. 5,384,261; the entire teachings of which are incorporated by reference herein. In another example, linear arrays can be utilized.

Once an oligonucleotide array is prepared, a nucleic acid of interest is allowed to hybridize with the array. Detection of hybridization is a detection of a particular allele in the nucleic acid of interest.

15 Hybridization and scanning are generally carried out by methods described herein and also in, e.g., published PCT Application Nos. WO 92/10092 and WO 95/11995, and U.S. Patent No. 5,424,186, the entire teachings of each of which are incorporated by reference herein. In brief, a target nucleic acid sequence, which includes one or more previously identified polymorphic markers, is amplified by well-known amplification techniques (e.g., PCR). Typically this involves the use of primer sequences that are

20 complementary to the two strands of the target sequence, both upstream and downstream, from the polymorphic site. Asymmetric PCR techniques can also be used. Amplified target, generally incorporating a label, is then allowed to hybridize with the array under appropriate conditions that allow for sequence-specific hybridization. Upon completion of hybridization and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The

25 hybridization data obtained from the scan is typically in the form of fluorescence intensities as a function of location on the array.

Although primarily described in terms of a single detection block, e.g., for detection of a single polymorphic site, arrays can include multiple detection blocks, and thus be capable of analyzing multiple, specific polymorphisms (e.g., multiple polymorphisms of a particular haplotype (e.g., an

30 haplotype)). In alternate arrangements, it will generally be understood that detection blocks can be grouped within a single array or in multiple, separate arrays so that varying, optimal conditions can be used during the hybridization of the target to the array. For example, it will often be desirable to provide for the detection of those polymorphisms that fall within G-C rich stretches of a genomic sequence, separately from those falling in A-T rich segments. This allows for the separate optimization of

35 hybridization conditions for each situation.

Additional descriptions of use of oligonucleotide arrays for detection of polymorphisms can be found, for example, in U.S. Patent Nos. 5,858,659 and 5,837,832, the entire teachings of both of which are incorporated by reference herein.

Other methods of nucleic acid analysis can be used to detect a particular allele at a

40 polymorphic site associated with cancer (e.g., a polymorphic site associated with the genomic segment

on Chr8q24.21 whose nucleic acid sequence is represented by the sequence set forth in SEQ ID NO:1 and SEQ ID NO:2). Representative methods include, for example, direct manual sequencing (Church and Gilbert, *Proc. Natl. Acad. Sci. USA*, 81: 1991-1995 (1988); Sanger, F., *et al.*, *Proc. Natl. Acad. Sci. USA*, 74:5463-5467 (1977); Beavis, *et al.*, U.S. Patent No. 5,288,644); automated fluorescent
5 sequencing; single-stranded conformation polymorphism assays (SSCP); clamped denaturing gel electrophoresis (CDGE); denaturing gradient gel electrophoresis (DGGE) (Sheffield, V., *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:232-236 (1989)), mobility shift analysis (Orita, M., *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:2766-2770 (1989)), restriction enzyme analysis (Flavell, R., *et al.*, *Cell*, 15:25-41 (1978); Geever, R., *et al.*, *Proc. Natl. Acad. Sci. USA*, 78:5081-5085 (1981)); heteroduplex analysis; chemical
10 mismatch cleavage (CMC) (Cotton, R., *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:4397-4401 (1985)); RNase protection assays (Myers, R., *et al.*, *Science*, 230:1242-1246 (1985); use of polypeptides that recognize nucleotide mismatches, such as *E. coli* mutS protein; and allele-specific PCR.

In another embodiment of the invention, diagnosis of cancer or a susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, melanoma) can be made
15 by examining expression and/or composition of a polypeptide encoded by cancer-associated nucleic acid in those instances where the genetic marker(s) or haplotype described herein results in a change in the composition or expression of the polypeptide. Thus, diagnosis of a susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, melanoma) can be made by examining expression and/or composition of one of these polypeptides, or another polypeptide
20 encoded by cancer-associated nucleic acid, in those instances where the genetic marker or haplotype described herein results in a change in the composition or expression of the polypeptide. The haplotypes and markers described herein that show association to cancer may play a role through their effect on one or more of these nearby genes. Possible mechanisms affecting these genes include, e.g., effects on transcription, effects on RNA splicing, alterations in relative amounts of alternative splice
25 forms of mRNA, effects on RNA stability, effects on transport from the nucleus to cytoplasm, and effects on the efficiency and accuracy of translation.

The *c-myc* gene on Chr8q24.21 encodes the c-MYC protein that was identified over 20 years ago as the cellular counterpart of the viral oncogene *v-myc* of the avian myelocytomatosis retrovirus (Vennstrom *et al.*, *J. Virology* 42:773-79 (1982)). The c-MYC protein is a transcription factor that is
30 rapidly induced upon treatment of cells with mitogenic stimuli. c-MYC regulates the expression of many genes by binding E-boxes (CACGTG) in a heterodimeric complex with a protein named MAX. Many of the genes regulated by c-MYC are involved in cell cycle control. c-MYC promotes cell-cycle progression, inhibits cellular differentiation and induces apoptosis. c-MYC also has a negative effect on double strand DNA repair (Karlsson, A., *et al.*, *Proc. Natl. Acad. Sci. USA* 100(17):9974-79 (2003)). c-
35 MYC also promotes angiogenesis (Ngo, C.V., *et al.*, *Cell Growth Differ.* 11(4):201-10 (2000); Baudino T.A., *et al.*, *Genes Dev.* 16(19):2530-43 (2002)).

The *c-myc* gene is highly tumorigenic *in vitro* and *in vivo*. c-MYC synergizes with proteins that inhibit apoptosis such as BCL, BCL-X_L or with the loss of p53 or ARF in lymphomagenesis in transgenic mice (Strasser *et al.*, *Nature* 348:331-333 (1990); Blyth, K., *et al.*, *Oncogene* 10:1717-23 (1990); Elson, A., *et al.*, *Oncogene* 11:181-90 (1995); Eischen, C.M., *et al.*, *Genes Dev.* 13:2658-69 (1999)).
40

Amplification and overexpression of the *c-myc* gene is seen in prostate cancer and is often associated with aggressive tumors, hormone independence and a poor prognosis (Jenkins, R.B., *et al.*, *Cancer Res.* 57(3):524-31 (1997); El Gedaily, A., *et al.*, *Prostate* 46(3):184-90 (2001); Saramaki, O., *et al.*, *Am. J. Pathol.* 159(6):2089-94 (2001); Bubendorf, L., *et al.*, *Cancer Res.* 59(4):803-06 (1999)). *c-myc* and the Chr8q24.21 region is furthermore gained in prostate, breast and lung tumors and in melanoma (Blancato J., *et al.*, *Br. J. Cancer* 90(8):1612-9 (2004); Kubokura, H., *et al.*, *Ann. Thorac. Cardiovasc. Surg.* 7(4):197-203 (2001); Treszl, A., *et al.*, *Cytometry* 60B(1):37-46 (2004); Kraehn, G.M., *et al.*, *Br. J. Cancer* 84(1):72-79 (2001)). In addition, many other tumor types show a gain of this region including colon, liver, ovary, stomach, intestinal and bladder cancer. Combining all tumor types shows that Chr8q24.21 is the most frequently gained chromosomal region with gain in approximately 17% of all tumor types (www.progenetix.com).

The oncogene is involved in Burkitt's lymphoma as a result of translocations that juxtapose *c-myc* to immunoglobulin enhancers, thereby activating expression of the gene (Dalla-Favera, R., *et al.*, *Proc. Natl. Acad. Sci. USA* 79(24):7824-27 (1982); Taub, R., *et al.*, *Proc. Natl. Acad. Sci. USA* 79(24):7837-41 (1982)). It is also involved in cervical cancer by Human papillomavirus (HPV) integration close to the gene. In most cases, HPV integrations occur in a region spanning 500 kb centromeric and 200 kb telomeric of the *c-myc* gene (Ferber, J.M., *et al.*, *Cancer Genetics Cytogenetics* 154:1-9 (2004); Ferber, M.J., *et al.*, *Oncogene* 22:7233-7242 (2003)).

Two fragile sites, FRA8C and FRA8D, lie centromeric and telomeric to *c-myc*, respectively, on Chr8q24.21. Fragile sites are prone to breakage in the presence of agents that arrest DNA synthesis. Replication of fragile sites is thought to occur late in S-phase and upon induction even later. The involvement of fragile sites in chromosomal amplification, translocation and/or viral insertion may relate to the late replication of these sites and that a break is initiated at or close to stalled replication forks (Hellman, A., *et al.*, *Cancer Cell* 1:89-97 (2002)).

It is possible that markers or haplotypes described here within LD Block C (SEQ ID NO:1) or LD Block C' (SEQ ID NO:2) or in strong LD with LD Block C (SEQ ID NO:1) or LD Block C' (SEQ ID NO:2) (e.g., as measured by r^2 greater than 0.2 and/or $|D'| > 0.8$) could affect the stability of the region leading to gene amplifications of the *c-myc* gene or other nearby genes. That is, a person could inherit a particular variant form of the region as set forth in SEQ ID NO:1 or SEQ ID NO:2 from one or both parents and therefore be more likely to have a somatic mutational event later in one or more cells leading to progression of cancer to a more aggressive form. Thus, in one embodiment, identification of a marker or haplotype of the invention (e.g., a marker or haplotype associated with SEQ ID NO:2 or SEQ ID NO:1) may be used to diagnose a susceptibility to a somatic mutational event, which can lead to progression of cancer to a more aggressive form.

In one embodiment, the marker or haplotype does not comprise a marker that is located within the *c-myc* open reading frame (i.e., chr8:128,705,092-128,710,260 bp in NCBI Build 34). In another embodiment, the marker or haplotype does not comprise a marker that is located within the *c-myc* promoter or open reading frame. In yet another embodiment, the marker or haplotype does not comprise a marker that is located within the *c-myc* promoter, enhancer or open reading frame. In still

other embodiments, the marker or haplotype does not comprise a marker that is located within 1 kb, 2 kb, 5 kb, 10 kb, 15 kb, 20 kb or 25 kb of the *c-myc* open reading frame.

A variety of methods can be used to make such a detection, including enzyme linked immunosorbent assays (ELISA), Western blots, immunoprecipitations and immunofluorescence. A test sample from a subject is assessed for the presence of an alteration in the expression and/or an alteration in composition of the polypeptide encoded by a Chr8q24.21-associated nucleic acid and/or a nucleic acid associated with LD Block C (SEQ ID NO:1) or LD Block C' (SEQ ID NO:2). An alteration in expression of a polypeptide encoded by such a nucleic acid can be, for example, an alteration in the quantitative polypeptide expression (i.e., the amount of polypeptide produced). An alteration in the composition of a polypeptide encoded by the nucleic acid is an alteration in the qualitative polypeptide expression (e.g., expression of a mutant polypeptide or of a different splicing variant). In one embodiment, diagnosis of a susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, colon cancer lung cancer, melanoma) is made by detecting a particular splicing variant encoded by a cancer-associated nucleic acid as described herein (e.g., a Chr8q24.21-associated nucleic acid, a LD Block C (SEQ ID NO:1)-associated nucleic acid, and/or a LD Block C' (SEQ ID NO:2)-associated nucleic acid), or a particular pattern of splicing variants.

Both such alterations (quantitative and qualitative) can also be present. An "alteration" in the polypeptide expression or composition, as used herein, refers to an alteration in expression or composition of a polypeptide in a test sample, as compared to the expression or composition of the polypeptide in a control sample. A control sample is a sample that corresponds to the test sample (e.g., is from the same type of cells), and is from a subject who is not affected by, and/or who does not have a susceptibility to, cancer (e.g., a subject that does not possess a marker or haplotype as described herein). Similarly, the presence of one or more different splicing variants in the test sample, or the presence of significantly different amounts of different splicing variants in the test sample, as compared with the control sample, can be indicative of a susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, colon cancer, lung cancer, melanoma). An alteration in the expression or composition of the polypeptide in the test sample, as compared with the control sample, can be indicative of a specific variant (e.g., marker allele or haplotype) in the instance where the variant alters a splice site relative to the reference in the control sample. Various means of examining expression or composition of a polypeptide encoded by a nucleic acid are known to the person skilled in the art and can be used, including spectroscopy, colorimetry, electrophoresis, isoelectric focusing, and immunoassays (e.g., David *et al.*, U.S. Pat. No. 4,376,110) such as immunoblotting (see, e.g., Current Protocols in Molecular Biology, particularly chapter 10, *supra*).

For example, in one embodiment, an antibody (e.g., an antibody with a detectable label) that is capable of binding to a polypeptide encoded by a nucleic acid associated with cancer as described herein can be used. Antibodies can be polyclonal or monoclonal. An intact antibody, or a fragment thereof (e.g., Fv, Fab, Fab', 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 labeled secondary antibody (e.g., a fluorescently-labeled secondary antibody) and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin.

In one embodiment of this method, the level or amount of polypeptide encoded by a nucleic acid associated with cancer (e.g., prostate cancer) in a test sample is compared with the level or amount of the polypeptide in a control sample. A level or amount of the polypeptide in the test sample that is higher or lower than the level or amount of the polypeptide in the control sample, such that the difference is statistically significant, is indicative of an alteration in the expression of the polypeptide encoded by the nucleic acid, and is diagnostic for a particular allele responsible for causing the difference in expression. Alternatively, the composition of the polypeptide in a test sample is compared with the composition of the polypeptide in a control sample. In another embodiment, both the level or amount and the composition of the polypeptide can be assessed in the test sample and in the control sample.

In another embodiment, the diagnosis of a susceptibility to cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), breast cancer, colon cancer, lung cancer, melanoma) is made by detecting at least one marker or haplotype of the invention (e.g., the markers as set forth in Table 5A, 5B and 5C, and markers or haplotypes in linkage disequilibrium therewith) in combination with an additional protein-based, RNA-based or DNA-based assay (e.g., other cancer diagnostic assays including, but not limited to: PSA assays, carcinoembryonic antigen (CEA) assays, BRCA1 assays BRCA2 assays). Such cancer diagnostic assays are known in the art, and include also other genetic risk factors for cancer known to the skilled person. The methods of the invention can also be used in combination with an analysis of a subject's family history and risk factors (e.g., environmental risk factors, lifestyle risk factors).

As is known in the art, and as described herein, PSA testing has aided early diagnosis of prostate cancer, but it is neither highly sensitive nor specific (Punglia *et al.*, *N. Engl. J. Med.* 349(4):335-42 (2003)). Accordingly, PSA testing alone leads to a high percentage of false negative and false positive diagnoses, which results in both many instances of missed cancers and unnecessary follow-up biopsies for those without cancer. In one embodiment, the diagnosis of prostate cancer or a susceptibility to prostate cancer is made by detecting at least one Chr8q24.21-associated allele and/or LD Block C-associated allele in combination with a PSA assay.

Kits

Kits useful in the methods of the invention comprise components useful in any of the methods described herein, including for example, hybridization probes, restriction enzymes (e.g., for RFLP analysis), allele-specific oligonucleotides, antibodies that bind to an altered polypeptide encoded by a nucleic acid of the invention as described herein (e.g., a genomic segment comprising at least one polymorphic marker and/or haplotype of the present invention) or to a non-altered (native) polypeptide encoded by a nucleic acid of the invention as described herein, means for amplification of a nucleic acid

associated with cancer, means for analyzing the nucleic acid sequence of a nucleic acid associated with cancer, means for analyzing the amino acid sequence of a polypeptide encoded by a nucleic acid associated with cancer, etc. The kits can for example include necessary buffers, nucleic acid primers for amplifying nucleic acids of the invention (e.g., one or more of the polymorphic markers associated
5 with cancer, e.g., the markers set forth in Tables 5A, 5B and 5C), and reagents for allele-specific detection of the fragments amplified using such primers and necessary enzymes (e.g., DNA polymerase). Additionally, kits can provide reagents for assays to be used in combination with the methods of the present invention, e.g., reagents for use with other cancer diagnostic assays.

In one embodiment, the invention is a kit for assaying a sample from a subject to aid in the
10 detection of a particular cancer (e.g., prostate cancer (e.g., aggressive prostate cancer), lung cancer, colon cancer, breast cancer, melanoma) or a susceptibility to cancer (e.g., prostate cancer, lung cancer, colon cancer, breast cancer, melanoma) in a subject, wherein the kit comprises reagents necessary for selectively detecting at least one allele of at least one polymorphism of the present invention in the genome of the individual. In a particular embodiment, the reagents comprise at least one contiguous
15 oligonucleotide that hybridizes to a fragment of the genome of the individual comprising at least one polymorphism of the present invention. In another embodiment, the reagents comprise at least one pair of oligonucleotides that hybridize to opposite strands of a genomic segment obtained from a subject, wherein each oligonucleotide primer pair is designed to selectively amplify a fragment of the genome of the individual that includes at least one polymorphism, wherein the polymorphism is selected from the
20 group consisting of the polymorphisms as set forth in Table 5A, 5B and 5C, and polymorphic markers in linkage disequilibrium therewith. In yet another embodiment the fragment is at least 20 base pairs in size. Such oligonucleotides or nucleic acids (e.g., oligonucleotide primers) can be designed using portions of the nucleic acid sequence flanking polymorphisms (e.g., SNPs or microsatellites) that are indicative of cancer. In another embodiment, the kit comprises one or more labeled nucleic acids
25 capable of allele-specific detection of one or more specific polymorphic markers or haplotypes associated with cancer, and reagents for detection of the label. Suitable labels include, e.g., a radioisotope, a fluorescent label, an enzyme label, an enzyme co-factor label, a magnetic label, a spin label, an epitope label.

In particular embodiments, the polymorphic marker or haplotype to be detected by the reagents
30 of the kit comprises one or more markers, two or more markers, three or more markers, four or more markers or five or more markers selected from the group consisting of the markers set forth in Table 5A, Table 5B and Table 5C. In another embodiment, the marker or haplotype to be detected comprises the markers set forth in Table 4A and Table 4B. In another embodiment, the marker or haplotype to be detected comprises at least one marker from the group of markers in strong linkage disequilibrium, as
35 defined by values of r^2 greater than 0.2, to at least one of the group of markers consisting of the markers listed in Table 4A, and Table 4B. In a preferred embodiment, the marker or haplotype to be detected comprises rs16901979, and markers in linkage disequilibrium therewith. In another preferred embodiment, the marker or haplotype to be detected comprises HapC (rs1456314 allele G, rs17831626 allele T, rs7825414 allele G, rs6993569 allele G, rs6994316 allele A, rs6470494 allele T, rs1016342
40 allele C, rs1031588 allele G, rs1016343 allele T, rs1551510 allele G, rs1456306 allele C, rs1378897 allele G, rs1456305 allele T, and rs7816535 allele G).

In one preferred embodiment, the kit for detecting the markers of the invention comprises a detection oligonucleotide probe, that hybridizes to a segment of template DNA containing a SNP polymorphisms to be detected, an enhancer oligonucleotide probe and an endonuclease. As explained in the above, the detection oligonucleotide probe comprises a fluorescent moiety or group at its 3' terminus and a quencher at its 5' terminus, and an enhancer oligonucleotide, is employed, as described by Kutyavin *et al.* (*Nucleic Acid Res.* **34**:e128 (2006)). The fluorescent moiety can be Gig Harbor Green or Yakima Yellow, or other suitable fluorescent moieties. The detection probe is designed to hybridize to a short nucleotide sequence that includes the SNP polymorphism to be detected. Preferably, the SNP is anywhere from the terminal residue to -6 residues from the 3' end of the detection probe. The enhancer is a short oligonucleotide probe which hybridizes to the DNA template 3' relative to the detection probe. The probes are designed such that a single nucleotide gap exists between the detection probe and the enhancer nucleotide probe when both are bound to the template. The gap creates a synthetic abasic site that is recognized by an endonuclease, such as Endonuclease IV. The enzyme cleaves the dye off the fully complementary detection probe, but cannot cleave a detection probe containing a mismatch. Thus, by measuring the fluorescence of the released fluorescent moiety, assessment of the presence of a particular allele defined by nucleotide sequence of the detection probe can be performed.

The detection probe can be of any suitable size, although preferably the probe is relatively short. In one embodiment, the probe is from 5-100 nucleotides in length. In another embodiment, the probe is from 10-50 nucleotides in length, and in another embodiment, the probe is from 12-30 nucleotides in length. Other lengths of the probe are possible and within scope of the skill of the average person skilled in the art.

In a preferred embodiment, the DNA template containing the SNP polymorphism is amplified by Polymerase Chain Reaction (PCR) prior to detection, and primers for such amplification are included in the reagent kit. In such an embodiment, the amplified DNA serves as the template for the detection probe and the enhancer probe.

Certain embodiments of the detection probe, the enhancer probe, and/or the primers used for amplification of the template by PCR include the use of modified bases, including modified A and modified G. The use of modified bases can be useful for adjusting the melting temperature of the nucleotide molecule (probe and/or primer) to the template DNA, for example for increasing the melting temperature in regions containing a low percentage of G or C bases, in which modified A with the capability of forming three hydrogen bonds to its complementary T can be used, or for decreasing the melting temperature in regions containing a high percentage of G or C bases, for example by using modified G bases that form only two hydrogen bonds to their complementary C base in a double stranded DNA molecule. In a preferred embodiment, modified bases are used in the design of the detection nucleotide probe. Any modified base known to the skilled person can be selected in these methods, and the selection of suitable bases is well within the scope of the skilled person based on the teachings herein and known bases available from commercial sources as known to the skilled person.

In one of such embodiments, the presence of the marker or haplotype is indicative of a susceptibility (increased susceptibility or decreased susceptibility) to cancer (e.g. prostate cancer (e.g.,

aggressive prostate cancer), lung cancer, colon cancer, breast cancer, melanoma). In another embodiment, the presence of the marker or haplotype is indicative of response to a therapeutic agent for cancer. In another embodiment, the presence of the marker or haplotype is indicative of cancer prognosis in an individual. In yet another embodiment, the presence of the marker or haplotype is
5 indicative of progress of cancer treatment. Such treatment may include intervention by surgery, medication or by other means (e.g., lifestyle changes).

Diagnosis of disease associated with variants of the invention

Although the methods of the invention have been generally described in the context of
10 diagnosing susceptibility to cancer (e.g. prostate cancer), the methods can also be used to diagnose cancer associated with the polymorphic markers of the present invention. For example, an individual having cancer or who is at risk for developing cancer can be assessed to determine whether the presence in the individual of a polymorphism or haplotype of the present invention could have been a contributing factor to the diagnosed cancer in the individual. In one embodiment, identification of a
15 cancer associated with the markers and/or haplotypes of the present invention facilitates treatment planning. For example, preventive treatment to minimize the occurrence of the individual developing cancer can be administered. Such preventive treatment can also include assessment of (i) whether the individual is heterozygous or homozygous for the at-risk variant; (ii) the individual's age, and (iii) the individual's sex, since the variants of the present invention have been shown to be associated with
20 lower age at onset of coronary artery disease and myocardial infarction. In other embodiments of the invention, therapies can be designed and therapeutics selected to target the appropriate gene or protein associated with the polymorphism and/or haplotype of the present invention.

In one embodiment of the invention, diagnosis of a cancer associated with the markers and/or haplotypes of the present invention is made by detecting a polymorphism or haplotype of the present
25 invention. Particular polymorphisms are described herein. In particular embodiments, the polymorphic marker or haplotype to be detected comprises one or more markers, two or more markers, three or more markers, four or more markers or five or more markers selected from the group consisting of the markers set forth in Table 5A, Table 5B and Table 5C. In another embodiment, the marker or haplotype to be detected comprises the markers set forth in Table 4A and Table 4B. In another embodiment, the
30 marker or haplotype to be detected comprises at least one marker from the group of markers in strong linkage disequilibrium, as defined by values of r^2 greater than 0.2, to at least one of the group of markers consisting of the markers listed in Table 4A, and Table 4B. In a preferred embodiment, the marker or haplotype to be detected comprises rs16901979, and markers in linkage disequilibrium therewith. In another preferred embodiment, the marker or haplotype to be detected comprises HapC (rs1456314
35 allele G, rs17831626 allele T, rs7825414 allele G, rs6993569 allele G, rs6994316 allele A, rs6470494 allele T, rs1016342 allele C, rs1031588 allele G, rs1016343 allele T, rs1551510 allele G, rs1456306 allele C, rs1378897 allele G, rs1456305 allele T, and rs7816535 allele G).

A test sample of genomic DNA, RNA, or cDNA, is obtained from a subject having cancer to determine whether the disease is associated with one or more polymorphisms of the invention. The

DNA, RNA or cDNA sample is then examined to determine whether a specific allele of a polymorphism or a specific haplotype of the invention is found to be present in the sample. If the nucleic acid sample is found to contain the specific allele of the polymorphism or the haplotype, then the presence of the allele or haplotype is indicative of the cancer being associated with the polymorphism and/or haplotype.

- 5 Methods known to the person skilled in the art, as described in further detail herein in the methods and kits of the present invention, can be used to detect the polymorphism. .

Therapeutic agents

- 10 Variants of the present invention (e.g., the markers and/or haplotypes of the invention, e.g., the markers listed in Tables 5A, 5B and 5C, and markers in linkage disequilibrium therewith, e.g., the markers listed in Table 4A and 4B) can be used to identify novel therapeutic targets for cancer (e.g., prostate cancer). For example, genes containing, or in linkage disequilibrium with, variants (markers and/or haplotypes) associated with cancer, or their products, as well as genes or their products that are directly or indirectly regulated by or interact with these variant genes or their products, can be targeted
- 15 for the development of therapeutic agents to treat cancer, or prevent or delay onset of symptoms associated with cancer. In one embodiment, the gene is *c-myc*. Therapeutic agents may comprise one or more of, for example, small non-protein and non-nucleic acid molecules, proteins, peptides, protein fragments, nucleic acids (DNA, RNA), PNA (peptide nucleic acids), or their derivatives or mimetics which can modulate the function and/or levels of the target genes or their gene products.

- 20 The nucleic acids and/or variants of the invention, or nucleic acids comprising their complementary sequence, may be used as antisense constructs to control gene expression in cells, tissues or organs. The methodology associated with antisense techniques is well known to the skilled artisan, and is described and reviewed in *Antisense Drug Technology: Principles, Strategies, and Applications*, Crooke, ed., Marcel Dekker Inc., New York (2001). In general, antisense nucleic acid
- 25 molecules are designed to be complementary to a region of mRNA expressed by a gene, so that the antisense molecule hybridizes to the mRNA, thus blocking translation of the mRNA into protein. Several classes of antisense oligonucleotide are known to those skilled in the art, including cleavers and blockers. The former bind to target RNA sites, activate intracellular nucleases (e.g., RNaseH or RNase L), that cleave the target RNA. Blockers bind to target RNA, inhibit protein translation by steric
- 30 hindrance of the ribosomes. Examples of blockers include nucleic acids, morpholino compounds, locked nucleic acids and methylphosphonates (Thompson, *Drug Discovery Today*, 7:912-917 (2002)). Antisense oligonucleotides are useful directly as therapeutic agents, and are also useful for determining and validating gene function, for example by gene knock-out or gene knock-down experiments. Antisense technology is further described in Lavery *et al.*, *Curr. Opin. Drug Discov. Devel.* 6:561-569
- 35 (2003), Stephens *et al.*, *Curr. Opin. Mol. Ther.* 5:118-122 (2003), Kurreck, *Eur. J. Biochem.* 270:1628-44 (2003), Dias *et al.*, *Mol. Cancer Ther.* 1:347-55 (2002), Chen, *Methods Mol. Med.* 75:621-636 (2003), Wang *et al.*, *Curr. Cancer Drug Targets* 1:177-96 (2001), and Bennett, *Antisense Nucleic Acid Drug.Dev.* 12:215-24 (2002)

The variants described herein can be used for the selection and design of antisense reagents that are specific for particular variants. Using information about the variants described herein, antisense oligonucleotides or other antisense molecules that specifically target mRNA molecules that contain one or more variants of the invention can be designed. In this manner, expression of mRNA molecules that contain one or more variant of the present invention (markers and/or haplotypes) can be inhibited or blocked. In one embodiment, the antisense molecules are designed to specifically bind a particular allelic form (i.e., one or several variants (alleles and/or haplotypes)) of the target nucleic acid, thereby inhibiting translation of a product originating from this specific allele or haplotype, but which do not bind other or alternate variants at the specific polymorphic sites of the target nucleic acid molecule.

As antisense molecules can be used to inactivate mRNA so as to inhibit gene expression, and thus protein expression, the molecules can be used to treat a disease such as cancer, including prostate cancer (e.g., aggressive prostate cancer), lung cancer, colon cancer, breast cancer, melanoma. The methodology can involve cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Such mRNA regions include, for example, protein-coding regions, in particular protein-coding regions corresponding to catalytic activity, substrate and/or ligand binding sites, or other functional domains of a protein.

The phenomenon of RNA interference (RNAi) has been actively studied for the last decade, since its original discovery in *C. elegans* (Fire *et al.*, *Nature* 391:806-11 (1998)), and in recent years its potential use in treatment of human disease has been actively pursued (reviewed in Kim & Rossi, *Nature Rev. Genet.* 8:173-204 (2007)). RNA interference (RNAi), also called gene silencing, is based on using double-stranded RNA molecules (dsRNA) to turn off specific genes. In the cell, cytoplasmic double-stranded RNA molecules (dsRNA) are processed by cellular complexes into small interfering RNA (siRNA). The siRNA guide the targeting of a protein-RNA complex to specific sites on a target mRNA, leading to cleavage of the mRNA (Thompson, *Drug Discovery Today*, 7:912-917 (2002)). The siRNA molecules are typically about 20, 21, 22 or 23 nucleotides in length. Thus, one aspect of the invention relates to isolated nucleic acid molecules, and the use of those molecules for RNA interference, i.e. as small interfering RNA molecules (siRNA). In one embodiment, the isolated nucleic acid molecules are 18-26 nucleotides in length, preferably 19-25 nucleotides in length, more preferably 20-24 nucleotides in length, and more preferably 21, 22 or 23 nucleotides in length.

Another pathway for RNAi-mediated gene silencing originates in endogenously encoded primary microRNA (pri-miRNA) transcripts, which are processed in the cell to generate precursor miRNA (pre-miRNA). These miRNA molecules are exported from the nucleus to the cytoplasm, where they undergo processing to generate mature miRNA molecules (miRNA), which direct translational inhibition by recognizing target sites in the 3' untranslated regions of mRNAs, and subsequent mRNA degradation by processing P-bodies (reviewed in Kim & Rossi, *Nature Rev. Genet.* 8:173-204 (2007)).

Clinical applications of RNAi include the incorporation of synthetic siRNA duplexes, which preferably are approximately 20-23 nucleotides in size, and preferably have 3' overlaps of 2 nucleotides. Knockdown of gene expression is established by sequence-specific design for the target mRNA.

Several commercial sites for optimal design and synthesis of such molecules are known to those skilled in the art.

Other applications provide longer siRNA molecules (typically 25-30 nucleotides in length, preferably about 27 nucleotides), as well as small hairpin RNAs (shRNAs; typically about 29 nucleotides in length). The latter are naturally expressed, as described in Amarzguioui *et al.* (*FEBS Lett.* 579:5974-81 (2005)). Chemically synthetic siRNAs and shRNAs are substrates for *in vivo* processing, and in some cases provide more potent gene-silencing than shorter designs (Kim *et al.*, *Nature Biotechnol.* 23:222-226 (2005); Siolas *et al.*, *Nature Biotechnol.* 23:227-231 (2005)). In general siRNAs provide for transient silencing of gene expression, because their intracellular concentration is diluted by subsequent cell divisions. By contrast, expressed shRNAs mediate long-term, stable knockdown of target transcripts, for as long as transcription of the shRNA takes place (Marques *et al.*, *Nature Biotechnol.* 23:559-565 (2006); Brummelkamp *et al.*, *Science* 296: 550-553 (2002)).

Since RNAi molecules, including siRNA, miRNA and shRNA, act in a sequence-dependent manner, the variants of the present invention (e.g., the nucleotide sequence of markers set forth in Table 5A, 5B and 5C, and markers in linkage disequilibrium therewith, e.g., the markers set forth in Table 4A and 4B) can be used to design RNAi reagents that recognize specific nucleic acid molecules comprising specific alleles and/or haplotypes (e.g., the alleles and/or haplotypes of the present invention), while not recognizing nucleic acid molecules comprising other alleles or haplotypes. These RNAi reagents can thus recognize and destroy the target nucleic acid molecules. As with antisense reagents, RNAi reagents can be useful as therapeutic agents (i.e., for turning off disease-associated genes or disease-associated gene variants), but may also be useful for characterizing and validating gene function (e.g., by gene knock-out or gene knock-down experiments).

Delivery of RNAi may be performed by a range of methodologies known to those skilled in the art. Methods utilizing non-viral delivery include cholesterol, stable nucleic acid-lipid particle (SNALP), heavy-chain antibody fragment (Fab), aptamers and nanoparticles. Viral delivery methods include use of lentivirus, adenovirus and adeno-associated virus. The siRNA molecules are in some embodiments chemically modified to increase their stability. This can include modifications at the 2' position of the ribose, including 2'-O-methylpurines and 2'-fluoropyrimidines, which provide resistance to Rnase activity. Other chemical modifications are possible and known to those skilled in the art.

The following references provide a further summary of RNAi, and possibilities for targeting specific genes using RNAi: Kim & Rossi, *Nat. Rev. Genet.* 8:173-184 (2007), Chen & Rajewsky, *Nat. Rev. Genet.* 8: 93-103 (2007), Reynolds, *et al.*, *Nat. Biotechnol.* 22:326-330 (2004), Chi *et al.*, *Proc. Natl. Acad. Sci. USA* 100:6343-6346 (2003), Vickers *et al.*, *J. Biol. Chem.* 278:7108-7118 (2003), Agami, *Curr. Opin. Chem. Biol.* 6:829-834 (2002), Lavery, *et al.*, *Curr. Opin. Drug Discov. Devel.* 6:561-569 (2003), Shi, *Trends Genet.* 19:9-12 (2003), Shuey *et al.*, *Drug Discov. Today* 7:1040-46 (2002), McManus *et al.*, *Nat. Rev. Genet.* 3:737-747 (2002), Xia *et al.*, *Nat. Biotechnol.* 20:1006-10 (2002), Plasterk *et al.*, *curr. Opin. Genet. Dev.* 10:562-7 (2000), Boshier *et al.*, *Nat. Cell Biol.* 2:E31-6 (2000), and Hunter, *Curr. Biol.* 9:R440-442 (1999).

A genetic defect leading to increased predisposition or risk for development of a disease, including cancer, or a defect causing the disease, may be corrected permanently by administering to a subject carrying the defect a nucleic acid fragment that incorporates a repair sequence that supplies the normal/wild-type nucleotide(s) at the site of the genetic defect. Such site-specific repair sequence may
5 concompass an RNA/DNA oligonucleotide that operates to promote endogenous repair of a subject's genomic DNA. The administration of the repair sequence may be performed by an appropriate vehicle, such as a complex with polyethelenimine, encapsulated in anionic liposomes, a viral vector such as an adenovirus vector, or other pharmaceutical compositions suitable for promoting intracellular uptake of the administered nucleic acid. The genetic defect may then be overcome, since the chimeric
10 oligonucleotides induce the incorporation of the normal sequence into the genome of the subject, leading to expression of the normal/wild-type gene product. The replacement is propagated, thus rendering a permanent repair and alleviation of the symptoms associated with the disease or condition.

The present invention provides methods for identifying compounds or agents that can be used to treat cancer. Thus, the variants of the invention are useful as targets for the identification and/or
15 development of therapeutic agents. Such methods may include assaying the ability of an agent or compound to modulate the activity and/or expression of a nucleic acid that includes at least one of the variants (markers and/or haplotypes) of the present invention, or the encoded product of the nucleic acid. This in turn can be used to identify agents or compounds that inhibit or alter the undesired activity or expression of the encoded nucleic acid product. Assays for performing such experiments can be
20 performed in cell-based systems or in cell-free systems, as known to the skilled person. Cell-based systems include cells naturally expressing the nucleic acid molecules of interest, or recombinant cells that have been genetically modified so as to express a certain desired nucleic acid molecule.

Variant gene expression in a patient can be assessed by expression of a variant-containing nucleic acid sequence (for example, a gene containing at least one variant of the present invention,
25 which can be transcribed into RNA containing the at least one variant, and in turn translated into protein), or by altered expression of a normal/wild-type nucleic acid sequence due to variants affecting the level or pattern of expression of the normal transcripts, for example variants in the regulatory or control region of the gene. Assays for gene expression include direct nucleic acid assays (mRNA), assays for expressed protein levels, or assays of collateral compounds involved in a pathway, for
30 example a signal pathway. Furthermore, the expression of genes that are up- or down-regulated in response to the signal pathway can also be assayed. One embodiment includes operably linking a reporter gene, such as luciferase, to the regulatory region of the gene(s) of interest.

Modulators of gene expression can in one embodiment be identified when a cell is contacted with a candidate compound or agent, and the expression of mRNA is determined. The expression level
35 of mRNA in the presence of the candidate compound or agent is compared to the expression level in the absence of the compound or agent. Based on this comparison, candidate compounds or agents for treating cancer can be identified as those modulating the gene expression of the variant gene. When expression of mRNA or the encoded protein is statistically significantly greater in the presence of the candidate compound or agent than in its absence, then the candidate compound or agent is identified
40 as a stimulator or up-regulator of expression of the nucleic acid. When nucleic acid expression or

protein level is statistically significantly less in the presence of the candidate compound or agent than in its absence, then the candidate compound is identified as an inhibitor or down-regulator of the nucleic acid expression.

The invention further provides methods of treatment using a compound identified through drug
5 (compound and/or agent) screening as a gene modulator (i.e. stimulator and/or inhibitor of gene expression).

In a further aspect of the present invention, a pharmaceutical pack (kit) is provided, the pack comprising a therapeutic agent and a set of instructions for administration of the therapeutic agent to humans diagnostically tested for one or more variants of the present invention, as disclosed herein. The
10 therapeutic agent can be a small molecule drug, an antibody, a peptide, an antisense or RNAi molecule, or other therapeutic molecules. In one embodiment, an individual identified as a carrier of at least one variant of the present invention is instructed to take a prescribed dose of the therapeutic agent. In one such embodiment, an individual identified as a homozygous carrier of at least one variant of the present invention is instructed to take a prescribed dose of the therapeutic agent. In another embodiment, an
15 individual identified as a non-carrier of at least one variant of the present invention is instructed to take a prescribed dose of the therapeutic agent.

Methods of assessing probability of response to therapeutic agents, methods of monitoring progress of treatment and methods of treatment

As is known in the art, individuals can have differential responses to a particular therapy (e.g., a
20 therapeutic agent or therapeutic method). Pharmacogenomics addresses the issue of how genetic variations (e.g., the variants (markers and/or haplotypes) of the present invention) affect drug response, due to altered drug disposition and/or abnormal or altered action of the drug. Thus, the basis of the differential response may be genetically determined in part. Clinical outcomes due to genetic variations
25 affecting drug response may result in toxicity of the drug in certain individuals (e.g., carriers or non-carriers of the genetic variants of the present invention), or therapeutic failure of the drug. Therefore, the variants of the present invention may determine the manner in which a therapeutic agent and/or method acts on the body, or the way in which the body metabolizes the therapeutic agent.

Accordingly, in one embodiment, the presence of a particular allele at a polymorphic site or
30 haplotype is indicative of a different, e.g. a different response rate, to a particular treatment modality. This means that a patient diagnosed with cancer (e.g. prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, colon cancer, melanoma), and carrying a certain allele at a polymorphic or haplotype of the present invention (e.g., the at-risk and protective alleles and/or haplotypes of the invention) would respond better to, or worse to, a specific therapeutic, drug and/or
35 other therapy used to treat the cancer. Therefore, the presence or absence of the marker allele or haplotype could aid in deciding what treatment should be used for a the patient. For example, for a newly diagnosed patient, the presence of a marker or haplotype of the present invention may be assessed (e.g., through testing DNA derived from a blood sample, as described herein). If the patient is

positive for a marker allele or haplotype at (that is, at least one specific allele of the marker, or haplotype, is present), then the physician recommends one particular therapy, while if the patient is negative for the at least one allele of a marker, or a haplotype, then a different course of therapy may be recommended (which may include recommending that no immediate therapy, other than serial
5 monitoring for progression of the disease, be performed). Thus, the patient's carrier status could be used to help determine whether a particular treatment modality should be administered. The value lies within the possibilities of being able to diagnose the disease at an early stage, to select the most appropriate treatment, and provide information to the clinician about prognosis/aggressiveness of the disease in order to be able to apply the most appropriate treatment.

10 The present invention also relates to methods of monitoring progress or effectiveness of a treatment for a particular cancer (e.g. prostate cancer (e.g., aggressive prostate cancer), breast cancer, lung cancer, colon cancer, melanoma). This can be done based on the genotype and/or haplotype status of the markers and haplotypes of the present invention, i.e., by assessing the absence or presence of at least one allele of at least one polymorphic marker as disclosed herein, or by monitoring
15 expression of genes that are associated with the variants (markers and haplotypes) of the present invention. The risk gene mRNA or the encoded polypeptide can be measured in a tissue sample (e.g., a peripheral blood sample, or a biopsy sample). Expression levels and/or mRNA levels can thus be determined before and during treatment to monitor its effectiveness. Alternatively, or concomitantly, the genotype and/or haplotype status of at least one risk variant for the cancer as presented herein is
20 determined before and during treatment to monitor its effectiveness.

Alternatively, biological networks or metabolic pathways related to the markers and haplotypes of the present invention can be monitored by determining mRNA and/or polypeptide levels. This can be done for example, by monitoring expression levels or polypeptides for several genes belonging to the network and/or pathway, in samples taken before and during treatment. Alternatively, metabolites
25 belonging to the biological network or metabolic pathway can be determined before and during treatment. Effectiveness of the treatment is determined by comparing observed changes in expression levels/metabolite levels during treatment to corresponding data from healthy subjects.

In a further aspect, the markers of the present invention can be used to increase power and effectiveness of clinical trials. Thus, individuals who are carriers of at least one at-risk variant of the
30 present invention, i.e. individuals who are carriers of at least one allele of at least one polymorphic marker or haplotype conferring increased risk of developing cancer may be more likely to respond to a particular treatment modality. In one embodiment, individuals who carry at-risk variants for gene(s) in a pathway and/or metabolic network for which a particular treatment (e.g., small molecule drug) is targeting, are more likely to be responders to the treatment. In another embodiment, individuals who
35 carry at-risk variants for a gene, which expression and/or function is altered by the at-risk variant, are more likely to be responders to a treatment modality targeting that gene, its expression or its gene product. This application can improve the safety of clinical trials, but can also enhance the chance that a clinical trial will demonstrate statistically significant efficacy, which may be limited to a certain sub-group of the population. Thus, one possible outcome of such a trial is that carriers of certain genetic
40 variants, e.g., the markers and haplotypes of the present invention, are statistically significantly likely to

show positive response to the therapeutic agent, i.e. experience alleviation of symptoms associated with cancer when taking the therapeutic agent or drug as prescribed.

In a further aspect, the markers and haplotypes of the present invention can be used for targeting the selection of pharmaceutical agents for specific individuals. Personalized selection of treatment modalities, lifestyle changes or combination of the two, can be realized by the utilization of the at-risk variants of the present invention. Thus, the knowledge of an individual's status for particular markers of the present invention, can be useful for selection of treatment options that target genes or gene products affected by the at-risk variants of the invention. Certain combinations of variants may be suitable for one selection of treatment options, while other gene variant combinations may target other treatment options. Such combination of variant may include one variant, two variants, three variants, or four or more variants, as needed to determine with clinically reliable accuracy the selection of treatment module.

In addition to the diagnostic and therapeutic uses of the variants of the present invention, the variants (markers and haplotypes) can also be useful markers for human identification, and as such be useful in forensics, paternity testing and in biometrics. The specific use of SNPs for forensic purposes is reviewed by Gill (*Int. J. Legal Med.* 114:204-10 (2001)). Genetic variations in genomic DNA between individuals can be used as genetic markers to identify individuals and to associate a biological sample with an individual. Genetic markers, including SNPs and microsatellites, can be useful to distinguish individuals. The more markers that are analyzed, the lower the probability that the allelic combination of the markers in any given individual is the same as in an unrelated individual (assuming that the markers are unrelated, i.e. that the markers are in perfect linkage equilibrium). Thus, the variants used for these purposes are preferably unrelated, i.e. they are inherited independently. Thus, preferred markers can be selected from available markers, such as the markers of the present invention, and the selected markers may comprise markers from different regions in the human genome, including markers on different chromosomes.

In certain applications, the SNPs useful for forensic testing are from degenerate codon positions (i.e., the third position in certain codons such that the variation of the SNP does not affect the amino acid encoded by the codon). In other applications, such for applications for predicting phenotypic characteristics including race, ancestry or physical characteristics, it may be more useful and desirable to utilize SNPs that affect the amino acid sequence of the encoded protein. In other such embodiments, the variant (SNP or other polymorphic marker) affects the expression level of a nearby gene, thus leading to altered protein expression.

Nucleic acids and polypeptides

The nucleic acids and polypeptides described herein can be used in methods and kits of the present invention, as described in the above.

An "isolated" nucleic acid molecule, as used herein, is one that is separated from nucleic acids that normally flank the gene or nucleotide sequence (as in genomic sequences) and/or has been completely or partially purified from other transcribed sequences (e.g., as in an RNA library). For example, an isolated nucleic acid of the invention can be substantially isolated with respect to the complex cellular milieu in which it naturally occurs, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material can be purified to essential homogeneity, for example as determined by polyacrylamide gel electrophoresis (PAGE) or column chromatography (e.g., HPLC). An isolated nucleic acid molecule of the invention can comprise at least about 50%, at least about 80% or at least about 90% (on a molar basis) of all macromolecular species present. With regard to genomic DNA, the term "isolated" also can refer to nucleic acid molecules that are separated from the chromosome with which the genomic DNA is naturally associated. For example, the isolated nucleic acid molecule can contain less than about 250 kb, 200 kb, 150 kb, 100 kb, 75 kb, 50 kb, 25 kb, 10 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of the nucleotides that flank the nucleic acid molecule in the genomic DNA of the cell from which the nucleic acid molecule is derived.

The nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated. Thus, recombinant DNA contained in a vector is included in the definition of "isolated" as used herein. Also, isolated nucleic acid molecules include recombinant DNA molecules in heterologous host cells or heterologous organisms, as well as partially or substantially purified DNA molecules in solution. "Isolated" nucleic acid molecules also encompass *in vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention. An isolated nucleic acid molecule or nucleotide sequence can include a nucleic acid molecule or nucleotide sequence that is synthesized chemically or by recombinant means. Such isolated nucleotide sequences are useful, for example, in the manufacture of the encoded polypeptide, as probes for isolating homologous sequences (e.g., from other mammalian species), for gene mapping (e.g., by *in situ* hybridization with chromosomes), or for detecting expression of the gene in tissue (e.g., human tissue), such as by Northern blot analysis or other hybridization techniques.

The invention also pertains to nucleic acid molecules that hybridize under high stringency hybridization conditions, such as for selective hybridization, to a nucleotide sequence described herein (e.g., nucleic acid molecules that specifically hybridize to a nucleotide sequence containing a polymorphic site associated with a marker or haplotype described herein). Such nucleic acid molecules can be detected and/or isolated by allele- or sequence-specific hybridization (e.g., under high stringency conditions). Stringency conditions and methods for nucleic acid hybridizations are well known to the skilled person (see, e.g., *Current Protocols in Molecular Biology*, Ausubel, F. *et al*, John Wiley & Sons, (1998), and Kraus, M. and Aaronson, S., *Methods Enzymol.*, 200:546-556 (1991), the entire teachings of which are incorporated by reference herein.

The percent identity of two nucleotide or amino acid sequences can be determined by aligning the sequences for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first

sequence). The nucleotides or amino acids at corresponding positions are then compared, and the percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions x 100). In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, at least 40%,
5 at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95%, of the length of the reference sequence. The actual comparison of the two sequences can be accomplished by well-known methods, for example, using a mathematical algorithm. A non-limiting example of such a mathematical algorithm is described in Karlin, S. and Altschul, S., *Proc. Natl. Acad. Sci. USA*, 90:5873-5877 (1993). Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0), as described
10 in Altschul, S. *et al.*, *Nucleic Acids Res.*, 25:3389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., NBLAST) can be used. See the website on the world wide web at ncbi.nlm.nih.gov. In one embodiment, parameters for sequence comparison can be set at score=100, wordlength=12, or can be varied (e.g., W=5 or W=20).

Other examples include the algorithm of Myers and Miller, CABIOS (1989), ADVANCE and
15 ADAM as described in Torellis, A. and Robotti, C., *Comput. Appl. Biosci.* 10:3-5 (1994); and FASTA described in Pearson, W. and Lipman, D., *Proc. Natl. Acad. Sci. USA*, 85:2444-48 (1988).

In another embodiment, the percent identity between two amino acid sequences can be accomplished using the GAP program in the GCG software package (Accelrys, Cambridge, UK).

The present invention also provides isolated nucleic acid molecules that contain a fragment or
20 portion that hybridizes under highly stringent conditions to a nucleic acid that comprises, or consists of, the nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:2, or a nucleotide sequence comprising, or consisting of, the complement of the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:2, wherein the nucleotide sequence comprises at least one polymorphic allele contained in the markers and haplotypes described herein. The nucleic acid fragments of the invention are at least
25 about 15, at least about 18, 20, 23 or 25 nucleotides, and can be 30, 40, 50, 100, 200, 500, 1000, 10,000 or more nucleotides in length.

The nucleic acid fragments of the invention are used as probes or primers in assays such as those described herein. "Probes" or "primers" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of a nucleic acid molecule. In addition to DNA and RNA, such
30 probes and primers include polypeptide nucleic acids (PNA), as described in Nielsen, P. *et al.*, *Science* 254:1497-1500 (1991). A probe or primer comprises a region of nucleotide sequence that hybridizes to at least about 15, typically about 20-25, and in certain embodiments about 40, 50 or 75, consecutive nucleotides of a nucleic acid molecule. In one embodiment, the probe or primer comprises at least one allele of at least one polymorphic marker or at least one haplotype described herein, or the complement
35 thereof. In particular embodiments, a probe or primer can comprise 100 or fewer nucleotides; for example, in certain embodiments from 6 to 50 nucleotides, or, for example, from 12 to 30 nucleotides. In other embodiments, the probe or primer is at least 70% identical, at least 80% identical, at least 85% identical, at least 90% identical, or at least 95% identical, to the contiguous nucleotide sequence or to the complement of the contiguous nucleotide sequence. In another embodiment, the probe or primer is
40 capable of selectively hybridizing to the contiguous nucleotide sequence or to the complement of the

contiguous nucleotide sequence. Often, the probe or primer further comprises a label, e.g., a radioisotope, a fluorescent label, an enzyme label, an enzyme co-factor label, a magnetic label, a spin label, an epitope label.

The nucleic acid molecules of the invention, such as those described above, can be identified
5 and isolated using standard molecular biology techniques well known to the skilled person. The amplified DNA can be labeled (e.g., radiolabeled) and used as a probe for screening a cDNA library derived from human cells. The cDNA can be derived from mRNA and contained in a suitable vector. Corresponding clones can be isolated, DNA can be obtained following *in vivo* excision, and the cloned insert can be sequenced in either or both orientations by art-recognized methods to identify the correct
10 reading frame encoding a polypeptide of the appropriate molecular weight. Using these or similar methods, the polypeptide and the DNA encoding the polypeptide can be isolated, sequenced and further characterized.

In general, the isolated nucleic acid sequences of the invention can be used as molecular weight markers on Southern gels, and as chromosome markers that are labeled to map related gene
15 positions. The nucleic acid sequences can also be used to compare with endogenous DNA sequences in patients to identify cancer (e.g., prostate cancer) or a susceptibility to cancer (e.g., prostate cancer), and as probes, such as to hybridize and discover related DNA sequences or to subtract out known sequences from a sample (e.g., subtractive hybridization). The nucleic acid sequences can further be used to derive primers for genetic fingerprinting, to raise anti-polypeptide antibodies using immunization
20 techniques, and/or as an antigen to raise anti-DNA antibodies or elicit immune responses.

Antibodies

Polyclonal antibodies and/or monoclonal antibodies that specifically bind one form of a gene product but not to the other form of the gene product are also provided. Antibodies are also provided
25 which bind a portion of either the variant or the reference gene product that contains the polymorphic site or sites. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain antigen-binding sites that specifically bind an antigen. A molecule that specifically binds to a polypeptide of the invention is a molecule that binds to that polypeptide or a fragment thereof, but does not substantially
30 bind other molecules in a sample, e.g., a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind to a polypeptide of the invention. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population
35 of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of a polypeptide of the invention. A monoclonal antibody composition thus typically displays a single binding affinity for a particular polypeptide of the invention with which it immunoreacts.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a desired immunogen, e.g., polypeptide of the invention or a fragment thereof. The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules
5 directed against the polypeptide can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein, *Nature* 256:495-497
10 (1975), the human B cell hybridoma technique (Kozbor *et al.*, *Immunol. Today* 4: 72 (1983)), the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, Inc., pp. 77-96) or trioma techniques. The technology for producing hybridomas is well known (see generally *Current Protocols in Immunology* (1994) Coligan *et al.*, (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a
15 mammal immunized with an immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds a polypeptide of the invention.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating a monoclonal antibody to a polypeptide of the invention
20 (see, e.g., *Current Protocols in Immunology*, *supra*; Galfre *et al.*, *Nature* 266:55052 (1977); R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); and Lerner, *Yale J. Biol. Med.* 54:387-402 (1981)). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods that also would be useful.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody to a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with the polypeptide to thereby isolate immunoglobulin library members that bind the polypeptide. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia *Recombinant Phage Antibody System*,
30 Catalog No. 27-9400-01; and the Stratagene *SurfZAP™* Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO
35 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs *et al.*, *Bio/Technology* 9: 1370-1372 (1991); Hay *et al.*, *Hum. Antibod. Hybridomas* 3:81-85 (1992); Huse *et al.*, *Science* 246: 1275-1281 (1989); and Griffiths *et al.*, *EMBO J.* 12:725-734 (1993).

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA

techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art.

In general, antibodies of the invention (e.g., a monoclonal antibody) can be used to isolate a polypeptide of the invention by standard techniques, such as affinity chromatography or immunoprecipitation. A polypeptide-specific antibody can facilitate the purification of natural polypeptide from cells and of recombinantly produced polypeptide expressed in host cells. Moreover, an antibody specific for a polypeptide of the invention can be used to detect the polypeptide (e.g., in a cellular lysate, cell supernatant, or tissue sample) in order to evaluate the abundance and pattern of expression of the polypeptide. Antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. The antibody can be coupled to a detectable substance to facilitate its detection. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibodies may also be useful in pharmacogenomic analysis. In such embodiments, antibodies against variant proteins encoded by nucleic acids according to the invention, such as variant proteins that are encoded by nucleic acids that contain at least one polymorphic marker of the invention, can be used to identify individuals that require modified treatment modalities.

Antibodies can furthermore be useful for assessing expression of variant proteins in disease states, such as in active stages of a cancer (e.g., prostate cancer), or in an individual with a predisposition to cancer related to the function of the protein, in particular prostate cancer. Antibodies specific for a variant protein of the present invention that is encoded by a nucleic acid that comprises at least one polymorphic marker or haplotype as described herein can be used to screen for the presence of the variant protein, for example to screen for a predisposition to cancer (e.g., prostate cancer) as indicated by the presence of the variant protein.

Antibodies can be used in other methods. Thus, antibodies are useful as diagnostic tools for evaluating proteins, such as variant proteins of the invention, in conjunction with analysis by electrophoretic mobility, isoelectric point, tryptic or other protease digest, or for use in other physical assays known to those skilled in the art. Antibodies may also be used in tissue typing. In one such embodiment, a specific variant protein has been correlated with expression in a specific tissue type, and antibodies specific for the variant protein can then be used to identify the specific tissue type.

Subcellular localization of proteins, including variant proteins, can also be determined using antibodies, and can be applied to assess aberrant subcellular localization of the protein in cells in various tissues. Such use can be applied in genetic testing, but also in monitoring a particular treatment

modality. In the case where treatment is aimed at correcting the expression level or presence of the variant protein or aberrant tissue distribution or developmental expression of the variant protein, antibodies specific for the variant protein or fragments thereof can be used to monitor therapeutic efficacy.

5 Antibodies are further useful for inhibiting variant protein function, for example by blocking the binding of a variant protein to a binding molecule or partner. Such uses can also be applied in a therapeutic context in which treatment involves inhibiting a variant protein's function. An antibody can be for example be used to block or competitively inhibit binding, thereby modulating (i.e., agonizing or antagonizing) the activity of the protein. Antibodies can be prepared against specific protein fragments
10 containing sites required for specific function or against an intact protein that is associated with a cell or cell membrane. For administration *in vivo*, an antibody may be linked with an additional therapeutic payload, such as radionuclide, an enzyme, an immunogenic epitope, or a cytotoxic agent, including bacterial toxins (diphtheria or plant toxins, such as ricin). The *in vivo* half-life of an antibody or a fragment thereof may be increased by pegylation through conjugation to polyethylene glycol.

15 The present invention further relates to kits for using antibodies in the methods described herein. This includes, but is not limited to, kits for detecting the presence of a variant protein in a test sample. One preferred embodiment comprises antibodies such as a labelled or labelable antibody and a compound or agent for detecting variant proteins in a biological sample, means for determining the amount or the presence and/or absence of variant protein in the sample, and means for comparing the
20 amount of variant protein in the sample with a standard, as well as instructions for use of the kit.

The present invention is now illustrated by the following Examples, which are not intended to be limiting in any way.

25 **EXAMPLE 1. IDENTIFICATION OF THE LD BLOCK C REGION ASSOCIATED WITH PROSTATE CANCER**

Patients involved in the genetics study

A population based list of all prostate cancer patients that were diagnosed with prostate cancer in Iceland from 1955 to 2005 form the basis for this study. Patients have been invited to join the study since 2001 on an ongoing basis. As of October 2006, blood samples from 1564 prostate cancer patients
30 have been collected. Of that collection 1455 prostate cancer patients in addition to 4182 control individuals were genotyped with the Illumina 317K Bead chip.

Statistical Methods for Association and Haplotype Analysis

For single marker association to the disease, Fisher exact test was used to calculate a two-sided P-value for each individual allele. When presenting the results, we used allelic frequencies rather
35 than carrier frequencies for SNPs and haplotypes. Haplotype analyses were performed using a

computer program we developed at deCODE called NEMO (NEsted MOdels) (Gretarsdóttir, *et al.*, *Nat Genet.* 2003 Oct;35(2):131-8). NEMO was used both to study marker-marker association and to calculate linkage disequilibrium (LD) between markers, and for case-control haplotype analysis. With NEMO, haplotype frequencies are estimated by maximum likelihood and the differences between patients and controls are tested using a generalized likelihood ratio test. The maximum likelihood estimates, likelihood ratios and P-values are computed with the aid of the EM-algorithm directly for the observed data, and hence the loss of information due to the uncertainty with phase and missing genotypes is automatically captured by the likelihood ratios, and under most situations, large sample theory can be used to reliably determine statistical significance. The relative risk (RR) of an allele or a haplotype, *i.e.*, the risk of an allele compared to all other alleles of the same marker, is calculated assuming the multiplicative model (Terwilliger, J.D. & Ott, J. A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* 42, 337-46 (1992) and Falk, C.T. & Rubinstein, P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann. Hum. Genet.* 51 (Pt 3), 227-33 (1987)), together with the population attributable risk (PAR).

In the haplotype analysis, it may be useful to group haplotypes together and test the group as a whole for association to the disease. This is possible to do with NEMO. A model is defined by a partition of the set of all possible haplotypes, where haplotypes in the same group are assumed to confer the same risk while haplotypes in different groups can confer different risks. A null hypothesis and an alternative hypothesis are said to be nested when the latter corresponds to a finer partition than the former. NEMO provides complete flexibility in the partition of the haplotype space. In this way, it is possible to test multiple haplotypes jointly for association and to test if different haplotypes confer different risk. As a measure of LD, we use two standard definitions of LD, D' and R^2 (Lewontin, R., *Genetics*, 49:49-67 (1964) and Hill, W.G. and A. Robertson, *Theor. Appl. Genet.*, 22:226-231 (1968)) as they provide complementary information on the amount of LD. For the purpose of estimating D' and R^2 , the frequencies of all two-marker allele combinations are estimated using maximum likelihood methods and the deviation from linkage disequilibrium is evaluated using a likelihood ratio test. The standard definitions of D' and R^2 are extended to include microsatellites by averaging over the values for all possible allele combinations of the two markers weighted by the marginal allele probabilities.

The number of possible haplotypes that can be constructed out of the dense set of markers genotyped over the whole genome is very large and even though the number of haplotypes that are actually observed in the patient and control cohort is much smaller, testing all of those haplotypes for association to the disease is a formidable task. It should be noted that we do not restrict our analysis to haplotypes constructed from a set of consecutive markers, as some markers may be very mutable and might split up an otherwise well conserved haplotype constructed out of surrounding markers

Results

As described herein, a region on chromosome 8q24.21 (LD block C) was identified that confers an increased risk for particular cancers (e.g., prostate cancer (e.g., aggressive prostate cancer). Particular haplotypes and markers associated with an increased risk of prostate cancer are depicted in

Table 1. As indicated in Table 1, the haplotype (HapC) involves the following markers (e.g., SNPs) and alleles: rs1456314 3 allele, rs17831626 4 allele, rs7825414 3 allele, rs6993569 3 allele, rs6994316 1 allele, rs6470494 4 allele, rs1016342 2 allele, rs1031588 3 allele, rs1016343 4 allele, rs1551510 3 allele, rs1456306 2 allele, rs1378897 3 allele, rs1456305 4 allele, rs7816535 3 allele. By examining results for the Caucasian (CEU) samples in the HapMap project allele 1 of the SNP marker rs16901979 was shown to be strongly correlated with HapC ($D' = 1$, $r^2 = 1$). The genotyping of rs16901979 in the Icelandic samples confirmed its correlation with HapC ($D' = 0.98$, $r^2 = 0.70$ in controls). The markers and haplotypes are located in what we call LD block C between positions 128,032,278 and 128,094,256 bp (NCBI Build 34) and positions of the individual markers are indicated in Table 2.). Allele codes for SNPs are as follows: 1= A, 2=C, 3=G, 4=T. Note that an increased risk is seen in aggressive prostate cancer as compared to all prostate cancer. The aggressive phenotype is determined by prostate cancer with combined Gleason grades of 7 or higher AND/OR stage T3 or higher AND/OR node positive AND/OR metastasis positive disease AND/OR death because of prostate cancer. Note that it is sufficient to have one of these criteria to be determined aggressive prostate cancer. These clinical parameters are well known surrogates for increased aggressiveness of the disease.

Table 1. Association of SNP markers on Chr8q24 to prostate cancer (PrCa) in Iceland

Affected Phenotype	# of Individuals		p-value	RR	Allele frequency		Allele or haplotype	PAR
	Affected	Controls			Affected	Controls		
PrCa	1455	4182	7.86E-09	1.72	0.072	0.043	1 rs16901979	0.059
PrCa	1455	4182	3.89E-12	2.05	0.063	0.032	HapC	0.064
Aggressive PrCa	745	4182	4.82E-07	1.82	0.076	0.043	1 rs16901979	0.067
Aggressive PrCa	745	4182	3.69E-10	2.26	0.069	0.032	HapC	0.076

Alleles for the marker rs16901979 and SNPs in same LD block at 8q24.21 are shown and the corresponding numbers of cases and controls (N), allelic frequencies of variants in affected and control individuals, the odds-ratio (OR) and two-sided P values. Affected individuals are those diagnosed with prostate cancer (ICD10 = C61); or those diagnosed with prostate cancer (ICD10: C61) and have aggressive cancer as determined by a combined phenotype including cancers with Gleason combined grade of 7 or higher or cancers with stage T3 or higher or node positive disease or metastasis positive disease or men that have died because of prostate cancer. Controls are unaffected population based controls. Allele codes for SNPs are as follows: 1=A, 2=C, 3=G, 4=T. PAR is population attributable risk that can be explained by these variants.

Results for allele 1 of rs16901979 are generated by using genotypes for rs16901979 as well as other correlated SNPs in HapC to increase genotype yield.

HapC: 3 rs1456314 4 rs17831626 3 rs7825414 3 rs6993569 1 rs6994316 4 rs6470494 2 rs1016342 3 rs1031588 4 rs1016343 3 rs1551510 2 rs1456306 3 rs1378897 4 rs1456305 3 rs7816535.

As shown in table 1, both marker rs16901979 1 allele and the 14 marker HapC haplotype give significant association to prostate cancer. The population frequency of the haplotype is slightly lower than that of the rs16901979 1 allele, and the corresponding relative risk somewhat higher. The population attributable risk of these variants in Iceland is 5.9-6.4% in all prostate cancer and slightly higher or 6.7-7.6% in the aggressive prostate cancer patients.

The opposite allele of SNP marker rs16901979 or allele 2 shows significant protection to prostate cancer in carriers. These results are shown in Table 2.

Table 2. Variants on Chr8q24 that confer a protection to prostate cancer in Iceland

Affected Phenotype	# of Individuals		p-value	RR	Allele frequency		Allele
	Affected	Controls			Affected	Controls	
Prostate Cancer	1446	4182	1.63E-08	0.59	0.928	0.957	2 rs16901979
Aggressive PrCa	745	4182	4.82E-07	0.55	0.924	0.957	2 rs16901979

Alleles for the marker rs16901979 at 8q24.21 are shown and the corresponding numbers of cases and controls (N), allelic frequencies of variants in affected and control individuals, the odds-ratio (OR) and two-sided P values. Affected individuals are those diagnosed with prostate cancer (ICD10 = C61); or those diagnosed with prostate cancer (ICD10: C61) and have aggressive cancer as determined by a combined phenotype including cancers with Gleason combined grade of 7 or higher or with stage T3 or higher or node positive disease or metastasis positive disease or have died because of prostate cancer. Controls are unaffected population based controls. Allele codes for SNPs are as follows: 1=A, 2=C, 3=G, 4=T

Table 3 lists SNP markers in HapC in the LD Block C region. Location of the SNPs is given with respect to NCBI Build34 of the human genome assembly. The relative position of the markers in
5 basepair position is indicated.

Table 3: Genomic location of SNPs in HapC and marker rs16901979

rs name	BUILD34
rs1456314	128032278
rs17831626	128037011
rs7825414	128038109
rs6993569	128040685
rs6994316	128041127
rs6470494	128044492
rs1016342	128049043
rs1031588	128049865
rs1016343	128049885
rs1551510	128050067
rs1456306	128073088
rs1378897	128079247
rs16901979	128081505
rs1456305	128083840
rs7816535	128094256

SNP markers in HapC and
rs16901979 with genomic locations
(base pairs) in NCBI Build 34.

To find this association of LD block C, as manifested by the marker and haplotypes tested and indicated in Table 1, we performed a genome-wide scan using Infinium HumanHap300 SNP chips from
10 Illumina for assaying approximately 317,000 single nucleotide polymorphisms (SNPs) on a single chip. We genotyped 1455 patients and 4182 controls in this manner. The controls represent an unaffected Icelandic population sample. Association analysis was done with single markers, two markers and haplotypes representing all consecutive markers within each and every LD block in the human genome. The SNP marker rs16901979 1 allele was found to give highly significant association to prostate cancer.

The marker rs16901979 is not on the Illumina Hap300 chip and was genotyped independently. Two marker haplotypes of markers within LD block C also gave highly significant association to prostate cancer. When haplotypes consisting of all consecutive SNP markers within each LD block throughout the whole genome were tested (LD block haplotypes) for their association to prostate cancer, the most significant haplotype was on Chr8q24 at location 128.414-128.506 Mb (NCBI Build34). This finding is already described by us (Amundadottir et.al. Nat Genet. 2006 Jun;38(6):652-8.). Much to our surprise the second best LD block haplotype was located only a few hundred thousand basepairs away on Chr8q24 or at 128.032-128.095 Mb (NCBI Build34). We can call this haplotype and LD block haplotype C (HapC) and LD block C. Due to the LD block (haplotype block) structure of the human genome, we proceeded to investigate the LD block C, within which rs16901979 resides.

There are 30 SNPs that are in strong LD with rs16901979 and would therefore be just as good in detecting the association to prostate cancer described herein. The LD between these SNPs and rs16901979 as measured by r^2 is greater than 0.5. These markers are listed in Table 4A. Table 4B lists markers in LD with rs16901979 in the different HapMap populations with values of r^2 to rs 16901979 of at least 0.2 in that population.

Table 4A: SNPs that are correlated with rs16901979

rs name	BUILD34
rs10453084	128069367
rs10505483	128081783
rs1551512	128080714
rs16901948	128063689
rs16901949	128063741
rs16901950	128063831
rs16901952	128063858
rs16901953	128064817
rs16901959	128066118
rs16901966	128066840
rs16901967	128066865
rs16901969	128068685
rs16901970	128069303
rs16901984	128087549
rs6470498	128072308
rs6983561	128063468
rs6987640	128069616
rs6987723	128069447
rs6988257	128068044
rs6989838	128085960
rs6990420	128065313
rs7001069	128067234
rs7010450	128065267
rs7013255	128087075
rs7817677	128082092
rs7824451	128071049
rs7824785	128071298
rs7826388	128066364

rs7830341 128066518
rs7844219 128075403

All SNPs are listed with a $r^2 > 0.5$ with rs16901979.
SNP names are listed with genomic locations
(base pair start) in NCBI Build 34.

Table 4B. Markers in strong linkage disequilibrium to marker rs16901979. Markers that are correlated to rs16901979 by values of r^2 of at least 0.2 in the Caucasian CEPH population CEU, African Yuroba population YRI, Japanese population JPT and Chinese population CHB are shown. Shown are values for the degree of LD of each marker with rs16901979 (D' , r^2 and p-value), as well as the location of the markers in NCBI Builds 34, 35, 36 and SEQ ID NO:2 (LD Block C'), as well a reference to their SEQ ID NO.

CEU HapMap								
Marker	D'	R^2	p	Position (B34)	Position B35	Position (B36)	Position SEQ ID NO:2	SEQ ID NO
rs12682344	1	1	1.18E-07	128063373	128175966	128175966	34261	43
rs6983561	1	1	1.18E-07	128063469	128176062	128176062	34357	21
rs16901948	1	1	1.18E-07	128063690	128176283	128176283	34578	9
rs16901949	1	1	1.18E-07	128063742	128176335	128176335	34630	10
rs16901950	1	1	1.18E-07	128063832	128176425	128176425	34720	11
rs16901952	1	1	1.24E-07	128063859	128176452	128176452	34747	12
rs16901953	1	1	1.18E-07	128064818	128177411	128177411	35706	13
rs7010450	1	1	1.18E-07	128065268	128177861	128177861	36156	28
rs6990420	1	0.74026	1.28E-06	128065314	128177907	128177907	36202	26
rs16901959	1	1	1.18E-07	128066119	128178712	128178712	37007	14
rs7826388	1	1	1.18E-07	128066365	128178958	128178958	37253	33
rs7830341	1	1	1.18E-07	128066519	128179112	128179112	37407	34
rs16901966	1	1	1.18E-07	128066841	128179434	128179434	37729	15
rs16901967	1	1	1.18E-07	128066866	128179459	128179459	37754	16
rs7001069	1	1	1.18E-07	128067235	128179828	128179828	38123	27
rs6988257	1	1	1.18E-07	128068045	128180638	128180638	38933	24
rs16901969	1	1	1.18E-07	128068686	128181279	128181279	39574	17
rs16901970	1	1	1.18E-07	128069304	128181897	128181897	40192	18
rs10453084	1	1	1.18E-07	128069368	128181961	128181961	40256	6
rs6987723	1	1	1.18E-07	128069448	128182041	128182041	40336	23
rs6987640	1	1	1.18E-07	128069617	128182210	128182210	40505	22
rs7824451	1	1	1.18E-07	128071050	128183643	128183643	41938	31
rs7824785	1	1	1.18E-07	128071299	128183892	128183892	42187	32
rs6470498	1	1	1.18E-07	128072309	128184902	128184902	43197	20
rs7844219	1	1	1.18E-07	128075404	128187997	128187997	46292	35
rs1551512	1	1	1.18E-07	128080715	128193308	128193308	51603	8
rs10505483	1	1	1.18E-07	128081784	128194377	128194377	52672	7
rs7817677	1	1	1.18E-07	128082093	128194686	128194686	52981	30
rs6989838	1	1	1.18E-07	128085961	128198554	128198554	56849	25
rs7013255	1	1	1.18E-07	128087076	128199669	128199669	57964	29
rs16901984	1	1	1.45E-07	128087550	128200143	128200143	58438	19
YRI HapMap								
M1	D'	r^2	p	Location (B34)	Location B35	Location (B36)	Position SEQ ID NO:2	SEQ ID NO
rs1456315	0.583779	0.237876	9.12E-07	128060526	128173119	128173119	31414	46
rs13254738	0.767679	0.260913	4.36E-08	128060932	128173525	128173525	31820	45
rs7006390	0.923431	0.652818	8.36E-21	128060956	128173549	128173549	31844	63
rs1073997	0.86296	0.720149	1.90E-22	128061776	128174369	128174369	32664	41
rs7012442	0.966424	0.933975	3.18E-32	128062320	128174913	128174913	33208	65
rs6983561	0.962864	0.782622	1.44E-25	128063469	128176062	128176062	34357	21
rs16901948	1	0.507692	7.86E-19	128063690	128176283	128176283	34578	9
rs16901949	1	0.507692	7.86E-19	128063742	128176335	128176335	34630	10
rs16901950	1	0.507692	7.86E-19	128063832	128176425	128176425	34720	11
rs16901952	1	0.507692	2.04E-18	128063859	128176452	128176452	34747	12
rs16901953	1	0.507692	7.86E-19	128064818	128177411	128177411	35706	13
rs7010450	1	0.507692	1.27E-18	128065268	128177861	128177861	36156	28
rs12544977	0.93185	0.31233	2.91E-10	128065876	128178469	128178469	36764	42
rs16901959	1	0.507692	7.86E-19	128066119	128178712	128178712	37007	14
rs7826337	0.95706	0.604515	1.40E-19	128066163	128178756	128178756	37051	69

rs7826388	1	0.294728	6.60E-12	128066365	128178958	128178958	37253	33
rs7830341	1	0.489879	2.94E-18	128066519	128179112	128179112	37407	34
rs16901966	1	0.507692	7.86E-19	128066841	128179434	128179434	37729	15
rs16901967	1	0.507692	7.86E-19	128066866	128179459	128179459	37754	16
rs7000910	0.954702	0.55904	4.30E-18	128067194	128179787	128179787	38082	62
rs7001069	1	0.507692	7.86E-19	128067235	128179828	128179828	38123	27
rs7006409	0.932011	0.31983	6.19E-10	128068018	128180611	128180611	38906	64
rs6988257	1	0.507692	7.86E-19	128068045	128180638	128180638	38933	24
rs16901969	1	0.507692	7.86E-19	128068686	128181279	128181279	39574	17
rs16901970	1	0.507692	7.86E-19	128069304	128181897	128181897	40192	18
rs10453084	1	0.507692	7.86E-19	128069368	128181961	128181961	40256	6
rs6987723	1	0.489879	2.94E-18	128069448	128182041	128182041	40336	23
rs6987640	1	0.507692	7.86E-19	128069617	128182210	128182210	40505	22
rs7824451	1	0.507692	7.86E-19	128071050	128183643	128183643	41938	31
rs7824785	1	0.507692	7.86E-19	128071299	128183892	128183892	42187	32
rs6470498	1	0.507692	7.86E-19	128072309	128184902	128184902	43197	20
rs7827234	1	0.610001	1.92E-21	128074695	128187288	128187288	45583	70
rs7844219	0.964897	0.930545	3.58E-30	128075404	128187997	128187997	46292	35
rs6470499	1	0.934871	5.99E-34	128079983	128192576	128192576	50871	56
rs1551512	1	0.934871	5.99E-34	128080715	128193308	128193308	51603	8
rs10505483	1	1	7.04E-38	128081784	128194377	128194377	52672	7
rs7817677	1	0.525988	2.02E-19	128082093	128194686	128194686	52981	30
rs6989838	1	1	7.04E-38	128085961	128198554	128198554	56849	25
rs7013255	1	1	7.04E-38	128087076	128199669	128199669	57964	29
rs16901984	1	0.525043	6.71E-18	128087550	128200143	128200143	58438	19
rs7000307	1	0.334526	3.19E-13	128089665	128202258	128202258	60553	61
rs7840773	1	0.334526	3.19E-13	128091923	128204516	128204516	62811	71
rs7824364	1	0.334526	3.19E-13	128091954	128204547	128204547	62842	68

JPT HapMap

M1	D'	r2	p	Location (B34)	Location B35	Location (B36)	Position SEQ ID NO:2	SEQ ID NO
rs16901932	0.871703	0.309945	0.000085	128028107	128140700	128140700		48
rs16901935	0.668518	0.254464	0.000841	128033134	128145727	128145727	4022	49
rs6470494	0.654239	0.236706	0.000813	128044493	128157086	128157086	15381	55
rs1016343	1	0.52	3.40E-11	128049886	128162479	128162479	20774	37
rs7841060	1	0.52	3.40E-11	128053066	128165659	128165659	23954	72
rs7006390	1	0.921053	4.71E-17	128060956	128173549	128173549	31844	63
rs1073997	1	1	7.61E-17	128061776	128174369	128174369	32664	41
rs7012442	1	1	1.14E-18	128062320	128174913	128174913	33208	65
rs12682344	1	1	1.14E-18	128063373	128175966	128175966	34261	43
rs6983561	1	1	9.10E-18	128063469	128176062	128176062	34357	21
rs16901948	1	1	1.14E-18	128063690	128176283	128176283	34578	9
rs16901949	1	1	1.14E-18	128063742	128176335	128176335	34630	10
rs16901950	1	1	1.14E-18	128063832	128176425	128176425	34720	11
rs16901952	1	1	1.14E-18	128063859	128176452	128176452	34747	12
rs16901953	1	1	1.14E-18	128064818	128177411	128177411	35706	13
rs7010450	1	1	1.14E-18	128065268	128177861	128177861	36156	28
rs6990420	1	0.736842	1.81E-14	128065314	128177907	128177907	36202	26
rs16901959	1	1	1.14E-18	128066119	128178712	128178712	37007	14
rs7826337	1	0.238919	3.93E-07	128066163	128178756	128178756	37051	69
rs7826388	1	1	1.14E-18	128066365	128178958	128178958	37253	33
rs7830341	1	1	9.10E-18	128066519	128179112	128179112	37407	34
rs16901966	1	1	1.14E-18	128066841	128179434	128179434	37729	15
rs16901967	1	1	1.14E-18	128066866	128179459	128179459	37754	16
rs7000910	1	0.238919	3.93E-07	128067194	128179787	128179787	38082	62
rs7001069	1	1	9.10E-18	128067235	128179828	128179828	38123	27
rs6988257	1	1	9.10E-18	128068045	128180638	128180638	38933	24
rs16901969	1	1	1.14E-18	128068686	128181279	128181279	39574	17
rs16901970	1	1	1.14E-18	128069304	128181897	128181897	40192	18
rs10453084	1	1	1.14E-18	128069368	128181961	128181961	40256	6
rs6987723	1	1	1.14E-18	128069448	128182041	128182041	40336	23
rs6987640	1	1	9.10E-18	128069617	128182210	128182210	40505	22
rs7824451	1	1	9.10E-18	128071050	128183643	128183643	41938	31
rs7824785	1	1	9.10E-18	128071299	128183892	128183892	42187	32
rs6470498	1	1	1.14E-18	128072309	128184902	128184902	43197	20
rs7827234	1	0.238919	3.93E-07	128074695	128187288	128187288	45583	70
rs7844219	1	1	1.14E-18	128075404	128187997	128187997	46292	35
rs6470499	1	1	9.10E-18	128079983	128192576	128192576	50871	56
rs1551512	1	1	9.10E-18	128080715	128193308	128193308	51603	8

rs10505483	1	1	1.26E-17	128081784	128194377	128194377	52672	7
rs7817677	1	1	1.14E-18	128082093	128194686	128194686	52981	30
rs6989838	1	1	9.10E-18	128085961	128198554	128198554	56849	25
rs7013255	1	1	9.10E-18	128087076	128199669	128199669	57964	29
rs16901984	1	1	8.21E-15	128087550	128200143	128200143	58438	19
rs16901997	1	0.387755	5.84E-07	128128090	128240683	128240683		50
rs283710	0.634312	0.209223	0.002854	128246180	128358773	128358773		52
rs283719	0.657568	0.254825	0.000865	128266367	128378960	128378960		53
rs7017081	0.576954	0.264459	0.003929	128275936	128388529	128388529		66
rs6984136	0.570098	0.235283	0.002359	128276727	128389320	128389320		58
rs1995613	0.570098	0.235283	0.002359	128278532	128391125	128391125		51

CHB HapMap

M1	D'	r2	p	Location (B34)	Location B35	Location (B36)	Position SEQ ID NO:2	SEQ ID NO
rs10216560	0.562661	0.293676	0.001256	127962278	128074871	128074871		38
rs6470494	0.649644	0.224207	0.00107	128044493	128157086	128157086	15381	55
rs1016342	1	0.246652	2.77E-08	128049044	128161637	128161637	19932	36
rs1031589	1	0.235197	5.23E-08	128049571	128162164	128162164	20459	40
rs1016343	0.77074	0.361993	9.71E-06	128049886	128162479	128162479	20774	37
rs1031587	1	0.235197	5.23E-08	128050061	128162654	128162654	20949	39
rs1551510	1	0.22807	1.07E-07	128050067	128162660	128162660	20955	47
rs4871008	1	0.224138	9.74E-08	128050130	128162723	128162723	21018	54
rs6981122	1	0.224138	9.74E-08	128051049	128163642	128163642	21937	57
rs13252298	1	0.206124	3.66E-07	128051745	128164338	128164338	22633	44
rs7841060	0.77074	0.361993	9.71E-06	128053066	128165659	128165659	23954	72
rs6997559	1	0.235197	5.23E-08	128054167	128166760	128166760	25055	60
rs7006390	1	0.895105	2.06E-21	128060956	128173549	128173549	31844	63
rs1073997	1	1	2.42E-25	128061776	128174369	128174369	32664	41
rs7012442	1	1	2.42E-25	128062320	128174913	128174913	33208	65
rs12682344	1	1	2.42E-25	128063373	128175966	128175966	34261	43
rs6983561	1	1	2.42E-25	128063469	128176062	128176062	34357	21
rs16901948	1	1	2.42E-25	128063690	128176283	128176283	34578	9
rs16901949	1	1	2.42E-25	128063742	128176335	128176335	34630	10
rs16901950	1	1	2.42E-25	128063832	128176425	128176425	34720	11
rs16901952	1	1	4.86E-25	128063859	128176452	128176452	34747	12
rs16901953	1	1	2.42E-25	128064818	128177411	128177411	35706	13
rs7010450	1	1	2.42E-25	128065268	128177861	128177861	36156	28
rs6990420	1	0.729647	5.85E-18	128065314	128177907	128177907	36202	26
rs12544977	1	0.325	4.30E-10	128065876	128178469	128178469	36764	42
rs16901959	1	1	2.42E-25	128066119	128178712	128178712	37007	14
rs7826337	1	0.40625	7.40E-12	128066163	128178756	128178756	37051	69
rs7826388	1	1	2.42E-25	128066365	128178958	128178958	37253	33
rs7830341	1	1	2.42E-25	128066519	128179112	128179112	37407	34
rs16901966	1	1	2.42E-25	128066841	128179434	128179434	37729	15
rs16901967	1	1	2.42E-25	128066866	128179459	128179459	37754	16
rs7000910	1	0.40625	7.40E-12	128067194	128179787	128179787	38082	62
rs7001069	1	1	2.42E-25	128067235	128179828	128179828	38123	27
rs7006409	1	0.3125	1.46E-09	128068018	128180611	128180611	38906	64
rs6988257	1	1	2.42E-25	128068045	128180638	128180638	38933	24
rs16901969	1	1	2.42E-25	128068686	128181279	128181279	39574	17
rs16901970	1	1	2.42E-25	128069304	128181897	128181897	40192	18
rs10453084	1	1	2.42E-25	128069368	128181961	128181961	40256	6
rs6987723	1	1	2.42E-25	128069448	128182041	128182041	40336	23
rs6987640	1	1	2.42E-25	128069617	128182210	128182210	40505	22
rs7824451	1	1	2.42E-25	128071050	128183643	128183643	41938	31
rs7824785	1	1	2.42E-25	128071299	128183892	128183892	42187	32
rs6470498	1	1	2.42E-25	128072309	128184902	128184902	43197	20
rs7827234	1	0.40625	7.40E-12	128074695	128187288	128187288	45583	70
rs7844219	1	1	2.42E-25	128075404	128187997	128187997	46292	35
rs6470499	1	1	2.42E-25	128079983	128192576	128192576	50871	56
rs1551512	1	1	2.42E-25	128080715	128193308	128193308	51603	8
rs10505483	1	1	2.42E-25	128081784	128194377	128194377	52672	7
rs7817677	1	1	2.42E-25	128082093	128194686	128194686	52981	30
rs6989838	1	1	2.42E-25	128085961	128198554	128198554	56849	25
rs7013255	1	1	2.42E-25	128087076	128199669	128199669	57964	29
rs16901984	1	1	2.38E-24	128087550	128200143	128200143	58438	19
rs7816535	1	0.283608	3.74E-09	128094257	128206850	128206850	65145	67
rs6990483	0.510459	0.260568	0.000619	128655867	128768460	128768460		59

The LD structure in the LD Block C area of the markers and haplotypes that associates with prostate cancer is shown in FIG. 1. The structure was derived from HAPMAP data release 19. This LD block (LD block C) is located between markers rs1456314, located at 128,032,278 bp, and rs7816535, located at 128,094,256 bp (NCBI build 34), and is almost 65 kb in length. The LD structure is seen as a block of DNA that has a high r^2 and $|D'|$ between markers as indicated in the figure. Markers are represented equidistantly, i.e. with the same distance between any two markers.

It is possible that other markers residing within the LD block C region also are associated with prostate cancer, since such markers are in linkage disequilibrium with rs16901979. Table 5A provides a list of publicly known SNP markers in the LD Block C spanned by LD block C (positions 128032278 - 128094256 bp in NCBI Build 34). The Table 5B provides a list of all public SNP markers in LD block C' (corresponding to the sequence set forth in SEQ ID NO:2), and Table 5C provides a list of all microsatellite markers within that region.

15

Table 5A: All known SNPs within LD block C (SEQ ID NO:1)

rs name	BUILD34
rs1456314	128032278
rs1551516	128032477
rs3940781	128032803
rs16901935	128033133
rs12542102	128034109
rs11987219	128035269
rs11988135	128035997
rs2124598	128036063
rs2392726	128036067
rs2392727	128036184
rs2392728	128036231
rs2392729	128036254
rs1014656	128036271
rs2392730	128036282
rs2392731	128036328
rs1014655	128036427
rs2392732	128036522
rs2893603	128036545
rs9656965	128036836
rs17831626	128037011
rs16901938	128037088
rs28451013	128037234
rs28385355	128037374
rs11785961	128037423
rs13267056	128037675
rs7824674	128037681
rs7824679	128037698
rs7824923	128037726
rs7843300	128037819
rs7824957	128037848
rs7825394	128038037

rs10308400	128038109
rs7825414	128038109
rs11994932	128038486
rs13248907	128038933
rs13282331	128038957
rs13282364	128038997
rs12548394	128039290
rs10089648	128039437
rs13363429	128039437
rs11985630	128040216
rs6993569	128040685
rs6994316	128041127
rs12550334	128041512
rs11998124	128042068
rs6999589	128042234
rs11998248	128042343
rs7005114	128043346
rs11319438	128043465
rs6991325	128043514
rs1902431	128043664
rs6470494	128044492
rs7843737	128044654
rs10095770	128045158
rs6987760	128046899
rs10306042	128047254
rs6470495	128047254
rs4581008	128047453
rs7461435	128047550
rs4620246	128047550
rs12544839	128047970
rs12544844	128048146
rs10956351	128048147
rs1016342	128049043
rs1551511	128049402
rs1031589	128049570
rs1378896	128049642
rs1031588	128049865
rs1016343	128049885
rs1031587	128050060
rs1551510	128050066
rs4871008	128050129
rs11363193	128050445
rs6981122	128051048
rs7001504	128051071
rs13252298	128051744
rs7841060	128053065
rs4540377	128053502
rs6997559	128054166
rs11554368	128055276
rs7007694	128055754
rs17831913	128055754
rs13257371	128055830
rs4571699	128055857
rs1840709	128056043

rs3037753	128056058
rs7837848	128056120
rs3837167	128056513
rs3837166	128056618
rs11993508	128056664
rs17832021	128056764
rs3999809	128056896
rs3999810	128056955
rs28663168	128056997
rs3999811	128057050
rs3999812	128057095
rs3999806	128057100
rs3857883	128057194
rs2392733	128057260
rs1456317	128057294
rs1456316	128057436
rs16901946	128057513
rs28735726	128057517
rs9656813	128057552
rs9656814	128057565
rs9656815	128057591
rs10505484	128057900
rs11994653	128057931
rs6470496	128058617
rs6470497	128058918
rs7844454	128059089
rs7826322	128059141
rs12677860	128059213
rs12677861	128059215
rs7000321	128059607
rs12682421	128059739
rs7000967	128059948
rs7001895	128060267
rs1456315	128060525
rs5013678	128060567
rs7463708	128060643
rs13254738	128060931
rs7006390	128060955
rs1073997	128061775
rs7012407	128062181
rs7012442	128062319
rs7016828	128062820
rs7016830	128062823
rs12682344	128063372
rs6983561	128063468
rs16901948	128063689
rs1869931	128063724
rs16901949	128063741
rs16901950	128063831
rs16901952	128063858
rs7005144	128064232
rs16901953	128064817
rs4871009	128065004
rs7010450	128065267

rs6990420	128065313
rs6994359	128065511
rs17832285	128065581
rs7825340	128065717
rs12544977	128065875
rs16901959	128066118
rs7826337	128066162
rs7826388	128066364
rs7830341	128066518
rs7830381	128066713
rs16901966	128066840
rs16901967	128066865
rs7000910	128067193
rs7001069	128067234
rs17765137	128067402
rs11781162	128067484
rs11774449	128067620
rs7006076	128067826
rs7005862	128067930
rs7006251	128067931
rs7006409	128068017
rs6988257	128068044
rs16901969	128068685
rs16901970	128069303
rs10453084	128069367
rs6987723	128069447
rs6987640	128069616
rs12675027	128069785
rs13259163	128070111
rs13256837	128070171
rs13259247	128070181
rs13256863	128070189
rs13257829	128070218
rs13259520	128070224
rs10106712	128070449
rs3956788	128070480
rs7824451	128071049
rs7824785	128071298
rs13268116	128071384
rs6470498	128072308
rs1011829	128072331
rs6999725	128072340
rs28686871	128073068
rs1456306	128073088
rs13260084	128073418
rs12544678	128074274
rs7827234	128074694
rs7844219	128075403
rs6982324	128075644
rs10098867	128075845
rs28667456	128076744
rs28368838	128076745
rs12676925	128076746
rs12216801	128076898

rs10956352	128077241
rs17866275	128077520
rs7464849	128077876
rs4341134	128077876
rs12549692	128078335
rs7837051	128078437
rs1378898	128078955
rs4593501	128078994
rs1378897	128079247
rs2392734	128079338
rs13259145	128079368
rs6470499	128079982
rs7842760	128080058
rs16901976	128080672
rs1551512	128080714
rs16901979	128081504
rs10505483	128081783
rs7817677	128082092
rs10505482	128082437
rs1456305	128083840
rs17184796	128085047
rs12678505	128085094
rs16901983	128085630
rs6989838	128085960
rs7341677	128086973
rs7013255	128087075
rs11274077	128087187
rs16901984	128087549
rs10956353	128087878
rs10098156	128088379
rs6995291	128088912
rs7000307	128089664
rs12545722	128089822
rs16901985	128090565
rs13253958	128090894
rs7840773	128091922
rs7824364	128091953
rs28645998	128094025
rs7816535	128094256

SNP names in LD Block C with genomic locations (base pair start) in NCBI Build 34. Based on db SNP 125.

Table 5B. SNP markers within the region defined by SEQ ID NO:2 (LD block C'). Shown are marker names, position in NCBI builds 34, 35 and 36, SEQ ID NO:2 and SEQ ID NO:1, as well as SEQ ID of the marker, as applicable.

Marker	Pos Build 34	Pos Build 35 & 36	Pos SEQ ID NO:2	Pos SEQ ID NO:1	SEQ ID NO
rs1031586	128029114	128141707	2		
rs17764031	128030209	128142802	1097		
rs35569027		128142871	1166		
rs10091694	128030343	128142936	1231		
rs1456314	128032279	128144872	3167	280	
rs1551516	128032478	128145071	3366	479	
rs3940781	128032804	128145397	3692	805	
rs16901935	128033134	128145727	4022	1135	49
rs35664936		128146550	4845	1958	
rs12542102	128034110	128146703	4998	2111	
rs35732005		128146827	5122	2235	
rs36032667		128146879	5174	2287	
rs11987219	128035270	128147863	6158	3271	
rs11988135	128035998	128148591	6886	3999	
rs2124598	128036064	128148657	6952	4065	
rs2392726	128036068	128148661	6956	4069	
rs2392727	128036185	128148778	7073	4186	
rs2392728	128036232	128148825	7120	4233	
rs2392729	128036255	128148848	7143	4256	
rs1014656	128036272	128148865	7160	4273	
rs2392730	128036283	128148876	7171	4284	
rs2392731	128036329	128148922	7217	4330	
rs1014655	128036428	128149021	7316	4429	
rs2392732	128036523	128149116	7411	4524	
rs2893603	128036546	128149139	7434	4547	
rs34138989		128149265	7560	4673	
rs9656965	128036837	128149430	7725	4838	
rs17831626	128037012	128149605	7900	5013	
rs16901938	128037089	128149682	7977	5090	
rs28451013	128037235	128149828	8123	5236	
rs28385355	128037375	128149968	8263	5376	
rs35755774		128149994	8289	5402	
rs11785961	128037424	128150017	8312	5425	
rs13267056	128037676	128150269	8564	5677	
rs7824674	128037682	128150275	8570	5683	
rs7824679	128037699	128150292	8587	5700	
rs7824923	128037727	128150320	8615	5728	
rs35346020		128150398	8693	5806	
rs7843300	128037820	128150413	8708	5821	
rs7824957	128037849	128150442	8737	5850	
rs7825394	128038038	128150631	8926	6039	
rs10308400	128038110	128150703	8998	6111	
rs7825414	128038110	128150703	8998	6111	
rs35954086		128150967	9262	6375	
rs11994932	128038487	128151080	9375	6488	
rs13248907	128038934	128151527	9822	6935	
rs13282331	128038958	128151551	9846	6959	
rs13282364	128038998	128151591	9886	6999	
rs35832881		128151824	10119	7232	
rs12548394	128039291	128151884	10179	7292	
rs35409776		128151899	10194	7307	
rs10089648	128039438	128152031	10326	7439	
rs13363429	128039438	128152031	10326	7439	
rs34895815		128152438	10733	7846	
rs11985630	128040217	128152810	11105	8218	
rs6993569	128040686	128153279	11574	8687	
rs35520562		128153449	11744	8857	
rs6994316	128041128	128153721	12016	9129	
rs34629453		128153986	12281	9394	
rs12550334	128041513	128154106	12401	9514	
rs35090438		128154122	12417	9530	
rs11998124	128042069	128154662	12957	10070	
rs6999589	128042235	128154828	13123	10236	

Table 5B. SNP markers within the region defined by SEQ ID NO:2 (LD block C'). Shown are marker names, position in NCBI builds 34, 35 and 36, SEQ ID NO:2 and SEQ ID NO:1, as well as SEQ ID of the marker, as applicable.

Marker	Pos Build 34	Pos Build 35 & 36	Pos SEQ ID NO:2	Pos SEQ ID NO:1	SEQ ID NO
rs11998248	128042344	128154937	13232	10345	
rs35837627		128155023	13318	10431	
rs35450468		128155214	13509	10622	
rs34658436		128155215	13510	10623	
rs34985891		128155403	13698	10811	
rs7005114	128043347	128155940	14235	11348	
rs11319438	128043466	128156059	14354	11467	
rs6991325	128043515	128156108	14403	11516	
rs1902431	128043665	128156258	14553	11666	
rs34371194		128156400	14695	11808	
rs6470494	128044493	128157086	15381	12494	55
rs7843737	128044655	128157248	15543	12656	
rs10095770	128045159	128157752	16047	13160	
rs41388647		128158164	16459	13572	
rs6987760	128046900	128159493	17788	14901	
rs10306042	128047255	128159848	18143	15256	
rs6470495	128047255	128159848	18143	15256	
rs34872094		128159883	18178	15291	
rs4581008	128047454	128160047	18342	15455	
rs4620246	128047551	128160144	18439	15552	
rs7461435	128047551	128160144	18439	15552	
rs36002700		128160340	18635	15748	
rs12544839	128047971	128160564	18859	15972	
rs12544844	128048147	128160740	19035	16148	
rs10956351	128048148	128160741	19036	16149	
rs36012530		128160820	19115	16228	
rs34375527		128160828	19123	16236	
rs35719509		128160838	19133	16246	
rs1016342	128049044	128161637	19932	17045	36
rs1551511	128049403	128161996	20291	17404	
rs1031589	128049571	128162164	20459	17572	40
rs1378896	128049643	128162236	20531	17644	
rs1031588	128049866	128162459	20754	17867	
rs1016343	128049886	128162479	20774	17887	37
rs1031587	128050061	128162654	20949	18062	39
rs1551510	128050067	128162660	20955	18068	47
rs4871008	128050130	128162723	21018	18131	54
rs11363193	128050446	128163039	21334	18447	
rs35534753**		128163334	21629	18742	
rs6981122	128051049	128163642	21937	19050	57
rs7001504	128051072	128163665	21960	19073	
rs13252298	128051745	128164338	22633	19746	44
rs35201402		128165598	23893	21006	
rs7841060	128053066	128165659	23954	21067	72
rs35024323		128165800	24095	21208	
rs4540377	128053503	128166096	24391	21504	
rs6997559	128054167	128166760	25055	22168	60
rs11554368	128055277	128167870	26165	23278	
rs1802259		128168243	26538	23651	
rs17831913	128055755	128168348	26643	23756	
rs7007694	128055755	128168348	26643	23756	
rs13257371	128055831	128168424	26719	23832	
rs4571699	128055858	128168451	26746	23859	
rs1840709	128056044	128168637	26932	24045	
rs3037753	128056059	128168652	26947	24060	
rs7837848	128056121	128168714	27009	24122	
rs3837167	128056513	128169106	27401	24514	
rs3837166	128056618	128169211	27506	24619	
rs11993508	128056665	128169258	27553	24666	
rs17832021	128056765	128169358	27653	24766	
rs3999809	128056897	128169490	27785	24898	
rs3999810	128056956	128169549	27844	24957	
rs28663168	128056998	128169591	27886	24999	
rs3999811	128057051	128169644	27939	25052	
rs3999812	128057096	128169689	27984	25097	

Table 5B. SNP markers within the region defined by SEQ ID NO:2 (LD block C'). Shown are marker names, position in NCBI builds 34, 35 and 36, SEQ ID NO:2 and SEQ ID NO:1, as well as SEQ ID of the marker, as applicable.

Marker	Pos Build 34	Pos Build 35 & 36	Pos SEQ ID NO:2	Pos SEQ ID NO:1	SEQ ID NO
rs3999806	128057101	128169694	27989	25102	
rs3857883	128057195	128169788	28083	25196	
rs2392733	128057261	128169854	28149	25262	
rs1456317	128057295	128169888	28183	25296	
rs3999807***		128169888	28183	25296	
rs1456316	128057437	128170030	28325	25438	
rs16901946	128057514	128170107	28402	25515	
rs28735726	128057518	128170111	28406	25519	
rs9656813	128057553	128170146	28441	25554	
rs9656814	128057566	128170159	28454	25567	
rs9656815	128057592	128170185	28480	25593	
rs10505484	128057901	128170494	28789	25902	
rs11994653	128057932	128170525	28820	25933	
rs6470496	128058618	128171211	29506	26619	
rs6470497	128058919	128171512	29807	26920	
rs7844454	128059090	128171683	29978	27091	
rs7826322	128059142	128171735	30030	27143	
rs12677860	128059214	128171807	30102	27215	
rs12677861	128059216	128171809	30104	27217	
rs7000321	128059608	128172201	30496	27609	
rs12682421	128059740	128172333	30628	27741	
rs7000967	128059949	128172542	30837	27950	
rs7001895	128060268	128172861	31156	28269	
rs1456315	128060526	128173119	31414	28527	46
rs5013678	128060568	128173161	31456	28569	
rs7463708	128060644	128173237	31532	28645	
rs34446482		128173246	31541	28654	
rs13254738	128060932	128173525	31820	28933	45
rs7006390	128060956	128173549	31844	28957	63
rs1073997	128061776	128174369	32664	29777	41
rs7012407	128062182	128174775	33070	30183	
rs7012442	128062320	128174913	33208	30321	65
rs7016828	128062821	128175414	33709	30822	
rs7016830	128062824	128175417	33712	30825	
rs12682344	128063373	128175966	34261	31374	43
rs6983561	128063469	128176062	34357	31470	21
rs16901948	128063690	128176283	34578	31691	9
rs1869931	128063725	128176318	34613	31726	
rs16901949	128063742	128176335	34630	31743	10
rs16901950	128063832	128176425	34720	31833	11
rs16901952	128063859	128176452	34747	31860	12
rs7005144	128064233	128176826	35121	32234	
rs16901953	128064818	128177411	35706	32819	13
rs4871009	128065005	128177598	35893	33006	
rs7010450	128065268	128177861	36156	33269	28
rs6990420	128065314	128177907	36202	33315	26
rs35365584		128177908	36203	33316	
rs6994359	128065512	128178105	36400	33513	
rs17832285	128065582	128178175	36470	33583	
rs7825340	128065718	128178311	36606	33719	
rs12544977	128065876	128178469	36764	33877	42
rs16901959	128066119	128178712	37007	34120	14
rs7826337	128066163	128178756	37051	34164	69
rs7826388	128066365	128178958	37253	34366	33
rs7830341	128066519	128179112	37407	34520	34
rs7830381	128066714	128179307	37602	34715	
rs16901966	128066841	128179434	37729	34842	15
rs16901967	128066866	128179459	37754	34867	16
rs7000910	128067194	128179787	38082	35195	62
rs7001069	128067235	128179828	38123	35236	27
rs17765137	128067403	128179996	38291	35404	
rs11781162	128067485	128180078	38373	35486	
rs11774449	128067621	128180214	38509	35622	
rs7006076	128067827	128180420	38715	35828	
rs7005862	128067931	128180524	38819	35932	

Table 5B. SNP markers within the region defined by SEQ ID NO:2 (LD block C'). Shown are marker names, position in NCBI builds 34, 35 and 36, SEQ ID NO:2 and SEQ ID NO:1, as well as SEQ ID of the marker, as applicable.

Marker	Pos Build 34	Pos Build 35 & 36	Pos SEQ ID NO:2	Pos SEQ ID NO:1	SEQ ID NO
rs7006251	128067932	128180525	38820	35933	
rs7006409	128068018	128180611	38906	36019	64
rs6988257	128068045	128180638	38933	36046	24
rs16901969	128068686	128181279	39574	36687	17
rs16901970	128069304	128181897	40192	37305	18
rs10453084	128069368	128181961	40256	37369	6
rs6987723	128069448	128182041	40336	37449	23
rs6987640	128069617	128182210	40505	37618	22
rs12675027	128069786	128182379	40674	37787	
rs13259163	128070112	128182705	41000	38113	
rs13256837	128070172	128182765	41060	38173	
rs13259247	128070182	128182775	41070	38183	
rs13256863	128070190	128182783	41078	38191	
rs13257829	128070219	128182812	41107	38220	
rs13259520	128070225	128182818	41113	38226	
rs10106712	128070450	128183043	41338	38451	
rs3956788	128070481	128183074	41369	38482	
rs7824451	128071050	128183643	41938	39051	31
rs7824785	128071299	128183892	42187	39300	32
rs13268116	128071385	128183978	42273	39386	
rs6470498	128072309	128184902	43197	40310	20
rs1011829	128072332	128184925	43220	40333	
rs6999725	128072341	128184934	43229	40342	
rs34965819		128185532	43827	40940	
rs28686871	128073069	128185662	43957	41070	
rs1456306	128073089	128185682	43977	41090	
rs13260084	128073419	128186012	44307	41420	
rs35270158		128186496	44791	41904	
rs12544678	128074275	128186868	45163	42276	
rs7827234	128074695	128187288	45583	42696	70
rs34512164		128187312	45607	42720	
rs7844219	128075404	128187997	46292	43405	35
rs6982324	128075645	128188238	46533	43646	
rs10098867	128075846	128188439	46734	43847	
rs34171547		128189021	47316	44429	
rs34646118		128189070	47365	44478	
rs28667456	128076745	128189338	47633	44746	
rs28368838	128076746	128189339	47634	44747	
rs12676925	128076747	128189340	47635	44748	
rs34425474		128189416	47711	44824	
rs12216801	128076899	128189492	47787	44900	
rs10956352	128077242	128189835	48130	45243	
rs17866275	128077521	128190114	48409	45522	
rs4341134	128077877	128190470	48765	45878	
rs7464849	128077877	128190470	48765	45878	
rs12549692	128078336	128190929	49224	46337	
rs7837051	128078438	128191031	49326	46439	
rs1378898	128078956	128191549	49844	46957	
rs4593501	128078995	128191588	49883	46996	
rs1378897	128079248	128191841	50136	47249	
rs2392734	128079339	128191932	50227	47340	
rs13259145	128079369	128191962	50257	47370	
rs6470499	128079983	128192576	50871	47984	56
rs7842760	128080059	128192652	50947	48060	
rs16901976	128080673	128193266	51561	48674	
rs1551512	128080715	128193308	51603	48716	8
rs16901979	128081505	128194098	52393	49506	73
rs10505483	128081784	128194377	52672	49785	7
rs7817677	128082093	128194686	52981	50094	30
rs35045168		128194764	53059	50172	
rs10505482	128082438	128195031	53326	50439	
rs1456305	128083841	128196434	54729	51842	
rs35997767		128196523	54818	51931	
rs17184796	128085048	128197641	55936	53049	
rs12678505	128085095	128197688	55983	53096	

Table 5B. SNP markers within the region defined by SEQ ID NO:2 (LD block C'). Shown are marker names, position in NCBI builds 34, 35 and 36, SEQ ID NO:2 and SEQ ID NO:1, as well as SEQ ID of the marker, as applicable.

Marker	Pos Build 34	Pos Build 35 & 36	Pos SEQ ID NO:2	Pos SEQ ID NO:1	SEQ ID NO
rs16901983	128085631	128198224	56519	53632	
rs34933805		128198512	56807	53920	
rs6989838	128085961	128198554	56849	53962	25
rs7341677	128086974	128199567	57862	54975	
rs7013255	128087076	128199669	57964	55077	29
rs11274077	128087187	128199780	58075	55188	
rs16901984	128087550	128200143	58438	55551	19
rs34826080*		128200278	58573	55686	
rs10956353	128087879	128200472	58767	55880	
rs10098156	128088380	128200973	59268	56381	
rs6995291	128088913	128201506	59801	56914	
rs34537182		128201519	59814	56927	
rs7000307	128089665	128202258	60553	57666	61
rs12545722	128089823	128202416	60711	57824	
rs16901985	128090566	128203159	61454	58567	
rs34720002		128203299	61594	58707	
rs35610862		128203459	61754	58867	
rs13253958	128090895	128203488	61783	58896	
rs7840773	128091923	128204516	62811	59924	71
rs7824364	128091954	128204547	62842	59955	68
rs28645998	128094026	128206619	64914	62027	
rs7816535	128094257	128206850	65145	62258	67
rs16901988	128095149	128207742	66037		
rs7821533	128095346	128207939	66234		
rs7821299	128095370	128207963	66258		
rs34074310		128208988	67283		
rs12335235	128100703	128213296	71591		
rs34933261		128214191	72486		
rs2124596	128101759	128214352	72647		
rs2124595	128101824	128214417	72712		
rs4624993	128101824	128214417	72712		
rs10111496	128102406	128214999	73294		
rs13364435	128102406	128214999	73294		
rs35623297		128215361	73656		
rs34035606		128215465	73760		
rs10101884	128102897	128215490	73785		
rs6470500	128103187	128215780	74075		
rs6987409	128106750	128219343	77638		
rs35775018		128219607	77902		
rs36080589		128219616	77911		
rs17446747	128107241	128219834	78129		
rs17446776	128108261	128220854	79149		
rs10094434	128110394	128222987	81282		
rs7017820	128110462	128223055	81350		
rs7018238	128110470	128223063	81358		
rs7018243	128110480	128223073	81368		
rs11286339	128110587	128223180	81475		
rs13253040	128111938	128224531	82826		
rs13252570	128111939	128224532	82827		
rs13252598	128111964	128224557	82852		
rs2466040	128112253	128224846	83141		
rs7822987	128113252	128225845	84140		
rs7822995	128113277	128225870	84165		
rs2516280	128115044	128227637	85932		
rs13257144	128115979	128228572	86867		
rs3100857	128116021	128228614	86909		
rs34946000		128228837	87132		
rs35282781		128228987	87282		
rs28824784	128116644	128229237	87532		
rs34195077		128230348	88643		
rs2392724	128118125	128230718	89013		
rs2925465	128118125	128230718	89013		
rs4427134	128118125	128230718	89013		
rs36021065		128230793	89088		
rs2392725	128118217	128230810	89105		

Table 5B. SNP markers within the region defined by SEQ ID NO:2 (LD block C'). Shown are marker names, position in NCBI builds 34, 35 and 36, SEQ ID NO:2 and SEQ ID NO:1, as well as SEQ ID of the marker, as applicable.

Marker	Pos Build 34	Pos Build 35 & 36	Pos SEQ ID NO:2	Pos SEQ ID NO:1	SEQ ID NO
rs2954005	128118217	128230810	89105		
rs4593502	128118217	128230810	89105		
rs2893602	128118242	128230835	89130		
rs4552862	128118242	128230835	89130		
rs7459511	128118242	128230835	89130		
rs35636657		128230862	89157		
rs2445605	128118533	128231126	89421		
rs7843464	128118533	128231126	89421		
rs17374770	128119309	128231902	90197		
rs17446916	128119563	128232156	90451		
rs11341635	128119823	128232416	90711		
rs35532717		128232429	90724		
rs5894869	128120238	128232831	91126		
rs33965163		128232832	91127		
rs3037747	128120240	128232833	91128		
rs3758058	128122147	128234740	93035		
rs13275200	128123145	128235738	94033		
rs34106304		128237228	95523		
rs34285962		128238577	96872		
rs13269873	128126198	128238791	97086		
rs10111439	128126446	128239039	97334		

* (-/TATT polymorphism)

** (A/AGAA polymorphism)

*** (-/GTTT polymorphism)

Table 5C. Microsatellite markers within the SEQ ID NO:2 (LD Block C') region. Shown is the marker name, start and end positions of a 400bp genomic segment centered around each microsatellite marker in NCBI Build 34, and the corresponding location in SEQ ID NO:2

Marker name	Start B34	End B34	Start SEQ ID NO:2	End SEQ ID NO:2
DG8S502	128029888	128030288	776	1176
DG8S815	128043386	128043621	14274	14509
DG8S345	128055889	128056158	26777	27046
DG8S1334	128056004	128056111	26892	26999
DG8S625	128064928	128065358	35816	36246
DG8S551	128071376	128071525	42264	42413
DG8S911	128071591	128071748	42479	42636
DG8S1179	128075716	128076097	46604	46985
DG8S1508	128087019	128087367	57907	58255
DG8S1088	128094303	128094579	65191	65467
DG8S733	128097441	128097570	68329	68458
DG8S402	128099742	128100125	70630	71013
DG8S1431	128108504	128108934	79392	79822
DG8S1230	128119990	128120400	90878	91288

Genes and predicted genes that map to LD block C on chromosome 8q24.21 of the human genome include a retrotransposed gene that is an intronless copy of the SRRM1 gene on Chr1p36 (Genbank Accession No. BC017315), as well as predicted genes (e.g., NT_008046.701, chr8_1173.1, chr8.129.001.a, vugee.bDec03, kloger, keebly). The underlying variation in markers or haplotypes associated with the LD block C region and cancer may affect expression of genes within the LD block C, but also nearby genes, such as NSE2, POU5FLC20, c-MYC, PVT1, and/or other known, unknown or predicted genes in the area. Furthermore, such variation may affect RNA or protein stability or may have structural consequences, such that the region is more prone to somatic rearrangement in haplotype carriers. This is in accordance with LD Block C being amplified in a large percentage of cancers, including, but not limited to, prostate cancer (www.progenetix.com). In fact, Chr8q21-24 is the most frequently gained chromosomal region in all cancers combined (about 17%) and in prostate cancer (about 20%) (www.progenetix.com). Thus, the underlying variation could affect uncharacterized genes directly linked to the haplotypes described herein, or could influence neighbouring genes not directly linked to the haplotypes described herein.

Table 6. Key to sequence ID's provide herein.

SEQ ID NO	Marker/Sequence
1	LD Block C
2	LD Block C'
3	DG5S802
4	DG5S803-F
5	DG5S803-R
6	rs10453084
7	rs10505483
8	rs1551512
9	rs16901948
10	rs16901949
11	rs16901950
12	rs16901952
13	rs16901953
14	rs16901959
15	rs16901966
16	rs16901967
17	rs16901969
18	rs16901970

SEQ ID NO	Marker/Sequence
19	rs16901984
20	rs6470498
21	rs6983561
22	rs6987640
23	rs6987723
24	rs6988257
25	rs6989838
26	rs6990420
27	rs7001069
28	rs7010450
29	rs7013255
30	rs7817677
31	rs7824451
32	rs7824785
33	rs7826388
34	rs7830341
35	rs7844219
36	rs1016342
37	rs1016343
38	rs10216560
39	rs1031587
40	rs1031589
41	rs1073997
42	rs12544977
43	rs12682344
44	rs13252298
45	rs13254738
46	rs1456315
47	rs1551510
48	rs16901932
49	rs16901935
50	rs16901997
51	rs1995613
52	rs283710
53	rs283719
54	rs4871008
55	rs6470494
56	rs6470499
57	rs6981122
58	rs6984136
59	rs6990483
60	rs6997559
61	rs7000307
62	rs7000910
63	rs7006390
64	rs7006409
65	rs7012442
66	rs7017081
67	rs7816535
68	rs7824364
69	rs7826337
70	rs7827234
71	rs7840773
72	rs7841060
73	rs16901979

Discussion

As described herein, a locus on chromosome 8q24.21 has been demonstrated to play a role in
5 cancer (e.g., prostate cancer (e.g., aggressive prostate cancer). Particular markers and haplotypes

(e.g., HapC, haplotypes containing one or more of the markers depicted in Table 3) are present at a higher than expected frequency in subjects having prostate cancer. Based on the haplotypes described herein, which are associated with a propensity for particular forms of cancer, genetic susceptibility assays (e.g., a diagnostic screening test) can be used to identify individuals at risk for cancer.

- 5 The markers and haplotypes described herein do have a higher relative risk in the aggressive prostate cancer as compared to the whole prostate cancer group (Table 1), thereby indicating an increased risk for aggressive, fast growing prostate cancer. Given that a significant percentage of prostate cancer is a non-aggressive form that will not spread beyond the prostate and cause morbidity or mortality, and treatments of prostate cancer including prostatectomy, radiation, and chemotherapy all
10 have side effects and significant cost, it would be valuable to have diagnostic markers, such as those described herein, that show greater risk for aggressive prostate cancer as compared to the less aggressive form(s).

15 EXAMPLE 2: VERIFICATION OF ASSOCIATION WITH PROSTATE CANCER IN SEVERAL COHORTS

- Additional analysis further supports the presence of the variants associated with prostate cancer on chromosome 8q24. As depicted in Table 7, both the rs16901979 1 allele and HapC were found to be associated with prostate cancer in two cohorts of Caucasian ancestry and one cohort of
20 African ancestry.

- Replication study of association of single markers and haplotypes within LD Block C with prostate cancer. The cohorts are comprised of samples of Caucasian origin from Chicago, U.S. (Northwestern University) and Spain (Zaragoza University Hospital), and an African American cohort from Michigan, U.S. (University of Michigan). Allele codes for SNPs are as follows: 1= A, 2=C, 3=G,
25 4=T, and X=any allele

Table 7. Association of SNP markers on Chr8q24 to prostate cancer (PrCa) in Spain and the United States

Affected Phenotype	# of Individuals		p-val	RR	Allele frequency			
	Affected	Controls			Affected	Controls	Allele	PAR
PrCa-Chicago	419	237	0.006	2.39	0.054	0.023	1 rs16901979	0.061
PrCa-Spain	385	892	0.005	1.71	0.066	0.040	1 rs16901979	0.054
PrCa-JHU	373	372	0.005	1.35	0.500	0.420	1 rs16901979	0.240

Alleles for the marker rs16901979 and SNPs in same LD block at 8q24.21 are shown and the corresponding numbers of PrCa cases and controls (N), allelic frequencies of variants in affected and control individuals, the odds-ratio (OR) and two-sided P values. Affected individuals are those diagnosed with prostate cancer (ICD10 = C61). Controls are unaffected population based controls. Allele codes for SNPs are as follows: 1=A, 2=C, 3=G, 4=T. PAR is population attributable risk. The prostate cancer patients and controls from Spain and Chicago are of Caucasian ethnicity and the cohort from John Hopkins University (JHU) is of African American ethnicity.
1 rs16901979: using information from other SNPs to increase genotype yield.

The population frequency of the at-risk variants (approximated by the frequency in controls) in the Caucasian samples is comparable to that of the Icelandic population. However, it is noted that the population frequency in African Americans is much greater. The population attributable risk in the African American cohort is thus much higher or about 24% as compared to about 5-6% in the two
5 Caucasian origin cohorts. This suggests that a greater percentage of prostate cancer in African Americans can be explained by the at-risk variants of the present invention than in Caucasians.

Materials and Methods

The Caucasian U.S. study population consisted of 419 prostate cancer patients (ICD10 C61),
10 who underwent surgery at the Urology Department of Northwestern Memorial Hospital, Chicago, and 237 population based controls enrolled at the Department of Human Genetics, University of Chicago. Medical records were examined to retrieve clinical information including stage and biopsy Gleason score. The mean age at diagnosis was 59 years for patients. Both patients and controls were of self-reported European American ethnicity. This was confirmed by the estimation of genetic ancestry using
15 30 microsatellite markers distributed randomly throughout the genome (see below). The mean and median portion of European ancestry in this cohort were both greater than 0.99 (see methods described below for details). The study protocols were approved by the Institutional Review Boards of Northwestern University and the University of Chicago. All subjects gave written informed consent.

The Spanish study population consisted of 390 prostate cancer patients, recruited from the
20 Oncology Department of Zaragoza Hospital from June 2005 to June 2006. Patients were recruited by Dr. Jose I. Mayordomo and collaborating oncologists at the Division of Medical Oncology, University Hospital in Zaragoza, Spain. During the 12- month interval when the study samples were collected, 700 patients were eligible. Of these, about 600 (~ 85 %) patients were approached of whom 440 enrolled (73 % participation rate). All patients were of Caucasian ethnicity. The median time interval from prostate
25 cancer diagnosis to collection of blood samples was 5 months (mean 7 months, range 1 - 67 months). Clinical information including age at onset, grade and stage was collected from medical records. The mean age at onset for the patients is 69 years (median 71 years) and the range is from 45-83 years. The 892 Spanish controls were approached at the University Hospital in Zaragoza, Spain for all other diseases than cancer, questioned to rule out prior cancers, before drawing the blood sample. Study
30 protocols were approved by the Institutional Review Boards of Zaragoza University Hospital. All subjects gave written informed consent.

Evaluation of genetic ancestry

The program Structure (Pritchard, J.K. *et al.*, *Genetics* 115:945-59 (2000)) was used to estimate the
35 genetic ancestry of individuals. Structure infers the allele frequencies of K ancestral populations on the basis of multilocus genotypes from a set of individuals and a user-specified value of K, and assigns a proportion of ancestry from each of the inferred K populations to each individual. The analysis of the

data set was run with K=3, with the aim of identifying the proportion of African and European ancestry in each individual. The statistical significance of the difference in mean European ancestry between African American patients and controls was evaluated by reference to a null distribution derived from 10000 randomized datasets.

5 To evaluate genetically estimated ancestry of the study cohorts from the US, 30 unlinked microsatellite markers were selected from about 2000 microsatellites genotyped in a previously described (Pritchard, J.K. *et al.*, *Genetics* 115:945-59 (2000)) multi-ethnic cohort of 35 European Americans, 88 African Americans, 34 Chinese, and 29 Mexican Americans. Of the 2000 microsatellite markers the selected set showed the most significant differences between European Americans, African
10 Americans, and Asians, and also had good quality and yield: D1S2630, D1S2847, D1S466, D1S493, D2S166, D3S1583, D3S4011, D3S4559, D4S2460, D4S3014, D5S1967, DG5S802, D6S1037, D8S1719, D8S1746, D9S1777, D9S1839, D9S2168, D10S1698, D11S1321, D11S4206, D12S1723, D13S152, D14S588, D17S1799, D17S745, D18S464, D19S113, D20S878 and D22S1172. The following primer pairs were used for DG5S802: DG5S802-F: CAAGTTTAGCTGTGATGTACAGGTTT
15 (SEQ ID NO: 46) and DG5S802-R: TTCCAGAACCAAAGCCAAAT (SEQ ID NO: 47).

Discussion

In summary, significant association of prostate cancer risk to the 1 allele of rs16901979 has been demonstrated in three cohorts of European ancestry from Iceland, Spain and Chicago. The
20 corresponding population attributable risk (PAR) ranges from about 5-6% when considering all prostate cancer in these populations and is slightly increased in aggressive prostate cancer or to about 7-8%. The association was replicated between prostate cancer and the 1 allele of rs16901979 in an African American cohort with a relative risk of 1.35 ($P = 0.005$). Due to the increased frequency of these variants in the African American sample, the PAR in this population is much higher as compared to the
25 Caucasian samples or about 24%.

The variants described herein were identified through a genome wide association analysis. Over 300,000 SNPs were typed in 1455 prostate cancer patients and 4182 controls. Our attention focused to this region as it is the second most significant haplotype contained within an individual LD block in the whole genome (LD block haplotype).

30 Of importance is the greatly increased frequency of this haplotype in African Americans as compared to Caucasians which may explain the higher incidence of prostate cancer in this group.

THE CLAIMS DEFINING THE INVENTIONS ARE AS FOLLOWS:

1. A method for diagnosing a susceptibility to prostate cancer in a human individual, comprising determining the presence or absence of at least one allele of at least one polymorphic marker in a nucleic acid sample obtained from the individual, within the genomic region with sequence as set forth in SEQ ID NO:2, selected from the group consisting of rs16901979, and markers in linkage disequilibrium therewith, and wherein the presence of the at least one allele is indicative of a susceptibility to the cancer.
2. The method of Claim 1, wherein the at least one marker comprises at least one marker within Chr8q24.21 in strong linkage disequilibrium, as defined by $|D'| > 0.8$ and/or $r^2 > 0.2$ with marker rs16901979.
3. The method of Claim 1 or Claim 2, wherein the markers in linkage disequilibrium with rs16901979 are selected from the group consisting of the markers listed in Table 4A and 4B.
4. The method of any of the preceding claims, wherein the susceptibility is increased susceptibility with a relative risk of at least 1.7
5. The method of claim 4, wherein the at least one marker is the marker rs16901979 allele 1.
6. The method of any of the claims 1 to 3, wherein the susceptibility is decreased susceptibility.
7. The method of claim 6, wherein the at least one marker is the marker rs16901979 allele 2.
8. The method of any one of the preceding Claims, wherein the prostate cancer is an aggressive prostate cancer as defined by a combined Gleason score of 7(4+3)-10.
9. The method of Claim 8, wherein the presence of the at least one marker is indicative of a more aggressive prostate cancer and/or a worse prognosis.
10. A method of identification of a marker for use in assessing susceptibility to prostate cancer, the method comprising
 - a. identifying at least one polymorphic marker within SEQ ID NO:2;
 - b. determining the genotype status of a sample of individuals diagnosed with, or having a susceptibility to, prostate cancer; and
 - c. determining the genotype status of a sample of control individuals;
 wherein a significant difference in frequency of at least one allele in at least one polymorphism in individuals diagnosed with, or having a susceptibility to, prostate cancer, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one polymorphism being useful for assessing susceptibility to cancer.

11. The method of Claim 10, wherein the at least one polymorphic marker is in linkage disequilibrium, as characterized by numerical values of r^2 of greater than 0.2 and/or $|D'|$ of greater than 0.8 with marker rs16901979.
- 5 12. The method of Claim 10 or claim 11, wherein an increase in frequency of the at least one allele in the at least one polymorphism in individuals diagnosed with, or having a susceptibility to, prostate cancer, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one polymorphism being useful for assessing increased susceptibility to prostate cancer.
- 10 13. The method of any of the Claims 10 to 12, wherein a decrease in frequency of the at least one allele in the at least one polymorphism in individuals diagnosed with, or having a susceptibility to, prostate cancer, as compared with the frequency of the at least one allele in the control sample is indicative of the at least one polymorphism being useful for assessing decreased susceptibility to, or protection against, prostate cancer.
- 15 14. The method of any of the preceding Claims, further comprising analyzing non-genetic information to make risk assessment, diagnosis, or prognosis of the individual.
- 20 15. The method of Claim 14, wherein the non-genetic information is selected from age, gender, ethnicity, socioeconomic status, previous disease diagnosis, medical history of subject, family history of prostate cancer, biochemical measurements, and clinical measurements.
- 25 16. The method of Claim 14 or claim 15, further comprising calculating overall risk.
- 30 17. Use of an oligonucleotide probe in the manufacture of a diagnostic reagent for diagnosing and/or assessing susceptibility to prostate cancer in a human individual, wherein the probe hybridizes to a segment of a nucleic acid whose nucleotide sequence is given by SEQ ID NO: 2 that comprises at least one polymorphic site, wherein the fragment is 15-500 nucleotides in length.
18. The use according to Claim 17, wherein the polymorphic site is selected from the polymorphic markers set forth in Table 5A, 5B and 5C.
- 35 19. The use according to Claim 17 or 18, wherein the polymorphic site is rs16901979 (SEQ ID NO: 73).
20. An apparatus when used in a method according to any one of the claims 1 to 16, comprising:
 - 40 a computer readable memory; and
 - a routine stored on the computer readable memory;

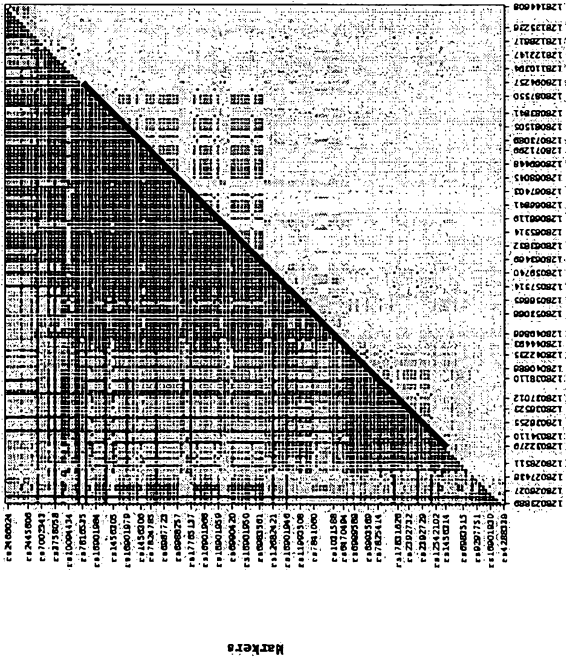
2007310412 29 Jan 2013

- 5 wherein the routine is adapted to be executed on a processor to analyze marker and/or haplotype information for at least one human individual with respect to at least one polymorphic marker selected from the markers rs16901979, and markers in linkage disequilibrium therewith, as defined by numerical values of r^2 of at least 0.2 and/or values of $|D'|$ of at least 0.8, and generate an output based on the marker or haplotype information, wherein the output comprises a risk measure of the at least one marker or haplotype as a genetic indicator of prostate cancer for the human individual.
- 10 21. The apparatus of Claim 20, wherein the routine further comprises an indicator of the frequency of at least one allele of at least one polymorphic marker or at least one haplotype in a plurality of individuals diagnosed with prostate cancer, and an indicator of the frequency of at the least one allele of at least one polymorphic marker in a plurality of reference individuals, and wherein a risk measure is based on a comparison of the at least one marker status for the human individual to the indicator of the frequency of the at least one marker information for the plurality
- 15 of individuals diagnosed with prostate cancer.
22. A method according to claim 1, substantially as hereinbefore described.
23. A method according to claim 10, substantially as hereinbefore described.
- 20 24. A use according to claim 17, substantially as hereinbefore described.
25. An apparatus according to claim 20, substantially as hereinbefore described.

1/2

A.

CEU



B.

YRI

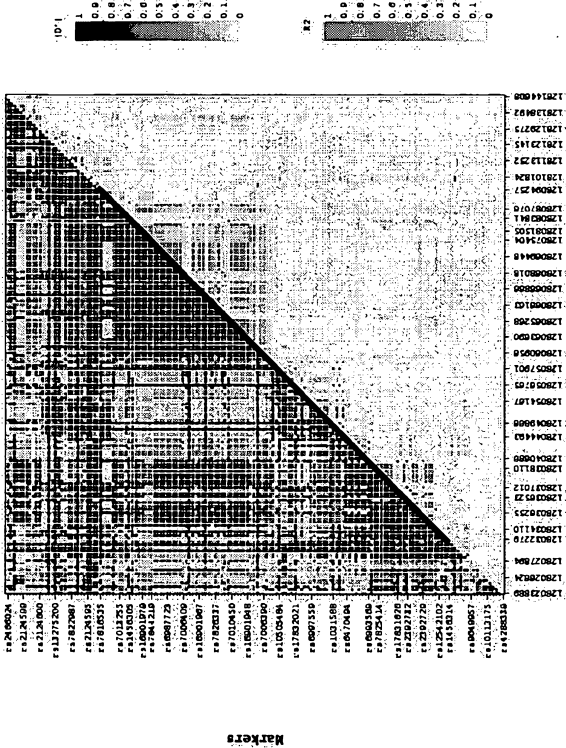


Figure 1

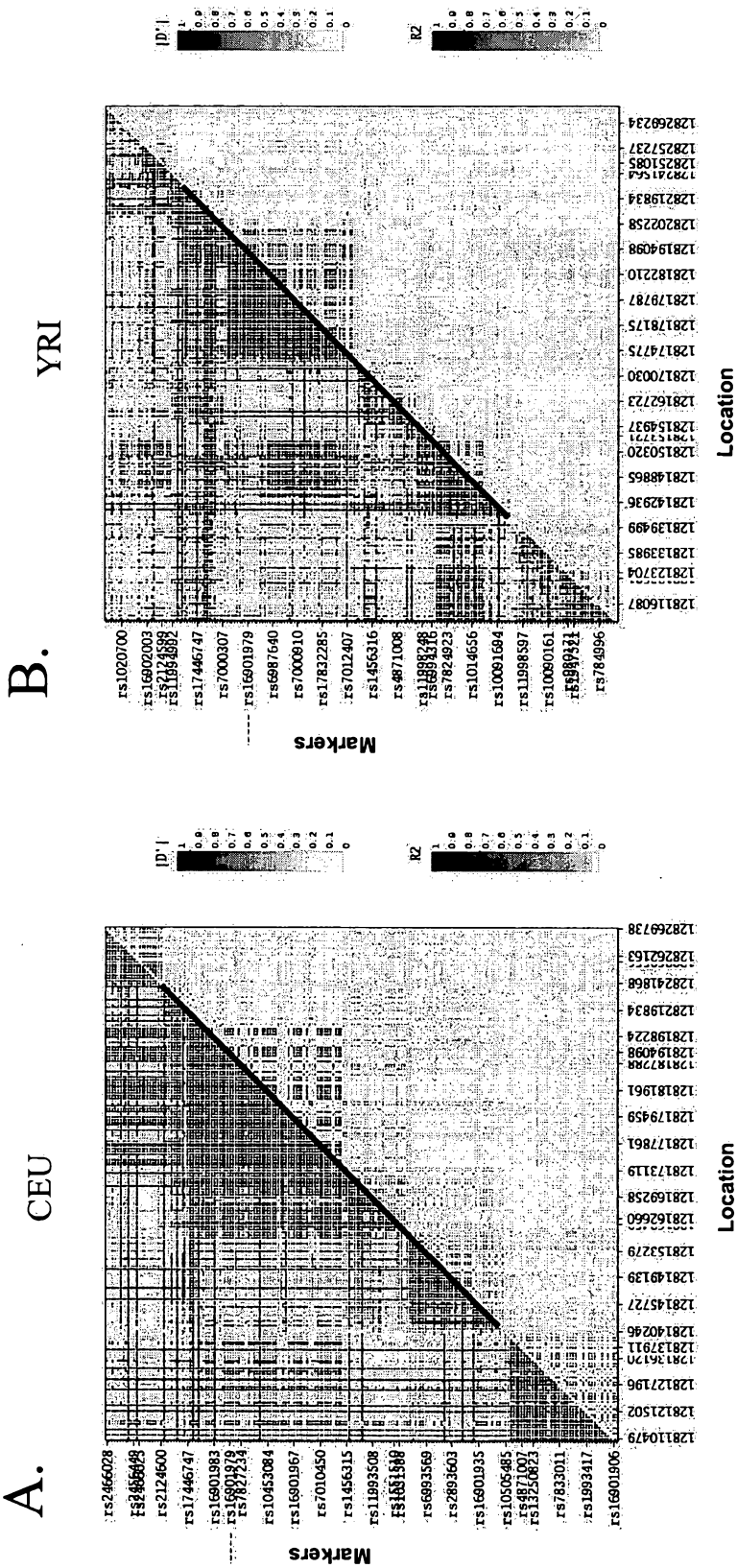


Figure 2

SEQUENCE LISTING

<110> deCODE Genetics

<120> CANCER SUSCEPTIBILITY VARIANTS ON CHR8Q24.21

<130> 2501PC00

<160> 73

<170> PatentIn version 3.4

<210> 1

<211> 63001

<212> DNA

<213> Homo sapiens

<400> 1

gataacatgc	agtaactttta	agtaatactg	tatatagaga	agtagatcga	aataggtaaa	60
ctattaacat	aaattttcc	gtagtggaa	cccaattaaa	gcaataatgt	agagcaataa	120
gaagcaaaa	ctgcttggat	ccaaatcatt	tctgaacttt	gtaactttgg	tcgtgttgct	180
ttctgtgaat	cagtttcttc	atttaaaaaa	gttggttatga	ggatagaaca	tgctagtatt	240
tataaaatgc	tcataaatcc	ttggcacaaa	gtaagctata	gactcaaaaa	tggtgattcc	300
agacctcaag	aaataagaac	ttgaatttca	gagaaatcct	cccttattat	ccctgaggtc	360
cactccatag	cttgatcaaa	gaagtgggat	actatgtagt	agactagtgt	ggcagcacct	420
taaaccacaca	aataagtttg	gctgttgtaa	agagaaggga	aacccaaaatt	atttcaggac	480
ctactatgtg	cagagttttt	atctgctata	tacataggca	tatgctattc	catcagtggg	540
tcagagggca	taaacagtta	ataaagctat	ctcagggatg	attaaattga	gagttgctca	600
gggtcataca	atTTTTgagt	caggaattaa	accaggggct	atttcattcc	aaggcccatg	660
ttactttctg	gtaaaacaaa	taaccatgaa	atataatata	ttttctttga	atccctaata	720
gtgggtgcta	ttataccccc	aattacttaa	attgatcctg	tacattgcc	agagattaaa	780
ttataaatta	ttgagttcaa	aaagttaatt	tcttcctcca	aagagatggc	atcttacagt	840
ttaattactc	tgacagcttt	ttttttttat	tttttctcct	agtgaattat	caacagaaac	900
tacctagaga	aacctgtaac	ttcagatcct	ctgattaatt	gattttcttt	gggtgctgct	960
tgacagtaac	agaatgggtc	atggggacag	tttttcagt	aatgttttca	acaatgatag	1020
gaacacagag	ctaagcagag	tgtcaagaga	catttagtgc	agtagagaca	tctcatctgt	1080
ttttctgcag	aacttgcata	aaattagcat	ctgattatta	tcacctcaga	atgaaccgga	1140
acccgtaaa	atgcttttga	tcgtaaaaagt	ggggaggaga	ccctgatgga	agaagagaag	1200
acttcagcct	cacaatagaa	aagactgtgt	ttatgcacct	aaatctaact	tagacaccat	1260
ctaattcaag	aaacaggcac	atttagggaa	ataaagaaaa	gaggcagaaa	gagatgaagc	1320
accttggtg	aatcagagaa	attaggggtt	aatgttccag	actgatgact	tgaaggaggg	1380
tgtgcatttt	catgcaagg	accattctgt	agaagaagca	gcacattcac	ccttaaatga	1440
gcaaagagaa	tccaatcac	tctccctctc	accctccctc	tttccctcat	ttctctagct	1500
ctttatttcc	aagaaggcca	gcccttctct	tcagccacat	ggtatttgga	cagcctaggc	1560
tcaatccctg	cagggaaagc	tcaatcagcc	agctgagcca	gtcactgaac	aaccaagtct	1620
ggtaagaaca	ggggtgaagt	atttaagaag	cctatcgtgg	ctggagggct	aaagccttgt	1680
tctcaaata	gctttatagg	taataaaact	gacattaaag	caaaaacaaa	taagagtaat	1740
aaggcttctt	ctgtgtgttt	cttgggtttt	gtttctcatt	tgattccaaa	tggggagaca	1800
agggtatatt	tatcctaccc	ttcccccagc	ctcaacagg	taaatgggtt	aatacatgaa	1860
aagcatcaat	cacagtgtct	ggagcattat	cagtatcagt	aaatgttaac	tattactttg	1920
aggctggccc	atgtggctcc	tggcattaca	ggcaccagag	gaattccagg	gagaccagga	1980
ggattatagt	ctgggggagt	atggaggatg	tgtactgtg	tggcagggtc	tttgcaaaact	2040
gcattaaaa	agtcttgata	ttttcagaat	ctaacgggat	cccatatact	atgattctga	2100
aattcacttg	cgcttcttta	gatttttgcat	ctacatgatg	catagtaaca	gttcttaatt	2160
tggtttatct	ctagcttatt	tatgggtgtc	attatttcaa	gagcctagta	aagttgaaat	2220
caatttttag	actacaaaaa	aaaaaaaaaa	ggaagagatg	tgagtagggt	ctcactgtgt	2280
tgcacgtaaa	tgcattgaca	tgggaagattt	aaaacacctg	atgagtttta	ttctgtttct	2340
ttatttctta	ttaaaggtag	tactgtgcaa	cattccctac	cctctatttt	tgacactgac	2400
gttgggctta	ttttgattgt	ctaagagggtc	ctgtgctaca	agctggcagg	gcttgctggg	2460
atgaagtga	aagcaaagat	ttgctatatg	gtatagccag	gggcagagac	agggtttctg	2520
ggggcatgag	acttctacag	tggggagaca	ctattaagaa	atagaataca	aaattatgac	2580
tacaaaatga	gaggcccca	agcttatgtt	tcattagctt	cacagggagc	tcacttctgc	2640
ccagagccaa	ggactgtgga	catgacctgg	taatggcatc	aatgattcc	ctttattgcc	2700
tcccctgggg	cagaaatggt	ggcagcaaa	caacttccaa	tcacaactat	cttctttcat	2760
ttgacacagg	aaggggaagc	tccctcttta	atctctaata	gaaacattaa	acagcaggcc	2820
atcacaattt	ctagctaaag	agttttcata	ggacctttgc	aaaggttgca	ttcctccctt	2880
cccttggtgg	agtatgcata	atgtaccttt	taggaatata	aattcagtgt	cttggaacac	2940

atgatttaga	atcagtgttc	agttaccaac	tatcaaataa	aaaaataaat	ccagcatatg	3000
actaacggaa	agaagtcaat	ctgaaatgg	tacatattat	atgattccaa	ctatatgaca	3060
ttttgggaaa	ggcaagacta	tggagacagt	gaaaaggcca	gtgggttgcca	ggggttagaa	3120
ggaagggaga	gatgaatcgg	ggaagcacag	aggattttta	gggcagcaaa	actgttctgt	3180
atgatactat	aatgggtgat	acatgtcatt	gtacatttgt	caaaacacat	atgtcttggt	3240
ccatttaggc	tgctctaaga	aaatatctta	gactgggtgg	cttataaaca	acagcaattg	3300
gtttcttaca	gttctgaagt	ccacgatgaa	ggcactggca	gattcagtgt	ctgggtgcagg	3360
cgatcctctt	gctttaactt	cactgggtgga	aggggcaaa	cagctgtcta	gggcttctta	3420
cataaaggta	ttaatctcat	tcatgagggc	tctgccctca	tgacctagtc	acctctaaaa	3480
ggtcccacct	cttaaaacca	tcacactggg	ctttagggtt	agtggggatg	aattttgtgg	3540
ggaacccaaa	taacagacca	aaattcactg	ggaattgttt	tggggagaca	taaacattca	3600
gaccacagaa	gcacagcaca	tacaacacca	agagtaaatg	tttatgtaaa	ctgcagattt	3660
ttagtgatgt	cacggagaaa	aacttctcct	cgactctctt	aggtttagtg	cttgggaacc	3720
agtaaattaa	actaactaaa	gacaagttaa	caagagaaaa	agcacaagtt	tttattgata	3780
tttacatgca	taggagttca	tggaaaagaa	atgaaattca	aagatgcaat	tagactcaca	3840
ggggttatat	accattttta	caaaggaaa	gtttggggct	tcaaattgatg	ctaaattgtg	3900
gggaagtgac	taggaaagat	acagggaaac	aaaggaaaga	taagggttat	tttactaagg	3960
tctgtttatg	taaacttgtc	tgactctcaa	tctccagtgg	taagagtggc	tctcctcttc	4020
ccagttcagg	agagggagac	accttcacaa	gaggaaatgt	atgccctgat	tttaattctga	4080
taagaagagg	gcagagaacc	cttctacat	ttgggtgttt	tcaattgcct	tcagctcaaa	4140
atggtcctta	cgtggaagtg	acataatttt	atgggtggtg	tataattctga	tctctttcaa	4200
taataattgt	gtgtcaatga	gtctcattga	ttataataga	tataccgctc	tggtgcagga	4260
tttgatagta	agagaggttg	tacttgctg	gaggcaggat	gagtaggtga	atgaactttc	4320
tataattctc	actcaatttt	tgtttgaatt	taatatgtct	ctaaacataa	aatctgttca	4380
ataaataaag	taaaaaacaa	aacaataaaa	taaagtaaaa	aaaacaaatc	tagactgagg	4440
taagtagaag	ctttattgga	aaagattatt	acagaggggg	aaggggactcc	aactataaaa	4500
cagcaagctt	ctcagagctc	aggcagaaaa	aggcttttct	tttatacgaa	ggagagaaca	4560
aggctagaaa	gaaccagatg	taagtaagtg	agggtggcctg	attggatagt	agatcagaga	4620
acgttttacc	ctcaggccag	ccgcttctctg	gaaggagcca	ttgagaagga	gctaaattct	4680
ggctcgactt	agaatgcac	agagttcaag	agtctgggga	aaagaaagaa	ccgtaaccaa	4740
aattttttca	ggtcagggtta	ataggcattc	tgttctaata	gatcactggg	gacaataaat	4800
ttagctaate	atttatgaag	ccaagaacag	gagtttggag	cgtctgtatc	tggtttgtgc	4860
ataggtaaac	aatgccggca	atccttgagc	ctcatctaac	tcatatcggg	aagggtgggc	4920
ctttggaata	agctattctc	tgaacacgtt	tagtctaagt	acaatgtatt	tcaccagcaa	4980
ataccttgat	tgtgccttgg	aaggaaaggg	ttgggtgatgg	taagtctgga	aaaggccaga	5040
caaagagagg	gagaattctg	gtccctgaag	agacatccca	tatgctctcg	ggtgagattg	5100
tgagaaatgc	tggagaatg	aagtggatga	ctaccaagcc	catcttattt	caatgtacat	5160
agataatact	tcattttgca	aaaatgaaaa	gtatgggaca	aataactaat	cagacagggc	5220
actgggataa	gaggcgcgaa	ttaggatatg	tagtcattta	ctgtactaat	ttgcatgaat	5280
agaaatgcag	agagagctct	ataaaagaaa	aataattttgt	gtaatctatt	agtatagaaa	5340
ctcgtaatga	atttactaag	atataacgat	ttattcgtca	tgtaggcaa	cgtattttca	5400
atgtcatctt	atggagaatt	gtaaagattt	tcttgtatgt	taaaggctct	tggtcaccat	5460
cattattcac	acgattgtta	ttcacctgg	tatcattatt	cctcatcatt	aaggacacta	5520
gtaatctttt	aaacctttct	ggctctttct	ggggaaaagg	gcaggcttaa	tacaggcatt	5580
cctcatttta	gttagcttca	ctagcttcac	tttattgtgc	ctcacagata	ctatgttttt	5640
cacaaattga	aggtttgtgg	caacctgccc	tcaagcaaac	ctatcagcac	cattcttcca	5700
acaacatggg	ctcatttctt	gtttccacgt	cacatttggg	taattctgac	aataatttcaa	5760
aatgtattat	tattatagct	gttatgatag	tcagtgaagca	gtgatctttg	atgttactct	5820
tgtacttgat	tggggaacca	tgaaccacaa	ccatattagg	cagcaagctt	aattgataaa	5880
tggtattgtg	agtctaaccg	ctcctccaaa	tggccattcc	cccattctct	tctctctgct	5940
caggcctccc	tattccctga	gacacaaaaa	tatttaaat	atgctaatta	ataacccaac	6000
aatagcctct	aagtgttcaa	ataacaagaa	gagtcacacg	tctctcactt	taaatcaaaa	6060
ccttggcata	attaaggtta	ctgaggaggg	catatagaaa	gctgagacag	acaaaaagct	6120
agacttctta	tgcagaacag	ttaaccatgt	tgtgaatgca	aaagaaacgt	tcttgaaaga	6180
aataaaaaat	agaagatgct	acctctata	attctatgaa	ggttcagaga	ggtgaggaag	6240
ctgcaaagga	gaagttggaa	gctagcagaa	gttggttcat	gagattttaag	gaaataagcc	6300
acctctataa	tataaaagta	caaggagaag	cagcaagtgc	tgatgtagaa	ggtgcagcaa	6360
gttatccaga	agatgtagct	aagatcattg	atgaagggtg	ctacagtaag	taacagattt	6420
ttaatgtaga	caaacaacc	ttatgttgaa	agaagacacc	atctaggact	ttcataccta	6480
tagaggagaa	gtcaatgcct	ggcttcaaag	cttcaaaaga	taagctgact	ctctgttag	6540
gggctattgt	atctgggtggc	tttaagtcaa	agccaatgtt	catttaccat	tctgaaaatc	6600

aaagagtcct	tcagaattat	gctaaatcta	ctctgcctgt	gctctgtaag	cagaacaaac	6660
tggatagcag	cacgtctgtt	tatagcatgg	cttactcaat	cttttaagcc	cactgttgaa	6720
aattaatgct	cagaaaaata	agatttattt	caaaatatta	ctgcttattg	acaatgcacc	6780
tagttgccca	agagctctga	tggagttgta	caaggagatt	aatgttgttt	tcatgcctgc	6840
taatacaaca	ttcattccaa	agcctatgga	ccaatgagca	atttggaatt	tcaaatctta	6900
ttatttaaaa	aatacatttt	gtaaggctat	agtttctata	gatagtgatt	cctctgacgg	6960
atctcagcaa	agtacagaac	aacatggcct	tgaactgtgc	tggtccacct	atacatagat	7020
ttttatctgc	ctctgccacc	cctgagacag	caaggccac	acctcctcct	cctcagtcta	7080
ctcaacgtga	agatgatgag	gatgaagacc	tttatgatga	tccacccccc	ttaataaata	7140
gtaaatatat	gttttcttcc	ttatgatttt	cttaatacat	tttcttttct	ctagcttact	7200
ttattctaaa	aatccattat	ataatacata	tgacataaaa	aatatgttaa	ttgacggttt	7260
atgttatcca	taaggcttct	gttctacatt	aggccattag	tttagttttt	gggaagtcaa	7320
agttatatac	aaattttcaa	ctgtgcaggt	ggcccaaadc	cctaactccc	acattactca	7380
agggtccact	gtaagttgaa	aactcctgga	aaggattcat	cactgtagat	attagggtct	7440
tcatgattca	tggggggaag	tcaaaatttc	aatattaaca	ggggtatggg	agaagttgat	7500
tccaaccctt	atcgatgact	ttgaagggtt	caaagattta	gtggtggaag	gagctgcaga	7560
tgtggtagaa	atagcaagag	aggcctggcg	tggtagctca	cgcctgtaat	cccagcactt	7620
tgggaggcca	acgtggtgga	tcacaaggtc	aggaaatcga	gaccatcttg	gtcaacatgg	7680
tgaacctca	tctttactaa	aatacaaaaa	attagccggg	tatggtgacg	catgcctata	7740
atcccagcta	cttgggaggc	tgaggcaggg	gaatcgcttg	aaaccaggag	gtggagggtt	7800
cagtgaactg	agatcacgcc	actgttctcc	agcctggcca	cagagcaaga	ctccatccaa	7860
aaaaaaaaaa	aaaaagaaga	aagtaagaaa	gaaagagcaa	gagaactaga	agtgggtcct	7920
gaagatgtga	ctgaattgct	acaatctcat	ggtaaaactt	gaatgaataa	ggagttgctt	7980
tgtctgaatc	agcaaagaaa	gtggttgcc	gagatggaat	ctacttctgg	tgcacaggct	8040
ttgaacattg	ttgaaataaa	gacaaaggat	ttagaatatt	gtacataaac	ttaattgata	8100
aagtggtagc	agagttcaag	aggattgact	ccaaatttga	aagaagtttt	attgtagata	8160
aaacgttaaa	tagcagcaca	ttctacagag	aaacctttca	ggaaaggaag	agtcaattta	8220
tgtggcaaac	ttcattgttg	tcataatttt	agaaattgcc	acagacacac	cagctattag	8280
caactaccac	cctgttcagt	cagcagccat	caacactgag	gcaaaatctt	ccaccagcaa	8340
aatgaatatg	agttgctgaa	gactcagaca	tctttagcat	tttttagcaa	taaagtttta	8400
actaagatag	tacattttta	agacatatgc	tattgcacat	tttatacact	acagtatagg	8460
gtaaacataa	cttttatatg	tattggaac	aaaaaaaaat	tatatgactc	actttattgt	8520
aatattggct	ttaccgtggg	gatctggaca	cagaatccac	agtatctctg	aggaatgcct	8580
gtagtttgaa	tatgagtgat	gcaaagaagg	atttttagct	attgtagtat	ctctggagt	8640
atttcccaca	atcttttttg	gttatagttg	ctcttgatc	tgttttgcct	ttcccaatag	8700
tgattgcact	tagcagaatt	tgtgctagga	aacattgtta	ttgttattga	tacaaaagta	8760
attttctgta	agaatagat	cattttgtct	gattcattat	tttagtggct	ctctgaaaag	8820
tatttttctt	catttccact	ctgaaacaga	atctagaaaa	aaaaaatacc	atcagctgag	8880
acatttagaa	acatctgtac	tttcacacaa	tgcatagaca	acaaaccctt	acagaattat	8940
ttttaggaac	tctattgaaa	ttagatattt	tctaaaagta	tttgctaaaa	gagaaatgca	9000
ttttagctta	catcatattg	tttaatttta	aatctatcat	attgtgacaa	taacaagaat	9060
aaaaacaatt	tcaccctcag	tgtaaaactt	acagattttg	ttgccgagaa	acttcaaaac	9120
atacctctga	atagtaagaa	gtgtagtaaa	atttaaatga	aatgttgggc	caaatataga	9180
atgattgaaa	aatcaccctt	gaattataaa	aatttgtgct	catcttgtga	atactttaaa	9240
tgtcttgctc	atggtaagga	cttcgtgcta	tctttgtaca	attaacatat	agggcaaatt	9300
tcaataccaa	tgacagtaga	tgtaatttca	catatcaagt	agtttcccta	ataatttagg	9360
acttttttca	ctgacttctt	gtcatatatg	attctgtcat	tactttttta	tatagctcaa	9420
aatatggctg	atggagtgtc	tgtgcccttc	catggttttt	tcatttggtg	tcactagtat	9480
tatatttatt	tattcttttt	ttccaggaca	tttgcaagac	attcgtcaag	ttttgaggtt	9540
tatttaaggt	ataactagat	aatgtattat	gtaaatcagc	ctaaccagca	cgcataaact	9600
aaaccatgtc	acaggatgcc	aatagtgaac	gacggagtgc	tgatgcctct	ggaaatctct	9660
tcacattgtg	gagtttctcc	tcttctcttg	ggatatcttc	tctattttct	gtgacagatg	9720
cctatttgct	agatcaactg	agagcatcag	tgccatatagg	taatgaaata	caaataaaac	9780
tcaatttggt	ttatgggtctg	aatgtttata	ttcctccaaa	attcatatgt	cgaaatccgg	9840
tccccagtgt	gttgggtatta	agagtgcagg	cttttgggaa	gtgacagggt	catgagagtg	9900
gagcacttgt	gaacaggatt	agtgggtctta	tgaaagatgg	agcgtctgtt	tttcccttcc	9960
cccatgggag	tgggaaaata	cggatagaag	ccactatcca	tggggaacaa	gcctccacca	10020
aaaactgaat	ctgctggcac	cttgatcttg	gacttcccag	cctccagaac	gctgaacaac	10080
acatttctgt	tgtttgtaaa	ttacctaatc	taagacagca	gcctgagtg	actaggacaa	10140
ttgggttctt	ttttgtttag	gtagcaataa	ctagaacacg	aatttctaca	gaggcaacac	10200
acacaaaaat	aggtgtgatt	ttatcaaatt	aaaatacttc	tgcacagtaa	aagaaacaat	10260

taccaaagtg	gaaaaaaaaa	aacctatggg	gtgggagaaa	atatttgtaa	accatgtata	10320
tgataaagcg	ttaatatcca	aaatgtatga	gaaattcctc	caagtcaata	gcaaacaaac	10380
caatagaaaa	ccagaagcaa	aagaaaaaag	aaaataagtt	aaccaatttg	aaaaatagac	10440
aaagaaactg	aatagacatc	tctcgaaaga	agacatacaa	atgggtcaaca	gatatctgaa	10500
aaagtgtgga	tatcactaat	caggaaaaatg	caagttaata	tcacaagata	ccacctcaca	10560
gctgttagga	taactattag	caaaaaaaac	aaaagcggcc	ccttgctttg	tgtctgccag	10620
aatccattat	ggctgccact	atagttctcc	aggagtctgg	atagcagtc	cggctgggaa	10680
tgtctggagt	ctgcaggccc	cagtcactat	ggacagtttt	tccttcagta	ttggggttct	10740
aagaggcaga	aattagaaat	aggaatgcag	atgggtgcag	agtgtgggtg	gggggttagag	10800
cggcgccctg	gaaaaaaaaa	aaaaaaaaaag	atgagtgttg	gaaatgatgt	gcagaaaagg	10860
gaaccccgat	atattattag	tggaaatgta	aattgggtgca	gcctctatga	aataatgtat	10920
atttcaatac	ataataatgt	atattccaaa	attgttttta	tagagtagat	tttagatgtt	10980
ctcaccacaa	ataaatgata	agtatgtgaa	atgatgcatg	tgttagcctg	ttttaatcat	11040
tccacaatgt	acacaagtat	aacattttaca	ttataccaca	taaatatata	ctctatatat	11100
tactatata	tatatatact	actgtttgtc	catttaagta	attttttaca	aataattgtt	11160
ctaaataata	agattatctt	ataaaggaga	cagtaggttt	gttttgctcc	agatggtaga	11220
gttaatccct	gaagctagaa	gatgcttagg	gtgaagtttg	tgtcaaaat	atttttctaa	11280
aaatagagca	gactgcctca	aagggtgaggt	atggtgcaga	aaaaagtcac	tgttaaggat	11340
gtgtgggatg	ctgtttgggg	ataatcaggg	actcaggata	gggaactaag	gagtttaggt	11400
tgtctgctgt	atagaggaaa	ctagaggaga	tagtttatcc	agccatcatt	aattcattca	11460
ttcagtaaac	attactgaac	acctgtgtgc	caggtgctgt	gattgacact	ggagatttaa	11520
agttgaataa	tatctagtca	ctggggccct	gatgtcaagg	cacatgctct	agtggaagag	11580
aggcacacgt	agctaagagt	gtgtggaaa	tgttatgagg	aaagtatgtt	caggatagtg	11640
tagaggctga	ggaagaggat	tcccatctca	cctgctgcgg	gcaaagaaag	gtggatggta	11700
ttcattaggg	attatcacag	gagggatcaa	ctgaactaat	ccttaaaata	taatacagta	11760
ggagccagcc	cacaaaatgc	taatagaggc	agtggggagg	gtggtggaaa	aaaaaaatag	11820
gtatgagtca	aaggagagaa	ttccagacag	atgaaagctg	taaaaatagt	ttagttttac	11880
tgtaaaaagg	tcaaagccag	aagtaagggg	ctgcaaagtc	catttaagct	gagaacttta	11940
ttacagctcc	tgttccaagc	ccatgctctg	tgcgactgca	gatgggatat	tctaatacca	12000
cttgtggaag	tcaatctttt	tggacaaaga	ccaagagaga	cccatgttgt	gattcactct	12060
ctatggcaac	ccaaaaggaa	gggtgaatgc	ctgcagatga	gagttgcctc	ttcatgccag	12120
ccacttaata	ggctatggaa	aaggggaagag	cagtccccat	ggtgggccta	ataggaatca	12180
taatctaact	gattatcaca	gcagtgactc	tttcaaatct	ggcctaactg	aagctagcac	12240
cacaccagga	tctctgtctg	gcacacacca	atcacccagg	aggtcagtat	catccccatt	12300
ttacagatgt	gaaagctgag	gcacaaggta	aataacagtt	atgcagaagt	cttgcttcag	12360
tgtctaactt	cctcatctca	ctttattctg	tttttcaata	ggagaagaga	tataatctca	12420
tgatgaaaag	tgccatctaa	agtggcacct	tagacacagt	aagcaaactg	gattggaagt	12480
gaagaagtca	gtctcaagat	acttgacaat	gtctcctatt	tgtagattgt	ttagcaaaat	12540
gctttcacct	gagttatttc	atttgtttgt	cacaaccaat	ctgtctggta	ggaaaggcag	12600
atgatattgt	cacctccaga	taagaaaaat	gaaattcaga	gaggccaagg	ggcttgctca	12660
agattccaca	gtgaagaatc	tacttgcagg	aattttaaca	aaggttttca	aattcagaag	12720
tccacaggat	gactgccctt	gaaatccagt	ggtaacatat	aataagggtg	ctttgagcag	12780
gtaagagagg	accaaccttt	tctaaagcct	ttctcctctt	gctctctgtg	cgccaacctg	12840
agcagcctcc	tgagagattc	tggaaacatgc	tagaattttt	cttgcttttag	ggccacagtt	12900
catgtcactc	cttcttcctg	gaatgtttct	cacaagcact	gtgtggctcc	ttcttatcct	12960
tcaaatactt	gctgaaatgt	tcccttctta	gaaaggactt	gctgaccacc	ctgatggaag	13020
tgccctctcc	agctcacctc	tttactgact	ctgtttttat	tccttttgca	agcacatttt	13080
ataatctata	tgtagtgttt	aatgcaacag	tttgctaagg	actgaaattt	gcattcctcc	13140
agcaatcaaa	accactgcg	atggtttgtg	gaggtgggac	ttttaggggg	taattaggtt	13200
tagatgaggt	tgtgaggggt	gggccccctca	taatggaatt	agtgttctta	taagaagaga	13260
cagagagaac	gaagctctct	ctctttgtgt	gtgtgtgtct	ctctctctct	aagacatggg	13320
aggacacaat	gagaagacaa	ctgtctacaa	gccaggaaaa	gggacgtcac	caggaaaccga	13380
gcctgctgac	atcttgatct	tagactttcc	aacctcccaa	actgttagaa	atcattatct	13440
gttgtttaag	ccaccagac	tgtgatattt	tgccatgatg	cctgaactga	ctaataataca	13500
gtttacctgt	acattgtttg	ttcctctaga	ttataagctc	tccaaggcag	ggaccatata	13560
gttctttaaa	agataatgtc	atggctataa	gtttactgct	atttccccat	cacatgggaa	13620
aaagaattgt	cagagtgtgc	agccattaga	gtttattttac	cttaagggtg	tctgacgaaa	13680
tgatctgatt	atctccttaa	cttcctaaat	agctcaaact	tcataatgtt	tgaatttgga	13740
aatatctgtt	gattcaagag	agacagatta	catggtctct	cttgaacttt	ggacatctct	13800
tccacaaact	tccaaatatt	aggccatcta	ttaccagacc	ttctgaactt	aaacaaatct	13860
tactttcaac	cccactttag	ccaccttcac	catggcctaa	cttttgccct	tgttattttgt	13920

tatccatagt	tttctaaact	tagaatcctt	aatcaactggc	attatagttg	tttacttcaa	13980
cctccaacac	ttacaggtcc	ctcaaatacct	cattttattca	acccttgttt	tttgtcatct	14040
tataaaactgc	aaattcttga	aaatcacatg	tcctctctat	aaatcagatc	ccttcatatc	14100
atgttttcaa	tctgagtttg	gctccatgat	taatcttcca	aatcaaaggt	ttacaaactg	14160
tggcctgtgg	aataaatctg	gtctgcacct	gttttagtaa	ataaagtttt	actggaacac	14220
agtcatactc	atccattttac	atatttttta	tatccatttt	catgttacia	tggcacaggt	14280
gactagtggc	aacagaatct	gtgtgggtca	caaagcctag	aacattttct	atctgtcctt	14340
tatagaaaga	agttaccaac	ccctaataca	aatcaataat	ctatgaagta	gaaagtaaaa	14400
gatgatcttt	tgggtgctgac	atgaacatat	tatgattaaa	aaaacacttt	attgaggttt	14460
gattagcata	cagaagctgt	acataacaca	tacaacttga	agggtttggg	gatcagtatg	14520
caatcatgaa	accatcacta	tgatgaatgt	cataaacata	tccatcaccc	ccaaaggttt	14580
cctctgcctt	atttattttat	ctattttata	gaatacttaa	cataagattt	gctaccctct	14640
tagcaaattt	agcaactcct	tgagtatgca	atataatata	gttaactata	gacactatgc	14700
tgacacataga	tttctaggac	ttatttaact	tgcgtagcaa	aaactttggt	ctctttgccc	14760
aataacctccc	tgtttccccc	tccttcatcc	cctggtaacc	atcattctac	tgtctgcttc	14820
tatgagtttg	actgttttag	atttctcata	tgagtgggat	catttagtat	ttgtccttcg	14880
atgtctggct	tatttcactt	agcataatgt	tcaccaggtt	tatcaatggt	gttgcaaatg	14940
acaagattcc	tttttttatt	attaaggcta	gataatatcc	cattgcaaat	atataccaca	15000
ttttctttgt	ctattcatct	gtagatgaac	attcaggttg	ctttcatgtc	ttgggtattg	15060
cgagtaatgt	tgcaatggac	atagaagaac	aggtatctct	ttgacatact	aatttcatgc	15120
ctttgggtaa	ataccagaa	gtgaaattac	taaatcacat	ggtagttcta	ttttttgggt	15180
tttgaggagc	ctccatacag	ttttccataa	tggtgttact	aatttacatt	ctcaccaaca	15240
gtgtaaaaag	gttccttttt	ctccacttct	tcaccaacat	ttaaatttca	tctttttgat	15300
aatagccatt	ctgacagatc	tgaggtgata	tttaattgtg	gtttcaattt	gcatttcccc	15360
aatgattagt	gatagggata	ttgggctttg	ttattaataa	cccactggcc	atztatgtgt	15420
cttcttttga	gaaatatctg	ttcaagtcac	ttgactatct	ttaatgtaat	cacttggttt	15480
cttattattg	agttgtttta	tttctctctg	tattttggat	attagagccc	cttaacagat	15540
gtattatttg	cacatatttt	tctcttaacc	tatgggttgt	ctctttattt	tgtaaattgt	15600
ttccttttgc	gtgcgtaagc	tttttagttt	gatgcaatac	aataactaat	agattaatgg	15660
tctctttttg	tttttgttgc	ctgtgctttt	agggctcatg	tcaaaaaaat	ctttgtccag	15720
atcagtgtgt	ggagcttttc	tcttacgttt	tttatttcta	atagttttat	agtttcagat	15780
cttatattta	agttctcaac	ccactttgag	ttgattcttg	tatatgggat	gtgatgtgtt	15840
caatttcatt	cttcttcatg	aggacatcca	gttttttcaa	caccatttgt	tgagaattca	15900
ataaagtgtg	agtatataaa	atcaacgtaa	aaaatcagta	ggtttttcat	acactgacaa	15960
tgaactatct	ggaaaagaaa	ttaagaaaat	aatcccatct	taataccata	gcaaaaatac	16020
tagtggtaaa	tttaaccaag	caagaatatg	gaaatatcta	tattctaaaa	actataaaac	16080
attgatcaaa	gaaattgaag	attacacaaa	taaatggaaa	tatgtctcat	atccataagt	16140
tgaaaaccga	tattgttgaa	atgtccatac	taaatgtgat	ctaaagagtc	catgtgattt	16200
ctataaaaaa	tccaacgtca	tctttttatc	aaattgaaaa	aaaaatgcta	aaatttatat	16260
ggaaccagag	aagacctgaa	taaacaaagc	aatcttgagc	cataagaaca	aagttggagg	16320
catcacacta	cctgattttca	aaatatattg	caaaactata	gtaatcaaaa	tggcatagta	16380
ctggtgaaca	aacagatata	aaaaccaatg	gaataggata	gagagcccag	gaataaatcc	16440
acaaattaaa	aaccaactga	tttttgacaa	agatgccaa	aatggggaag	ggagagtttc	16500
tttcattatt	tattctctgt	attcttttca	tctcaggaaa	aatatccagt	ttcctcaatt	16560
gtatatccat	tcaactcccc	gataatttct	tcaattcact	atattgtttat	tcatttaatc	16620
tcaattgttt	gttcattctg	taaatattca	gatttctttt	tatgcatttt	ctcagaattg	16680
gaagcataat	actgaacaaa	ataactataa	atctcagcct	cccactccca	tatttacagt	16740
ttgattaggg	aggcacattt	agatatgcag	tgataattgc	tttgcttaga	gaaattcaag	16800
gtgatagaat	gcattggtgac	acctaaccac	gactggtaga	gaaagggaat	tcttccactg	16860
ggaatgacat	gattatctaa	ataagtaggc	tcaatcaggt	caggaaagg	cctgaaagac	16920
tatttcaagc	agagggaagg	tatttgccaa	ggccagggtg	tgtagtggag	agaatgggca	16980
gtggcagaga	attatgaagt	gttccaatga	ctaaaagtaa	agtaacaagt	tccttgtcca	17040
agagcttggg	ttgtatccta	actgaaatga	gtatacacta	agtgtttgaa	gaagagggat	17100
gaaatgggtca	agttttcatt	acacaaaaat	aacctgttcc	tttcattttt	tgtttatttt	17160
tttttttaat	tttctgactg	ctcctttctg	gaaatctcaa	atztatattt	gccaaatatt	17220
gtcacatttt	cgatggagaa	tacaaactaa	gaatgggtta	gggaactgag	tcagaaagtc	17280
cctgttgtac	aattcaatca	tgtttttcta	aggatgtgct	tttggacatt	atggaaacta	17340
tcttaggctc	tcaattggat	cctagaaaag	aaggcacctg	ttaaaaggaa	tgtccagccc	17400
cacctaat	ggcagctgcc	cctccaagct	aatgacatta	gggatgtagt	ggattcaaga	17460
ggctgatgat	gccatctggt	caactcatgt	gactatttct	atcagcttta	tttctgcaac	17520
atcctgtgcc	cacagagggg	acacaaaaat	gctttataat	tccttcacca	tgtgaagata	17580

cagatacacc	cagaacctta	aaggcaagaa	tatgattgaa	atgtcaaagt	gggacttggt	17640
gatctaaatt	atgtccccc	aaagccaatg	tcttgccacc	accagtgcc	tatgggtgga	17700
gtttctaaac	agattactca	aaacacaaac	tttcaaaaag	ggaaagtc	aaccctctag	17760
tcatcagggc	aattacggaa	taacattgct	ggagtaaggt	tttctcaatg	cccaagagat	17820
gagctggcaa	tgccacaaca	atgtccaatt	cttagtgggt	ccaagaccat	gtgttacatt	17880
tccctcccat	gattactcac	agcttcacag	ttctgctgtc	ctcttcgcct	ctctgccacc	17940
tcttaactgc	acctttgacc	tcctacccct	aagattcaac	cctgtgagat	tacttgtctt	18000
ttcatctaca	ctctggtcac	tctgaccccc	attctttaga	ttcaggattt	tctcttttcc	18060
tggcttactc	tccagcacia	gtagaaaaat	attgtgcttc	attggaaaa	gcatgttggt	18120
tgaatcacac	tctttcagat	tatacaattg	tagtctttca	ttatctttga	gccattttat	18180
aatgctgtaa	actaatatta	atgcatataa	ttcttgtcta	tagacatgta	aattgtgtcc	18240
aggggtgatt	taattgactt	ttcttcccc	ttgtggaaaa	ggctagtagt	tttgcctcca	18300
tttgcccat	tatttctata	ttcctgatct	gtaaatgc	ttctgggttt	tcctttggac	18360
tcattctaac	atctcttctt	tttctctcatt	aattaattaa	attaaatctt	taacacctct	18420
tcatattttt	gtcagtatga	gagtataaaa	aaaaatcttt	ttttgaaaat	tatcttttaa	18480
taggtgacaa	aaagaataaa	cactgaacaa	gaagtcctga	gctcctacta	caatccaaag	18540
tgaactagtt	gcagtgccat	aggatgatgt	ttctcctgag	acttctttac	taggatcatt	18600
tagactgtta	taaatgcaga	ttcctggggc	tctccttaga	cccactgaat	cagacactct	18660
aagaggagg	tcagaaatct	ccatcttagc	aagttaatgt	ccaggagagt	tagagactca	18720
ctatgacgta	tgataaaaa	tagaaggaag	gactttccag	caccttaaac	tacttggaaa	18780
gatagattaa	tgtctagatg	aaacttgatg	agactttaga	ctaataatgt	acataacat	18840
caggagatgc	atgtataaac	caaataccaa	taacccaaag	caaaatttct	tttagaaata	18900
gtgataaata	aatgagtga	gagtttgtca	ctcacatccc	ccaggtaaaa	catacctttt	18960
tagcctaaat	aaatgctata	gtttgaatgt	gttccctcca	aaattcaggt	actgatagtt	19020
aacagcagcc	gatatgatgg	tattacacaa	aatagtaggt	tcttttaggag	gttggttaggc	19080
catgagtgcc	tcccccaaga	atgggatgaa	ggacctcata	aaaggagcct	ctcacagcat	19140
tgtgacctct	agccctccca	ccttccataa	gtgaggacac	agtgttctct	ccctctggag	19200
gatgcagcaa	caaagtgtca	tcttggaagc	agaggagcca	tcaccagaca	acagaaccag	19260
ccaacagctt	gatcctggac	atctcatttt	ctccagaact	gggataaaa	aaattcctgt	19320
tttttaataa	atttcccaat	ttcaggtatt	ttgtgatagc	agcaaaaaca	gactaagaca	19380
actagtatga	aaataacat	taacaaataa	aattaatcat	aataatgttt	tgcttttaaa	19440
agaaaaataa	taagccaata	tgctttctgt	tgattgattg	atttactaaa	cattgatagg	19500
ccatctccac	tggggatatg	gcatttaaga	gatctctttg	atcttagtac	ttttactgct	19560
ttttaaatag	gatcaaatat	acccaaggta	aaaaatagaa	cacactatac	gttacatttt	19620
ggaactgtta	gaaattcctt	tgaagctaaa	attactgcta	tcatttgaca	acttttacc	19680
ctaaaataat	gtgggtgctca	ccagcttgct	taagttacag	cacttgctgt	cttctcagat	19740
acaatatcag	aaacttataa	tccaagaaaa	atctaaatgg	caagtgtgag	ttaatggaag	19800
cctcataaag	caagaggtgt	tttggaagt	tatggaagac	atcaataaat	gatatgtata	19860
acatcaagt	caaaagtgtg	tgctaggaaa	gttcaaaaaa	gaaaaaaaat	atggtaaggt	19920
aagaccagag	attggggagt	atatagcttt	tgggaattca	ggaaatgcta	acgtctatgg	19980
agaatttgca	taggtgaaa	atcagattgg	aagcctttct	ctgtggaagc	attgtggatc	20040
taagtccaga	agtgatcctg	agccacctac	tatggaaagg	tgtcagtga	caagagactg	20100
tctgacaaa	gtggaagctg	agcagacttc	tactgcgc	cgcctatgta	caggccagat	20160
tccaagggt	gatattacac	tgctagtttg	atctttctca	gatagctggg	tagagaggga	20220
aaagtttcac	cccaaatacc	agatgcctcc	aaacatctag	atgcttggtg	ttatgtattt	20280
caaacagcag	gtttgataaa	gctgttttta	ctctccagtt	ggatgttggt	tgtcaaggct	20340
gtcattaact	tctgggagtt	cccaaatacc	cagagagaga	gaaataagca	gttctgcctg	20400
caatcagaag	ctggctttga	gccccacatg	gcaacatgcc	tgtatttaat	tagaagtggc	20460
ttagttcttg	ggatttactg	tggattgagt	ctacgaagct	ggaagggttt	atcctgagta	20520
ttccactgtc	tcctcatgtg	gattttatca	aatttctcct	ttaaccatat	aagcgctttc	20580
aaagttgaca	tatcacactt	agttatgagg	ggaacctaga	agtatgattg	tgacgtgcat	20640
tgttgagaga	ccatgctttg	taagtgttaa	taattaaaat	tattttgaac	actaattatg	20700
ctgaagactg	gattttctac	tccttcagat	gttttccagt	ggacataatt	tgccctttta	20760
ataaattgat	tgcaacttta	tcttcatttc	acctttatga	tgttatacct	tttcattttt	20820
gtcttctcat	agctagggtc	tcctgggtcc	caaagttaga	cacacatctt	acaccaatcc	20880
cagagccatt	ttgtataaga	gccaccatgg	atttaaccag	ctttagctcc	agtatttgaa	20940
cataatgttc	agcatcatca	cctggccacc	aaatcaaaac	tgagcaccct	ttaatccatc	21000
aacaagttct	ctgcagccat	gcaagggttat	gaaatgggca	cagacatcaa	tatacagttc	21060
ttgtgtttta	gaggttcatg	gtctacctga	gaaatgcac	tttaaaccct	aagtagacgc	21120
tctgtttatt	ccataaatga	tttttaagca	tcaatgggat	atcaagcact	gtactggctt	21180
ctgggctata	ataaatatat	aaagaccaca	agtttgaatt	tcatgacatt	gaactataat	21240

gtttaaatgt	tataataatc	atagtaaatg	tccttgagga	gctacggaag	attcctgcat	21300
gaagcagaaa	caagaagctg	aagaaaaaac	aactggcttt	gggggctata	taaatataac	21360
cctcaaatta	aaaactcaat	agtttgattg	attaccaa	caatacagcc	agaaaataaa	21420
tgtattaccc	caaaagcttg	agtgaagaaa	gacttctgta	agttactaga	aagcgctaag	21480
gaagaaataa	agaatgata	ttcctgagag	ccagggtgctt	tctgtaggac	acacagatac	21540
agaatgaggg	aaaacagaaa	attctatggt	tgtggattca	aaatagaggc	atcaccatgt	21600
ctatctcatc	agggtttaac	cagaggaaac	caaaccagta	agatatctgt	attaagagat	21660
ttcttgcaaa	gaattgacct	atgtgattgt	gggcactggc	taggtaaatc	ctaagtccac	21720
agagcaggca	ggaagaaaac	aggctgggac	ttgtaggcac	aggatgaagc	tgcaatactc	21780
atgtggatgc	tgctcttctt	cagggaagac	ttggctctgc	tctcaaggac	tttcagggtga	21840
ttaaatcagg	cccactcaca	ttatctaaaa	taatctccct	tactcaaaac	caagtgatta	21900
tggaacttta	tcacatctat	aaaatatcat	tatagtaaca	cctaaattag	tgtttgaata	21960
actgagagtt	gtaactgata	tggtttggct	atgtcccccac	ccaaatctca	tcttgaatta	22020
tagttcttat	aatccccatg	tgttgtggga	ggaaccagggt	ggagataatt	gaatcatagg	22080
ggcagtttct	cccattctgt	tctcatgata	gtgagttagt	tctcaggaga	tctcatgggt	22140
ttataaaggg	cttccctctt	tgctccactt	tcattctcct	tcttccctgct	gccatgtgaa	22200
gaaggacatg	tttgcttccc	cttctgccat	gattgtgaagt	ttcctgcggc	ctccccagcc	22260
acgctgaact	gtgagtcaat	taaatctctt	ttctgtataa	attaccaggt	cttgggcagt	22320
tctttatagc	agcatgagaa	tggaactaata	cagtaacttt	accaagtggg	cacataaaac	22380
tgatcattac	aatgtacagt	gaatatttgg	tgagtttaata	gatataattca	taactgaatg	22440
aaagaggatg	gtgattccta	cttcagggtg	gtattatgag	agttaaaagg	gttagcatag	22500
atagaacact	tttctatgat	tgatctaagg	ttggcttttta	gggactacat	tatacatatg	22560
ggttttgtct	tcatgttttt	gcttgcattt	ctctctctgc	ccactcttgg	ggctattttg	22620
atgtcacctc	ttcttcttct	gagctccagt	ctatgtattc	cctgttccat	catcacacc	22680
actgaagtct	acactcctca	ttagattcag	ggaaccagga	tgccagccaa	agacatccct	22740
ttccctcatg	accagttga	aaaaattctc	cacccttctg	ttaattctag	aatatttagg	22800
agtaaagatc	ttttccctca	gatattctgtc	aattcccggc	cttataatctc	aaagcccctt	22860
ttcatccagg	tttcctagat	catttcttgt	cgtccctgtg	ctataactca	ttaaccagc	22920
cataaaaactc	atcagctact	aatgtcctct	tccatcccag	actccctca	caccaataat	22980
tcttgaatag	agcaacccaa	attgagaatt	ttaagtaaaa	ttgaaggtaa	atttaattgc	23040
aggaggacat	acaacattaa	ccttttaata	gagatcctaa	ttcttaaaaa	aaaaagtcct	23100
cattgatctc	ggggccggtt	cacgcagggtg	cttttctaaa	tcatcaagggt	tatcttcagc	23160
ttcactttca	gtctccttct	tcggctcttg	cgtgccact	ggctcttctc	gtgctaattgt	23220
ggttgtggca	gctgcaatgg	ctgcagggtg	cacagcagct	gcagcagctg	cagccacagc	23280
cttttccttc	ttgtgttttt	tgtgcttctt	gtgcttctta	tcctttttgt	gtttcctgtc	23340
cttctttttc	ttctttttct	ttccatctcc	ttcttgatct	gaatttcttg	gcagtgatgg	23400
gctcgggtgt	gggcttttgg	cctttttgac	cggtagagct	ggtgaccagt	ttgtagacgg	23460
tgactgagac	tggaatgggg	atggagatgc	tgggggcttt	ttagctgctg	gctcaggaga	23520
cccagagaca	gatcgggagg	atgagaccct	ccttacagac	tgtgggcttg	gggaagcagc	23580
ctttttttat	cttttttaggt	tccggagtcc	tggaagactct	cctaattggg	ctagtacttg	23640
gagacgggga	ctgccttctt	tggggtgatg	acgacgctcc	tcttcgaacg	ggtggaggac	23700
ttgaggtctg	aggagctcga	ggcgtgatg	agggcgaaatg	ccatttggtt	ggatgcgggtg	23760
atcgggcctc	ccgggtagag	cggcttgggg	aagacccttt	cctatgcttg	gatgatagtg	23820
aaggtgaacg	tctcttggtg	actggggagc	ttctttggaa	ggtggagaat	gggagacccg	23880
ccgctttggg	ggtggaatg	gtgatgctct	tcgttttagga	gggggaggag	gtgaagccgt	23940
tcttctcttt	ggaggtggag	aaggagagta	tctcctctgt	attggaggag	agtatcttct	24000
aggagaaggt	gagcgccgac	gtaggggagg	agaaggagtc	cttcgctcgtg	gtggtggtgg	24060
tgtgggagtc	ctgagccgtc	gaggatgagg	ggcgggagaa	ggagaccgtc	gccttctggt	24120
gggtggtggg	gatggacttc	tcctccgtct	accatgaggg	gaagtctctt	tttgggtgctt	24180
tcgtggtgat	ggagaagcac	tccgggaagg	ggaatgtctc	cgcgcgttgc	caacctcacc	24240
attcttcaca	tgggatctct	tgggtcggtc	atcttctgag	gagaaggagg	aaccagagtc	24300
agatgaagac	tgctgggttt	gtcgtctgta	ttggcgtctc	tgctgagatc	aatgagggaag	24360
gcccagtgt	ccccacagtc	ttagggggaa	atgtttgtta	tgatgtaaat	tttatttggt	24420
ttgtacgcag	ttcaatttca	aaattgctaa	aatgtgtttg	agcttttagac	tataacattt	24480
gttgtaataa	ttgctaggtt	gaagttcaac	atgtaaaaaa	agggggcatg	gattttacatt	24540
gcaaaagggtg	tccacagtgt	attagtgaca	ttctttcatt	gacagctgac	ataattcatt	24600
gagtgaataa	ttttaagcca	aaaaaaattc	ccttttttaa	aaagggggtt	taaatactgt	24660
tgacactttt	atggttcctt	taaatgctct	ggctattccc	agaggggttt	ttttgtttgt	24720
ttttttgggt	ttgatttgct	ttttgttttt	ctttcttctt	cttacatttt	tttccatttg	24780
agtcttagct	cccatttaag	ttatgcttct	gaccttgat	ggtctgtaag	cttgcccaga	24840
aataagacca	ctgttttgaa	ctaccacaaa	agtataaatg	aatattttta	tgccacagtc	24900

tttctgttg	cctgtggagt	ctctgctgaa	atgaatcagg	attcgagctc	taggataaga	24960
cagaaaatga	aagcatgttg	tttgccagga	cactgtgggt	ttatatattgat	gtgtaacaac	25020
ttgattttgga	acactggact	ctcattctgt	tcttctgggt	ttgttttttt	gttttgtttt	25080
tttcttttgt	aaaggccatg	aactagtccc	agaaaggatt	ccttcagtta	catacaattt	25140
gtttaatgaa	atgtcatggc	tctgttcata	tttttgtctt	gttcttccaa	ttgggtgtata	25200
caactttcag	agcctcttgt	atlttggaagg	ctggaagggc	ccagactttg	gaatagtgtc	25260
tcgggtttcac	tgtttttgtt	ttgatttttt	ttttttattt	tttttaaact	aaagctatat	25320
aaagcttgtg	gattaaacag	aataaatttc	taaattttaa	aattttaaaa	aaaaaaagtc	25380
tattgtcttc	cctcccctac	cctaagcaat	atgcaatagt	ggctcttcaa	tagtcccaga	25440
ctcttcttct	cttccctggac	tgcccatctc	ctgatcaacc	cttaatttct	cttcccttctc	25500
tcacccttct	tttcaggatt	gaattaatga	atcctttctt	ctcactcatg	cagagtaagt	25560
ttctgctctc	ctgggtcttt	ctgtttactg	accgcaacaa	cttcagatta	tacctcttct	25620
actccaagtg	ctttcaaaga	aagtcctctg	ccaagacaaa	ttcattacgt	tttttccctc	25680
tacctgtttg	cctttattct	cttttgtatt	tcactcttct	atctagattg	aataatcttt	25740
gagagcacag	atgtttattt	atatttttcc	tttccatttc	tactcagcat	gaggtgtcca	25800
ttgaacaaac	ttgatgaatt	tttattgctt	aatatcttgc	tagaggtggg	gagagaggtt	25860
gggggctgtt	aaggaactat	cagctagcct	aggagatatt	agagctgcag	agatttggct	25920
atcttgttca	acgttatatc	cctagggatt	agtacatagg	cttgcaaata	gcagggtatga	25980
ataaaaaatt	attgaatgag	taaatgaatt	taaaatataa	gttacttagg	cggtatcttc	26040
aggcatatct	gtgtttatgt	ggtattcaat	ggcccacaaa	tgtctacatc	ctaattccta	26100
agatctgtaa	acattaattt	gcatgacaaa	agagacttta	cagatgtgat	taaatgaaag	26160
gattttgaca	tgagataat	atcctgtatt	cttcatgtgg	aaccaatgta	tttacaaggg	26220
tccttataag	taaaacagag	aagcaggaaa	atgagggtcg	caaaaaaaaa	aaaaaaaaaca	26280
aacatgaaga	cagagaagag	gttagagtga	tggtggcttt	agagatggaa	ggagtcacaa	26340
gctgtcttaa	aggaataaga	caagctgtct	taaaggaatt	gttataaagg	aatagctgaa	26400
gctgggtaat	ttatttttaa	aaggtttatt	ttgctcacta	ttctcatgtc	tggaaaagtt	26460
taatattggg	tagctgcac	tggcaagggc	ctcaggctgt	ttccactcat	gtcagaacgt	26520
aaaggggagc	tggtgtgttt	agagatcacg	tggggagaga	ggaagcaaga	gagagggagg	26580
agggggccagg	ctttttttta	acaaccagct	ctttttttta	aaaaaaatta	tactttaagt	26640
tctggtatac	atgtgcagaa	tgtgcagggt	tgctacatag	gaatacacat	gccatagtgg	26700
tttgctgcac	ccattaacct	gtcttctaca	ttaggtattt	ctcctaattgc	tatcccttcc	26760
ttaaccccca	aaccctgaca	ggccctgggt	tgtgatgttc	ccctccctgt	gtccatgtgt	26820
tgtcattgtt	catctcccac	ttatgaatga	taacatgcag	tggtgggttt	tctgtccttg	26880
tgatagtttg	ctgagaatga	tgattttccag	cttcatccat	gtccctgcaa	aggacatgaa	26940
cttatccttt	ttatggttgc	atagtattcc	atgggtgtata	tgtgccacat	tttctttatc	27000
cactctatca	ttgatgggca	tttgggttgg	ttccaagtct	ttgctattat	gaacagtgtc	27060
gcagtaaaca	tacgtgtgca	tgtgtctttg	tagtacagtg	atttataatc	ctttgggtat	27120
atacacagta	atggaattgc	tgagtcaaat	ggtatttctg	gttctagatc	cttgagggaat	27180
tgcaacattg	tcttccacaa	tggacaacac	agttctctta	agaataaaaag	tgagaactca	27240
cttccctggc	cccagagaga	gcaaaagcaa	ttcatcccca	tcacccaaac	acctccctatt	27300
aggccctacc	tccaacattg	ggatcaaatt	tcaacatgag	gttttttaggg	gacaaacatc	27360
caaactatgt	cacaagacaa	ttcatgtaag	cagcctctaa	aagatggaag	aggcaaggaa	27420
acagattctc	ccctaaagcc	cacagaagga	aagcagccct	gccaacttct	tgattttaac	27480
ccagtaagac	ccgttttggg	ctttggacac	caagagctat	aagatgatta	tggtgtttta	27540
agccattaag	cttgaggcaa	tttgttacag	cagcaatttg	aaactaatac	agatcacatt	27600
ctaattcaat	tagtattgtt	cccagttctc	tggacctcag	atttcttttc	tgaaaaacat	27660
taaaaataat	acctgaaagt	tttgcacacg	agtgcagagt	gcctattttac	tagagagatc	27720
agcatttgtt	taggctctga	atagatttga	ggatgaaatt	aaatagcata	aataaagttc	27780
ctagtgatgc	ttctgataaa	aaaatatctc	cttcaaaatg	ccagaggcag	gtcctaataa	27840
cccacaagc	aggtgaactg	gcaaaagact	gtaaaaagca	aagtagaggt	tctcttctca	27900
agacttttct	ctccatctaa	ttaggaataa	atagtaacct	atcttagaaa	caaaatttat	27960
tcaaagacct	gtgctaacat	tctgaaatat	ctgtagccg	taataaataa	atcgatgtac	28020
tttatgttct	tagctccac	aatttaacct	aaatatgtgc	cctggcatgc	ttatactggg	28080
ccaagcaagc	attaggtcat	agcctgttcc	tcttctttat	tttaagggtg	ttttaccttt	28140
gtcagcatgc	cacaagttac	ttcctccttc	ctttgttctc	ctctgccttt	gactctttta	28200
aagagtccta	agttgctagc	caatcaggac	aaatacagaa	tgtgaggtcc	cgttttcagcc	28260
aatggaaagt	ggacacagca	ggaagggtgga	tgggtcagg	tataaatgac	cctgtctcct	28320
ttgttcggtg	tactcttgtg	gcaaaactgc	tggcaagtgt	accttttctg	caataggtaa	28380
aaactgcctt	gctgaggaaa	ttaaatttat	gttcaagtgc	tatttcttta	tggcaccggg	28440
gaacaagcat	ttctaacaag	actatgtaat	ttaatctcag	gaacctaaaa	aagtgggatg	28500
aagaactgag	gttgctaata	aatctataca	acttataagt	aaatatttta	tttactaaca	28560

tataataata	aagacatcat	tgtaagacaa	tggtaaaaaca	ttttacacat	tttaaagtgtg	28620
caatagtaaa	tccttcacta	ttcagggtatt	atgttggaatc	ccttggtcacc	agaagctctt	28680
aaggaaataa	cttctacttc	gttgcaataa	tggtcttggtc	ttagttgagg	taatgcaaat	28740
actagaatac	ttgtttgttt	aacagcttat	tcttcctga	agctgttcct	ccagtcctctg	28800
ccagtgggat	cttatgtctc	caggagtact	taacacccct	aatagcccca	tcttttaagc	28860
ctccctggga	cctgccctcg	cagtacctct	tatacctact	ccacttcctc	ctcatggcct	28920
cctgcagaat	gccattctaa	aattagggttc	tattttcctc	gcccgcattc	tcttttgcaa	28980
agcctccaaa	aaatttactt	tgcttctctg	cgctgtcttt	atctctatctt	tctacactcg	29040
ctccttcttt	ttctaattat	ctataatagg	cgtcacaaaa	tttgcatctg	ttggaacca	29100
aatttccatg	gttgccctcaa	aatatacaga	tgtaaatttg	catataatta	aattttgcat	29160
aagggaact	ctcatttggtg	gagatatgca	atgcccaata	aatggcagtt	tccttcaatg	29220
tccccaggcc	agcctcccag	tctgtgtgtt	tccccctggc	tgcaagctacc	aggactctgc	29280
tctggggatt	tacggacaag	ggtatcaagt	tttaattaaa	ctaaccctct	caaactgaat	29340
gagtggctta	aaatcttctt	gtaaagaaac	cgcaaaataa	taatgctggc	attgagaagt	29400
aagaaaagag	cgagccagca	ccccacccc	ccaaatcctg	tgacaagggtg	tatttttggtg	29460
ttttgttttt	tttctttggc	agcatttatgg	gggaaaagca	atgatgatct	aatgagatct	29520
gataagaagt	tagcccaaaa	caaggaaatt	gttgagggtt	ctctttgaag	tatggattta	29580
taccaccaa	ccttagctgc	gaaccttacc	tcaagtgtta	cctgtgcctt	gagatgttct	29640
ctggtcatag	tactaagcta	tcataatgag	caagacattc	aataagcaag	tgtgatggct	29700
atgaggacag	atcttaacag	gttttttttt	ctggaaggct	taaaatcatg	cattactcaa	29760
tctaatactt	cacgaaattt	cagtaaaacc	taatgataat	atagaagctt	gtgttgtagt	29820
tttgtaatca	acagcaaaac	ataaaattta	aaaaaaacat	acattactgg	ggctgtatcc	29880
tgctacaata	ataaggctga	cataatagat	ggagaacaat	atggtaacaa	gccaaaatgt	29940
attacttcat	ccacaaatag	tatcgtgcta	tatatagaca	gacttggtta	aaatttaaag	30000
aaatacacaa	tcaattacac	aatagaaaat	ttgctatatg	gtgcatgggtg	gcatgcaact	30060
acaaatgttt	ctaacatggt	tctcttcata	ggattttctg	aattttcatt	taatattcaa	30120
gcacatcaaa	aacacctttt	cagggtgtgat	cctatacagc	aaagctgtcc	tcacaaacaa	30180
tagttgacta	aataaacaca	tggcttttatg	gaagaaatgt	gtaagtatag	ccattgtttg	30240
agcagatgct	ctgcttaaaa	agaaaaaaa	taagttaaag	ttatagatct	caccatgctt	30300
atttactgct	taagtcatag	ccaattttatt	gcaccaaaag	tgaagttcaa	agcataaaga	30360
atactatata	taatgcaatt	aatgagggtg	atgtccctaa	aacaagagag	aattagtaaa	30420
tgttttacaa	tagttttcat	gagatgggaa	atgacaatag	aatctttgt	tacaatgcag	30480
attttattgt	ggaaatgatc	tcatggcaaa	gttttttaaag	aggctgcaga	acaataaaga	30540
gagataaac	attttgcttt	tatgaaaaag	ccgatgttcc	aaatttgctg	accttctctg	30600
taaataagaa	gtgactgtca	gtagccagca	gatgtgttta	taaaaataac	ccacttgatc	30660
tgtccctgca	aggtaagagt	ggcatttttca	aagtgcata	gagaactgct	ctcaagagaa	30720
aatcatactg	taaagacagc	atttcaaaaa	cattttatttg	gaaatattta	aatgatgtga	30780
tctattttatt	tatttggtca	aaaatgccca	acttgccctaa	cattattttta	ttccttaaaa	30840
cacagatcaa	gcaacagatc	ctacagttag	actcttctga	tcaccctgag	agttagttac	30900
ccctttcttt	ggtttcttct	ctttaacgtc	ctgcatatat	gctgccattg	tgtttatagc	30960
acaatatcag	agtgtgttat	ttatacgtgc	atctccctca	ttagattggg	agtttcttga	31020
tgtcagaaca	cagcaattgt	cccaccagaa	taaaaagaat	attgccaaaa	gatgctctat	31080
aaatgttttg	ccattgggtt	aaaaaacaaa	taatggacca	atgggctcaa	aagcaaaactg	31140
gctaataataa	aaataataat	aacaaccaca	ataataaaat	aagggtcaac	acttaaaccgg	31200
ttgtgtactc	actatgtact	aggcactgat	caaagtactt	tgtacatatt	ttcttattta	31260
atattcgcta	ccatcatatt	acaatatact	gttattaacc	ccaatgtata	gatgtagggtg	31320
aagaaacttg	tcacaaatca	tacagctagt	tgtctgagat	gcaatccatg	tgatttggtc	31380
acagagctca	ggttctgtga	agcgggtaaa	aacaaaattt	ggcatccagt	ttcaaaagga	31440
gaattgcaaa	ctaatagaac	atatagcaca	aaatgattat	atcaatagaa	tgctaattgc	31500
atatcaagga	tatttggtat	aatacaaaat	attctacctt	aaacatatgg	aaatttggtg	31560
tccatgatgt	tgtagattct	atcttcccac	tctgcatttt	caaaggcata	tggtattgac	31620
tcattcgatt	aattggttga	tagtctttat	tatagactaa	atcatagaat	aaatacatgg	31680
atacatgcac	gaatattata	tctcaagggtc	tttacaatagt	tcattatctc	acttcatagt	31740
caaaacaaac	ctactgatag	ttccaatgca	aagcctagaa	cgctttggct	tagagaggcc	31800
caagtctttt	ctcagtgtctg	cactgctggt	acgtggcgtg	gtccctctc	ttctctcagt	31860
acacactacc	catgcagact	atcactctca	gtcttggtta	tctcaaatac	agagggtata	31920
actaactgga	atgtatccag	aacagtggag	ccaaagtgtg	gggaagctcc	ttaaccatgc	31980
tgctgcatga	ggaacagctg	gagagactga	gaacatgagg	cctaaagagg	agactcaggg	32040
agatgggatc	acaatcttca	aatattttaa	agacatcaag	gggaaaagag	attaacaag	32100
gtaatgtagc	tctagagagc	aaatccaaga	gtgttgagtg	gaagtgaag	ggaggctggt	32160
ttcagtcaga	tagtaggaag	aactttctag	tatttggtac	tacaatggga	aagactattt	32220

tgtgagattt	ttttaaat	tttttaatt	atactttaag	ttctagggta	catgtgcaca	32280
acgtgcaggt	ttgttacata	cgtatacatg	tgccatgttg	gtgtgctgta	cccattaact	32340
catcatttaa	cattaggtat	atctccta	gctatccctc	cccgtccccc	ccaccccaca	32400
acaggccccc	gtgtgtgatg	ttcccccttc	tgtgtccaag	tgttctcact	gttcaattcc	32460
cacctatgag	tgagaacatg	cggtgttttg	ttttttgtcc	ttgcaatagt	ttgctgagaa	32520
tgatgggttc	cagcttcctc	catgttccta	caaaggacat	gaactcatca	ttttttatgg	32580
ctgcatagta	ttccatgggtg	tatatgtgcc	acattttctt	aatccagtct	atcattgttg	32640
gacatttggg	ttggttccaa	gtctttgcta	ttgtgaatag	tgccctcaata	aacatacatg	32700
tgcatgtgtc	tttatagcag	catgacttaa	aatccttttg	gtatataccc	agtaatggga	32760
tggctgggtc	aaatggtatt	tctagttcta	gatccctgag	gaataaatga	ccaactattg	32820
agaaattgca	gggtagtccc	tacatgaggg	ttaggtagaa	ttgacctgct	ttctgcctca	32880
taaatttttag	aaaattaata	agataattta	ttacgggggtg	gtgtttgttc	cctcagtact	32940
ttatcatcta	tggtgataat	gttaataaatt	aattgcataa	ttaacaaata	gcaaattatt	33000
gtgggggtgt	gtgtgtgtgt	gtgtgtgtgt	gttttagacag	ggctcttgctg	tgccacccag	33060
gctggagtgc	agtggcgtga	tctcggctca	ctgcaacctg	tgccctccag	gttcaagcca	33120
tcatcctgcc	tcagcctccc	tagtagctgg	gattcacaggc	gcctgccacc	atgccagct	33180
aattttttgta	tttttaatatg	aaatgggatt	tcaccatggt	ggctaggcta	gtcttgaact	33240
cctgacatca	ggtgatccat	ccgactcatt	tcccaaagt	ctgggattac	aggcatgggc	33300
catcatgcct	ggcccgcata	ttgttgttat	ttataactct	tcaatccaaa	tcatcagtgt	33360
ctatgttgtt	tccttaacta	tcaaagtatg	ataataatag	taccttcttc	ataagatagt	33420
tgaagggttt	ttaatatcca	tatgggtactg	agaatgatgc	ctgaaacata	gtaactaccc	33480
cattttttatt	atattttctgt	taataataat	acataccatt	attgctcttg	cataccatat	33540
tgctcttgca	taccatatat	gctcttgcta	tatgctacac	acagtatttc	athtagcct	33600
cactatgtcc	ctgatgtagg	cattaatatc	tttattttgc	aaatgagaaa	acagtctgta	33660
ccttgatgc	catgctgcta	ttgtttatct	gtttgaatct	caagcaaata	tgcttgataa	33720
ttggtacca	aataagcctt	tttctgggta	aggaatctga	tattgtgttt	taaaaaacac	33780
acattttaatc	ctggggctgc	tgcattactc	ctgctgcccc	atcctactgt	gatcaaaggc	33840
acatacatga	gatggtgagt	tgtccccttg	ccaatgaggg	tttggtaga	aaggaaagt	33900
cagtacttct	ttgtttctga	attgcaagta	tgtgtgggtt	agagggggag	gctgaatatg	33960
aaggctctgg	gacagccac	caggtatccc	atgagacttt	gcaaaggaaa	aggaggtgag	34020
tgacagccca	gggtccaata	ggatagaagg	aaaagccagg	ccatggagt	cctcagacct	34080
gctttctaa	ggcaactcta	ccacctcagc	aagccattga	acttctctga	gctcagctct	34140
ttcatttata	aaatgggggtg	acagtgtctca	catgccagga	atacaaaagg	attgaaagat	34200
aaaacacgta	attaagcacc	tgttggtaca	catctgtcag	ggaccccaat	aaggctcagct	34260
gtcttcctgt	tgacttctgt	tcttggtggt	tctccaagat	cataccttcc	atcaacattt	34320
accgtcactc	ccccacccca	tgcccaatac	tgaacagtgg	agggacgctt	cacctacagt	34380
tataatgttg	aaacttcaac	ccaaagcaag	tactgttagg	atctctggaa	actttccctc	34440
aaataaggga	tttgaatggg	acaagaagaa	gttttacaga	tagccaatgg	agatgattta	34500
atgggggttat	gatagaaacg	agaaagtaaa	acaaacccat	gctttaaagt	ctaccatttc	34560
aggtccatat	tttcgcttga	aaattgagat	tcctattaaa	caatgacatt	tacacaaaaa	34620
agtagaggag	ttggttgaag	gacagggtaa	tgccaggagg	aattgggaat	ttgagagtca	34680
agtcaaagga	ctgaaatact	cagaatacta	agggcacctc	agggctctac	caaggacacg	34740
tagaagcttt	gaatttgtag	caccacccta	atthaacgag	ctacctcagc	acgtagtggg	34800
gccttgga	acagatgtca	caaactctca	ttagattgtc	aaacattttc	cagcatttcc	34860
tctcccatca	tagctggtta	tcaagatata	tagacacaca	cgtgcataca	cataaatacc	34920
ttgataagtt	actagagaaa	gcagaaaaat	gtctgacagt	ttaatgagat	ttgggtgaaa	34980
gaaaattcta	tatttcat	ttttccaggc	actagaaata	attcatcaat	gtttctaaga	35040
ctcattcagc	gtggctgcat	tttttaaaat	atthtcataa	atthttgagga	gcaaatacca	35100
ttattaggca	ctaaaaaggt	tgaagtctaa	tagattagcc	gcttcactct	ccttcactca	35160
gctcagcatt	cgttcaactg	gctcttactg	gttaacatcc	acacgcctcc	tgactggcta	35220
ctcagtgccg	atgacatttc	cttcacacac	agggctggtt	ttaagataca	ttgaggtgac	35280
atcaggtggc	ctgtaaaagt	gtcatttttag	gatatcctat	tcaaagacat	ctgtggaagt	35340
gtggaccaat	ttattgatga	ataacagtga	aggggtttcc	accagcaagt	aacataaatt	35400
tttacaatga	tgatgctgaa	gtagaaagag	tttctagtca	gggactggac	aaatcaattt	35460
gcagacgatt	tttaggaaga	aaaacattgc	aacagtaaat	tgtaattgat	aacttctaga	35520
gccactttta	gtactgctat	tttaggattc	tgaggggaag	aaagtgttct	gcaaagcaat	35580
aagcaaagt	atthgttcca	agcccaaat	ttaagcagtt	tgagaggtaa	aaagagtcat	35640
taccaatgtg	ggtatagaac	atgtgctagg	caaactctct	tcacatacat	gtggggaggt	35700
aataaaaatta	taatttgagg	ccgggcacgg	tggtctcatgc	ctgtaatccc	agcacttttg	35760
gaggctgagg	tggttggtatc	acgagatcaa	gagattgaga	ccatcctggc	caacatggta	35820
aaacctcgtc	tctactaaaa	atacaaaaaa	ttagctgggc	atggtggcac	acacctgtag	35880

tcccagctgc	tcgggaggct	gaggtgggag	aatcacttga	acctgggagg	cagaggttgc	35940
attgagttga	aatcacgcc	ctgcactcca	gcctggtgac	agagcaagac	tccatctcaa	36000
aaacaaacaa	aaattatagt	atgaaatagg	cattaaaata	ttgtgtat	tagaggagac	36060
tgaggattgg	aggctgaaga	attactctaa	attaatcagc	ttgtgtactt	cagagctaag	36120
atagctcttt	gggttctaaa	ttctgtgatc	ttctttttga	ttctctttgg	agcaataatg	36180
aaggcaaaac	atcaataaac	ataacaaact	gggtaaggga	gaccattgag	aaggactaag	36240
gacaccttca	aagttctgag	tgagttttaa	aagaagaatg	atgaaaactt	tgatagaaat	36300
aggaaaaaaa	gtagaggaac	ttgtttggct	tgaaaacttct	taatgttttag	gctaattata	36360
ttgaagatga	cagtggcat	tgagaaaaa	aaatcccca	agcaattttg	gaataagagc	36420
caacatttaa	tacttaccag	acaactattc	taagtatttt	actatattca	ctcatagcaa	36480
ctctaaaaag	caggtagtat	taacagagaa	aatgaggcac	agtgaggtta	aatagctggt	36540
ccgaggctac	acagctaate	agtgggagag	ttgggactta	gacccagagg	tccagttttg	36600
aagtcacac	ttttagccat	tacactacaa	tggaagaaa	tttagaagat	atacacagaa	36660
aactataggc	acatagatta	ggggttagta	gaatgctctg	ggcagttaaa	ggaactcttc	36720
ttaaaggagg	taaagcttga	atgagactgt	tagtaagcta	tttttcactc	attggtgaat	36780
gatgttttgt	gcagtgtgtt	tttttcccca	tagaaaaata	agaaagaaaa	gaaaattgag	36840
aactctctct	ataaaaatgt	gtaacatata	tcatattcca	agagatcctt	ttggtagtat	36900
taatttttat	ctgctcacag	tactggcttc	attatttggg	gttaaaaatt	aactcaacca	36960
gataaaaaaa	tcagtgtgtt	gtatttgttt	atctttcaaa	tctgtgttct	aatttttaaaa	37020
agttatttaa	cagaacgaag	ctatcagcta	agacaatggc	aaagccgtaa	acaaacatag	37080
gttgcgttta	tgcaaatggt	caggtccaaa	gtagatgcag	aatatgccag	gttcaactaat	37140
tttaatccct	attcagccca	ggactatgta	ccataagatt	actgctagtg	ttttctgaaa	37200
atgatgtatc	aaggcatttt	ctgtagaaat	acgaaacagt	gacatacagt	agggagagct	37260
ggattgaggc	agagtagtat	agatggaagt	ttcctgaaag	cattttgggg	aaacatcttt	37320
tgggtatggt	tcttggtatga	agagttgatt	tatttagtact	ggaagggtgt	atgggagaga	37380
ggaagtgaga	ggttatgaga	gaatgaccct	cccgtgatgg	tgagtgggag	aattattgca	37440
gtatgtacgt	tagcattgct	atgtggtgaa	gttcttggga	tttcttgggg	tccgtgctgg	37500
acagcatgct	tagccaccag	tcacatgtgg	gcactgagca	ctgacaatgt	gggtagtctg	37560
aactgggata	tgctgtaagt	gtaaaaataca	aactggactc	caaagattta	gtatgaaaaa	37620
aagaatttga	aatatctcat	taatgatgtg	tatttggttt	catattgaaa	acaacttttg	37680
tattatatat	tgagttaaat	aaaatgtcat	taaaatttaa	ttttacttaa	actaaaattt	37740
aaaattctat	ttttactctt	tttttttttt	tttttttttt	ttgaggtgcc	gtttcactct	37800
tgttgtccag	gcttgagtgc	aatggtgcca	tcttggtcca	ccacaacctc	cgtctcctgg	37860
gttcaagcga	ttctcctgcc	tcagcctccc	tggtagctgg	gattgcaggc	acgcaccacc	37920
acgcctggct	aattttttat	tttttagtaga	gacagggttt	ctccatgttg	gtcaggctgg	37980
tctcaaactc	ccaacctcag	gtgatccgtc	cacctcgccc	tcccaaaggg	ctgggattac	38040
agggtgtgagc	caccgtgcct	ggcctatctt	taccttttta	aggtagtgac	tagcaacttt	38100
aagattcata	tgtggctcat	gctgtatttc	tattatggaa	ctgccttatg	actttcaatg	38160
ggtaggatgg	acacatcctt	ggtgggatgg	agaaatctat	catagcagct	ggtccttgaa	38220
gtgggtgggg	atatgataat	aacttaggtg	gggaggctca	ggaggactca	cagaataact	38280
ggcaaccctg	cccctgtctg	taaaaacccc	atcctggagg	aaataagtta	ggaaaagggt	38340
ttgcattttg	tggaatgaaa	agtctgttgc	atctagcttg	agacagagca	aaaagagttg	38400
attgtcagct	tcatgaagac	cagggggtct	aaaagaccca	gggatcaaca	accaatgaga	38460
gcagcatgga	ggccaagaac	caggcaaaat	gctgattcca	ggactgggat	tcaggatgat	38520
ttccttctat	gcaataatct	gctccttgaa	agggtatcta	attgggcatt	gcttttactt	38580
gctgctttca	attcttttat	gttcttttct	agtaaatatt	ttttcttaat	ttcattgcag	38640
ctcgtattta	tcttggaac	agagagaaat	gtttcacaag	cttaagccag	tcttttaaaa	38700
ggagaatggc	aggactccaa	aaacagacat	gctgatatgt	actggggaat	ttttaagtgc	38760
tgaaacctcc	aagacaaaag	agactgtgtc	tttattgttc	tctgaattac	tcgtacccag	38820
ctcggtacct	ggaacatgat	agggatccca	tagtggtttg	atgaataaat	tagtgactcc	38880
aagagtaaag	taatcctcag	gaggacaaa	gcagatagct	tcccttccct	atcagaatgt	38940
acttctctta	aagcttttct	tggtataaatt	cttgagagaat	tttgcccttac	agaagtcaaa	39000
tcacatacca	aagtgaanaac	tggatcttct	acaaataatg	gaagaatcaa	ctctatcaaa	39060
acaacaatta	tacatatgat	caatggaggg	gttgtcacga	gccaggctaa	gagctttaca	39120
tatattatct	cattctgtct	atgccagaga	atcaactatg	acatatgtaa	cattaaatct	39180
cattttatag	atgcaaaaac	tggggtgtaa	agaagtcaaa	gaatcagcca	gaatgtacag	39240
aattagcaaa	ggtggaactg	ggatttgaat	tcagacagtc	tgactccaga	cgccatctcc	39300
gaattatgca	taattatatt	tcaattatta	acattcataa	attgaaatat	gagggataat	39360
gtaccttttc	atgaaagctt	tgctcgttgt	gtggatgagt	gtgtgtacat	gtaactgctt	39420
atgtgtgcta	tactgaggt	agaagacatc	tctctctctc	tctctctctc	tctctctctg	39480
ttttggtcta	cttttagtaa	gacttgtatt	tgattgagtt	cagaagtttg	attatctttt	39540

taactaacct	gtttgtttta	attatattaa	aaattagtca	ctttcaacat	at ttgcatag	39600
gtaattgtta	gggtgatctt	ttggatgata	agatgtaata	tactactact	acacacagac	39660
acccagacac	acacacacac	acacacacac	acacgcatac	agacacatcc	ttgagctcaa	39720
agagctttct	caggactaca	tactttacat	ttacaacaag	tcatttagaa	aactatcaat	39780
cctattcaaa	tctcagcaaa	acagagatac	caggctctgc	ctttctcttt	aactgctctt	39840
tttttgctgt	agacaaagct	gtttcttgca	ctgctacata	tataaaaaag	tgacaaatcc	39900
ttactgtca	aacaagagaa	atagtttgat	aaatataata	attccataag	atggcacatt	39960
atacactact	aaattgtaag	gacagtaaa	ttactgttaa	gtacaaaaaa	gtcatgatta	40020
aatgttaagt	gaaaaaatag	aatatagcta	gatttgaatt	tgaatattca	atctgtatac	40080
cagtatgtat	agaaggaaga	gtatatacca	aatagtaaga	gtatctatct	gttttataat	40140
ttgatataat	acaaattatt	ctaccttaaa	catatgagaa	tttggtggtcc	atgatgttgt	40200
agattctatc	ttctcaccct	gcatttccga	agacatatgg	tattggctca	ttagactatt	40260
tggtgaatag	tctttattct	attatcatag	aaaaaataaa	tgagtgcata	tatccatata	40320
caaaatagag	gtctgttctt	cctgtatata	tttatactaa	aaaaactgag	actttttttt	40380
acagttgtat	atatataaac	atatttgttt	atttatatac	acatatataa	atcaatttta	40440
tgtacatgtg	ggtatacata	catccatgca	tataactctg	aagtgtgac	tctctaaaga	40500
aagcccaggt	attggtcaga	attcatgctc	ggctcaggag	tatagaatta	agagatacaa	40560
acctcaaaaa	agagggaaac	gaatcttcaa	atctgagcca	ccttacaaga	at ttttaagg	40620
taactgtttt	aagtgtaaac	attatggcaa	tgtaatatga	taatttgtat	ggcacaagat	40680
ggagtctctg	tggccagaac	tgagtatggg	aatagtcagt	gctaattctt	gtgcaaagca	40740
caaaggaaaa	ttggtatata	aggcctgaga	gagaggtcaa	ggaagcaa	actataaagt	40800
cccagaaggc	agttggatag	gacaagggga	tcggaagagc	aggatcaaga	tcagagagca	40860
gggaagatgg	ggctagctaa	tctagggaa	ggagggagaa	gggataccca	gaagcaggat	40920
tcagggggtt	agagcactta	caccatgctc	atggccatgc	ccatggacag	gagaagcata	40980
tgatgggtgc	acatgggtga	gctgggtctaa	agaagagggg	gattcatatc	acagtaaaat	41040
taagtgaat	ttatccacag	agcagaggag	gttgtgaaga	gcaggcatag	gctat ttttc	41100
atgtctgaat	gctctcaaga	cctcttagtg	ttgtaggtag	atgacgcaca	ataaatattc	41160
cttgatttga	attggcataa	aagg tcaaat	cagaaagcca	gaagt tcttt	aagg tttcaa	41220
actagattct	aagaagtcca	aaggcatctc	agaggtcatc	tcaggggagaa	aagg ggaagc	41280
aatgaggtag	aactctgaac	ctccactcta	cctaaacaag	gcagccctgc	tgtgctccga	41340
gtcatgtatg	gcgacatcag	ggaagatgta	aaataggatt	gtgataggaa	acagcat tttg	41400
aaagccattt	ctttacgaag	ggaagcgtag	aacctctttc	cccttgtag	ccaattcaaa	41460
tgaagacttt	tgggagcttt	tgaagagaaa	gacaggattt	ctagggagat	gtttcagaag	41520
cagcctaact	ataccctgtg	cttcagaaag	agcagtgctg	tctcagaagt	aatcaccttc	41580
atcaaccagc	aggtcagtg	gggtctcctg	aagagcccg	acaaccacgg	gaagcgacat	41640
ccactgttgt	gcagtcaaaa	gaatctttgc	tctcat tttc	tccactctct	ttccttccca	41700
aataggggat	cataggaaga	tcctgccttt	cttccagttc	caacatttat	gaagt gaaat	41760
ttccatcaga	cagttgcttt	tgacaaacaa	agattgacta	gaagcttctg	tgagagagcc	41820
ttaggaagtt	ctctggggaa	gccctgcctt	tgtttgggtt	tctttgcctt	tggctctgtg	41880
atgtttgctg	taaaaaatga	ttactttcag	gtgattaaaa	gtggggaaga	atgg tttcaa	41940
gcttttcatg	tagcaataaa	tatccctgtc	tgtaggatta	ctttagttga	aaaaaaaaca	42000
tggcttcagt	gatgccttct	caatgtacaa	gattcagagg	aatggaaaga	aaatgaaata	42060
aggccggcct	cgggtggctca	cacctgtaat	cccgggaagc	cgaggcggtg	ggatcacgag	42120
gtcaggagtt	caagaccagc	ctggccaaga	tggtgaaacc	ccatctctac	taaaaatata	42180
aaaattagcc	aggcatggtg	gcaggcgctt	atactcccag	ctactcagga	ggctaaggca	42240
gagaattgct	tgaaccccg	aggcgagggt	tgcagtgagc	caagatcatg	ccactgcagt	42300
ccagcctggg	cgacagagca	agactccatc	tcaaaaacaa	aacaaaacaa	aacaaaaggaa	42360
aagaaatctg	cagttaatat	tttggaagc	tttcttctact	tgtatgcatt	tttaaaatgc	42420
taatgttaat	aacagttcgg	gacttctaac	ttctatat t	aagcaacaaa	taaataaatt	42480
gtcagatgg	acttcatcat	ccttctctcc	catcttctta	gaaatataaa	ttgcttttag	42540
tgggaatgct	ataatttttag	accagaaaa	acatgccaga	tgtctcttat	atgaagcgt	42600
cccgcccaag	gatatatata	tgcttagtgc	attaggatgt	gttctaaata	atactgcaaa	42660
gcccttgga	ggatgggtct	gaacactcac	ttatat ttaa	ctgctggcat	gttgctttgt	42720
ccctgtgtct	tgtgtacta	tttccattga	tgtaaaggaa	gcaccaatta	aataacactc	42780
cattattaga	gaaccaggca	caagtcagct	gaggcaggag	acccgccttc	ttttccagaa	42840
acaatgtaaa	gcctgggtgg	gtgagggtct	ctgggcttcc	gccgtgcctt	gcttttgaca	42900
ttctccagca	cacctataaa	acatgtctaa	ggctgtcctg	tttagtctga	ttattcaaac	42960
tatattgtcc	agggtagagc	aaagggaac	ctagctgaac	cctggagatg	acagcaggga	43020
gagagagagg	ggcaaagaag	ggcaaaacgg	gaaaaacagg	aaacaggcta	gtgagaagag	43080
taaaaacgct	cagggtgagg	aagcagggtt	tctaagctct	ctaactctcc	ctgtgcagct	43140
ggcttgctgt	atgg tttata	caaatccagt	ggtgatctct	gtgcaacgtg	gtatcacctg	43200

tttaaagagg	tctcatcttc	attttcaaag	aggaatacat	gtttttttac	ttactcttct	43260
gcatggctga	ctccttttca	tgcttttaagt	ctcaatctta	atgccacctc	ctccttccag	43320
acgttcccag	ctaaagtggc	acttcccagc	cccattactc	tctatgttta	ttgcctgcat	43380
agctcttatt	tgtaatgatt	tcgtaatagt	ttgatgatga	tcgatgatgaa	tattacttta	43440
cctatttatg	gcctctcttt	tagtattaaa	ttctgtaagc	cacatgagca	tggggacaca	43500
tctcctttgt	cactgccccca	ttgctggcat	ttagcacaaag	catgggtctat	aatagatacc	43560
aaacaaatat	gtattaatca	tgtaaataag	taaatccatg	aatgaatcta	tcagacagtg	43620
tagatagcag	cacataaagg	aaagggaatg	tagtaaattt	ttcatttttcc	ttgaagatgt	43680
agctatgtat	taggaatttg	aaaaatacat	tatcaaacac	aaagctaaat	tatgccagct	43740
aatgactact	aaatataata	aaatcggtcg	ggcacgggtg	ctcgcacctg	taatcccagc	43800
actttgggag	gccgaggtgg	gtggatcacg	aggtcaggag	attgagacca	tcctggcaaa	43860
catggtgaaa	cgctgtctct	actaaaaata	caaaaaatta	gctgggcatg	gtagcaggca	43920
cctgtaatacc	cagctatactg	ggaggctgag	gcaggagaat	cgcttgaacc	caggaggcag	43980
agcttgcaagt	gagcagagat	cacaccactg	cactccagcc	tgggcgacag	agtgaacctc	44040
tgtctcaaaa	ataaataaat	aaatataata	aagtatgtag	aaagtcaaaa	atcttggggg	44100
ttatattgca	aagaattttcc	actatattga	taatggagaa	aggcttttaa	tattatattt	44160
tttgaatatt	aagaaattgg	catctactca	ccagtttgga	catcgctttt	aaaatacaca	44220
ctaaacgaaa	gccattttgt	acttataagt	gctagattaa	attcctgcat	aggctgaaaa	44280
aggctctctt	ccatgccttt	ccaaattttac	aattaacaaa	gagttaatta	ttctcagagt	44340
catttcttcc	aattcaccaa	ttaggatgag	ggctattttgt	tacaatcata	aaagaggaaa	44400
tggtgcatgg	gcaagaagaa	atttggaaaag	gaaatgtgat	tggaggaatt	atattgaaaag	44460
gtgaaacaag	ggagaaaaga	taaagagaag	aaaaattaga	aattggaaac	aaagttattc	44520
cagccccctc	ctaataacta	ctactctttg	gaacaaggga	agcagtacct	gacaagaaat	44580
tttttttctt	ttatttttta	tttttattat	tatactttaa	gttttagggg	acatgtgcac	44640
aatgtgcagg	ttgtttacat	atgtatacat	gtgccgtgct	ggtgtgctac	acccattaac	44700
tcgtcattta	gcattagtta	tatctcccaa	tgctatccct	ccccctccc	cccacccac	44760
agcagtcccc	agagtgtgat	gttccccctc	ctgtgtccat	gtgttctcat	tgttcaattc	44820
ccatctatga	gtaagaacat	gcagtgtttg	gttttttgtc	cttgggatag	tttactgaga	44880
atgatgattt	ccaatttcat	ccatgtccct	acaaaggaca	tgaactcatc	attttttatg	44940
gctgcatagt	attccatggg	gtatatgtgc	cacattttct	taatccagtc	tatcattgtt	45000
ggacatttgg	gttggttcca	agtctttgct	atttgtgaata	gtgccacaat	aaatatacgt	45060
tgcatgtgtg	ctttatagca	gcatgattta	tagtcctttg	ggtatatacc	cagtaatggg	45120
atgtctgggt	caaatagat	ttctagtctt	agatccctga	ggaatcacca	cactgacttc	45180
cacaatggtt	gaactagttt	acagtcccac	caacagtgtg	aaagtgtccc	tatttctcca	45240
cagcctctcc	agcacctggt	gtttcctgac	tttttaatga	ttgccattct	aactgggtgtg	45300
agatggtatc	tcattgtggg	tttgatttgc	atttctctga	tggccagtg	tggtgagcat	45360
tttttcatgt	gtcttttggc	tgcaaaaatg	tcttcttttg	agaagtgtct	gttcataatc	45420
tccgcccact	ttttgatggg	gttggtttgt	tttttcttgt	aaatttgttt	gagttcattg	45480
tagattctgg	atattagccc	tttgtcagat	gagtaggttg	cgaaaatttt	ctcccatttt	45540
gtaggttgcc	tgttcactct	gatggtagtt	tcttttgctg	tgcagaagct	cttttagttta	45600
attagatccc	atttgtcaat	tttgtctttt	gttgccattg	cttttggtgt	tttagacatg	45660
aagtccttgc	ccatgcctat	gtcctgaatg	gtaatgccta	ggttttcttc	taggggttctt	45720
atggttttag	gtctaacatt	taagtcttta	atccatcttg	aattaatttt	tgtataagggt	45780
gtaaggaagg	gatccagttt	cagctttctc	catatggcta	gccagttttc	ccagcaccat	45840
ttattaaata	gggaatcctt	tccccattgc	ttgtttttct	cagggtttgtc	aaagatcaga	45900
tagttgtaga	tatggggcgt	tatttctgag	ggctctgttc	tgttccattg	atctatatct	45960
ctgttttggg	accagtagca	tgctgtttgg	gttactgtag	ccttgtagta	tagtttgaag	46020
tcaggtaatg	tgatgcctcc	agttttgttc	tttttgctta	ggattgactt	ggcgatgcgg	46080
gctctttttt	ggtgccatat	gaactttaaa	gtagtttttt	ccaattctgt	gaagaaagtc	46140
attggtagct	tgatggggat	ggcattgaat	ctataaatta	ccttgggcag	tatggccatt	46200
ttcacgatat	tgattcttcc	tacccatgag	catggaatgt	tcttccattt	gtttgtatcc	46260
tcttttattt	tattgagcag	tggtttgtag	ttctccttga	agaggctcct	cacgtccctt	46320
gtaagttgga	ttcctaagta	ttttattctc	tttgaagcaa	ttgtgaatgg	gagttcactc	46380
atgatttggc	tctctgtctg	ttattgggtg	ataagaatgc	ttgtgatttt	tgtacattga	46440
ttttgtatcc	tgagactttg	ctgaagttgc	ttatcagctt	aaggagatac	tggcaaaaac	46500
cacatgatta	tctcaataga	tgcaaaaaag	gccttgacaa	aattcaacaa	cccttcatgc	46560
caaaaactct	caataaatta	ggtattgatg	ggacatatct	caaaaataata	agagctatct	46620
atgacaaacc	cacagccaat	atcatactga	atgggcaaaa	actggaagca	ttccctttga	46680
aaactggcac	aagacaggga	tgccctctct	taccactcct	attcaacatc	gtgttggaag	46740
ttctggccag	ggcaattagg	caggagaagg	aaataaaggg	tattcagtta	ggaaaagagg	46800
aagtcaaatt	gtccctgttt	gcagacgaca	tgattgtata	tctagaaaac	ctcattgtct	46860

gacaagaaat	tttatagtct	gatgaaaggg	attctaaaga	gtcagggggc	acaggtctca	46920
ggcttcgact	ggatgtgatc	atgtctgagg	cctttcgatc	ctcactttcc	ttatctggaa	46980
aacaagaata	gctgaatctc	cttctaaggg	cgtttgtgat	atgaactgag	atcttgcata	47040
tgactgcacc	aagtctagct	caattcgcac	tagttccctc	cattataccc	ctccctcgag	47100
ctttacccag	actcagaaga	aagccaggca	acatttctac	ttctctatat	gcaaaaacaa	47160
aagcaaacaa	gtggaaaacc	tcacaaaaac	agttaacttc	aacatttggg	cttacacaaa	47220
caattcaaaa	atctcttttt	atttcatccg	ccatgattat	agttattttt	ctaaagtgaa	47280
tgattctact	tcccaaagtc	agtaaaccce	ctgttaaaga	tagttaattt	tcctctagat	47340
gattgtggcc	cttgaaagtc	atcaaggtca	tatttttaat	tatttcccca	gaatttttcc	47400
tgaacacgtg	tccttttgtc	taaatcaatc	caagtagggt	ttagcattag	tcataaagag	47460
ggtgctgtca	acaaagaaat	caactgagtg	gaagtgatat	tataatgtaa	ataacttgac	47520
ataggaatac	ataacactca	taaatattta	ttgattttatt	taataaatta	aaaattaatg	47580
cttatgatat	gtaaacaaag	tataacaagg	cagtcaagaa	taactctctt	tagttcatat	47640
gattttttcc	catacgtaac	tgaatagcaa	gaaaacaagt	aaccagttt	gaaaatggac	47700
aaaaaactga	aatagatatt	tctcaaaaga	agacatacaa	atggccaata	ggatattttt	47760
taaatgttac	tagtcatcaa	ggaaatgcaa	atcaaaatga	caatgaacta	tcaccttaca	47820
cttgtagtaa	tggttactag	caaaaaaaga	caagggataa	caagggttgg	aatgatgtag	47880
agaaaaggga	atccttgtac	attgttggag	ggaatgtaaa	ttagtatagt	cactatggaa	47940
aactgcatgg	aggagcttca	aaaaactgaa	aataagccta	ccatgtgatc	ctaatactgg	48000
gtatatatcc	aaaggatttg	aatcaatat	gttgaagaga	tatctgcatt	cccatgttcg	48060
ctgcagcctt	attcacat	gccaagtatg	aaattggctt	gagtgtccat	caacagatga	48120
atggctatag	aaaacatata	cacagtggaa	tactattcag	ccttaaaaaa	gaaggcaatc	48180
ctgtcatttg	caacaacatg	atgaacctgt	aggacattgt	gctgagtaaa	ataagcctgt	48240
cacagaaaga	caaatactgt	ataatctcat	atgcagaatc	ttcaaaaagt	tgaacttata	48300
aaggtagaga	gtagagtgat	gtttaccaga	gggtgggggtg	gagaggggtg	gggtacaggg	48360
aatgggagaa	tggttggtcaa	agagtacaaa	gttttagtta	gacatgacaa	ataagttttc	48420
aatgctattg	cacagtgtgg	tgaccataat	taacaataat	gtcttttata	tctcaaaatt	48480
gctgaaagaa	taggccttaa	atgttttcag	tatagaaaag	tatatgagat	gatgccta	48540
ttaattagct	tgatataatc	attccacaat	gtatacatat	atgaaaacat	cacattgtat	48600
cccataaata	tatacaatta	ttatttgtta	attacaagta	aaagtttaaa	aaatggcaag	48660
taaatcatgt	agcccttgag	atagatggat	atgtgagcct	agcttgaaac	aaatgtcatc	48720
atataccac	catttacaaa	gactttgagc	agactgaatg	ctcacagaac	atattgggag	48780
gagatatttg	gcagctgaag	tgccaaatag	tcattttcca	agggacaac	aacagttag	48840
aggtttccag	ttaaagttgc	agagtttctg	cagagtctct	agcagtgtg	gatccaagg	48900
tatgcggtgt	atccaagtag	ctcttgagga	aaccacaggc	acatcctggc	atggggagca	48960
cctcaggagc	acatcctgag	tttcagggca	tttgaaatgg	atgtgcagtc	acatcccacc	49020
caactgcaag	ggataccag	cctacatgca	gaggtcagga	aagctgcca	cattaagaca	49080
ttgcatgcaa	taggcccctc	cccactagga	gtctatggag	acagaaatgc	attttgagga	49140
gcaatttcat	gcagtcattg	gttaagtga	ccaagtgagc	tatgaagcca	gattttccct	49200
cctggggccac	atatttcaga	ggcacataac	tcaagcttgc	aacacgtatt	caaaagagac	49260
cagctacact	tggtagagac	agccatagga	aagtgaatg	accctagggt	ttagtaaagc	49320
cagctgtttc	cacttctgaa	aataataaaa	tgaataata	aaataaattt	aaaatgatac	49380
aaagttcaaa	gtttaacaaa	tacatttgaa	gccatttgca	acaaatacat	ctgaagctaa	49440
ttgctggctc	tagaaaagtgt	gggtctttg	ttgtggagca	gtgttaatga	tttagcatta	49500
cttatctctg	gcaaaggtga	tttttgagat	aacatgttat	ggaagaaagt	gaactgaact	49560
tgaagtttg	aagatctcga	ttgaagtatc	atcttctgct	caactacttg	cattaacttg	49620
tacaagtcat	tcaaccgctc	tgaacataat	ggaaaaatgg	gatgagaata	catgttgtat	49680
actctccaaa	gacagggaga	ctgctgatat	aagagggcac	ttttagtaac	tgatggagca	49740
aaatgttggt	atatgagtgt	cagcataggg	ccctgggctt	acaacgggtc	catgagcctt	49800
agaacagagg	aaggacagct	atagcaatga	aaggactagt	gcagattcag	aaaaataaga	49860
agacagaaac	caaggtgtag	taacatgttt	tagtatggag	gggaaggcag	ttatagaaac	49920
ttgaattaca	taatttgtac	atcttctggga	gatagaaggt	aaagatagca	gctaattggag	49980
acaggacagg	actggtactt	gattatggaa	gaaaggagggt	aaatagaaga	gacaaaaagg	50040
gagagaagag	atgtcaactg	cctactctgg	tagcctctgt	atccaaaagg	ttgactcaaa	50100
cattcgctca	taactttgtc	tggtttaatc	ctgctcatcc	cagcagactt	atctcaagtg	50160
tctccacgtt	ttgggaagtc	atcactcact	tctctgggct	ttcatatggg	agagcattta	50220
attctgttga	aaaactat	aatactacat	ctacctttct	ctatggactc	tgagcttctt	50280
gagggcatgt	atcatgtatg	ttctattctg	aagcaccat	acctagaaca	aagcttagca	50340
catagtagga	acttaataaa	tatttcggag	ttgaataact	agccttatgt	aatcctcaca	50400
acaaccctaa	gctggagact	caaacaaggc	tggaaataag	taggtgccaa	gaagaactga	50460
gattcagaca	catatttgca	ggtaaaacat	aggaacactg	aacattcact	gagaactgac	50520

aacttggtggg	gttggtgtag	gatattgtgac	cagagactct	tgaatgccag	tctctgtacc	50580
tgtaccatgt	tggctaacaa	gaatcgcatg	gaatccttgc	tgaaaataca	gaccttagaa	50640
gttttctcaa	atctggagag	actgtactat	ggtttgaata	tggtttgtcg	ccaccagaac	50700
tcatgttgag	gcttggtcct	caatgcagct	gtgttgggag	gtgggacctt	gagggaagtg	50760
tttgggtcca	gtgggcagat	ccctcatgaa	cagataaata	ttgtcttgtg	ggagtggatg	50820
acttctgtct	tgcaggactg	gatgaattac	cacaagaatg	agttgttgtg	aagcttctcc	50880
tcatgttttg	ctgtgttttg	acgctgtctc	ttgacatttt	tcttcttttg	tatgttgtaa	50940
ggcagcacat	aaccctctgc	aagctgagca	gatgccagt	ccatgctctt	ggattttcta	51000
gccactagag	ttgtgagcca	aataacattt	tttttcttta	taaattaccc	agatcaggta	51060
ttctgttaga	gcaacactaa	agggactaag	acactcttat	tctcaccaa	tctttatttt	51120
ggtaatgatt	tctcacacct	attcatttgc	tccagaaaag	gtagtatttc	tccatagtct	51180
atcttcatct	tccacttcat	gcttattcaa	tccattacca	attcctgtca	atatacttc	51240
ctaaatatct	ctttcaacca	ccaacttttc	tatcctcact	attactatcc	ttttacaagc	51300
acatagacca	ctgccaggaa	cccttgactt	gccgacctga	ctctatacta	gttcttcttc	51360
ctactgcagc	cggagcaatc	tttttaatac	aagctatgac	tcaacattta	ctctcttgat	51420
gaaatctcct	agaagacctt	ttgtagctct	tagaaciaag	actgaaataa	aaactctata	51480
ctatagtcta	aaagttcctt	gttggtctca	cctctccagc	ttcctccctc	tgtgctccaa	51540
ccacatgggc	ttgccttcaa	tgtctcacat	atcaccacgc	ttcaaattcc	cttctggctg	51600
cagggccttg	gcacaaactg	ttttctctgc	ctgatgtttt	ccacccttcc	acctacattc	51660
atcagtttga	cttctactta	tcttttggag	ctcagctcag	acaaggttag	atcctgccat	51720
tgacatactc	attagcacc	tgaaatat	cttaatacaca	gcgtatgatt	atataattat	51780
ttgtgtgatt	aactgataaa	tgactgtctg	acccctcctc	cctgctttta	gactataagt	51840
tgtattaatt	cagggcctat	gttagcttta	ctcagtaact	tgtaccaaat	gcccacccca	51900
gcatcctcac	aagtaagggt	gttcagtaca	tgggtgttga	atagatgcat	gaataatata	51960
ataagtcaac	aatttgtctc	aggaatctca	ataattttta	tgctatcaaa	gtgatttgat	52020
gcagatttgg	gaaaaattat	ctaaaaaatt	catcccaagc	taagatccta	tgattcttag	52080
ctatcacaga	atctgtgatt	ctgtgctact	cctctgcgct	tctcatgtac	acttacatgg	52140
gtatacccat	gaaaaatggt	tgttgggttg	tttgtttgtt	ttgagacaga	gtttaactct	52200
tgttgtccag	gctggagtgc	aatggtgcaa	tctcggtcca	ctgcaaactc	cacctcctgg	52260
cctcaaggga	cttctcctgc	ccagcctcca	gagtagctga	gattacaggc	gcatgtcact	52320
acactcagct	aatttttata	tttttagtag	agacagagtt	tcaccatggt	ggccagggtg	52380
gtgtcaaaact	cctgatctca	ggtgatctac	tgcctgagcc	tcccaaagtg	ctgggattac	52440
aggcgtgagc	caccgtgccc	agccgaaaaa	tgttttaaa	catctaggat	ccttggcagg	52500
ccttttagcac	actgcacaga	agggacattc	tgtgcctgtc	actggaatga	ccagcaactc	52560
tggttccctg	ctttgcccag	actgtttcta	tgtcccttcc	agtttagttc	agttcaacaa	52620
ttatctagt	agcactttct	ctgagcgggg	catctgcttt	gtgctaagt	taagccctgc	52680
ctccaagaac	tatttgcata	aggagagaca	cacacatgaa	aaccaactaa	ttgtgattca	52740
gtgtaaaata	cgtagtaatt	gacaaatgca	tatagtttca	tgacagccct	gtagaaggag	52800
taggcaagt	ttctagtatg	gcttcaaggga	ggaaatgtag	cttaaaacag	gttcggagg	52860
atgagtaaga	gtttaccata	cttttaagg	ggttgggagg	aatatattga	aatgaaaatt	52920
acattttgca	aatgcagtat	gaggagccgt	ggtacagttt	aatgtgttta	gagaacaacg	52980
agtaactgta	tggcagaagc	agtaaatatg	acgaaggagg	aagctgggtg	gtttggagaa	53040
ggctgaggga	tcatgaggca	tattaccttc	ctttaaaagc	catgccctat	tctccctctc	53100
ctgccacttc	aaattcagg	tcaccattta	tatgtctatt	agtcctggtg	tctttctcat	53160
gctctttgat	tagtccttaa	tccacaagcg	caacattgca	atacttgcc	agcatatttc	53220
acaggcagg	gacctaatgt	ccctgtgaga	acctatcttg	ctgagattgt	cgctggcaga	53280
tttacttcca	gtgtgattgt	tgcaagaatt	tgtctaacag	aatgaatgat	caaccttgag	53340
cagaagagat	tatgaaaaac	ttaatagcat	tgtagcaatg	tggctgttaa	tgaaatacac	53400
ttggctgctc	ccgctgtttg	gcaccaacca	acctgacact	gtcaacatca	caatacgata	53460
tttattccca	attattttac	ggcaacaact	gaaatacaat	gtgttattaa	tcatatttat	53520
tataagtatc	aatttgagaa	atttctgaca	tgccagaaga	taaatagggt	tattatgaaa	53580
agcagttctg	cttgggtgat	gctggctgct	gctgtgtaat	aaatagcctc	tgtggggaaa	53640
gtttttttaa	agaaataaag	caaaaaata	gcaactgaaa	cagaaaggaa	gcatcaaaac	53700
tcttcaata	cctgctgtgt	ccattggtca	agcacattca	ggacatcgca	tgcttttaga	53760
actccagcag	gttccaacag	ctagtaggac	attctagact	ctgagagaga	gcaaggagg	53820
ttttatgact	ggggacaaag	aaaagagaca	ctgaaggcga	aggacaatct	ctgaaaatgc	53880
agtaccctcc	agactgctcc	tcctctcaca	aaaacacctt	cccagcatgc	actgctttag	53940
ggactatgat	tataccattg	attctgtcca	gaaaacctgt	gtcctgaata	tattacagg	54000
ctcattcctt	cacttctttc	aggtgcctac	tcagggtattt	ccttatcaga	acagtctttc	54060
gaacgacccc	attaaaaaaa	tagtctgtgc	aacctatgt	taacaatttt	atttattttt	54120
attatttgtt	aacaatacat	aatagggtgca	tatattttgg	gggtacatat	aataatttga	54180

tacattcata	ttgtgcataa	agattgaatc	ggagtaattg	ggatatccat	tgcccttaagt	54240
gtttttacctt	ttcttttatgc	tgtggacatt	caaattactt	tctaactttt	tgaaatatac	54300
aatagaagaa	tgtaaactat	aatcacccta	ttgatctatc	aaatgctaga	tcttattttct	54360
tctaactata	tattgtacct	attaatctgt	aattccacaa	ctatatttac	ttcttataact	54420
tttccccttc	taggctataa	accaaagag	agctgagcat	ctggttggtt	cactgcccac	54480
cacatgcatg	cctactacat	ggcagtcaaa	atattttgtg	aataaatgaa	tgaatgaaaa	54540
aaaaaagaaa	tagatgaatg	aatcatggat	gaatgaatca	aatcagtcag	caatgtcttt	54600
ctaaacaaaa	tttgatgat	tttgatgat	tacgcctctt	aaaaatattt	cttcattttcc	54660
taccccaatt	tagttttctac	tcaggacttt	ttcaatatct	tccaaacctt	ttgttctttt	54720
ttatttggtt	gctttttgag	gcaaggctct	gctctgttgc	cagggtctaga	gtgcagtggt	54780
gtgatcacag	ctcactaaag	cctccaactc	ttgggttcaa	gtgattctcc	acctcagcct	54840
cccaaatagc	tgggatgaaa	gtgtacacca	caatgcccgg	agaattattt	catttctttct	54900
ttgtagagat	tgagtcttac	tctgttgacc	aggttgcttt	cgaactcctg	gcctcaagcc	54960
atccttccac	ctcagccttc	ccaagtgcct	ggattacagg	cgcgagccaa	cttgcccagc	55020
cctggaattt	ttgagcctgt	tcaattctaa	ctattgtcac	caaaagtaac	cttaagaaaa	55080
aaaatgcatt	atctccttgc	ttcattgcac	cattaaaatc	tttcctaaat	tttccatggt	55140
aaagatgaag	ctcaaaatcc	tcagcatagc	atacaaaaac	cttcataatc	agatgcctct	55200
tcaaatacct	cctatcagaa	tggctctctt	gactaccctt	ttaaaaaat	tcccccaac	55260
cctatttttta	aattattttac	ttattttttat	tatatttttg	atacataata	gatggacaca	55320
ttcaaacagt	gcccccaaaa	ctggggcagc	agaaacaggt	ccttgcttat	tttctcagct	55380
tcacctcctg	cctccacccc	atctgtactg	ctgggtccaga	cattcctaca	gagggtgtcct	55440
cctaagttgg	tctcttcctc	tcctgcttca	gaggctttgc	cctgctcttc	tctgcctctt	55500
gaggctctgt	cctgctcttc	tctgcgtctt	gtgggttgaa	tgctgtctt	tttctactg	55560
aagatctgga	tgcttaaac	ataatgtaa	attgctgctt	tttacttcca	tttacagcag	55620
agaaattcct	cctctggcct	ctcctcttct	ctgtgtttct	ttcttcataa	tttttattta	55680
tttatatatt	tatttattta	tttattttat	tatttatttt	cattgagatg	gagtctcggt	55740
ctgtcgccca	ggctagagtg	cagtgggtgtg	atctcagctc	actgcaatct	ccaccttcca	55800
ggttcaagcg	aacctcttgc	ttcagcctcc	ctcctgtagc	tgggactaca	ggtgcccgcc	55860
accacccctg	gctagttttc	atattttttag	tagagacagg	gtttcaccat	ggtggccagg	55920
ctatttctcaa	actcctgac	tcaagtgate	tgcccacctt	ggcctcccaa	agtgtcggga	55980
ttacagagct	gagtcaccgc	acccagcctc	tctctttata	attttctctac	tgttcacctg	56040
catcaaatct	ctgaattctg	tcatgcaact	ggaacagtaa	gagggaaaaa	catggagctc	56100
aaagaaagat	ggttgagaaac	gtagagttgc	atagaattta	ctgtataaga	atggaatctg	56160
tcaagtcaga	caagcgacag	agacctat	acaaagagac	ccagtgaaaa	ttactggaga	56220
aataataaag	agaaatgctg	tgactttgaa	ataaataatg	ttcaaaaagtc	acctgcaata	56280
tttaggatag	tgtctgaaac	agatacaaat	atttctcagc	agtaaaaagaa	ttttgtattt	56340
agtctagtc	tggaaatagta	gtcagttgtc	actgaggaag	cactttgggg	tagaagaagc	56400
atltgaatgt	gtttgaagtc	tgaggcagca	ggtgaggttt	gattttatat	ttttgaaaat	56460
ggatctatca	gatggtggag	ctaccctcat	aaaagatttg	taatacgcct	gtttacctac	56520
aagattaaat	caagtgtcat	ttcttcaggg	aagattgccc	ttcaccatgt	gaaatataca	56580
taaagtatat	gtcacaatgt	gtgacagttc	tttgcctaca	tatatatttc	tccacttaaa	56640
ggagaatttt	ttttgagata	tgatcttgtt	ctgttatcca	ggatgggggtg	caatggggca	56700
atcacggctc	attgcagcct	tcacctcctg	ggttcaagtg	gtcctcccac	atcagcctcc	56760
tgagtagctg	agactacaag	tgtgcaactat	caagcctggc	taatgtttta	tttttagtag	56820
agacaaggtc	ttgctacatt	tcccaggtctg	gtcttgaatt	cctggcctca	agtgatcctc	56880
ccaccttagc	ccagaaggag	tcttattttca	ttcatcgtat	attcttagta	tctacctgtc	56940
atcaggccta	tagtagatac	tcagtagatg	tagattgagg	tttgaagaat	aagagatagc	57000
tcaccaagta	gaacactgga	tggatattgga	actaatgtat	tcctttattg	tcagcagaat	57060
ggaccatgca	catagaaata	ataaaatggg	agaattgatt	gccatggtct	aaattttgtg	57120
ccccctacaa	ttcatatgtt	gagaccctga	cccccaaggt	gccggtacta	agaagtggg	57180
cctttgggca	gtgataaggt	tgtgaggggtg	gagccctcac	aaatgggatt	tatgccctca	57240
taaaagaaac	tgagagaac	tagttaaac	cttctactat	ctgaggacat	ggcaagaagc	57300
tgtgtttctc	tgaacctgga	tagaggacct	cactagacgc	tgaccagcgc	tttcatcttg	57360
gatttcccag	gtcccaaaac	tgtgagaaat	aagtttctgt	tgtttataag	ctaccaggtt	57420
taaggatatt	tgtacagcaa	cccaaagagt	ctgagaccat	aatgaagcca	ttggaatggt	57480
gggaaggcaa	cttcatgtga	gtaactacag	taaagccagg	tgctggtaac	agtcatgttg	57540
cccatagagc	agatcctact	attacagtgc	ctagcacatt	acctgcatat	gatgatattg	57600
gatcaattag	ttaactgatt	agtttatgaa	tcagtctgcc	aaaaactagg	gcagaaattg	57660
atagcacatt	aaaataaata	tgccctaaag	tttgcaagga	gaccctatta	actgcgcact	57720
gttttctttt	tattttcttt	tttttctttt	tgagacaggg	tctcactctg	tcaccagggc	57780
tggagtgc	tggcacagtc	ttggctcatt	gcaacctcca	cttcccgggt	tcaagcgatt	57840

cttgtgcctc	agcctcccaa	gtagttggga	atacaggtgt	gcaccaccac	acctggctaa	57900
tttttatatt	tttagtagat	aggggggtttc	ccatattagc	cagcctggtc	ctgaactcct	57960
ggcctcaagt	gatctacctg	ccttggcctc	tcaaagtgtc	gggattacag	gcgtgagcca	58020
ctgcacctgg	ctagccactc	actgttttca	tgttaggcta	agtaagcttt	tttgaagacc	58080
attaacataa	atatacaacc	taaatgtatt	ttacctgaat	aattttactc	atgtccacag	58140
cttggttcttt	cataggctgc	catgatgagg	aagaacagag	attagtagta	gcactattca	58200
tttctgatat	ttttgcagta	gtggttctaa	ttctcactcc	agtttagaaa	agatctgtag	58260
gaaatcacag	gtctacattt	cgtcctctaa	actactctgt	tgggtagaat	ttattttgca	58320
aagacttatg	taggatcact	tttttactac	aggttttgtc	atatgggatt	tttacaacct	58380
ttttttctgc	tgaacaaaat	ggcttttaat	ccttaaaagg	gcagggctat	attttccttc	58440
aaacattttt	aaaataaact	aagagaatta	attttttagta	atagcaagtg	aagaacattt	58500
taatcctaga	gcttaagaaa	ggggaggccc	aataaccaga	tgctggaaat	ctattgaggt	58560
tttttttaaa	ttccagtatc	cagacaattg	gcatgaaaat	aaaggagcct	agaaaaaatg	58620
ttgaaaatga	aaacaataaa	agtgacaact	aaacatatta	ttttgtattt	gcaaagcact	58680
tccacaaaatg	gccacaaaatg	tcattctctt	aaattttatt	cctgagttga	gtaggaaaata	58740
aaccatttgg	agattcaagg	tttagcttaa	gatatgggtg	tttcttgaac	ctgggaggtg	58800
gaggtttcag	tgagctgaga	tcattgccact	gcactccagc	ctagatgaca	gagcgagagt	58860
ttgtctcaaa	aaaaaaaaaa	aaaaaaaaaa	aagatatgga	tgttgtcaat	acaatcgggg	58920
gaaaggaata	ctttgaacta	ctttgtttgga	aggagtttga	aatcgttgag	gactcagcag	58980
catgaagtag	agaaattcac	aattggtaga	aaggactatt	gtccttcaac	cttcattaag	59040
gttaactatt	caaccttcac	taaaaacaga	aagtgacaat	ttcacagcaa	attctagaac	59100
tttagatcaa	aagtcaactc	aatatggggg	atztatataa	gaaagagtta	aaaaaaagac	59160
gaaatgtaat	atctatgtta	ttgcaagtga	aaggaaaaca	ggaagataaa	tatcacaaga	59220
agacaaaaat	gtatctaaca	ttttgggaca	agattgtggg	atccacagaa	aattggaact	59280
tggaaacttcc	tgttccacag	agataagaaa	tacacttgct	tttatctcac	ttctcaaaaa	59340
aagtaagatg	aatgggggtt	taggccccag	agagaaattg	tagctgcaat	caattgtact	59400
atctgagtaa	aaattgtcct	cagaggaaaag	tgagtaggga	gctgtctgaa	gggacaggtt	59460
attaacaaaa	gagaggggata	atggattgcg	tttgcaagtg	cagttggggc	taacatcaat	59520
gccatcttca	tagctgggtc	aaaaaaatat	tctggattct	tttagtgtct	tggttcttac	59580
ctggttggtg	tgcaagaagg	ataaatgtac	ccttaaaaga	gattagggag	agaagtgcct	59640
cccacagcac	cacgaccaga	aagggaagag	gaaggacagg	caatagccaa	ggactcctgg	59700
cagtgaactc	atgtccacat	caagatctaa	tgagcttgca	ctcaactcat	ttctagctct	59760
gccttggaag	ctggagctcc	tgcactgact	atcaatgtga	gcccctgagt	aggagcagct	59820
tggtagagtt	gaaagaccat	tgatctgggt	caacagactc	tggttcctgt	ctcagcagtg	59880
ctgtaatcaa	tccaagtcaa	atatcatctc	tgggacttaa	tttgctaaat	ttaaaatgaa	59940
aagaaaaaca	aaaatagaac	aattagacta	gatcaggatt	cggcaaacca	aagcctgctt	60000
atcaaatctg	gcataccacc	tgtttctgtc	tatacaattg	tattgaaact	cagccacact	60060
cattcatttg	catattgtcc	atgaaagaag	cttttgcgct	gcctcaggag	atctgagtag	60120
tggccacaga	gatgttgcag	tggaccatgt	tgcaacattg	tccaaaatac	ctaaaatatt	60180
tacttcttgt	tttgagagagt	ttgctgactg	gcaccagaga	aatctatggt	ctaaaatcat	60240
ctaaaaatth	aagcacatat	gtgtgaccac	acatttcgaa	atgccgtctt	cccaatctag	60300
aacacagcac	atgaatcagt	aggtaaacct	atccatgtca	attctcaaat	ttgaaattca	60360
ttcacttgaa	aatccagcaa	tttttcatgc	ttcatatcat	atctgggttt	ggaataaaga	60420
agtgtgaggg	gagaaaaatt	ccctgagcat	tttaaattcta	atttcacctt	tattatgaga	60480
ctactgagtc	ttttcttgag	caaaggagag	agtgtgaaat	agaataaggt	gctcaaaaca	60540
atagattaaa	tttattgaaa	ggatgagtaa	ttggagtaat	gttacagaat	attaagcaga	60600
ttatttagat	agcatataca	tttccagttt	gattaagtca	attcacaggc	catcaaaaag	60660
tacacagaaa	aatagagggt	ttgatccgat	agccttctgc	gttagaatgt	agtcatttgc	60720
tctctttagc	tatttaattct	cttctgcact	ttcagaggga	gaaaatgagt	gatgagagag	60780
agaagtaata	agatcatact	gcaaattcctt	acatcgataa	gataacagaa	ggcttttcac	60840
actggctaatt	atttcatggt	gtattttacag	ctcttgtcca	tgtacagatt	tggggctcatt	60900
aacagaggtg	ttgatcagca	aagtcttaga	gtgagttctag	ggaaatggtt	gccaaatata	60960
tctgggtatt	ggaaccatca	acagagcctg	tgccccagac	ttactgaatc	atggtctctg	61020
gtggtggatc	ccaggtagtc	acagctttta	ctgagtgatt	aggagattca	ctgtgcagcc	61080
agttcgggaa	ccaccagtga	aatccaaccc	cttcattttat	tagtggttga	accgtaactt	61140
tctgaagatg	taccactgtc	ccagccactt	cacctgtatt	cccagctcat	ttctctcaga	61200
gaaaactgcc	aagtccccac	agtggccttg	catgggccat	ttgtgatgtc	acctctctga	61260
cctcatctcc	tactttcctt	caccoggttg	tgtctagcta	ttgccatgct	ctttctccaa	61320
cacaccaagc	acactcctgc	atcagaggct	gtgaattttc	catttcctgc	tttgttcctc	61380
cccaatattt	gcatgaacca	ctttcccaact	taatctaatt	ttctgttcag	aagtcacttt	61440
cacagaaaaag	tcttctctcag	ccacatgtat	tttcttcttg	gtgcttatta	tcacctgaca	61500

ttttatatat	ctatgtttgt	ttcttgtttg	tttctactc	cattcggtaa	gctctaattg	61560
agaacagaaa	agctctgttt	attcactgtt	gtatcccca	catctatatc	tagcacatca	61620
ttgattctca	acaactattt	gtaaatgaat	gaataaatgg	ctctgctaga	tttttgtcac	61680
ctggattgcc	agaccagtgc	aataatggaa	agctaagtaa	tgtaaagagc	tttgcaatca	61740
gacagctctg	gcttgcatte	ttcatgtgtg	atthttgagca	agttactgat	tcttgctgag	61800
catcagtttt	ctatgtgtaa	agtggaggca	tggacactgg	aggatcatgg	tgagaatttc	61860
atagagcagg	ccaataagac	ctccttgggg	aaagaaggct	aaagtggggc	tcgggttttag	61920
gtttcccca	ctgggagttc	caaggggcta	tggagttgat	aggattcccc	atthgggctg	61980
cagaggagcc	agaaacatgc	tttagttttc	taatcctgaa	ggagtggggg	attgtgtaa	62040
cctcattggg	gcccttttga	accacatttc	atcaagatat	tgtgaaatgg	ggagcagcag	62100
gcagccttct	ctggccaaca	tgtatcaggg	cctcgtgctt	agtgggatgt	gccagagta	62160
gacagggaa	ccctggcagc	cttgcatagt	ccctctcatg	cctttgcttt	catggggctg	62220
gctggaaca	ggccggtaag	ccccatggaa	ctaccgggac	aaattaccaa	gcaatctccc	62280
atggccatgt	agaacagcca	tgtagaacag	gatgtctccc	aaggtaagat	tttaaggtttg	62340
tggtttgag	aaaagagaag	ggaggaaata	aaaaagcact	gacaatcatt	gaaagcccac	62400
gttctaggta	actggctact	cactttcaca	gccagtaaac	aattttgctt	aatcctatac	62460
atctataaat	aggattattg	tcttcatttt	ccagataagg	aaacaagact	tggcaaatgt	62520
gtagtgtagt	gtttgacagt	attatgtagt	tgttttagcat	gcacatgaac	ttgcaagacc	62580
ttgctcttcc	ctaataattc	tgtcctttac	taattctgtg	atcttacaag	ctgtgcctta	62640
tgtattttgt	gttctatttt	gtaaaacct	atgacaacaa	cataataata	ccttctcata	62700
gaacaattat	tgtgataaat	taagctgttg	agagtcaggc	actcaacca	acaacatgcc	62760
cataatgtgt	actcaataaa	tgcttgctag	tatttttaat	aatattgtta	ataaaataac	62820
ccagccagag	gtagaaccca	agattccgac	tctagttgac	ctgattctaa	gccactctt	62880
ctacagcaac	acatcggccc	attgatttag	acagcatgga	actcatagcc	cagggaattc	62940
acactcttaa	agaacttaaa	cagcacagct	taatattggtc	ccttccaaac	atacgactca	63000
g						63001

<210> 2
 <211> 97335
 <212> DNA
 <213> Homo sapiens
 <400> 2

aaatgaatta	atthttacaga	cattacactg	agttaaagaa	gacagacaca	atgaggacct	60
attatatgaa	gttctacaac	aagcaaagtg	aatctgtggg	ctatgggtgat	agaaggaaaa	120
aagggtattac	ttttttgtgg	ttgcagaaaa	acttactgga	tagggccatg	agagaactat	180
ctggagtgat	ggacgtattc	tatatcttga	tccatatgtg	gttacatgat	tatatatatt	240
tgccaaaatg	tattgaaact	aattattagg	ggaaaaattc	agggtacata	aatgctttct	300
taaatatata	gggtcaagat	aagaaactct	ggataagaat	ccaagttttc	taatatcagg	360
tcatatactc	ttgtcttgaa	gccataggtc	ttcggcctaa	agctactaga	gaatccgggt	420
tgctgacccc	tagacaaatg	ccatgactct	agtagccagt	gtaacccttt	tccctttgaa	480
gttgtagcta	ataaagtcta	gagttcaaag	tcatgtcctg	catgggcaaa	gtgaatgttc	540
tagaataaga	gaaactatgc	tgagaatgaa	agagtttttt	tgagacaata	gaataatttt	600
aggctcctgaa	ccatgtgaga	aaagtgttct	aagaaaagct	ctcaagatct	agaatgaggc	660
agataaccaa	gacagcttga	aataaacctt	gtaaatggac	tggggtagct	tatcaggatt	720
tagagtcatt	ctaattgacac	ttttgacaac	agtgtccaca	tgattgatgc	tataaaactg	780
ataacggaag	gcgatgccct	tacaggaatg	gcattttggc	aaaactcaat	atcttctcac	840
agttcatttt	ttataagatt	ctattcatgg	actgcatact	tatacaatct	cattagtctg	900
ccattaatca	tttctatgtc	tgtattctgt	agttttcaag	acttgagttt	gcatacatga	960
aaacaatatg	ctccattcat	ataacatttt	ggtttacaag	gcatttggga	tctgaatcat	1020
gagacattca	ggagaggaga	ggtgatgact	tctgtacct	tctgtagggtc	gttgggtgag	1080
ctaaagccat	gatggagagt	cagggtcccat	gagggtccatg	ctcaaatagaa	acctgaacaa	1140
ctaaaggaat	gggaaaaaca	gagatgctgg	ctttgagggt	gatgaatttg	tgagcactca	1200
ggctctctga	tgctgtgatt	ttgttactcc	tctgtaatca	gcttatgcat	aataatgtca	1260
cacctctctt	tagaatacat	tagataatac	atthtagaata	cattataacct	ccacacaggt	1320
ataacgtatc	ctggaagaca	taaagtgtcc	tcttaccttg	tcctgaagta	tttgcccttat	1380
gcacagcatt	ccaagccagg	aagctcttgc	taagtttaaa	aatggatgga	tatgggccgg	1440
gcgtggtggc	taacgcctgt	aatcccagca	ctttggggagg	ccgaagaggg	cggatcatga	1500
ggtcaggaga	ttgagaccat	cctggctaac	actgtgaaac	cccgtctcta	ctaaaaatac	1560
aaaacattag	ccaggcgggc	acctgtagtc	ccagctactt	gggagggtga	ggcaggagaa	1620
caggagaatg	gtgtgaactc	gggaggcgga	gcttgcatgtg	agccgagatc	atgccactgc	1680
actccagcct	gggcgacaga	gcaagactcc	gtctcaggaa	aaaaaaaaaa	aaaaaaaaat	1740

ggatggatat	gaaggtggag	gcagagttgg	gttaaggaat	cattacacat	gaagtatcct	1800
agttaccaga	aaaaggtcac	tggactgaag	caggaaactt	cttgttttgg	ttctttaagg	1860
tattctactc	agcataacag	ctcattctaa	tactctgtat	tagtttgcta	aggggtgcat	1920
aacaacattc	cacaaaccta	gtggcttaaa	caacagaaat	gtattgtctt	gtagttctgg	1980
aggccagaag	tcacgatgtc	gctgggtgtg	gctcctcctg	agggctgtga	gggaagggtc	2040
tattcctgat	ctttctttct	ggtttgtgga	tgcccttttt	cctgtgtctc	ctcacaccat	2100
cttccctctg	tacatgtctt	atccaaattt	cccctttgta	taagaacact	aatcatgttg	2160
gattaaggtc	caccctaata	atttcatttt	aacttgacta	ccactgtaca	gactctatct	2220
ccaaatacgg	tcacactctg	gggcagtga	ggttaggact	acaacgtatt	tgttttgggt	2280
gacacaattc	aaccggtaac	atatacctgat	ccttggtctc	ctcatctgca	aaacaagtat	2340
atttacctta	cactgttatt	taaaaattga	aaagcagcca	atgtatttag	aagaacttag	2400
taaactgaag	attctacata	agaacaaggc	agtaaaaaa	gtgagcgata	acagcataga	2460
tatacataat	aataagagct	ctctcttact	gagaacagcc	atgtgcctgg	cagttctaaa	2520
tgtattcact	tatgaatcta	tgagatctcc	tatgaagatt	gatggtagca	ttatctctac	2580
tttccagatg	aggaaactga	aggacagcaa	gtttgggcaa	tttgctaacg	acacagcttg	2640
taagtgggat	ggccacttgt	aagccactcc	attagaagcc	agagcctgta	agcttaatca	2700
cgtccctctc	ccacatacag	atactagtga	gagaaccaac	tttaaatttc	cagaataaag	2760
attgtttcaa	aaatagaaat	agatattaag	aaataaaaa	agaaattatt	tctatagaca	2820
gctaggtaaa	tggaaaagca	atagttaaaa	tgtgaacata	ttttaagtat	gatatactgt	2880
agctaaagat	aacatgcagt	acttttaagt	aatactgtat	atagagaagt	agatcgaaat	2940
aggtaaacta	ttaacataaa	ttttcctgta	gtggaatccc	aattaaagca	ataatgtaga	3000
gcaataagaa	gcaaaatctg	cttggatcca	aatcatttct	gaactttgta	actttggtcg	3060
tgttgctttc	tgtgaatcag	tttcttcatt	taaaaaagtt	gttatgagga	tagaacatgc	3120
tagtatttat	aaaatgctca	taaattcctg	gcacaaagta	agctatagac	tcaaaaatgg	3180
tgattccaga	cctcaagaaa	taagaacttg	aatttcagag	aaatcttccc	ttattatccc	3240
tgagggccac	tccatagctt	gatcaaagaa	gtgggatact	atgtagtaga	ctagtgtggc	3300
agcaccttaa	acccacaaat	aagtttggct	gttgtaaaga	gaagggaac	caaaattatt	3360
tcaggaccta	ctatgtgcag	agtttttatc	tgctatatac	ataggcatat	gctattccat	3420
cagtggttca	gagggcataa	acagttaata	aagctatctc	agggatgatt	aaattgagag	3480
ttgctcaggg	tcatacaatt	tttgagtcag	gaattaaacc	cagggctatt	tcattccaag	3540
gccatgttta	ctttctggta	aaacaaaata	ccatgaaata	taatatattt	tctttgaatc	3600
cctaatagtg	ggtgctatta	tacccccaat	tacttaaat	gatcctgtac	attgccaaga	3660
gattaaaatta	taaattattg	agttcaaaaa	gttaatttct	tcctccaaag	agatggcatc	3720
ttacagttta	attactctga	cagctttttt	tttttatttt	ttctcctagt	gaattatcaa	3780
cagaaactac	ctagagaaac	ctgtaacttc	agatcttctg	attaattgat	tttctttggg	3840
tgctgcttga	cagtaacaga	atggtccatg	gggacagttt	ttcagtgaat	gttttcaaca	3900
atgataggaa	cacagagcta	agcagagtgt	caagagacat	ttagtgcagt	agagacatct	3960
catctgtttt	tctgcagaac	ttgcataaaa	ttagcatctg	attattatca	cctcagaatg	4020
aaccggaacc	cgtaaagatg	tctttgatcg	taaaagtggg	gaggagaccc	tgatggaaga	4080
agagaagact	tcagcctcac	aatagaaaag	actgtgttta	tgacacctaa	tctaacttag	4140
acaccatcta	attcaagaaa	caggcacatt	tagggaaata	aagaaaagag	gcagaaagag	4200
atgaagcacc	ttggctgaat	cagagaaatt	aggggttaat	gttcagact	gatgacttga	4260
aggaggggtg	gcattttcat	gcaagggacc	attctgtaga	agaagcagca	cattcacccct	4320
taaatgagca	aagagaatcc	caatcactct	ccctctcacc	ctcccttttt	ccttcatttc	4380
tctagctctt	tattttccaag	aaggccagcc	cttctcttca	gccacatggg	atttggacag	4440
cctaggtctc	atccctgcag	ggaaagctca	atcagccagc	tgagccagtc	actgaacaac	4500
caagtctggg	agaacagggt	gtgaagtatt	taagaagcct	atcgtggctg	gagggctaaa	4560
gccttgttct	caaatcagct	ttataggtaa	taaaactgac	attaaagcaa	aaacaaataa	4620
gagtaataag	gcttcttctg	tgtgtttctt	gggttttgtt	tctcatttga	ttccaaatgg	4680
ggagacaagg	gtatatttat	cctacccttc	ccccagcctc	aacaggttaa	atgggttaat	4740
acatgaaaag	catcaatcac	agtgtctgga	gcattatcag	tatcagtaaa	tgttaactat	4800
tactttgag	ctggcccatg	tggctcctgg	cattacaggc	accagaggaa	ttccaggggag	4860
accaggagga	ttatagtctg	ggggagtatg	gaggatgtgc	tactgtgtgg	caggtgcttt	4920
gcaaactgca	ttaaaatagt	cttgatatatt	tcagaatcta	acgggatccc	atatactatg	4980
attctgaaat	tcacttgccg	ttcttttagat	tttgcatcta	catgatgcat	agtaacagtt	5040
cttaattttg	tttattccta	gcttatctat	ggttgtcatt	atttcaagag	cctagtaaag	5100
ttgaaatcaa	ttttagaact	acaaaaaaaa	aaaaaaagga	agagatgtga	gtagggctctc	5160
actgtgttgc	acgtaaatgc	attgacatgg	aagatttaaa	acacctgatg	agttttattc	5220
tgtttcttta	ttctctatta	aaggtagtac	tgtgcaacat	tcctaccct	ctatttttga	5280
cactgacgtt	gggcttattt	tgattgtcta	agaggtcctg	tgctacaagc	tggcagggct	5340
tgctgggatg	aagtgacaag	caaagatttg	ctatatggta	tagccagggg	cagagacagg	5400

ttttctgggg	gcatgagact	tctacagtgg	ggagacacta	ttaagaaata	gaatacaaaa	5460
ttatgactac	aaaatgagag	gccccaaagc	ttatgtttca	ttagcttcac	agggagctca	5520
cttctgcccc	gagccaagga	ctgtggacat	gacctggtaa	tggcatcaaa	tgattccctt	5580
tattgcctcc	cctggggcag	aaatggtggc	agcaaagcaa	cttccaatca	caactatctt	5640
ctttcatttg	acacaggaag	gggaagctcc	ctctttaatc	tctaatagaa	acattaaaca	5700
gcaggccatc	acaattttcta	gctaaagagt	tttcatagga	cctttgcaaa	ggttgcatte	5760
ctcccttccc	ttgtggcagt	atgcataatg	taccttttag	gaatataaat	tcagtgtctt	5820
ggaacacatg	attagaatc	agtgttcagt	taccaactat	caaataaaaa	aataaatcca	5880
gcatatgact	aacggaaaga	agtcaatctg	aaatggttac	atattatatg	attccaacta	5940
tatgacattt	tgggaaaggc	aagactatgg	agacagtga	aaggtcagtg	gttgccaggg	6000
gttagaagga	agggagagat	gaatcgggga	agcacagagg	attttttaggg	cagcaaaact	6060
gttctgtatg	atactataat	ggttgatata	tgtcattgta	catttgtaaa	aacacatatg	6120
tcttggtcca	tttaggctgc	tctaagaaaa	tatcttagac	tgggtggctt	ataaacaaca	6180
gcaattgggt	tcttacagtt	ctgaagtcca	cgatgaaggc	actggcagat	tcagtgtctg	6240
gtgcaggcga	tcctcttgct	ttaacttcac	tggtggaagg	ggcaaagcag	ctgtctaggg	6300
cttcttacat	aaaggtatta	atctcattca	tggggctct	gccctcatga	cctagtccacc	6360
tctaaaaggt	cccacctctt	aaaaccatca	cactgggctt	taggtttagt	ggggatgaat	6420
tttgtgggga	acccaaataa	cagaccaaaa	ttcactggga	attgttttgg	ggagacataa	6480
acattcagac	cacagaagca	cagcacatac	aacaccaaga	gtaaatgttt	atgtaaactg	6540
cagattttta	gtgatgtcac	ggagaaaaac	ttctcctcga	ctctcttagg	tttagtgctt	6600
gggaaccagt	aaattaaact	aactaaagac	aagttaacaa	gagaaaaagc	acaagttttt	6660
attgatattt	acatgcatag	gagttcatgg	aaaagaaatg	aaattcaaag	atgcaattag	6720
actcacaggg	gttatatacc	attttaacaa	aggaaagggt	tggggcttca	aatgatgcta	6780
aattgtgggg	aagtgactag	gaaagataca	gggaaacaaa	ggaaagataa	gggttatttt	6840
actaaggctc	gtttatgtaa	acttgtctga	ctctcaatct	ccagtggtaa	gagtggctct	6900
cctcttccca	gttcaggaga	gggagacacc	ttcacaagag	gaaatgtatg	ccctgatttt	6960
aatctgataa	gaagagggca	gagaaccctt	cctacatttg	gttgttttca	attgccttca	7020
gctcaaaatg	gtccttacgt	ggaagtgaca	tatttttatg	gtggtgggat	attctgatct	7080
ctttcaataa	taattgtgtg	tcaatgagtc	tcattgatta	taatagatat	accgctctgg	7140
tgcaggattt	gatagtaaga	gaggttgtag	ttgctgtagg	gcaggatgag	taggtgaatg	7200
aactttctat	attctccact	caatttttgt	ttgaatttaa	tattgctcta	aacataaaat	7260
ctgttcaata	aataaagtaa	aaaacaaaaa	aataaaaaaa	agtaaaaaaa	acaaacttag	7320
actgaggtaa	gtagaagctt	tattggaaaa	gattattaca	gaggggggag	ggactccaac	7380
tataaaacag	caagcttctc	agagctcagg	cagaaaaagg	cttttctttt	atacgaagga	7440
gagaacaagg	ctagaaagaa	ccagatgtaa	gtaagtgagg	tggcctgatt	ggatagtaga	7500
tcagagaacg	ttttaccctc	aggccagccg	cttctctggg	ggagccattg	agaaggagct	7560
aaattctggc	tcgacttaga	atgcatcaga	gttcaagagt	ctggggaaaa	gaaagaaccg	7620
taaccaaaaat	tttttcagggt	caggtttaata	ggcattctgt	tctaatagat	cactggggac	7680
aataaattta	gctaatacatt	tatgaagcca	agaacaggag	tttggagcgt	ctgtatctgg	7740
ctttgtcata	ggtaaacaaat	gccggcaatc	cttgagcctc	atctaactca	tatcggaag	7800
ggtggtcctt	tggaaataagc	tattctctga	aacagtttag	tctaagtaca	atgtatttca	7860
ccagcaaaata	ccttgattgt	gccttggaag	gaaggggttg	gtgatggtaa	gtctggaaaa	7920
ggccagacaa	agagagggag	aattctgggtc	cctgaagaga	catcccatat	gctctcgggt	7980
gagattgtga	gaaatgctgg	aagaatgaag	tggatgacta	ccaagcccat	cttatttcaa	8040
tgtacataga	taatacttca	ttttgcaaaa	atgaaaagta	tgggacaaat	actaaatcag	8100
acagggcact	gggataagag	gcgcgaatta	ggatatgtag	tcattttactg	tactaatttg	8160
catgaataga	aatgcagaga	gagctctata	aaagaaaaat	attttggtga	atctattagt	8220
atagaaactc	gtaatttaatt	tactaagata	taacgattta	ttcgtcatgt	taggcaacgt	8280
attttcaatg	tcattcttatg	gagaattgta	aagattttct	tgtatgttaa	aggtctttgg	8340
tcaccatcat	tattcacacg	attgttatc	acctggttat	cattattcct	catcattaag	8400
gacactagta	atctttttaa	cctttctggc	tctttctggg	gaaaagggca	ggcttaatac	8460
aggcattcct	catttttagtt	agcttcacta	gcttcacttt	attgtgcctc	acagatacta	8520
tgttttttcac	aaattgaagg	tttgtggcaa	ccctgcctca	agcaaacctc	tcagcaccat	8580
tcttccaaca	acatgggctc	atttcctgtt	tccacgtcac	atttgggtaa	ttctgacaat	8640
atttcaaaaat	gtattattat	tatagctgtt	atgatagtca	gtgagcagtg	atctttgatg	8700
ttactcttgt	acttgattgg	ggaaccatga	accacaacca	tattaggcag	caagcttaat	8760
tgataaatgt	tatgtgtagt	ctaaccgctc	ctccaaatgg	ccattccccc	atctctttct	8820
ctctgctcag	gcctccctat	tccttgagac	acaaaaatat	ttaaattatg	ctaattaata	8880
acccaacaat	agcctctaag	tgttcaaaata	acaagaagag	tcccacgtct	ctcactttaa	8940
atcaaaacct	tggcataatt	aaggttactg	aggagggcat	atagaaagct	gagacagacc	9000
aaaagctaga	cttcttatgc	agaacagtta	accatgttgt	gaatgcaaaa	gaaacgttct	9060

tgaagaaat	aaaaaataga	agatgctacc	ctctataatt	ctatgaaggt	tcagagaggt	9120
gaggaagctg	caaaggagaa	gttggaagct	agcagaagtt	ggttcatgag	atttaaggaa	9180
ataagccacc	tctataatat	aaaagtacaa	ggagaagcag	caagtgctga	tgtagaaggt	9240
gcagcaagtt	atccagaaga	tgtagctaag	atcattgatg	aaggtggcta	cagtaagtaa	9300
cagattttta	atgtagacca	aacaacctta	tgttgaaaga	agacaccatc	taggactttc	9360
atacctatag	aggagaagtc	aatgcctggc	ttcaaagctt	caaaagataa	gctgactctc	9420
ttgttagggg	ctattgtatc	tggtaggctt	aagtcaaagc	caatgttcat	ttaccattct	9480
gaaaatcaaa	gagtccttca	gaattatgct	aaatctactc	tgctgtgct	ctgtaagcag	9540
aacaaactgg	atagcagcac	gtctgtttat	agcatggctt	actcaatctt	ttaagcccac	9600
tgttgaaaa	taatgctcag	aaaaataaga	tttatttcaa	aatattactg	cttattgaca	9660
atgcacctag	ttgcccaga	gctctgatgg	agttgtacaa	ggagattaat	gttgttttca	9720
tgctgtctaa	tacaacattc	attccaaagc	ctatggacca	atgagcaatt	tggaatttca	9780
aatcttatta	tttaaaaaat	acattttgta	aggctatagt	ttctatagat	agtgaattct	9840
ctgacggatc	tcagcaaagt	acagaacaac	atggctttga	actgtgctgg	tccacctata	9900
catagatttt	tatctgcctc	tgccaccctt	gagacagcaa	ggccacacc	tcctcctcct	9960
cagtctactc	aacgtgaaga	tgatgggat	gaagaccttt	atgatgatcc	accccatcta	10020
ataaatagta	aatatatgtt	ttcttcctta	tgattttctt	aatacatttt	cttttctcta	10080
gcttacttta	ttctaaaaat	ccattatata	atacatatga	cataaaaaat	atgttaattg	10140
acggtttatg	ttatccataa	ggcttctgtt	ctacattagg	ccattagttt	agtttttggg	10200
aagtcaaagt	tatatacaaa	ttttcaactg	tgtaggtggc	ccaaatccct	aactcccaca	10260
ttactcaagg	gtccactgta	agttgaaaac	tcctggaaaag	gattcatcac	tgtagatatt	10320
agggctttca	tgattcatgg	ggggaagtca	aaatttcaat	attaacaggg	gtatgggaga	10380
agttgattcc	aacctttatc	gatgactttg	aagggttcaa	agatttagtg	gtggaaggag	10440
ctgcagatgt	ggtagaaata	gcaagagagg	cctggcgtgg	tagctcacgc	ctgtaatccc	10500
agcacttttg	gaggccaacg	tggtagatca	caaggtcagg	aaatcgagac	catcttggtc	10560
aacatggtga	aacctcatct	ttactaaaat	acaaaaaatt	agccgggtat	ggtgacgcat	10620
gcctataatc	ccagctactt	gggagggtga	ggcaggggaa	tcgcttgaaa	ccaggagggtg	10680
gaggttgag	tgagctgaga	tcacgccact	gttctccagc	ctggccacag	agcaagactc	10740
catccaaaaa	aaaaaaaaaa	aagaagaaaag	taagaaagaa	agagcaagag	aactagaagt	10800
gggtcctgaa	gatgtgactg	aattgctaca	atctcatggt	aaaacttgaa	tgaataagga	10860
gttgctttgt	ctgaatcagc	aaagaaagtg	gttgccctgag	atggaatcta	cttctggtgc	10920
acaggctttg	aacattgttg	aaataaagac	aaaggattta	gaatattgta	cataaactta	10980
attgataaag	tggtagcaga	gttcaagagg	attgactcca	aatttgaaaag	aagttttatt	11040
gtagataaaa	cgttaaatag	cagcacattc	tacagagaaa	cctttcagga	aaggaagagt	11100
caatttatgt	ggcaaacttc	attgttgtca	tattttcaga	aattgccaca	gacacaccag	11160
ctattagcaa	ctaccacctt	gttcagtcag	cagccatcaa	cactgaggca	aaatcttcca	11220
ccagcaaaat	gaatatgagt	tgtgaagac	tcagacatct	ttagcatttt	ttagcaataa	11280
agttttaact	aagatagtac	atttttaaga	catatgctat	tgcacatttt	atacactaca	11340
gtatagggta	aacataactt	ttatatgtat	tggaacaaaa	aaaaatttat	atgactcact	11400
ttattgtaat	attggcttta	ccgtgggtgat	ctggacacag	aatccacagt	atctctgagg	11460
aatgcctgta	gtttgaatat	gagtgatgca	agaaggattt	tttagctatt	gtagtatctc	11520
tggatgtatt	tcccacaatc	tttttgggtt	atagttgctc	ttgtatctgt	tttgcctttc	11580
ccaatagtga	ttgcacttag	cagaatttgt	gctaggaaac	attgttattg	ttattgatac	11640
aaaagtaatt	ttctgtaaag	aatagatcat	tttgtctgat	tcattatttt	agtggctctc	11700
tgaaaagtat	ttttcttcat	ttcacttctg	aaacagaatc	tagaaaaaaa	aaataccatc	11760
agctgagaca	tttagaaaca	tctgtacttt	cacacaatgc	atagacaaca	aaccttaca	11820
gaattatttt	taggaactct	attgaaatta	gatattttct	aaaagtattt	gctaaaagag	11880
aatgcatttt	tagcttacat	catattgttt	aatttttaaat	ctatcatatt	gtgacaataa	11940
caagaataaa	aacaatttca	ccctcagtgt	aaaacttaca	gattttgttg	ccgagaaaat	12000
tcaaaacata	cctctgaata	gtaagaagtg	tagtaaaaat	taaatgaaat	gttgggccaa	12060
atatagaatg	attgaaaaat	cacccttgaa	ttataaaaaa	ttgtgctcat	cttgtaata	12120
ctttaaatgt	cttgctcatg	gtaaggactt	cgtgctatct	ttgtacaatt	aacatataag	12180
gcaaatttca	ataccaatga	cagtagatgt	aatttcacat	atcaagtagt	ttccctaata	12240
atttaggact	tttttcaactg	acttcttgct	atatatgatt	ctgtcattac	ttttttatat	12300
agctcaaaat	atggctgatg	gagtgcctgt	gcccttccat	ggttttttca	tttggatatca	12360
ctagtattat	atttatttat	tctttttttc	caggacattt	gcaagacatt	cgtcaagttt	12420
tgaggtttat	ttaagggtata	actagataat	gtattatgta	aatcagccta	accagcacgc	12480
ataaactaaa	ccatgtcaca	ggatgccaat	agtgacagac	ggagtgcctga	tgctctgga	12540
aatctcttca	cattgtggag	tttctcctct	tctcttggga	tatcttctct	attttctgtg	12600
acagatgcct	atttgctaga	tcaactgaga	gcacagtgct	ctataggtaa	tgaaatacaa	12660
ataaaactca	atttgtgtta	tggctctgaat	gtttatatct	ctccaaaatt	catatgtcga	12720

aatccggtcc	ccagtgtggt	ggtattaaga	gtcgaggcct	ttgggaagtg	acagggtcat	12780
gagagtggag	cacttgtgaa	caggattagt	ggtcttatga	aagatggagc	gtctgttttt	12840
cccttcccc	atgggagtgg	gaaaatacgg	atagaagcca	ctatccatgg	ggaacaagcc	12900
tccaccaaaa	actgaatctg	ctggcacctt	gatcttggac	ttcccagcct	ccagaacgct	12960
gaacaacaca	tttctgttgt	ttgtaaatta	cctaatactaa	gacagcagcc	tgagtggact	13020
aggacaattg	gtttcttttt	tgtttaggta	gcaataacta	gaaacagaat	ttctacagag	13080
gcaacacaca	caaaaatagg	tgtgatttta	tcaaattaaa	atacttctgc	acagtaaaaag	13140
aaacaattac	caaagtggaa	aaaaaataac	ctatgggggtg	ggagaaaata	tttgtaaacc	13200
atgtatatga	taaagcggtt	atatccaaaa	tgtatgagaa	attcctccaa	gtcaatagca	13260
aacaaaccaa	tagaaaacca	gaagcaaaaag	aaaaaagaaa	ataagttaac	caatttgaaa	13320
aatagacaaa	gaaactgaat	agacatctct	cgaaagaaga	catacaaatg	gtcaacagat	13380
atctgaaaaa	gtgtggatat	cactaatcag	gaaaatgcaa	gttaatatca	caagatacca	13440
cctcacagct	gttaggataa	ctattagcaa	aaaaaacaaa	agcggccccct	tgctttgtgt	13500
ctgccagaat	ccattatggc	tgccactata	gttctccagg	agtctggata	gcagtcccgg	13560
ctgggaatgt	ctggagtctg	caggccccag	tcactatgga	cagtttttcc	ttcagtattg	13620
gggttctaag	aggcagaaat	tagaaatagg	aatgcagatg	gttgccagagt	gtgggtgggg	13680
gttagagcgg	cgccctggaa	aaaaaaaaaa	aaaaaagatg	agtgttggaa	atgatgtgca	13740
gaaaagggaa	ccccagtata	ttattagtgg	aaatgtaaat	tgggtgcagcc	tctatgaaat	13800
aatgtatatt	tcaatacata	ataatgtata	ttccaaaatt	gtttttatag	agtagatttt	13860
agatgttctc	accacaaata	aatgataagt	atgtgaaatg	atgcatgtgt	tagcctgttt	13920
taatcattcc	acaatgtaca	caagtataac	atttacatta	taccacataa	atatatactc	13980
tatatatata	ctatatatat	atatactact	gtttgtccat	ttaagtaatt	ttttacaaat	14040
aattgttcta	aataataaga	ttatcttata	aaggagacag	taggtttgtt	ttgctccaga	14100
tggtagagtt	aatccctgaa	gctagaagat	gcttaggggtg	aagtttgtgc	tcaaaaatatt	14160
tttctaaaaa	tagagcagac	tgccctcaaag	gtgagggtatg	ttgcagaaaa	aagtcaactgt	14220
aaggtagtg	tgggatgctg	tttggggata	atcaggggact	caggataggg	aactaaggag	14280
tttaggttgt	ctgctgtata	gaggaaaacta	gaggagatag	tttattcagc	catcattaat	14340
tcattcattc	agtaaacatt	actgaacacc	tgtgtgccag	gtgctgtgat	tgacactgga	14400
gattttaaagt	tgaataatat	ctagtcaactg	gggccctgat	gtcaaggcac	atgctctagt	14460
ggaagagagg	cacacgtagc	taagagtgtg	tggaaagtgt	tatgaggaaa	gtatgttcag	14520
gatagtgtag	aggctgagga	agaggattcc	catctcacct	gctgcgggca	aagaaagggtg	14580
gatggtattc	attagggatt	atcacaggag	ggatcaactg	aactaatcct	taaaataata	14640
tacagttagga	gccagcccac	aaaatgctaa	tagaggcagt	ggggagggtg	gtggaaaaaaa	14700
aaaataggta	tgagtcaaaag	gagagaattc	cagacagatg	aaagctgtaa	aaatagttta	14760
gttttactgt	aaaaagggtca	aagccagaag	taaggggctg	caaagtccat	ttaagctgag	14820
aactttatta	cagctcctgt	tccaagccca	tgctctgtgc	gactgcagat	gggatattct	14880
aataccactt	gtggaaatca	atcttttttg	acaaagacca	agagagaccc	atgttgtgat	14940
tcactctcta	tggcaaccca	aaaggaaggg	tgaatgcctg	cagatgagag	ttgcctcttc	15000
atgccagcca	cttaataggc	tatggaaaag	ggaagagcag	tccccatgg	gggcctaata	15060
ggaatcataa	tctaactgat	tatcacagca	gtgactcttt	caaactctggc	ctaactgaag	15120
ctagcaccac	accaggatct	ctgctggtgca	cacaccaatc	acccaggagg	tcagtatcat	15180
ccccatttta	cagatttgaa	agctgahtgca	caaggtaaat	aacagttatg	cagaagtctt	15240
gcttcagtgt	ctaacttcct	catctcactt	tattctgttt	ttcaatagga	gaagagatat	15300
aatctcatga	tgaaaagtgc	catctaaagt	ggcaccttag	acacagtaag	caaactggat	15360
tggaaagtga	gaagtcagtc	tcaagatact	tgacaatgtc	tcctatttgt	agattgttta	15420
gcaaaatgct	ttcacctgag	ttatttcatt	tgttgctcac	aaccaatctg	tctggtagga	15480
aaggcagatg	atatgttcac	ctccagataa	gaaaaatgaa	attcagagag	gccaaaggggc	15540
ttgctcaaga	ttccacagtg	aagaatctac	ttgcaggaat	tttaacaaaag	gttttcaaat	15600
tcagaagtcc	acaggatgac	tgcccttgaa	atccagtggt	aacatataat	aagggtgtctt	15660
tgagcaggta	agagaggacc	aaccttttct	aaagcctttc	tcctcttgct	ctctgtgcgc	15720
caacctgagc	agcctcctga	gagattctgg	aacatgctag	aatttttctt	gctttagggc	15780
cacagttcat	gtcactcctt	cttcctggaa	tgttcttcac	aagcactgtg	tggctccttc	15840
ttatccttca	aatatctgct	gaaatgttcc	cttcttagaa	aggacttgct	gaccacctg	15900
atggaagtgc	cctctccagc	tcacctcttt	actgactctg	tttttattcc	ttttgcaagc	15960
acattttata	atctatatgt	agtgtttaat	gcaacagttt	gctaaggact	gaaatttgca	16020
ttcctccagc	aatcaaaacc	cactgcgatg	gtttgtggag	gtgggacttt	tagggggtaa	16080
ttaggtttag	atgaggttgt	gaggggtggg	cccctcataa	tgggaattagt	gttcttataa	16140
gaagagacag	agagaacgaa	gctctctctc	tttgtgtgtg	tgtgtctctc	tctctctaag	16200
acatgggagg	acacaatgag	aagacaactg	tctacaagcc	aggaaaagg	acgtcaccag	16260
gaaccgagcc	tgttgacatc	ttgatcttag	actttccaac	ctcccaaact	gttagaaatc	16320
attatctgtt	gtttaagcca	cccagactgt	gatatttttg	catgatgcct	gaactgacta	16380

atatacagtt	tacctgtaca	ttgtttgttc	ctctagatta	taagctctcc	aaggcagggg	16440
ccatatagtt	ctttaaaaga	taatgtcatg	gctataagtt	tactgctatt	tccccatcac	16500
atgggaaaaa	gaattgtcag	agttgtcagc	cattagagtt	tatttacctt	aagggtgtct	16560
gacgaaatga	tctgattatc	tccttaactt	cctaaatagc	tcaaacttca	taatgtttga	16620
atttggaat	atctgttgat	tcaagagaga	cagattacat	ggctctctct	gaactttgga	16680
catctcttcc	acaaacttcc	aaatattagg	ccatctatta	ccagaccttc	tgaacttaaa	16740
caaactttac	tttcaacccc	acttttagcca	ccttcaccat	ggcctaactt	ttgcccttgt	16800
tatttgttat	ccatagtttt	ctaaacttag	aatccttaat	cactggcatt	atagttgttt	16860
acttcaacct	ccaacactta	caggtccctc	aaatcctcat	ttattcaacc	cttgtttttt	16920
gtcatcttat	aaactgcaaa	ttcttgaaaa	tcacatgtcc	tctctataaa	tcagatccct	16980
tcatatcatg	ttttcaatct	gagtttggct	ccatgattaa	tcttccaaat	caaagggtta	17040
caaactgtgg	cctgtggaat	aaatctggtc	tgcacctgtt	ttagtaaata	aagttttact	17100
ggaacacagt	catactcatc	catttacata	ttttttatat	ccattttcat	gttacaatgc	17160
cacaggtgac	tagtggcaac	agaatctgtg	tggtcacaaa	agcctagaac	attttctatc	17220
tgtcctttat	agaaagaagt	taccaacccc	taatcaaaat	caataatcta	tgaagtagaa	17280
agtaaaagat	gatcttttgg	tgctgacatg	aacatattat	gattaaaaaa	acactttatt	17340
gaggtttgat	tagcatacag	aagctgtaca	taacacatac	aacttgaagg	gtttggtgat	17400
cagtatgcaa	tcatagaaacc	atcactatga	tgaatgtcat	aaacatatcc	atcaccccca	17460
aaggtttcct	ctgccttatt	tatttatcta	tttataagaa	tacttaacat	aagatttgct	17520
accctcttag	caaatttagc	aactctttga	gtatgcaata	taatatagtt	aactatagac	17580
actatgctgc	acatagattt	ctaggactta	tttaacttgc	gtagcaaaaa	ctttgttctc	17640
tttgcccaat	acctccctgt	ttccccctcc	ttcatccctt	ggtaaccatc	attctactgt	17700
ctgcttctat	gagtttgact	gttttagatt	tctcatatga	gtgggatcat	ttagtatttg	17760
tccttcgatg	tctggcttat	ttcacttagc	ataatgttca	ccagggttat	caatgttggt	17820
gcaaatagaca	agattccctt	ttttattatt	aaggctagat	aatattccat	tgcaaataata	17880
taccacattt	tctttgtcta	ttcatctgta	gatgaacatt	caggttgctt	tcattgtcttg	17940
ggtattgcca	gtaatgttgc	aatggacata	gaagaacagg	tatctctttg	acataactaat	18000
ttcatgcctt	tgggtaaaata	cccagaagtg	aaattactaa	atcacatggg	agttctattt	18060
tttggttttt	gaggagcctc	catacagttt	tccataatgg	ctgtactaat	ttacattctc	18120
accaacagtg	taaaaagggt	cctttttctc	cacttcttca	ccaacattta	aatttcatct	18180
ttttgataat	agccattctg	acagatctga	ggatgatatt	aattgtgggt	tcaatttgca	18240
ttttcccaat	gattagtgat	agggatattg	ggctttgtta	tttaataacc	actggccatt	18300
tatgtgtcct	cttttgagaa	atatctgttc	aagtcatttg	actattttta	atgtaatcac	18360
ttgttttctt	attattgagt	tgtttaattt	ctctctgtat	tttggtattt	agagcccctt	18420
aacagatgta	ttatttgcac	atatttttct	cttaacctat	gggttgtctc	tttattttgt	18480
aaattgtttc	ctttgctgtg	cgtaagcttt	ttagtttgat	gcaatacaat	aactaataga	18540
ttaatggctc	ctttttgttt	ttgttgcttg	tgcttttagg	gtcatgggtc	aaaaaatctt	18600
tgtccagatc	agtgtgtgga	gctttcctct	tacgtttttt	atcttctaata	gttttatagt	18660
ttcagatctt	atattttaagt	cttcaaccca	ctttgagttg	attcttgtat	atgggatgtg	18720
atgtgttcaa	tttcattctt	cttcatgagg	acatccagtt	ttttcaacac	catttgttga	18780
gaattcaata	aagttgcagt	atataaaatc	aacgtaaaaa	atcagtaggt	ttttcataca	18840
ctgacaatga	actatctgga	aaagaaatta	agaaaataat	cccattttaa	taccatagca	18900
aaataacttag	tggtaaattt	aaccaagcaa	gaatatggaa	atatctatat	tctaaaaact	18960
ataaaacatt	gatcaaagaa	attgaagatt	acacaaataa	atggaaatat	gtctcatatc	19020
cataagttga	aaaccgatat	tgttgaaatg	tccatactaa	atgtgatcta	aagagtccat	19080
gtgatttcta	taaaaaatcc	aacgtcatct	tttatcgaaa	ttgaaaaaaa	aatgctaaaa	19140
tttatatgga	accagagaag	acctgaataa	acaaagcaat	cttgagccat	agaacaaaag	19200
ttggaggcat	cacactacct	gattttcaaaa	tatattgcaa	aactatagta	atcaaaatgg	19260
catagtactg	gtgaacaaac	agatacaaaa	accaatggaa	taggatagag	agcccaggaa	19320
taaatccaca	aattaaaaac	caactgattt	ttgacaaga	tgccaagaat	ggggaaggga	19380
gagtttcttt	cattattttat	tctctgtatt	cttttcatct	caggaaaaat	atccagtttc	19440
ctcaattgta	tatccattca	ctcacccgat	aatttcttca	ttcacttatt	tgtttattca	19500
tttaattctca	attgtttgtt	cattctgtaa	atattcagat	ttctttttat	gcattttctc	19560
agaattggaa	gcataatact	gaacaaaata	actataaatc	tcagcctccc	actcccatat	19620
ttacagtttg	attagggagg	cacattttaga	tatgcagtga	taattgcttt	gcttagagaa	19680
attcaagggtg	atagaatgca	tggtgacacc	taaccagac	tggtagagaa	agggaattct	19740
tccactggga	atgacatgat	tatctaaata	agtaggctca	atcagggtcag	gaaaggccct	19800
gaaagactat	ttcaagcaga	gggaagggtat	ttgccaaggc	cagggtgtgt	agtggagaga	19860
atgggcagtg	gcagagaatt	atgaagtgtt	ccaatgacta	aaagtaaagt	aacaagttcc	19920
ttgtccaaga	gcttggaattg	tatcctaact	gaaatgagta	tacactaagt	gtttgaagaa	19980
gagggatgaa	atgggtcaagt	tttcattaca	caaaaataac	ctgttccttt	cattttatgt	20040

ttattttattt	ttttaattttt	ctgactgctc	ctttctggaa	atctcaaatt	tatatttgcc	20100
aaatattgtc	acatttttcga	tggagaatac	aaactaagaa	tgggttaggg	aactgagtc	20160
gaaagtcctt	gttgatcaat	tcaatcatgt	ttttctaagg	atgtgctttt	ggacattatg	20220
gaaactatct	taggctctca	cttgatcctt	agaaaagaag	gcacctgtta	aaaggaatgt	20280
ccagccccac	ctaatttggc	agctgcccct	ccaagcta	gacattaggg	atgtagtgg	20340
ttcaagaggc	tgatgatgcc	atctgggtcaa	ctcatgtgac	tatttctatc	agctttattt	20400
ctgcaacatc	ctgtgcccac	agaggggaca	caaaatcgct	ttataattcc	ttcaccatgt	20460
gaagatacag	atacaccacg	aaccttaaa	gcaagaatat	gattgaaatg	tcaaattggg	20520
acttggtgat	ctaaattatg	tcccccaaaa	gccaatgtct	tgccaccacc	agtgccttat	20580
gggtggagtt	tctaaacaga	ttactcaaaa	cacaaacttt	caaaaaggga	aagtcataac	20640
cctctagtca	tcagggcaat	tacggaataa	cattgctgga	gtaagggttt	ctcaatgccc	20700
aagagatgag	ctggcaatgc	cacaacaatg	tccaatttct	agtgggtcca	agaccatgtg	20760
ttacattttc	ctcccatgat	tactcacagc	ttcacagttc	tgtgtctctc	ttcgctctct	20820
tgccacctct	taactgcacc	tttgacctcc	tacccttaag	attcaaccct	gtgagattac	20880
ttgtcttttc	atctacactc	tggctactct	gacccccatt	ctttagattc	aggattttct	20940
cttttctctg	cttactctcc	agcacaagta	gaaaaatatt	gtgcttcatt	ggaaaatgca	21000
tgttgtttga	atcacactct	ttcagattat	acaattgtag	tctttcatta	tctttgagcc	21060
attttataat	gctgtaaaat	aatattaatg	catataattc	ttgtctatag	acatgtaaat	21120
tgtgtccagg	ggtgatttaa	ttgacttttc	ttccccattg	tggaaaaggc	tagtagtttt	21180
gcctccattt	gcccatttat	ttctatatct	ctgatctgta	aatgcatttc	tgggttttcc	21240
cttggaactc	ttctaacatc	tcttcttttt	cctcattaat	taattaaatt	aaatctttta	21300
cacctcttca	tatttttgtc	agtatgagag	tataaaaaaa	aatctttttt	tgaaaattat	21360
cttttaatat	gtgacaaaaa	gaataaacac	tgaacaagaa	gtcctgagct	cctactacaa	21420
tccaaagtga	actagttgca	gtgccatagg	atgatgtttc	tccctgagact	tctttactag	21480
gatcattttg	actgtttata	atgcagattc	ctgggcctct	ccttagaccc	actgaatcag	21540
acactctaag	aggagggtca	gaaatctcca	tcttagcaag	ttaatgtcca	ggagagttag	21600
agactcacta	tgacgtatga	taaaaagtag	aaggaaggac	tttccagcac	cttaaactac	21660
ttggaaagat	agattaatgt	ctagatgaaa	cttgatgaga	ctttagacta	atatgttaca	21720
tatacatcag	gagatgcatg	tataaaccaa	atccaaataa	cccaaagcaa	aatttctttt	21780
agaaatagt	ataaataaat	gagtgaggag	tttgtcactc	acatccccc	ggtaaaacat	21840
acctttttag	cctaataaaa	tgtatagtgt	tgaatgtgtt	ccctccaaaa	ttcagggtact	21900
gatagttaac	agcagccgat	atgatgggat	tacacaaaat	agtagggtct	ttaggaggtt	21960
gttaggccat	gagtgccctc	cccaagaatg	ggatgaagga	cctcataaaa	gaggcttctc	22020
acagcattgt	gacctctagc	cctcccacct	tccataagtg	aggacacagt	gttctctccc	22080
tctggaggat	gcagcaacaa	agtgtcatct	tggaaagcag	ggagccatca	ccagacaaca	22140
gaaccagcca	acagcttgat	cctggacatc	tcattttctc	cagaactggg	ataaaaataa	22200
ttcctgtttt	ttaataaaat	tcccaatttc	aggatatttg	tgatagcagc	aaaaacagac	22260
taagacaact	agtatgaaaa	tatacattaa	caaataaaat	taatcataat	atatgtttgc	22320
ttttaaaaga	aaataaataa	gccaatatgc	tttctgttga	ttgattgatt	tactaaacat	22380
tgattggcca	tctccactgg	ggatatggca	tttaagagat	ctctttgatc	ttagtacttt	22440
tactgctttt	taaataggat	caaatacacc	caaggtaaaa	aatagaacac	actatacggt	22500
acatttttga	actgtttaga	attcctttga	agctaaaatt	actgctatca	tttgacaact	22560
tttaccctta	aaataatgtg	gtgctcacca	gcttgcttaa	gttacagcac	ttgctgtctt	22620
ctcagatata	atatcagaaa	cttataatcc	aagaaaaatc	taaatggcaa	gtgtgagtta	22680
atggaagcct	cataaagcaa	gaggtgtttt	ggaagtgtat	ggaagacatc	aataaatgat	22740
atgtataaca	tcaagtgc	aagtgtgtgc	taggaaagtt	caaaaaagaa	aaaaaatatg	22800
gtaaggtaag	accagagatt	ggggagtata	tagcttttgg	gaattcagga	aatgctaacg	22860
tctatggaga	atttgcatag	gtgaaagatc	agattggaag	cctttctctg	tggaagcatt	22920
gtggatctaa	gtccagaagt	gatcctgagc	cacctactat	ggaaagggtg	cagtgaacaa	22980
gagactgtct	gacaaagggt	gaagctgagc	agacttctac	tgcgcacgc	ctatgtacag	23040
gccagattcc	aagggtgat	attacactgc	tagtttgatc	tttctcagat	agctgggtag	23100
agagggaaaa	gtttcacccc	aaataaccaga	tgcctccaaa	catctagatg	cttggtgtta	23160
tgtatttcaa	acagcaggtt	tgataaagct	gtttttactc	tccagttgga	tgttggttgt	23220
caaggctgtc	attaacttct	gggagttccc	aaatcctcag	agagagagaa	ataagcagtt	23280
ctgcctgcaa	tcagaagctg	gctttgagcc	ccacatggca	acatgcctgt	atttaattag	23340
aagtggctta	gttcttggga	tttactgtgg	attgagtcta	cgaagctgga	aggttttatc	23400
ctgagtatct	cactgtctcc	tcatgtggat	tttatcaaat	ttctccttta	accatataag	23460
cgctttcaaa	gttgacatat	cacacttagt	tatgagggga	acctagaagt	atgattgtga	23520
cgtgcattgt	tgagagacca	tgttttgtaa	gtgttaataa	ttaaaattat	tttgaacact	23580
aattatgctg	aagactggat	tttctactcc	ttcagatgtt	ttccagtgga	catattttgc	23640
ccttttaata	aattgattgc	gaacttatct	tcatttcacc	tttatgatgt	tatacctttt	23700

catttttgtc	ttctcatagc	taggggtctcc	tggtccccaa	atgtagacac	acatcttaca	23760
ccaatcccag	agccattttg	tataagagcc	accatggatt	taaccagctt	tagctccagt	23820
atttgaacat	aatgttcagc	atcatcacct	ggccaccaa	tcaaaactga	gcacccttta	23880
atccatcaac	aagttctctg	cagccatgca	aggttatgaa	atgggcacag	acatcaatat	23940
acagtctttg	tgtttaagag	gttcatggtc	tacctgagaa	atgcatcttt	aaacctaaag	24000
tagacgctct	gtttattcca	taaatgattt	ttaagcatca	atggtatatc	aagcactgta	24060
ctggcttctg	ggctataata	aatatataaa	gaccacaagt	ttgaatttca	tgacattgaa	24120
ctataatgtt	taaatgttat	aataatcata	gtaaatgtcc	ttgaggagct	acggaagatt	24180
cctgcatgaa	gcagaaacaa	gaagctgaag	aaaaaacaac	tggttttggg	ggctatataa	24240
atataaccct	caaattaaaa	actcaatagt	ttgattgatt	accaaataca	tacagccaga	24300
aaataaatgt	attaccccaa	aagcttgagt	gaagaaagac	ttctgtaagt	tactagaaag	24360
cgctaaggaa	gaaataaaag	aatgatattc	ctgagagcca	ggtgctttct	gtaggacaca	24420
cagatacaga	atgagggaaa	acagaaaatt	ctatggttgt	ggattcaaaa	tagaggcatc	24480
accatgtcta	tctcatcagg	gtttaaccag	aggaaaccaa	accagtaaga	tatctgtatt	24540
aagagatttc	ttgcaaagaa	ttgacctatg	tgattgtggg	cactggctag	gtaaatccta	24600
agtccacaga	gcaggcagga	agaaaacagg	ctgggacttg	taggcacagg	atgaagctgc	24660
aatactcatg	tgatgctgc	tcttcttcag	ggaagacttg	gctctgctct	caaggacttt	24720
caggtgatta	aatcaggccc	actcacatta	tctaaaataa	tctcccttac	tcaaaaccaa	24780
gtgattatgg	actttaatca	catctataaa	atatcattat	agtaaacact	aaattagtg	24840
ttgaataact	gagagttgta	actgatatgg	tttggctatg	tccccacca	aatctcatct	24900
tgaattatag	ttcttataat	ccccatgtgt	tgtgggagga	accaggtgga	gataattgaa	24960
tcataggggc	agtttctccc	attctgttct	catgatagt	agttagttct	caggagatct	25020
catggtttta	taaagggtct	ccctctttgc	tccactttca	ttctccttct	tctgctgcc	25080
atgtgaagaa	ggacatgttt	gcttccccct	ctgccatgat	tgtaagtttc	ctgcggcctc	25140
cccagccacg	ctgaactgtg	agtcaattaa	atctcttttc	tgtataaatt	accaggtctt	25200
gggcagttct	ttatagcagc	atgagaatgg	actaatcacg	taactttacc	aagtggacac	25260
ataaaactga	tcattacaat	gtacagtga	tatttgggtga	gttaatagat	atattcataa	25320
ctgaatgaaa	gaggatgggtg	attcctactt	cagggtggta	ttatgagagt	taaaagggtt	25380
agcatagata	gaacactttt	ctatgattga	tctaagggtg	gcttttaggg	actacattat	25440
acatatgggt	tttgtcttca	tgtttttgtc	tgcatttctc	tctctgcccc	ctcttggggc	25500
tatttgatg	tcacctcttc	ttcttctgag	ctccagctca	tgtattccct	gttccatcat	25560
cacaaccact	gaagctctaca	ctcctcatta	gttcagggga	accagatgc	cagccaaaga	25620
catccctttc	cctcatgacc	cagttgaaaa	aattctccac	ccttctgtta	attctagaat	25680
atttaggagt	aaagatcttt	tccctcagat	atctgtcaat	tcccggcctt	atatctcaaa	25740
gccccctttc	atccaggttt	cctagatcat	ttcttgtcgt	ccctgtgcta	taactcatta	25800
accagggcat	aaaactcatc	agctactaat	gtcctcttcc	atcccagact	cccctcacac	25860
caataattct	tgaatagagc	aacccaaatt	gagaatttta	agtaaaattg	aaggtaaatt	25920
taattgcagg	aggacataca	acattaacct	ttaaatagag	atcctaattc	ttaaaaaaaa	25980
aagtcctcat	tgatctcggg	gccgtttcac	gcagggtgct	ttctaaatca	tcaaggttat	26040
cttcagcttc	actttcagtc	tccttcttcg	gctctggcgc	tgccactggc	tcttctgtg	26100
ctaattgtgt	tgtggcagct	gcaatggctg	caggggtcac	agcagctgca	gcagctgcag	26160
ccacagcctt	ttccttcttg	tgttttttgt	gcttcttgtg	cttcttatcc	tttttgtgtt	26220
tctgtcctt	ctttttcttc	tttttctttc	catctccttc	ttgatctgaa	tttcttggca	26280
gtgatgggct	cgggtgttggg	cttttggcct	ttttgaccgg	tacagctggg	gaccagtttg	26340
tagacgggtga	ctgagactgg	atgggggatg	gagatgctgg	gggcttttta	gctgctggct	26400
caggagaccc	agagacagat	cgggaggatg	agaccctcct	tacagactgt	gggcttgggg	26460
aagcagcctt	tttttatctt	tttaggttcc	ggagtctctg	agactctcct	aatgggccta	26520
gtacttggag	acggggactg	ccttcttttg	ggtgatgacg	acgctcctct	tcgaacgggt	26580
ggaggacttg	aggtctgagg	agctcgaggc	cgtgatgagg	gcgaatgcca	tttgtttgga	26640
tgcggtgatc	gggcctccc	ggtagagcgg	cttggggaag	accctttcct	atgcttggt	26700
gatagtgaag	gtgaacgtct	cttgggtgact	ggggagcttc	tttgggaagg	ggagaatggg	26760
agaccgcgcg	ctttggtggt	ggaaatggtg	atgctcttcg	tttaggaggg	ggaggaggtg	26820
aagccgttct	tctctttgga	ggtggagaag	gagagtatct	cctctgtatt	ggaggagagt	26880
atcttctagg	agaagggtgag	cgcgcagcta	ggggaggaga	aggagtctt	cgtcgtgggtg	26940
gtggtggtgt	gggagtctctg	agccgtcgag	gatgaggggc	gggagaagga	gaccgtcgcc	27000
ttctggtggg	tggtggggat	ggacttctcc	tccgtctacc	atgaggggaa	gtctcttttt	27060
ggtgctttcg	tggtgatgga	gaagcactcc	gggaagggga	atgtctccgc	cgcttgccaa	27120
cctcaccatt	cttcacatgg	gatctcttgg	gtcgttcata	ttctgaggag	aaggagggaac	27180
cagagtcaga	tgaagactgc	tggttttgtc	gtctgtattg	gcgtctctgc	tgagatcaat	27240
gaggaaggcc	caagtgtccc	cacagtctta	gggggaaatg	tttgttatga	tgtaaatttt	27300
atttggtttg	tacgcagttc	aatttcaaaa	ttgctaaaaa	gtggttgagc	tttagactat	27360

aacatttgtt	gtaataattg	ctaggttgaa	gttcaacatg	taaaaaaagg	gggcatggat	27420
ttacattgca	aaaggtgtcc	acagtgtatt	agtgacattc	tttcattgac	agctgacata	27480
attcattgag	tgaaatat	taagccaaaa	aaaattccct	ttttaaaaaa	gggggtttta	27540
atactgttga	cacttttatg	gttcctttta	atgctctggc	tattcccaga	gggggttttt	27600
tgtttgtttt	tttggttttg	atlttgctttt	tgtttttctt	tcttcttctt	acattttttt	27660
ccatttgagt	cttagctccc	atlttaagtta	tgcttctgac	cttgatggg	ctgtaagctt	27720
gcccagaaat	aagaccactg	ttttgaacta	ccacaaaagt	ataaatgaat	atltttaatgc	27780
cacagtcttt	cctgttgcc	gtggagtctc	tgctgaaatg	aatcaggatt	cgagctctag	27840
gataagacag	aaaatgaaag	catgttggtt	gccaggacac	tgtgggttta	tattgatgtg	27900
taacaacttg	atlttggaaca	ctggactctc	attctgttct	tctgggtttg	tttttttgtt	27960
ttgttttttt	cttttgtaaa	ggccatgaac	tagtcccaga	aaggattcct	tcagttacat	28020
acaatttgtt	taatgaaatg	tcattggctct	gttcatat	ttgtcttgtt	cttccaattg	28080
gtgtatacaa	ctttcagagc	ctcttgat	tggaaggctg	gaagggccca	gactttggaa	28140
tagtgtctcg	gtttcactgt	ttttgttttg	atlttttttt	tttatttttt	ttaaactaaa	28200
gctatataaaa	gcttggtgat	taaacagaat	aaatttctaa	atlttaaaaat	ttaaaaaaa	28260
aaaagtctat	tgtcttccct	cccctaccct	aagcaatatg	caatagtggc	tcttcaatag	28320
tcccagactc	ttcttctctt	cctggactgc	ccatctcctg	atcaaccctt	aatlttctctt	28380
ccttctctca	cccttctttt	caggattgaa	ttaatgaatc	cttcttctct	actcatgcag	28440
agtaagtttc	tgctccctg	ggtctttctg	tttactgacc	gcaacaactt	cagattatac	28500
ctcttctact	ccaagtgtt	tcaaagaaag	tctctgccca	agacaaattc	attacgtttt	28560
ttccctctac	ctgtttgctt	ttattctctt	ttgtatttca	tcttctcatc	tagattgaat	28620
aatctttgag	agcacagatg	tttattttata	tttttctttt	ccatttctac	tcagcatgag	28680
gtgtccattg	aacaaacttg	atgaattttt	attgcttaat	atcttgctag	aggtggggag	28740
agaggttggg	ggcgggttaag	gaactatcag	ctagcctagg	agatattaga	gctgcagaga	28800
tttggtctatc	ttgttcaacg	ttatatccct	agggattagt	acataggctt	gcaaatagca	28860
ggtatgaata	aaaaattatt	gaatgagtaa	atgaatttaa	aatataagtt	acttaggcgg	28920
tatcttcagg	catatctgtg	tttatgtggt	attcaatggc	ccacaaatgt	ctacatccta	28980
attcctaaga	tctgtaaaca	ttaatlttgca	tgacaaaaga	gactttacag	atgtgattaa	29040
atgaaaggat	tttgacatgc	agataatatc	ctgtattctt	catgtggaac	caatgtattt	29100
acaaggggtcc	ttataagtaa	aacagagaag	caggaaaatg	aggttcgcaa	aaaaaaaaaa	29160
aaaaacaacac	atgaagacag	agaagaggtt	agagtgtatg	tggctttaga	gatggaagga	29220
gtcacacgct	gtcttaaaag	aataagacaa	gctgtcttaa	aggaattgtt	ataaaggaat	29280
agctgaagct	gggtaatltta	ttttaaaaaag	gtttattttg	ctcactattc	tcatgtctgg	29340
aaaagtttaa	tattgggtag	ctgcatctgg	caagggcctc	aggctgtttc	cactcatgtc	29400
agaacgtaaa	ggggagctgg	tgtgttttaga	gatcacgtgg	ggagagagga	agcaagagag	29460
agggaggagg	ggccaggctt	tttttaaaaa	accagctctt	tttttaaaaa	aaaattatac	29520
tttaagttct	ggtatacatg	tgcagaatgt	gcagggtttg	tacataggaa	tacacatgcc	29580
atagtgggtt	gctgcaccca	ttaacctgtc	ttctacatta	ggtatttctc	ctaattgctat	29640
cccttccctta	acccccaaac	cctgacaggc	cctgggtgtg	gatgttcccc	tccctgtgtc	29700
catgtgttgt	cattgttcat	ctcccactta	tgaatgataa	catgcagtgt	tgggttttct	29760
gtccttgtga	tagtttgctg	agaatgatga	tttccagctt	catccatgtc	cctgcaaagg	29820
acatgaactt	atccttttta	tgggtgcata	gtattccatg	gtgtatatgt	gccacatttt	29880
ctttatccac	tctatcattg	atgggcattt	gggttggttc	caagtctttg	ctattatgaa	29940
cagtgtgca	gtaaacatac	gtgtgcatgt	gtctttgtag	tacagtgatt	tataatcctt	30000
tgggtatata	cacagtaatg	gaattgctga	gtcaaatggg	atlttctggt	ctagatcctt	30060
gaggaattgc	aacattgtct	tccacaatgg	aacaaccagt	tctcttaaga	ataaaaagtga	30120
gaactcactt	ccctggcccc	agagagagca	aaagcaattc	atccccatca	cccaaacacc	30180
tcccattagg	ccctacctcc	aacattggga	tcaaatttca	acatgagggt	tttaggggac	30240
aaacataccaa	actatgtcac	aagacaattc	atgtaagcag	cctctaaaaag	atggaagagg	30300
caaggaaaaca	gattctcccc	taaagcccac	agaaggaaaag	cagccctgcc	aacttcttga	30360
ttttaacccta	gtaagaccg	ttttggactt	tggacaccaa	gagctataag	atgattatgt	30420
tgtttttaagc	cattaagctt	gaggcaattt	gttacagcag	caattggaaa	ctaatacaga	30480
tcacattcta	attcaattag	tattgttccc	agttctctgg	acctcagatt	tctttcttga	30540
aaaacatttaa	aaataatacc	tgaaagtltt	gcacacgagt	gcagagtgcc	tatttactag	30600
agagatcagc	atlttgtttag	gctctgaata	gatttgagga	tgaaattaaa	tagcataaat	30660
aaagtctcta	gtgatgcttc	tgataaaaaa	atatctcctt	caaaatgccca	gaggcagggtc	30720
ctaaaaacccc	acaaagcagg	tgaactggca	aaagactgta	aaaagcaaag	tagaggttcc	30780
tcttcaaaga	ctttcctctc	catctaatta	ggaataaata	gtaacttatc	ttagaacaaa	30840
aatlttattca	aagacctgtg	ctaacattct	gaaatatctg	ctagccgtaa	taaataaatc	30900
gatgtacttt	atgttcttag	ctcccacaat	ttaacctaaa	tatttgccct	ggcatgctta	30960
tactggtcca	agcaagcatt	aggtcatagc	ctgttcctct	tctttatttt	aagggtgttt	31020

tacctttgtc	agcatgccac	aagttacttc	ctccttcctt	tgtttctcctc	tgcccttgac	31080
tcttttaaaag	agtcctaagt	tgctagccaa	tcaggacaaa	tacagaatgt	gaggtcccgt	31140
ttcagccaat	ggaaagtggg	cacagcagga	aggtggatgg	gtcaggttat	aaatgaccct	31200
gtctcctttg	ttcgggtgtac	tcttgtggca	aaactgctgg	caagtgtacc	ctttctgcaa	31260
taggtaaaaa	ctgccttgct	gaggaaatta	aatttatgtt	caagtgtctat	ttcttttatgg	31320
caccggggaa	caagcatttc	taacaagact	atgtaattta	atttcaggaa	cctaaaaaag	31380
tgggatgaag	aactgaggtt	gctaataaat	ctatacaact	tataagttaa	tatttaattt	31440
actaacatat	aataataaag	acatcattgt	aagacaatgt	taaaacattt	tacacatttt	31500
aatgtgcaa	tagtaaatcc	ttcactattc	agggattatt	tggaatccct	tgtcaccaga	31560
agctcttaag	gaaataaact	ctacttcgtt	gcaaataatgt	tcttggettta	gttgaggtaa	31620
tgcaaatact	agaataacttg	tttgttttaac	agcttattct	tccctgaagc	tgttcctcca	31680
gtccctgcca	gtgggatcct	atgtctccag	gagtacttaa	caccctaat	agccccatct	31740
tttaagcctc	cctgggacct	gccctcgag	tacctcttat	acctactcca	cttcctcctc	31800
atggcctcct	gcagaatgcc	attctaaaat	taggttctat	tttcctcgcc	cgcattctct	31860
tttgcaaagc	ctccaaaaaa	tttactttgc	ttctctgcgc	ctgctttatc	tctattttct	31920
acactcgctc	cttctttttc	taattatcta	taataggcgt	cacaaaaattt	gcattttgttg	31980
gaacccaaat	ttccatgggt	gcctcaaaat	atacagatgt	aaattttgcat	ataattaaat	32040
tttgcataag	ggaaactctc	atltggggag	atatgcaatg	cccaataaat	ggcagtttcc	32100
ttcaatgtcc	ccaggccagc	ctcccagctc	gtgtgtttcc	ccctggctgc	agctaccagg	32160
actctgctct	ggggatttac	ggacaagggt	atcaagtttt	aattaaacta	accctctcaa	32220
actgaatgag	tggtctaaaa	tcttcctgta	aagaaaccgc	aaaataataa	tgctggcatt	32280
gagaagtaag	aaaagagcga	gccagcacc	ccacccccca	aatcctgtga	caaggtgtat	32340
ttttgtgttt	tggttttttt	ctttggcagc	attatggggg	aaaagcaatg	atgatcta	32400
gagatctgat	aagaagttag	cccaaaacaa	ggaaattgtt	gagggttctc	tttgaagtat	32460
ggattttatac	ccaccaacct	tagctgcgaa	ccttacctca	agtgttacct	gtgccttgag	32520
atgtttcctg	gtcatagtac	taagctatca	taatgagcaa	gacattcaat	aagcaagtgt	32580
gatggctatg	aggacagatc	ttaacagggt	tttttttctg	gaaggcttaa	aatcatgcat	32640
tactcaatct	aataacttcac	gaaattttcag	taaaacctaa	tgataatata	gaagcttggtg	32700
ttgtagtttt	gtaatcaaca	gcaaaacata	aaattttaaa	aaaacataca	ttactggggc	32760
tgtatcctgc	tacaataata	aggctgacat	aatagatgga	gaacaatatg	gtaacaagcc	32820
aaaatgtatt	acttcatcca	caaatagtat	cgtgctatat	atagacagac	ttgttaaaaa	32880
tttaagaaaa	tacacaatca	attacacaat	agaaaatttg	ctatatgggtg	catggtggca	32940
tgcaactaca	aatgttttcta	acatgtttct	cttcatagga	ttttctgaat	tttcatttaa	33000
tattcaagca	catcaaaaac	accttttcag	gtgtgatcct	atacagcaaa	gctgtcctca	33060
caaacaatag	ttgactaaat	aaacacatgg	ctttatggaa	gaaatgtgta	agtatagcca	33120
ttgttgaggc	agatgctctg	cttaaaaaga	aaaaaaataa	gttaaagtta	tagatctcac	33180
catgcttatt	tactgcttaa	gtcatagcca	atlttattgca	ccaaagttag	agttcaaagc	33240
ataaagaata	ctatatataa	tgcaattaat	gaggttgatg	tccctaaaac	aagagagaat	33300
tagtaaatgt	tttacaatag	ttttcatgag	atgggaaatg	acaatagaaa	tctttgttac	33360
aatgcagatt	ttattgtgga	aatgatctca	tggaagaagt	tttaagagg	ctgcagaaca	33420
ataaagagag	ataacacatt	ttgcttttat	gaaaaagccg	atgttccaaa	tttgctgacc	33480
ttctctgtaa	ataagaagtg	actgtcagta	gccagcagat	gtgtttataa	aaataaccca	33540
cttgatctgt	ccctgcaagg	taagagtggc	atlttcaaa	tgcatgagag	aactgctctc	33600
aagagaaaaat	catactgtaa	agacagcatt	tcaaaaacat	ttatttggaa	atattttaa	33660
gatgtgatct	atlttatttat	ttggtcaaaa	atgcccact	tgccaaacat	tattttattc	33720
cttaaaaacac	agatcaagca	acagatccta	cagttagact	cttctgatca	ccctgagagt	33780
tagttacccc	tttctttgggt	ttcttctctt	taacgtcctg	catatatgct	gccattgtgt	33840
ttatagcaca	atatcagagt	gtgttattta	tacgtgcac	tccctcatta	gattgggagt	33900
ttcttgatgt	cagaacacag	caattgtccc	accagaataa	aaagaatatt	gccaaaagat	33960
gctctataaa	tggttggtcca	ttggttttaa	aaacaaataa	tggaaccaatg	ggctcaaaag	34020
caaactggct	aatataaaaa	taataataac	aaccacaata	ataaaataag	ggctaacact	34080
taaacgggtg	tgtactcact	atgtactagg	cactgatcaa	agtactttgt	acatattttc	34140
ttattttaata	ttcgctacca	tcatattaca	atatactgtt	attaacccca	atgtatagat	34200
gtaggtgaag	aaacttgtca	caaatacatc	agctagtgtg	ctgagatgca	atccatgtga	34260
tttgttcaca	gagctcaggt	tctgtgaagc	gggtaaaaac	aaaatttggc	atccagtttc	34320
aaaaggagaa	ttgcaaaacta	atagaacata	tagcacaaaa	tgattatata	aatagaatgc	34380
taattgcata	tcaaggatat	ttggtataat	acaaattatt	ctaccttaaa	catatggaaa	34440
tttggtggtcc	atgatgttgt	agattctatc	ttcccactct	gcattttcaa	aggcatatgg	34500
tattgactca	ttcgattaat	tggttgatag	tctttattat	agactaaatc	atagaataaa	34560
tacatggata	catgcacgaa	tattatatct	caagggcttt	acatagttca	ttatctcact	34620
tcatagtcaa	aacaaaccta	ctgatagttc	caatgcaaag	cctagaacgc	tttggtcttag	34680

agaggcccaa	gtctttttctc	agtgtctgcac	tgtgtggtacg	tggcgtgggtc	ccctctcttct	34740
tctcagtaca	cactacccat	gcagactatc	actctcagtc	ttgtttatct	caaatacaga	34800
gggtataact	aactggaatg	tatccagaac	agtgaggcca	aagtgtgggg	aaagctcctta	34860
accatgtctgc	tgcattgagga	acagctggag	agactgagaa	catgaggcct	aaagaggaga	34920
ctcaggggaga	tgggatcaca	atcttcaa	atcttcaa	catcaagggg	aaaagagatt	34980
aaacaaggta	atgtagctct	agagagcaaa	tccaagagt	ttgagtggaa	gtgaaagggg	35040
ggctgggttc	agtcagatag	taggaagaac	tttctagtat	ttggtactac	aatgggaaag	35100
actattttgt	gagatTTTTT	taaatTTTTT	tttaattata	ctttaagttc	taggggtacat	35160
gtgcacaacg	tgcaggtttg	ttacatacgt	atacatgtgc	catgttgggtg	tgtgtacccc	35220
attaactcat	catttaacat	taggtatata	tcctaatgct	atccctcccc	gtcccccca	35280
ccccacaaca	ggccccgggtg	tgtgatgttc	cccttctctgt	gtccaagtgt	tctcactgtt	35340
caattcccac	ctatgagtga	gaacatgcgg	tgtttggttt	tttgtccttg	caatagtttg	35400
ctgagaatga	tggtttccag	cttcatccat	gttctctaca	aggacatgaa	ctcatcattt	35460
tttatggctg	catagtattc	catgggtgat	atgtgccaca	tttcttaaat	ccagtctatc	35520
attgttggac	atttgggttg	gttccaagtc	tttgctattg	tgaatagtgc	ctcaataaac	35580
atacatgtgc	atgtgtcttt	atagcagcat	gacttaaaat	cctttgggta	tatacccagt	35640
aatgggatgg	ctgggtcaaa	tgggtatttct	agttctagat	ccctgaggaa	taaatgacca	35700
actattgaga	aattgcaggg	tagtccctac	atgagggtta	ggtagaattg	acctgctttc	35760
tgcctcataa	attttagaaa	attaataaga	taatttatta	cgggggtgggtg	tttgttccct	35820
cagtacttta	tcatctatgt	tgataatgtt	aataattaat	tgcataatta	acaaatagca	35880
aattattgtg	gggggtgtgtg	tgtgtgtgtg	tgtgtgtgtt	tagacagggt	cttgctgtgt	35940
cacccaggct	ggagtgcagt	ggcgtgatct	cggctcactg	caacctgtgc	cttccagggt	36000
caagccatca	tcctgcctca	gcctccctag	tagctgggat	tacaggcgcc	tgccaccatg	36060
cccagcta	ttttgtattt	ttaatagaaa	tgggatttca	ccatgttggc	taggctagtc	36120
ttgaactcct	gacatcagggt	gatccatccg	actcatttcc	caaagtgtctg	ggattacagg	36180
catgggccat	catgcctggc	ccgcaaattg	ttgttattta	taactcttca	atccaaatca	36240
tcagtgtcta	tgttgtttcc	ttaactatca	aatgatgata	ataatagtac	cttcttcata	36300
agatagttga	aaggttttta	atatccatat	ggtactgaga	atgatgcctg	aaacatagta	36360
actaccccat	ttttattata	tttctgttaa	taataatata	taccattatt	gctcttgcat	36420
accatattgc	tcttgcatat	catatatgct	cttgctatat	gctacacaca	gtatttccatt	36480
taggcctcac	atgtccctg	atgtaggcat	taatactttt	attttgcaaa	tgagaaaaca	36540
gtctgtacct	tgtatgccat	gctgtctattg	tttatctgtt	tgaatctcaa	gcaaactctgc	36600
ttgataattg	gtacaaaaat	aagccttttt	ctgggttaagg	aatctgatata	tgtgttttaa	36660
aaaacacaca	tttaatcctg	gggctgtctgc	attactcctg	ctgccccatc	ctactgtgat	36720
caaaggcaca	tacatgagat	gggtagttgt	ccccttgcca	atgagggttt	ggtaagaaag	36780
gaaagtgcag	tacttctttg	tttctgaatt	gcaagtatgt	gtgggttaga	gggggaggct	36840
gaatatgaag	gtcctgggac	agcccaccag	gtatcccatg	agactttgca	aaggaaaagg	36900
aggtgagtga	cagcccaggg	tccaatagga	tagaaggaaa	agccaggcca	tggagtccct	36960
cagacctgct	ttctaagggc	aactctacca	cctcagcaag	ccattgaact	tctctgagct	37020
cagtcctttc	atttataaaa	tgggggtgaca	gtgctcacat	gccaggaata	caaagggatt	37080
gaaagataaa	acacgtaatt	aagcacctgt	tgttacacat	ctgtcaggga	ccccataaag	37140
gtcagctgtc	ttcctgttga	cttctgttct	tgggtggttct	ccaagatcat	accttccatc	37200
aacattttacc	gtcactcccc	caccccatgc	ccaatactga	acagtggagg	gacgcttcac	37260
ctacagttat	aatgttgaaa	cttcaaccca	aagcaagtac	tgttaggata	tctggaaact	37320
ttccctcaaa	taagggtattt	gaatgggaca	agaagaagt	ttacagatag	ccaatggaga	37380
tgattttaatg	gggttatgat	agaaacgaga	aagtaaaaca	aacccatgct	ttaaagtcta	37440
ccattttcagg	tccatatttt	cgttgaaaaa	ttgagattcc	tattaaacaa	tgacattttac	37500
acaaaaaagt	agaggagtgtg	gttgaaggac	agggtaatgc	caggagggaat	tgggaatttg	37560
agagtcaagt	caaaggactg	aaatactcag	aatactaagg	gcacctcagg	gctctaccaa	37620
ggacacgtag	aagctttgaa	tttgacgac	caccctaatt	taacgagcta	cctcagcacg	37680
tagtggagcc	ttgaaaaaca	gatgtcacaa	actctcatta	gattgtcaaa	cattttccag	37740
catttctctct	cccatcatag	ctggttatca	agatatatat	acacacacgt	gcatacacat	37800
aaataccttg	ataagttact	agagaaagca	gaaaaatgtc	tgacagttta	atgagatttg	37860
ggtgaaagaa	aattctatat	ttcattgttt	tccaggcact	agaaataatt	catcaatgtt	37920
tctaagactc	attcagcgtg	gctgcatttt	ttaaaatatt	ttcataaatt	ttgaggagca	37980
aataccatta	ttaggcacta	aaaagggttg	agtctaata	attagccgct	tcatcctcct	38040
tactcagct	cagcatcctg	tcaactggct	cttactgggt	aacatccaca	cgcctcctga	38100
ctggctactc	agtgccgatg	acatttccct	cacacacagg	gctgggttta	agatacattg	38160
aggtgacatc	aggtggcctg	taaagtggtc	atttttaggat	atcctattca	aagacatctg	38220
tggaaagtgtg	gaccaattta	ttgatgaata	acagtgaagg	ggtttccacc	agcaagtaac	38280
ataatttttt	acaatgatga	tgtgtgaagta	gaaagagtgtt	ctagtccagg	actggacaaa	38340

tcaatttgca	gacgattttt	aggaagaaaa	acattgcaac	agtaaattgt	aattgataac	38400
ttctagagcc	actttaagta	ctgctatttt	aggattctga	ggggaagaaa	gtgttctgca	38460
aagcaataag	caaagtgatt	tggtccaagc	cccaaattta	agcagtttga	gaggtaaaaa	38520
gagtcattac	caatgtgggt	atagaacatg	tgctaggcaa	atctctttca	catacatgtg	38580
gggaggtaat	aaaattataa	tttgaggccg	ggcacggtgg	ctcatgcctg	taatcccagc	38640
actttgggag	gctgaggtgg	gtggatcacg	agatcaagag	attgagacca	tcctggccaa	38700
catggtaaaa	cctcgtctct	actaaaaata	caaaaaatta	gctgggcatg	gtggcacaca	38760
cctgtagtcc	cagctgctcg	ggaggctgag	gtgggagaat	cacttgaacc	tgggaggcag	38820
aggttgcaat	gagttgaaat	cacgccactg	cactccagcc	tgggtgacaga	gcaagactcc	38880
atctcaaaaa	caaacaaaaa	ttatagtatg	aaataggcat	taaaatattg	tgtatttttag	38940
aggagactga	ggattggagg	ctgaagaatt	actctaaatt	aatcagcttg	tgtacttcag	39000
agctaagata	gctctttggg	ttctaaattc	tgtgatcttc	tttttgattt	ctcttggagc	39060
aataatgaag	gcaaaacatc	aataaacata	acaaactggg	taagggagac	cattgagaag	39120
gactaaggac	accttcaaag	ttctgagtga	gtttaaaaag	aagaatgatg	aaaactttga	39180
tagaaatagg	aaaaaaagta	gaggaacttg	tttggcttga	aacttcttaa	tgttttaggt	39240
aatttatattg	aagatgacag	tggtcattga	gaaaacaaaa	tccccaagc	aattttggaa	39300
taagagccaa	catttaatac	ttaccagaca	actattctaa	gtattttact	atattcactc	39360
atagcaactc	taaaaagcag	gtagtattaa	cagagaaaaat	gaggcacagt	gaggttaaat	39420
agctggtccg	aggctacaca	gctaatacag	gggagagttg	ggacttagac	ccagagggtcc	39480
agttttgaag	tccacacttt	tagccattac	actacaatgg	aaagaaattt	agaagatata	39540
cacagaaaac	tataggcaca	tagattaggg	gttagtagaa	tgctctgggc	agttaaagga	39600
actcttctta	aaggaggtaa	agcttgaatg	agactgttag	taagctattt	ttcactcatt	39660
ggtgaatgat	gttttgtgca	gtgtgttttt	ttccccatag	aaaaataaga	aagaaaagaa	39720
aattgagaac	tctctctata	aaaatgtgta	acatatctca	tattccaaga	gatccttttg	39780
gtagtattaa	tttttatctg	ctcacagtac	tggtctcatt	atttggagtt	aaaaattaac	39840
tcaaccagat	aaaaaaatca	gtgctgtgta	tttgtttatc	tttcaaactc	gtgttctaata	39900
tttaaaaagt	tatttaacag	aacgaagcta	tcagctaaga	caatggcaaa	gccgtaaaca	39960
aacataggtt	gcgttttatgc	gaatggtcag	gtccaaagta	gatgcagaat	atgccagggt	40020
cactaatttt	aatccctatt	cagcccagga	ctatgtacca	taagattact	gctagtgttt	40080
tctgaaaatg	atgtatcaag	gcattttctg	tagaaatacg	aaacagtgac	atacagtagg	40140
gagagctgga	ttgaggcaga	gtagtataga	tggaagtttc	ctgaaagcat	tttggggaaa	40200
catcttttgg	gtatggttct	tggatgaaga	gttgatttat	tagtactgga	aggggtgatg	40260
ggagagagga	agtgaagagt	tatgagagaa	tgaccctccc	gtgatgggtg	gtgggagaat	40320
tattgcagta	tgtacgttag	cattgctatg	tggtgaagtt	cttgggattt	cctggggctc	40380
gtgctggaca	gcattgcttag	ccaccagtca	catgtgggca	ctgagcactg	acaatgtggg	40440
tagtctgaac	tgggatatgc	tgtaagtgtg	aaatacaaac	tggactccaa	agatttagta	40500
tgaaaaaaag	aatttgaaat	atctcattaa	tgatgtgtat	ttggtttcat	attgaaaaca	40560
actttgggat	tatatattga	gttaaaataa	atgtcattaa	aattaaattt	tacttaaact	40620
aaaattttaa	attctatctt	tacctttttt	tttttttttt	tttttttttg	aggtgccgtt	40680
tcactcttgt	tgtccaggct	tgagtgcatt	ggtgcgatct	tggtccacca	caacctccgt	40740
ctcctgggtt	caagcgattc	tcctgcctca	gcctccctgg	tagctgggat	tgcaggcacg	40800
caccaccacg	cctggctaata	tttttatatt	tagtagagac	agggtttctc	catgttgggtc	40860
aggctgggtc	caaactccca	acctcaggtg	atccgtccac	ctcggcctcc	caaagggctg	40920
ggattacagg	tgtgagccac	cgtgcctggc	ctatctttac	ctttttaagg	tagtgactag	40980
caactttaag	attcatatgt	ggctcatgct	gtattttctat	tatggaactg	ccttatgact	41040
ttcaatgggt	aggatggaca	catccttggt	gggatggaga	aatctatcat	agcagctggt	41100
cttgaagggt	ggtggggata	tgataataac	ttagggtggg	aggctcagga	ggactcacag	41160
aataactggc	aaccctgccc	ctgtctgtaa	aaaccccatc	ctggaggaaa	taagttagga	41220
aaaggttttg	cattttgtgg	aatgaaaagt	ctgttgcata	tagcttgaga	cagagcaaaa	41280
agagttgatt	gtcagcttca	tgaagaccag	ggggtctaaa	agacccaggg	atcaacaacc	41340
aatgagagca	gcatggaggc	caagaaccag	gcaaaatgct	gattccagga	ctgggattca	41400
ggatgatttc	cttctatgca	ataatctgct	ccttgaaagg	gtatctaatt	gggcattgct	41460
tttacttgct	gctttcaatt	cttttatgtt	ctttcctagt	aaatattttt	tcttaatttc	41520
attgcagctc	gtatttatcc	tgggaacaga	gagaaatggt	tcacaagctt	aagccagtct	41580
tttaaaaagga	gaatggcagg	actccaaaaa	cagacatgct	gatatgtact	ggggaatttt	41640
taagtgtctga	aacctccaag	acaaaagaga	ctgtgtcttt	attgttctct	gaattactcg	41700
taccagctc	ggtacctgga	acatgatagg	gatcccatag	tggtttgatg	aataaattag	41760
tgactccaag	agtaaagtaa	tcctcaggag	gacaaaggca	gatagcttcc	cttccctatc	41820
agaatgtact	tctcttaaag	cttttcttgg	tataattctt	ggagaatttt	gccttacaga	41880
agtcaaatca	cataccaaag	tgaaaactgg	atcttctaca	aataatggaa	gaatcaactc	41940
tatcaaaaaca	acaattatac	atatgatcaa	tggaggggtt	gtcacgagcc	aggctaagag	42000

ctttacatat	attatctcat	tctgtctatg	ccagagaatc	aactatgaca	tatgtaacat	42060
taaatctcat	tttatagatg	caaaaactgg	ggtgtaaaga	agtcaaagaa	tcagccagaa	42120
tgtacagaat	tagcaaaggt	ggaactggga	tttgaattca	gacagtctga	ctccagacgc	42180
catctccgaa	ttatgcataa	ttatatattca	attattaaca	ttcataaatt	gaaatatgag	42240
ggataatgta	ccttttcatg	aaagctttgc	tcgttgtgtg	gatgagtgtg	tgtacatgta	42300
actgcttatg	tgtgctatca	ctgaggtaga	agacatctct	ctctctctct	ctctctctct	42360
ctctctgttt	tggctacttt	ttagtaagac	ttgtatttga	ttgagttcag	aagtttgatt	42420
atcttttttaa	ctaacctggt	tgtttttaatt	atattaaaaa	ttagtcactt	tcaacatatt	42480
tgcataaggta	attgtttagg	tgatctttttg	gatgatcaga	tgtaatatac	tactactaca	42540
cacagacacc	cagacacaca	cacacacaca	cacacacaca	cgcatgcaga	cacatccttg	42600
agctcaaaga	gcttttctcag	gactacatac	tttacattta	caacaagtca	tttagaaaac	42660
tatcaatcct	attcaaattct	cagcaaaaaca	gagataccag	gctctgcctt	tctcttttaac	42720
tgctcttttt	ttgctgtaga	caaagctggt	tcttgcactg	ctacatatat	aaaaaagtga	42780
caaatcctta	actgtcaaac	aagagaaaata	gtttgataaa	tataataatt	ccataagatg	42840
gcacattata	cactactaaa	ttgtaaggac	agtaaagtta	ctgttaagta	ccaaaaagtc	42900
atgattaaat	gttaagtga	aaaatagaat	atagctagat	ttgaatttga	atattcaatc	42960
tgtataccag	tatgtataga	aggaagagta	tataccaaat	agtaagagta	tctatctggt	43020
ttataatttg	atataatata	aattatttcta	ccttaaacat	atgagaattt	gtggtccatg	43080
atgttgtaga	ttctatcttc	tcaccttgca	tttccgaaga	catatgggat	tggctcatta	43140
gactatttgt	tgaatagtct	ttattctatt	atcatagaaa	aaataaatga	gtgcatatat	43200
ccatatacaa	aatagaggtc	tgttcttctc	gtatataatt	atactaaaaa	aactgagact	43260
ttttttttaca	gttgtatata	tacaaacata	tttgtttatt	tatatacaca	tatataaatc	43320
aatttttatgt	acatgtgggt	atacatatat	ccatgcatat	aactctgaag	tgtgactct	43380
ctaaagaaa	cccaggtatt	ggtcagaatt	catgctcggc	tcaggagtat	agaattaaga	43440
gatacaaac	tcaaaaaaga	gggaaccgaa	tcttcaaate	tgagccacct	tacaagaatt	43500
tttaaggtaa	ctgtttttaag	tgtaaacatt	atggcaatgt	aatagtataa	tttgtatggc	43560
acaagatgga	gtcctggtgg	ccagaactga	gtatgggaat	agtcagtgct	aatctttgtg	43620
caaagcacaa	aggaaaattg	gtatacaagg	cctgagagag	aggtcaagga	agcaaatact	43680
ataaagtccc	agaaggcagt	tggataggac	aaggggatcg	gaagagcagg	atcaagatca	43740
gagagcagg	aagatggggc	tagctaattct	agggaaatga	gggagaagg	ataccagaa	43800
gcaggattca	gggggttaga	gcacttacac	catgctcatg	gccatgcca	tggacaggag	43860
aagcatatga	tgggtgcaca	tgggtggagct	ggctctaaaga	agagggggat	tcatatcaca	43920
gtaaaattaa	gtgcaattta	tccacagagc	agaggaggtt	gtgaagagca	ggcataggct	43980
atttttcatg	tctgaatgct	ctcaagacct	cttagtggtg	taggtagatg	acgcacaata	44040
aatattcctt	gatttgaatt	ggcataaaaag	gtcaaatcag	aaagccagaa	gttcttttaag	44100
gtttcaaact	agatttctaag	aagtccaaag	gcatctcaga	ggcatctca	gggagaaaag	44160
gggaagcaat	gaggtagaac	tctgaacctc	cactctacct	aaacaaggca	gccctgctgt	44220
gctccgagtc	atgtatggcg	acatcaggga	agatgtaaaa	taggattgtg	ataggaaaca	44280
gcatttgaaa	gccattgctt	tacgaaggga	agcgtagaac	ccttttcccc	ttgttagcca	44340
attcaaataga	agacttttgg	gagctagtga	agagaaagac	aggatttcta	gggagatgtt	44400
tcagaagcag	cctaactata	cccgtgtctt	cagaaagagc	agtgtcgtct	cagaagtaat	44460
caccttcac	aaccagcagg	tcagtgtggg	tctcctgaag	agcccgaaca	accacgggaa	44520
gcgacatcca	ctgttgtgca	gtcaaaaagaa	tctttgtctct	cattttctcc	actctctttc	44580
cttcccaaatt	agggtatcat	aggaagatcc	tgcttttctt	ccagttccaa	catttatgaa	44640
gtgaaatttc	catcagacag	ttgcttttga	caaacaaaga	ttgactagaa	gcttctgtga	44700
gagagcctta	ggaagtcttc	tggggaagcc	ctgcctttgt	ttggttttct	ttgccttttg	44760
ctctgtgatg	tttgcgtgta	aaaatgatta	ctttcagggtg	attaaaagtg	gggaagaatg	44820
gtttcaagct	tttcatgtag	caaataatat	ccctgtctgt	aggattactt	tagttgaaaa	44880
aaaaacatgg	cttcagtgat	gccttctcaa	tgtacaagat	tcagagggaat	ggaaagaaaa	44940
tgaataaagg	ccggcctcgg	tggctcacac	ctgtaatccc	gggaagccga	ggcgggtgga	45000
tcacgaggtc	aggagttcaa	gaccagcctg	gccaagatgg	tgaaacccca	tctctactaa	45060
aaatacaaaa	attagccagg	catggtggca	ggcgctcata	ctcccagcta	ctcaggaggc	45120
taaggcagag	aattgcttga	accccgagg	cggagggttg	agtgagccaa	gatcatgcca	45180
ctgcagtcca	gcctgggcca	cagagcaaga	ctccatctca	aaaacaaaaac	aaaacaaaaac	45240
aaaggaaaaag	aaatctgcag	ttaatatttt	ggcaagcttt	cttcacttgt	atgcattttt	45300
aaaatgctaa	tgttaataac	agttcgggac	ttetaacttc	tatatttaag	caacaaataa	45360
ataaattgtc	agatgggtact	tcatcatcct	tctctcccat	cttcttagaa	atataaattg	45420
ctttaggtgg	gaatgctata	attttagacc	agaaaatata	tgccagatgt	ctcttatatg	45480
aagccgtccc	gccaaggat	atatatatgc	cttagtcatt	aggatgtgtt	ctaaataata	45540
ctgcaaagcc	cttggaagga	tgggtctgaa	cactcactta	tatttaactg	ctggcatgtt	45600
gctttgtccc	tgtgtcttgt	gctactattt	ccattgatgt	aaagggaagca	ccaattaaat	45660

aacactccat	tattagagaa	ccaggcacaa	gtcagctgag	gcaggagacc	cgccttcttt	45720
tccagaaaca	atgtaaagcc	tgggtgggtg	agggtctctg	ggcttccgcc	gtgccttgct	45780
tttgacattc	tccagcacac	cctataaaca	tgtctaaggc	tgtcctgttt	agtctgatta	45840
ttcaaaactat	attgtccagg	gtagagcaaa	gggaaacct	gctgaaccct	ggagatgaca	45900
gcaggggagag	agagaggggc	aaagaagggc	aaaacgggaa	aaacaggaaa	caggctagtg	45960
agaagagtaa	aaacgctcag	ggtgaggaag	cagggtttct	aagctctcta	atctcccctg	46020
tgcagctggc	ttgctgtatg	gtttatacaa	atccagtggt	gatctctgtg	caacgtggta	46080
tcacctgttt	aaagaggtct	catcttccatt	ttcaaagagg	aatacatggt	tttttactta	46140
ctcttctgca	tggctgactc	cttttcatgc	tttaagtctc	aatcttaatg	ccacctcctc	46200
cttccagacg	ttcccagcta	aagtggcaact	tcccagcccc	attactctct	atgtttattg	46260
cctgcatagc	tcttattttgt	aatgatttctg	taatagtttg	atgatgatca	tgatgaatat	46320
tactttacct	atltatggcc	tctcttttag	tattaaattc	tgtaagccac	atgagcatgg	46380
ggacacatct	cctttgtcac	tgccccattg	ctggcattta	gcacaagcat	ggtctataat	46440
agataccaaa	caaataatgta	ttaatcatgt	aaatgactaa	atccatgaat	gaatctatca	46500
gacagtgtag	atagcagcac	ataaaggaaa	gggaatgtag	taaatttttc	attttccttg	46560
aagatgtagc	tatgtattag	gaatttgaaa	aatacattat	caaacacaaa	gctaaattat	46620
gccagctaata	gactactaaa	tataataaaa	tcggctgggc	acggtggctc	gcacctgtaa	46680
tcccagcaact	ttgggaggcc	gaggtgggtg	gatcacgagg	tcaggagatt	gagaccatcc	46740
tggcaaaca	ggtgaaacgc	tgtctctact	aaaaatacaa	aaaattagct	gggcatggta	46800
gcaggcacct	gtaatcccag	ctactcggga	ggctgaggca	ggagaatcgc	ttgaacccag	46860
gaggcagagc	ttgcagttag	cagagatcac	accactgcac	tccagcctgg	gcgacagagt	46920
gaacctctgt	ctcaaaaata	aataaataaa	tataataaag	tatgtagaaa	gtcagaaatc	46980
ttggggatta	tattgcaaag	aatttccact	atattgataa	tggagaaagg	cttttaatat	47040
tatatttttt	gaatattaag	aaattggcat	ctactcacca	gtttggacat	cgctttttaa	47100
atacacacta	aacgaaagcc	atlttgtact	tataagtgtc	agattaaatt	cctgcatagg	47160
ctgaaaaagg	tctctttcca	tgcctttcca	aattttacaat	taacaaagag	ttaattattc	47220
tcagagtcata	ttcttccaat	tcaccaatta	ggatgagggc	tatttgttac	aatcataaaa	47280
gaggaaatgg	tgcattggga	agaagaaatt	tggaaaggaa	atgtgattgg	aggaattata	47340
ttgaaagggtg	aaacaaggga	gaaaagataa	agagaagaaa	aattagaaat	tggaaacaaa	47400
gttattccag	ccccctctcta	ataactacta	ctctttggaa	caagggaagc	agtacctgac	47460
aagaaatttt	ttttttttta	tttttttatt	ttattattat	actttaagtt	ttaggggtaca	47520
tgtgcacata	gtgcaggttt	gttacatatg	tatacatgtg	ccgtgctggt	gtgctacacc	47580
cattaaactcg	tcattttagca	ttagttatat	ctcccaatgc	tatccctccc	ccctccccc	47640
acccacagc	agtccccaga	gtgtgatgtt	ccccctcctg	tgtccatgtg	ttctcattgt	47700
tcaattccca	tctatgagta	agaacatgca	gtgtttgggt	ttttgtcctt	gggatagttt	47760
actgagaatg	atgattttcca	atlttcatcca	tgtccctaca	aaggacatga	actcatcatt	47820
ttttatggct	gcatagtatt	ccatgggtgta	tatgtgccac	atlttcttaa	tccagtctat	47880
cattgtttgga	catttgggtt	ggttccaagt	ctttgctatt	gtgaatagtg	ccacaataaa	47940
tatacgtgtg	catgtgtctt	tatagcagca	tgatttatag	tcctttgggt	atatacccag	48000
taatgggatg	tctgggtcaa	atgatatttc	tagttctaga	tccttgagga	atcaccacac	48060
tgacttccac	aatggttgaa	ctagttttaca	gtcccaccaa	cagtgtaaaa	gtgtccctat	48120
ttctccacag	cctctccagc	acctgttgtt	tcctgacttt	ttaatgattg	ccattctaac	48180
tgggtgtgaga	tggatatctca	ttgtgggttt	gatttgcatt	tctctgatgg	ccagtgatgg	48240
tgagcatttt	ttcatgtgtc	ttttggctgc	aaaaatgtct	tcttttgaga	agtgtctgtt	48300
catatcctcc	gcccactttt	tgatgggggt	gtttgttttt	ttcttgtaaa	tttgtttgag	48360
ttcattgtag	attctggata	ttagcccttt	gtcagatgag	taggttgcca	aaattttctc	48420
ccattttgta	ggttgccctgt	tcactctgat	ggtagtttct	tttgctgtgc	agaagctctt	48480
tagtttaatt	agatcccatt	tgtcaatttt	gtcttttggt	gccattgctt	ttgggtgttt	48540
agacatgaag	tccttgccca	tgccatgtgc	ctgaatggta	atgcctaggt	tttcttctag	48600
ggttcttatg	gttttaggtc	taacatttaa	gtctttaatc	catcttgaat	taatttttgt	48660
ataagggtgta	aggaagggat	ccagtttcag	ctttctccat	atggctagcc	agttttccca	48720
gcaccatttta	ttaaataggg	aatcctttcc	ccattgcttg	ttttctcag	gtttgtcaaa	48780
gatcagatag	ttgtagatat	ggggcggttat	ttctgagggc	tctgttctgt	tccattgatc	48840
tatatctctg	ttttgggtacc	agtacatgc	tgtttgggtt	actgtagcct	tgtagtatag	48900
tttgaagtca	ggtaatgtga	tgcctccagt	tttgttcttt	tggcttagga	ttgacttggc	48960
gatgcgggct	cttttttgggt	gccatatgaa	cttttaaagta	gttttttcca	attctgtgaa	49020
gaaagtcatt	ggtagcttga	tggggatggc	attgaatcta	taaattacct	tgggcagtat	49080
ggccattttc	acgatattga	ttcttcctac	ccatgagcat	ggaatgttct	tccatttgtt	49140
tgtatcctct	tttattttat	tgagcagtggt	tttgtagttc	tccttgaaga	ggtccttcac	49200
gtcccttgta	agttggattc	ctaagtattt	tattctcttt	gaagcaattg	tgaatgggag	49260
ttcactcatg	atlttggctct	ctgtctgtta	ttgggtgtata	agaatgcttg	tgatttttgt	49320

acattgattt	tgtatcctga	gactttgctg	aagttgctta	tcagcttaag	gagatactgg	49380
caaaaaccac	atgattatct	caatagatgc	agaaaaggcc	ttgacaaaat	tcaacaacccc	49440
ttcatgccaa	aaactctcaa	taaattaggt	attgatggga	catatctcaa	aataataaga	49500
gctatctatg	acaaacccac	agccaatatc	atactgaatg	ggcaaaaact	ggaagcattc	49560
cctttgaaaa	ctggcacaag	acagggatgc	cctctcttac	cactcctatt	caacatcgtg	49620
ttggaagttc	tggccagggc	aattagggcag	gagaaggaaa	taaaggggat	tcagttagga	49680
aaagaggaag	tcaaattgtc	cctgtttgca	gacgacatga	ttgtatatct	agaaaacctc	49740
attgtctgac	aagaaatfff	atagtctgat	gaaagggatt	ctaaagagtc	agggggccaca	49800
ggtctcaggc	ttcgactgga	tgtgatcatg	tctgaggcct	ttcgatcctc	actttcctta	49860
tctggaaaac	aagaatagct	gaatctcctt	ctaagggcgt	ttgtgatatg	aactgagatc	49920
ttgcatatga	ctgcaccaag	tctagctcaa	ttcgcattag	ttccctccat	tatacccctc	49980
cctcgagctt	tacccagact	cagaagaaaag	ccaggcaaca	tttctacttc	tctatatgca	50040
aaaacaaaag	caaacaagtg	gaaaacctca	caaaaacagt	taacttcaac	atltgggctt	50100
acacaaaaca	ttcaaaaatc	tctttttatt	tcatccgccca	tgattatagt	tattttttcta	50160
aagtgaatga	ttctacttcc	caaatgcagt	aaacccactg	ttaaagatag	ttaatltttcc	50220
tctagatgat	tgtggccctt	gaaagtcate	aaggtcataat	ttttaattat	ttccccagaa	50280
tttttcctga	aacagtgtcc	ttttgtctaa	atcaatccaa	gtagggttta	gcattagtca	50340
taaagagggg	gctgtcaaca	aagaaatcaa	ctgagtggaa	gtgatattat	aatgtaaaata	50400
acttgacata	ggaatacata	acactcataa	atattttattg	atttatttta	taaattaaaa	50460
attaatgctt	atgatatgta	aacaagatat	aacaaggcag	tcaagaataa	ctctcttttag	50520
ttcatatgat	tttttcccat	acgtaactga	atagcaagaa	aacaagtaac	ccagtttgaa	50580
aatggacaaa	aaactgaaat	agatattttct	caaaagaaga	catacaaatg	gccaatagga	50640
tatttttttaa	atgttactag	tcatcaagga	aatgcaaate	aaaatgacaa	tgaactatca	50700
ccttacactt	gttagaatgg	ttactagcaa	aaaaagacaa	gggataacaa	ggttggaat	50760
gatgtagaga	aaaggggaatc	cttgtacatt	gttgaggagg	atgtaaatta	gtatagtcac	50820
tatggaaaac	tgcattggagg	agcttcaaaa	aactgaaaat	aagcctacca	tgtgatccta	50880
atactgggta	tatatccaaa	ggattggaaa	tcaatatggt	gaagagatat	ctgcattccc	50940
atgttcgctg	cagccttatt	cacaattgcc	aagtatgaaa	ttggccttgag	tgtccatcaa	51000
cagatgaatg	gctatagaaa	acatatacac	agtggaaatc	tattcagcct	taaaaaagaa	51060
ggcaatcctg	tcatlttgcaa	caacatgatg	aacctgtagg	acattgtgct	gagtaaaaata	51120
agcctgtcac	agaaagacaa	atactgtata	atctcatatg	cagaatcttc	aaaaagttga	51180
acttataaaag	gtagagagta	gagtgatgtt	taccagaggg	tgggggtggag	aggggtgggg	51240
tacaggggaat	gggagaatgt	tgggtcaaaga	gtacaaagtt	ttagttagac	atgacaaaata	51300
agttttcaat	gctattgcac	agtgtggtga	ccataattaa	caataatgtc	ttttatatct	51360
caaaattgct	gaaagaatag	gccttaaatg	ttttcagtat	agaaaagtat	atgagatgat	51420
gcctaatttta	attagcttga	tataatcatt	ccacaatgta	tacatatatg	aaaacatcac	51480
attgtatccc	ataaatatat	acaattatta	tttgtttaatt	acaagtaaaa	gtttaaaaaaa	51540
tggcaagtaa	atcatgtagc	ccctgagata	gatggatatg	tgagcctagc	ttgaaacaaa	51600
tgtcatcata	tatccaccat	tacacaagac	tttgagcaga	ctgaatgctc	acagaacata	51660
ttgggaggag	atattttggca	gctgaagtgg	caaatagtca	ttttccaagg	gaacaacaac	51720
agtagagagg	tttccagtta	aagttgcaga	gtttctgcag	agtctctagc	agtgtctggat	51780
ccaaggggat	gcggtgtatc	caagtagctc	ttgaggaaac	cacaggcaca	tcctggcatg	51840
gggagcacct	caggagcaca	tcctgagttt	cagggcattt	gaaatggatg	tgcagtcaca	51900
tcccacccaa	ctgcaaggga	taccagcct	acatgcagag	gtcaggaaaag	ctgccacat	51960
taagacattg	catgcaatag	gccccctccc	actaggagtc	tatggagaca	gaaatgcatt	52020
ttgaggagca	atttcatgca	gtcatgggtt	aagtgaacca	agtgagctat	gaagccagat	52080
tttccctcct	gggccacata	tttcagaggc	acataactca	agcttgcaac	acgtattcaa	52140
aagagaccag	ctacacttgg	tagagacagc	cataggaaaag	tgaaatgacc	ctagggttta	52200
gtaaagccag	ctgtttccac	ttctgaaaat	aataaaaatga	aataataaaa	taaattttaaa	52260
atgatacaaa	gttcaaagtt	taacaaatac	atttgaagcc	atttgcacaa	aatacatctg	52320
aagctaattg	ctggctctag	aaagtgtggg	gtctttgttg	tggagcagtg	ttaatgattt	52380
agcattactt	atctctggca	aatggtatlt	ttgagataac	atgttatgga	agaaagtga	52440
ctgaacttgg	aagtttggaag	atctcgattg	aagtatcatt	tctgcctcaa	ctacttgcac	52500
taacttgtac	aagtcattca	accgctctga	acataatgga	aaaatgggat	gagaatacat	52560
gttgtatact	ctccaaagac	agggagactg	ctgatataag	agggcacttt	tagtaactga	52620
tggagcaaaa	tgttgttata	tgagtgtcag	catagggcc	tgggcttaca	acgggtgccat	52680
gagccttaga	acagagggaag	gacagctata	gcaatgaaag	gactagtgc	gattcagaaa	52740
aataagaaga	cagaaaccaa	ggtgtagtaa	catgttttag	tatggagggg	aaggcagtta	52800
tagaaacttg	aattacataa	tttgtacatt	tctgggagat	agaaggtaaa	gatagcagct	52860
aatggagaca	ggacaggact	ggtacttgat	tatggaagaa	aggaggtaaa	tagaagagac	52920
aaaaagggag	agaagagatg	tcaactgcct	actctggtag	cctctgtatc	caaaaggttg	52980

actcaaacat	tcgctcataa	ctttgtctgg	cttaatcctg	ctcatcccag	cagacttatt	53040
tcaagtgtct	ccacgttttg	ggaagtcata	actcaattct	ctgggctttc	atatgggaga	53100
gcatttaatt	ctgttgaaaa	actatttaat	actacatcta	cctttctcta	tggactctga	53160
gcttcttgag	ggcatgtatc	atgtatgttc	tattctgaag	cacccatacc	tagaacaaga	53220
cttagcacat	agtaggaact	taataaatat	ttcggagttg	aataactagc	cttatgtaat	53280
cctcacaaca	accctaagct	ggagactcaa	acaaggctgg	aaataagtag	gtgccaaaga	53340
gaactgagat	tcagacacat	atgtgcaggt	aaaacatagg	aacactgaac	attcactgag	53400
aactgacaac	ttgtgggggt	gttgtaggat	atgtgaccag	agactcctga	atgccagtct	53460
ctgtacctgt	accatgttgg	ctaacaagaa	tcgcatggaa	tccttgctga	aaatacacag	53520
cctagaagtt	ttctcaaate	tggagagact	gtactatggg	ttgaatatgg	tttgtcgcca	53580
ccagaactca	tgttgaggct	tggctcctca	tgcagctgtg	ttgggagggt	ggacctagag	53640
ggaagtgttt	gggtccagtg	ggcagatccc	tcatgaacag	ataaatattg	tcttgtggga	53700
gtggatgact	tctgtcttgc	aggactggat	gaattaccac	agaatgagt	tgttgtgaag	53760
cttctcctca	tgttttgctg	tgtttgcacg	ctgtctcttg	acatttttct	tctttgctat	53820
gttgtaaggc	agcacataac	cctctgcaag	ctgagcagat	gccagtggca	tgctcttggg	53880
ttttctagcc	actagagttg	tgagccaaat	aacatttttt	ttctttataa	attaccagag	53940
tcaggtattc	tgttagagca	acactaaagg	gactaagaca	ctcttattct	caccaaattct	54000
ttatttttgt	aatgattttct	cacacctatt	catttgctcc	agaaaaggta	gttattctcc	54060
atagtctatc	ttcatcttcc	acttcatgct	tattcaatcc	attaccaatt	cctgtcaata	54120
tatcttctca	aatatctctt	tcaaccacca	acttttctat	cctcactatt	actatccttt	54180
tacaagcaca	tagaccactg	ccaggaaccc	ttgacttgcc	gacctgactc	tatactagt	54240
cttcttctca	ctgcagccgg	agcaatcttt	ttaatcaaag	ctatgactca	acatttactc	54300
tcttgatgaa	atctcctaga	agaccttttg	tagctcttag	aacaaagact	gaaataaaaa	54360
ctctatacta	tagtctaaaa	gttccttggt	ggctctacct	ctccagcttc	ctccctctgt	54420
gctccaacca	catgggcttg	ccttcaatgc	ttcacatata	accagcttc	aaattccctt	54480
ctggctgcag	ggccttggca	caaactgttt	tctctgcctg	atgttttcca	cccttccacc	54540
tacattcatc	agtttgactt	ctacttatct	tttgaggctc	agctcagaca	aggtagatc	54600
ctgccattga	catactcatt	agcaccctga	aatattttct	aatcacagcg	tatgattata	54660
tattttattg	tgtgattaac	tgataaatga	ctgtctgacc	cctcctccct	gctttaagac	54720
tataagttgt	attaattcag	ggcctatggt	agctttactc	agtactctgt	acccaatgcc	54780
caccccagca	tcctcacaag	taagggtggt	cagtacatgg	gtgttgaata	gatgcatgaa	54840
taatacaata	agtcaacaat	tggctctcag	aatctcaata	attttaatgc	tatcaaagt	54900
atttgatgca	gatttgggaa	aaattatcta	aaaaattcat	cccaagctaa	gatcctatga	54960
ttcttagcta	tcacagaatc	tgtgattctg	tgctactcct	ctgcgcttct	catgtacact	55020
tacatgggta	tacctatgaa	aatgttttgt	tggtttgttt	gtttgttttg	agacagagtt	55080
taactcttgt	tgtccaggct	ggagtgcatt	gggtgcaatct	cggctcactg	caaactccac	55140
ctcctggcct	caagggaactc	tcctgcccc	gcctccagag	tagctgagat	tacaggcgca	55200
tgtcactaca	ctcagctaat	ttttatat	ttagtagaga	cagagtttca	ccatgttggc	55260
cagggttggtg	tcaaaactcct	gatctcaggt	gatctactgc	ctgagcctcc	caaagtgtctg	55320
ggattacagg	cgtgagccac	cgtgcccagc	cgaaaaatgt	tttaaagcat	ctaggatcct	55380
tggcaggcct	ttagcacact	gcacagaagg	gacattctgt	gcctgtcact	ggaatgacca	55440
gcaactctgg	ttccctgctt	tgccagact	gtttctatgt	ccccttcagt	ttagttcagt	55500
tcaacaatta	tctagtgage	actttctctg	agcggggcat	ctgctttgtg	ctaagtgtaa	55560
gccctgcctc	caagaactca	ttgcataagg	agagacacac	acatgaaaac	caactaattg	55620
tgattcagtg	taaaatacgt	agtaattgac	aaatgcatat	agtttcatga	cagccctgta	55680
gaaggagtag	gcaagtgttc	tagtatggct	tcaaggagga	aatgtagctt	aaaacagggt	55740
cggaggggatg	agtaagagtt	taccatactt	ttaagggggg	tgggaggaat	atattgaaat	55800
gaaaattaca	ttttgcaaat	gcagtatgag	gagccgtggt	acagtttaat	gtgttttagag	55860
aacaacgagt	aactgtatgg	cagaagcagt	aaatatgacg	aaggaggaag	ctggtgtgtt	55920
tggagaaggc	tgagggatca	tgaggcatat	taccttcctt	taaaagccat	gccctattct	55980
ccctctcctg	ccacttcaaa	ttcagggtca	ccatttatat	gtctattagt	cctggtgtct	56040
ttctcatgct	ctttgattag	tccttaatcc	acaagcgcaa	cattgcaata	cttgcttagc	56100
atatttcaca	ggcaggggac	ctaattgtccc	tgtgagaacc	catcttgctg	agattgtcgc	56160
tggcagattt	acttccagtg	tgattgttgc	aagaatttgt	ctaacagaat	gaatgatcaa	56220
ccttgagcag	aagagattat	gaaaaactta	atagcattgt	agcaatgtgg	ctgttaatga	56280
aatacagttg	gctgctcccg	ctgtttggca	ccaaccaacc	tgacactgtc	aacatcacaa	56340
tacgatattt	attcccaatt	attttacggc	aacaactgaa	atacaatgtg	ttattaatca	56400
tattttattat	aagtatcaat	ttgagaaatt	tctgacatgc	cagaagataa	ataggtttat	56460
tatgaaaagc	agttctgctt	ggtgcatgct	ggctgctgct	gtgtaataaa	tagcctctgt	56520
ggggaaagtt	ttttaaaaga	aataaagcaa	aaaaatagca	ctgaaaacag	aaaggaagca	56580
tcaaaactct	tcaaatacct	gctgtgtcca	ttgggtcaagc	acattcagga	catcgcatgc	56640

ctttagaact	ccagcaggtt	ccaacagcta	gtaggacatt	ctagactctg	agagagagca	56700
agggaggttt	tatgactggg	gacaaagaaa	agagacactg	aaggcgaagg	acaatctctg	56760
aaaatgcagt	accctccaga	ctgctcctcc	tctcacaana	acaccttccc	agcatgcact	56820
gctttaggga	ctatgattat	accattgatt	ctgtccagaa	aacctgtgtc	ctgaatatat	56880
tacagggttc	attccttcac	ttctttcagg	tgccactca	ggtatttcct	tatcagaaca	56940
gtctttcgaa	cgaccccat	aaaaaaatag	tctgtcaac	cctatgttaa	caattttatt	57000
tatttttatt	atttgtaaac	aatacataat	aggtgcata	atcttggggg	tacatataat	57060
aatttgatac	attcatattg	tgcataaaga	ttgaatcgga	gtaattggga	tatccattgc	57120
cttaagtgtt	ttaccttttc	tttatgctgt	ggacattcaa	attactttct	aactttttga	57180
aatatacaat	agaagaatgt	taactataat	cacctatttg	atctatcaaa	tgctagatct	57240
tatttcttct	aactatata	tgtacctatt	aatctgtaat	tccacaacta	tatttacttc	57300
ttatactttt	ccccttctag	gctataaacc	aaatgagagc	tgagcatctg	tttggttcac	57360
tgcccaacac	atgcatgcct	actacatggc	agtcaaaata	tttgtggaat	aatgaatga	57420
atgaaaaaaa	aaagaaatag	atgaatgaat	catggatgaa	tgaatcaaat	cagtcagcaa	57480
tgtctttcta	aacaaaattt	ggatgatttt	ggatgattac	gcctcttaaa	aatatttctt	57540
catttcctac	cccaattttg	tttctactca	ggactttttc	aatatcttcc	aaacctattg	57600
ttctttttta	tttgtttgct	ttttgaggca	aggtcttgct	ctgttgccag	ggctagagtg	57660
cagtgggtgtg	atcacagctc	actaaagcct	ccaactcttg	ggttcaagtg	attctccacc	57720
tcagcctccc	aaatagctgg	gatgaaagtg	tacaccacaa	tgcccggaga	attatttcat	57780
ttcttctttg	tagagattga	gtcttactct	gttgaccagg	ttgctttcga	actcctggcc	57840
tcaagccatc	cttccacctc	agccttccca	agtgttagga	ttacaggcgc	gagccaactt	57900
gccagccct	ggaatttttg	agcctgttca	attctaacta	ttgtcaccaa	aagtaacctt	57960
aagaaaaaaa	atgcattatc	tccttgcttc	attgcaccat	taaaatcttt	cctaaatttt	58020
ccatgttaaa	gatgaagctc	aaaatcctca	gcatagcata	caaaacactt	cataatcaga	58080
tgccctttca	aatacctcct	atcagaatgg	tctctttgac	tacccttcta	aaaaaattcc	58140
ccccaaccct	attttttaaa	tattttactta	tttttattat	atttttgata	cataatagat	58200
ggacacattc	aaacagtggc	cccaaaaactg	gggcagcaga	aacagggtcct	tgcttatttt	58260
ctcagcttca	cctcctgcct	ccaccccatc	tgtactgctg	gtccagacat	tcctacagag	58320
gtgtcctcct	aagttggtct	cttccctctcc	tgtctcagag	gctttgccct	gctcttctct	58380
gcctcttgag	gctctgtcct	gctcttctct	gcgtcttggt	gttggaatgc	ctgtcttttt	58440
cctactgaag	atctggatgc	ctaaaccata	atgtaaaatt	gctgcttttt	acttccattt	58500
acagcagaga	aattcctcct	ctggcctctc	ctcttctctg	tgtttctttc	ttcataattt	58560
ttattttatt	atatatttat	ttattttatt	attttattat	ttattttcat	tgagatggag	58620
tctcgtttctg	tcgcccaggc	tagagtgcag	tgggtgtgatc	tcagctcact	gcaatctcca	58680
ccttccaggt	tcaagcgaac	ctcttgcttc	agcctccctc	ctgtagctgg	gactacaggt	58740
gcccgccacc	accctgggct	agttttcata	tttttagtag	agacagggtt	tcaccatggt	58800
ggccaggcta	ttctcaaact	cctgatctca	agtgatctgc	ccaccttggc	ctcccaaagt	58860
gctgggatta	caggcgtgag	tcaccgcacc	cagcctctct	ctttataatt	ttcctactgt	58920
tcacctgcat	caaactcctg	aattctgtca	tgcaactgga	acagtaagag	ggaaaaacat	58980
ggagctcaaa	gaaagatggt	gagaaacgta	gagttgcata	gaatttactg	tataagaatg	59040
gaatctgtca	agtcagacaa	gcgacagaga	cctattttaca	aagagaccca	gtgaaaatta	59100
ctggagaaat	aataaagaga	aatgctgtga	ctttgaaata	aataatgttc	aaaagtcacc	59160
tgcaatattt	aggatagtgt	ctgaaacaga	tacaaatatt	tctcagcagt	aaaagaattt	59220
tgtattttagt	ctagtcatgg	aatagtagtc	agttgtcact	gaggaagcac	tttggggtag	59280
aagaagcatt	tgaatgtggt	tgaagtctga	ggcagcaggt	gaggtttgat	tttatatttt	59340
tgaaaatgga	tctatcagat	ggtggagcta	ccctcataaa	agatttgtaa	tacgcctggt	59400
tacctacaag	attaaatcaa	gtgtcatttc	ttcagggaag	attgcccttc	accatgtgaa	59460
atatacataa	agtatatgtc	acaatgtgtg	acagttcttt	gctcacatat	atatttctcc	59520
acttaaagga	gaattttttt	tgagatatga	tcttgttctg	ttatccagga	tggggtgcaa	59580
tggggcaatc	acggctcatt	gcagccttca	cctcctgggt	tcaagtgggt	ctcccacatc	59640
agcctcctga	gtagctgaga	ctacaagtgt	gcactatcaa	gcctggctaa	tgtttaattt	59700
ttagtagaga	caaggtcttg	ctacatttcc	caggctgggt	ttgaattcct	ggcctcaagt	59760
gatcctccca	ccttagccca	gaaggagtct	tatttcattc	atcgatatatt	cttagtatct	59820
acctgtcatc	aggcctatag	tagatactca	gtagatgtag	attgaggttt	gaagaataag	59880
agatagctca	ccaagtagaa	cactggatgg	tattggaact	aatgtattcc	tttattgtca	59940
gcagaatgga	ccatgcacat	agaaataata	aaatgggaga	attgattgcc	atgggtctaaa	60000
ttttgtgccc	cctacaattc	atatgttgag	accctgaccc	ccaagggtgc	ggtactaaga	60060
agtggggcct	ttgggcagtg	ataagggtgt	gaggggtggag	ccctcacaaa	tgggatttat	60120
gccctcataa	aagaaactgc	agagaactag	ttaaaccctt	ctactatctg	aggacatggc	60180
aagaagctgc	tgttctctga	acctggatag	aggacctcac	tagacgctga	ccagcgcttt	60240
catcttggat	ttcccaggct	ccaaaactgt	gagaaataag	tttctgttgt	ttataagcta	60300

ccaggttttaa	ggtatttttgt	acagcaaccc	aaagagtctg	agaccataat	gaagccattg	60360
gaatggtggg	aaggcaactt	catgtgagta	actacagtaa	agccagggtg	tggtaacagt	60420
catgttgccc	atagagcaga	tcctactatt	acagtgccta	gcacattacc	tgcatatgat	60480
gatatgtgat	caattagtta	actgattagt	ttatgaatca	gtctgccaaa	aactagggca	60540
gaaattgata	gcacattaaa	ataaatatgc	cttaaagttt	gcaaggagac	cctattaact	60600
gcgactgtt	ttctttttat	tttctttttt	tttcttttga	gacaggggtc	cactctgtca	60660
cccaggctgg	agtgcaatgg	cacagtcttg	gctcattgca	acctccactt	ccggggttca	60720
agcgattctt	gtgcctcagc	ctcccaagta	gttgggaata	cagggtgtgca	ccaccacacc	60780
tggctaattt	ttatatTTTT	agtagatagg	gggtttccca	tattagccag	cctgggtcctg	60840
aactcctggc	ctcaagtgat	ctacctgcct	tggcctctca	aagtgtctgg	attacaggcg	60900
tgagccactg	cacctggcta	gccactcact	gttttcatgt	taggctaagt	aagctttttt	60960
gaagaccatt	aacataaata	tacaacctaa	atgtatttta	cctgaataat	tttactcatg	61020
tccacagctt	gttcttttcat	aggctgccat	gatgaggaag	aacagagatt	agtagtagca	61080
ctattcattt	ctgatatttt	tgtagtagtg	gttctaattc	tactccaggt	ttagaaaaga	61140
tctgtaggaa	atcacaggtc	tacatttcgt	cctctaaact	actctgttgg	gtagaattta	61200
ttttgcaaa	acttatgtag	gatcactttt	ttactacagg	ttttgtcata	tggtattttt	61260
acaacctttt	tttctgctga	aacaaatggc	ttttaatcct	taaaagggca	gggctatatt	61320
ttccttcaaa	cattttttaaa	ataacttaag	agaatttaatt	tttagtaata	gcaagtgaag	61380
aacattttta	tcctagagct	taagaaaggg	gaggcccaat	aaccagatgc	tggaatctta	61440
ttgaggtttt	ttttaaatct	cagtatccag	acaattggca	tgaaaataaa	ggagcctaga	61500
aaaaatgttg	aaaatgaaaa	caataaaagt	gacaactaaa	catatttttt	tgtatttgca	61560
aagcacttcc	acaaatggcc	acaaatgtca	tctcttttaa	ttttattcct	gagttgagta	61620
ggaaataaac	catttgagga	ttcaagggtt	agcttaagat	atgggtgttt	cttgaacctg	61680
ggaggtggag	gtttcagtga	gctgagatca	tgccactgca	ctccagccta	gatgacagag	61740
cgagagtttg	tctcaaaaaa	aaaaaaaaaa	aaaaaaaaag	atatggatgt	tgtcaataca	61800
atcgggggaa	aggaatactt	tgaactactt	tgttgggaag	agtttgaaat	cgttgaggac	61860
tcagcagcat	gaagtagaga	aattcacaa	tggtagaaag	gactattgtc	cttcaacctt	61920
cattaaggtt	aactattcaa	ccttcattaa	aaacagaaag	tgacaatttc	acagcaaatt	61980
ctagaacttt	agatcaaaa	tcaactcaat	atgggggatt	tatataagaa	agagttaaaa	62040
aaaagacgaa	atgtaatatc	tatgttattg	caagtgaag	gaaaacagga	agataaatat	62100
cacaagaaga	caaaaatgta	tctaactttt	tgggacaaga	ttgtgggatc	cacagaaaat	62160
tggaaacttg	aacttcctgt	tccacagaga	taagaaatac	acttgctttt	atctcacttc	62220
tcaaaaaaag	taagatgaat	gggggttttag	gccccagaga	gaaattgtag	ctgcaatcaa	62280
ttgtactatt	tgagtaaaaa	ttgtcctcag	aggaaagtga	gtagggagct	gtctgaagg	62340
acaggttatt	aacaaaagag	agggataatg	gattgcgttt	gcaagtgcag	ttggggcttaa	62400
catcaatgcc	atcttcatag	ctgggttcaa	aaaatattct	ggattctttt	agtgctcttg	62460
ttcttacctg	ttgtggttgc	agaaggtata	aatgtaccct	taaaagagat	taggggagaga	62520
agtgcctccc	acagcaccac	gaccagaaag	ggaagaggaa	ggacaggcaa	tagccaagga	62580
ctcctggcag	tgaactcatg	tccacatcaa	gatctaata	gcttgcaact	aactcatttc	62640
tagctctgcc	ttggaagctg	gagctcctgc	actgactatc	aatgtgagcc	cctgagtagg	62700
agcagcttgg	tagagttgaa	agaccattga	tctgggtcaa	cagactctgg	ttcctgtctc	62760
agcagtgtcg	taatcaatcc	aagtcaata	tcatctctgg	gacttaattt	gctaaattta	62820
aatgaaaag	aaaaacaaaa	atagaacaat	tagactagat	caggattcgg	caaaccaaag	62880
cctgcttatt	aaatctggca	taccacctgt	ttctgtctat	acaattgtat	tgaaactcag	62940
ccacactcat	tcattttgcat	attgtccatg	aaagaagctt	ttgcgctgcc	tcaggagatc	63000
tgagtagtgg	ccacagagat	gttgtagtgg	accatgttgc	aacattgtcc	aaaataccta	63060
aatattttac	ttcttggttt	ggagagtttg	ctgactggca	ccagagaaat	ctatggtcta	63120
aatcatctta	aaaattttaag	cacatatgtg	tgaccacaca	tttcgaaatg	ccgtcttccc	63180
aatctagaac	acagcacatg	aatcagtagg	taaacctatc	catgtcaatt	ctcaaatttg	63240
aaattcattc	acttgaaaat	ccagcaattt	ttcatgcttc	atatcatatc	tgggtttgga	63300
ataaagaagt	gtgaggggag	aaaaattccc	tgagcatttt	aaatctaatt	tcacctttat	63360
tatgagacta	ctgagtcttt	tcttgagcaa	aggagagagt	gtgaaataga	ataaggtgct	63420
caaaacaata	attgaaattt	attgaaagga	tgagtaattg	gagtaattgt	acagaatatt	63480
aagcagatta	tttagatagc	atatacat	ccagtttgat	taagtcaatt	cacaggccat	63540
caaaaagtac	acagaaaaat	agaggggttg	atccgatagc	cttctgcgtt	agaatgtagt	63600
catttgctct	cttttagctat	ttaatctctt	ctgcatgttc	agagggagaa	aatgagtgat	63660
gagagagaga	agtaataaga	tcatactgca	aatccttaca	tcgataagat	aacagaaggc	63720
ttttcacact	ggctaataat	tcatgttgta	tttacagctc	ttgtccatgt	acagatttgg	63780
ggtcattaac	agaggtgttg	atcagcaaag	tcttagagtg	agtctaggga	aatgggtgcc	63840
aatatatctt	gggtattgga	accatcaaca	gagcctgtgc	cccagactta	ctgaatcatg	63900
gtctctggtg	gtggatccca	ggtagtcaca	gctttaactg	agtgattagg	agattcactg	63960

tgcagccagt	tcggaacca	ccagtgaat	ccaaccctt	catttattag	tgttgaacc	64020
gtaactttct	gaagatgtac	cactgtccca	gccacttcac	ctgtattccc	agctcatttc	64080
tctcagagaa	aactgccaa	tccccacagt	ggccttgcat	gggccatttg	tgatgtcacc	64140
tctctgacct	catctcctac	tttcccttcac	ccggttggtg	ctagctattg	ccatgctctt	64200
tctccaacac	accaagcaca	ctcctgcac	agaggctgtg	aattttccat	ttcctgcttt	64260
gttccctccc	aatttttgca	tgaaccactt	tcccacttaa	tctaattttc	tgttcagaag	64320
tcactttcac	agaaaagtct	tcctcagcca	catgtatttt	cttcttggtg	cttattatca	64380
cctgacattt	tatatatcta	tgtttgtttc	ttgtttgttt	cctactccat	tcggtaagct	64440
ctaattgaga	acagaaaagc	tctgtttatt	cactgttgta	tccccaacat	ctatatctag	64500
cacatcattg	attctcaaca	actattttgta	aatgaatgaa	taaatggctc	tgctagattt	64560
ttgtcacctg	gattgccaga	ccagtgcaat	aatggaaagc	taagtaatgt	aaagagcttt	64620
gcaatcagac	agctctggct	tgcattcttc	atgtgtgatt	ttgagcaagt	tactgattct	64680
tgctgagcat	cagttttcta	tgtgtaaagt	ggaggcatgg	acactggagg	atcatggtga	64740
gaatttcata	gagcaggcca	ataagacctc	cttggggaaa	gaaggctaaa	gtgggcctcg	64800
ggtttaggtt	tccccactg	ggagttccaa	ggggctatgg	agttgatagg	attccccatt	64860
tgggctgcag	aggagccaga	aacatgcttt	agtttcttaa	tcctgaagga	gtgggggatt	64920
gtgtaagcct	cattggtgcc	cttttgaacc	acatttcac	aagatattgt	gaaatgggga	64980
gcagcaggca	gccttctctg	gccaacatgt	atcagggcct	cgtgcttagt	gggatgtgcc	65040
cagagtagac	agggaatccc	tggcagcctt	gcatagtccc	tctcatgcct	ttgctttcat	65100
ggggctggct	ggaaacaggc	cggtaagccc	catggaacta	ccgggacaaa	ttaccaagca	65160
atctcccatg	gccatgtaga	acagccatgt	agaacaggat	gtctcccaag	gtaagattta	65220
aggtttgtgg	tttgacagaaa	agagaaggga	ggaaataaaa	aagcactgac	aatcattgaa	65280
agcccacgtt	ctaggttaact	ggctactcac	tttcacagcc	cagtaacaat	tttgcttaat	65340
cctatacatc	tataaatagg	attattgtct	tcattttcca	gataaggaaa	caagacttgg	65400
caaagtgtga	gtgtagtgtt	tgacagtatt	atgtagtgtt	ttagcatgca	catgaacttg	65460
caagaccttg	ctcttcccta	ataattctgt	cctttactaa	ttctgtgatc	ttacaagctg	65520
tgccttatgt	attttgtgtt	ctatttttga	aaacctaatg	acaacaacat	aataatacct	65580
tctcatagaa	caattattgt	gataaattaa	gctgttgaga	gtcaggcact	caacccaaca	65640
acatgcccac	aatgtgtact	caataaatgc	ttgctagtat	ttttaataat	attgttaata	65700
aaataaccac	gccagaggta	gaacccaaga	ttccgactct	agttgacctg	attctaagcc	65760
aactcttcta	cagcaacaca	tcggcccat	gatttagaca	gcatggaact	catagcccag	65820
ggaattcaca	ctcttaaaga	acttaaacag	cacagcttaa	tatggctcct	tccaaacata	65880
cgactcagta	attccagtag	gagggctaaa	caacctattc	tgagaggctc	ccaggaaagg	65940
aagcctggag	tcttccctgg	aaacatgttc	ccttatcgca	cctcccttct	cttcaggatg	66000
ctgaggacct	ctggctccat	ggacttcac	ctcccataca	attggttccc	ctcacaatta	66060
gtttctggca	ctgtcatcaa	agtgatccct	ctccaaggac	atggaacagg	atgaaaaaca	66120
aattgtaatt	gtaaaactga	agtagcatca	gggcaggcac	agcaaaggag	gaagggttac	66180
ttgtgggaat	caggggtcaa	caataaaagc	cagcctgctg	cctgtcccca	ccaggcaatt	66240
ttctgggctg	ttcatgtccc	cagttatgcc	cccaaacatt	cccatgacag	caatgcctaa	66300
ggaaaataac	ctctgaagat	tgagcctaaa	actagaatca	tttttctctc	tcatttctaat	66360
tcatagcagg	gaagtggact	ggctccttaa	atccagaaag	tatcagactt	atctgatcaa	66420
atcattaggg	caactgtgtt	ttccctcac	tcagacagc	atggatgagg	agaaggagcc	66480
aaaagacaca	gcatctctct	tcaaataaat	actcaacttc	aaagaagaaa	gtgaccactg	66540
ttaaggaggt	gaagtcatag	cctcggcaaa	cctttgaagc	tgggggagaa	gatgtaaatt	66600
aagcaccgac	aagcagggtg	taaaaataaa	atccttgaag	gaaaagcaaa	actgctttgc	66660
aatgggtgtg	aaagtaaagg	aagctggcca	aggccagagg	aggtaaacag	ggttgcatga	66720
tgtaaccaca	ctaggaaaag	gcaaactagt	tgagaaatat	agagaaaata	catggaagtt	66780
gtaaattgtc	tattttatct	gagaacttat	agaatcaaaa	gtggaaacaa	ggtcaatttt	66840
cttaatatct	aggccattat	aaacccttgt	ggggaacagg	aaacttgctt	ttatttgctg	66900
tttctgcaat	agtgttggtt	gggtacacga	caggaccatg	gaactctata	atcttgatga	66960
gggatacatg	ccaagtgcc	ttcctgattc	ctttgtagtt	tttttttttt	cttatgtctc	67020
ctattagttg	aataattcat	tttttgttta	tttattcatt	cattcaataa	actgcttttg	67080
tgcttacagt	ataccagaca	ctctactaga	tgctaaataa	gcaaagatga	acatggcctt	67140
gtccttacct	ttaaagaacc	ataactctat	ataaaaacac	atttaatttg	taaattagag	67200
gacaggatag	ttagtttgtg	atagagaaaa	ccaaatggct	ccaaagggaag	acaaagtata	67260
aagttttgaa	aactatcacc	ccaaaattca	catctacctt	gaagtgcaca	atggcaactt	67320
gtttggaaat	acggctcttg	cagatgtaat	tagttaaggt	catagcagat	taggatgtgt	67380
tctaaatgca	atgactgata	ttcttaaaaa	ggagagagca	tgactcaga	cacacacagg	67440
gaaaaaatcc	atgtgatgat	ggaagcagag	actgaagcca	agaaaagaca	aggattgaca	67500
ggagccacca	gaaactaaaa	agaggcaagg	aagcattttc	cctagaaagg	agcatggctc	67560
cactggcact	gtaatttcag	gttgctagcc	tctagaacta	taaaacaata	aattttattgg	67620

ttttcttaaa	ctaccaaat	ggtgttactt	tgtcatagca	gccctaggaa	gctaacatac	67680
aggtgaagga	catttaaagc	tatagtagta	agagaagaga	tgtcctaaga	tgtttctttg	67740
gtgacatcat	gtgattttta	atatggattt	tgtctagggg	aattatctct	gtcactgcca	67800
atcaagggtac	cttgctccac	tcatcaaggg	gtggcccata	tccaagctgg	agctttttga	67860
ttctctcttt	ctgcagtttg	gatattttatg	ctgatcatag	ctcccagaaa	aaaaaaaaata	67920
aattggccgt	tttgcttccct	tgattccagg	agctgtccca	gctcttgtgc	atcctacaac	67980
tcagttcttt	ggcttttccct	tctttgtttg	cttaagtga	ccaaagatgg	catttattgc	68040
ttgcagcaaa	gaagtctaac	tgatatagga	cccatgatta	atgttctggg	tcttctttta	68100
cagagaacat	gatagtgttc	tctaatttca	tctttttttc	tatttttttt	gttctgttgc	68160
aagaacaaat	ctgctatatt	tatggcaact	tgtggcaata	atggaagatt	ctaagataca	68220
tttttataaa	catgttttag	ccctggataa	aataacacct	taattcatca	tggctctctc	68280
atctggaatc	ccctttcttt	taaccaat	cacataccaa	aaattatttt	ttgaccacag	68340
aaatgcctta	agggataaat	aaataatata	gtcatcagag	tttaactaaa	ctgggcttta	68400
actgaggtgc	ctgcaatcaa	ttgtatatgg	tctaaatttg	gacttagcct	tcctgagcca	68460
tgggcttccc	aggtcatatc	caccaatggg	tcatctctga	aatcagaaaa	gtatttatg	68520
gcaggtactc	agcccagtag	cgcacatata	gagatgctaa	aaatattgtt	tttctctctt	68580
tctttcttct	tccaatcctg	acccctccaa	ggtcatcctg	actcttctca	aagcctgtcc	68640
ttttttctca	agtggacctg	tagctacaga	aagttgagaa	tgataaattgg	ataaatgagg	68700
acagatgtct	cttttagcccc	cttctctctt	agcttgctgg	ggtctcctga	ctggattact	68760
ctcagctcaa	ggaacccctc	cctcaacctc	aaaaagagaa	gacaagattc	atttttatac	68820
aatgttggtt	gtcaattcaa	tccaatctca	aagtacgttt	tgcttctga	aatctaaatg	68880
ttcacgtttc	tttttccaag	gtatctatat	tcatttaatg	gggaaaaaac	aatacctgga	68940
gtcattatat	gtgcatctat	gtctttctaa	atatgcacaa	gcattttact	tgaatatttc	69000
caggtcctgc	aagaaattct	gaaattgatt	ttcatttggtg	tggcttcagg	attcatttca	69060
atggcagccc	aaacattctt	ccagcagaat	ggccattcag	acctgcttaa	tgttgtgaat	69120
ccattgatat	ttttaccctt	ccaaaagcat	acactattga	ttctgcctct	ctcttaattg	69180
cacagaagaa	aaactgactt	ctcatgaaca	caacttgata	tctcatttga	tatcttcatt	69240
ttcataccag	tgtaactgag	ttatcaatgt	ttgcttgagc	cagactatta	aaacagccac	69300
ttggatagtt	tgggttttta	aaaagtctgg	caaggatatt	gcctcattgt	tattatttgg	69360
aaatttctag	aaatatgtca	tctgtaatat	cagtgtactg	gcccaccaac	tgggttagaa	69420
gtttctgttt	gatgttcttc	ccccaggact	cccatcctat	tctatcttac	ctctagggtg	69480
gcaatgatca	cattgtccca	gaaaaatgac	catataactt	tatttttcaa	actggaactt	69540
tcaagagtga	aaggaggatg	ggttattaat	taagctagga	aaagggaata	aactcagaat	69600
gtcctaagta	aaaataggtc	actctgattt	tcaccattag	ttcagtgtctg	tgtctcttcc	69660
tctagacatt	aagcccctgg	ccatatttag	cttgctctct	gctctctctc	tagcatctga	69720
catagtacac	agggaaggat	ctgcacttca	gtcatgggtc	tgtgtcatgg	caagaacaga	69780
taaaacaatc	tccctcaaac	ttacaactct	tttggtgtaa	tcactagaat	atatttatatg	69840
ttggtatcac	tcaaagtagc	aattggccaa	tcatctcttc	ttttctggct	aattgcttat	69900
ccttcaagac	acctgatgtc	aactgcattg	aggtacattc	ttatatattgt	gtagcattct	69960
gtcttgatat	ctttgtgctt	gatcatgttc	tattataatt	gtttctttaa	tagacagctg	70020
agaactgctg	tgaactccat	ttcttctcta	taccagtatt	tagctcggag	cctgtcacag	70080
aataaggaaa	tagtacataa	caactgaatg	aatgtatgtg	cagattcgtg	aatgaaaaaa	70140
tgaaatatcc	agattggaca	ttagaatatt	ctatctgtca	actgctgttt	ggtttgctg	70200
atactgatta	tttgccctact	tggtaaatatt	tatatagcta	gaatcatgga	cccacagact	70260
actataaact	aagagacaat	tccagacatt	agtaagacag	aaaagttttc	tagtatatca	70320
taaaatatta	aatctgatag	tagcactacc	aatggcatga	tatatttaaa	ggtgaacata	70380
gagctaattgt	tcatataaaa	atcatcctgt	gataatttgtt	gctgcttaac	atttttaatt	70440
caatttcatg	ggagaacaaa	agagtgatta	agagctagaa	aatagaattc	atgagaaaag	70500
gttagaggaa	ttgaaattat	ttagcccaga	ggaaaaaagt	atatgtagcc	attcagtgat	70560
tcagtgaaac	ctcatttatcc	aatatcagtt	tattcaaacg	atggcataag	ttaaactgac	70620
agctatgttt	ccaatgaaca	aatgacctac	catttcttaa	aaaataaaaa	aaaaagataa	70680
tcagaaacca	tttaacttaa	atatcttttt	tttttctgtc	tgatctcgtc	tatttgggat	70740
tataaaggat	ttattttaca	tgaatcttcc	tgggtcacac	tttgcattgg	agtagaggta	70800
gtagatcagg	tcaaagtggg	ttgtgcccc	attttttttt	ttctcttaag	tttctatttt	70860
ctaaaatgct	gttgtgtttt	cattcattca	acaagtattt	attgaggaac	tatgcatgtc	70920
agcgtccatg	ctaagcacga	gggacataga	agctatgaca	cacaagctat	ttgcattatt	70980
ataagatgtt	tcagtttaagt	tcctccctta	ccgaacttaa	gtgtgttggg	aaacattaaa	71040
aagtcatacc	cctgacacag	tgagatggta	tctcactgtg	attttgattt	tcatttctct	71100
aatgaccaga	gatgatgagc	tttttttcat	atgtttgttg	gctgcataaa	tatcttcttt	71160
tgagaagtgt	ctgttcatat	catttgccca	ctttttgatg	gggttggttg	tttttttctt	71220
tgtaaatttg	tttaagtctt	ttgtagattc	tggatagtag	ccctttgtca	gatggataga	71280

ttgcaagaat	tttctcccat	tatgtagggt	gcctgtatac	tctgatgata	gtttcttttg	71340
ctgtgcagaa	gctctttagt	ttaattaggt	cattaaaaacg	tcaggaaaaa	accgatgttg	71400
gaaaggatgt	ggagaaatag	gaatgctttt	acactgttgg	taggagtgtg	aattagttca	71460
gccattgttg	aagacagtgt	ggcaattcct	caaggatcta	gaaccagaaa	taccatttga	71520
cccagcaatt	ccattactgg	gtatataccg	aaaggattat	aaatcattct	actataaaga	71580
cacatgcaca	ggtatgtttc	ttgcagcact	gttcacaata	gcaaagactt	tgaaccagcc	71640
caaatgccca	tcagtgatag	atgggataaa	gaaaatgtgg	caaatataca	ccatggaata	71700
ctatgcagcc	ataaaaaagg	atatgttcat	gtcattcaca	gggacatgga	tgaagctgga	71760
aaccatcatc	ttcagcaaac	ttacacagga	acagaaaaaca	aaacaccaca	tggtctcact	71820
cataagtggg	agctgaagaa	tgagaacaca	tggatacagg	gaggagaaca	tcacacactg	71880
gggcctgtcg	gaggggtggg	ggttagggaa	gggatagcat	taggagaaat	acctgatgta	71940
gatgacaggt	tgatgggtgc	agcaaacgac	catggcatat	gtataactat	gtaacctgca	72000
cgttctgctc	atgtatccca	gaacttaaa	tataataata	ataatagtaa	agatatcacc	72060
ccccaaaaaa	gtcatacccg	taggttaagg	catctctaaa	gtacacgtat	tcttatgttg	72120
ggggcaagga	agtagcaaat	taagctatta	tctgtattgt	ttatggaatt	atttttcaga	72180
tgggttccca	ttgaagagaa	aatgggatat	agatcaaatg	agctttgaag	gataagaaaa	72240
aagaaatgag	ctttgaagga	tgaaaaaagg	aatttttttt	aaaaagacaa	tcagtgtggg	72300
aaaactggat	atccatgtac	agaagaataa	taatagactc	ttatctaacc	acttttacaa	72360
aaatcgactc	aaaatggatt	aacagcttaa	atgtaacacc	tgaatctata	aagctactag	72420
aagaaaaatgt	agggggaaag	cttcattata	ttggttttag	gtatgacttc	aaaagcacag	72480
gcagcaaaaag	caaaaaataa	gaaatggaat	tgcatcaaac	taaaaagctt	ttgcatagca	72540
caagaaacaa	ttacagagtg	aagagatcac	ccacagatta	ggagaaaata	tttgcaaata	72600
atattgcaag	tcataactgt	taaagagcta	atatcccaat	tagacaggga	actcaaaact	72660
attcaattac	aaggaaacag	agaacccaat	tttaaaaaacg	ggcaaagaac	cggaatagac	72720
atctctcaaa	agatgaaata	caaatggcca	acagctatat	gaaaaaaatc	cttaaaattt	72780
ttaatcatag	aaatgcaaac	taaaaccata	ctaagatatc	aactattata	ttggctatta	72840
tcaaactgat	gaaagataag	tggtggcgag	gatgtggaga	aaagtgaatc	catacacact	72900
gttggtggca	ttgtaaatta	gtacaggcat	tttgaaaaat	agtatggagg	ttcctcaaaa	72960
aactaaaaaac	agaattatcc	atatgatcca	gcaatcccac	aactggatat	gtacccaagt	73020
aaatagaagt	cagtatacca	aaggagatac	ctgcactccc	atgttcattg	cagcattatt	73080
cacaatagcc	aagatatgaa	aaaaatctgt	gtccaaaaat	gaatgaatgg	attttttaaaa	73140
tgcagtacat	ctacacaatg	gaatactatt	ctgcctgaaa	aaataaaaca	ggaaattcta	73200
tcattttaca	cagtcacact	tgaaggatat	tacgttaagt	gaaataagcc	aggcacactt	73260
acatgtggaa	gctaaaaaaa	cgggactcat	agaagtagag	agtagaactg	tggttatcag	73320
aggcaggtgg	gggtagggag	ggtagtgtat	ggggagatga	tggtcagcag	gagatgttgg	73380
tcaacaggtc	caaagttaca	atgttaggtt	tgtgcaacag	taattgcatt	tttgccatta	73440
aaagtaatga	caaaacgcac	ttatccactt	gcttcaaata	ataattgcta	ggaggaataa	73500
gttctagagt	tctattgcaa	tgcgagatga	ctataattaa	taataatgta	tattttcaaaa	73560
tagctaagag	atatgacttt	cgatgttctc	accctaaaga	aatgattaat	ctttgagatg	73620
gtgaatatgc	taattacctt	gatttatctt	gctatcttgc	agtgcctaca	tgcgtcaaaa	73680
atcacattgt	actttataaa	tatatagttg	tcaactaaaa	attaattttt	aaaaatatta	73740
ttaaaaataa	aggttattaa	aaataaatat	tttattatat	aaaatgttat	tttaagaatt	73800
tagtattaaa	ttcttaaaat	ttaaatttct	aatttaggat	taaaaaagtt	agagctataa	73860
agttcataat	attaagctga	gaaatataaa	attcacagaa	gtcagttgac	aagttaaacac	73920
aataagttag	taattactgt	gactatactt	ccggctctcag	atatcttact	atctagtact	73980
cttttcttct	tccaaattag	tatttgcatt	tatttactaa	tagtaactat	ttgtatatgg	74040
ttgtgaataa	cagaacccaa	ctatgatggc	ttaagcacat	aatatttatc	cacctacatc	74100
aaatccacag	ctaagcagtc	catagctgct	ataacagtgc	cacaaagtca	tcagagacct	74160
agactcctgt	cattatgctt	tacccttttt	gttatgtagc	tgtaatcttt	tttttttttt	74220
tttttttgag	acgggatttc	gctctgtcgc	ccaggctgga	gtgcagtggc	gtgatctcgg	74280
ctcattgcaa	gctccgcctc	ccgggttcac	gctattctcc	tgccctcagcc	tcctgagtag	74340
ctgggactac	aggcgctgc	caacacgccc	ggctaatttt	ttgtattttt	agtagagacg	74400
gggtttccac	gcattagcca	gaatggtctc	gatctcctga	cctcgtgac	tgcccgctc	74460
ggcctcacia	agtgcctggg	ttacaggcgt	gagccaccgc	gcctggccta	tgtagctata	74520
ctgtttaagg	ctacaagaag	gctgcttgac	atataatttag	ttttgttgct	aagtgaaaaa	74580
aaaaatattg	ggtaggcaac	ttgttatctt	taccacatcg	ttgttaagag	acaattttct	74640
atgggtctcc	agaatttctg	cacatcttgt	agacagaggc	actgcctctc	tttgttccga	74700
agtgattttc	caaggttgtg	tataagcctt	ggaagctaga	gatcatgtgt	cccttcaaag	74760
cagagagaag	tttctttact	gtccaatata	ataaaggtaa	tgtctgtctc	tctggggcga	74820
agattaggca	tgagcactgc	cagttataaa	agattcaaat	ttgctaaact	cagtatttct	74880
ttcctgtaac	acgatgcac	atgtgtgcag	gcacacatg	gccctattca	tgtcaccccta	74940

tgggaattga	ggcccagga	ctggcttgag	aaaatgaagt	tactggttac	tggtattggt	75000
gtaagtaacc	agatgtcttt	tgtctctgac	tctgaaatat	catgccatgt	gccgatacct	75060
atgacactgc	agcaggctac	cttggttagct	tgcaagtaag	tttcagaccc	ttcacagttc	75120
ttgacagtca	tccataactc	gctttgtatg	ctataatatc	taaggtgtct	aactccaagg	75180
tgtctagaac	agagtaggtg	actaatacat	acttaaataa	atagagagtc	ccaaaaggca	75240
ttttaatgag	tcagaccagt	tgacataaag	tcccagatga	taccactgag	gagccttatt	75300
accttccaca	aaatcactta	ccctctctaa	cctcagcttc	ttaaattagt	agtctgcatt	75360
ttctccttcc	ccttgctcaa	ctggttgttt	tttcctaact	tactgcta	tggttatggg	75420
cctcatccct	tcagagaaac	tggtcttcca	gggtctctag	tgacccacta	actggcacag	75480
cctgctgacc	ctttccactc	tgtatctcac	tccctgctgt	tctgttaaca	tttaactcct	75540
taaaactctt	tcttttcttg	ccttccatga	aacttttctc	tccctggttt	tcttatatct	75600
ccaggaaact	atztatccgt	ctcttcttcc	attattttact	cttaaacatt	gatatgccat	75660
aaatgcccac	cctccctcca	gccctcttct	cactctgtag	aataaagttt	ataccttttt	75720
tttaaaagaa	aaaattgggt	ttgcttacca	tggttggtgt	tataactttt	gaatttctat	75780
ctccagtcga	gatttttctc	ctaagcttca	aatctatatt	tttaaccccc	aattcgttac	75840
ctctacctgg	atgtgttatt	ggaatttcaa	atgaaaacat	gtttgaagtt	acacttatta	75900
tttattttcg	cacctcccaa	actggttagag	ctggcatccc	tttctcagtg	aatgggtatca	75960
ccattcatga	tttttttttt	ttattttata	aacctatcga	ctctaaggcc	aataacattc	76020
ccatccagta	ctatggctgt	ctcatcaata	cttctcaaaa	cccttgcttt	tcttcccact	76080
gcctcccat	taatttaggc	cctcactatc	tattacgctg	gccccatgtt	caaaccgtgt	76140
ctccagaatc	tgtcttactg	gatcataaaa	gccccagaa	taaactttca	gaaccaaact	76200
cacataatgc	tgctgccctg	attaaaaatc	ttctttggct	tttctcggtc	agaagaatgg	76260
gagaaagttc	aaactcttta	gtatgataca	cacaaccctc	cttaatctga	cagttcctgc	76320
ctcttcagct	tcacttccta	tttctctcca	tctcacagcc	tgtgccccag	ccataataaa	76380
ccactcacat	tttccagcac	gtgccctgtt	aatatagacc	ttgcaccgct	aaatagactg	76440
attcttctac	ctgctatttg	acatcacccc	caaatccagt	ctagggaatc	tttcatcctt	76500
caaaagctat	agcttaactt	tttctagccc	tggaagttt	acccttaact	atcaaacccc	76560
tcactcttaa	gactggtgga	ttactgccta	cctggtgctt	ctctctacct	tgaaggattt	76620
ttttttttat	cacattttatc	attgctgtac	catactaact	cgtaacata	cctgaaacca	76680
ctaataaaaag	ttgtcttatt	tatcgttata	tgtccaacac	agagcaciaa	tccagacgca	76740
gagtggaggc	tcggtaaaat	tatttaaaga	gtaagaaagc	tggagaattt	gagtaatacc	76800
agtaaaaagg	cagttactga	agtggaatga	aaggaaagaa	gcacagaatt	tacctttttc	76860
atagtttgcc	ttaagtaaac	tctggaatgt	tcttaggcaa	tgatcaccc	cctgggcccc	76920
aaaatgagaa	atcaagttca	ccaggaagta	gccaattggt	aatggcacat	acaaggacag	76980
tgctcttcac	ctcccagtcg	gtccaaagat	aagccagtgg	ggaaaagggc	agctacaggg	77040
caatggtttg	tcttgaacta	caaacctaaa	gaacaaacta	tgaccattgg	gaaactttct	77100
agttaaggag	cacacagttc	ggcccagact	gagtttgcca	gattaggaga	gagagagcac	77160
tgggcaggag	taaaggctag	aaggcttggt	tgcagggtgc	tctgctaacc	tcaccactcc	77220
tgggtccttg	gacagtttct	ttcacttctc	taaacctctg	tgtcccccata	tgcaaaatgg	77280
aaatgcaaaa	tacacatttt	accgaattca	tgagtttggt	aagagaatta	aaaaagagaa	77340
tgtgtatgga	atgttttcac	agtgcccaga	acagggttaag	tgctcagtaa	atgttaattt	77400
ttgtttgtaa	tgaaatacac	aaaacactca	agtaggactc	agctctgttc	cagctattct	77460
tcaagagggtg	tgacgaccaa	aatgtgtgta	aaattattaa	ttatttggtc	caaacacctt	77520
tcctcttcaa	tgtgtagtgg	agttccttta	tagattttatt	acaaaaaaaa	taagtttaaa	77580
caactaaata	cacctgtggg	aatgaaaatc	aaaaatagtt	actctggtat	taaatagcta	77640
ctgtagcatt	aattgctaaa	gaaattacat	agagaacaac	acacaggcat	aatgtaccat	77700
agcaataaccg	ttaaattcca	ggccctcctt	catgaaatgt	ctcttttctg	gcctcccttt	77760
taatcttctc	actttcttca	tgcacatcct	aaaagtccac	tccgtgtctc	tataaccccc	77820
acactttcat	ctcagattgt	ttatactcat	gtagttgtct	cttttattac	tagttttttt	77880
ctaactgcaa	tgaactctcg	gttttttttt	tttttttttt	gtaatcattt	tcttagagtt	77940
tgcacttttag	gtgcagtttt	tatttaatttt	attataaata	ttcacttgcc	taattgctat	78000
tcccactggg	gtattccaca	gagaacttaa	atttcacaag	tctcaatttc	aactcatttt	78060
cttcttctta	atgcaaatte	ttctcctatg	acttctaatt	gcaaacacaa	acctcatacc	78120
caggatagta	cacaccaaag	ccagattatt	tgaatttaac	tcccagatct	aatagtcagt	78180
tgctgtgtga	cctggaataa	attattaacc	ctctgtgcct	ccttcttctc	atctgatatt	78240
gatattactt	aactcatagg	gtctttatga	gggttaaatt	aattatcacg	tgtaaact	78300
cagaatagct	ccaaaaacat	ggcaattgtt	acatatgcaa	gtagcaactc	accaagaaaa	78360
actaatctga	aagccatctc	tgaactccta	cttccattta	gtcacttttt	tttttttttt	78420
ggagagagtc	ttgctctgtc	gcccgcactg	gagtgcagtg	gtgtgatctt	ggctcactgc	78480
aacctccatc	tcccagggtc	aagtgattct	cctgcctcag	cctcccaggt	aactgggggt	78540
acagggtgtgc	accaccacgc	ctggctaatt	tttgtatttt	tagtagagac	agggcttcac	78600

catattggcc	aggctggtct	tgaactcctg	acctcaagtg	atctgcctgc	ttcagccttc	78660
caaagtgcctg	ggattacagg	catgagccac	catgcctagc	cacttctttt	ctaaatattt	78720
taaatagatc	catatcagga	tatctctact	ttcattacca	ttgtttaaac	ccaatcttct	78780
ttaacctaaa	tagtgtaaca	ggctctgaaa	tggttcactc	atttgttcat	ccatcatgta	78840
acaggtaaat	attcacttct	tacttttaac	ctgacggtgg	gaagtattaa	tgtatgggat	78900
agtcatgctc	tcctccttct	ggtcttgctt	cccctaatac	agaataatca	gacagtgcc	78960
ttgccccatg	atccccaatg	ccttcatgtc	aaagtcaaaa	catttactgt	gccatgatct	79020
gacccttatt	taccactcca	tcctctctgc	tttccattcc	ttgtttgaat	ctcagctgta	79080
agcactcgtc	tccttgcttt	gtagagaagg	agccatgttg	ctactttcct	ctgggacttt	79140
gcacctgtca	tttactctga	tttccctcca	agtccaccct	atccacacac	ttttgaggag	79200
ctagttccta	ttggcgggtt	cagcaagacg	tcttccctgg	ttcccagacc	tatgttctga	79260
taacatgtag	aactcagtc	tatcatgaca	caaaattgca	ttgagttgag	ttttctggtt	79320
tattcatctg	cccgccatc	tgtaaagggg	atttggttaag	gaaattctaa	ccaaagggaac	79380
aaataaaact	tgaacttca	gtggcttaat	ataggtttta	tatatatgta	tatacaaaat	79440
gtatagatta	tatatattat	atatgtacag	tatgtaatat	tttagaaaata	atacactata	79500
caaaaatagg	ttatatacat	attctatagt	aatatatata	catatttata	tattatatat	79560
gagatatata	tctagaatat	atgttatata	tttatagatt	gtatattata	cataaaatgg	79620
tagaatattt	ttataaaaata	gaaaatata	aatcatatat	aatatatatt	tatatattat	79680
atattatatt	tacacataat	atataaatac	atttataaat	aatatatatta	tatatattat	79740
tatttatata	ttatatataa	atattatata	tttccctgat	tcatttatta	taatatgtac	79800
atgtgcatgt	ccctttaaga	gaagagaggt	atttctggaa	aggagttgtc	caagcagaga	79860
tgacttttcc	tagcctccct	gataatcata	tggagtcagg	gactgtgtct	tgcttattgt	79920
ttgttttctc	tgccatacat	agtacacagc	tcagaggatg	ttctcaataa	ataagtatga	79980
gataaattga	tactttacct	ctctggttgc	agcttgtctt	tttgctaatt	acacagatat	80040
gggaagggtg	taggattcaa	ataaatttta	caataagctg	tttccctttt	aaacttagag	80100
agcatctctc	aagtctaggg	gcatgtcact	tttccatcag	cttccctgct	atgaacactt	80160
tgtgaatcat	cctaaattac	atcaaaaaga	aagacatcat	gttctcttct	gcttggtgat	80220
acaagagggc	ttttccactg	gaaaaaatac	taaattttct	atttccctgtg	tggtaccagc	80280
aaaaactttt	ttcacaagct	gctgctttct	gggttagagt	gaataagtaa	ctaaccaatt	80340
agctcgttca	tttccaattc	aaggacatca	acaatcctct	catctattgg	tggcaaatc	80400
ttgtcctatg	tgaagtcaac	actttaatct	gtctatggca	agcattacta	attattcata	80460
cttacctttt	agccctgtct	gtggcttcag	ataaatttcc	aacgcaacca	tccaaatttt	80520
gaccacagc	gtgttacatt	tggcattgctg	gtgactcagt	tcctcatctt	tagtgtactc	80580
cttcatcggt	tttgagata	taaagaagtt	aaaggaaaag	aaggtaaaag	aactacaaag	80640
aacatgcaag	atgttggtag	caaagcaaaa	cgcaaaattt	tgggattctg	atcctcagcc	80700
actatttttt	ttttaaccac	ttggaaatag	acatctaaag	cagaaaagaa	tggtccagaa	80760
gccttggtgt	gtggggggtg	agtggatgtg	gctgtatgcc	ctaaagacaa	attttaagtc	80820
tcacaacttg	gctgcgggca	agctctggtg	tttagcaagg	tcttgtattt	tgggtttaca	80880
agatgctcat	ttgtttctca	atggttccag	cttgtccttt	cacctggaat	gttgtcagga	80940
acctggggcc	tagaaacagc	tgataaagat	aagctttgag	ccactgagtg	tccttgggca	81000
agttggcatg	ctgctccggg	cctcagttgc	ctgctcgaca	aaatatggtg	ttgggttttt	81060
gaggtaaat	tagattaact	tgtgtgaagc	actgaatcca	atatatggtg	tcttttaggtc	81120
caacctcagg	tctccctaca	tgttgtcaga	aaaagacatt	tgagcatttt	aagagtgaag	81180
tcaaaagtgc	actaccaagg	tttctaacag	cccagttctg	tcctggcttt	tcttccctgca	81240
actgttttca	gcttggtgga	gtcacttttc	tctctggacc	tccattttct	cacctgtaaa	81300
gtaaggagat	tggaaaatga	tcttaaggcc	ctttcatctt	tgaatatcta	tgattctgtg	81360
cctctctgta	tctgagtcct	cagaattctt	tttcttggaa	tatctccctc	agttgagaga	81420
aaattataag	aagggtgct	tcccaggttt	attttgggtg	cagtggccag	tctcccttga	81480
acttggcatt	tgtactaat	ctctataagc	aaagtgtgct	ccaattatga	cttcacccta	81540
ctggcccagc	agagctattg	ggtgctgcca	gtagggataa	aggagtgttc	cagccatgag	81600
gcaactattt	tgttacaatt	aatggacatg	gggttgagag	acatgtccct	caggcattct	81660
cttgtgccac	cttgaagagg	tcacctcagg	taaaagtctt	ctgaagagaa	atattggcag	81720
atacatatgg	caaaacacag	cctccaactc	cagtcaaacc	accaccacaa	gcacagtga	81780
aagcagaaaa	gaactgttgg	gtgtagtagg	gtggaggtga	agaccttgg	cctgccttta	81840
tgaagcatac	atcacttcgc	accaagcatc	acgcttggtg	ctggacacaa	agagatgatg	81900
aagatgcaaa	atgtgggtccc	tgtttttaag	gagatcacag	ttttctcagg	cagtcttaca	81960
gtaactatgg	cagtcaagat	attaataata	atatatatct	tggacattca	cctagagatt	82020
aaaatgatga	ttctagcaag	ataagtcata	aggtaattta	gattataggt	gggaataatg	82080
aatctttcag	aattactaag	tggggccttc	catagaagag	tataattcaa	gacacattgg	82140
tttcaccagt	ggtgacctcc	cctcaaagtc	agatacaggg	gagcagactt	aatattgttg	82200
gaggaaaaga	gagtcaaaac	ctaccagaag	gagcatcaca	aggagacaa	agaattgaat	82260

tgtttctgag	ttggtaatgt	gaaagttgct	agaaataatc	cttgctcagg	tgctctggga	82320
ctagtttatg	tgccagctgt	gtgcacaaag	tgaaaggctt	aattgctaga	actgtattac	82380
ttttggctct	agtctgctgg	accttccatc	tcccaaaactg	tagctccctg	gagctttctg	82440
gaagtgcctt	tgtattcctg	gacttaccta	tagcatactt	ttaaaggcaa	gggaagcttc	82500
ttaaactctc	ttaaatcagc	taagctatgc	tggggtagta	ataaacacct	atatctcagg	82560
gacttaacac	agattttctca	cccacatcac	agtccagtg	gggcttctct	ccatcctaaa	82620
actttccccc	tggaactcaa	gtctctctat	gttgccctga	aagggaagaa	agagccgcag	82680
ggtcaggatt	taaggaccat	gcctagaatg	gcttacataa	tttccaccca	cattgcattg	82740
tcccaaattc	agtcatatga	tcccaactta	ttgtcaccaa	aatgctaggg	atttggtcta	82800
ggccctgctg	ctcacagcac	agaatgccaa	tcaaagagac	aatgagtatt	gccatggaag	82860
aaggctttta	ttggatgctg	cagctgagga	gttaggagat	cagtctcaaa	tttgtctttc	82920
aaattagcta	aaattagggg	tttacaatagc	aggggaagaa	tgtaaccatg	tataagaaaa	82980
caggaattag	ggaggggtaa	ggaagagaag	ttgggtgaca	ggaagtaggt	ggaaagttag	83040
gcaatcctca	tgggtgaggg	gtctggcatc	tctttgtcca	gatgcagtga	tctggtaagt	83100
ttcacctcct	tgatactgtc	tgggaggtgt	gatggttggg	ttcctgagaa	agaaactgag	83160
ataagacaaa	tgtaactttc	ttgagtttca	agactggggg	gatcaatttg	tatgtttatt	83220
caatagaaa	aatagacatt	agttctgtag	aaaacttgga	ccagtttcag	tctccccttt	83280
ctattttcag	ttcctcgatc	atgggaaatt	tggtcattga	tctttctggc	tgcttcatgc	83340
tgaggagggg	cattacaagc	agctccatac	taagggtgat	gaatcaaagt	ttaatcta	83400
actgcagtct	ccttctatga	cacaatcttt	ctctctctag	tctcccactt	ccaccaaaga	83460
caaatacatg	caggaccaac	ctacctgcaa	aattagcttc	agtcccatat	acttggcccg	83520
attaccaca	gaaagtacag	caagcatcat	catccacata	gggtctccaa	aattggcttt	83580
cctggaacca	ttcacaaggc	catttcagtc	aaagccctgg	gaaaataacc	agttcctcca	83640
actgtgtctc	attgtaaaag	aaaacagatt	attattgaac	ttatgtaagc	aaccatattg	83700
ccataaatta	agaatattca	caaatagttt	acaaattcta	cagaaatcag	gcagagagag	83760
aaatgtgctt	caaattctat	tgacaagagt	acactctact	caattgctaa	aggttgtaaa	83820
cagctcaaaa	gaaaaagtgt	tctccagact	ctgaaaaaca	aaacaaaaag	aatcagcaat	83880
gtttcaaaga	acaaaaaaa	aaaaaaaaaa	aaaaccataa	aaattatttc	tgtcctccat	83940
tagctcagtc	catgcaatca	actcccactc	tgcttcatat	tgggttagta	atctttaaga	84000
acacatcagc	ttttcaatta	atatcctgga	agttttctct	ctagcccaat	ggcaaatct	84060
ccaaagttaa	cagaaacctg	cattcaagga	ttctttta	gaattcccca	aaagaagcaa	84120
gttcagact	gtagctgatg	ataagcctgt	ttttaaaaa	aattgaagca	aaataattat	84180
ggatgagaaa	agtctcaaca	gacacaattg	acaagaaaat	ttggttattt	ctgtgacata	84240
caaaaattta	acataatcgt	aattattact	gataacgtat	attaagacat	cagaatttta	84300
ggaatctcat	gcattaatgt	taaagcta	tttaacaaaa	ccttataaac	aaatctaatt	84360
tgaatcagtt	tgatcataag	gtaagaagtc	ttttataacc	ttttacaatg	ttttttatta	84420
aagataaaat	cattgctcta	agaaaactgt	gttattccat	cacaatggcc	cagatactgg	84480
tattgcatca	gtgtgctttt	gatattaaca	tttaatttat	agaaaacctt	gaactaattt	84540
gtcccttaaa	attagccctt	aaaatctcac	atgcgcccac	atcttccaca	atagtctctg	84600
ggcctacagg	gattgaatag	tttcaatttc	tagccctgtg	tctcgggcaa	gttggttcatt	84660
ttgattgtca	ccttctcctg	ggtctgaaaa	caaggctttg	actactgtca	atgttcaaga	84720
tttggcatga	gtctgtgcct	tttttagacc	caagagtcaa	agtcctaact	taacagtaca	84780
aggactagtt	aataggacat	ttatattata	gaaagtccctg	tcattctccc	tcacatgtca	84840
caaattaaag	cactgtgaat	tgggtgtctag	tagttaatgc	ctgcagcact	ttaaatcact	84900
gtattaaagt	ggctagatta	ttccttgc	atatctaatt	tctaaccattg	tagtgacaga	84960
actgtgaccc	aaagcatcaa	aatgtgtgata	ggtccaatgc	caaacttatc	caagtaagac	85020
aattaacttt	attctccatc	atttaagaaa	atggtaaatg	caaatatcgg	ttttggaaat	85080
tcaatatgaa	gataaataat	ctctttttgc	ttaaatacta	tacagtga	cagggacaaa	85140
gtaaaaacaa	gcacagaata	tttttttctg	ctatttttaa	agagcatcat	catacatctt	85200
caagactggg	ttctagatac	ggtactgaca	actattaggt	agatttcaac	tccagaatct	85260
tcaaaacaga	ataatactag	aatatatttt	gttttcaagt	acacacatta	aggcctaata	85320
gtgataacag	atttgaata	aaaaagaatc	actagaaaag	ctgacatttt	tactactact	85380
taatccaagt	gaatgtcact	tagttttgat	agtggtta	acaactaaca	tagcttgaga	85440
gaaatcccag	tcaatataat	ttcttttaag	gacaaaagcta	gtctttcctg	aacattaaaa	85500
ctttgtaccc	ctatcacaa	ttttcctcat	tacctaaaga	aaaagatctg	aaaccgattc	85560
aaattattga	ttgaattgaa	tttccttagg	gaaaaaaact	gtctaaacat	ttcttctctt	85620
acctgctttt	ccagataaat	aagtaatcta	ctatttctgc	tcagaactta	tttaagaaa	85680
caagggtttt	tgtttttttag	gggttctttt	ttgaccata	gcttaaagct	cttgtagctc	85740
tctagatcat	cagagtcagc	aaaaccaatc	aaatttttaa	tggctgggtg	cccctatcca	85800
tttttgcagg	cctgacaaag	atagcttaag	aacttttagat	aaatagagca	aataatgaat	85860
tgttggaaat	gcataaggaaa	caaaatgact	attcatagaa	gaaaatacaa	gccttccatt	85920

aaaaactaaa	actatcaatg	gttttttata	tgtgtatata	ccagtaaaac	ccaaaggaga	85980
acaaatagca	aataattaaa	aattaaaagc	aaaaacaaac	tggaaaccaa	cccccaattt	86040
ttcacctact	cagtttactt	tgaaggttac	agtataatcc	agagcctaaa	acaaacatat	86100
gttgatatt	ttgttcctat	tatacaaatt	aatgtctcta	agtccagcaa	catcactata	86160
ccttctgtgc	aattaagaaa	tttacttcaa	gcacgtgacc	agtaagtact	ttagtgtctag	86220
tactgtctat	gcagaatagc	aaatacagtg	tgaacaaaag	caatgcaagc	ctgtatgtaa	86280
aatttggtc	catgctaaat	ctggattcat	gcttaagtat	gttaaaaaaa	aaaaaagaaa	86340
agaattgcca	agctgccaat	atattttctt	acagtatttc	ttaattttacc	ttcatcaata	86400
ctaagatctt	tagctatgag	caatgttaat	tagccaaatg	tctccaattt	tctatcagat	86460
tttaaagagt	attttactac	ttattttacac	aaaccattca	acatgcttgg	acttcgtttt	86520
tgtcctagat	tttctttctt	tctttgtttt	tttttttttt	aaaatgacca	attattttat	86580
tttaggacaa	aagttttacca	tgggtccataa	tttgtatcaa	cctttacata	acttatgaat	86640
tagacaaaat	aatttgtttt	tctctgtaaa	aacacatctt	ctctcacaca	ttttatatac	86700
agaataatat	attaattgcc	attctgattc	ttagtaactt	taaatttttag	tgaaaaccta	86760
ggaagtaaga	cattctgaac	tagctgtcag	atattaacat	ttcagagatg	agaacattcc	86820
acaattttta	aaaacatgct	tcccaacgtc	ataacccttt	ctttctctcc	tttctctttc	86880
tttcttttcc	ttccttcctt	ccttcccttc	tttctttctt	tctttctttc	tttctttctt	86940
tctttgtccc	tccctccctc	cccccccttc	tttctttctt	ttcttttttt	cttttctttt	87000
tttttgacgg	agttttgctc	ttgtcactca	ggctggaatg	caatggcatg	gtcttggtc	87060
actgcaacct	ccgctcctg	ggttcaagtg	attttctctg	ctcagccttc	caagaaactg	87120
ggattacagg	tgcctgccac	cgtgccacgc	taatttttgt	attttttagca	gagatggggt	87180
ttcaccatgt	tggccaggct	ggtcttgaac	tccggtctc	aggtgatctg	cctgccttgg	87240
cctcccaaag	tgtgagatt	acaggcctga	gccaccacat	ccggcccat	aatcctttct	87300
taattggaag	tgaccagat	ggccaaaaag	catctattat	tcaatcctaa	ataacttcaa	87360
gatttcaa	tacattaaaa	agtgcagcta	caagcattta	tcccattttac	atgtacatga	87420
ttctttcatt	ttttcaaacg	atttatctag	attacttctg	aaaactgaaa	tattagacaa	87480
agctaatacat	catttcaagt	tatttccttg	ttaacctttt	ttatagcctg	cgaagatcag	87540
atgttcacct	aagtaagaac	acttttaaa	ttaaatacat	gggtattgtg	tcaataattc	87600
agaaaattca	gctattttca	ttaaactaac	attgaattag	tcttacttat	tgaaaaaagt	87660
cacacaaaca	aagactat	agttttggct	gggtatattg	ttttataacc	ttttatgcta	87720
aaacccgaca	ccccaaaata	tctagcagag	acaaataata	aaacccagaca	aaaaggtag	87780
ctggcaattc	caaagacatt	tctattttta	ccttaccaat	aatttttaaag	ctagttttta	87840
aaatttaaca	tttactta	caagtgaatt	taaaaaatc	ttgaacttat	ttacttaatt	87900
tacaagcatt	cttttgctta	taagcaaagt	tggtagacac	aacatataat	aaatgtacat	87960
acacataaac	acatctaaat	atgtatacac	acacagacac	aaagatccaa	tagcattttac	88020
ctcaaaactt	tagctatgaa	atagcaatag	aaactcacca	gtttacaaac	aggttcacat	88080
ggctaaacta	tttttgcccc	aatacataat	caaatgaagg	ctgtgaacca	aaatttgggg	88140
tagagcagtt	ctcatggcag	tttgcttttt	aaaggccata	ccttcccaga	tgccaaagag	88200
cactaggtcc	agatagcacc	acagaaaaac	atcatctata	acctactaat	caggcccaac	88260
cctgcttaga	acagcagcgt	aggagtctga	ctacatggaa	tttcatcttg	ccttctcatt	88320
caacagcaaa	ctccagatcc	caaagaatac	tggggccagg	ccaagtgcag	tggctaacac	88380
ctgcaatctc	agcactttga	gaggctgatg	tgggaggacc	acttgagtcc	aagagttaga	88440
gaccagcctg	ggcaacatga	tgggaactta	tctctccaaa	aattaaaaaa	aaaaaaagct	88500
aggcatgatg	gcatgtacct	ttactcctag	ctacttggga	ggctgagctg	ggaggatccc	88560
ttgagcccag	cagttccagg	ctacagttag	ccatgatgac	actattgcac	tccagcctgg	88620
gcaacagcgt	aagatcctgt	cttaaaaaaa	aaatggaata	ttgaagtcaa	acagcattac	88680
agaagaatat	cagttatatc	agtttatcaa	actctaattt	cccatgacta	tatcaacaca	88740
cacaaccaca	aaaatataaa	ccaactgctg	taacaacaag	ctctaagagt	atccaaactg	88800
aggcagtcag	ggtgcttccc	tctctcagtt	gggcttgctt	aacctataaa	tggaaattct	88860
ttaaaaaatt	tcccagccag	gtgtggttag	tcatgcctgt	aatgccagca	ctttgggagg	88920
ccgaggcagg	gggatcacga	ggtcaggaga	tcgagaccat	cctggctaac	acggtgaaac	88980
cccgtttcta	ctaaaaatac	aaaaaaatta	gccgggcatg	gtggcaggct	cctgtagtcc	89040
cagctactca	ggaggctgag	gcaggagaat	ggagtgaacc	cgggaggcgg	agcttgcagt	89100
gagccgagat	cacgccattg	cactgtagcc	tgggcaacag	agccagactc	tgtctcaaaa	89160
aaaaaaaaaa	aaaaattccc	aaattgagag	aagcagatac	cgtctaggcc	cacaaaggac	89220
acttttacct	atccagatgc	agatgtctaa	tttctaaggc	tgtttttcct	aggtaatcag	89280
gaacatggtt	ggggccagca	gtggtggggc	tggacagaga	gagaaactga	gactcacctc	89340
tggccaaaaa	agggctctggc	acctgcttag	gagggtctcc	aaaacttttt	cagcctgtgg	89400
aaacaaaccc	acaagcaatg	tgttactgg	cagggaacta	aaatctgtta	cctaaatgcc	89460
aggggtttat	tctaggttct	gctgcttgca	gcgcagaaag	ccggtcactg	agacagtga	89520
gatagccaga	gaaggcttta	atcaggtgct	acagccaagg	agatgacaga	taagtttcaa	89580

atccatctcc	ccaagcaacc	aaaattaggg	gtttatatag	taggaaagaa	atgtacccat	89640
gtataggaaa	atagaaacta	gggaggggca	agaaagagaa	gttgggtcaac	aggaagcagg	89700
tggttggtta	cgcaatcatg	atgggtgagg	gggtcttatg	tttcattgtc	aagatgcagt	89760
gatctgctaa	gtttcagctc	cttgatacta	tctggaaggc	ctgatggttg	ttttcctgag	89820
aaaggaactc	agataagaca	aacataactt	tcttgagttt	taagactgga	gcatcaattt	89880
ctatgtttat	cccaaagaaa	ccatacacgt	tagttctatg	agacaactgg	gacaatttca	89940
taacctgcat	tagaggttgg	gaaatgtctt	cccatatttc	caagataaga	aaatgggtgcg	90000
ggtgccaaaa	ggcattttgtc	ccagccatac	tcccctgtaa	tgtacacttt	tctcaaacat	90060
tttccaaatt	cactgctaca	atgctctgaa	ccctattttc	tgcagatgca	actttaacac	90120
tgtatatcc	cgatctgata	tgaacaaaac	aacagcacca	cattttccct	ggatattggt	90180
ttgccaaaggc	tttgatcgcc	actgcaatgt	gccagtaagg	caagaggaaa	atgagagatc	90240
tttgagttca	atgatcttgt	ctaattaagc	tggagctctt	ccttggaacc	ccaagccatc	90300
atgtttgtgg	tcagcattgg	agctgtgaac	atccttgcaa	ctggtgtttg	gcacaacctc	90360
ctttgatcag	cttctgttca	tccaccagca	ttcctatgcc	atagagagtc	aatgaggtaa	90420
gactaagcca	gatggaccac	ctacaaatca	cctggagacc	tctgaatcca	tagcaaaagaa	90480
attaagacaa	agttgaggaa	atctgaactg	ctggttctcc	agctactgta	gtcaaatgaa	90540
gaggacacag	aagccctgaa	tttttgaatt	cttaggatga	aaagggacac	attgggtgta	90600
tgactaactt	tatagggaaa	gaccttcaat	acttataatg	tgaactcaat	aaatagtaat	90660
gtagcaaaaga	gaagtttcca	tggggaaaat	aacatgtaaa	gtcaaaaactt	aaaaaaaaaa	90720
aaaagtaaga	tttttccaag	ttgcatttca	gaagagggag	gagaaaaagc	aaaggaacat	90780
gaaacgatga	aaaacaaact	ataacatatt	cctcacccctt	caacaataat	ctggagggtca	90840
aaagctaaat	tctttatcac	caaagaatct	tttgaccttc	acattggctt	attgggtgta	90900
attgtatttc	cagagtcact	ctctgattcc	tgatactccc	ttataaagtg	gccatcaatg	90960
agctgaagaa	agaaagagtg	gagcatctgg	atcactgaac	aacgggctgg	tttgacctca	91020
gaaagattgt	aaaggtggcc	ttgtgttaaa	tgagtcacta	agacaagccc	acacttagtt	91080
gcaaagagcc	agaaggaaaag	cttgtctctg	ttgggtgtgaa	atgggtgcagt	cttctcatcc	91140
attcatgaaa	ggaggaaaag	actttgacca	ttgggtgcaga	gcaaagctaa	agaggaagac	91200
caagataatg	acatatttttc	tccacatgcc	tttcaaaaatg	gtatttttatt	tactcttctt	91260
tcaagtaacg	tgaataaagg	gcaccacttt	tcacaaaatg	caaaaatactt	tatggtttaa	91320
tcctctcagc	cattctcaaa	atgggtacca	atctctatct	tacctataag	gaaaccaggg	91380
ctcagacaga	gtgaagggtt	tgttcaaaaat	ctcccagtta	gtctgaagca	gagcaaaaaat	91440
cacacacaag	atgatctgag	ttcaaaaacac	aagctgtggt	cactaaatgg	tcagaatgag	91500
ctcctctgtg	ctctgagagc	ctggcaaaaga	ctgggaatga	aatggacttc	cttgcccttaa	91560
atggaagggtg	ttgaggaaga	aagagtttaag	gtctcacaga	gtttaactaa	aggcaagttt	91620
tctcttccct	caggaaggaa	tcgtggatca	ttctccagct	ttttctctag	accatgttct	91680
aataaccttt	ctctccactt	ggctcctgat	cttatttcaa	tttctgagtg	taggggcttt	91740
tcattctgag	cacattatag	tattccagct	tgcccatcac	ccatctgtcc	caggagtgtg	91800
tcttgagggc	ttaaatgctc	agtgtggtgg	aggaaagttt	agcactagtt	tccttttggtg	91860
agggaataat	ttgcttatca	aataattaac	agaaatgttt	tagcagggtta	ctagttcctc	91920
tgcagaatgg	aacactgcat	agtacttcat	taccagctaa	tttcttctgc	aaatgaatga	91980
ctactttctg	aatttaatta	aactatgtaa	acattattgt	gaggggtcca	ctccgcaaca	92040
gagctgtgct	ctgctaggag	atacaaaaat	gagtaactca	tagtctctgc	ctttgagacc	92100
tctcaattca	tttgagtaaa	gtcagtaagt	gagtagcatt	ccatggatgg	gggttctcat	92160
aaaggtgttt	tcgttcacct	tagatgtcat	tgaggaatat	cagcaagtcc	taaattatct	92220
gattttttaa	taatagagtc	ttttacacat	gtagctattt	ttcatgcatt	gcatgtaatt	92280
ttctttggta	tctatatgga	tatgctttgt	gttagctttg	atactgggtca	cagttctttc	92340
ctgaggacag	aatagcatat	aactattaat	aataatgcc	gcatttggtg	ggtattttatt	92400
atgcatcaga	tgtagagct	cttctagtag	agaatctgtg	gccagtgaat	gtcatgtact	92460
actgtggagc	tttattctcc	aagcaacaca	gatgataaat	atagctatta	ctccaatttt	92520
gcagataaag	aaactgagac	tcagaaagct	tattaacatg	cccaaagtca	cacaagaaga	92580
actgtggatc	cgtttgactt	tttctaagct	tattttttaa	ctttatttaag	actaatccag	92640
agatgttagt	tttgcatccc	tttaagtgtc	atcttgcac	tgtaacaaaa	tccttgaaca	92700
ttcagcaagt	tctctatcct	gtaggtggca	taaactcaac	caattcccca	ggtctgtgtg	92760
agccctggca	atttttcagc	ttagagttcc	ctgataattc	ttctttccac	aaagtttttc	92820
tttgccagc	ctcttggaa	ttggagcagt	ttatcttctc	cggtagcttg	ctccacaaat	92880
ttgagctatc	gctgcttctg	ccccaatgg	gtcttcatcc	tcaactcaga	ttgccaaagct	92940
gttttgatct	ctgttggttg	ccccacagtt	catagattgt	ctctaggcaa	aaagccagaa	93000
taattgtggg	gcttgtctca	tttgcttccc	ttctttgggg	gtctcattca	tgctctgtct	93060
gctgtcctgt	attggaaaac	tgttgtcaaa	tatttaattc	cattttctag	tcatttgtga	93120
taagaaagta	tgctgctagc	agcagacatt	ctcccaagat	gtaaaaatgt	tcgtgggtgt	93180
ttaagtga	aaaaaatca	cactttaatg	cgacatatat	agtttatgct	tattatggaa	93240

aaatatgtta	atattctatt	gaatggatct	aaaaatacag	acataaaatg	tttgagatgc	93300
ttgatatccc	aattatcatg	atltgatcac	tacacattat	atacacgtat	caaaatatca	93360
catgtactcc	ataatatgta	caactattat	ttatcgatta	aaaaatgtca	atagtagttt	93420
ctaggtgggt	aagattgttg	ccattttttt	cttttttttt	ttttttctga	gaggaaacct	93480
acattgctta	gataacatga	aaaggaaagc	ttcaggaaaa	attaaatgac	aaaatataca	93540
gtccttaaaa	gcagaattac	aactaaaaata	tggcatctat	tttcatcatt	agagctgagt	93600
gatgtcaatc	agaatcagag	ctatcctact	agaaaactta	atgggtactat	ttaagcattc	93660
aaagatcaat	cactgctatt	ttactttaacc	ccaattttctc	tctatctcag	actaaaaaaa	93720
aaaatgtatc	caaggtgttt	ctttatttgtt	cccaataatt	tccccaatg	gtagcaatgc	93780
ctccacttta	caaaggggtt	ctgcctaaag	tttttcttaa	cacctttgtc	cctctgagct	93840
gtgttggtaa	tttgttttaa	aattgctatt	ttatttttct	ctttctctgg	tctgttttat	93900
cctgagagga	agaaggaagt	tgacgtttac	cactatggat	ccagcactgt	gttctcattc	93960
atatacttga	ttttatttga	ttttcaaaat	aatcctgcaa	agtattttga	attcacttta	94020
ggttgaaagt	ggggagacag	tctcaaaaga	caaagcactt	gccaagatt	ataaaactag	94080
taagcaacaa	atctagaatt	gagctctggt	atctagatgt	attttccaat	ttcttgaaag	94140
actatttttt	aggaattagg	tagaatttcc	cctcccttta	tgctctctcc	tcaatttcca	94200
tcaaagtttt	ataatatgca	agttagttca	caaaaaacat	gtggctctct	cctcttccaa	94260
ccatgtgcct	gggcacaact	ctgcatagta	tatataccaa	aatagtctct	gagcatatat	94320
atgcagatth	gtgaggtcat	tttcatttca	gttatcatct	tctctctact	tccatagcac	94380
tcagttaatt	catttatata	ttctttcctt	taactgatat	gcctgggacc	caggttatgt	94440
agtaataagt	aagactcgat	ttctcctttc	aggggtgctca	gagtctaata	tgaacacaca	94500
tccactcagt	gtagtaagta	tagtgttaga	agtggccatg	gagtcctaca	gatgcataaa	94560
gaaagggctc	tttattttaat	ttgcagagga	gaggagatca	gaaatggctt	cttgggagaat	94620
atcataactg	agtcaaatat	caaaagctat	gtaaaagtga	ggcaggtggg	taacatagga	94680
aagaggattc	aggtgctggg	aatattacac	aaaaaatact	tgtaagcatg	gaagaacaca	94740
gtgtaagtga	aaactcatga	acattttaaa	ctagttgaag	aacaagtaag	atggcggaat	94800
ggtgagaagg	aagatggaaa	cttagacaca	attcagatct	taagatggaa	agttttacatc	94860
ttaaaccaaa	gattgaatag	tcattgaaga	ttttcaggaa	acaaatgatg	taatcagatc	94920
tcctcttagt	gaaatcatgt	aggataccaa	tataaagaat	gctcacatat	tttactaaaa	94980
gtattccaaa	agataaagag	ggaatactcc	ttaaattatt	ctataaggcc	agtatcatcc	95040
taataccaaa	accaggaaa	gatataacaa	aaatagaaaa	atacagacca	atatctctga	95100
tgaacataga	tgcaaaaaatc	ctcaacaaaa	tgctagctaa	ctgagtccaa	cagcatatca	95160
aaaagctaag	acaccatgat	cgagtgggtt	tcataccagg	aatgcaggga	tggttgaaaca	95220
tttgcaagtc	ataaaatgtg	atacatcata	aacagaatta	aaaacaaaat	tcactttaat	95280
agatgcataa	aaaggcattt	gacaaaattt	aacatccctt	tataattaaa	agcctcagca	95340
aaatcagctt	agaagggaca	tgctttaagg	taataaaaagc	catctatgac	aaacccacag	95400
ccaacattat	accgaacagg	aaaaaggtga	aaacattccc	ctgagaactg	gaacaaaaaca	95460
aggataccca	ctttcaccac	ttctttttta	catagtacta	gaagtcctag	ctagagcaat	95520
cagacaagag	aaagaaataa	agggcatcca	aatcagtaaa	gaggaagtca	aactgctgct	95580
attcactgat	gatatgatca	tatacatagc	aaaccctaaa	gactcatcca	caaagctcct	95640
agatttggtta	aatgaattca	gcaaatttcc	gggatacaaa	atcagtgtac	acatatcagt	95700
agcactgcta	tatgccaaaca	accaagctga	gaataaaaatc	gagtactcaa	cgcattttcat	95760
aacagctgca	aaaaaataaa	ataaaatact	taggaatata	ctttaccaag	gaggcaaaaag	95820
agttctgcaa	ggaaaactac	aaaacgttgc	tgaaagaaat	catccatgac	acaaaacagat	95880
ggaaacacat	cccatgttca	tggatgggtg	gaatcaaaaat	tgtgaaaatg	accatgctgc	95940
caaaaacagt	ctataaatte	agtacaattc	ccatcaaaaat	accatcatca	ttcttcacag	96000
aactataaca	aaaaatacta	aaattcatat	ggaatcaaca	cagagcccat	gtagccaaaag	96060
cctgactaag	cgaaaagaac	aatctgatg	gcatcacatt	acccgacttc	aaactatact	96120
acaaggctat	cattaccaga	acagcatggg	actggtataa	aaataagcag	gtagaccaat	96180
ggaacagaat	agagaacctt	gaaataaagc	caaatactta	cagccaactg	atctttgaca	96240
aagcaaaaca	aaacataaag	tgaggaaaag	acaccttatt	caacaaatgg	tgctgggata	96300
attggcaagc	cacatgtaga	agaatgaagc	tggatcctca	tctctcacct	tatacacaaa	96360
tcaactaaag	atgaatcaaa	tacttaaatc	taagacctga	aactataaaa	attctagaaa	96420
ataacattgg	aaaaactctt	ctagacattg	gcttaggcaa	agagtacatg	accaagaacc	96480
caaaagcaaa	tgcaacaaaa	acaaaaaata	aatagagggg	acctgattaa	actgaaaacc	96540
ttctgcacag	cagaataaat	aaacagcaga	gtaaaaagac	aacccgcaga	gtaaaaagac	96600
aacccacgga	gtgggagaaa	atattcacia	actatgcatc	cgacaaggga	ctaacatcca	96660
gaatctacaa	ggaactcata	ggaatcagca	agaaaaaaaa	acaaataatc	ccatcaagaa	96720
gtgggcaaa	gacgtgaata	gacaattttc	aaaagcagat	atacaaatgg	ccaatgagca	96780
tatgaaaaaa	atacttaaca	tccctaatta	tcagggaaat	acaaattaaa	accacgatga	96840
gataccacct	tcctcctgca	agaataacca	taattaaaaa	attgaaaaat	aacagatgat	96900

ggcatggatg	tgaggcaaaa	ggaacacttt	tacatggctg	aagggaattt	gaactagtac	96960
aaccactatg	ggaaacagta	tgagatttcc	aaatagaact	actatttgat	ccagcaatcc	97020
catatatatg	gcatttctat	atataccatg	gaatactaca	cagccataaa	aaggaatgaa	97080
ataacggcat	tcacagcaac	ctgaatggag	ttggagacca	ttattctaag	tgaaataact	97140
caagaatgga	aaaccaagca	tcgtatattc	tcacttataa	gtgggagcta	agctgtgagg	97200
atgcaaaggc	gtaagaatga	tataatggac	tttgaagact	tagggggaag	ggtggaagtg	97260
aggtgaggga	taaaagacta	cacattgggt	acagtgtaca	ctgctcaggt	gatgggtgca	97320
cctaaatctc	agaat					97335

<210> 3
 <211> 163
 <212> DNA
 <213> Homo sapiens
 <400> 3

caagtttagc	tgtgatgtac	aggtttctta	actattagat	ttctcagatt	ctaatatgaa	60
acaatgccaa	ttttcctgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgat	tgagaggtttg	120
ttggcaaatt	gatgctaaag	catatttggc	tttggttctg	gaa		163

<210> 4
 <211> 26
 <212> DNA
 <213> Homo sapiens
 <400> 4

caagtttagc	tgtgatgtac	aggttt				26
------------	------------	--------	--	--	--	----

<210> 5
 <211> 20
 <212> DNA
 <213> Homo sapiens
 <400> 5

ttccagaacc	aaagccaaat					20
------------	------------	--	--	--	--	----

<210> 6
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 6

aacaaacata	ggttgcgttt	atgcgaatgg	tcagggtccaa	agtagatgca	gaatatgccca	60
ggttcactaa	ttttaatccc	tattcagccc	aggactatgt	accataagat	tactgctagt	120
gtttttctgaa	aatgatgtat	caaggcattt	tctgtagaaa	tacgaaacag	tgacatacag	180
tagggagagc	tggtattgagg	cagagtagta	tagatggaag	tttcctgaaa	gcattttggg	240
gaaacatctt	ttgggtatgg	ttcttggatg	aagagttgat	ttattagtag	tggaagggttr	300
tatgggagag	aggaagttag	aggttatgag	agaatgaccc	tcccgtgatg	gtgagtggga	360
gaattattgc	agtatgtacg	ttagcattgc	tatgtggtga	agttcttggg	atttcctggg	420
gtccgtgctg	gacagcatgc	ttagccacca	gtcacatgtg	ggcactgagc	actgacaatg	480
tggttagtct	gaactgggat	atgctgtaag	tgtaaaatac	aaactgggact	ccaaagattt	540
agtatgaaaa	aaagaatttg	aaatatctca	ttaatgatgt	gtatttggtt	tcattattga	599

<210> 7
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 7

ggatacagag	gctaccagag	taggcagttg	acatctcttc	tctccctttt	tgtctcttct	60
atttacctcc	tttcttccat	aatcaagtac	cagtcctgtc	ctgtctccat	tagctgctat	120
ctttaccttc	tatctcccag	aaatgtacaa	attatgtaat	tcaagtttct	ataactgcct	180
tcccctccat	actaaaacat	gttactacac	cttgggtttct	gtcttcttat	ttttctgaat	240
ctgcactagt	cctttcattg	ctatagctgt	ccttcctctg	ttctaaggct	catggcaccr	300
ttgtaagccc	agggccctat	gctgacactc	atataacaac	attttgctcc	atcagttact	360
aaaagtgcc	tcttatatca	gcagtctccc	tgtctttgga	gagtatacaa	catgtattct	420
catcccattt	ttccattatg	ttcagagcgg	ttgaatgact	tgtacaagtt	aatgcaagta	480

gttgaggcag	aaatgatact	tcaatcgaga	tcttcaaact	tccaagttca	gttcactttc	540
ttccataaca	tgttatctca	aaaataccat	ttgccagaga	taagtaatgc	taaatcatt	599

<210> 8
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 8

gatgtgactg	cacatccatt	tcaaattgcc	tgaaactcag	gatgtgctcc	tgagggtgctc	60
cccatgccag	gatgtgcctg	tggtttccctc	aagagctact	tggtatacacc	gcataccctt	120
ggatccagca	ctgctagaga	ctctgcagaa	actctgcaac	tttaactgga	aacctctcta	180
ctgttggtgt	tcccttgga	aatgactatt	tgccacttca	gctgccaat	atctcctccc	240
aatatgttct	gtgagcattc	agtctgctca	aagtcttggtg	taatggtgga	tatatgatgm	300
catttgtttc	aagctaggct	cacatatcca	tctatctcag	gggtacatg	atttacttgc	360
cattttttta	acttttactt	gtaatttaaca	aataataatt	gtatatattt	atgggatata	420
atgtgatgtt	ttcatatatg	tatacattgt	ggaatgatta	tatcaagcta	attaaattag	480
gcatactctc	atatactttt	ctatactgaa	aacattttaag	gcctattctt	tcagcaattt	540
tgagatataa	aagacattat	tgtaatttat	ggtcaccaca	ctgtgcaata	gcattgaaa	599

<210> 9
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 9

gttctgtgaa	gcgggtaaaa	acaaaatttg	gcatccagtt	tcaaaaggag	aattgcaaac	60
taatagaaca	tatagcacia	aatgattata	tcaatagaat	gctaattgca	tatcaaggat	120
atttggtata	atacaaatat	ttctacctta	aacatatgga	aatttggtgt	ccatgatgtt	180
gtagattcta	tcttcccact	ctgcattttc	aaaggcatat	ggtattgact	cattcgatta	240
attgttggt	agtctttatt	atagactaaa	tcatagaata	aatacatgga	tacatgcacr	300
aatattatat	ctcaagggtc	ttacatagtt	catttatctca	cttcatagtc	aaaacaaacc	360
tactgatagt	tccaatgcaa	agcctagaac	gctttggctt	agagaggccc	aagtcttttc	420
tcagtgtctg	actgctggta	cgtggcgtgg	tcccctctct	tctctcagta	cacactaccc	480
atgcagacta	tcactctcag	tcttgtttat	ctcaaataca	gaggggtata	ctaactggaa	540
tgtatccaga	acagtgaggc	caaagtgtgg	ggaagctcct	taaccatgct	gctgcatga	599

<210> 10
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 10

ttgcaaacta	atagaacata	tagcacaaaa	tgattatatc	aatagaatgc	taattgcata	60
tcaaggatat	ttggtataat	acaaattatt	ctaccttaaa	catatggaaa	tttgtgtgtcc	120
atgatgttgt	agattctatc	ttcccactct	gcattttcaa	aggcatatgg	tattgactca	180
ttcgattaat	tggttgatag	tctttattat	agactaaatc	atagaataaa	tacatggata	240
catgcacgaa	tattatatct	caagggtctt	acatagttca	ttatctcact	tcatagtcam	300
aacaaaccta	ctgatagtgc	caatgcaaag	cctagaacgc	tttggtcttag	agaggcccaa	360
gtctttttct	agtgtgtcac	tgctggtacg	tggtgtgtgc	ccctctcttc	tctcagtaca	420
cactacccat	gcagactatc	actctcagtc	ttgttttatc	caaatacaga	gggtataact	480
aactggaatg	tatccagaac	agtgaggcca	aagtgtgggg	aagctcctta	accatgctgc	540
tgcatgagga	acagctggag	agactgagaa	catgaggcct	aaagaggaga	ctcaggggag	599

<210> 11
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 11

ctaccttaaa	catatggaaa	tttgtgtgtcc	atgatgttgt	agattctatc	ttcccactct	60
gcattttcaa	aggcatatgg	tattgactca	ttcgattaat	tggttgatag	tctttattat	120
agactaaatc	atagaataaa	tacatggata	catgcacgaa	tattatatct	caagggtctt	180
acatagttca	ttatctcact	tcatagtcaa	aacaaaccta	ctgatagtgc	caatgcaaag	240
cctagaacgc	tttggtcttag	agaggcccaa	gtctttttct	agtgtgtcac	tgctgtgtacr	300

tggcgtgggc	ccctctcttc	tctcagtaca	cactacccat	gcagactatc	actctcagtc	360
ttgtttatct	caaatacaga	gggtataact	aactggaatg	tatccagaac	agtgaggcca	420
aagtgtgggg	aagctcctta	accatgctgc	tgcagtagga	acagctggag	agactgagaa	480
catgaggcct	aaagaggaga	ctcagggaga	tgggatcaca	atcttcaa	atcttcaa	540
catcaagggg	aaaagagatt	aaacaaggta	atgtagctct	agagagcaaa	tccaagagt	599

<210> 12
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 12

tccatgatgt	tgtagattct	atcttcccac	tctgcatttt	caaaggcata	tggatttgac	60
tcattcgatt	aattgttgga	tagtctttat	tatagactaa	atcatagaat	aaatacatgg	120
atacatgcac	gaatattata	tctcaagggc	tttacaatag	tcattatctc	acttcatagt	180
caaaacaaac	ctactgatag	ttccaatgca	aagcctagaa	cgctttggct	tagagaggcc	240
caagtctttt	ctcagtgtct	cactgctggt	acgtggcggt	gtcccctctc	ttctctcagy	300
acacactacc	catgcagact	atcactctca	gtcttgttta	tctcaaatac	agagggtata	360
actaactgga	atgtatccag	aacagtggag	ccaaagtgtg	gggaagctcc	ttaaccatgc	420
tgctgcatga	ggaacagctg	gagagactga	gaacatgagg	cctaaagagg	agactcaggg	480
agatgggatc	acaatcttca	aatatttaaa	agacatcaag	gggaaaagag	attaacaaga	540
gtaatgtagc	tctagagagc	aaatccaaga	gtgttgagtg	gaagtgaag	ggaggctgg	599

<210> 13
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 13

atgatggttt	ccagcttcat	ccatgttcct	acaaaggaca	tgaactcatc	atcttttatg	60
gctgcatagt	attccatggt	gtatatgtgc	cacattttct	taatccagtc	tatcattggt	120
ggacattttg	gttggttcca	agtctttgct	attgtgaata	gtgcctcaat	aaacatacat	180
gtgcatgtgt	ctttatagca	gcatgactta	aaatcctttg	ggtatatacc	cagtaatggg	240
atggctgggt	caaagtgtat	ttctagtctt	agatccctga	ggaataaatg	accaactaty	300
gagaaattgc	agggtagtcc	ctacatgagg	gttaggtaga	attgacctgc	tttctgcctc	360
ataaatttta	gaaaattaat	aagataattt	attacggggt	ggtgtttggt	ccctcagtag	420
tttatcatct	atggttgata	tggttaataat	taattgcata	attaacaaat	agcaaattat	480
tgtgggggtg	tgtgtgtgtg	tgtgtgtgtg	tgttttagaca	gggtcttgct	gtgtcaccca	540
ggctggagtg	cagtggcggt	atctcggtc	actgcaacct	gtgccttcca	ggttcaagc	599

<210> 14
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 14

atcctactgt	gatcaaaggc	acatacatga	gatggtgagt	tgtccccttg	ccaatgaggg	60
tttggttaaga	aaggaaagtg	cagtacttct	ttgtttctga	attgcaagta	tgtgtggggt	120
agagggggag	gctgaatatg	aaggctcctg	gacagccac	caggatatccc	atgagacttt	180
gcaaaggaaa	aggaggtgag	tgacagccca	gggtccaata	ggatagaagg	aaaagccagg	240
ccatggagtt	cctcagacct	gctttctaa	ggcaactcta	ccacctcagc	aagccattgr	300
acttctctga	gctcagtcct	ttcattttata	aaatgggggtg	acagtgtctca	catgccagga	360
atacaaaggg	attgaaagat	aaaacacgta	attaagcacc	tggtgttaca	catctgtcag	420
ggaccccaat	aaggtcagct	gtcttctctg	tgacttctgt	tcttggtggt	tctccaagat	480
cataccttcc	atcaacattt	accgtcactc	ccccaccoca	tgcccaatac	tgaacagtgg	540
agggacgctt	cacctacagt	tataatgttg	aaacttcaac	ccaaagcaag	tactgttag	599

<210> 15
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 15

tttaaagtct	accatttccag	gtccatattt	tcgcttgaaa	attgagattc	ctattaaaca	60
atgacattta	cacaaaaaag	tagaggagtt	ggttgaagga	cagggtaatg	ccaggaggaa	120

ttgggaattt	gagagtcaag	tcaaaggact	gaaataactca	gaataactaag	ggcacctcag	180
ggctctacca	aggacacgta	gaagctttga	atttgcagca	ccaccctaata	ttaacgagct	240
acctcagcac	gtagtggagc	cttggaaaac	agatgtcaca	aactctcatt	agattgtcar	300
acattttcca	gcattttcctc	tcccatcata	gctggttatc	aagatatata	gacacacacg	360
tgcatacaca	taaatacctt	gataagttac	tagagaaagc	agaaaaatgt	ctgacagttt	420
aatgagattt	gggtgaaaga	aaattctata	tttcattggt	ttccaggcac	tagaaataat	480
tcatcaatgt	ttctaagact	cattcagcgt	ggctgcattt	tttaaaatat	tttcataaat	540
tttgaggagc	aaataccatt	attaggcact	aaaaagggtg	aagtctaata	gatttagccg	599

<210> 16
 <211> 549
 <212> DNA
 <213> Homo sapiens
 <400> 16

tattttcgct	tgaaaattga	gattcctatt	aaacaatgac	atttacacca	aaaagtagag	60
gagttggttg	aaggacaggg	taatgccagg	aggaattggg	aatttgagag	tcaagtcaaaa	120
ggactgaaat	actcagaata	ctaagggcac	ctcagggtct	taccaaggac	acgtagaagc	180
tttgaatttg	cagcaccacc	ctaattttaac	gagctacctc	agcacgtagt	ggagccttgg	240
aaaacagatg	tcacaaactc	tcattagatt	gtcaaacatt	ttccagcatt	tcctctcccr	300
tcatagctgg	ttatcaagat	atatagacac	acacgtgcat	acacataaat	accttgataa	360
gttactagag	aaagcagaaa	aatgtctgac	agtttaaatga	gatttggttg	aaagaaaatt	420
ctatatattca	ttgtttttcca	ggcactagaa	ataattcatc	aatgtttcta	agactcattc	480
agcgtggctg	catttttttaa	aatatatttca	taaattttga	ggagcaaata	ccattattag	540
gcactaaaa						549

<210> 17
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 17

acaaaatccc	caaagcaatt	ttggaataag	agccaacatt	taatacttac	cagacaacta	60
ttctaagtat	tttactatat	tcactcatag	caactctaaa	aagcaggtag	tattaacaga	120
gaaaatgagg	cacagtgagg	ttaaatagct	ggtcogaggc	tacacagcta	atcagtggga	180
gagttgggac	ttagaccag	aggtccagtt	ttgaagtcca	cacttttagc	cattacacta	240
caatggaaaag	aaatttagaa	gatatacaca	gaaaactata	ggcacataga	ttagggggtm	300
gtagaatgct	ctgggcagtt	aaaggaactc	ttcttaaagg	aggtaaagct	tgaatgagac	360
tgttagtaag	ctattttttca	ctcattgggtg	aatgatgttt	tgtgcagtg	gtttttttcc	420
ccatagaaaa	ataagaaaga	aaagaaaatt	gagaactctc	tctataaaaa	tgtgtaacat	480
atctcatatt	ccaagagatc	cttttggtag	tattaatttt	tatctgctca	cagtactggc	540
ttcattattt	ggagttaaaa	attaactcaa	ccagataaaa	aatcagtg	tgtgtattt	599

<210> 18
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 18

gttctaattt	taaaaagtta	tttaacagaa	cgaagctatc	agctaagaca	atggcaaagc	60
cgtaaacaaa	cataggttgc	gtttatgcga	atggctcaggt	ccaaagtaga	tgagaatat	120
gccaggttca	ctaatttttaa	tccctattca	gccaggact	atgtaccata	agattactgc	180
tagtgttttc	tgaaaatgat	gtatcaaggc	attttctgta	gaaatacgaa	acagtgcacat	240
acagttaggga	gagctggatt	gaggcagagt	agtatagatg	gaagtttcct	gaaagcattk	300
tggggaaaca	tcttttgggt	atgggttcttg	gatgaagagt	tgatttatta	gtactggaag	360
ggtgtatggg	agagaggaag	tgagaggtta	tgagagaatg	accctcccgt	gatggtgagt	420
gggagaatta	ttgcagtatg	tacgttagca	ttgctatgtg	gtgaagttct	tgggatttcc	480
tgggggtccgt	gctggacagc	atgcttagcc	accagtcaca	tgtgggcact	gagcactgac	540
aatgtgggta	gtctgaactg	ggatatgctg	taagtgtaaa	atacaaactg	gactccaaa	599

<210> 19
 <211> 599
 <212> DNA
 <213> Homo sapiens

```

<400> 19
cccccaacc ctatTTTTaa attatttact tatttttatt atatttttga tacataatag      60
atggacacat tcaaacagtg ccccaaaaac tggggcagca gaaacaggtc cttgcttatt      120
ttctcagctt cacctcctgc ctccacccca tctgtactgc tgggccagac attcctacag      180
aggtgtcctc ctaagttggg ctcttcctct cctgcttcag aggccttgcc ctgctcttct      240
ctgcctcttg aggcctctgc ctgctcttct ctgcgtcttg tggttggaat gcctgtctty      300
ttcctactga agatctggat gcctaaacca taatgtaaaa ttgctgcttt ttacttccat      360
ttacagcaga gaaattcctc ctctggcctc tcctcttctc tgtgtttctt tcttcataat      420
ttttatttat ttatatattt atttatttat ttattttatt atttattttc attgagatgg      480
agtctcggtt tgtcgcccag gctagagtgc agtgggtgta tctcagctca ctgcaatctc      540
caccttccag gttcaagcga acctcttgct tcagcctccc tcctgtagct gggactaca      599

```

<210> 20

<211> 599

<212> DNA

<213> Homo sapiens

<400> 20

```

gtcatgatta aatgttaagt gaaaaaatag aatatagcta gatttgaatt tgaatattca      60
atctgtatac cagtatgtat agaaggaaga gtatatacca aatagtaaga gtatctatct      120
gttttataat ttgatataat acaaattatt ctaccttaaa catatgagaa tttgtgggtcc      180
atgatgttgt agattctatc ttctcaccct gcatttccga agacatatgg tattggctca      240
ttagactatt tgttgaatag tctttattct attatcatag aaaaaataaa tgagtgcatr      300
tatccatata caaaatagag gtctgttctt cctgtatata ttatactaa aaaaactgag      360
actttttttt acagttgtat atatacaaac atatttgttt atttatatac acatatataa      420
atcaatttta tgtacatgtg ggtatacata catccatgca tataactctg aagtgcgtgac      480
tctctaaaga aagcccagggt attggtcaga attcatgctc ggctcaggag tatagaatta      540
agagatacaa acctcaaaaa agagggaacc gaatcttcaa atctgagcca ccttacaag      599

```

<210> 21

<211> 599

<212> DNA

<213> Homo sapiens

<400> 21

```

ataataaaat aagggctaac acttaaaccg ttgtgtactc actatgtact aggcactgat      60
caaagtactt tgtacatatt ttcttattta atattcgcta ccatcatatt acaatatact      120
gttattaacc ccaatgtata gatgtaggtg aagaaacttg tcacaaatca tacagctagt      180
tgtctgagat gcaatccatg tgatttggtc acagagctca gggtctgtga agcgggtaaa      240
aacaaaattt ggcattccagt ttcaaaagga gaattgcaaa ctaatagaac atatagcacm      300
aaatgattat atcaatagaa tgctaattgc atatcaagga tatttggtat aatacaaaatt      360
attctacctt aaacatatgg aaatttggtg tccatgatgt tgtagattct atcttcccac      420
tctgcatttt caaaggcata tgggtattgac tcatcgattt aattggttga tagtctttat      480
tatagactaa atcatagaat aaatacatgg atacatgcac gaatattata tctcaagggc      540
tttcatatagt tcattatctc acttcatagt caaaacaaac ctactgatag ttccaatgc      599

```

<210> 22

<211> 599

<212> DNA

<213> Homo sapiens

<400> 22

```

tttgggtatg gttcttggat gaagagttga tttattagta ctggaagggt gtatgggaga      60
gaggaagtga gaggttatga gagaatgacc ctcccgtgat ggtgagtggg agaattattg      120
cagtatgtac gttagcattg ctatgtgggt aagttcttgg gatttcctgg ggtccgtgct      180
ggacagcatg cttagccacc agtcacatgt gggcactgag cactgacaat gtgggtagtc      240
tgaactggga tatgctgtaa gtgtaaaata caaactggac tccaaagatt tagtatgaaw      300
aaaagaattt gaaatatctc attaatgatg tgtatttggt ttcatattga aaacaacttt      360
ggtattatat attgagttaa ataaaatgtc attaaaatta aattttactt aaactaaaat      420
ttaaaattct atctttacct tttttttttt tttttttttt ttttgagggt ccgtttccact      480
cttggtgtcc aggccttgagt gcaatgggtg gatcttggct caccacaacc tccgtctcct      540
gggttcaagc gattctcctg cctcagcctc cctggtagct gggattgcag gcacgcacc      599

```

<210> 23

<211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 23

tattcagccc	aggactatgt	accataagat	tactgctagt	gttttctgaa	aatgatgtat	60
caaggcattt	tctgtagaaa	tacgaaacag	tgacatacag	tagggagagc	tggattgagg	120
cagagtagta	tagatggaag	tttcctgaaa	gcattttggg	gaaacatctt	ttgggtatgg	180
ttcttggaatg	aagagttgat	ttattagtag	tgggaagggtg	tatgggagag	aggaagtga	240
aggttatgag	agaatgaccc	tcccgtgatg	gtgagtggga	gaattattgc	agtatgtacr	300
ttagcattgc	tatgtggtga	agttcttggg	atttcctggg	gtccgtgctg	gacagcatgc	360
ttagccacca	gtcacatgtg	ggcactgagc	actgacaatg	tgggtagtct	gaactgggat	420
atgctgtaag	tgtaaaatac	aaactggact	ccaaagattt	agtatgaaaa	aaagaatttg	480
aaatatctca	ttaatgatgt	gtatttggtt	tcatattgaa	aacaactttg	gtattatata	540
ttgagttaaa	taaaatgtca	ttaaaattaa	attttactta	aactaaaatt	taaaattct	599

<210> 24
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 24

tcccagcact	ttgggaggct	gagggtgggtg	gatcacgaga	tcaagagatt	gagaccatcc	60
tggccaacat	ggtaaaacct	cgtctctact	aaaaatacaa	aaaattagct	gggcatggtg	120
gcacacacct	gtagtccag	ctgctcgga	ggctgagggtg	ggagaatcac	ttgaacctgg	180
gaggcagagg	ttgcattgag	ttgaaatcac	gccactgcac	tccagcctgg	tgacagagca	240
agactccatc	tcaaaaacaa	acaaaaatta	tagtatgaaa	taggcattaa	aatattgtgy	300
atttttagagg	agactgagga	ttggaggctg	agaattact	ctaaattaat	cagcttgtgt	360
acttcagagc	taagatagct	ctttgggttc	taaattctgt	gatcttcttt	ttgatttctc	420
ttggagcaat	aatgaaggca	aaacatcaat	aaacataaca	aactgggtaa	gggagaccat	480
tgagaaggac	taaggacacc	ttcaaagttc	tgagtgaagt	taaaaagaag	aatgatgaaa	540
actttgatag	aaataggaaa	aaaagtagag	gaacttggtt	ggcttgaaac	ttcttaatg	599

<210> 25
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 25

aaaaaatagc	actgaaaaca	gaaaggaagc	atcaaaaactc	ttcaaatacc	tgtgtgtgcc	60
attggtcaag	cacattcagg	acatcgcatg	cctttagaac	tccagcaggt	tccaacagct	120
agtaggacat	tctagactct	gagagagagc	aaggagggtt	ttatgactgg	ggacaaagaa	180
aagagacact	gaaggcgaag	gacaatctct	gaaaatgcag	taccctccag	actgctcctc	240
ctctcacaaa	aacaccttcc	cagcatgcac	tgctttaggg	actatgatta	taccattgay	300
tctgtccaga	aaacctgtgt	cctgaatata	ttacagggct	cattccttca	cttctttcag	360
gtgcctactc	aggtatttcc	ttatcagaac	agtctttcga	acgaccccat	taaaaaaata	420
gtcctgtcaa	ccctatgtta	acaattttat	ttatttttat	tatttgttaa	caatacataa	480
taggtgcata	tattttgggg	gtacatataa	taatttgata	cattcatatt	gtgcataaag	540
attgaatcgg	agtaattggg	atatccattg	ccttaagtgt	tttacctttt	ctttatgct	599

<210> 26
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 26

tgtgtgtgtg	tgtgtgttta	gacagggtct	tgtgtgttca	cccaggctgg	agtgcagtgg	60
cgtgatctcg	gctcactgca	acctgtgcct	tccagggttca	agccatcatc	ctgcctcagc	120
ctccctagta	gctgggatta	caggcgctcg	ccaccatgcc	cagctaattt	ttgtattttt	180
aatagaaatg	ggatttcacc	atgttggtca	ggctagtctt	gaactcctga	catcagggtga	240
tccatccgac	tcatttccca	aagtgtctgg	attacaggca	tgggccatca	tgcttgcccy	300
gcaaattgtt	gttattttata	actcttcaat	ccaaatcatc	agtgtctatg	ttgtttcctt	360
aactatcaaa	tgatgataat	aatagtacct	tcttcataag	atagttgaaa	ggtttttaaat	420
atccatatgg	tactgagaat	gatgcctgaa	acatagtaac	taccccatth	ttattatatt	480
tctgttaata	ataatacata	ccattattgc	tcttgcatac	catattgctc	ttgcatacca	540

tatatgtctct tgctatatgc tacacacagt atttcattta ggccctcacta tgtccctga 599

<210> 27

<211> 599

<212> DNA

<213> Homo sapiens

<400> 27

gaaagcagaa	aaatgtctga	cagtttaatg	agatttgggt	gaaagaaaat	tctatatattc	60
attgttttcc	aggcactaga	aataattcat	caatgtttct	aagactcatt	cagcgtggct	120
gcatttttta	aaatatattc	ataaattttg	aggagcaaat	accattatta	ggcactaaaa	180
aggttgaagt	ctaatagatt	agccgcttca	tcctccttca	ctcagctcag	cattcgttca	240
actggctctt	actggttaac	atccacacgc	ctcctgactg	gctactcagt	gccgatgacr	300
tttccttcac	acacagggct	ggttttaaga	tacattgagg	tgacatcagg	tgccctgtaa	360
agtggtcatt	ttaggatatc	ctattcaaag	acatctgtgg	aagtgtggac	caatttattg	420
atgaataaca	gtgaaggggt	ttccaccagc	aagtaacata	attttttaca	atgatgatgc	480
tgaagtagaa	agagtttcta	gtcagggact	ggacaaatca	atttgcagac	gatttttagg	540
aagaaaaaca	ttgcaacagt	aaattgtaat	tgataacttc	tagagccact	ttaagtact	599

<210> 28

<211> 599

<212> DNA

<213> Homo sapiens

<400> 28

taattgcata	attaacaaat	agcaaattat	tgtgggggtg	tgtgtgtgtg	tgtgtgtgtg	60
tgttttagaca	gggtcttgct	gtgtcaccca	ggctggagtg	cagtggcggtg	atctcggctc	120
actgcaacct	gtgccttcca	ggttcaagcc	atcatcctgc	ctcagcctcc	ctagtagctg	180
ggattacagg	cgcttgccac	catgcccagc	taatttttgt	atttttaata	gaaatgggat	240
ttcaccatgt	tggctaggct	agtcttgaac	tcctgacatc	aggtgatcca	tccgactcak	300
ttcccaaagt	gctgggatta	caggcatggg	ccatcatgcc	tggcccgcga	attgttgta	360
tttataactc	ttcaatccaa	atcatcagtg	tctatgttgt	ttccttaact	atcaaatgat	420
gataataata	gtaccttctt	cataagatag	ttgaaagggt	tttaatatcc	atatggtact	480
gagaatgatg	cctgaaacat	agtaactacc	ccatttttat	tatatctctg	ttaataataa	540
tacataccat	tattgtctct	gcataccata	ttgtctctgc	ataccatata	tgctcttgc	599

<210> 29

<211> 599

<212> DNA

<213> Homo sapiens

<400> 29

ggtgtgatca	cagctcacta	aagcctccaa	ctcttgggtt	caagtgattc	tccacctcag	60
cctcccaaatt	agctgggatg	aaagtgtaca	ccacaatgcc	cggagaatta	tttcatttct	120
tctttgtaga	gattgagctt	tactctgttg	accaggttgc	tttcgaactc	ctggcctcaa	180
gccatccttc	cacctcagcc	ttcccaagtg	ctaggattac	aggcgcgagc	caacttgccc	240
agccctggaa	tttttgagcc	tgttcaattc	taactattgt	cacccaaaagt	aaccttaagr	300
aaaaaaaaatg	attatctcct	tgtttcattg	caccattaaa	atctttccta	aattttccat	360
gttaaagatg	aagctcaaaa	tcctcagcat	agcatacaaa	acacttcata	atcagatgcc	420
tcttcaaata	cctcctatca	gaatggtctc	tttgactacc	cctttaaaaa	aattcccccc	480
aaccctattt	ttaaattatt	tacttatttt	tatttatatt	ttgatacata	atagatggac	540
acattcaaac	agtgccccc	aaactggggc	agcagaaaca	ggtccttgct	tattttctc	599

<210> 30

<211> 599

<212> DNA

<213> Homo sapiens

<400> 30

agccttagaa	cagaggaagg	acagctatag	caatgaaagg	actagtgcag	attcagaaaa	60
ataagaagac	agaaaccaag	gtgtagtaac	atgttttagt	atggagggga	aggcagttat	120
agaaacttga	attacataat	ttgtacattt	ctgggagata	gaaggtaaag	atagcagcta	180
atggagacag	gacaggactg	gtacttgatt	atggaagaaa	ggaggtaa	agaagagaca	240
aaaagggaga	gaagagatgt	caactgccta	ctctggtagc	ctctgtatcc	aaaaggttgr	300
ctcaaacatt	cgctcataac	tttgtctggc	ttaatcctgc	tcatcccagc	agacttattt	360

caagtgtctc	cacgttttgg	gaagtcacac	ctcacttctc	tgggctttca	tatgggagag	420
catttaattc	tgttgaaaaa	ctatttaata	ctacatctac	ctttctctat	ggactctgag	480
cttcttgagg	gcatgtatca	tgtatgttct	attctgaagc	acccatacct	agaacaaagc	540
ttagcacata	gtaggaactt	aataaatatt	tcggagttga	ataactagcc	ttatgtaat	599

<210> 31
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 31

tttaagtgtc	gaaacctcca	agacaaaaga	gactgtgtct	ttattgttct	ctgaattact	60
cgtacctcag	tcggtacctg	gaacatgata	gggatcccat	agtggtttga	tgaataaatt	120
agtgactcca	agagtaaagt	aatcctcagg	aggacaaagg	cagatagctt	cccttcccta	180
tcagaatgta	cttctcttaa	agcttttctt	ggtataattc	ttggagaatt	ttgccttaca	240
gaagtcaaat	caacatacaa	agtgaaaact	ggatcttcta	caaataatgg	aagaatcaas	300
tctatcaaaa	caacaattat	acatatgata	aatggagggg	ttgtcacgag	ccaggctaag	360
agctttacat	atattatctc	attctgtcta	tgccagagaa	tcaactatga	catatgtaac	420
attaaatctc	attttataga	tgcaaaaact	ggggtgtaaa	gaagtcaaag	aatcagccag	480
aatgtacaga	attagcaaag	gtggaactgg	gatttgaatt	cagacagtct	gactccagac	540
gccatctccg	aattatgcat	aattatattt	caattattaa	cattcataaa	ttgaaatat	599

<210> 32
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 32

tcacatacca	aagtgaaaac	tggatcttct	acaaataatg	gaagaatcaa	ctctatcaaa	60
acaacaatta	tacatatgat	caatggaggg	gttgtcacga	gccaggctaa	gagctttaca	120
tatattatct	cattctgtct	atgccagaga	atcaactatg	acatatgtaa	cattaaatct	180
cattttatag	atgcaaaaac	tggggtgtaa	agaagtcaaa	gaatcagcca	gaatgtacag	240
aattagcaaa	ggtggaactg	ggatttgaat	tcagacagtc	tgactccaga	cgccatctcy	300
gaattatgca	taattatatt	tcaattatta	acattcataa	attgaaatat	gagggataat	360
gtaccttttc	atgaaagctt	tgctcgttgt	gtggatgagt	gtgtgtacat	gtaactgctt	420
atgtgtgcta	tcactgaggt	agaagacatc	tctctctctc	tctctctctg	tctctctctg	480
ttttggtcta	cttttagtaa	gacttgtatt	tgattgagtt	cagaagtttg	attatctttt	540
taactaacct	gtttgtttta	attatattaa	aaattagtca	ctttcaacat	atttgcata	599

<210> 33
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 33

agttcctcag	acctgctttc	taagggcaac	tctaccacct	cagcaagcca	ttgaacttct	60
ctgagctcag	tccttttcatt	tataaaatgg	ggtgacagtg	ctcacatgcc	aggaatacaa	120
agggattgaa	agataaaaaca	cgtaattaag	cacctgttgt	tacacatctg	tcagggaccc	180
caataaggtc	agctgtcttc	ctgttgactt	ctgttcttgg	tggttctcca	agatcatacc	240
ttccatcaac	atttaccgtc	actccccac	cccatgccca	atactgaaca	gtggagggay	300
gcttcaccta	cagttataat	gttgaaaactt	caacccaaag	caagtactgt	taggatctct	360
ggaaactttc	cctcaaataa	gggatttgaa	tgggacaaga	agaagtttta	cagatagcca	420
atggagatga	tttaatgggg	ttatgataga	aacgagaaag	taaaacaaac	ccatgcttta	480
aagtctacca	tttcagggtcc	atattttcgc	ttgaaaattg	agattcctat	taaacaatga	540
catttacacc	aaaaagtaga	ggagttgggt	gaaggacagg	gtaatgccag	gaggaattg	599

<210> 34
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 34

tggtgttaca	catctgtcag	ggaccccaat	aaggtcagct	gtcttctctg	tgacttctgt	60
tcttggtggg	tctccaagat	cataccttcc	atcaacattt	accgtcactc	ccccaccca	120
tgcccaatac	tgaacagtgg	agggacgctt	cacctacagt	tataatgttg	aaacttcaac	180

ccaaagcaag	tactgttagg	atctctggaa	actttccctc	aaataaggga	tttgaatggg	240
acaagaagaa	gttttacaga	tagccaatgg	agatgattta	atggggttat	gatagaaacr	300
agaaagtaaa	acaaacccat	gctttaaagt	ctaccatttc	aggtccatat	tttcgcttga	360
aaattgagat	tcctattaaa	caatgacatt	tacacaaaaa	agtagaggag	ttggttgaag	420
gacagggtaa	tgccaggagg	aattgggaat	ttgagagtca	agtcaaagga	ctgaaatact	480
cagaatacta	agggcacctc	agggctctac	caaggacacg	tagaagcttt	gaatttgacg	540
caccacccta	atttaacgag	ctacctcagc	acgtagtggg	gccttggaaa	acagatgtc	599

<210> 35
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 35

gggttttctaa	gctctctaat	ctccccctgtg	cagctggctt	gctgtatggg	ttatacaaat	60
ccagtgggtga	tctctgtgca	acgtgggtatc	acctgttttaa	agaggctctca	tcttcatttt	120
caaagaggaa	tacatgtttt	tttacttact	cttctgcatg	gctgactcct	tttcatgctt	180
taagtctcaa	tcttaatgcc	acctcctcct	tccagacgtt	cccagctaaa	gtggcacttc	240
ccagcccat	tactctctat	gtttatttgc	tgcatagctc	ttatttgtaa	tgatttcgtr	300
atagtttgat	gatgatcatg	atgaatatta	ctttacctat	ttatggcctc	tcttttagta	360
ttaaattctg	taagccacat	gagcatgggg	acacatctcc	tttgtcactg	ccccattgct	420
ggcattttagc	acaagcatgg	tctataatag	ataccaaaca	aatatgtatt	aatcatgtaa	480
atgactaaat	ccatgaatga	atctatcaga	cagtgtagat	agcagcacat	aaaggaaagg	540
gaatgtagta	aatttttcat	tttccttgaa	gatgtagcta	tgtatttaga	atttgaaaa	599

<210> 36
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 36

tagggaggga	catttagata	tgcatgata	attgctttgc	ttagagaaat	tcaagggtgat	60
agaatgcatg	gtgacacctc	acccagactg	gtagagaaag	ggaattcttc	cactgggaat	120
gacatgatta	tctaaataag	taggctcaat	caggtcagga	aagggcctga	aagactattt	180
caagcagagg	gaagggtatt	gccaaggcca	gggtgtgtag	tggagagaa	gggcagtgcc	240
agagaattat	gaagtgttcc	aatgactaaa	agtaaagtaa	caagttcctt	gtccaagagy	300
ttggattgta	tcctaactga	aatgagtata	cactaagtgt	ttgaagaaga	gggatgaaat	360
ggtcaagttt	tcattacaca	aaaataacct	gttcctttca	ttttatgttt	atttattttt	420
ttaattttct	gactgctcct	ttctggaaat	ctcaaattta	tatttgccaa	atattgtcac	480
attttcgatg	gagaatacaa	actaagaatg	ggttagggaa	ctgagtcaga	aagtccctgt	540
tgtacaattc	aatcatgttt	ttctaaggat	gtgcttttgg	acattatgga	aactatctt	599

<210> 37
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 37

accagaacc	ttaaaaggcaa	gaatatgatt	gaaatgtcaa	atggggactt	ggtgatctaa	60
attatgtccc	ccaaaagcca	atgtcttgcc	accaccagtg	ccctatgggt	ggagtttcta	120
aacagattac	tcaaaacaca	aactttcaaa	aagggaaggt	cataaccctc	tagtcatcag	180
ggcaattacg	gaataacatt	gctggagtaa	ggttttctca	atgccaaga	gatgagctgg	240
caatgccaca	acaatgtcca	attcttagtg	ggccaagac	catgtgttac	atttccctcy	300
catgattact	cacagcttca	cagttctgct	gtcctcttcg	cctctctgcc	acctcttaac	360
tgacaccttg	acctcctacc	cctaagattc	aacctgtgga	gattacttgt	cttttcatct	420
acactctggg	cactctgacc	cccatctctt	agattcagga	ttttctcttt	tcctggctta	480
ctctccagca	caagtagaaa	aatattgtgc	ttcattggaa	aatgcatgtt	gtttgaatca	540
cactctttca	gattatacaa	ttgtagtctt	tcattatctt	tgagccattt	tataatgct	599

<210> 38
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 38

aatagactgt	tgcattttga	atTTTTatatt	tgccataaatt	ttgattttttt	ttcttctctg	60
actactttct	tttgtagtct	catcagctca	gaagggtcaaa	gttgagatac	ttgtcaagac	120
agtcaaagtg	tagaacaatg	agcctatgac	agaagtgagg	catgatatta	attgtacatt	180
aagctttata	ggttttttga	tcatccatcc	ctctgacagt	ggtaatagga	atcctgagtg	240
ctatttatatt	tccttcacag	caaagattat	aaaaagcaag	aggaattaca	tgtgcttctr	300
ccgcatctta	tcagcactca	ggcaagggtt	actaatatgt	ataatcacia	tgacttccaa	360
aggttccacc	ttggtacttc	tgtaatttta	attcaatgta	acagacactt	agtgagcatc	420
taataagtgt	caggcattag	gcctagtgtc	agggatagag	aaatgaatac	acagatgtgg	480
actttgccag	taaaaataag	aagtagagtt	gtgtgcagtg	gatttatatg	ttttaggcta	540
aatatggact	tttcttagaa	gagggattca	gatggcattt	cagttcacat	ttagtacta	599

<210> 39
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 39

agtccaaggg	aaaaccacga	aatgcattta	cagatcagga	atatagaaat	aaatgggcaa	60
atggaggcaa	aactactagc	cttttccaca	atgggggaaga	aaagtcaatt	aaatcacccc	120
tggacacaa	ttacatgtct	atagacaaga	attatatgca	ttaatattag	tttacagcat	180
tataaaatgg	ctcaaagata	atgaaagact	acaattgtat	aatctgaaag	agtgtgattc	240
aaacaacatg	catttttcaa	tgaagcacia	tatttttcta	cttgtgctgg	agagtaagcy	300
aggaaaagag	aaaatcctga	atctaaagaa	tgggggtcag	agtgaccaga	gtgtagatga	360
aaagacaagt	aatctcacag	ggttgaatct	taggggtagg	aggtcaaagg	tgcatgtaag	420
agggtggcaga	gaggcgaaga	ggacagcaga	actgtgaagc	tgtgagtaat	catggggagg	480
aaatgtaaca	catggtcttg	gacccactaa	gaattggaca	ttgttgtggc	attgccagct	540
catctcttgg	gcattgagaa	aaccttactc	cagcaatggt	attccgtaat	tgccctgat	599

<210> 40
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 40

catggtcttg	gacccactaa	gaattggaca	ttgttgtggc	attgccagct	catctcttgg	60
gcattgagaa	aaccttactc	cagcaatggt	attccgtaat	tgccctgatg	actagagggt	120
tatgactttc	cctttttgaa	agtttgtgtt	ttgagtaatc	tgtttagaaa	ctccacccat	180
agggcactgg	tgggtggcaag	acattggcct	ttgggggaca	taatttagat	caccaagtcc	240
ccatttgaca	tttcaatcat	attcttgcct	ttaagggtct	gggtgtatct	gtatcttcay	300
atggtgaagg	aattataaag	cgattttgtg	tccctctctg	gggcacagga	tgttgacagaa	360
ataaagctga	tagaaatagt	cacatgagtt	gaccagatgg	catcatcagc	ctcttgaatc	420
cactacatcc	ctaattgtcat	tagcttggag	gggcagctgc	caaattaggt	ggggctggac	480
attcctttta	acagggtgct	tcttttctag	gatccaagtg	agagcctaag	atagtttcca	540
taatgtccaa	aagcacatcc	ttagaaaaac	atgattgaat	tgtacaacag	ggactttct	599

<210> 41
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 41

ggcagcatta	tgggggaaaa	gcaatgatga	tctaattgaga	tctgataaga	agttagccca	60
aaacaaggaa	attgttgagg	gttctctttg	aagtatggat	ttataccac	caaccttagc	120
tgcgaacctt	acctcaagtg	ttacctgtgc	cttgagatgt	ttcctggtca	tagtactaag	180
ctatcataat	gagcaagaca	ttcaataaag	aagtgtgatg	gctatgagga	cagatcttaa	240
cagggttttt	tttctggaag	gcttaaaatc	atgcattact	caatctaata	cttcacgaam	300
tttcagtaaa	acctaattgat	aatatagaag	cttgtgttgt	agttttgtaa	tcaacagcaa	360
aacataaaa	ttaaaaaaa	catacattac	tggggctgta	tcctgctaca	ataataaggc	420
tgacataata	gatggagAAC	aatatggtaa	caagccaaaa	tgtattactt	catccacaaa	480
tagtatcgtg	ctatatatag	acagacttgt	taaaaattta	aagaaataca	caatcaatta	540
cacaatagaa	aatttgctat	atggtgcatg	gtggcatgca	actacaaatg	tttctaaca	599

<210> 42
 <211> 599

<212> DNA
 <213> Homo sapiens
 <400> 42
 cacacagtat ttcatttagg cctcactatg tccctgatgt aggcattaat atctttatatt 60
 tgcaaatgag aaaacagtct gtaccttgta tgccatgctg ctattgttta tctgtttgaa 120
 tctcaagcaa atctgcttga taattggtac caaaataagc ctttttctgg gtaaggaatc 180
 tgatattgtg ttttaaaaaa cacacattta atcctggggc tgctgcatta ctcctgctgc 240
 cccatcctac tgtgatcaaa ggcacatata tgagatgggtg agttgtcccc ttgccaatgr 300
 gggtttggtg agaaaggaaa gtgcagtact tctttgtttc tgaattgcaa gtatgtgtgg 360
 gtttagagggg gaggctgaat atgaagggtc tgggacagcc caccagggtat cccatgagac 420
 tttgcaaagg aaaaggaggt gagtgcagac ccagggtcca ataggataga aggaaaagcc 480
 aggccatgga gttcctcaga cctgctttct aagggcaact ctaccacctc agcaagccat 540
 tgaacttctc tgagctcagt ccttttcattt ataaaatggg gtgacagtgc tcacatgcc 599

<210> 43
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 43
 ctctataaat gtttgccat tggtttaaaa aacaaataat ggaccaatgg gctcaaaagc 60
 aaactggcta atataaaaat aataataaca accacaataa taaaataagg gctaacactt 120
 aaacggttgt gtactcacta tgtactaggc actgatcaaa gtactttgta catattttct 180
 tatttaatat tgcgtaccat catattacaa tatactgtta ttaaccccaa tgtatagatg 240
 taggtgaaga aacttgctac aaatcatata gctagtgtgc tgagatgcaa tccatgtgak 300
 ttgttcacag agctcaggtt ctgtgaagcg ggtaaaaaca aaatttgga tccagtttca 360
 aaaggagaat tgcaaactaa tagaacatat agcacaaaat gattatatca atagaatgct 420
 aattgcatat caaggatatt tgggtataata caaattattc taccttaaac atatggaaat 480
 ttgtggtcca tgatgttgta gattctatct tcccactctg cattttcaaa ggcatatggt 540
 attgactcat tcgattaatt gttggatagt ctttattata gactaaatca tagaataaaa 599

<210> 44
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 44
 taaataagcc aatatgcttt ctgttgattg attgatttac taaacattga ttggccatct 60
 ccactgggga tatggcattt aagagatctc tttgatctta gtacttttac tgctttttaa 120
 ataggatcaa atacacccaa ggtaaaaaat agaacacact atacgttaca ttttggaact 180
 gttagaaatt ctttgaagc taaaattact gctatcattt gacaactttt acccctaaaa 240
 taatgtggtg ctcaccagct tgcttaagtt acagcacttg ctgtcttctc agatacaatr 300
 tcagaaactt ataatccaag aaaaatctaa atggcaagtg tgagttaatg gaagcctcat 360
 aaagcaagag gtgttttgga agtgtatgga agacatcaat aaatgatatg tataacatca 420
 agtgcaaaag tgtgtgctag gaaagttcaa aaaagaaaaa aaatatggta aggtaagacc 480
 agagattggg gagtatatag cttttgggaa ttcaggaaat gctaacgtct atggagaatt 540
 tgcataagggtg aaagatcaga ttggaagcct ttctctgtgg aagcattgtg gatctaagt 599

<210> 45
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 45
 ttcactattc agggattatt tggaatccct tgtcaccaga agctcttaag gaaataactt 60
 ctacttcgtt gcaaatatgt tcttggttta gttgaggtaa tgcaaatact agaatacttg 120
 tttgtttaac agcttattct tccctgaagc tgttctctca gtccctgccg gtgggatctt 180
 atgtctccag gactacttaa caccctaact agccccatct ttttaagcctc cctgggacct 240
 gccctcgcag tacctcttat acctactcca cttcctcctc atggcctcct gcagaatgcm 300
 attctaaaaat taggttctat tttcctcgcc cgcattctct tttgcaaagc ctccaaaaaa 360
 tttactttgc ttctctgcgc ctgctttatc tctattttct acactcgctc cttctttttc 420
 taattatcta taataggcgt cacaaaattt gcatgtgttg gaaccaaact ttccatgggt 480
 gcctcaaaat atacagatgt aaatttgcac ataattaaat tttgcataag ggaaactctc 540
 atttggggag atatgcaatg cccaataaat ggcagtttcc ttcaatgtcc ccaggccag 599

<210> 46
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 46
 ctccctggaga cataagatcc cactggcagg gactggagga acagcttcag ggaagaataa 60
 gctgttaaac aaacaagtat tctagtattt gcattacctc aactaagcca agaacatatt 120
 tgcaacgaag tagaagttat ttcccttaaga gcttctggtg acaagggatt ccaataaatc 180
 cctgaatagt gaaggattta ctattgcaca tttaaaatgt gtaaaatgtt ttaacattgt 240
 cttacaatga tgtctttatt attatatgtt agtaaattaa atatttactt ataagttgtr 300
 tagattttatt agcaacctca gttcttcac ccactttttt aggttcctga aattaaatta 360
 catagtcttg ttagaaatgc ttgttccccg gtgccataaa gaaatagcac ttgaacataa 420
 atttaatttc ctacagcaagg cagttttttac ctattgcaga aagggtacac ttgccagcag 480
 ttttgccaca agagtacacc gaacaaagga gacagggtca tttataacct gacccatcca 540
 ccttcctgct gtgtccactt tccattggct gaaacgggac ctacacattct gtatttgtc 599

<210> 47
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 47
 agaatgagtc caagggaaaaa cccagaaatg catttacaga tcaggaatat agaaataaat 60
 gggcaaattg aggcaaaact actagccttt tccacaatgg ggaagaaaag tcaattaaat 120
 cacccttgga cacaattttac atgtctatag acaagaatta tatgcattaa tattagttta 180
 cagcattata aaatggctca aagataatga aagactacaa ttgtataatc tgaaagagtg 240
 tgattcaaac aacatgcatt ttccaatgaa gcacaatatt tttctacttg tgctggagar 300
 taagccagga aaagagaaaaa tcctgaatct aaagaatggg ggtagagtg accagagtg 360
 agatgaaaag acaagtaatc tcacaggggt gaactcttag ggtaggaggt caaagggtgca 420
 gttaagaggt ggcagagagg cgaagaggac agcagaactg tgaagctgtg agtaatcatg 480
 ggagggaat gtaacacatg gtcttggacc cactaagaat tggacattgt tgtggcattg 540
 ccagctcatc tcttgggcat tgagaaaacc ttactccagc aatgttattc cgtaattgc 599

<210> 48
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 48
 tgctgacatc ccaattttata aaaataaaaa ggaaacaata tcagcaacat taacttgtca 60
 ctgccctccc attacctcat gaaagccaaa cccacttttc taacaaaccc caaaacttcg 120
 caccacttcc ttccctacag ggacctccca tctgccagtc acccaccctc tgtacagctg 180
 gctaacatac ttttcagtta tcatttctgg attgttattt aatagataat ttccagtctt 240
 ctctccctta tttgagatct taaataagct tctggaatgc aaagtgcagg tctctttty 300
 catagtgacc acaaatacac cgttttattt ttgagcatat aattcttata tatttttgtt 360
 ataaaacagt cacaatgtc cacaatgcat cgtagtatat gaaatgttcc ctgtccatat 420
 tcattaatcc aactaataat tttagaacaa tttctgggtg ccaagcatga ttctctacct 480
 agttgaagggt gtcaacgagg gcctctatga gaagggaaca ttttaagctga gacataaatg 540
 gaagaacaag cagacctgca gtaaaagagg gaacttgtca ggccccatca caggaataa 599

<210> 49
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 49
 acagtttaat tactctgaca gctttttttt tttatttttt ctccctagtga attatcaaca 60
 gaaactacct agagaaacct gtaacttcag atcttctgat taattgattt tctttgggtg 120
 ctgcttgaca gtaacagaat ggtccatggg gacagttttt cagtgaatgt tttcaacaat 180
 gataggaaca cagagctaag cagagtgtca agagacattt agtgagtag agacatctca 240
 tctgtttttc tgcagaactt gcataaaatt agcatctgat tattatcacc tcagaatgar 300
 ccggaacccg taaagatgtc tttgatcgta aaagtgggga ggagaccctg atggaagaag 360
 agaagacttc agcctcacia tagaaaagac tgtgtttatg cacctaaatc taacttagac 420

accatcta	at	tcaagaa	acac	ggcacatt	ta	gggaaata	aa	gaaaagag	gc	agaaagag	at	480
gaagcac	ctt	ggctga	atca	gagaaatt	ag	gggtta	atgt	tccagact	ga	tgacttga	ag	540
gaggggt	gtgc	attttc	atgc	aagggacc	at	tctgtaga	ag	aagcagca	ca	ttcaccct	t	599

<210> 50

<211> 599

<212> DNA

<213> Homo sapiens

<400> 50

tggccca	att	tagaaa	acca	ttttgt	gggg	gtccc	attag	gttcag	cttt	cagttg	cttg	60
aagagca	tgg	actgac	atat	gccacaa	agc	aggtc	agaaa	atgact	tttct	gtcaac	attc	120
atccct	actc	cacgct	cccc	acctct	acaa	catta	acaag	aaaatg	agaa	cgctg	ccca	180
gatgtc	ctgt	tggatg	tcac	tgggt	aaatt	atcttt	tatt	aagaat	ctgt	aattgt	tctt	240
cactgag	ccg	tgacct	aaat	ttagcg	gtgc	tgaag	agtga	gcccc	agcca	acagg	attty	300
cctgaga	aca	cgatga	agtt	cctgt	atcaa	ttcgat	tcac	taacta	gaaa	gaggg	aacaa	360
gcagcc	agct	tttttt	tttct	gcccc	attc	ttcagt	gaat	catttg	tttag	tttgac	caga	420
attttt	gtga	gatcc	agtt	acaatt	ctgt	aagaaaa	aaa	gacag	caaac	ctcatt	tcaaa	480
tacttct	gtgc	gacatta	aatt	ccgagg	caga	aaccaa	atta	catta	agcag	atctga	tgtt	540
cttgct	tatta	aagac	cttgg	gtttact	cac	tgagt	tttt	cagcct	gagg	agaaga	aat	599

<210> 51

<211> 599

<212> DNA

<213> Homo sapiens

<400> 51

tctgact	gct	caaaa	atgcc	tttgga	gctc	taaaa	attta	tacatt	cttg	ggccc	gacct	60
gattcag	tag	ttctag	tga	aagct	cagga	atctgc	ataa	aacttt	caga	gaatt	ctgat	120
atgttag	ctc	agttga	ggaa	cctct	actct	tgtct	cccat	agtcc	agtt	ccctg	ttttg	180
ttgtcat	att	tagcct	tgt	gggtg	tgcgc	ttcac	ctgac	agta	agctgg	tga	aggctgt	240
gttgcag	agg	atgag	cctgg	aga	acactgg	ctttt	cagtt	ggctg	atgtt	taa	attccam	300
ctttgg	tgt	ctagg	agg	tggg	actagg	gcact	cttggc	tgttt	cattt	tcct	catctt	360
taaatg	ggga	gaata	atatt	ttct	tataatt	accact	tttac	tactaa	atca	aggct	accag	420
ttactaa	att	tttgta	aatgg	ttct	aggata	agata	gtaag	ggct	tgaaag	aga	aggaaag	480
aacagag	cca	ttagat	gaaa	gaatg	gaaag	gttat	gctgt	cctct	cctca	ctttt	atttga	540
atgctga	ctg	tcgtag	gtct	tatga	gtagc	ggta	acacta	tattat	ctca	tttatt	ctt	599

<210> 52

<211> 599

<212> DNA

<213> Homo sapiens

<400> 52

tgtctt	cttg	ggaagg	ctta	acaca	agaga	gcagg	tggct	tcacg	ggcct	ttcatt	ttggg	60
taatga	gttg	gaacac	ttac	ccaa	agccaa	gcca	aggaa	ctccc	agctg	ctcc	acactg	120
gaagtt	gaca	cacag	taaca	tatat	gcata	tgtgt	gtgtg	ttgag	atggt	ggggg	gatca	180
tttttt	tgt	acaat	caaca	atctt	gtaac	tgttt	ctctg	ctatg	ttata	agcc	aggata	240
agagct	tcct	ggatc	acttg	atctc	ttttt	gtgtt	ggtaa	aaata	aatcaa	gcct	gtctty	300
aggca	attag	ccacct	cttc	aagcc	atcaa	ccttat	gcac	aatat	ccaag	gaact	cctcc	360
ttgcct	tttc	tgaac	tactg	tggga	agtga	ttgga	aggag	acagc	caaag	gcca	acaata	420
taggg	cgagg	cccct	gggca	tcctt	ctcag	ctctg	gggaa	ggtta	atgaa	ttgt	acttgg	480
ggaaca	aaaga	gaact	ttgac	aacag	agatt	ctgac	agttt	cata	aggcag	gaaat	gtggc	540
cactat	caac	ccagag	ctgg	gtgtg	caaac	acaat	ctatg	tgttt	atatata	cctt	ctcct	599

<210> 53

<211> 599

<212> DNA

<213> Homo sapiens

<400> 53

tcattct	tgt	gttac	ataat	aaggc	actga	ttagat	gctt	taggt	ttttga	attttt	tcct	60
tacc	cgattg	acca	actctc	taagt	ctcag	tttct	ccatc	cata	aaatag	aagta	ataat	120
gcctg	catgt	agagt	tgt	aaat	atataa	tgag	aaaaat	attgt	gaggt	acc	cagcata	180
gtagc	acca	cacag	tagac	aata	acctat	tttact	ttatc	tacct	tccaa	tcct	taagct	240

aactatctct	tagtcccaag	aataaaatct	caggattggg	aaaaggcttc	ttcattttcts	300
tttaggggccc	ccagatcatt	gagattcttc	tttcctatta	actaccaggt	tatggatttg	360
aaagaatctc	agttagaat	ataatagcat	cattctagac	agaagaatgg	aaaactagtg	420
ttctaaactg	gatcaaagg	tgtcttgtga	aataatgcag	tcattgttct	ctatagtttt	480
gctccaatag	gtcaaggata	tactagtata	tcaggggtag	ctatggatcc	tgggatttcc	540
cctttaccac	aaaggagatg	atgtaggtgt	gaaattccag	gcaccttgcc	catcttttg	599

<210> 54
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 54

gccacaacaa	tgtccaattc	ttagtgggtc	caagaccatg	tgttacattt	ccctcccatg	60
attactcaca	gcttcacagt	tctgctgtcc	tcttcgcctc	tctgccacct	cttaactgca	120
cctttgacct	cctaccccta	agattcaacc	ctgtgagatt	acttgtcttt	tcactctacac	180
tctggtcact	ctgaccccca	ttcttttagat	tcaggatttt	ctcttttcct	ggcttactct	240
ccagcacaag	tagaaaaata	ttgtgcttca	ttggaaaatg	catgttggtt	gaatcacacy	300
ctttcagatt	atacaattgt	agtctttcat	tatctttgag	ccattttata	atgctgtaaa	360
ctaataattaa	tgcataata	tcttgtctat	agacatgtaa	atttgtgtcca	ggggtgattt	420
aattgacttt	tcttcccat	tgtggaaaag	gctagtagtt	ttgcctccat	ttgcccat	480
atctctatat	tctgatctg	taaatgcatt	tctgggtttt	cccttggtact	cattctaaca	540
tctcttcttt	ttcctcatta	attaattaaa	ttaaattctt	aacacctctt	catattttt	599

<210> 55
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 55

atcacagcag	tgactctttc	aaatctggcc	taactgaagc	tagcaccaca	ccaggatctc	60
tgctgggcac	acaccaatca	cccaggaggt	cagtatcatc	cccattttac	agatttgaaa	120
gctgaggcac	aaggtaata	acagttatgc	agaagtcttg	cttcagtgtc	taacttcctc	180
atctcacttt	attctgtttt	tcaataggag	aagagatata	atctcatgat	gaaaagtgcc	240
atctaaagtg	gcaccttaga	cacagtaagc	aaactggatt	ggaagtgaag	aagtcagtcy	300
caagatactt	gacaatgtct	cctatttgta	gattgttttag	caaaatgctt	tcacctgagt	360
tatttcattt	gttgctcaca	accaatctgt	ctggtaggaa	aggcagatga	tatgttcacc	420
tccagataag	aaaaatgaaa	ttcagagagg	ccaaggggct	tgctcaagat	tccacagtga	480
agaatctact	tgcagggaat	ttaacaaagg	ttttcaaatt	cagaagtcca	caggatgact	540
gcccttgaaa	tccagtggta	acataataa	aggtgtcttt	gagcaggtaa	gagaggacc	599

<210> 56
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 56

cagtttgaaa	atggacaaaa	aactgaaata	gatattttctc	aaaagaagac	atacaaatgg	60
ccaataggat	attttttaaa	tgttactagt	catcaaggaa	atgcaaatca	aatgacaat	120
gaactatcac	cttaccttg	ttagaatggt	tactagcaaa	aaaagacaag	ggataacaag	180
gttggaatg	atgtagagaa	aagggaatcc	ttgtacattg	ttggagggaa	tgtaaattag	240
tatagtcact	atggaaaact	gcatggagga	gcttcaaaaa	actgaaaata	agcctaccay	300
gtgatcctaa	tactgggtat	atatccaaag	gattggaaat	caatatgttg	aagagatatt	360
tgcatccca	tggtcgctgc	agccttattc	acaattgcca	agtatgaaat	tggttgagt	420
gtccatcaac	agatgaatgg	ctatagaaaa	catatacaca	gtggaatact	attcagcctt	480
aaaaaagaag	gcaatcctgt	catttgcaac	aacatgatga	acctgtagga	cattgtgctg	540
agtaaaataa	gcctgtcaca	gaaagacaaa	tactgtataa	tctcatatgc	agaatcttc	599

<210> 57
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 57

gactttccag	caccttaaac	tacttggaag	gatagattaa	tgtctagatg	aaacttgatg	60
------------	------------	------------	------------	------------	------------	----

agactttaga	ctaatatggt	acatatatat	caggagatgc	atgtataaac	caaatccaaa	120
taacccaaag	caaaatttct	tttagaaata	gtgataaata	aatgagttag	gagtttgtca	180
ctcacatccc	ccaggtaaaa	catacctttt	tagcctaaat	aaatgctata	gtttgaatgt	240
gttcctcca	aaattcaggt	actgatagtt	aacagcagcc	gatatgatgg	tattacacam	300
aatagtaggt	tcttttaggag	gttggttaggc	catgagtgcc	tcccccaaga	atgggatgaa	360
ggacctcata	aaagaggctt	ctcacagcat	tgtgacctct	agccctccca	ccttccataa	420
gtgaggacac	agtgttcctc	ccctctggag	gatgcagcaa	caaagtgtca	tcttgggaagc	480
agaggagcca	tcaccagaca	acagaaccag	ccaacagctt	gatcctggac	atctcatttt	540
ctccagaact	gggataaaat	aaattcctgt	tttttaataa	atttcccaat	ttcagggtat	599

<210> 58
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 58

gatgtggaaa	gcatagacag	aaggcactgc	agaagaggag	agtaaatgac	tttaggtcat	60
tgaactggtt	ttttagataga	atcaagttaa	gaaaaataga	agtcaatagt	actaagtttg	120
aactttatct	ttctaaaaac	tgtttctgtg	ggcagagttc	atatgctgga	ggaatgaatc	180
ataatgacac	cctttggtat	gaccctgggg	agaataaatg	tgacttagtt	gaggtttctg	240
aggatgacat	tcagaaagaa	aacaagagag	ggaaataccc	tttctctctg	tcacactcty	300
aggttagaaa	ctcacttttag	gggccacagc	ctatgacaaa	tacactcatg	caaattctgt	360
gcttgcttcc	atgctccatt	tctttgccaa	ctatgtagct	aagcaaaatg	gaacaagaaa	420
tagtgtcttg	ggagaaataa	aattgctgga	ttagatacag	gccaccttta	tataagaaaa	480
aacgttaaata	tatgtactctg	gtgaaataag	tccatttggtg	ttattgtgga	attttatgtg	540
agactgagat	agctgattaa	gatgggtgatg	cttgaggtatt	atcatttttag	taataaata	599

<210> 59
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 59

ccctgggaaa	atcacgctc	tttgtgccta	aatgtcttca	tctacaaaag	gtggataaat	60
aatatattatc	tcatgaggtg	tttggagatt	taattaacac	ttgcaaagca	ctttgaatag	120
gagatgaaag	gccctgttaa	aggaactatt	agtgtcagca	cctagcgctc	tttggccact	180
gaggaaattt	aatcaggatg	gggaagtgcg	tgcaggcccc	tcctaggcca	tcctctcctc	240
caccagaagc	catattggcc	tttcattatc	ctctaagcct	ggaattctct	ccaccaagcy	300
gctggttatg	tcatttccct	gcctacaatt	ccctctcttc	aattttctcc	ctcctctttt	360
ccacgtatta	aatcctccag	gaccgcttca	agccccatct	gctctgcaaa	gtcttcctta	420
accattcatt	gactgattca	ttcatttact	tattcatttg	acaaggcttt	ggtgagagcc	480
tcctctgtcc	caggctccac	gaagagccct	gagggtgctg	taaattgcat	attcttgacc	540
tcgtgtgccc	gaaggggaga	ctggagctca	gacaggctac	aacgcgtcca	aggtcacac	599

<210> 60
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 60

aataatctcc	cttactcaaa	accaagtgat	tatggacttt	aatcacatct	ataaaatctc	60
attatagtaa	cacctaaatt	agtgtttgaa	taactgagag	ttgtaactga	tatggtttg	120
ctatgtcccc	acccaaatct	catcttgaat	tatagttctt	ataatcccca	tgtgttggtg	180
gaggaaccag	gtggagataa	ttgaatcata	ggggcagttt	ctccattctt	gttctcatga	240
tagtgagtta	gttctcagga	gatctcatgg	ttttataaag	ggcttccctc	tttgctccay	300
tttcattctc	cttcttcctg	ctgccatgtg	aagaaggaca	tgtttgcttc	cccttctgcc	360
atgattgtaa	gtttcctgcg	gcctccccag	ccacgctgaa	ctgtgagtca	attaaatctc	420
ttttctgtat	aaattaccca	gtcttgggca	gttctttata	gcagcatgag	aatggactaa	480
tacagtaact	ttaccaagtg	gacacataaa	actgatcatt	acaatgtaca	gtgaatattt	540
ggtgagttaa	tagatatatt	cataactgaa	tgaaagagga	tggtgattcc	tacttcagg	599

<210> 61
 <211> 599
 <212> DNA

```

<213> Homo sapiens
<400> 61
ccaggctcca aaactgtgag aaataagttt ctgttggtta taagctacca ggtttaaggt      60
atTTTgtaca gcaacccaaa gAgTctgaga ccataatgaa gccattggaa tggTgggaag    120
gcaacttcat gtgagtaact acagtaaagc caggTgctgg taacagtcAt gTtgcccata    180
gagcagatcc tactattaca gtgcctagca cattacctgc atatgatgat atgtgatcaa    240
ttagttaact gattagttta tgaatcagtc tgccaaaaaac tagggcagaa attgatagcr    300
cattaaaata aatatgcctt aaagtttgca aggagaccct attaaactgcg cactgttttc    360
TTTTtatttt cTTTTTTTTT cTTTTgagac agggTctcac tctgtcaccc aggctggagt    420
gcaatggcac agtcttggtc cattgcaacc tccacttccc gggTtcaagc gattcttgtg    480
cctcagcctc ccaagtagtt ggaatacag gtgtgcacca ccacacctgg ctaattttta    540
tatttttagt agataggggg tttcccatat tagccagcct ggtcctgaac tcctggcct    599

```

<210> 62

<211> 599

<212> DNA

<213> Homo sapiens

<400> 62

```

acacacgtgc atacacataa ataccttgat aagttactag agaaagcaga aaaatgtctg      60
acagtttaat gagatttggg tgaaagaaaa ttctatattt cattgttttc caggcactag    120
aaataattca tcaatgtttc taagactcat tcagcgtggc tgcatttttt aaaatatttt    180
cataaatttt gaggagcaaa taccattatt aggactaaa aaggTtgaag tctaatagat    240
tagccgcttc atcctccttc actcagctca gcattcgTtc aactggctct tactggTtaw    300
catccacacg cctcctgact ggctactcag tgccgatgac atttccttca cacacagggc    360
tggttttaag atacattgag gtgacatcag gtggcctgta aagtggTcat tttaggatat    420
cctattcaaa gacatctgtg gaagtgtgga ccaatttatt gatgaataac agtgaagggg    480
tttccaccag caagtaacat aattttttac aatgatgatg ctgaagtaga aagagtttct    540
agtCagggac tggacaaatc aatttgCaga cgatttttag gaagaaaaac attgcaaca    599

```

<210> 63

<211> 599

<212> DNA

<213> Homo sapiens

<400> 63

```

atcccttgTc accagaagct cttaaggaaa taactttctac ttcgTtgcaa atatgtttctt      60
ggcttagttg aggtaatgca aatactagaa tacttgTttg tttaacagct tattcttccc    120
tgaagctgtt cctccagTcc ctgccagTgg gatcttatgt ctccaggagt acttaacacc    180
cctaatagcc ccatctttta agcctccctg ggacctgccc tcgcagTacc tcttatacct    240
actccacttc ctctcatgTg cctcctgcag aatgccattc taaaattagg ttctatttttY    300
ctcgcccgca ttctcttttg caaagcctcc aaaaaattta ctttgcttct ctgcgcctgc    360
tttatctcta ttttctacac tcgctccttc tttttctaata tatctataat aggcgtcaca    420
aaatttgcat ttgttggaac caaaatttcc atggTtgccT caaaatatac agatgtaaat    480
ttgcatataa ttaaattttg cataagggaa actctcattt ggggagatat gcaatgccca    540
ataaatggca gtttccttca atgtccccag gccagcctcc cagtctgtgt gtttcccc    599

```

<210> 64

<211> 599

<212> DNA

<213> Homo sapiens

<400> 64

```

gccgggcacg gtggtcatg cctgtaatcc cagcactttg ggaggctgag gtgggtggat      60
cacgagatca agagattgag accatcctgg ccaacatggT aaaacctcgt ctctactaaa    120
aatacaaaaa attagctggg catggtggca cacacctgta gtcccagctg ctcgggaggc    180
tgaggtggga gaatcacttg aacctgggag gcagaggTtg cattgagTtg aaatcacgcc    240
actgcactcc agcctggTga cagagcaaga ctccatctca aaaacaaaca aaaattatar    300
tatgaaatag gcattaaaat attgtgtatt ttagaggaga ctgaggattg gaggctgaag    360
aattactcta aattaatcag cttgtgtact tcagagctaa gatagctctt tgggttctaa    420
attctgtgat cttctttttg atttctcttg gagcaataat gaaggcaaaa catcaataaa    480
cataacaaac tgggtaaggg agaccattga gaaggactaa ggacaccttc aaagtTctga    540
gtgagTttta aaagaagaat gatgaaaact ttgatagaaa taggaaaaaa agtagagga    599

```

```

<210> 65
<211> 599
<212> DNA
<213> Homo sapiens
<400> 65
atagaaaatt tgctatatgg tgcattggtgg catgcaacta caaatgtttc taacatgttt 60
ctcttcatag gattttctga attttcattt aatattcaag cacatcaaaa acaccctttc 120
aggtgtgatc ctatacagca aagctgtcct cacaaacaat agttgactaa ataaacacat 180
ggctttatgg aagaaatgtg taagtatagc cattgttgga gcagatgctc tgcttaaaaa 240
gaaaaaaaaa aagttaaagt tatagatctc accatgctta tttactgctt aagtcatags 300
caatttattg caccaaagtt gaagttcaaa gcataaagaa tactatatat aatgcaatta 360
atgagggttga tgtccctaaa acaagagaga attagtaaat gttttacaat agttttcatg 420
agatgggaaa tgacaataga aatctttgtt acaatgcaga ttttattgtg gaaatgatct 480
catggcaaag tttttaaaga ggctgcagaa caataaagag agataacaca ttttgctttt 540
atgaaaaagc cgatgttcca aatttgctga ctttctctgt aaataagaag tgactgtca 599

```

```

<210> 66
<211> 599
<212> DNA
<213> Homo sapiens
<400> 66
tcaaattctgc caatcagata ccagtaataa gccatctcct tagcataatg taagacacat 60
tactgatta tctaacttgt ggcaggcatt gtcctgggtc ctgaggaaga aaaattacaa 120
atttacaaat ctcatagttt acaaagcatt ttactgaagc cagatagacc tcaggataaa 180
tagcaaactt catctgagtt ggccataggg agtctgggca aaagaaacac tcttgatata 240
gaaagaaaga aaagatgggt tttctgaaag aaagaagaag cttgtgcagt atttggcagm 300
cccagccaac agcacaagag ggaagggaga aagagcacct cattcccaat tgtagggcaa 360
gaacagattc aatacagcta cataaggagc caagactgaa ttagatgtgt agggaggatg 420
tcctgttacc tgagcccaaa gtttaacacc ctttgatcac tttccgtcta gagagatcgc 480
tcagcaaagg tcattgggtt tgataattaa agtggttttc taaatagaag caatttcaag 540
aaggcaggtt gacctttggg agaatttcta gcagcaatca cttcatgatt gattggaaa 599

```

```

<210> 67
<211> 599
<212> DNA
<213> Homo sapiens
<400> 67
ataggattcc ccatttgggc tgcagaggag ccagaaacat gcttttagttt cttaatcctg 60
aaggagtggg ggatttgtga agcctcattg gtgccctttt gaaccacatt tcatcaagat 120
atttgtaaat ggggagcagc aggcagcctt ctctggccaa catgtatcag ggcctcgtgc 180
ttagtgggat gtgccagag tagacaggga atccctggca gccttgcata gtccctctca 240
tgcctttgct ttcattgggc tggctggaaa caggccggta agcccatgg aactaccggr 300
acaaattacc aagcaatctc ccatggccat gtagaacagc catgtagaac aggatgtctc 360
ccaaggtaag atttaagggt tgtgggtttgc agaaaagaga agggaggaaa taaaaaagca 420
ctgacaatca ttgaaagccc acgttctagg taactggcta ctcaacttca cagcccagta 480
acaattttgc ttaatcctat acatctataa ataggattat tgtcttcatt ttccagataa 540
ggaaacaaga cttggcaaat gtgtagtgtg gtgtttgaca gtattatgta gttgttttag 599

```

```

<210> 68
<211> 599
<212> DNA
<213> Homo sapiens
<400> 68
ccagaaaggg aagaggaagg acaggcaata gccaaaggact cctggcagtg aactcatgtc 60
cacatcaaga tctaattgagc ttgcaactca ctcatcttcta gctctgcctt ggaagctgga 120
gctcctgcac tgactatcaa tgtgagcccc tgagtaggag cagcttggtg gagttgaaag 180
accattgatc tgggtcaaca gactctggtt cctgtctcag cagtgtctga atcaatccaa 240
gtcaaatatc atctctggga cttaatttgc taaattttaa atgaaaagaa aaacaaaaay 300
agaacaatta gactagatca ggattcggca aaccaaagcc tgcttatcaa atctggcata 360
ccacctgttt ctgtctatac aattgtattg aaactcagcc acactcattc atttgcatat 420
tgtccatgaa agaagctttt gcgctgcctc aggagatctg agtagtggcc acagagatgt 480

```

tgcagtggac	catgtttgcaa	cattgttccaa	aatacctaaa	atattttactt	cttgtttttgg	540
agagttttgct	gactgggcacc	agagaaaatct	atgggtctaaa	atcatctaaa	aattttaagc	599

<210> 69
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 69

cccttgccaa	tgaggggtttg	gtaagaaagg	aaagtgcagt	acttctttgt	ttctgaattg	60
caagtatgtg	tgggttagag	ggggaggctg	aatatgaagg	tcctgggaca	gcccaccagg	120
tatcccatga	gactttgcaa	aggaaaagga	ggtgagtgac	agcccagggt	ccaataggat	180
agaaggaaaa	gccaggccat	ggagttcctc	agacctgctt	tctaagggca	actctaccac	240
ctcagcaagc	cattgaactt	ctctgagctc	agtcctttca	tttataaaat	ggggtgacar	300
tgctcacatg	ccaggaatac	aaagggattg	aaagataaaa	cacgtaatta	agcacctggt	360
gttacacatc	tgtcagggac	cccaataagg	tcagctgtct	tcctgttgac	ttctgttctt	420
ggtggtttct	caagatcata	ccttccatga	acattttaccg	tcactccccc	accccatgcc	480
caatactgaa	cagtggaggg	acgcttcacc	tacagttata	atggtgaaac	ttcaacccaa	540
agcaagtact	gttaggatct	ctggaaactt	tcctctcaaat	aagggatttg	aatgggaca	599

<210> 70
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 70

cacttgatg	cattttttaa	atgctaattg	taataacagt	tcgggacttc	taacttctat	60
atttaagcaa	caaataaata	aattgtcaga	tggtacttca	tcctccttct	ctcccatctt	120
cttagaaata	taaattgctt	taggtgggaa	tgctataatt	ttagaccaga	aaatacatgc	180
cagatgtctc	ttatatgaag	ccgtcccgc	caaggatata	tatatgcctt	agtcattagg	240
atgtgttcta	aataatactg	caaagccctt	ggaaggatgg	gtctgaacac	tcacttatay	300
ttaactgctg	gcatgttgct	ttgtccctgt	gtcttgtgct	actatttcca	ttgatgtaaa	360
ggaagcacca	attaaataac	actccattat	tagagaacca	ggcacaagtc	agctgaggca	420
ggagaccgcg	cttcttttcc	agaaacaatg	taaagcctgg	gtgggtgagg	gtctctgggc	480
ttccgcctgt	ccttgctttt	gacattctcc	agcacaccct	ataaacatgt	ctaaggctgt	540
cctgttttagt	ctgattattc	aaactatatt	gtccagggta	gagcaaaggg	aaacctagc	599

<210> 71
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 71

agggagagaa	gtgcctccca	cagcaccacg	accagaaagg	gaagaggaag	gacaggcaat	60
agccaaggac	tcctggcagt	gaactcatgt	ccacatcaag	atctaagag	cttgcactca	120
actcatttct	agctctgcct	tggaaagctg	agctcctgca	ctgactatca	atgtgagccc	180
ctgagtagga	gcagcttggt	agagttgaaa	gaccattgat	ctgggtcaac	agactctggt	240
tcctgtctca	gcagtgtgtg	aatcaatcca	agtcaaatat	catctctggg	acttaattty	300
ctaaatttaa	aatgaaaaga	aaaacaaaaa	tagaacaatt	agactagatc	aggattcggc	360
aaaccaaagc	ctgcttatca	aatctggcat	accacctggt	tctgtctata	caattgtatt	420
gaaactcagc	cacactcatt	cattttgcata	ttgtccatga	aagaagcttt	tgcgctgcct	480
caggagatct	gagtagtggc	cacagagatg	ttgcagtgga	ccatgttgca	acattgtcca	540
aaatacctaa	aatattttact	tcttgttttg	gagagtttgc	tgactggcac	cagagaaat	599

<210> 72
 <211> 599
 <212> DNA
 <213> Homo sapiens
 <400> 72

gattgcgaac	ttatcttcat	ttcaccttta	tgatgttata	ccttttcatt	tttgtcttct	60
catagctagg	gtctcctggt	ccccaaatgt	agacacacat	cttacaccaa	tcccagagcc	120
attttgtata	agagccacca	tggatttaac	cagcttttagc	tccagtattt	gaacataatg	180
ttcagcatca	tcacctggcc	accaaataca	aactgagcac	cctttaatcc	atcaacaagt	240
tctctgcagc	catgcaaggt	tatgaaatgg	gcacagacat	caatatacag	tctttgtgtk	300

taagaggttc	atggtctacc	tgagaaatgc	atcttttaaac	ctaaagtaga	cgctctgttt	360
attccataaa	tgatttttaa	gcatcaatgg	tatatcaagc	actgtactgg	cttctgggct	420
ataataaata	tataaagacc	acaagtttga	atttcatgac	attgaactat	aatgtttaaa	480
tgttataata	atcatagtaa	atgtccttga	ggagctacgg	aagattcctg	catgaagcag	540
aaacaagaag	ctgaagaaaa	aacaactggc	tttgggggct	atataaatat	aaccctcaa	599

<210> 73

<211> 599

<212> DNA

<213> Homo sapiens

<400> 73

ccacatattt	cagaggcaca	taactcaagc	ttgcaacacg	tattcaaaaag	agaccagcta	60
cacttggtag	agacagccat	aggaaaagtga	aatgacccta	gggttttagta	aagccagctg	120
tttcacttc	tgaaaataat	aaaatgaaat	aataaaaataa	atttaaaatg	atacaaagtt	180
caaagttaa	caaatacatt	tgaagccatt	tgcaacaaat	acatctgaag	ctaattgctg	240
gctctagaaa	gtgtggggtc	ttgtttgtgg	agcagtgtta	atgatttagc	attacttatm	300
tctggcaa	ggtatTTTTG	agataacatg	ttatggaaga	aagtgaactg	aacttggaag	360
tttgaagatc	tcgattgaag	tatcatttct	gcctcaacta	cttgcattaa	cttgtacaag	420
tcattcaacc	gctctgaaca	taatggaaaa	atgggatgag	aatacatgtt	gtatactctc	480
caaagacagg	gagactgctg	atataagagg	gcacttttag	taactgatgg	agcaaaatgt	540
tgttatatga	gtgtcagcat	agggccctgg	gcttacaacg	gtgccatgag	ccttagaac	599