

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
22 January 2009 (22.01.2009)

PCT

(10) International Publication Number
WO 2009/011911 A2

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

(74) Agents: **CORLESS, Peter, F.** et al.; Edwards Angell
Palmer & Dodge LLP, P.O. Box 55874, Boston, MA 02205
(US).

(21) International Application Number:
PCT/US2008/008798

(81) Designated States (*unless otherwise indicated, for every
kind of national protection available*): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA,
CH, 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, ST, SV, SY, TJ,
TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM,
ZW.

(22) International Filing Date: 18 July 2008 (18.07.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/959,954 18 July 2007 (18.07.2007) US

(71) Applicant (*for all designated States except US*): **THE
JOHNS HOPKINS UNIVERSITY** [US/US]; 3400 N.
Charles Street, Baltimore, MD 21218 (US).

(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, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL,
NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG,
CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **LUO, Jun** [CN/US];
6431 Western Star Run, Clarksville, MD 21029 (US).
ISAACS, William [US/US]; 2618 Butler Road, Glyndon,
MD 21071 (US). **DUNN, Thomas** [US/US]; 6282 Chin-
quapin Parkway, Baltimore, MD 21239 (US). **HU, Rong**
[CN/US]; 6 Bexleigh Ct., Apartment 303, Parkville, MD
21234 (US).

Published:

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

(54) Title: METHODS OF DETECTING PROSTATE CANCER

(57) Abstract: The present invention relates to methods of detecting the presence or absence of a nucleic acid segment in the (Phosphoserine phosphatase-like) *PSPHL* gene locus of a subject. The invention also relates to methods of identifying a subject at risk for developing prostate cancer, and methods of determining the prognosis of a subject with prostate cancer. The invention features kits for use in identifying a subject at risk for developing prostate cancer.



WO 2009/011911 A2

5 **METHODS OF DETECTING PROSTATE CANCER**

RELATED APPLICATIONS

This application claims priority from U.S. Provisional Application No. 60/959,954 filed July 18, 2007, and which is incorporated herein by reference in its entirety.

BACKGROUND

10 The incidence of prostate cancer has increased in recent years. On January 1, 2005, in the United States there were approximately 2,106,499 men alive who had a history of cancer of the prostate. This includes any person alive on January 1, 2005 who had been diagnosed
15 with cancer of the prostate at any point prior to January 1, 2005 and includes persons with active disease and those who are cured of their disease.

 It is estimated that 186,320 men will be diagnosed with and 28,660 men will die of cancer of the prostate in 2008 (http://seer.cancer.gov/csr/1975_2005/results_single/sect_01_table.01.pdf). From 2001-2005, the median age at diagnosis for cancer of the
20 prostate was 68 years of age, and in the US from 2001-2005, the median age at death for cancer of the prostate was 80 years of age (http://seer.cancer.gov/csr/1975_2005/results_single/sect_01_table.13_2pgs.pdf). The incidence of prostate cancer has been found to be population specific. The age-adjusted death rate was 26.7 per 100,000 men per year, based on patients who died in 2001-2005 in the US. Among White men, the rate
25 was 24.6 per 100,000 men; among Black men the rate was 59.4 per 100,000 men; among Asian/Pacific Islanders, the rate was 11.0 per 100,000 men; among American Indian/ Alaskan Natives, the rate was 21.1 per 100,000 men; and among Hispanics, the rate was 20.6 per 100,000 men.

 According to the National Cancer Institute, based on rates from 2003-2005, 15.78%
30 of men born today will be diagnosed with cancer of the prostate at some time during their lifetime. This number can also be expressed as 1 in 6 men will be diagnosed with cancer of the prostate during their lifetime. These statistics are called the lifetime risk of developing cancer. Lifetime risk may also be discussed in terms of the probability of developing or of dying from cancer. Based on cancer rates from 2003 to 2005, it was estimated that men had
35 about a 44 percent chance of developing cancer in their lifetimes, while women had about a

37 percent chance. of developing cancer. When looking at the probability of developing cancer of the prostate between two age groups, for example, 8.04% of men will develop cancer of the prostate between their 50th and 70th birthdays. (information available on the world wide web at seer.cancer.gov/statfacts/html/prost.html).

5 Prostate cancer is a latent disease. Many men carry prostate cancer cells without overt signs of disease. The progression of the disease usually, goes from a well-defined mass within the prostate to a breakdown and invasion of the lateral margins of the prostate, followed by metastasis to regional lymph nodes, and metastasis to the bone marrow. Cancer metastasis to bone is common and often associated with uncontrollable pain. Autopsies of individuals
10 dying of other causes show prostate cancer cells in 30% of men at age 50 and in 60% of men at age 80. Furthermore, prostate cancer can take up to 10 years to kill a patient after the initial diagnosis.

Current microarray technologies and the sequencing of the human genome have significantly enhanced the potential for investigations in all fields and particularly in the area
15 of cancer research. High-throughput gene expression profiling technologies offer an opportunity to uncover critical molecular events in the development and progression of various cancers and can be used to design improved prognostic testing and effective treatment strategies. High-density tissue microarrays (TMA) are useful for profiling protein expression in a large number of samples (Rubin M.A. et al., Am J Surg Pathol. 2002 Mar;26(3):312-9),
20 and previous transcriptome analyses in various malignancies have provided valuable information for the assessment of patient group classifications such as subgroups of patients that are likely to respond to a particular therapy (Sondak, V.K. Adjuvant therapy for melanoma. Cancer J7 Suppl 1, S24-7. (2001)). Particularly, in prostate cancer, microarray analysis may provide a useful way to examine a large number of clinical samples for putative
25 prostate cancer biomarkers.

Prostate cancer is typically diagnosed with biopsy examination following a digital rectal exam and/or prostate specific antigen (PSA) screening. An elevated serum PSA level can indicate the presence of PCA. PSA is used as a screening marker for prostate cancer because it is secreted only by prostate cells. A healthy prostate will release a stable amount--
30 typically below 4 nanograms per milliliter into the circulation, or a serum PSA reading of "4" or less--whereas cancer cells release escalating amounts that correspond with the severity of the cancer. A level between 4 and 10 may raise a doctor's suspicion that a patient has prostate cancer, while amounts above 50 may show that the tumor has spread elsewhere in the body.

When PSA or digital tests indicate a strong likelihood that cancer is present, a transrectal ultrasound (TRUS) may be used to map the prostate and show any suspicious areas. Biopsies of various sectors of the prostate are used to determine if prostate cancer is present. Treatment options depend on the stage, grade, and other clinical variables of the cancer. Men
5 with a 10-year life expectancy or less who have a low Gleason number and whose tumor has not spread beyond the prostate are often treated with watchful waiting (no treatment). Treatment options for more aggressive cancers include surgical treatments such as radical prostatectomy (RP), in which the prostate is completely removed (often with nerve sparing techniques to preserve potency and urinary functions) and radiation, applied through an
10 external beam that directs the dose to the prostate from outside the body or via low-dose radioactive seeds that are implanted within the prostate to kill cancer cells locally. For more aggressive prostate cancers, anti-androgen hormone therapy is also used, alone or in conjunction with surgery or radiation. Hormone therapy uses luteinizing hormone-releasing hormones (LH-RH) analogs, which block the pituitary from producing hormones that
15 stimulate testicular testosterone production, or by surgical removal of the testis, alone or in combination with chemical (anti-androgens) that block androgenic signaling.

While surgical and hormonal treatments are often effective for localized PCA, advanced disease remains essentially incurable. Androgen ablation is the most common therapy for advanced PCA, leading to massive apoptosis of androgen-dependent malignant
20 cells and temporary tumor regression. In most cases, however, the tumor reemerges with a vengeance and can proliferate independent of androgen signals.

The advent of prostate specific antigen (PSA) screening has led to earlier detection of PCA and significantly reduced PCA-associated fatalities. However, the impact of PSA screening on cancer-specific mortality is still unknown pending the results of prospective
25 randomized screening studies (Etzioni et al., J. Natl. Cancer Inst., 91:1033.1999); Maattanen et al., Br. J. Cancer 79:1210 1999; Schroder et al., J. Natl. Cancer Inst., 90:1817.1998). A major limitation of the serum PSA test is a lack of prostate cancer sensitivity and specificity especially in the intermediate range of PSA detection (4-10 ng/ml). Elevated serum PSA levels are often detected in patients with non-malignant conditions such as benign prostatic
30 hyperplasia (BPH) and prostatitis, and provide little information about the aggressiveness of the cancer detected. Coincident with increased serum PSA testing, there has been a dramatic increase in the number of prostate needle biopsies performed (Jacobsen et al., JAMA 274:1445 1995). This has resulted in a surge of equivocal prostate needle biopsies (Epstein and Potter J. Urol., 166:402 2001). Thus, development of additional serum and tissue

biomarkers or additional methods to detect a patient at risk for prostate cancer remains. There remains a need for determining those at risk for or susceptible to prostate cancer, early-stage prostate cancer prognosis, and early intervention.

5 SUMMARY

The present invention provides, for the first time, a novel structural variation of the (Phosphoserine phosphatase-like) *PSPHL* locus that is tightly linked to gene expression and demonstrates unusual patterns of population differentiation. Given the potential importance of genomic variations in the differential risk for diseases, and the invention provides an
10 association of the variation within the *PSPHL* locus with prostate cancer in the African American population.

The findings presented herein may have an important impact on the design of clinical trials focused upon the prevention of prostate cancer in subject populations, for example in high-risk individuals, on the implementation of community based outreach programs aimed at
15 early screening and timely treatment during the window of curability, or on individualized treatment of subjects with advanced diseases.

In a first aspect, the invention provides methods of detecting the presence or absence of a nucleic acid segment in the (Phosphoserine phosphatase-like) *PSPHL* gene locus of a subject, wherein the presence or absence of the nucleic acid segment in the gene locus
20 indicates an altered risk of cancer.

In preferred embodiments, the cancer is prostate cancer.

In one embodiment, the presence or absence of the nucleic acid segment in the *PSPHL* gene locus is detected in an African American subject. In another embodiment, the absence of the nucleic acid segment indicates an increased risk of prostate cancer in the
25 African American subject.

In a further embodiment, the nucleic acid segment comprises 133 base pairs of exon 1 of human *PSPHL* mRNA encoded by GenBank Accession No. AJ001612 corresponding to SEQ ID NO: 1.

In one embodiment, the nucleic acid segment comprises SEQ ID NO: 2.

30 In one embodiment, the nucleic acid segment comprises SEQ ID NO: 13.

In one embodiment, the nucleic acid segment comprises SEQ ID NO: 14.

In one embodiment, the nucleic acid segment comprises SEQ ID NO: 15.

In a further embodiment, the presence of the insertion allele of the *PSPHL* gene locus is correlated with the expression of the *PSPHL* gene product. In a related embodiment, the

absence of the insertion allele of the PSPHL gene locus is correlated with the absence of the PSPHL gene product.

In another embodiment, the deletion allele is associated with the expression of a set of genes.

5 In one embodiment, the subject is homozygous for a deletion in the *PSPHL* gene locus. In another embodiment, the subject is heterozygous for a deletion in the in the *PSPHL* gene locus. In a related embodiment, the homozygous deletion allele is associated with the expression of a set of genes. In another embodiment, In another embodiment, the heterozygous deletion allele is associated with the expression of a set of genes.

10 In one embodiment, the expression of the PSPHL gene product is associated with the expression of a set of genes.

In another aspect, the invention features a method of determining the ancestry of a subject comprising detecting the presence or absence of a nucleic acid segment in the *PSPHL* gene locus of a sample subject population, wherein the presence or absence of the variation
15 indicates the ancestry of the subject.

In one embodiment, the presence or absence of a nucleic acid segment is indicative of African, e.g., African American or European, e.g., European American, ancestry. In another embodiment, the absence of the nucleic acid segment identifies the population as an African American subject.

20 In one embodiment of any one of the above aspects, the method further comprises selecting subjects with an increased risk of developing prostate cancer. In a related embodiment, the method comprises obtaining a sample from the subjects.

In another aspect, the invention features a biomarker for prostate cancer in an African American subject comprising an insertion in the *PSPHL* gene locus, wherein the presence of
25 the biomarker is correlated with a decreased risk of prostate cancer.

In one embodiment, the insertion encodes a nucleic acid comprising 133 base pairs of exon 1 of human PSPHL mRNA encoded by GenBank Accession No. AJ001612 corresponding to SEQ ID NO: 1.

In another embodiment, the insertion encodes a nucleic acid comprising SEQ ID NO:
30 2.

In another related embodiment, the absence of the biomarker is correlated with an increased risk of prostate cancer in the African American subject.

In still another embodiment, the presence of the insertion in the PSPHL gene locus is correlated with the expression of the PSPHL gene product.

In one embodiment, the insertion allele is associated with the expression of a set of genes.

In another aspect, the invention features a method of identifying a subject at risk for developing prostate cancer comprising detecting the presence or absence of a nucleic acid segment in the *PSPHL* gene locus of a subject to determine the genotype of the subject, wherein the absence of the nucleic acid segment in the gene locus indicates an increased risk of prostate cancer.

In yet another aspect, the invention features a method of determining the prognosis of a patient with prostate cancer comprising: detecting the presence or absence of a nucleic acid segment in the *PSPHL* gene locus of a subject, wherein the absence of the variation determines the prognosis of a patient with prostate cancer.

In one embodiment, the prognosis determines the course of treatment.

In another embodiment of any one of the above aspects, the subject is homozygous for a deletion in the in the *PSPHL* gene locus. In still another embodiment of any one of the above aspects, the subject is heterozygous for a deletion in the in the *PSPHL* gene locus.

In a further embodiment of any one of the above aspects, the subject is selected from an African American population.

In another embodiment of any one of the above aspects, the absence of the nucleic acid segment indicates an increased risk of, or risk of recurrence of, prostate cancer.

In still another embodiment of any one of the above aspects, the nucleic acid comprises 133 base pairs of exon 1 of human *PSPHL* mRNA encoded by GenBank Accession No. AJ001612 corresponding to SEQ ID NO: 1.

In another related embodiment of any one of the above aspects, the nucleic acid comprises SEQ ID NO: 2. In another embodiment of any one of the above aspects, the nucleic acid comprises SEQ ID NO: 13, SEQ ID NO; 14 or SEQ ID NO; 15.

In one embodiment of any one of the above aspects, the presence of the insertion allele of the *PSPHL* gene locus is correlated with the expression of the *PSPHL* gene product. In another embodiment of any one of the above aspects, the absence of the insertion allele of the *PSPHL* gene locus is correlated with the absence of the *PSPHL* gene product. In a related embodiment of any one of the above aspects, the homozygous deletion allele is associated with the expression of a set of genes. In another related embodiment of any one of the above aspects, the heterozygous deletion allele is associated with the expression of a set of genes.

In another embodiment of any one of the above aspects, the presence or absence of a nucleic acid segment in the *PSPHL* gene locus is determined using a polymerase chain

reaction (PCR) assay. In a related embodiment, the PCR assay is a multiplexed PCR assay. In a further related embodiment, the PCR is carried out using primers comprising the nucleic acid sequences as set forth as SEQ ID NO: 3 and SEQ ID NO: 4 and primers comprising the nucleic acid sequences as set forth as SEQ ID NO: 5 and SEQ ID NO: 6. In another related
5 embodiment, the nucleic acid sequences as set forth as SEQ ID NO: 3 and SEQ ID NO: 4 amplify a 133 base pair fragment of the insertion sequence in exon 1 of the PSPHL gene. In still another related embodiment, the nucleic acid sequences as set forth as SEQ ID NO: 5 and SEQ ID NO: 6 generate an amplicon only if the insertion sequence is absent.

10 In another embodiment of any one of the above aspects, the subject has previously been treated for prostate cancer.

In still another embodiment of any one of the above aspects, the measurement is performed after surgery or therapy to treat prostate cancer.

In another aspect, the invention features an antibody to detect PSPHL protein in cells and tissues with PSPHL genotypes.

15 In one embodiment, the antibody is polyclonal. In another embodiment, the antibody is monoclonal. In a related embodiment, the polyclonal antibody is directed to the 72AA antigen of prostate cells corresponding to SEQ ID NO: 7.

In another aspect, the invention features a kit for use in identifying a subject at risk for developing prostate cancer comprising primers directed to amplify a 133 base pair sequence
20 of exon 1 of human PSPHL mRNA encoded by GenBank Accession No. AJ001612 corresponding to SEQ ID NO: 1, and instructions for use.

In one embodiment, the primers comprise the nucleic acid sequences as set forth as SEQ ID NO: 3 and SEQ ID NO: 4.

25 In another embodiment, the primers comprise the nucleic acid sequences as set forth as SEQ ID NO: 5 and SEQ ID NO: 6.

In another aspect, the invention features a kit comprising primers comprising the nucleic acid sequences set forth as SEQ ID NO: 3 and SEQ ID NO: 4, and instructions for use.

30 In another aspect, the invention features a kit comprising primers comprising the nucleic acid sequences set forth as SEQ ID NO: 5 and SEQ ID NO: 6, and instructions for use.

In another further aspect, the invention features a kit comprising primers designed against the nucleic acid sequence set forth as SEQ ID NO: 17, and instructions for use.

In still another aspect, the invention features a kit comprising primers designed against the nucleic acid sequence set forth as SEQ ID NO: 18, and instructions for use.

In still another further aspect, the invention features a kit comprising primers designed against the nucleic acid sequence set forth as SEQ ID NO: 19, and instructions for use.

5 In one embodiment of any one of the above aspects, the kits further comprise instructions for use in PCR assay. In one embodiment, the PCR is multiplexed PCR.

In another aspect, the invention features a kit for use in identifying a subject at risk for developing prostate cancer comprising: an antibody directed to a PSPHL antigen, and instructions for use.

10 In another aspect, the invention features a kit comprising an antibody directed to a PSPHL antigen.

In one embodiment of any one of the above aspects, the antibody is monoclonal.

In one embodiment of any one of the above aspects, the antibody is polyclonal. In a related embodiment, the polyclonal antibody is used to detect the 72AA antigen. In another
15 related embodiment, the polyclonal antibody is directed to a sequence encoded by SEQ ID NO: 7.

The above features and advantages of the present invention will be apparent from or are set forth in more detail in the accompanying drawings, which are incorporated in and form a part of this specification, and the following Detailed Description, which together serve
20 to explain by way of example the principles of the present invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a graph showing that PSPHL gene expression is higher in African American prostate cancer tissues when compared to European American prostate tissues. The
25 graph shows Expression microarray analysis of prostate cancer tissues by race. Surgical prostate cancer tissues from 12 European American cases were compared to those from 8 African American cases. In this "volcano" plot, each gene is represented by a dot, positioned by fold expression change (x axis) and p value (y axis), that were calculated by comparing the expression values between the two racial groups. Genes represented by red dots demonstrated
30 expression changes greater than 2 fold and $p < 0.05$. Using this cut-off criteria (2 fold change and $p < 0.05$), six genes are under-expressed in African American prostate tissue, and only gene, PSPHL, is over-expressed in African American prostate tissues.

Figure 2 shows the nucleotide sequence of the PSPHL mRNA (GenBank accession number AJ001612 represented by SEQ ID NO: 1) with the nucleotide numbers on the left. The

position of the P1 (underlined, SEQ ID NO: 12) and P2 (double underlined, SEQ ID NO: 13) primer pairs are shown.

Figure 3 shows RT-PCR validation of PSPHL expression in the same tissue samples used in expression microarray analysis, with the addition of one more African American

5 Tumor sample. The PSPHL gene was expressed in 3 of 12 European American prostate cancers and 7 of the 9 African American prostate cancers. Primer sets used in the assays are indicated in the parentheses to the right and annotated below. Note than primer set 2 amplified two products, because the primers spanned exon 3, which is alternatively spliced during gene transcription (see Figure 6) . GAPDH is a control for RT-PCR. P1: CTGGGAGAACCGGAA
10 GAATAACAT (forward), corresponds to nt 414-438 of the AJ001612 sequence, CCAATATTCACTGAAGGCTGCCGA (reverse), corresponds to nt 760-783 of the AJ001612 sequence. P2: CGGTCATCAGTGAAGAAGGAATCGGA (forward), corresponds to nt 304-329 of the AJ001612 sequence; TGGCGTTATCCTTGACTTGTTGCC (reverse), corresponds to nt 362-385 of the AJ001612 sequence.

15 Figure 4 shows RT-PCR validation of PSPHL expression in paired normal (N) and cancer tissues (T) from 8 European American prostate cancer patients. The 8 cases are independent of cases used in Figure 1 and 2. Primer sets used were identical to those used in Figure 2. Note the high concordance of PSPHL gene expression between the paired normal and tumor tissues from the same patient.

20 Figure 5 shows RT-PCR validation of PSPHL expression in paired normal (N) and cancer tissues (T) from 7 African American prostate cancer patients. Five of the 7 cases are independent of cases used in Figure 1 and 2, while two cases (1081, 1134) overlaps with those used in Figure 2. Primer sets used were identical to those used in Figure 2. Note again the high concordance of PSPHL gene expression between the paired normal and tumor tissues
25 from the same patient.

Figure 6 shows the nucleotide sequence of the alternative PSPHL mRNA. The nucleotide sequence of an alternative PSPHL mRNA (GenBank accession number BC065228) is shown with the nucleotide numbers on the left. The position of the P2 primer pair is shown double underlined. This sequence has an insertion of 122 bp (shown in capitals and bold)
30 compared to the sequence of AJ001612 (Figure 2), resulting in a PCR product that is 122 bp longer.

Figure 7 shows comparison of the human AJ001612 and BC065228 mRNA sequences with the predicted gene from the Chimnpanzee genome chromosome 7 contig (Genbank Accession No. NW 001237953.1).

Figure 8 shows the nucleotide sequence of AAIns sequence in the PSPHL mRNA. The nucleotide sequence of the PSPHL mRNA (GenBank accession number AJ001612) is shown with the nucleotide numbers on the left. The sequence of a 326 bp portion of the African American specific insertion (AAIns) that overlaps with the PSPHL mRNA sequence is shown in bold italics. The position of the P3 primer pair is shown heavy underlined and they amplify a product of 133 bp.

Figure 9A and B are schematic drawings. Figure 9A shows the predicted gene structure of PSPHL. Solid boxes represent the exons and the lines represents the introns. The size and the position of the introns, exons in PSPHL gene is predicted based on the assumption that the human PSPHL and Chimpanzee PSPHL genes are similar to each other. Figure 9B shows the two spliced variants of PSPHL mRNA. Numbers above the bars represent the nucleotide position of the exon boundary in the respective clones. Note that the 5 prime of the exon 1 sequence of BC065228 was not complete, and the last exon in BC065228 was shortened when compared to AJ001612 possibly due to alternative cleavage during mRNA synthesis.

Figure 10 shows that expression of the PSPHL Gene is Determined by the Presence of AAIns. The Figure shows the results of PCR. A total of 11 cases for which expression status of PSPHL were known were examined for the presence of AAIns in genomic DNA isolated from seminal vesicles. The 8 PSPHL mRNA positive cases (362, 731, 994, 1081, 1115, 1166, 1665, 1863) were also positive for the presence of AAIns, as detected by the primer set P3. Three cases that were know to be negative for PSPHL expression (554, 1134, 1957), plus 5 European American genomic DNA samples randomly picked from the case-control cohort (413, 414, 436, 441, 483), were all negative for AAIns (with the exception of a faint band for 1134, which may be caused by contamination). Three cDNA samples were used as positive (731, 1863) and negative controls (1957). The P3 primers: TCAGCTAAAGTGGCTGTTG GGTGT (Forward) nt 25-48 in AJ001612 sequence; AAGCTTCTGCGCTACCTTGCGAT (Reverse) nt 135-157 in AJ001612 sequence.

Figure 11 is a graph showing AAIns is associated with Cancer Risk in European Americans. The graph shows percentages of individuals (expressed as a fraction) with positive AAIns in European American controls (1), prostate cancer cases with Gleason 6 and below (2), and prostate cancer cases with Gleason 7 and above (3).

Figure 12 shows the results of Multiplexed PCR using primer sets P3 and P4 to genotype the AAIns locus. P3 primer pair, as described in Figures 7 and 9, amplified the allele with the AAIns sequence, while P4 primer pair, as designed below, amplified the allele

without the sequence. The validity of the assay is confirmed by the use of mixed genomic DNA samples known to have two copies of the AAIns sequence (1665, 1863), and zero copies of the AAIns sequence (1957, 1704). The P4 primer sequence was designed against a sequence 5' of the AAIns sequence, as determined by alignment of the Chimpanzee human genome assemblies as of 2006, and based on the prediction of the breakpoints in chromosome 7 human reference sequence. P4 primers: AGTCTTGCTATCTTGCCCAGGCTGAT (forward), nt 5419659-5419684 in chromosome 7 human reference assembly GTAGAGACTGGGTTTCACCATGTTGG (Reverse), nt 5421321-5421346 in chromosome 7 human reference assembly. Lane and corresponding samples are as follows: 1. Mixture of genomic DNA from sample 1665 and 1704; 2. Mixture of genomic DNA from sample 1665 and 1957; 3. Mixture of genomic DNA from sample 1863 and 1704; 4. Mixture of genomic DNA from sample 1863 and 1957; 5. Mixture of genomic DNA from sample 1704 and 1957; 6. Mixture of genomic DNA from sample 1665 and 1863; 7. No DNA template; 8. No DNA template.

Figure 13 is a schematic diagram of AAIns in relation to the assembled Human genome (HG) and Chimpanzee genome (CG) chromosome 7 sequences (not to scale). The human WGS and Trace sequences presented in Exhibit F and H were used to obtain the partial assembly of AAIns and their positions and IDs marked. Coordinates for the assembled genome were marked below the line positions and sizes of the exons and introns in both genomes and AAIns were marked above their respective spaced-out positions. The 5' breakpoint is chr7 55798228 in the assembled human genome, and separated by 9bp (GTGCGTCTA) from the 3' break point at Chr7 55798238 in the assembled human genome. The position of primer sets (P3 and P4) used to genotype the AAIns locus are marked with red vertical bars as shown. P3 primer set amplifies sequences in Exon1 which is part of AAIns, while P4 primer set amplifies the PSPHL-null allele. A Black triangle marks the site of a ~9kb viral insertion specific to the Chimpanzee genome, which resulted in size differences in intron 3 between the two genomes.

Figure 14 shows the alignment of AJ001612 with the human genome. Matching bases in AJ001612 and human genomic sequences are colored blue and capitalized. Light blue bases mark the boundaries of gaps in either sequence. The mRNA sequence (query sequence) is presented first, followed by the genomic sequence. Each sequence is marked by the nucleotide position (for mRNA) or coordinates (genome sequence based on the March 2006 assembly).

Figure 15 shows the alignment of AJ001612 with the assembled Chimpanzee genome.

Matching bases in AJ001612 and Chimpanzee genomic sequences are colored blue and capitalized. Light blue nucleotides mark the boundaries of predicted exons in either sequence. The mRNA sequence (query sequence) is presented first, followed by the genomic sequence. Each sequence is marked by the nucleotide position (for mRNA) or coordinates (genome
 5 sequence based on the March 2006 assembly).

Figure 16 shows the alignment of BC065228 with the assembled Chimpanzee genome. Matching bases in BC065228 and Chimpanzee genomic sequences are colored blue and capitalized. Light blue nucleotides mark the boundaries of predicted exons in either sequence. The mRNA sequence (query sequence) is presented first, followed by the genomic sequence.
 10 Each sequence is marked by the nucleotide position (for mRNA) or coordinates (genome sequence based on the March 2006 assembly). Note that the exon 1 sequence is not complete in BC065228.

Figure 17 shows partial assembly of AAIns based on WGS and Trace Archive. In the Figure, the 4 exon sequences in BC0065228 (SEQ ID NO: 8, Figure 16) were used to query
 15 the human whole genome shotgun (WGS) and Trace Archive. These queries identified 4 genomic sequences – Contig0, gnl|ti|226793227, gi|148184701|, gi|68978189| - each of which contain at least one exon sequence (red underlined; see Figure 13).

Figure 18 shows identification of the AAIns 3' breakpoint. The first 4kb of
 >gi|68978189|gb|AADB02010625.1| Homo sapiens chromosome 7 CRA_219000002701389
 20 whole genome shotgun (WGS) sequence was aligned with the assembled human genome. The matched sequence is blue capitalized and the light blue bold nucleotides border to small region of mismatch. Exon 3 in the WGS sequence is marked red underline. The nucleotide position (for WGS sequence) and coordinates (for human chr7 genome sequence) are marked to the right. Results of the alignment are ordered by the query sequence followed by the
 25 genome sequence. Therefore, the 3' breakpoint of AAIns (upstream black) is where the region of homology (blue) with the assembled human genome sequence ends.

Figure 19 shows identification of the AAIns 5' breakpoint. First, a 100bp non-AAIns sequence upstream to the AAIns 3' breakpoint from the assembled human genome (1, see Exhibit G above) was used to query the Trace Archive to obtain gnl|ti|1656600323 (2). The
 30 overlapping sequence is underlined. Next, the Trace sequence gnl|ti|1656600323 (2) was then used to query the assembled human genome sequence (3). The overlap between the gnl|ti|1656600323 sequence and the human genome sequence is shown. Therefore, the 5' breakpoint of AAIns (downstream black) is where the region of homology with the assembled human genome sequence ends. Finally, the Trace sequence gnl|ti|1656600323 (2)

was used to query the assembled Chimpanzee genome. The overlap between the gnl|ti|1656600323 sequence and the chimpanzee genome sequence upstream of the 5' AAIns breakpoint is shown in blue bold, and the downstream of the 5' AAIns breakpoint is shown in red italics. This confirms that the gnl|ti|1656600323 sequence contains the human AAIns sequence.

Figure 20 shows validation of the breakpoint position. A 200bp assembled human genome sequence flanking the breakpoint (5' at chr7 55798228, 3' at chr7 55798238) was aligned with the assembled Chimpanzee genome. The two matched sequences on the Chimpanzee genome spanned 52765 bp on chromosome 7. The bold nucleotides mark the boundaries of the AAIns insertion. The 52765 bp insertion includes exons 1, 2, and 3 (underlined) of the Chimpanzee PSPHL gene. Nucleotide positions are marked to the right. Again query sequence is followed by Chimpanzee genome sequence.

Figure 21 A and B shows differential expression of *PSPHL* detected in two independent samples sets using different array platforms. Panel A is a "Volcano" plot of expression data derived from the Agilent platform. Average expression ratios for each gene from the comparison of African American vs. European American prostate tumors were plotted on the X axis in log scale, and $-\log_{10}$ of p values comparing the two groups were plotted on the Y axis. Differentially expressed genes (red) were defined by fold expression greater than 2 and p value less than 0.05. Panel B is a heatmap for *PSPHL* expression ratios detected using cDNA microarrays in paired normal and tumor prostate tissues. Expression ratios of each of the 52 (13x4) samples were derived from the comparison of test sample vs a common BPH reference, and represented by red color if overexpressed relative to BPH. Paired N/T samples from each of the 26 cases were aligned vertically as shown. AA: African American; EA: European American; N: normal; T: tumor.

Figure 22 shows two alternatively spliced *PSPHL* transcripts. The ORF sequences were marked as color matched bars above numbered variant sequence positions, the translated amino acid sequences were similarly color coded according to their respective coding sequences. GenBank sequence associated with BC065228 is incomplete at the 5' end but no other putative start sites were identified from the additional sequence we revealed from RACE analysis. Variant 1 encodes SEQ ID NO: 9. Variant 2 encodes SEQ ID NO: 10.

Figure 23 A and B. Panel A shows the results of RT-PCR. The RT-PCR shows concordance between DNA and RNA in matched cases. Primers were designed to amplify a short stretch of DNA sequence within nt 50-200 of AJ00161 12, later confirmed to be within a single exon (see Figure 5), and correspond to the *PSPHL* mRNA sequences absent in the

reference genome. Primers used to examine gene expression by RT-PCR spanned the exons, as revealed after the exon structure was later defined, excluding the possibility of false positive RNA detection due to DNA contamination. Primers for GAPDH were similarly designed to detect DNA within a single exon, and to detect the transcript sequences spanning two exons. Panel B shows the results of RT-PCR. In Panel B The PSPHL gene structure, indel break points, and the genotyping assay. The top is a schematic diagram of the complete PSPHL gene structure, with 4 exons sized at 212bp, 113bp, 122bp, and 502 bp. Positions for TRACE and WGS sequences used to partially assemble the insertion allele were marked with light blue lines above the insertion allele track. The deletion allele sequence positions were defined by the reference genome. Sizes of the DNA segments, when available, are indicated by numbers shown above the tracks. The bottom shows representative genotyping results in African Americans and European Americans. The genomic positions of primer sets (P3 and P4) used to genotype the PSPHL locus were marked with red vertical bars as shown in A. Ins/Ins genotype: P3 signal only; Ins/Del genotype: both P3 and P4 signals, Del/Del genotype: P4 signal only.

Figure 24 is a graph that shows a summary of genotyping results in 3 populations. AA: African American; EA: European American.

Figure 25 is a graph that shows PSPHL genotype in cases and controls in the African American population.

Figure 26 is a Western blot that shows detection of the polyclonal anti 72AA antibody.

Other aspects of the invention are described *infra*.

DETAILED DESCRIPTION

Genetic variation refers to heritable DNA level differences that exist in all living organisms. There are 3 billion chemical base pairs that make up human DNA. One of the most common types of genetic variation is called single nucleotide polymorphisms (SNPs). Each SNP accounts for only one base pair difference, however there are millions of SNPs that account for an average genetic difference between humans in about 0.08% of the 3 billion chemical base pairs in the human genome. Structural variation is another type of genetic variation that each involves at least 1000 such chemical base pair codes. These longer stretches of DNA can be deleted, duplicated, or inserted, leading to change in DNA dosage (therefore also termed copy number variation). They can also be inverted in orientation or translocated to a different location in the genome. In recent years, knowledge regarding this

type of variation has increased, following the improvement of genomic technologies such as microarrays. These variations, by definition, are differences that can be inherited, not alterations such as the ERG-TMPRSS2 fusion that has been discovered in prostate cancer, which are not heritable.

5 Genomic variants of all sizes and types can contribute to genetic disease. From the twin studies the understanding has emerged that there is a strong genetic basis for human prostate cancer. Many geneticists have made use of the genetic variations such as SNPs to identify the risk factors that would elevate the chance of developing prostate cancer if inherited. Indeed, they have been very successful. However, these studies are not without
10 limitations. One such limitation is that the gene or genes that contribute to this elevated risk have not been definitively identified. Yet another limitation is that corresponding studies in African American men, the highest risk group, have not been conducted in similar scale with similar intensity. While commercial tests, some costing a few hundred dollars, are being developed to assess prostate cancer risk, with the hope to benefit humankind by early
15 diagnosis and timely treatment to reduce prostate cancer mortality, it remains to be established whether these tests are equally effective in identifying high risk African American men, already a high risk group for developing prostate cancer. Further, the use of structural variation for identification of prostate cancer genes is a research area that has not been explored.

20 Recently, the present inventors have discovered a novel genomic variation that has the potential to address the above-mentioned aspects of prostate cancer biology and research. Described herein is a locus on chromosome 7, termed PSPHL, that harbors a segment of DNA that can be either present or missing from the human genome. When it is present, the PSPHL gene is expressed in the prostate and may function through the expressed products.
25 When it is absent, the gene is not expressed in the prostate or any other tissues in the body because there is no genetic code to start with. Furthermore, it has been found that this segment of DNA is present in ~96% of healthy African Americans but deleted in most healthy Americans of European descent.

 Genetic differences between humans is shaped by the common ancestry of humans in
30 Africa some 50,000 years ago. Rare forms of genomic variations that seem to be skewed in its distribution among different populations are, in certain cases, more commonly associated with medically relevant traits. Many such genetic differences have been analyzed in detail. One clear example is the genes that confer resistance to malaria. Another example that may

be applied to human prostate cancer, is the use of ancestry informative markers to identify prostate cancer genes in admixture studies. .

The present invention provides methods of detecting the presence or absence of a nucleic acid segment in the (Phosphoserine phosphatase-like) *PSPHL* gene locus of a subject, wherein the presence or absence of the nucleic acid segment in the gene locus indicates an altered risk of cancer.

In certain embodiments, the invention provides biomarkers for prostate cancer in an African American subject comprising an insertion in the *PSPHL* gene locus, wherein the presence of the biomarker is correlated with a decreased risk of prostate cancer. Accordingly, the present invention presents a biomarker or biomarkers that are differentially present in samples of prostate cancer subjects and control subjects, or in subjects of different populations, or in subjects at different stages of cancer, e.g. prostate cancer, progression, and the application of this discovery in methods and kits for determining the presence of prostate cancer. These biomarkers are found in samples from prostate cancer subjects at levels that are different than the levels in samples from subject in whom prostate cancer is undetectable. Accordingly, the amount of the biomarker, or, one or more biomarkers, found in a test sample compared to a control, or the presence or absence of one or more markers in the test sample provides useful information regarding the cancer status of the subject.

In accordance with one embodiment of the invention, a set of genes whose expression correlates with expression of the *PSPHL* gene product were identified. Further, in preferred embodiments, the homozygous deletion allele is associated with the expression of a set of genes. In other embodiments, the heterozygous deletion allele is associated with the expression of a set of genes. These genes may also be useful to determine disease onset/progression, to determine prognosis, to determine risk of recurrence and to determine course of therapy in subjects having routine health screenings, routine prostate cancer screenings, in those suspected of having prostate cancer, for those with known a risk of prostate cancer, for those previously treated for prostate cancer. The genes described herein are also useful as novel therapeutic targets.

In one embodiment, the absence of the nucleic acid segment in the *PSPHL* gene product indicates an increased risk of prostate cancer in an African American subject. The absence of the nucleic acid segment as described herein is useful, for example, to predict disease progression. The claimed methods allow for earlier detection of disease recurrence/progression and therefore earlier treatment of subjects with recurrent/progressive disease.

In addition, knowledge of genetic changes that occur in prostate cancer enable the design and screening for targeted therapeutic agents that interact with the targets. The interaction may be direct or indirect. Therapeutic agents are agents that improve survival in subjects with disease, including advanced disease.

5 Provided herein are methods of detecting the presence or absence of a nucleic acid segment in the *PSPHL* gene locus of a subject, methods of identifying a subject at risk for developing prostate cancer, methods of determining the prognosis of a patient with prostate cancer, biomarkers for prostate cancer, and microarray technologies to identify molecular and genetic defects associated with prostate cancer onset or progression, and to correlate the
10 expression of the biomarkers with the presence or stage of disease, thus providing diagnostic and prognostic markers for this disease. Such markers are useful clinically to determine therapeutic strategies for subjects and guide subject treatment.

Definitions

15 Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988);
20 The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

 The term "set of genes" refers to the one or more genes. In certain embodiments, one or more genes is particularly expressed when the nucleic acid segment in the *PSPHL* gene
25 locus is present. In other embodiments, one or more genes is particularly expressed when the nucleic acid segment in the *PSPHL* gene locus is present. The set of one or more genes expressed when the nucleic acid segment in the *PSPHL* gene locus is present may be overlapping, may be the same, or may be different (e.g. the set of genes may have one, two three or more genes in common). The "set of genes" may refer to genes whose expression
30 level, alone or in combination with other genes, is correlated with cancer or prognosis of cancer, for example prostate cancer. The correlation may relate to either an increased or decreased expression of the gene. For example, the expression of the gene may be indicative of cancer, or lack of expression of the gene may be correlated with poor prognosis in a cancer patient.

The term "detect" refers to identifying the presence, absence or amount of an object or molecule.

The term "nucleic acid" or "nucleic acid segment" as used herein refers to any nucleic acid containing molecule, including but not limited to, DNA or RNA. The term encompasses sequences that include any of the known base analogs of DNA and RNA including, but not limited to, 4-acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinylcytosine, pseudoisocytosine, 5-(carboxyhydroxymethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethylaminomethyluracil, dihydrouracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarbonylmethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, and 2,6-diaminopurine. In preferred embodiments, the nucleic acid segment is part of the (Phosphoserine phosphatase-like) *PSPHL* gene locus of a subject. In further examples the nucleic acid segment comprises SEQ ID NO: 2.

The term "gene" refers to a nucleic acid (e.g., DNA) sequence that comprises coding sequences necessary for the production of a polypeptide, precursor, or RNA (e.g., rRNA, tRNA). The polypeptide can be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or functional properties (e.g., enzymatic activity, ligand binding, signal transduction, immunogenicity, etc.) of the full-length or fragment are retained. The term also encompasses the coding region of a structural gene and the sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb or more on either end such that the gene corresponds to the length of the full-length nRNA. Sequences located 5' of the coding region and present on the mRNA are referred to as 5' non-translated sequences. Sequences located 3' or downstream of the coding region and present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene that are transcribed into nuclear RNA (hnRNA); introns may contain regulatory elements such as

enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide. In certain preferred embodiments, the gene is Phosphoserine phosphatase-like (*PSPHL*).

5 The phrase "Phosphoserine phosphatase-like (*PSPHL*) gene locus" is meant to refer to a gene locus on chromosome 7 that harbors a segment of DNA that can be either present or missing from the human genome. When the gene locus is present, the *PSPHL* gene is expressed in the prostate. When the gene locus is absent, the gene is not expressed in the prostate. Human *PSPHL* mRNA is encoded by GenBank Accession No. AJ001612. In
10 certain embodiments, the *PSPHL* gene locus contains a nucleic acid segment comprising 133 base pairs of exon 1 of human *PSPHL* mRNA whose presence or absence corresponds to *PSPHL* expression.

As used herein the phrase "prostate cancer" refers to cancers of the prostate tissue and/or other tissues of the male genitalia, or reproductive or urinary tracts.

15 As used herein, the term "gene expression" refers to the process of converting genetic information encoded in a gene into RNA (e.g., mRNA, rRNA, tRNA, or snRNA) through "transcription" of the gene (i. e., via the enzymatic action of an RNA polymerase), and for protein encoding genes, into protein through "translation" of mRNA. Gene expression can be regulated at many stages in the process. "Up-regulation" or "activation" refers to regulation
20 that increases the production of gene expression products (i.e., RNA or protein), while "down-regulation" or "repression" refers to regulation that decrease production. In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences that are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to
25 the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers that control or influence the transcription of the gene. The 3' flanking region may contain sequences that direct the termination of transcription, post-transcriptional cleavage and polyadenylation.

As used herein, the term "primer" refers to an oligonucleotide, whether occurring
30 naturally as in a purified restriction digest or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid strand is induced, (i.e., in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in

amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

The phrase "determining a prognosis" or "providing a prognosis" refers to determining or providing information regarding the impact of the presence of cancer, for example prostate cancer, (e.g., as determined by the diagnostic methods of the present invention) on a subject's future health (e.g., expected morbidity or mortality, the likelihood of getting cancer, the risk of metastasis).

The term "measuring" means methods which include detecting the presence or absence of marker(s) in the sample, quantifying the amount of marker(s) in the sample, and/or qualifying the type of biomarker. Measuring can be accomplished by methods known in the art and those further described herein, including but not limited to microarray analysis (with Significance Analysis of Microarrays (SAM) software), SELDI and immunoassay. Any suitable methods can be used to detect and measure one or more of the markers described herein. These methods include, without limitation, mass spectrometry (e.g., laser desorption/ionization mass spectrometry), fluorescence (e.g. sandwich immunoassay), surface plasmon resonance, ellipsometry and atomic force microscopy.

"Detect" as used herein refers to identifying the presence, absence or amount of the object to be detected.

"Marker" or "biomarker" in the context of the present invention refer to a polypeptide (of a particular apparent molecular weight) or nucleic acid, which is differentially present in a sample taken from subjects having prostate cancer as compared to a comparable sample taken from control subjects (e.g., a person with a negative diagnosis or undetectable prostate cancer, normal or healthy subject). The term "biomarker" is used interchangeably with the term "marker." The biomarkers are identified by, for example, molecular mass in Daltons, and include the masses centered around the identified molecular masses for each marker, affinity binding, nucleic acid detection, etc.

A marker can be a polypeptide, which is detected at a higher frequency or at a lower frequency in samples of unaffected tissue from prostate cancer subjects compared to samples of affected tissue from prostate cancer subjects.

A marker can be a polypeptide, which is detected at a higher frequency or at a lower frequency in samples of human unaffected tissue from prostate cancer subjects compared to samples of control subjects.

5 A marker can be a polypeptide, which is detected at a higher frequency or at a lower frequency in samples of human affected tissue from prostate cancer subjects compared to samples of control subjects.

A marker can be differentially present in terms of quantity, frequency or both.

10 "Subject" as used herein refers to any animal (e.g., a mammal), including, but not limited to, humans, non-human primates, rodents, and the like, which is to be the recipient of a particular treatment. Typically, the terms "subject" and "patient" are used interchangeably herein in reference to a human subject.

15 "At risk for cancer" refers to a subject with one or more risk factors for developing a specific cancer. Risk factors include, but are not limited to, ancestry, gender, age, genetic predisposition, environmental expose, previous incidents of cancer, preexisting non-cancer diseases, and lifestyle.

"Unaffected tissue," as used herein refers to a tissue from a prostate cancer subject that is from a portion of tissue that does not have gross disease present, for example tissue that is about 1, 2, 5, 10, 20 or more cm from grossly diseased tissue.

20 A polypeptide is differentially present between two samples if the amount of the polypeptide or nucleic acid in one sample is statistically significantly different from the amount of the polypeptide or nucleic acid in the other sample. For example, a polypeptide or nucleic acid is differentially present between the two samples if it is present at least about 25%, at least about 50%, at least about 75%, at least about 100%, 120%, at least about 130%, at least about 150%, at least about 180%, at least about 200%, at least about 300%, at least
25 about 500%, at least about 700%, at least about 900%, or at least about 1000% greater than it is present in the other sample, or if it is detectable in one sample and not detectable in the other.

30 Alternatively or additionally, a polypeptide or nucleic acid is differentially present between two sets of samples if the frequency of detecting the polypeptide or nucleic acid in the cancer subjects' samples is statistically significantly higher or lower than in the control samples. For example, a polypeptide or nucleic acid is differentially present between the two sets of samples if it is detected at least about 25%, at least about 50%, at least about 75%, at least about 100%, at least about 120%, at least about 130%, at least about 150%, at least about 180%, at least about 200%, at least about 300%, at least about 500%, at least about

700%, at least about 900%, or at least about 1000% more frequently or less frequently observed in one set of samples than the other set of samples.

“Diagnostic” means identifying the presence or nature of a pathologic condition, i.e., cancer. Diagnostic methods differ in their sensitivity and specificity. The “sensitivity” of a diagnostic assay is the percentage of diseased individuals who test positive (percent of “true positives”). Diseased individuals not detected by the assay are “false negatives.” Subjects who are not diseased and who test negative in the assay, are termed “true negatives.” The “specificity” of a diagnostic assay is 1 minus the false positive rate, where the “false positive” rate is defined as the proportion of those without the disease who test positive. While a particular diagnostic method may not provide a definitive diagnosis of a condition, it suffices if the method provides a positive indication that aids in diagnosis.

A “diagnostic amount” of a marker refers to an amount of a marker in a subject’s sample that is consistent with a diagnosis of cancer. A diagnostic amount can be either in absolute amount (e.g., $\mu\text{g/ml}$) or a relative amount (e.g., relative intensity of signals).

A “control amount” of a marker can be any amount or a range of amount, which is to be compared against a test amount of a marker. For example, a control amount of a marker can be the amount of a marker in a person without cancer. A control amount can be either in absolute amount (e.g., $\mu\text{g/ml}$) or a relative amount (e.g., relative intensity of signals).

As used herein, the term “sensitivity” is the percentage of subjects with a particular disease.

For example, in the cancer group, the biomarkers of the invention have a sensitivity of about 80.0%-98.6%, and preferably a sensitivity of 85%, 87.5%, 90%, 92.5%, 95%, 97%, 98%, 99% or approaching 100%.

As used herein, the term “specificity” is the percentage of subjects correctly identified as having a particular disease i.e., normal or healthy subjects. For example, the specificity is calculated as the number of subjects with a particular disease as compared to non-cancer subjects (e.g., normal healthy subjects). The specificity of the assays described herein may range from about 80% to 100%. Preferably the specificity is about 90%, 95%, or 100%.

The terms “polypeptide,” “peptide” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an analog or mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. Polypeptides can be modified, e.g., by the addition of carbohydrate residues to form glycoproteins. The

terms “polypeptide,” “peptide” and “protein” include glycoproteins, as well as non-glycoproteins.

“Antibody” refers to a polypeptide ligand substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof, which specifically binds and recognizes an epitope (e.g., an antigen). The recognized immunoglobulin genes include the kappa and lambda light chain constant region genes, the alpha, gamma, delta, epsilon and mu heavy chain constant region genes, and the myriad immunoglobulin variable region genes. Antibodies exist, e.g., as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. This includes, e.g., Fab’ and F(ab)’₂ fragments. The term “antibody,” as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized de novo using recombinant DNA methodologies. It also includes polyclonal antibodies, monoclonal antibodies, chimeric antibodies, humanized antibodies, or single chain antibodies. “Fc” portion of an antibody refers to that portion of an immunoglobulin heavy chain that comprises one or more heavy chain constant region domains, CH1, CH2 and CH3, but does not include the heavy chain variable region.

The phrase “specifically (or selectively) binds” to an antibody or “specifically (or selectively) immunoreactive with,” when referring to a protein or peptide, refers to a binding reaction that is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein at least two times the background and do not substantially bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, polyclonal antibodies raised to marker “X” from specific species such as rat, mouse, or human can be selected to obtain only those polyclonal antibodies that are specifically immunoreactive with marker “X” and not with other proteins, except for polymorphic variants and alleles of marker “X”. This selection may be achieved by subtracting out antibodies that cross-react with marker “X” molecules from other species. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (see, e.g., Harlow & Lane, *Antibodies, A Laboratory Manual* (1988), for a description of immunoassay formats and conditions that can be used to determine specific

immunoreactivity). Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

PROSTATE CANCER MARKERS

5 The present invention is based upon the discovery that the presence or absence of a nucleic acid segment in the *PSPHL* gene locus of a subject indicates an altered risk of cancer, in particular prostate cancer, and the application of this discovery in methods and kits for determining the risk of prostate cancer. Some of these markers are found at an elevated level and/or more frequently in samples from prostate cancer subjects compared to a control (e.g.,
10 subjects with diseases other than prostate cancer). Accordingly, this novel structural variation of the *PSPHL* locus that is tightly linked to gene expression and demonstrates unusual patterns of population differentiation provides useful information regarding probability of whether a subject being tested is at risk for prostate cancer, and has prognostic value. The invention further provides biomarkers that find use in the diagnosis and
15 characterization (e.g. the determination of risk of developing) prostate cancer.

Detection

 The invention provides methods of detecting the presence or absence of a nucleic acid segment in the (Phosphoserine phosphatase-like) *PSPHL* gene locus of a subject, where the
20 presence or absence of the nucleic acid segment in the gene locus indicates an altered risk of cancer, for example prostate cancer.

 Prostate cancer disproportionately affects men of African descent. In certain embodiments of the invention, the presence or absence of the nucleic acid segment in the *PSPHL* gene locus is detected in an African American subject, where the absence of the
25 nucleic acid segment indicates an increased risk of prostate cancer in the African American subject.

 As described herein, the nucleic acid segment is in the *PSPHL* gene locus of a subject. In particular examples, the nucleic acid segment comprises 133 base pairs of exon 1 of human *PSPHL* mRNA encoded by GenBank Accession No. AJ001612 corresponding to SEQ
30 ID NO: 1. In more particular examples, the nucleic acid segment comprises SEQ ID NO: 2. SEQ ID NO: 2 is set forth below:

SEQ ID NO: 2

aagccacaggctccctggctggcgctcagctaaagtggctgttgggtgtccgcaggcttct
 61 gcctggccgcccgcctataagctaccaggaggagctttacgacttcccgtcctgcggg
 121 aagtggcgggcacgatcgcaaggtagcgcagaagcttctcaatggccagcgcagctgca
 181 gccccggggcgcaactcgctcacctgagcctgggaggaaaattctccaaggatgatct
 5 241 cccactcagagctgaggaagcttttctactcagcagatgctgtgtgtttgatgttgaca
 301 gcacggatcatcagtgaagaaggaatcggatgctttcattggatttgaggaaatgtgatc
 361 aggcaacaagtcaaggataacgccaaatggtatatcactgattttgtagagctgctggga
 421 gaaccggaagaataacatccattgtcatcacgctccaaacaacttcagatgaattttac
 481 aagttacacagattgatactgtttgcttacaattgcctattacaacttgctataaaaagt
 10 541 tggtagagatgatctgcactgtcaagtaactacagtttaggaatcctcaaagattggttt
 601 gtttgttttaactgtagtccagtattatatgatcactatcgatttcctggagagtttt
 661 gtaatctgaattctttatgtatatcctagctatatttcatacaaaagtgtttaagagt
 721 gagagtcaattaaacacctttactcttaggaatatagattcggcagccttcagtgaatat
 781 tggttttttccctttggtatgtcaataaaagtattcatgtgtcagaaaaaaaaa

15 SEQ ID NO: 2 is the transcribed mRNA sequence. However, genomic sequences that
 are not transcribed can also be used to detect the presence of the insertion allele. For example,
 SEQ ID NO: 13, SEQ ID NO: 14 and SEQ ID NO: 15 as set forth below can be used to as
 markers of the presence of the insertion allele.

20 SEQ ID NO: 13
 CAGTGCAGCTGAGATCAGACTTCCATGTGTAACTCCCACTACCCTACCAG
 GATGCCTTTTCATAAAGGTAAGAAATGTAAATTTGGCCTTAATATACAAA
 GTTGCCAGGGCAGCACTGGGTCAGTTCTACATATAGTACTTCTACGTTTCAT
 CAGCGGAAACTTTAAGGGAAGGTGAAAATGCTTCTAGAAGGCGACTGGA
 25 CACCAGCGCCTTTGGGCTCTTCTTCTAAGGCCAATAGTGACCTAAATTATT
 GACTGACTGCTCCAATCAAGTGGGCAAAAGGGTACCAAGGCCGCCAACAT
 CAGACAAATTCACCTTGAGGGCCTACCTATGTGCTTTGAAAGACAAAACCTG
 CTGTTGTGAAGGACACTGTATTTTCAGAAAAACATAATCATATTAACAACCT
 AGTAACAATGTAAAATGCTGATGTGTTGAATGCTACTTTAGAAAAACATG
 30 TAAAATCTACAAAAAAAATTTATGATACAAAACCTACGTTATCAATCAT
 CTAGCTAGCTAACTATCTACAGACATGGTTTTTCATTTCTGTTGCTCAGGAT
 GGAAAGCAGTGGGATGATCATAGCTCACTGCAGCCTTGAGCTCCTGGCCT
 CAAGTGATCCTCCTGCCTCCTAAGTAGCTGGGGCCACAGGTGGACACAGT

GACACCTGGGTTTTTTCTTTTGTAGAGACAGGGTTTCACTACACTGCCCAG
GCTGGTGTCAAACCTTTGGAGTCTCGCTGTGTCACCCAGGCTGGGGTGCAGT
GGTGGGGATCTCGGCTCACTGCAACCTCTGCCACCTGGGTTCAAGCGATT
ACTTCCTGTATCAGCCTCCCGAGTAGCTGGGACTACAGGCATGTGCCACC
5 ACACCCGGCTAATTTTTGTATTTTTGTAGAGAACTGGGTTTCACCATGTT
GGCCAGGCTGGTCTCAAACCTCCTGACCTCAGGTGATCCACCTGCCTCGGCC
TCCCAAAGTGCTGGGATTACAGTCATGAGCCGATGAACACTTTCTTATGCT
ATTAAATGGCCTAACCCAGGCGGGTCGTGGTGGCTCACGCCTGTAATCCC
GAAAACCTCGGATGGCCAAGGTAAGAAGATCACTTGAACCTAGGAATTCCA
10 AACTGGCCTGGGCAACATAGCGAGACTCCCATCCCTACAAAAGATACAAA
AATTAGGCCTGGCGCACACCACGCTCGGCTAGTTTTTTGTACCTTTTGTAGA
GACAGGGTTGCGTCATGTTGCCTAAGCCGGTCTCGAACTCCTGAACCCAA
GCCATCCATCCTCCCGCCTCGGCCTCCCAAAGTGCTGGAGATTACAGGGG
CCCAGCCAGCCTCATGTTTTCTTTTAAGCAGTCCCTTCCCTGTTGCACACT
15 TGGGTTAGTTTTCTTTTTAATTTTTTTAAAACAGGGGGTTACCTCAATCTCG
CAGCCTGGAGTGCTGTGGTGGGATCACAGCTCATTGGAGCCTTGAACCTT
GGGGTTCAAGTAGCTGGGGGGCTGAGGTAGGACTACAGAGATGGGGTCC
CGCCATGTTGCCAGGCTGCTCTTGGCCTGAAGGGCATCTCCCGCTGGCCGT
GCCCCGACATAGTTTTCTATTTTTGACCGACATAAACACTGTGCTGAGTCG
20 GTGTTTGTCAACACACAGGACCTGGCGGGGAGGTCGCGGTTACCAGGCTC
CACTCTAAGTAGAAGACTGCCCAGCTCCAAGCACTGTACCTCCCGGTGAC
GTCGCCGAACGCCCGCCCTGTGACGATACCTAAGGCCACCTTCATGACG
CCGCCGAAGGCCCGCCCCTGTGACGCCGCGGAGGCCCGCCCCTCACGCG
GAGCCAATCGGAACTCGAGGCGGGGCTGTTGGGTCTTCGGGAGCGCGCAT
25 GCGCGGGGGGCCACAGGCTCCCTGGCTGGCGTCAGCTAAAGTGGCTGTTG
GGTGTCCGAGGCTTCTGCCTGGCCGCCGCCCTATAAGCTACCAGGAG
GAGCTTTACGACTTCCCGTCCTGCGGGAAGTGGCGGGCACGATCGCAAGG
TAGCGCAGAAGCTTCTCAATGGCCAGCGCCAGCTGCAGCCCCGGCGGCGC
ACTCGCCTCACCTGAGCCTGGgtACGTGCAGCCCCACAACACCTTCCCCAG
30 CCAGGGCCCCGGGGACCCCGGGAGCGTCCCCCGCCACCTGGCGCCGCTCAT
ACCTGGGCAAGGGTGGGACCCCACTGAGGCCCGCCACGCATTAGGGAGCT
TGCACTTCCCGAGTTTTTGACCTCTGACGGGCAGTTGTAATAGCATTAAAGT
TTTTGAAATTTTGTAGCGGGGGTAGAAGGGGCTTGAAAGGGAAGAAAAC
ATCTTTTAAAATATAACGTTCCGGCCGGGCCCCGGGGGTAAACCCTTTAATC

CACATTTGGGAGGCCAGGCAGGGGATTACGAGGTAGGAGTTAAGACCAC
CTGGCCGCATGGGGAACCCTGTTTTACTAAA

5 SEQ ID NO: 13 contains exon 1 and is assembled from sequences gnl|ti|945553014,
gnl|ti|949228349, gi|41830605|gb|AADD01081234.1, and gnl|ti|954837275.

SEQ ID NO: 14

10 GAGATATAGATGGGGAGATTTTTAATTTAGTTTTTTATTTAAAATTGTGTT
TATAAGGAAAGAGATTATTATGTNTTTTGATAGAGAATTTGGGAGAGTTT
GTTTAATTTAGAGAGAGATGGGTAGNAAGATTTATAGGGTTGGANTGTG
ATTTANGAAAAGATTTAAGGTGTTTGGAGAATAATTTTTGGGGAGGAATT
TTTGAATAAAAAAGAAAANTTTGGTAAAAANGGTGGATTTGGTTTTANAT
GAAATAAAAAAATTAACCGGATGTGGTTGCACACGCCCTGTATNCCCAGC
TACTCAGGAGGCTGAGGCACGAGAATCACTAGAACCCAGGAGGTGGAGG
15 TTGCAGTAAGCCAAGTTCGTGCCACTACCCTCCAGCCTGGGCAACAGAGT
AAGACTCCATCTAAAAAAAAAATGAAGAAGAAGAAATTAGTGTAGTGTG
GGAAGTGAAAAAAAAAAAAAGAAAAGGAAAAGAAAAATGATTGAATTC
ATGAATACACTCTTATGTGGCCTGCACCGACTTTGACACAAATTAGATTGG
CTTAGTAGGCAAGGGTGGGATCTTTTCATAATTTTATTTGATGTCTAAAT
20 ACATTTATCTTTTTTTCTTATAGGAGGAAAATTCTTCCAAGGATGATCTCC
CACTCAGAGCTGAGGAAGCTTTTCTACTCAGCAGATGCTGTGTGTTTTGAT
GTTGACAGCACGGTCATCAGTGAAGAAGGAATCGGTGAGCTAGCCAAAAT
CTGTGGTGTAGAGGACGCGGTGTCAGAAATGTAGGGATAGCATTTATTCA
CTTTATGAAATGATAAAGAATTTTTTTTTTTTTTTTTTTTGAGACAGAGTCT
25 CACTCTATTGCCCAGACTGGAGTGCAGTGGCACAATCTTGGCTTACTGCAA
CCTCCACCTCCCAGGTTCAAGCGATTATCCTGCCTCAGCCTCCTGAGTAGC
TGGGATTACAGGCGTGTGCAACCACATCCGGCTAATTTTTTTATTTTTAGT
AGAGATGGAGTTTCACCATGTTGGTCAGGCTGGTCTCAAACCTCCTGACCTC
GTGATCCACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGC
30 CACCATGCCCGGCCATGAAATGATAAAGAATTTCTAAAGGGTGGCTGTTT
TGGATGAAGTGCTGGACCCTGGCTATAAAAGATGAGCACTAGGCTTTTTTC
TCTACCCCTTAGAGTTGAATTTGATATATTGAGAACTNTGCTATCGCTCC
NN

SEQ ID NO: 14 comprises a Trace sequence that has exon 2 (red underlined), and the underlined 3' sequence overlaps with gi|148184701|gb|ABBA01007462.1| Homo sapiens CTG_1103276812568, whole genome shotgun sequence.

5 SEQ ID NO: 15
 CACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCAT
 GCCCGGCCATGAAATGATAAAGAATTTCTAAAGAGTGGCTGTTTTGGATG
 AAGTGCTGGACCCTGGCTATGGAACATGAGCACTAGGCTTTTTCTCTCACC
 CCTTAGAGTTGAATTTGATATATTCATTCATTTATATAAATATTTATGGCA
 10 AAAAAATATGGATAAGACATGTTCTTAGCTTGATTAGGGGAATGATGACC
 ATCACTAACCAGGAAGATGGCCTAGGGGGCAGGGCTCAAGTGTCAGTG
 GCTTTGTGTCATGGTTGAAAGTGTGGCCTTAGTGGTTCCTTGGAAGGTTC
 AGAGTCTCTTGGTGGATCCTGGGTCAGAGCCCCTCTCTCCCTCCCTCCCTC
 CTGCCCTCCCACCTCCTGCCCTGCAGCTGGGCACCACCCTCTGCAGCCCCA
 15 GTCCCCCAGTCATGCACCATGTCATTTTCTTTTTTTTTTTTTTTTTCagGACGG
 AGTCTCGCTCTGTCACCAGGCTGGAGTGCAATGGTGCAATCTCGGCTCACT
 GCAACCTCCGCCTCCTGGGTTCAAGCAATTCTCCTGCCTCCACCTCCTGAG
 TAGCTGAAACTACAG_{gc}ACGCACCACCACACCCGGCTTATTTTTGTATTTTT
 AGTAGAGATGGTGTTCACCATGTTGGCCAGCCTGGTCTTGAACCTCCTGAC
 20 CTCGTGATCTGCCGGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATG
 AGCCACCACACTCGGCCACCATTTTATTTTCAACTCCCTTTCATGGAAGAA
 CGTTTAGCCTTTGGTCTCTTTCTTGATTTATGACAGCTCGCGGCTTTCAAGA
 AAACTACCTATGAATAGGCTGTGTAACCTTTTATTTATTTATTTATTTTGA
 GATGGAGTCTCTGTCTTCTAGGCTGGGGTGCAGTGGCATGATCTCGGCTCA
 25 CTGCAACCTCCCCCTCCCAGGTTCCAGCAATTCTCCTGCCTCAGCTTCCCA
 AGTAGCTGGGATTACAGGCATGCACCACCACACCTGGCTAATTTTTGCATT
 TTTTTTTTATTAGAGATGGGGTTTTGTTCATGTTGGCCAGGCTGGTCTCGAA
 CTCCTGGCTTCAGATGAGCTGCCCACCTCAGCTTCCCAAAGTTTTGGCATT
 ATAGGCATGAGTCACTGCACCCAGCTTGTTAATGTTATTTTCAAGCACACC
 30 TTCAAAAGTTTATTCCAAGGCCTTGCTCTCATAGCAGCAAAGCCTGTCTGT
 TGCATAGTGGGGCTCTTGTTAGCATCTTCCTTTTGGTGGATGTTGTGTAGA
 CCCGGGGCAGTAAGGCATGTAAACCTCGAGGACATTGGACCTGGCTGGCC
 TGACTGTTGGGTCTCCCTCCTAGGACATGGCGAGCCATGGGTGGGGCAGT
 GTCTTTCAAAGCTGCTCTCACGGAGCACTTAGCCCCAATCCAGCCCTCCAG

GGAGCAGGTGCAGAGACTCATAGCAGAGCACCCCCACACCTGACCCCCA
GCATAAGTAAGAGGAGCCGCTGCTCCAGGTGTATTTTCAGCACCAAGTGTTG
GGGAGCATGTCCTCCTGAGAGCATCTAACGATTGCTTTAGAGAGCCCTCT
GGGTGGTTTTATTTATTTTCATTTTTATTTTTTAATTTTATTATTTTAAAAATA
5 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTGAGGCAGAGTC
TCGCTCTGTGCCCCAGACTGGAGTGCAATGGTACGATATCGGCTCACTGC
AACTTCTGCCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCTCCTGAGTA
GTTGGGATTTTCAGGCACCCACCATGCCAGGCTAGTTTTTGTATTTTTAGTA
GAGACAGGATTTACACCATGTTGTTTAGGCTGGTCTCGAACTCCTGACCTCA
10 GGTGGTCTGCCCCGCTCAGACTCCCAAAGTGCTGGGATCACAGGCGTGAG
CCACTGCGCCTGGCAAAATTTTATTTATTTATTTTTTTTTGAGACAAAGTCTC
ACTCTGTGCCCAGGCTGGAGTGCAAGTGGTGCAATCTCAACTCACTGCAA
CTTTTGCCTCCCAGGTTTCACACGATTCTTCTGCCTCAGCCTCAGGAGTAGC
TGGGATTACAGGCGCCTGTCACCACACCTGGCTAATTTTTGTATTTTTAGT
15 AGAGATGGGTTTCGCCATGTTGGCCAGGCTGGTCTCGAACTCTTCATCTCA
AGTGATCCACCTGCCTCGGCCTCCCAACGTGCTGAGATTACAGGTGTGAG
TCACCACACCTGGCTTATTATTTTTTTTTTTAGATTTCAGAGTCTCATTCTGT
TGCCCAGGTTGGAGTACAGTGGTATAATCATGGCCCACTGCAGCCTCGAA
CTCCTGGGGTCAAGTAGCCCTCCACCTTAGCCTCCTTAGTAGTAGGGACT
20 GCAGGTGAATGCCACCGTGCCCAGCTAATTAAAGAATTTTTTTTTTTTTCT
GAAAGTCTTGCTTTGTCACCCAGGCTGGAGTGCAAGTGGCATGAACACAGC
TCACTGCAGCATCAATCTTCTGGGCTCAAGGGATCCTCCTGCCTCTGCTTC
CTAAGTAGCTGGGACTACAGACATAGGCCACCATGCCAATTTTTTTTTTTTT
TTTTTTGAGATGGAGTCTCGCTCTGTTGCCAGGCTGGAGTGCAAGTGGCGC
25 CATCTTGGCTCACTGCAAGCTCCACCTCCTGGGTTACGCCATTCTCTTCC
CTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCTGCCACCACACCTGAC
TAATTTTTTGTATTTTAAATAGAGACGGGGTTTCACCATGTTAGCCAGGAT
GGTCTCAATCTCCTGACCTTGTGATCCACCTGCCTCAGCCTCTCAAAGTGC
TGGGATTACAGGCGTGAGTCACCGCGCCTGGCCCACCATGCCTAATTTTA
30 AAAAAAAAAATTTTTTTATAGACATAGCATCTCACTACATTGCCAGGCTGG
TCTTGAACCTTGTGGGCTTAAGTGACCAATTTTCACAACTCAGCCTCCCAA
GTGTTGGAATTATAGGTTTAAGACACTATACCCAGCCTTTAAGAATCTTCT
TGTAGAGATGGAATCTTGCTGTGTTGCCAGGCTGTTCTCAAGCTCCTTGC
CTCAAGTGATTTTCTACCTTGGCCTCCCAAAGTGCTGGAATTAAAGGCATG

GACTGTGTGGTTCTTCTGGCACTTACACGTGGTCTTGTCTGGCCGGTTGTC
TGGTCCTGTCTGTTTCTGCCTTTCCTCTTTCTCCAGGGAAAACCTAAGCTTT
CCTTTTTGTCCTCATCTTGTGTTTTTCTGGGTCCATGGGCAGAGTAGAGTTC
TAGAATGGTTTCCTAAAGCAGCCAAGCCCTACCCTTTGATTTCTAAGTACA
5 TTTAGAAAAACACATATAAGGCTGGGCATGGTGGCTCACGCCTGTAATCC
CAGCACTTTGGGAGGCCGAAGTGGATGGATCACCAGGTCAGGAGTTCGAG
ACCAGGCTAACCAACATGATGAAACCCTGTCTCTACTAAAAATACAAAAA
TAAGCCGGGCACGGTAGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCC
AAGATGGGTGTATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAAC
10 ATGGTGAAACCCAGTCTCTACTAAAAATACAAAAAAAGTAGCTGGGCGTG
GTGATAGGCGCCTGTAATCCCAGCTACCCAGGAGGCTGAGGGAGGAAAAT
CACTGGAACCCGGGAGGCAGAGGTTGCTGTGAGCCGAGATCATGCCAGTG
CACTCCAGCTAGGGCAACAGAGCAAGATTCCATCTCAAAAAAAAAAAAAA
AAAAAAATTAGCCGGGCATTGTGGCATGGGCCTATAATACTCAGGAGGCT
15 GAGGCAGGAGAAGCGCTTGAACCCGGGAGGTAGAGGTTGCAGTGAGCCA
AGATTGCAGCCTTGCACTCCAGCCTGGGTGACAAGAGTGAAACTCTGTCT
CAAAAAATAAAGAGAAAAACATATAGAAAACATTAACACCCCAGGCAGT
ATACCTTGTCAAACATACCTCAGGCAAATGCATTCAGGAGAAGAAAATAC
ATCTTATTTCCCTCTTCATGTTTCGTTTTTTTTTTTTTGTCTTTTGTATTAC
20 TCAGTGTTGGGTATTTTGTATTTTATTTTGCAGGGAGCTGGTAAGTCGCCT
ACAGGAGCGAAATGTTTCAGGTTTTCTTAATATCTGGTGGCTTTAGGAGTAT
TATAGAGCATGTTGCTTCAAAGCTCAATATCCCAGCAACCAATGTATTTGC
CAGTAGGCTGAAATTCTACCTTAATGGTAAGATGTAAACGGTAACATGTTC
CCTTTCTTAGCAGTTCATTATTAGTATTCTGGGTAATGTCTTTTGGAATG
25 CAACTTGAACAGTCACATCAGAGTTAAATATTGAGATGAATGGTTCTCTTT
TTTTTTTTTTTTTTTTTTTTTGTAGACGGAGTCTCGCTCTGTCGCCCAGGCTGG
GGTGCAGTGGCGCGATCTCGGCTCACTGCAAGCTCCCCCTCCCGGGTTCAC
GCCATTCTCCTGCCTCAGCCTCCCCAGTAGCTGGGACTACAGGCGCCCGCT
ACCATGTCAGGATAATTTTTTGTATTTTGTAGTAGAGACGGGGTTTCTCCGT
30 GTTAGCCAGGATGGTGTGATCTCCTGACCTCGTGATCCGCCCGCCTCGGC
CTCACAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGGCCGAGAT
GAATGGTTCTCTTAATTGATGTCTTTTGCCCTTTGGTACCCTTTGCTCAGCA
AAACAGACTTAGATCATCACCTGTCTTAGCTTTATTATCTATAACATCACC
TGTCTGTACATAAATGTCTGCATTTTCG

SEQ ID NO: 15 contains exon 3 (nucleotides 504-625) and 3' sequence. The underlined sequence overlaps with the 3' end of gnl|ti|226793227 (SEQ ID NO: 14).

The presence of the insertion sequence can be detected by PCR of any of these sequences as set forth herein. In certain cases, the absence of the insertion sequence can be detected by absence of signal. In other cases, the absence of the insertion sequence can be detected by the presence of the deletion allele.

In still other cases, the presence of the deletion allele can be detected by PCR.

Primers can be designed to the following exemplary sequences:

SEQ ID NO: 16

GTGGGCTCAGCCATCCTCCCAACTGAGCCTCCTGAGTAGCTGGGACTACA
GGTGTGAGCCATCACACTCAACTGGTGTAGCTATTTTAGAAGACAAACTG
GCAGTTTCTCAAAAGGCTAAACATACAGTCATCATATAATGCAACAATTT
CACTCCTAGGCATATATCCCAGAGAAATAGAAATATATGTCCACACAAAA
ACTTGTACAGCAATCTTCATAGCAGCATTGTTTCATAATAGCCAATACGTGG
AAACAACCCAAATGTCCATCAACTGATGAACAGATAAAACAAAATGCAGTG
TGTCTCTACCATGGAATATTATTCAGCCACAGAAGAAATGAAATACTGAT
ACACACTATGACATAAAGGAACTTTGAAAACATTGTGCTAAGAGGGAAAA
AAAAGCCA

SEQ ID NO: 17

GTGGGCTCaAGCCATCCTCCCAACTGAGCCTCCTGAGTAGCTGGGACTACA
GGTGTGAGCCATCACACTCAACTGGTGTAGCTATTTTAGAAGACAAACTG
GCAGTTTCTCAAAAGGCTAAACATACAGTCATCATATAATGCAACAATTT
CACTCCTAGGCATATATCCCAGAGAAATAGAAATATATGTCCACACAAAA
ACTTGTACAGCAATCTTCATAGCAGCATTGTTCTgTAATAGCCAATACGTGG
AAACAACCCAAATGTCCATCAACTGATGAACAGATAAAACAAAATGCAGTG
TGTCTCTACCATGGAATATTATTCAGCCACAGAAGAAATGAAATACTGAT
ACACACTATGACATAAAGGAACTTTGAAAACATTGTGCTAAGAGGGAAAA
AAAAGCCA

SEQ ID NO: 18

GTGGGCTCaAGCCATCCTCCCAACTGAGCCTCCTGAGTAGCTGGGACTACA
 GGTGTGAGCCATCACACcCAACTGGTGTAGCTATTTTAGAAGACAAACTG
 GCAGTTTCT CAAAAGGCTA AACATACAGT CATCATATAA
 TGCAACAATTTCACTCCTAGGCATATATCCCAGAGAAATAGAAATATATG
 5 TCCACACAAAAACTTaTACAGCAATCTTCATAGCAGCATTaTTCATAATAG
 CCAATACGTGGAAACAACCCAAATGTCCA TCAACTGATG AACAGATAAA
 CAAAATGCAGTGTGTCTCTA CCATGGAATA TTATTCAGCC
 ACAGAAGAAA TGAAATACTGATACACACTA TGACATAAAG
 GAACTTTGAA AACATTGTGC TAAGAGGGgAAAAAAAGCCA

10

Methods of the invention for determining the prostate cancer status, or the risk of developing prostate cancer of a subject, include for example, obtaining a biomarker profile from a sample taken from the subject; and comparing the subject's biomarker profile to a reference biomarker profile obtained from a reference population, wherein the comparison is
 15 capable of classifying the subject as belonging to or not belonging to the reference population; wherein the subject's biomarker profile and the reference biomarker profile comprise one or more markers as described herein.

The method may further comprise repeating the method at least once, wherein the subject's biomarker profile is obtained from a separate sample taken each time the method is
 20 repeated.

Samples from the subject may be taken at any time, for example, the samples may be taken 24 hours apart or any other time determined useful.

Such comparisons of the biomarker profiles can determine prostate cancer status or risk of prostate cancer in the subject with an accuracy of at least about 60%, 70%, 80%, 90%,
 25 95%, and approaching 100% as shown in the examples which follow.

The reference biomarker profile can be obtained from a population comprising a single subject, at least two subjects, at least 20 subjects or more. The number of subjects will depend, in part, on the number of available subjects, and the power of the statistical analysis necessary.

30 The invention includes methods of qualifying prostate cancer status in a subject comprising:

- (a) measuring at least one biomarker in a sample from the subject, and
- (b) correlating the measurement with prostate cancer status.

The method may also comprise the step of measuring the at least one biomarker after subject management.

In a preferred embodiment, any one of the markers described herein or contemplated by the instant invention are used to make a correlation with the presence or absence of prostate cancer, wherein the prostate cancer may be any type or subtype of prostate cancer.

In another example, the biomarker is an insertion sequence corresponding to a nucleic acid segment in the PSPHL gene locus.

In another example, the biomarker is an insertion sequence set forth in SEQ ID NO: 2.

Optionally, the methods of the invention may further comprise generating data on immobilized subject samples on a biochip, by subjecting the biochip to laser ionization and detecting intensity of signal for mass/charge ratio; and transforming the data into computer readable form; and executing an algorithm that classifies the data according to user input parameters, for detecting signals that represent biomarkers present in prostate cancer subjects and are lacking in non-prostate cancer subject controls.

In some embodiments, the present invention provides methods for detection of the presence or absence of the nucleic acid segment in the *PSPHL* gene locus as described herein, wherein the absence of the sequence is associated with prostate cancer. In some embodiments, the presence or absence of the nucleic acid segment is detected in tissue samples (e.g., biopsy tissue). In other embodiments, detection is carried out in bodily fluids (e.g., including but not limited to, plasma, serum, whole blood, mucus, and urine). Exemplary methods are described below.

Direct sequencing Assays

In some embodiments of the present invention, a nucleic acid segment, for example, but not only limited to a nucleic acid segment in the PSPHL gene locus, is detected using a direct sequencing technique. In these assays, DNA samples are first isolated from a subject using any suitable method. In some embodiments, the region of interest is cloned into a suitable vector and amplified by growth in a host cell (e.g., a bacteria). In other embodiments, DNA in the region of interest is amplified using PCR.

Following amplification, DNA in the region of interest (e.g., the region containing the insertion, the region containing the SNP) is sequenced using any suitable method, including but not limited to manual sequencing using radioactive marker nucleotides, and automated sequencing. The results of the sequencing are displayed using any suitable method. The sequence is examined and the presence or absence of a given SNP is determined.

PCR Assay

In some embodiments of the present invention, the presence or absence of a nucleic acid segment is detected using a PCR-based assay. In some embodiments, the PCR assay comprises the use of oligonucleotide primers that hybridize only to the insertion or deletion allele (e.g., to the region of polymorphism). Both sets of primers are used to amplify a sample of DNA. In certain embodiments, the subject is homozygous for a deletion in the *PSPHL* gene locus. In other embodiments, the subject is heterozygous for a deletion in the in the *PSPHL* gene locus.

Hybridization Assays

In preferred embodiments of the present invention, the presence or absence of a nucleic acid segment is detected using a hybridization assay. In a hybridization assay, the presence or absence of a given SNP is determined based on the ability of the DNA from the sample to hybridize to a complementary DNA molecule (e.g., a oligonucleotide probe). A variety of hybridization assays using a variety of technologies for hybridization and detection are available. A description of a selection of assays is provided herein.

(1) Direct Detection of Hybridization. In some embodiments, hybridization of a probe to the sequence of interest (e.g., a SNP or mutation) is detected directly by visualizing a bound probe (e.g., a Northern or Southern assay; See e.g., Ausabel et al. (eds.), Current Protocols in Molecular Biology, John Wiley & Sons, NY [1991]). In a these assays, genomic DNA (Southern) or RNA (Northern) is isolated from a subject. The DNA or RNA is then cleaved with a series of restriction enzymes that cleave infrequently in the genome and not near any of the markers being assayed. The DNA or RNA is then separated (e.g., on an agarose gel) and transferred to a membrane. A labeled (e.g., by incorporating a radionucleotide) probe or probes specific for the SNP or mutation being detected is allowed to contact the membrane under a condition or low, medium, or high stringency conditions. Unbound probe is removed and the presence of binding is detected by visualizing the labeled probe.

(2) Detection of Hybridization Using "DNA Chip" Assays. In some embodiments of the present invention, the nucleic acid segment is detected using a DNA chip hybridization assay. In this assay, a series of oligonucleotide probes are affixed to a solid support. The oligonucleotide probes are designed to be unique to a given SNP or mutation. The DNA sample of interest is contacted with the DNA "chip" and hybridization is detected.

In some embodiments, the DNA chip assay is a GeneChip (Affymetrix, Santa Clara, Calif.; See e.g., U.S. Pat. Nos. 6,045,996; 5,925,525; and 5,858,659; each of which is herein incorporated by reference) assay. The GeneChip technology uses miniaturized, high-density arrays of oligonucleotide probes affixed to a "chip." Probe arrays are manufactured by
5 Affymetrix's light-directed chemical synthesis process, which combines solid-phase chemical synthesis with photolithographic fabrication techniques employed in the semiconductor industry. Using a series of photolithographic masks to define chip exposure sites, followed by specific chemical synthesis steps, the process constructs high-density arrays of oligonucleotides, with each probe in a predefined position in the array. Multiple probe arrays
10 are synthesized simultaneously on a large glass wafer. The wafers are then diced, and individual probe arrays are packaged in injection-molded plastic cartridges, which protect them from the environment and serve as chambers for hybridization.

The nucleic acid to be analyzed is isolated, amplified by PCR, and labeled with a fluorescent reporter group. The labeled DNA is then incubated with the array using a fluidics
15 station. The array is then inserted into the scanner, where patterns of hybridization are detected. The hybridization data are collected as light emitted from the fluorescent reporter groups already incorporated into the target, which is bound to the probe array. Probes that perfectly match the target generally produce stronger signals than those that have mismatches. Since the sequence and position of each probe on the array are known, by complementarity,
20 the identity of the target nucleic acid applied to the probe array can be determined.

In other embodiments, a DNA microchip containing electronically captured probes (Nanogen, San Diego, Calif.) is utilized (See e.g., U.S. Pat. Nos. 6,017,696; 6,068,818; and 6,051,380; each of which are herein incorporated by reference). Through the use of microelectronics, Nanogen's technology enables the active movement and concentration of
25 charged molecules to and from designated test sites on its semiconductor microchip. DNA capture probes unique to a given SNP or mutation are electronically placed at, or "addressed" to, specific sites on the microchip. Since DNA has a strong negative charge, it can be electronically moved to an area of positive charge.

First, a test site or a row of test sites on the microchip is electronically activated with a
30 positive charge. Next, a solution containing the DNA probes is introduced onto the microchip. The negatively charged probes rapidly move to the positively charged sites, where they concentrate and are chemically bound to a site on the microchip. The microchip is then washed and another solution of distinct DNA probes is added until the array of specifically bound DNA probes is complete.

A test sample is then analyzed for the presence of target DNA molecules by determining which of the DNA capture probes hybridize, with complementary DNA in the test sample (e.g., a PCR amplified gene of interest). An electronic charge is also used to move and concentrate target molecules to one or more test sites on the microchip. The electronic concentration of sample DNA at each test site promotes rapid hybridization of sample DNA with complementary capture probes (hybridization may occur in minutes). To remove any unbound or nonspecifically bound DNA from each site, the polarity or charge of the site is reversed to negative, thereby forcing any unbound or nonspecifically bound DNA back into solution away from the capture probes. A laser-based fluorescence scanner is used to detect binding.

In still further embodiments, an array technology based upon the segregation of fluids on a flat surface (chip) by differences in surface tension (ProtoGene, Palo Alto, Calif.) is utilized (See e.g., U.S. Pat. Nos. 6,001,311; 5,985,551; and 5,474,796; each of which is herein incorporated by reference). Protogene's technology is based on the fact that fluids can be segregated on a flat surface by differences in surface tension that have been imparted by chemical coatings. Once so segregated, oligonucleotide probes are synthesized directly on the chip by ink-jet printing of reagents. The array with its reaction sites defined by surface tension is mounted on a X/Y translation stage under a set of four piezoelectric nozzles, one for each of the four standard DNA bases. The translation stage moves along each of the rows of the array and the appropriate reagent is delivered to each of the reaction site. For example, the A amidite is delivered only to the sites where amidite A is to be coupled during that synthesis step and so on. Common reagents and washes are delivered by flooding the entire surface and then removing them by spinning.

DNA probes unique for the SNP or mutation of interest are affixed to the chip using Protogene's technology. The chip is then contacted with the PCR-amplified genes of interest. Following hybridization, unbound DNA is removed and hybridization is detected using any suitable method (e.g., by fluorescence de-quenching of an incorporated fluorescent group).

In yet other embodiments, a "bead array" is used for the detection of polymorphisms (Illumina, San Diego, Calif.; See e.g., PCT Publications WO 99/67641 and WO 00/39587, each of which is herein incorporated by reference). Illumina uses a BEAD ARRAY technology that combines fiber optic bundles and beads that self-assemble into an array. Each fiber optic bundle contains thousands to millions of individual fibers depending on the diameter of the bundle. The beads are coated with an oligonucleotide specific for the detection of a given SNP or mutation. Batches of beads are combined to form a pool specific

to the array. To perform an assay, the BEAD ARRAY is contacted with a prepared subject sample (e.g., DNA). Hybridization is detected using any suitable method.

(3) Enzymatic Detection of Hybridization.]In some embodiments, hybridization of a bound probe is detected using a TaqMan assay (PE Biosystems, Foster City, Calif.; See e.g.,
5 U.S. Pat. Nos. 5,962,233 and 5,538,848, each of which is herein incorporated by reference). The assay is performed during a PCR reaction. The TaqMan assay exploits the 5'-3' exonuclease activity of DNA polymerases such as AMPLITAQ DNA polymerase. A probe, specific for a given allele or mutation, is included in the PCR reaction. The probe consists of an oligonucleotide with a 5'-reporter dye (e.g., a fluorescent dye) and a 3'-quencher dye.

10 During PCR, if the probe is bound to its target, the 5'-3' nucleolytic activity of the AMPLITAQ polymerase cleaves the probe between the reporter and the quencher dye. The separation of the reporter dye from the quencher dye results in an increase of fluorescence. The signal accumulates with each cycle of PCR and can be monitored with a fluorimeter.

In still further embodiments, polymorphisms are detected using the SNP-IT primer
15 extension assay (Orchid Biosciences, Princeton, N.J.; See e.g., U.S. Pat. Nos. 5,952,174 and 5,919,626, each of which is herein incorporated by reference). In this assay, SNPs are identified by using a specially synthesized DNA primer and a DNA polymerase to selectively extend the DNA chain by one base at the suspected SNP location. DNA in the region of interest is amplified and denatured. Polymerase reactions are then performed using
20 miniaturized systems called microfluidics. Detection is accomplished by adding a label to the nucleotide suspected of being at the SNP or mutation location. Incorporation of the label into the DNA can be detected by any suitable method (e.g., if the nucleotide contains a biotin label, detection is via a fluorescently labeled antibody specific for biotin). Numerous other assays are known in the art.

Other Detection Assays

Additional detection assays that are suitable for use in the present invention include, but are not limited to, enzyme mismatch cleavage methods (e.g., Variagenics, U.S. Pat. Nos. 6,110,684, 5,958,692, 5,851,770, herein incorporated by reference in their entireties);
30 polymerase chain reaction; branched hybridization methods (e.g., Chiron, U.S. Pat. Nos. 5,849,481, 5,710,264, 5,124,246, and 5,624,802, herein incorporated by reference in their entireties); rolling circle replication (e.g., U.S. Pat. Nos. 6,210,884 and 6,183,960, herein incorporated by reference in their entireties); NASBA (e.g., U.S. Pat. No. 5,409,818, herein incorporated by reference in its entirety); molecular beacon technology (e.g., U.S. Pat. No.

6,150,097, herein incorporated by reference in its entirety); E-sensor technology (Motorola, U.S. Pat. Nos. 6,248,229, 6,221,583, 6,013,170, and 6,063,573, herein incorporated by reference in their entireties); INVADER assay, Third Wave Technologies; See e.g., U.S. Pat. Nos. 5,846,717, 6,090,543; 6,001,567; 5,985,557; and 5,994,069; each of which is herein
5 incorporated by reference; cycling probe technology (e.g., U.S. Pat. Nos. 5,403,711, 5,011,769, and 5,660,988, herein incorporated by reference in their entireties); Dade Behring signal amplification methods (e.g., U.S. Pat. Nos. 6,121,001, 6,110,677, 5,914,230, 5,882,867, and 5,792,614, herein incorporated by reference in their entireties); ligase chain reaction (Barnay Proc. Natl. Acad. Sci USA 88, 189-93 (1991)); and sandwich hybridization methods
10 (e.g., U.S. Pat. No. 5,288,609, herein incorporated by reference in its entirety).

Mass Spectroscopy Assay

In some embodiments, a MassARRAY system (Sequenom, San Diego, Calif.) is used to detect the presence or absence of the nucleic acid segments as described herein. U.S. Pat.
15 Nos. 6,043,031; 5,777,324; and 5,605,798; each of which is herein incorporated by reference, described Mass Spectroscopy assay. DNA is isolated from blood samples using standard procedures. Next, specific DNA regions containing the mutation or SNP of interest, about 200 base pairs in length, are amplified by PCR. The amplified fragments are then attached by one strand to a solid surface and the non-immobilized strands are removed by standard
20 denaturation and washing. The remaining immobilized single strand then serves as a template for automated enzymatic reactions that produce genotype specific diagnostic products.

Very small quantities of the enzymatic products, typically five to ten nanoliters, are then transferred to a SpectroCHIP array for subsequent automated analysis with the SpectroREADER mass spectrometer. Each spot is preloaded with light absorbing crystals that
25 form a matrix with the dispensed diagnostic product. The MassARRAY system uses MALDI-TOF (Matrix Assisted Laser Desorption Ionization-Time of Flight) mass spectrometry. In a process known as desorption, the matrix is hit with a pulse from a laser beam. Energy from the laser beam is transferred to the matrix and it is vaporized resulting in a small amount of the diagnostic product being expelled into a flight tube. As the diagnostic
30 product is charged when an electrical field pulse is subsequently applied to the tube they are launched down the flight tube towards a detector. The time between application of the electrical field pulse and collision of the diagnostic product with the detector is referred to as the time of flight. This is a very precise measure of the product's molecular weight, as a molecule's mass correlates directly with time of flight with smaller molecules flying faster

than larger molecules. The entire assay is completed in less than one thousandth of a second, enabling samples to be analyzed in a total of 3-5 second including repetitive data collection. The SpectroTYPER software then calculates, records, compares and reports the genotypes at the rate of three seconds per sample.

5

Data Analysis

In some embodiments, a computer-based analysis program is used to translate the raw data generated by the detection assay (e.g., the presence, absence, or amount of a given nucleic acid segment) into data of predictive value for a clinician. The clinician can access the predictive data using any suitable means. Thus, in some preferred embodiments, the present invention provides the further benefit that the clinician, who is not likely to be trained in genetics or molecular biology, need not understand the raw data. The data is presented directly to the clinician in its most useful form. The clinician is then able to immediately utilize the information in order to optimize the care of the subject.

The present invention contemplates any method capable of receiving, processing, and transmitting the information to and from laboratories conducting the assays, information provides, medical personal, and subjects. For example, in some embodiments of the present invention, a sample (e.g., a biopsy or a serum or urine sample) is obtained from a subject and submitted to a profiling service (e.g., clinical lab at a medical facility, genomic profiling business, etc.), located in any part of the world (e.g., in a country different than the country where the subject resides or where the information is ultimately used) to generate raw data. Where the sample comprises a tissue or other biological sample, the subject may visit a medical center to have the sample obtained and sent to the profiling center, or subjects may collect the sample themselves (e.g., a urine sample) and directly send it to a profiling center. Where the sample comprises previously determined biological information, the information may be directly sent to the profiling service by the subject (e.g., an information card containing the information may be scanned by a computer and the data transmitted to a computer of the profiling center using an electronic communication systems). Once received by the profiling service, the sample is processed and a profile is produced (i.e., expression data), specific for the diagnostic or prognostic information desired for the subject.

30

The profile data is then prepared in a format suitable for interpretation by a treating clinician. For example, rather than providing raw expression data, the prepared format may represent a diagnosis or risk assessment (e.g., likelihood of cancer being present or the subtype of cancer) for the subject, along with recommendations for particular treatment

options. The data may be displayed to the clinician by any suitable method. For example, in some embodiments, the profiling service generates a report that can be printed for the clinician (e.g., at the point of care) or displayed to the clinician on a computer monitor.

5 In some embodiments, the information is first analyzed at the point of care or at a regional facility. The raw data is then sent to a central processing facility for further analysis and/or to convert the raw data to information useful for a clinician or patient. The central processing facility provides the advantage of privacy (all data is stored in a central facility with uniform security protocols), speed, and uniformity of data analysis. The central processing facility can then control the fate of the data following treatment of the subject. For
10 example, using an electronic communication system, the central facility can provide data to the clinician, the subject, or researchers.

In some embodiments, the subject is able to directly access the data using the electronic communication system. The subject may chose further intervention or counseling based on the results. In some embodiments, the data is used for research use. For example,
15 the data may be used to further optimize the inclusion or elimination of markers as useful indicators of a particular condition or stage of disease.

Antibodies

Antibodies are well known to those of ordinary skill in the science of immunology.
20 As used herein, the term “antibody” means not only intact antibody molecules, but also fragments of antibody molecules that retain immunogen binding ability. Such fragments are also well known in the art and are regularly employed both *in vitro* and *in vivo*. Accordingly, as used herein, the term “antibody” means not only intact immunoglobulin molecules but also the well-known active fragments F(ab')₂, and Fab. F(ab')₂, and Fab fragments which lack the
25 Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding of an intact antibody (Wahl et al., *J. Nucl. Med.* 24:316-325 (1983). The antibodies of the invention comprise whole native antibodies, bispecific antibodies; chimeric antibodies; Fab, Fab', single chain V region fragments (scFv) and fusion polypeptides.

30 In one embodiment, an antibody that binds PSPHL polypeptide (e.g., PSPHL or a PSPHL variant) is monoclonal. Alternatively, the anti-PSPHL antibody is a polyclonal antibody. The preparation and use of polyclonal antibodies are also known the skilled artisan. The invention also encompasses hybrid antibodies, in which one pair of heavy and light chains is obtained from a first antibody, while the other pair of heavy and light chains is

obtained from a different second antibody. Such hybrids may also be formed using humanized heavy and light chains. Such antibodies are often referred to as “chimeric” antibodies.

5 In general, intact antibodies are said to contain “Fc” and “Fab” regions. The Fc regions are involved in complement activation and are not involved in antigen binding. An antibody from which the Fc’ region has been enzymatically cleaved, or which has been produced without the Fc’ region, designated an “F(ab’)₂” fragment, retains both of the antigen binding sites of the intact antibody. Similarly, an antibody from which the Fc region has been enzymatically cleaved, or which has been produced without the Fc region, designated an
10 “Fab” fragment, retains one of the antigen binding sites of the intact antibody. Fab’ fragments consist of a covalently bound antibody light chain and a portion of the antibody heavy chain, denoted “Fd.” The Fd fragments are the major determinants of antibody specificity (a single Fd fragment may be associated with up to ten different light chains without altering antibody specificity). Isolated Fd fragments retain the ability to specifically
15 bind to immunogenic epitopes.

Antibodies can be made by any of the methods known in the art utilizing PSPHL gene product (e.g. polypeptides gene product) or immunogenic fragments thereof, as an immunogen. In embodiments, a synthetic PSPHL protein sequence is used to generate the PSPHL antibody, In other embodiments, said sequence correspond to SEQ ID NO: 7
20 (MASASCSPGGALASPEPGRKILPRMISHSELRKLFYSADAVCFDVDSTVISEEGIGCFHWIWRKCDQATSQG).

One method of obtaining antibodies is to immunize suitable host animals with an immunogen and to follow standard procedures for polyclonal or monoclonal antibody production. The immunogen will facilitate presentation of the immunogen on the cell surface.
25 Immunization of a suitable host can be carried out in a number of ways. Nucleic acid sequences encoding a PSPHL polypeptide, or immunogenic fragments thereof, can be provided to the host in a delivery vehicle that is taken up by immune cells of the host. The cells will in turn express the receptor on the cell surface generating an immunogenic response in the host. Alternatively, nucleic acid sequences encoding a PSPH: polypeptide, or
30 immunogenic fragments thereof, can be expressed in cells *in vitro*, followed by isolation of the receptor and administration of the receptor to a suitable host in which antibodies are raised.

Using either approach, antibodies can then be purified from the host. Antibody purification methods may include salt precipitation (for example, with ammonium sulfate), ion exchange chromatography (for example, on a cationic or anionic exchange column preferably run at neutral pH and eluted with step gradients of increasing ionic strength), gel
5 filtration chromatography (including gel filtration HPLC), and chromatography on affinity resins such as protein A, protein G, hydroxyapatite, and anti-immunoglobulin.

Antibodies can be conveniently produced from hybridoma cells engineered to express the antibody. Methods of making hybridomas are well known in the art. The hybridoma cells can be cultured in a suitable medium, and spent medium can be used as an antibody
10 source. Polynucleotides encoding the antibody of interest can in turn be obtained from the hybridoma that produces the antibody, and then the antibody may be produced synthetically or recombinantly from these DNA sequences. For the production of large amounts of antibody, it is generally more convenient to obtain an ascites fluid. The method of raising ascites generally comprises injecting hybridoma cells into an immunologically naive
15 histocompatible or immunotolerant mammal, especially a mouse. The mammal may be primed for ascites production by prior administration of a suitable composition; e.g., Pristane.

Monoclonal antibodies (Mabs) produced by methods of the invention can be "humanized" by methods known in the art. "Humanized" antibodies are antibodies in which at least part of the sequence has been altered from its initial form to render it more like human
20 immunoglobulins. Techniques to humanize antibodies are particularly useful when non-human animal (e.g., murine) antibodies are generated. Examples of methods for humanizing a murine antibody are provided in U.S. Patent Nos. 4,816,567, 5,530,101, 5,225,539, 5,585,089, 5,693,762 and 5,859,205.

25 Kits

The invention features kits for use in identifying a subject at risk for developing prostate cancer. In preferred examples, the kits for use in identifying a subject at risk for developing prostate cancer comprise primers directed to amplify a 133 base pair sequence of exon 1 of human PSPHL mRNA encoded by GenBank Accession No. AJ001612
30 corresponding to SEQ ID NO: 1, and instructions for use.

In certain embodiments, the invention features kits for the detection of AAIns or PSPHL. For example, a kit for detecting AAIns might include reagents for genomic DNA extraction, PCR reagents, and AAIns specific primers. A kit for detecting PSPHL gene expression might include reagents for mRNA isolation, RT-PCR reagents and PSPHL

specific primers. A kit for detection of PSPHL protein expression may include primary antibodies against the PSPHL antigen coupled with general detection methods for specific binding.

In preferred embodiments, the kits of the invention feature primers for use in detecting a nucleic acid segment in the (Phosphoserine phosphatase-like) *PSPHL* gene locus of a subject. Thus, in certain examples, the kits preferably comprise primers. For example, the primers, in certain embodiments, comprise the nucleic acid sequences as set forth as SEQ ID NO: 3 and SEQ ID NO: 4. In other embodiment, the primers comprise the nucleic acid sequences as set forth as SEQ ID NO: 5 and SEQ ID NO: 6.

In other preferred embodiment, the kits comprise the nucleic acid sequences set forth as SEQ ID NO: 3 and SEQ ID NO: 4, and instructions for use or the kits comprise the nucleic acid sequences set forth as SEQ ID NO: 5 and SEQ ID NO: 6, and instructions for use.

The kits may further comprise instructions for use in PCR assay, for example in multiplexed PCR.

Other kits that are featured in the invention are kits for use in identifying a subject at risk for developing prostate cancer comprising an antibody directed to a PSPHL antigen, and instructions for use. The antibody may be monoclonal or polyclonal. The polyclonal antibody may be used to detect the 72AA antigen, for example the polyclonal antibody comprising a sequence encoded by SEQ ID NO: 7.

Tissue sources for the detection of the AAIns genomic DNA, the expressed products including mRNA and protein, may include any tissue sources where genomic DNA, mRNA, or protein can be retrieved.

It is possible in certain embodiments that the kits of this invention could include a solid substrate having a hydrophobic function, such as a protein biochip (e.g., a CIPHERGEN ProteinChip array) and a buffer for washing the substrate, as well as instructions providing a protocol to measure the biomarkers of this invention on the chip and to use these measurements to diagnose prostate cancer.

In one aspect, the invention provides kits for detecting a biomarker for prostate cancer in an African American subject. The invention provides kits for detecting the presence (or absence) of a nucleic acid segment in the (Phosphoserine phosphatase-like) *PSPHL* gene locus of a subject, wherein the presence or absence of the nucleic acid segment in the gene locus indicates an altered risk of cancer. The kits include PCR primers for at least one marker, preferably the nucleic acid comprising SEQ ID NO: 2 as described herein. However, the kit may include identification of more than one biomarker as described herein. The kit

may further include instructions for use and correlation of the biomarker with disease status. The kit may also include a DNA array containing the complement of one or more of the biomarkers, reagents, and/or enzymes for amplifying or isolating sample DNA. The kits may include reagents for PCR, for example, probes and/or primers, and enzymes.

5 The kits of the invention have many applications. For example, the kits can be used to differentiate if a subject has prostate cancer or does not have prostate cancer (a negative diagnosis).

10 In one embodiment, a kit comprises: (a) a substrate comprising an adsorbent thereon, wherein the adsorbent is suitable for binding a biomarker, and (b) instructions to detect the marker or markers by contacting a sample with the adsorbent and detecting the biomarker or markers retained by the adsorbent. In some embodiments, the kit may comprise an eluant (as an alternative or in combination with instructions) or instructions for making an eluant, wherein the combination of the adsorbent and the eluant allows detection of the biomarkers using gas phase ion spectrometry.

15 Such kits can be prepared from the materials described above, and the previous discussion of these materials (e.g., probe substrates, adsorbents, washing solutions, etc.) is fully applicable to this section and will not be repeated.

20 In another embodiment, the kit may comprise a first substrate comprising an adsorbent thereon (e.g., a particle functionalized with an adsorbent) and a second substrate onto which the first substrate can be positioned to form a probe, which is removably insertable into a gas phase ion spectrometer. In other embodiments, the kit may comprise a single substrate, which is in the form of a removably insertable probe with adsorbents on the substrate. In yet another embodiment, the kit may further comprise a pre-fractionation spin column (e.g., Cibacron blue agarose column, anti-HSA agarose column, K-30 size exclusion
25 column, Q-anion exchange spin column, single stranded DNA column, lectin column, etc.).

30 In another embodiment, a kit comprises (a) an antibody that specifically binds to a biomarker; and (b) a detection reagent. Such kits can be prepared from the materials described above, and the previous discussion regarding the materials (e.g., antibodies, detection reagents, immobilized supports, etc.) is fully applicable to this section and will not be repeated. Optionally, the kit may further comprise pre-fractionation spin columns. In some embodiments, the kit may further comprise instructions for suitable operation parameters in the form of a label or a separate insert.

Optionally, the kit may further comprise a standard or control information so that the test sample can be compared with the control information standard to determine if the test amount of a biomarker detected in a sample is consistent with a diagnosis of prostate cancer.

Reference cells may be normal cells (cells that are not prostate cancer cells) or prostate cells at a different stage from the prostate cancer cells being compared to. The reference cells may be primary cultured cells, fresh blood cells, established cell lines or other cells determined to be appropriate to one of skill in the art.

10 EXAMPLES

This invention is further illustrated by the following examples, which should not be construed as limiting. All documents mentioned herein are incorporated herein by reference.

The incidence and mortality rates of prostate cancer vary considerably among different geographic areas and ethnic groups. African American men have the highest incidence and mortality from prostate cancer in the world (Ries 2007, Chu 2003). They have a 60% increased risk of developing prostate cancer, twice the risk of developing distant disease, and over twice the mortality relative to Americans of European descent (Jemal 2007, Clegg 2002). Multiple reasons have been postulated to explain these findings including access to care, attitudes about care, socioeconomic and education differences, differences in type and aggressiveness of treatment, dietary, and genetic factors. Yet etiological factors accounting for this disparity remain elusive (Freedland and Isaacs 2005). For a long period of time, the consistent risk factors identified for prostate cancer in addition to race, have been age and family history.

Recently, a number of common genetic variants joined and expanded the known list of risk factors for human prostate cancer (Zheng 2008, Eeles 2008, Thomas 2008). This most recent finding was the result of combined efforts of multiple large-scale genome-wide association studies, in which single nucleotide polymorphism (SNP) allele frequencies were compared between prostate cancer cases and unaffected controls. Because treatments for early stage prostate cancer are effective, these newly established genetic risk variants, most of which have cumulative effect, may be useful for seeking out high-risk individuals for early screening and timely intervention, as the inherited risk alleles can be tested inexpensively anytime during the lifespan. However, whether these genetic findings can be translated to better control of prostate cancer in African American men is still largely unknown, owing to

the relative scarcity of corresponding studies at similar scales in the African American population. Further, it is still unknown whether any of the established genetic variants may lead to phenotypic differences in gene expression or gene function. In the absence of functional validation and follow-up studies of these SNPs, it is presently not clear how they contribute to prostate cancer risk.

To better understand the potential genetic-based contribution to prostate cancer disparity, the experiments described herein employ an alternative approach focusing on gene expression studies to identify candidate genes that can be followed up for the assessment of related genetic variations contributing to the development of prostate cancer. The experiments describe a genome-wide gene expression analysis on surgical human prostate specimens and reveal differences in the mRNA expression of the PSPHL (Phosphoserine Phosphatase-Like) gene when comparing prostate tumors from patients of African and European descent. The cDNA for this gene was originally isolated from fibroblasts derived from a patient with Fanconi's anemia by cDNA differential display technique and was described as a homologue to L-3-phosphoserine-phosphatase (PSPH) (Planitzer 1998). As described herein, the expression of the PSPHL gene was completely shut down in approximately 70% of prostate tumors from patients of European descent, but expressed at readily detectable levels in 80- 90% of prostate tumors from patients of African descent. This observation was recently confirmed by another group in a published study that examined expression differences between prostate tumors derived from patients of European and African populations (Wallace 2008). Further, detailed follow-up genomic analysis revealed that the expression status of the PSPHL gene is 100% concordant with the presence or absence of an insertion allele at the PSPHL gene locus in the assayed genomes (i.e., the corresponding germ line DNA). The insertion sequence is absent in the genomes of individuals with negative prostate PSPHL expression and also absent in the current version of the assembled reference human genome.

The experiments described herein subsequently defined the insertion/deletion breakpoints based on TRACE DNA sequence data and developed a robust genotyping assay. Genotyping results in populations of different geographical ancestries validated the unusual pattern of population differentiation of this novel structural variation. The studies described herein are based on this novel finding.

Example 1. Expression microarray analysis using Agilent Whole Genome Expression

Microarray was used to compare gene expression in surgical prostate cancer tissues derived from 12 European American patients and 8 African American patients (Figure 1). Using a cut-off criteria of 2-fold change in expression and $p < 0.05$, six genes were found to be under-expressed in African American prostate tissues, and one gene, PSPHL, was found to be over-expressed in African American prostate tissues.

The PSPHL probe from the Agilent Whole Genome Expression Microarray was mapped to two homologous genes on chromosome 7p11.2; the two genes were PSPH and PSPHL, and further validation using gene-specific primers identified the gene as PSPHL, and not PSPH. PSPHL (Phosphoserine Phosphatase-Like) was previously identified by Planitzer et al (1998) as a human L-3-phosphoserine-phosphatase homologue that is significantly upregulated in FA fibroblasts. The sequence of PSPHL transcript can be found in the NCBI GenBank database under accession number AJ001612 (SEQ ID NO: 1), shown below.

SEQ ID NO: 1

AAGCCACAGGCTCCCTGGCTGGCGTCAGCTAAAGTGGCTGTTGGGTGTCC
GCAGGCTTCTGCCTGGCCGCCGCCGCTATAAGCTACCAGGAGGAGCTTT
ACGACTTCCCGTCCTGCGGGAAGTGGCGGGCACGATCGCAAGGTAGCGCA
GAAGCTTCTCAATGGCCAGCGCCAGCTGCAGCCCCGGCGGCGCACTCGCC
TCACCTGAGCCTGGGAGGAAAATTCTTCCAAGGATGATCTCCCACTCAGA
GCTGAGGAAGCTTTTCTACTCAGCAGATGCTGTGTGTTTTGATGTTGACAG
CACGGTCATCAGTGAAGAAGGAATCGGATGCTTTCATTGGATTTGGAGGA
AATGTGATCAGGCAACAAGTCAAGGATAACGCCAAATGGTATATCACTGA
TTTTGTAGAGCTGCTGGGAGAACCGGAAGAATAACATCCATTGTCATACA
GCTCCAAACAACCTTCAGATGAATTTTACAAGTTACACAGATTGATACTGT
TTGCTTACAATTGCCTATTACAACCTTGCTATAAAAAGTTGGTACAGATGAT
CTGCACTGTCAAGTAAACTACAGTTAGGAATCCTCAAAGATTGGTTTGT
GTTTTTAAGTGTAGTTCCAGTATTATATGATCACTATCGATTTCCCTGGAGA
GTTTTGTAATCTGAATTCTTTATGTATATTCCTAGCTATATTCATACAAAG
TGTTTTAAGAGTGGAGAGTCAATTAAACACCTTTACTCTTAGGAATATAGA
TTCGGCAGCCTTCAGTGAATATTGGTTTTTTTCCCTTTGGTATGTCAATAAA
AGTTTATCCATGTGTCAGAAAAAAAAAAAAA

The race-specific expression difference in PSPHL was further validated by RT-PCR using the gene-specific primer pairs P1 and P2 (Figures 2 and 3), using RNA isolated from

the same tissue samples used in the microarray analysis (Figure 3). This confirmed that there is a higher frequency of PSPHL expression in tissues from African Americans (7 out of 9) compared to European Americans (3 of 12). Additional analyses showed that PSPHL is expressed in 12 of 14 African American cases, but only in 6 of 20 European American cases (data not shown).

Analysis of paired tissue samples determined the patient-specific expression pattern of PSPHL (Figures 4 and 5). Patients with a positive value for PSPHL expression in cancer tissues also had a positive value for the paired normal tissues, and vice versa. These results shows that the expression pattern of PSPHL is all-or-none and the terms either “positive” or “negative” can be used to describe its expression. Further, its expression is individual-specific, i.e, if it is positive, it is positive in both normal and tumor tissues from the same individual and vice versa.

Example 2. PSPHL Gene Analysis

Planitzer et al (1998) described the overexpression of the PSPHL mRNA (the cDNA referred to by Plantizer as CO9) in fibroblasts isolated from a few Fanconi’s anemia patients, and published the PSPHL cDNA sequence (GenBank Accession number AJ001612;SEQ ID NO: 1 as above). Two variants of the PSPHL mRNA were cloned, one is identical to AJ001612, and the other (identified in GenBank under Accession Number BC065228; SEQ ID NO: 8) has an additional insertion of 122bp, that is in an equivalent position to nt 327 of AJ001612 (Figure 6). SEQ ID NO: 8 is shown below. Thus the two variants, if translated, would lead to two distinctive protein products. The function of the protein products is unknown.

SEQ ID NO: 8

```
CCGCAGGCTTCTGCCTGGCCGCCGCCCTATAAGCTACCAGGAGGAGCT
TTACGACTTCCCGTCCTGCGGGAAGTGGCGGGCACGATCGCAAGGTAGCG
CAGAAGCTTCTCAATGGCCAGCGCCAGCTGCAGCCCCGGCGGCGCACTCG
CCTCACCTGAGCCTGGGAGGAAAATTCTTCCAAGGATGATCTCCCACTCA
GAGCTGAGGAAGCTTTTCTACTCAGCAGATGCTGTGTGTTTGTATGTTGAC
AGCACGGTCATCAGTGAAGAAGGAATCGGACGGAGTCTCGCTCTGTCACC
AGGCTGGAGTGCAATGGTGCAATCTCGGCTCACTGCAACCTCCGCCTCCT
GGGTTTCAGGCAGTTCTCCTGCCTCCACCTCCTGAGTAGCTGAAACTACAGG
ATGCTTTCATTGGCTTTGGAGGAAATGTGATCAGGCAACAAGTCAAGGAT
```

AACGCCAAATGGTATATCACTGATTTTGTAGAGCTGCTGGGAGAACCGGA
 AGAATAACATCCATTGTCATACAGCTCCAAACAACCTTCAGATGAATTTTAA
 CAAGTTACACAGATTGATACTGTTTGCTTACAATTGCCTATTACAACCTTGC
 TATAGAAAGTTGGTACAGATGATCTGCACTGTCAAGTAACTACAGTTAG
 5 GAATCCTCAAAGATTGGTTTGTGTTTGTGTTTAACTGTAGTTCCAGTATTATAT
 GATCACTATTGATTTTCCTGGAGAGTTTTGTAATCTGAATTCTTTATGTATAT
 TCCTAGCTATATTTTCATACAAAGTGTTTTAAGAGTGGAGAGTCAATTAAAC
 ACCTTTACTCTTAGGAAAAAAAAAAAAAAAAAAAA

10 BLAST searches of the AJ001612 sequence against the assembled Human and
 Chimpanzee genome sequences using the UCSC genome browser identified matches for nt
 327-830 of AJ001612 in both genomes assembled on March 2006 (Figure 14 and Figure 15).
 This matched sequence is the last exon of the gene based on the absence of gaps in this
 matched sequence in either genome. Alignment of the mRNA sequence AJ001612 to human
 15 genomic DNA shows that nt 1-326 of AJ001612 had no strong match in the human genomic
 DNA (see below), but matched, with more than 98% homology, to DNA sequences in the
 Chimpanzee genome (Figure 15). The first stretch of matched sequence corresponds to nt 3-
 214 of AJ001612, suggesting that nt 3-214 of AJ001612 is the sequence for the first exon.
 Likewise, the second exon is likely to be 113bp in size and corresponds to nt 215-327 of
 20 AJ001612 (Figure 15). Furthermore, the unique 122bp insertion present in BC065228 (Figure
 6) does not match the assembled human genome sequence (not shown) but matched a
 continuous sequence in the Chimpanzee genome between predicted exon 2 and the last exon
 (Figure 16), suggesting that this 122bp insertion is exon 3 that is likely alternatively spliced
 out in AJ001612. Therefore, the human PSPHL has 4 predicted exons based on the alignment
 25 of AJ001612 and BC065228 with the assembled Chimpanzee genome (Figure 7).

The results from the alignment of the AJ001612 sequence with the assembled Human
 and Chimpanzee genomes (Figures 14 - 16 and Figure 7) suggested that there is a genomic
 DNA indel (insertion or deletion) of unknown size that is present in the assembled
 Chimpanzee genome, but not present in the assembled Human genome. This segment of
 30 DNA (hereafter named AAIns for African American specific insertion; see below) aligns
 with nt 1-326 of the AJ001612 sequence (Figure 8) and includes the first 3 exons of the
 PSPHL gene and possibly 5 prime regulatory regions and part of intron 3. Deletion of this
 sequence may lead to the absence of PSPHL gene expression and may explain the correlation
 with PSPHL gene expression that is observed.

The predicted exon structure and the mRNA variants are shown in Figure 9. Both the predicted exon structure and the involvement of AAIns were validated by the identification of the breakpoint and partial assembly of AAIns as described herein.

5 **Example 3. Partial Assembly of AAIns and Identification of the Breakpoints**

Although AAIns was not assembled into the human genome, it was reasoned that the partial or whole AAIns sequence may be assembled from the Celera Whole Genome Shotgun (WGS) sequence database as well as the Trace Archive that contained DNA sequences not assembled due to lack of consensus. Next, the Celera whole genome shotgun (WGS) sequence database was queried using the mRNA sequences from BC065228, which has sequences from all 4 exons.

First, a contig (Contig0, Figure 17) was assembled based on the WGS and Trace sequences as shown (Figure 17). This contig contains exon 1 and flanking genomic sequences. This sequence, which is part of AAIns, was not assembled into the human genome as previously mentioned. Second, a Trace genomic sequence of 1200 bp that contained part of intron 1, exon 2, and part of intron 2 was identified (gnl|ti|226793227, Figure 17). This sequence, which is part of AAIns, again was not assembled into the human genome as previously mentioned. Third, a WGS sequence of 4906bp that overlaps with gnl|ti|226793227 and contains the 122bp exon 3 was identified (gi|148184701|, Figure 17). The size of intron 2 is therefore determined to be 831bp and consistent with predictions from the alignment with the assembled Chimpanzee genome (Figure 9). This sequence, which is part of AAIns, again was not assembled into the human genome. Fourth, a WGS sequence of 96797bp was identified (gi|68978189|, Figure 17) that contains part of exon 3, complete intron 3 (11954bp), and complete exon 4. This sequence matched the assembled human genome at chr7: 55798238 (Figure 18), at a point which starts 1820 bp downstream of the exon 3 boundary. This is the 3' breakpoint for AAIns.

Finally, the Trace Archive with a 100bp non-AAIns sequence from upstream of the 3' prime AAIns breakpoint from the assembled human genome was queried, and a Trace sequence was identified (gnl|ti|1656600323, Figure 19). The assembled human genome sequence breaks apart from this Trace sequence at chr7: 55798228 (Figure 19). This is the 5' AAIns breakpoint. The distance between the two breakpoints, in the assembled human genome, is 9 base pairs (55798228-55798238). To ensure that the Trace sequence (gnl|ti|1656600323, as shown in Figure 17) contains AAIns sequence following the breakpoint, this sequence was aligned with the assembled Chimpanzee genome. This

alignment resulted in nearly perfect match with the Chimpanzee genome sequence (Figure 19).

To further validate the position of the breakpoint, the Chimpanzee was queried genome with a 200bp assembled human sequence flanking the breakpoints (Figure 20). As predicted, the human sequence was broken into two pieces with a sequence of 52765 bp in size separating the two pieces in the corresponding Chimpanzee genome. The three exons of the PSPHL gene are within this 52765 bp AAIns sequence. While the complete sequence of the human counterpart of the AAIns) is not fully assembled, the gene structure of the human PSPHL locus has been largely decoded based on this bioinformatics approach.

Example 4. PSPHL gene expression and the detection of AAIns in prostate cancer cases and controls

Based on the data presented, which suggested that the PSPHL gene was expressed in African Americans in higher frequency compared to European Americans, it was thought that (1) the expression of the gene is controlled by the presence of absence of this AAIns; (2) AAIns is present in higher proportion of African Americans when compared to European Americans.

Next, a new primer set, P3, was designed that specifically amplifies a 133 bp fragment of the AAIns (Figures 8 and 10). As Figure 10 shows, the presence of the AAIns is perfectly correlated with the expression of PSPHL in the African American Samples tested. Furthermore, absence of PSPHL expression appears to correlate perfectly with the absence of AAIns (with the exception of a potential positive negative in sample 1134), which is very common in European Americans. This confirms that the expression of PSPHL is regulated by the structural variation of the PSPHL gene and, in particular, gene sequences associated with the AAIns that are required for expression of the PSPHL gene. Next, a case-control cohort comprised of 321 individuals from both racial groups were examined for the presence of AAIns using the same P3 PCR primers. When grouped according to race, 149 of 156 (95%) African Americans are AAIns positive, while only 49 of 165 European Americans (30%) were AAIns positive. This confirms that AAIns is present in higher proportion of African Americans when compared to European Americans. There is no difference between prostate cancer case and controls identified so far within the African American group. However, within the European American group, AAIns is detected in 35.2% of low grade cases (19/54), in 33% of high grade cases (18/55), but in only 21.4 (12/56) of healthy control individuals (Figure 11).

Therefore, one conclusion that can be reached is that the absence of AAIns in the assembled human genome sequence is probably associated with a lack of representation of African Americans in the DNA samples pooled for the Human Genome Project,

5 **Example 5. Genotyping of the AAIns Locus**

The results provided herein indicate that the presence of the AAIns sequence, and thus expression of the PSPHL gene product, is associated with populations with a modulated risk of developing prostate cancer. Thus, assaying for the presence of AAIns through molecular diagnostic techniques provides a method for identifying people with a modulated risk of developing prostate cancer. Furthermore, because AAIns is more prevalent in African Americans, a detection of AAIns through molecular diagnostic techniques provides a genetic test for determining ancestral origin. However, Primer set P3, while reliably detects the presence of AAIns, can not differentiate one copy (heterozygous) versus two copies (homozygous) of the AAIns sequence in the genome. The identification of the breakpoint allowed the design of primer set P4 (Figure 12), which spans the entire AAIns, to detect the PSPHL-null allele. Here, the PSPHL-null allele refers to the allele that does not have the AAIns sequence. P4 primers would not detect the allele with the AAIns sequence because the amplicon would exceed 50kb.

The primer sets P3 and P4 are set forth by SEQ ID NOs: 3, 4, 5 and 6 as follows:

P3

SEQ ID NO:3

Forward. TCAGCTAAAGTGGCTGTTGG GTGT

SEQ ID NO: 4

Reverse. AAGCTTCTGCGCTACCTTGCGA

P4

SEQ ID NO: 5

Forward. AGTCTTGCTATCTTGCCCAGGCTGAT

SEQ ID NO: 6

Reverse. GTAGAGACTGGGTTTCACCATGTTGG

The validity of the genotype assay for the AAIns locus was demonstrated by multiplex PCR as shown in Figure 12. The multiplexed PCR allows the detection of both the AAIns allele and the PSPHL-null allele in a single reaction, in samples that are heterozygous

for the locus. First, samples were identified that were homozygous for the locus. Two samples, 1704 and 1957, are negative for AAIns (as they lack expression of PSPHL as shown in Figure 3) and therefore known to be homozygous for the PSPHL-null allele (both chromosomes lack the AAIns sequence). Two other samples, 1665 and 1863, are positive for AAIns as detected by the P3 primers (Figure 10), but negative for the PSPHL-null allele because there was no signal from the P4 primers (Figure 12, lane 6) and are therefore homozygous for the AAIns allele. A 1:1 mixture of samples with the homozygous AAIns and the homozygous PSPHL-null allele would mimic a heterozygous sample. As shown in Figure 12 (lanes 1-4), all AAIns heterozygous mixtures are detected positive by both the P3 and P4 primer.

Figure 13 is a schematic diagram showing the AAIns locus relative to the PSPHL gene in the human and chimpanzee genomes. Figure 13 shows the position of the AAIns relative to the assembled genomic sequences and the PSPHL gene, the sizes of exons and introns when known, the human WGS and Trace sequences used to assemble the AAIns, the positions of primers used for genotyping, and the precise location of the breakpoints.

Example 7. Molecular Diagnostic Kits for Determining the Risk of Prostate Cancer and Ancestral Origin

Assaying for the presence of AAIns, or genotype of the locus, or the PSPHL gene products can be done by any molecular method known to those skilled in the art. The AAIns sequence can be detected by a number of methods including, but not limited to, Southern blotting, RFLP analysis, PCR, genome arrays (SNP microarrays, Array CGH), Sequenom Assays, and DNA sequencing.

For example, the presence or absence of AAIns is detected by PCR using primers specific to the AAIns genomic sequence and primers specific to the PSPHL-null allele (which lacks the AAIns sequence), and using genomic DNA from any tissue sample. In the example shown here, the P3 primer pairs identified an amplification product of 133 bp that is specific to AAIns, and P4 primers detected the presence of the PSPHL-null allele. However, any other primer pairs that are specific to the AAIns, or that lies on opposite sides of the AAIns (so that the amplicon would span the breakpoints), can be used. The PCR product can be detected by gel electrophoresis or the PCR primers may be fluorescently labeled for detection by any method for measuring fluorescence that is known in the art.

Primer pairs that align with any sequence within the AAIns genomic sequence, with any sequence in the vicinity of the breakpoints (up to 10kb from the breakpoints), or with any sequence that spans the breakpoints can also be used in the methods of the invention as described herein.

5 PSPHL gene expression can be detected by a number of methods including, but not limited to, Northern Blotting, RT-PCR, real-time PCR, in-situ hybridization, microarrays, or any method that detects the gene product, such as Western Blotting, ELISA, mass spectrometry, immunohistochemistry, or protein arrays.

10 In preferred examples, PSPHL gene expression is detected by RT-PCR using primers specific to PSPHL coding sequence and using RNA from any tissue sample where the gene is expressed. In the example shown here, PSPHL gene expression was detected using either the P1 or P2 primer pairs for RT-PCR, although any other primer pairs specific to PSPHL can be used. The RT-PCR product can be detected by gel electrophoresis or the PCR primers may be fluorescently labeled for detection by any method for measuring fluorescence that is known
15 in the art. Also anticipated are primer pairs that differentially amplify the different PSPHL expression products such as the P2 primer pairs described herein.

The genome-wide gene expression analysis on surgical human prostate specimens revealed differences in the mRNA expression of the PSPHL (Phosphoserine Phosphatase-Like) gene when comparing prostate tumors from patients of African and European descent.
20 The cDNA for this gene was originally isolated from fibroblasts derived from a patient with Fanconi's anemia by cDNA differential display technique and was described as a homologue to L-3-phosphoserine-phosphatase (PSPH) (Planitzer 1998). It was found that the expression of the PSPHL gene was completely shut down in approximately 70% of prostate tumors from patients of European descent, but expressed at readily detectable levels in 80- 90% of prostate
25 tumors from patients of African descent. This observation was recently confirmed by another group in a published study that examined expression differences between prostate tumors derived from patients of European and African populations (Wallace 2008). Most recently, a detailed follow-up genomic analysis revealed that the expression status of the PSPHL gene is 100% concordant with the presence or absence of an insertion allele at the PSPHL gene locus
30 in the assayed genomes (i.e., the corresponding germ line DNA). The insertion sequence is absent in the genomes of individuals with negative prostate PSPHL expression and also absent in the current version of the assembled reference human genome. Subsequently, the insertion/deletion breakpoints based on TRACE DNA sequence data were defined, and a robust genotyping assay was developed. Genotyping results in populations of different

geographical ancestries validated the unusual pattern of population differentiation of this novel structural variation. The studies described herein are based on this novel finding.

Example 8. Differential *PSPHL* Gene Expression in Prostate Tumors from African

Americans Compared to those from European Americans

In studies previously funded partially by the Howard/Hopkins Partnership (see attached letter from Dr. William Nelson), two sets of expression data were generated using two different microarray platforms to analyze expression differences in human prostate cancer specimens from patients of different geographic ancestries. Surgical prostate tissue specimens representing African Americans were procured and processed. The first expression dataset using surgical prostate TUMOR tissues prepared from 12 European American cases and 9 African American cases. As shown in the “volcano” plot in Figure 21A, genes differentially expressed between the two tumor groups can be identified by applying cut-off values of fold expression change and empirically derived p values. Using cut-off criteria of expression changes greater than 2 fold (in either directions) and $p < 0.05$, six genes were found to be under-expressed in African American prostate tumors (red dots at the top left side), while only one gene, PSPHL (Phosphoserine Phosphatase-Like), was shown over-expressed in African American prostate tumors. The second expression dataset consists of 13 African American prostate cancer cases and 13 European American cases, each represented by paired NORMAL and TUMOR samples, and profiled using an in-house manufactured 20K cDNA microarray that we described in a previous publication (Dunn 2006). As shown in the extracted “heatmap” for the PSPHL gene (Figure 21B), 11 out of 13 African American prostate tissue pairs expressed visually apparent high expression relative to a BPH common reference, while 9 out of 13 European American prostate tissue pairs expressed background levels of PSPHL similar to that in the BPH reference sample.

Example 9. Cloning, sequencing, and mapping of the PSPHL transcripts

The PSPHL (also named CO9) cDNA was initially cloned from fibroblasts of a Fanconi’s anemia patient, sequenced, and described as a homologue to L-3-phosphoserine-phosphatase (PSPH) (Planitzer 1998). In that original study the PSPHL transcript was also detected in Raji cells, Burkitt lymphoma cells from an African American donor (Planitzer 1998). Published mRNA sequence (AJ001612) suggested a coding region of 216 bp (72 amino acids) with partial N-terminal homology to PSPH (Planitzer 1998). Due to lack of

follow-up confirmatory studies, the accuracy of the published mRNA sequences was confirmed. 5' RACE (rapid amplification of cDNA ends) was performed using mRNA samples from prostate tissues positive for PSPHL mRNA expression. RACE products were subcloned and sequenced. Sequences were obtained for two alternative transcripts, termed
5 variant 1 and variant 2 (Figure 22), which differed by a contiguous stretch of 122bp sequence (shown in yellow). This 122bp was later confirmed as a single alternatively spliced exon (see Figure 23B). The sequence of variant 1 matched GenBank accession number AJ001612 over a contiguous stretch of sequence that included the full open reading frame (72 amino acids) (Figure 22). An antibody to this variant has been reported and used in a recent study (Kuo
10 2007). The sequence of variant 2 matched GenBank accession number BC065228, also over a contiguous stretch of sequence that included an open reading frame (91 amino acids) (Figure 22). The two open reading frames (marked with colored bars, Figure 22) started from the same consensus starting site but differed at the site of the spliced 122bp exon (Figure 22). In order to define the exon structure of these sequences, the two variant sequences were
15 blasted against the reference human genome. nt 1-326 of the variant 1 sequence (red) and nt 1-447 of the variant 2 sequence (red and yellow) do not match the genomic DNA sequences in the reference human genome assembly (HG1 8). This unexpected finding strongly suggested that part of the PSPHL gene is deleted in the genome represented in the reference assembly, as there is no sequence gap in this region. Therefore, despite the previous cloning
20 and our validation of the PSPHL cDNA, the complete coding sequence is absent in the corresponding genomic locus. The lack of information regarding the PSPHL gene structure, combined with the fact that the transcripts may not be detectable in many samples particularly those derived from individuals of European descent (Figure 21), may have contributed to the lack of literature on this gene and its expression products other than the
25 three aforementioned studies (Planitzer 1998, Kuo 2007, Wallace 2008).

Example 10. The presence or absence of a PSPHL DNA segment in the genome determines expression status of the PSPHL gene in prostate tissue.

The results presented herein suggest that the expression of the PSPHL gene in the
30 human prostate was associated with a potential deletion/insertion variation, one form of structural DNA variation, in the human genome. Next a total of 24 prostate cancer cases were identified from our banked fresh frozen prostate tumors, for which genomic DNA samples have also been prepared previously using blood or seminal vesicle samples from the matched

individuals. Samples from African Americans (n=12) were intentionally enriched. For the genomic DNA samples, PCR was performed to detect the presence or absence of the DNA sequences corresponding to AJ001612, part of the segment that was absent in the reference genome. For the prostate tumor samples, RNA was extracted and RT-PCR was performed to detect the PSPHL mRNA. As shown in Figure 23, all nine cases negative for the targeted genomic DNA were also negative for the PSPHL transcript, while all of the 15 cases detected positive for the targeted segment of PSPHL genomic DNA were also positive for the PSPHL transcript. These results suggested that a novel structural variation of the PSPHL locus, likely in the form of insertion/deletion (indel), was responsible for the observed expression differences. The observed expression pattern in microarray studies (Figure 21) also suggested that the deletion allele was the predominant allele in European Americans while the insertion allele was the predominant allele in African Americans. All these speculations were confirmed by the subsequent identification and confirmation of the insertion/deletion breakpoints as well as the genotyping results in 3 populations (see Figures 23A, 24 below).

Example 11. *In silico* assembly of the PSPHL genomic sequence and development of a genotyping assay

Next, *in silico* analyses of genomic DNA sequences in the Trace Archive and other genomic sequences derived from whole genome shotgun sequencing was performed. Partial assembly of the insertion allele was possible, and the exon structure of the gene was defined, as well as the 5' and 3' breakpoints of the insertion/deletion locus. Figure 23B showed a schematic diagram comparing the insertion allele with the deletion allele (not to scale). Coordinates for the assembled reference genome, which is represented by the deletion allele, were marked below the line positions. Sizes of the exons and introns were marked above their respective positions in the insertion allele. The 5 prime breakpoint is chr7 55798228 in the assembled human genome (deletion allele), and separated by 9bp (GTGCGTCTA) from the 3 prime break point at Chr7 55798238. However, the sizes and sequences of the remaining gaps are still not known. Mapping of the break points however allowed the design of a multiplexed genotyping assay for the PSPHL locus using primer sets (P3 and P4) at their respective sites marked with red vertical bars. Representative genotyping results in African Americans and European Americans were also shown. The genotyping assay differentiates the three different genotypes, each with 0 copy (homozygous deletion, detected by single lower band for the smaller amplicon, a predominant genotype in European Americans), 1 copy (heterozygous, detected by the presence of both amplicons), and 2 copies

(homozygous insertion, detected by a single upper band for the larger amplicon, a predominant genotype in African Americans) of the insertion allele. Because the reference genome sequence is from a deletion allele, previous genome wide analysis of structural variation that depended on the reference genome for experimental design and analysis would not detect this novel structural variation.

Example 12. PSPHL genotyping in 3 populations

Next, the above mentioned genotyping assay was used to type an expanded set of genomic DNA samples derived from Yoruba trios (purchased from Coriell Institute for Medical Research), African Americans (AA), and European Americans (EA). Figure 24 summarizes the results from an initial typing of 999 genotypes. More than 96% of African Americans (n=335) and 100% of 30 Yoruba trios (n=90 genotypes) have at least one copy of the insertion allele (insertion/insertion and insertion/deletion), therefore should express the PSPHL gene, at least in the prostate, according to Figure 23. In sharp contrast, approximately 70% European Americans (n=574) are homozygous deletion (deletion/deletion) genotype. The allele frequency distributions are in accordance with Hardy Weinberg Equilibrium in each of the 3 populations (data not shown) and consistent with Mendelian pattern of inheritance in the Yoruba trios (data not shown), thus further validating the accuracy of the genotyping results. PSPHL genotype frequencies in prostate cancer cases and controls. The distribution of structural variation between populations, like all other forms of genomic variation such as SNPs, is dominated and shaped by the common ancestry of humans in Africa some 50,000 years ago. Rare forms of genomic variations shared among different populations but with high level of population differentiation, such as the PSPHL alleles, may be more likely to be associated with differential risk to diseases as shown in previous such examples (Gonzalez 2005, Stefansson 2005). To determine whether the PSPHL genotype is associated with risk for prostate cancer, 1541 prostate cancer cases were genotyped and 574 controls in the European American population were genotyped. Preliminary analysis did not reveal statistically significant association between the PSPHL alleles and the risk for prostate cancer in the European American population. Parallel genotyping and analysis in prostate cancer cases (n=356) and controls (n=335) from the African American population, however, revealed an interesting trend. As shown in Figure 25, the deletion allele was more frequently detected in cases than in the controls. The frequencies for the deletion/deletion (D/D), deletion/insertion (D/I), and insertion/insertion (I/I) genotypes were 6.46%, 32.58, and 60.96 in prostate cancer case (n=356), and 3.6%, 27.76, and 68.66% in the unaffected controls

(n=335), respectively. Using the additive model, the deletion allele was found to be statistically associated with prostate cancer ($p=0.03$), prior to adjustment for individual ancestry. Further studies will therefore focus on assessment of this potential association using 700 cases and 700 controls following control for population stratification.

5

Example 13. Association of the PSPHL gene and prostate cancer in a case-control African American population

The experiments described herein are aimed at determining the genetic contribution to prostate cancer disparity by further analysis of the PSPHL gene. The gene expression based approaches described herein have resulted in the discovery of a novel structural variation at the PSPHL locus that is tightly linked to gene expression and geographic ancestry. Given the potential importance of such genomic variations in the differential risk for diseases, and important leads from our preliminary data, the experiments described herein address the association and functional link between this novel structural human genome variation and disease risk in Americans of African descent.

In a recent report, a two-gene prostate tumor expression signature was shown to accurately differentiate prostate tumors from African-American and European American patients (Wallace 2008). One of these two genes was PSPHL. This finding was consistent the results described herein. The results described herein demonstrate that the deletion allele of the P SPHL gene was putatively associated with prostate cancer specifically in the African American case -control population (see, e.g. Figure 25), but no significant association was found between prostate cancer and the variants in the European American population. Thus, further experiments as described assess the potential association of the PSPHL genotype with the risk for developing human prostate cancer in an expanded African American case-control cohort. Such association studies present take into account population stratification that can be augmented in the setting of ancestral variants and an admixed population, resulting in spurious associations. However, in a single gene association study, this confounding may be more effectively controlled. Accordingly, controls for population stratification will be applied to mitigate population stratification.

The experiments described herein make use well-established collection of DNA samples from cases and controls from the African American population (described below).

Study population and Case Control Samples

Cases for this study have been drawn from the large number of men undergoing radical prostatectomy (RP) for treatment of prostate cancer in the Department of Urology at

the Johns Hopkins Hospital. A standard tissue processing procedure has been in place that takes portions of the non-cancer seminal vesicles, surgically excised along with the prostate, for storage at -80C and extraction of high quality DNA. Each of these cases is assigned an anonymous tissue bank number and entered into a database that contains all the pathological information from the radical prostatectomy specimen, demographic information including age at surgery and self-reported racial information, and any follow-up information that are updated annually. Of the over 1100 men undergoing this surgery at Hopkins annually, ~6% are African American; over 750 African American men have undergone RP at Hopkins and have seminal vesicle tissues available. Approximately 55% of the African American cases are from either Baltimore or Maryland, with the remainder primarily from the Mid -Atlantic region (VA, DE, WV, NJ, NC). Currently, over 500 DNA samples are being prepared from African American prostate cancer cases and are in various stages of being integrated into the study, and it is anticipated that an additional 200 cases could be available by the time the proposed studies begin as we have intensified our DNA extraction effort from these cases. A summary of information regarding cases from which DNA has already been extracted is presented in Table 1, below.

Table 1

	age	sdev	range	PSA	sdev
	(avg)			(ng/ml)	
controls (n=611)	58.2	9.3	45-92	1.25	0.82
cases (n=514)	56.7	7.1	36-74		
	Gleason Sum	%		Stage	%
	5	1.1		T2	62
	6	52.3		T3A	25.5
	7	37.7		T3B	6.2
	8	5.8		N1	3.4
	9	2.6			

The most common case has org an confined pathologic stage (T2, 60%) and pathologic Gleason Sum 6 (5 2%). Controls for this study have been drawn from men undergoing screenings for prostate cancer in Baltimore and throughout Maryland. Over 600 blood samples from African American men who have serum PSA values below 4ng/ml and ages at screening 45 years or older have been obtained from Dr. Partin and DNA samples has been prepared. An additional 100 samples are to be procured and prepared prior to the start of the proposed study.

Genotyping

The current genotyping assay as described (Figure 23B) is a multiplexed PCR assay based on the partial assembly of the PSPHL locus. Primer set P3 (Forward.

TCAGCTAAAGTGGCTGTTGG GTGT, Reverse. AAGCTTCTGCGCTACCTTGCGA)

(SEQ ID NO:3 and SEQ ID NO: 4) was designed to specifically amplify a 133 bp fragment of the insertion sequence in exon 1 of the PSPHL gene. The detection of the exon DNA

sequence indicates the presence of the insertion allele (Figure 23B). The deletion allele will

be detected by primer set P4 (Forward. A GTCTTGCTA TCTTGCCCA GGCTGA T;

Reverse. GTA GA GA CTGGGTTTCA CCA TGTGTTGG) (SEQ ID NO: 5 and SEQ ID NO:

6) that would generate an amplicon only if the insertion sequence is absent. For genotyping

quality control, we will test the Hardy-Weinberg proportions (HWP). Maximum likelihood

estimates of allele frequencies will be tested for departures from HWP using the chi-square

goodness of fit tests.

Data analysis

One concern regarding the use of an ancestry associated genotype for association study is that it may lead to false positive associations, in which allele frequency differences detected among cases and controls are simply the result of unequal distributions of subgroups

between the cases and controls. This confounding due to population stratification can be

minimized if the cases and controls are drawn from the same genetic group. However, it is

difficult to genetically classify individuals accurately based solely on self-report. The

proposed study will incorporate information on ancestry proportions and population

substructure by matching on ancestry. The study will generate data of a panel of 120

Ancestry -informative Markers (AIMs) for the set of prepared DNA samples as part of his

DOD-funded project not related to this proposal. Individual ancestral proportions estimated

from STRUCTURE using these AIMs will be used as a covariate in the analysis. Each

association test will be adjusted for covariates such as age and individual ancestral

proportions. Descriptive statistics will be performed using SAS statistical package (Version

9.1.3; SAS Institute Inc., Cary, NC). Genotype and allele frequencies will be estimated by

gene counting using the SAS statistical package. Odds ratios, 95% confidence intervals and P

values will be determined by logistic regression analyses from comparisons of genotype

between individuals with prostate cancer and healthy controls with adjustment for age and

ancestry as covariates, using SAS statistical package (Version 9.1.3; SAS Institute Inc., Cary,

NC). Association will be tested under the assumption of a dominant, additive and recessive

model of inheritance. Probability of significance for all analyses will be set at $P < 0.05$.

While multiple genes in a causal pathway, working additively or interactively together to increase an individual's risk, is one scenario for prostate carcinogenesis, it is also possible that a single gene can impose a risk of prostate cancer. The results presented in Figure 25, though not adjusted for individual ancestry, suggest a high likelihood of a positive association between the PSPHL deletion allele with prostate cancer in the African American population. Further studies as described can be used to confirm the positive association, an subsequently genome wide structural variation analysis in the same samples could help to identify multiple variants that has cumulative effects when combined. Therefore, the significance of the study is not limited to the targeted PSPHL gene only. Rather, technically feasible means are used to address the question of genetic contribution to prostate cancer disparity that has the potential to be translated to better control of prostate cancer in the African American and the general population.

Example 14. A functional link between the PSPHL gene and prostate cancer development and progression

Detailed functional studies on genes known to be functionally different among populations of different geographic ancestry constitute an alternative approach for investigating genetic contributions to prostate cancer disparity. For example, individuals homozygous for the inactivating DARC (Duffy Antigen Receptor for Chemokines) variant (Tournamille 1995) do not express the DARC protein, leading to the Duffy null phenotype that is prevalent in West Africa. While loss of DARC function has been functional linked to resistance to malaria (Hamblin 2002), recent data also implicated its role in prostate cancer progression. The precise mechanism of function may involve the molecular interaction between DARC and a tumor suppressor named KAI 1 that leads to subsequent cancer cell senescence and inhibition of a critical step in the metastatic cascade: extravasation (Bandyopadhyay 2006). These functional data suggest that the inactivation of DARC as seen in individuals of African descent, while protective for malaria, may allow for prostate cancer spread/metastasis due to the absence of interactions between KAI1 and DARC. In this aim, we will investigate the potential functional contribution of the PSPHL variants in human prostate cancer using two complementing approaches.

Experiments will be carried out to determine the gene expression correlates of the different PSPHL genotypes using whole genome expression analysis in paired normal and tumor prostate tissues derived from both African and European Americans. In vitro genetic manipulation of the gene in prostate cancer cell lines will be performed to interrogate their

putative functional roles. Together, the results will determine whether P SPHL is functionally linked to prostate cancer and whether its expression correlates are altered during the prostate cancer development and progression.

Prostate samples for gene expression analysis

- 5 Samples representing each genotype category for whole genome expression analysis as shown in the Table 2 will be collected.

Table 2

	European American		African American		Total #samples
Tissue Type	Normal	Tumor	Normal	Tumor	
I/I Genotype	I/I	I/I	I/I	I/I	40
I/D Genotype	I/D	I/D	I/D	I/D	40
D/D Genotype	D/D	D/D	D/D	D/D	40
# samples	30	30	30	30	120

* I/I: homozygous insertion; I/D: heterozygous insertion/deletion; D/D: homozygous deletion.

- 10 In one experiment, prostate tumor RNA samples from 214 cases of prostate cancer specimens, about half of which were paired with normal appearing epithelial tissues from matched cases are collected. The majority of the 214 RNA cases can be matched to DNA samples based on tumor bank number, allowing us to genotype the corresponding individual before using these existing RNA samples. Identification of the required number of cases (n=10) may be a
 15 challenge for two of the genotype groups, African American cases with deletion/deletion genotype and European American case with the insertion/insertion genotype, owing to relative small percentage of these specified genotypes in the corresponding populations (Figure 23B).

Gene expression analysis

- 20 One possibility is that the different PSPHL genotypes are correlated to the expression signature of a set of genes in addition to PSPHL gene itself. Using published microarray data (Wallace 2008) as well as microarray data generated by these inventors, the results suggested coordinated gene expression changes as a function of the PSPHL genotype in African
 25 American cases but not in European American cases. The Agilent whole human genome oligo arrays may be used to systemically analyze expression correlates of the PSPHL variants in paired normal and tumor samples from both African and European American cases. The Agilent array contains 44k 60mers in the sense orientation and over 95% of the probes target regions within 600 bp of 3' ends of the transcripts. Hybridization will be carried out using the conditions specified in the Agilent system, using the two-color design in which all test
 30 samples will be compared to a single reference samples with the D/ITM genotype. Following

image analysis, processed ratio data will be imported to GeneSpring(Agilent Technologies) for further processing, visualization, and analysis. Both supervised and unsupervised approaches will be used to analyze the data, with the primary goals to (1) identity a set of genes that are correlated to the PSPHL genotype and PSPHL gene expression; (2) assess the concordance of the correlated genes in both normal tissues and cancer tissues, and in both African and European cases; (3) assess whether the correlated genes contribute to the differences seen in normal and tumor tissues. Detailed analytical approaches are not presented here due to space limitations yet our previous studies and publications have demonstrated our expertise in gene expression analysis.

Cell line based functional studies

Recombinant antigens corresponding to the open reading frames presented in Figure 22 have been generated. Polyclonal antibody against one of the spliced variants (variant 1) has been made.

The synthetic PSPHL protein sequence used to generate the PSPHL antibody is shown below, and corresponds to SEQ ID NO: 7:

SEQ ID NO: 7

MASASCSPGGALASPEPGRKILPRMISHSELRKLFYSADAVCFDVDSTVISEEG
IGCFHWIWRKCDQATSQG

This polyclonal antibody has been used to detect the 72AA antigen in prostate cells with D/D genotype (REPE-2) and transfected with cDNAs corresponding to the ORF region (ORFonly) or the full length cDNA (FLcDNA) (ORF plus the 5' and 3' untranslated regions) as shown in the inserted western blot. Monoclonal antibodies can also be used. The antibodies will first be screened using the recombinant antigen and then used to detect the PSPHL protein in cell lines and tissue samples with known PSPHL genotypes. It is possible that, if the protein is present in low abundance and the antibody affinity is not high enough, the antigen will be concentrated through immunoprecipitation in order for detection in biological specimens. Gene manipulation followed by functional analysis are then performed, to assess whether experimental manipulation of PSPHL gene expression levels will have biological consequences in the form of altered cell proliferation, cell migration, survival, and adaptation to stress. The genotypes of commonly used human prostate cancer cells have been characterized(data not shown). Experimental depletion of PSPHL expression can be readily achieved using routine stable gene knockdown techniques in MDA -PCa-2b cells (D/I

genotype), an AR positive cell line derived from an African American donor. The PSPHL protein will be expressed in E006AA cells, also a prostate cell line derived from an African American donor but with the rare D/D genotype, followed by the same functional assays. The effect of altered expression levels of P SPHL in the stable clones will be examined using a suite of assays including MTS (Promega) assay to examine overall cell growth, the cell adhesion assay, the anoikis resistance assay (BD Biosciences) to examine the effect of PSPHL on anchorage-independent survival, as well as Annexin V-FITC kit (BD Biosciences) based apoptosis assays to examine the effect of PSPHL in apoptosis. In addition, global gene expression changes will be examined and validated by quantitative RT-PCR following altered P SPHL expression and establishment of the stable clones.

Expected results and alternative approaches

By inclusion of multiple groups of samples including paired normal and tumor samples from both African and European American cases the functional read-out of the PSPHL genotypes in the form of altered expression profiles will be assessed. It is possible that expression differences shown among the different genotypes in African Americans may not be mirrored in prostate specimens from European Americans, and that genes differentially expressed as a function of the genotype in tumor samples may not be replicated in the corresponding normal samples. The results will be utilized to provide functional insight regarding the differential impact of the genotype in different tissue histologies (i.e., normal or tumor) and different risk groups (i.e., European American and African American). The signature set of genes can be further analyzed by using the OntoExpress software (Khatri, 2002) or Gene Set enrichment analysis (Subramanian 2005) or other gene ontology tools and pathway analysis tools to identify key gene expression changes in specific functional categories. From the function of these concordant and discordant genes we may be able to draw mechanistic inferences regarding context-dependent genetic contribution of the PSPHL genotype to the development of prostate cancer. This analysis will also determine whether there is any expression change, within cases with the same genotypes, in the transition from normal prostate epithelium to prostate cancer. The PSPHL transcript levels were markedly different between normal and tumor samples derived from the same population as recently reported (Wallace 2008). Although our microarray data did not show difference between normal and cancer samples (Figure 2), the definitive assessment could be best achieved by follow-up RT-PCR due to limitations of microarray analysis when a D/D genotype (negative expression) was used to generate expression ratios. The results from the tissue based studies can be further corroborated by those from cell lines based studies and functional alterations as

a result of manipulation of the PSPHL expression levels. Results from these efforts will provide mechanistic insights into the functional role of PSPHL in prostate cancer disparity.

Other Embodiments

5 From the foregoing description, it will be apparent that variations and modifications may be made to the invention described herein to adopt it to various usages and conditions. Such embodiments are also within the scope of the following claims.

10 The recitation of a listing of elements in any definition of a variable herein includes definitions of that variable as any single element or combination (or subcombination) of listed elements. The recitation of an embodiment herein includes that embodiment as any single embodiment or in combination with any other embodiments or portions thereof.

 All patents and publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent patent and publication was specifically and individually indicated to be incorporated by reference.

REFERENCES

- Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Robertson-Lowe C, Marshall AJ, Petretto E, Hodges MD, Bhargal G, Patel SG, Sheehan-Rooney K,
 5 Duda M, Cook PR, Evans DJ, Domin J, Flint J, Boyle JJ, Pusey CD, Cook HT. Copy number polymorphism in *Fcgr3* predisposes to glomerulonephritis in rats and humans. *Nature*. 2006 Feb 16;439(7078):851-5.
- Bandyopadhyay S, Zhan R, Chaudhuri A, Watabe M, Pai SK, Hirota S, Hosobe S, Tsukada T, Miura K, Takano Y, Saito K, Pauza ME, Hayashi S, Wang Y, Mohanta S, Mashimo T,
 10 Iizumi M, Furuta E, Watabe K. Interaction of KAI1 on tumor cells with DARC on vascular endothelium leads to metastasis suppression. *Nat Med*. 2006 Aug;12(8):933-8
- Chu KC, Tarone RE, Freeman HP. Trends in prostate cancer mortality among black men and white men in the United States. *Cancer*. 2003 Mar 15;97(6):1507-16
- Clegg L, Li FP, Hankey BF, Chu K, Edwards BK. Cancer survival among US whites and
 15 minorities: a SEER (Surveillance, Epidemiology, and End Results) Program population-based study. *Arch Intern Med*. 2002;162(17):1985-93.
- Conrad D.F., and Hurles M.E. The population genetics of structural variation. *Nature Genetics* 39 (7s): S30-S36, 2007.
- Dunn T., Chen S., Faith D., Hicks J., Platz E. , Chen Y., Ewing C., Sauvageot J., Isaacs W.,
 20 De Marzo A., Luo J. A Novel Role of Myosin VI in Human Prostate Cancer. *American Journal of Pathology*. 169(5): 1843-54. 2006.
- Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK, Mulholland S, Leongamornlert DA, Edwards SM, Morrison J, Field HI, Southey MC, Severi G, Donovan JL, Hamdy FC, Dearnaley DP, Muir KR, Smith C, Bagnato M, Arden-Jones AT, Hall AL,
 25 O'Brien LT, Gehr-Swain BN, Wilkinson RA, Cox A, Lewis S, Brown PM, Jhavar SG, Tymrakiewicz M, Lophatananon A, Bryant SL; UK Genetic Prostate Cancer Study Collaborators; British Association of Urological Surgeons' Section of Oncology; UK ProtecT Study Collaborators, Horwich A, Huddart RA, Khoo VS, Parker CC, Woodhouse CJ, Thompson A, Christmas T, Ogden C, Fisher C, Jamieson C, Cooper CS, English DR, Hopper

JL, Neal DE, Easton DF. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet.* 2008 Mar;40(3) :316-21.

Freedland SJ, Isaacs WB. Explaining racial differences in prostate cancer in the United States: sociology or biology? *Prostate.* 2005 Feb 15;62(3):243-52.

- 5 Gonzalez, E. et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307, 1434–1440 (2005).

Hamblin, M. T.; Thompson, E. E.; Di Rienzo, A. Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet.* 70: 369-383, 2002.

- 10 Jakobsson J, Ekström L, Inotsume N, Garle M, Lorentzon M, Ohlsson C, Roh HK, Carlström K, Rane A. Large differences in testosterone excretion in Korean and Swedish men are strongly associated with a UDP-glucuronosyl transferase 2B17 polymorphism. *J Clin Endocrinol Metab.* 2006 Feb;91(2): 687-93.

Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. Cancer statistics, 2007. *CA Cancer J Clin.* 2007 Jan-Feb;57(1):43-66.

- 15 Khatri P, Draghici S, Ostermeier GC, Krawetz SA. Profiling gene expression using onto-express. *Genomics* 2002; 79(2):266-270.

- 20 Kittles RA, Chen W, Panguluri RK, Ahaghotu C, Jackson A, Adebamowo CA, Griffin R, Williams T, Ukoli F, Adams-Campbell L, Kwagyan J, Isaacs W, Freeman V, Dunston GM. CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification? *Hum Genet.* 2002 Jun;110(6):553-60

Kuo CH, Miyazaki D, Nawata N, Tominaga T, Yamasaki A, Sasaki Y, Inoue Y. Prognosis-determinant candidate genes identified by whole genome scanning in eyes with pterygia *Invest Ophthalmol Vis Sci.* 2007 Aug;48(8):3566-75.

- 25 Luo, J., Duggan, D.J., Chen, Y., Sauvageot, J., Ewing, C.M., Bittner, M.L., Trent, J.M., and Isaacs, W.B. Human Prostate Cancer and Benign Prostatic Hyperplasia: Molecular Dissection by Gene Expression Profiling. *Cancer Research* 61(12): 4683-4688, 2001.

Luo, J., Zha, S., Gage, W.R., Dunn, T.A., Hicks, J., Bennett, C.J., Ewing, C.M., Platz, E.A., Ferdinandusse, S., Wanders, R.J., Trent, J.M., Isaacs, W.B., and De Marzo, A.M. -

Methylacyl-CoA Racemase: A New Molecular Marker for Prostate Cancer. *Cancer Research* 62(8): 2220-2226, 2002.

Park J, Chen L, Ratnashinge L, Sellers TA, Tanner JP, Lee JH, Dossett N, Lang N, Kadlubar FF, Ambrosone CB, Zachariah B, Heysek RV, Patterson S, Pow-Sang J. Deletion

5 polymorphism of UDP-glucuronosyltransferase 2B 17 and risk of prostate cancer in African American and Caucasian men. *Cancer Epidemiol Biomarkers Prev.* 2006 Aug; 15(8): 1473-8.

Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revenga L, Tran CW, Scheffer

A, Steinfeld I, Tsang P, Yamada NA, Park HS, Kim JI, Seo JS, Yakhini Z, Laderman S, Bruhn L, Lee C. The fine-scale and complex architecture of human copy-number variation.

10 *Am J Hum Genet.* 2008 Mar;82(3):685-95.

Planitzer SA, Machl AW, Rueckels M, Kubbies M. Identification of a novel c-DNA overexpressed in Fanconi's anemia fibroblasts partially homologous to a putative L-3-phosphoserine-phosphatase. *Gene.* 1998 Apr 1 4;2 10(2) :297-306.

Ries LAG, Melbert D, Krapcho M, Mariotto A, Miller BA, Feuer EJ, Clegg L, Horner MJ,

15 Howlader N, Eisner MP, Reichman M, Edwards BK (eds). *SEER Cancer Statistics Review, 1975-2004*, National Cancer Institute. Bethesda, MD, 2007.

Sebat J Major changes in our DNA lead to major changes in our thinking. *Nature Genetics* 39 (7s): S3-S5, 2007

Stefansson, H. et al. A common inversion under selection in Europeans. *Nat. Genet.* 37, 129–

20 137 (2005).

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005; 102(43):15545-15550. Thomas G, Jacobs KB,

Yeager M, Kraft P, Wacholder S, Orr N, Yu K, Chatterjee N, Welch R, Hutchinson A,

25 Crenshaw A, Cancel-Tassin G, Staats BJ, Wang Z, Gonzalez-Bosquet J, Fang J, Deng X, Berndt SI, Calle EE, Feigelson HS, Thun MJ, Rodriguez C, Albanes D, Virtamo J, Weinstein S, Schumacher FR, Giovannucci E, Willett WC, Cussenot O, Valeri A, Andriole GL, Crawford ED, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover R, Hayes RB, Hunter DJ,

Chanock SJ. Multiple loci identified in a genome - wide association study of prostate cancer. *Nat Genet.* 2008 Mar;40(3):310 -5.

Tournamille, C.; Colin, Y.; Cartron, J. P.; Le Van Kim, C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals.

5 *Nature Genet.* 10: 224-228, 1995.

Wallace TA, Prueitt RL, Yi M, Howe TM, Gillespie JW, Yfantis HG, Stephens RM, Caporaso NE, Loffredo CA, Ambs S. Tumor immunobiological differences in prostate cancer between African-American and European-American men. *Cancer Res.* 2008 Feb 1 ;68(3):927-36.

10

Zheng SL, Sun J, Wiklund F, Smith S, Stattin P, Li G, Adami HO, Hsu FC, Zhu Y, BŠlter K, Kader AK, Turner AR, Liu W, Bleecker ER, Meyers DA, Duggan D, Carpten JD, Chang BL, Isaacs WB, Xu J, Gršnberg H. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med.* 2008 Feb 28;358(9):910-9. Epub 2008 Jan 16

What is claimed is:

1. A method of detecting the presence or absence of a nucleic acid segment in the (Phosphoserine phosphatase-like) *PSPHL* gene locus of a subject, wherein the presence or
5 absence of the nucleic acid segment in the gene locus indicates an altered risk of cancer.
2. The method of claim 1, wherein the cancer is prostate cancer.
3. The method of claim 1, wherein the presence or absence of the nucleic acid segment
10 in the *PSPHL* gene locus is detected in an African American subject.
4. The method of claim 1, wherein the absence of the nucleic acid segment indicates an increased risk of prostate cancer in the African American subject.
- 15 5. The method of claim 1, wherein the nucleic acid segment comprises 133 base pairs of exon 1 of human *PSPHL* mRNA encoded by GenBank Accession No. AJ001612 corresponding to SEQ ID NO: 1.
6. The method of claim 1, wherein the nucleic acid segment comprises SEQ ID NO: 2.
20
7. The method of claim 1, wherein the nucleic acid segment comprises SEQ ID NO: 13.
8. The method of claim 1, wherein the nucleic acid segment comprises SEQ ID NO: 14.
- 25 9. The method of claim 1, wherein the nucleic acid segment comprises SEQ ID NO: 15.
10. The method of claim 1, wherein the presence of the insertion allele of the *PSPHL* gene locus is correlated with the expression of the *PSPHL* gene product.
- 30 11. The method of claim 1, wherein the absence of the insertion allele of the *PSPHL* gene locus is correlated with the absence of the *PSPHL* gene product.
12. The method of claim 1, wherein the deletion allele is associated with the expression of a set of genes.

13. The method of claim 1, wherein the subject is homozygous for a deletion in the *PSPHL* gene locus.

5 14. The method of claim 1, wherein the subject is heterozygous for a deletion in the in the *PSPHL* gene locus.

15. The method of claim 13, wherein the homozygous deletion allele is associated with the expression of a set of genes.

10

16. The method of claim 14, wherein the heterozygous deletion allele is associated with the expression of a set of genes.

17. The method of claim 10, wherein the expression of the *PSPHL* gene product is
15 associated with the expression of a set of genes.

18. A method of determining the ancestry of a subject comprising detecting the presence or absence of a nucleic acid segment in the *PSPHL* gene locus in a sample from a subject, wherein the presence or absence of the variation indicates the ancestry of the subject.

20

19. The method of claim 18, wherein the presence or absence of a nucleic acid segment is indicative of African American or European American ancestry.

20. The method of claim 18, wherein the absence of the nucleic acid segment identifies
25 the subject as having an African American ancestry.

21. The method of any one of claims 1 – 20, further comprising selecting subjects with an increased risk of developing prostate cancer.

30 22. The method of claim 21, further comprising obtaining a sample from the subjects.

23. A biomarker for prostate cancer in an African American subject comprising an insertion in the *PSPHL* gene locus, wherein the presence of the biomarker is correlated with a decreased risk of prostate cancer.

24. The biomarker of claim 23, wherein the insertion encodes a nucleic acid comprising 133 base pairs of exon 1 of human PSPHL mRNA encoded by GenBank Accession No. AJ001612 corresponding to SEQ ID NO: 1.

5

25. The biomarker of claim 23, wherein the insertion encodes a nucleic acid comprising SEQ ID NO: 2.

10

26. The biomarker of claim 23, wherein the absence of the biomarker is correlated with an increased risk of prostate cancer in the African American subject.

27. The biomarker of claim 23, wherein the presence of the insertion in the PSPHL gene locus is correlated with the expression of the PSPHL gene product.

15

28. The biomarker of claim 23, wherein the insertion allele is associated with the expression of a set of genes.

20

29. A method of identifying a subject at risk for developing prostate cancer comprising: detecting the presence or absence of a nucleic acid segment in the *PSPHL* gene locus of a subject to determine the genotype of the subject, wherein the absence of the nucleic acid segment in the gene locus indicates an increased risk of prostate cancer.

25

30. A method of determining the prognosis of a patient with prostate cancer comprising: detecting the presence or absence of a nucleic acid segment in the *PSPHL* gene locus of a subject, wherein the absence of the variation determines the prognosis of a patient with prostate cancer.

30

31. The method of claim 30, wherein prognosis determines the course of treatment.

32. The method of claim 29 or 30, wherein the subject is homozygous for a deletion in the *PSPHL* gene locus.

33. The method of claim 29 or 30, wherein the subject is heterozygous for a deletion in the in the *PSPHL* gene locus.

34. The method of claim 29 or 30, wherein the subject is selected from an African
5 American population.

35. The method of claim 29 or 30, wherein the absence of the nucleic acid segment indicates an increased risk of, or risk of recurrence of, prostate cancer.

10 36. The method of claim 29 or 30, wherein the nucleic acid comprises 133 base pairs of exon 1 of human *PSPHL* mRNA encoded by GenBank Accession No. AJ001612 corresponding to SEQ ID NO: 1.

37. The method of claim 29 or 30, wherein the nucleic acid comprises SEQ ID NO: 2.
15

38. The method of claim 29 or 30, wherein the presence of the insertion allele of the *PSPHL* gene locus is correlated with the expression of the *PSPHL* gene product.

39. The method of claim 29 or 30, wherein the absence of the insertion allele of the
20 *PSPHL* gene locus is correlated with the absence of the *PSPHL* gene product.

40. The method of claim 32, wherein the homozygous deletion allele is associated with the expression of a set of genes.

25 41. The method of claim 33, wherein the heterozygous deletion allele is associated with the expression of a set of genes.

42. The method of claim 29 or 30 wherein the presence or absence of a nucleic acid segment in the *PSPHL* gene locus is determined using a polymerase chain reaction (PCR)
30 assay.

43. The method of claim 42, wherein the PCR assay is a multiplexed PCR assay.

44. The method of claim 42, wherein the PCR is carried out using primers comprising the nucleic acid sequences as set forth as SEQ ID NO: 3 and SEQ ID NO: 4 and primers comprising the nucleic acid sequences as set forth as SEQ ID NO: 5 and SEQ ID NO: 6.
- 5 45. The method of claim 44, wherein the nucleic acid sequences as set forth as SEQ ID NO: 3 and SEQ ID NO: 4 amplify a 133 base pair fragment of the insertion sequence in exon 1 of the PSPHL gene.
- 10 46. The method of claim 44, wherein the nucleic acid sequences as set forth as SEQ ID NO: 5 and SEQ ID NO: 6 generate an amplicon only if the insertion sequence is absent.
47. The method of claims 29 or 30, wherein the subject has previously been treated for prostate cancer.
- 15 48. The method of claims 29 or 30, wherein the measurement is performed after surgery or therapy to treat prostate cancer.
49. An antibody to detect PSPHL protein in cells and tissues with PSPHL genotypes.
- 20 50. The antibody of claim 49, wherein the antibody is polyclonal.
51. The antibody of claim 49, wherein the antibody is monoclonal.
52. The antibody of claim 50, wherein the polyclonal antibody is directed to the 72AA
25 antigen of prostate cells corresponding to SEQ ID NO: 7.
53. A kit for use in identifying a subject at risk for developing prostate cancer comprising: primers directed to amplify a 133 base pair sequence of exon 1 of human PSPHL mRNA encoded by GenBank Accession No. AJ001612 corresponding to SEQ ID NO: 1, and
30 instructions for use.
54. The kit of claim 53, wherein the primers comprise the nucleic acid sequences as set forth as SEQ ID NO: 3 and SEQ ID NO: 4.

55. The kit of claim 53, wherein the primers comprise the nucleic acid sequences as set forth as SEQ ID NO: 5 and SEQ ID NO: 6.

56. A kit comprising primers comprising the nucleic acid sequences set forth as SEQ ID NO: 3 and SEQ ID NO: 4, and instructions for use.

57. A kit comprising primers comprising the nucleic acid sequences set forth as SEQ ID NO: 5 and SEQ ID NO: 6, and instructions for use.

58. A kit comprising primers designed against the nucleic acid sequence set forth as SEQ ID NO: 17, and instructions for use.

59. A kit comprising primers designed against the nucleic acid sequence set forth as SEQ ID NO: 18, and instructions for use.

60. A kit comprising primers designed against the nucleic acid sequence set forth as SEQ ID NO: 19, and instructions for use.

61. The kit of any one of claims 54 – 60, further comprising instructions for use in PCR assay.

62. The kit of claim 61, wherein the PCR is multiplexed PCR.

63. A kit for use in identifying a subject at risk for developing prostate cancer comprising: an antibody directed to a PSPHL antigen, and instructions for use.

64. A kit comprising an antibody directed to a PSPHL antigen.

65. The kit of claim 63 or 64, wherein the antibody is monoclonal.

66. The kit of claim 63 or 64, wherein the antibody is polyclonal.

67. The kit of claim 66, wherein the polyclonal antibody is used to detect the 72AA antigen.

68. The kit of claim 67, wherein the polyclonal antibody is directed to a sequence encoded by SEQ ID NO: 7.

PSPHL Gene Expression is Higher in African American Prostate Cancer Tissues When Compared to European American Prostate Tissues

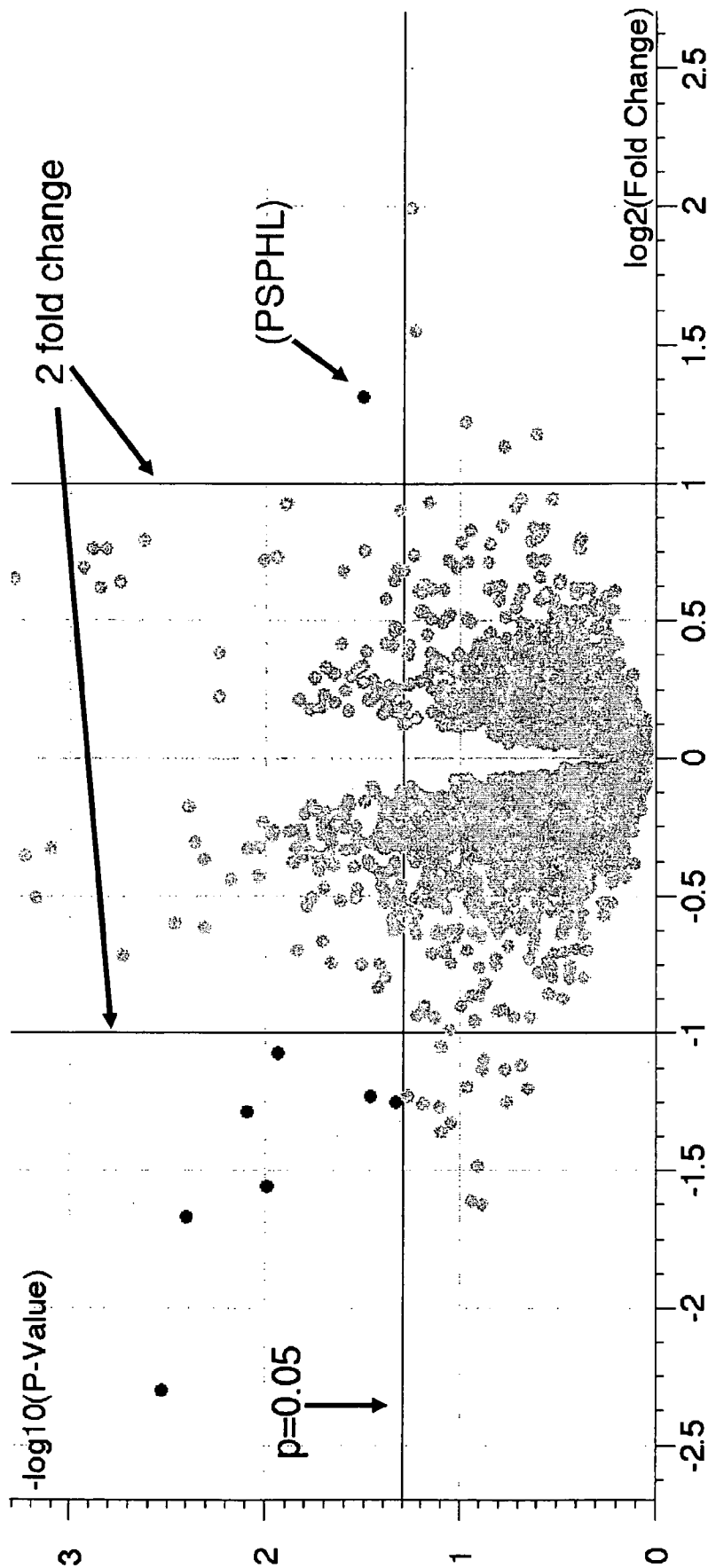


Figure 1: Expression microarray analysis of prostate cancer tissues by race.

Surgical prostate cancer tissues from 12 European American cases were compared to those from 8 African American cases. In this "volcano" plot, each gene is represented by a dot, positioned by fold expression change (x axis) and p value (y axis), that were calculated by comparing the expression values between the two racial groups. Genes represented by red dots demonstrated expression changes greater than 2 fold and $p < 0.05$. Using this cut-off criteria (2 fold change and $p < 0.05$), six genes are under-expressed in African American prostate tissue, and only one gene, PSPHL, is over-expressed in African American prostate tissues.

PSPHL mRNA Nucleotide Sequence

```

1  aagccacaggctccctggctggcgctcagctaaagtggctgttgggtgtccgcaggcttct
61  gcctggccgcgcgcctataaagctaccaggagagctttacgacttcccgtcctgcggg
121  aagtggcgggcacgatacgcaaggtagcgcaagcttctcaatggccagcgcagctgca
181  gccccggcggcgcactcgccctcacctgagcctgggaggaatacttccaaggatgatct
241  cccactcagagctgaggaagcttttctactcagcagatgctgtgtgttttgatgttgaca
301  gcacggtcatcagtgaaagaaatcggatgctttcatgtgatttgaggaaatgtgatac
361  agccaacaagtcaaggataaacgcccaaatggatatatcactgattttgtagagctgctggga
421  gaaccggaaataaacatccattgtcatacagctccaacaacttcagatgaatttttac
481  aagttacacagattgatactgtttgcttacaattgcctattacaacttgctataaaaagt
541  tggtaacagatgatctgcactgtcgaagtcaaaactacagttaggaatcctcaaatgtggttt
601  gtttggttttaactgtagttccagtatatatgatcactatcgatttccctggagagtgtt
661  gtaatctgaattctttatgtataattcctagctataattcatacaaaagtgttttaagagtg
721  gagagtcaattaaacacaccttactcttaggaaatatagattcggcagccttcagtgaaatat
781  tgggtttttccctttgggtatgtcaataaaagtattatccatgtgtcagaaaaaa

```

Figure 2: Nucleotide sequence of PSPHL mRNA.

The nucleotide sequence of the PSPHL mRNA (GenBank accession number AJ001612) is shown with the nucleotide numbers on the left. The position of the P1 (underlined) and P2 (double underlined) primer pairs are shown.

RT-PCR validation of PSPHL expression

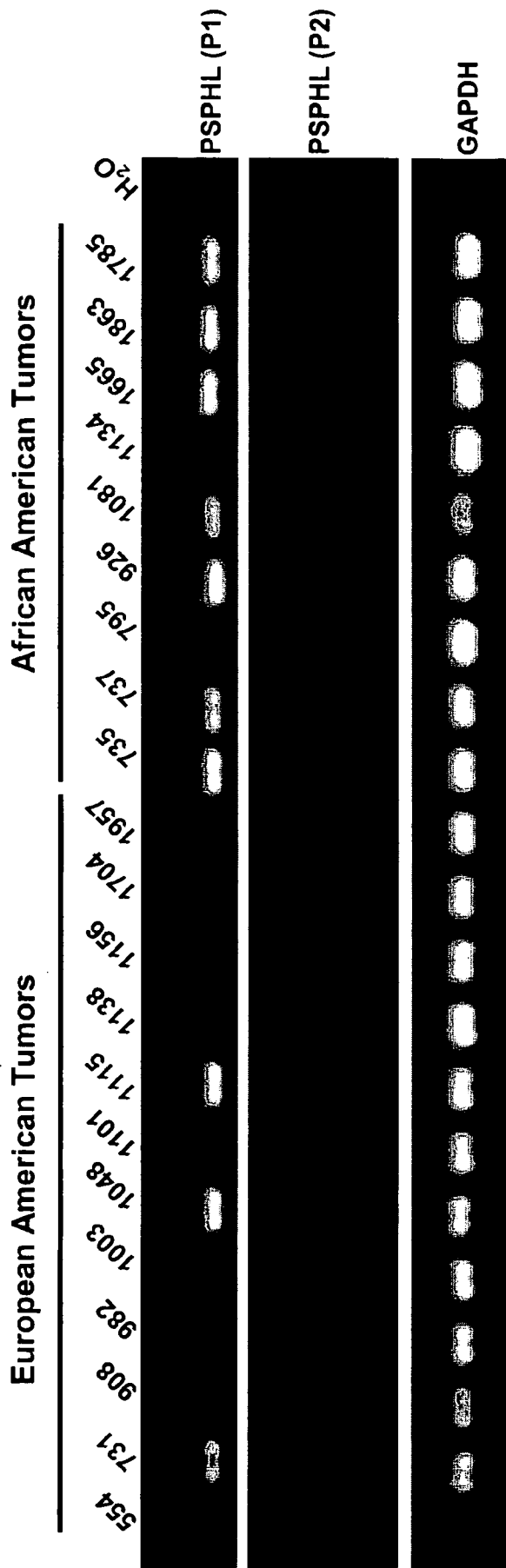


Figure 3: RT-PCR validation of PSPHL expression in the same tissue samples used in expression microarray analysis, with the addition of one more African American Tumor sample.

The PSPHL gene was expressed in 3 of 12 European American prostate cancers and 7 of the 9 African American prostate cancers. Primer sets used in the assays are indicated in the parentheses to the right and annotated below. Note that primer set 2 amplified two products, because the primers spanned exon 3, which is alternatively spliced during gene transcription (see Figure 6). GAPDH is a control for RT-PCR.

P1: GCTGGGAGAACCGGAATACAT (forward), corresponds to nt 414-438 of the AJ001612 sequence,
CCAATATTCACCTGAAGGCTGCCGA (reverse), corresponds to nt 760-783 of the AJ001612 sequence.

P2: CGGTCATCAGTGAAGGAATCGGA (forward), corresponds to nt 304-329 of the AJ001612 sequence; TGGCGTTATCCTTGACTTGTTGCC (reverse), corresponds to nt 362-385 of the AJ001612 sequence.

RT-PCR validation of PSPHL expression in paired normal (N) and tumor (T) prostate specimens revealed a patient-specific expression pattern

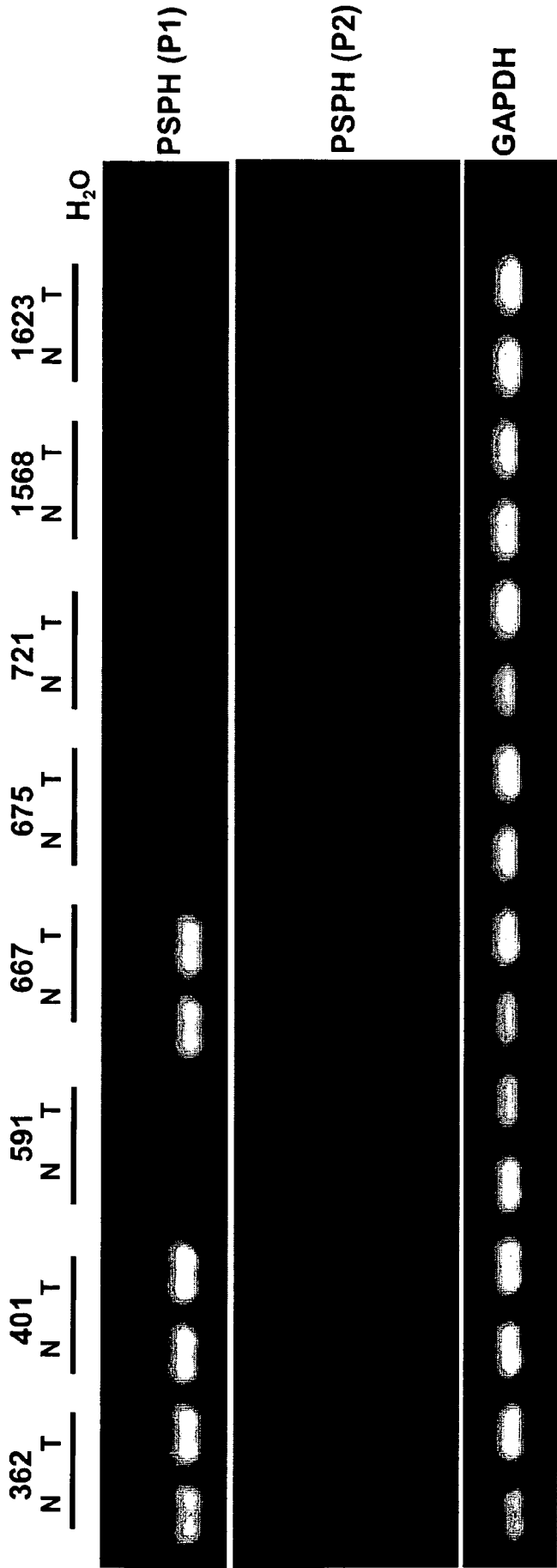


Figure 4: PSPHL expression in paired normal (N) and cancer tissues (T) from 8 European American prostate cancer patients.

The 8 cases are independent of cases used in Figure 1 and 2. Primer sets used were identical to those used in Figure 2. Note the high concordance of PSPHL gene expression between the paired normal and tumor tissues from the same patient.

RT-PCR validation of PSPHL expression in paired normal (N) and tumor (T) prostate specimens revealed a patient-specific expression pattern

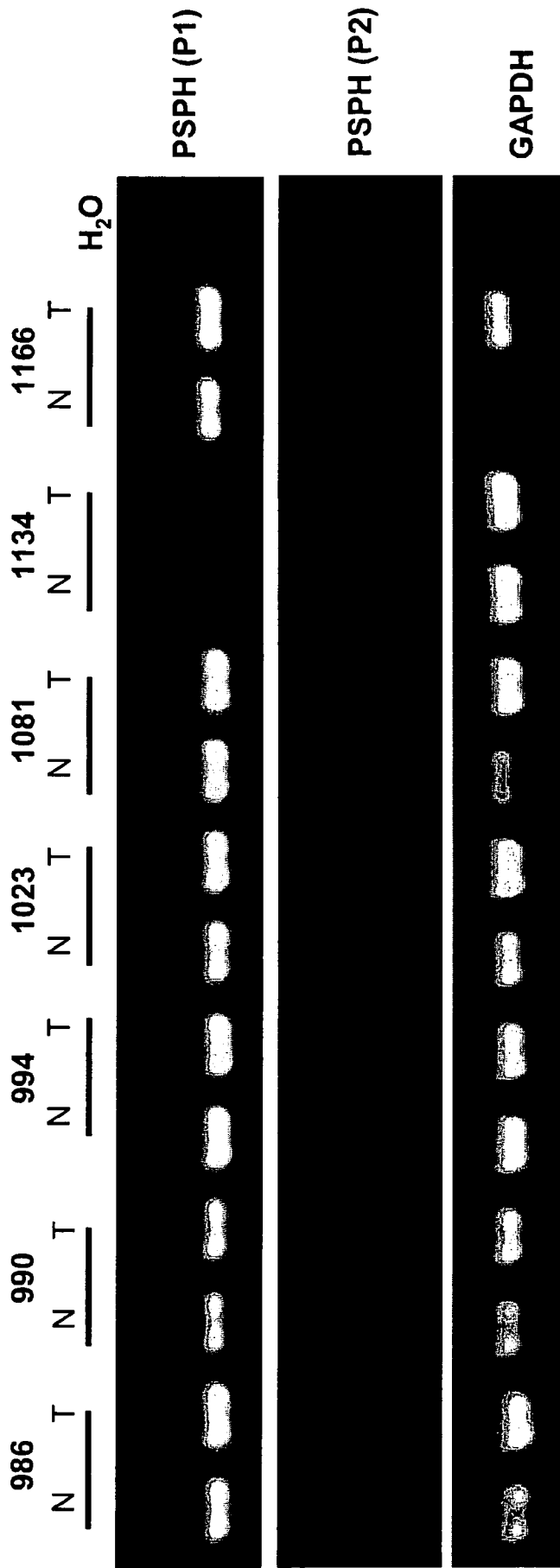


Figure 5: PSPHL expression in paired normal (N) and cancer tissues (T) from 7 African American prostate cancer patients.

Five of the 7 cases are independent of cases used in Figure 1 and 2, while two cases (1081, 1134) overlaps with those used in Figure 2. Primer sets used were identical to those used in Figure 2. Note again the high concordance of PSPHL gene expression between the paired normal and tumor tissues from the same patient.

Alternative PSPHL mRNA Nucleotide Sequence

```

1   ccgcaggcttctgcctggccgcccgcctataagctaccaggaggagctttacgacttc
61  ccgtcctgcgggaagtggcgggcacgatacgcaaggtagcgagaaagcttctcaatggcca
121 gcgccagctgcagccccggcgccactcgccctcacctgagcctgggagaaaaattcttc
181 caaggatgatctccactcagagctgaggaagcttttctactcagcagatgctgtgtgtt
241 ttgatgttgacagcacggtcatcagtgaggaaggaatcGGACGGAGTCTCGCTCTGTAC
301 CAGGCTGGAGTGCAATGGTGCAATCTCGGCTCACTGCACCTCCGCCCTCCCTGGGTTCAGG
361 CAGTCTCCTGCCCTCCACCTCCTGAGTAGCTGAACACTACAggatgctttcattggctttg
421 gaggaatgtgatcaggcaacaaagtcaaggataacgccaatgggtatatcactgatatttg
481 tagagctgctggagaacccggaagaataacatccattgtcatcacagctccaaacaacttc
541 agatgaatttttacaagttacacagattgatactgtttgcttacaattgcctattacaac
601 ttgctatagaaaagttggtacagatgatctgcactgtcaagtaaaactacagtaggaatcc
661 tcaaagattggtttgtttgttttaactgtagttccagtatatatgatcactatttgatt
721 tcctggagagttttgtaatctgaaattctttatgtatatccctagctatatatttcatacaaa
781 gtgttttaagagtgagagtgcaattaaacacctttactcttaggaaaaaaaaaaaaaa
841 a

```

Figure 6: Nucleotide sequence of the alternative PSPHL mRNA.

The nucleotide sequence of an alternative PSPHL mRNA (GenBank accession number BC065228) is shown with the nucleotide numbers on the left. The position of the P2 primer pair is shown double underlined. This sequence has an insertion of 122 bp (shown in capitals and bold) compared to the sequence of AJ001612 (Figure 2), resulting in a PCR product that is 122 bp longer.

Figure 7. Comparison of the human AJ001612 and BC065228 mRNA sequences with the predicted gene from the Chimpanzee genome chromosome 7 contig (GenBank accession number NW 001237953.1).

In the human mRNA sequences, the AAIins sequence is underlined and the position of the P2 primer pair is shown in bold capitals. The predicted sequences of the PSPHL exons are shown in alternating blue and black. Exon 3 of the predicted human gene, which is not present in AJ001612, is shown in red in the BC065228 sequence.

AJ001612 - human mRNA

aagccacaggctccctggctggcgctcagctaaagtggctggtgggtgtccgcaggcttctgcctggccgcc
gccgcctataagctaccaggaggagctttacgacttcccgtcctgcgggaagtggcgggcacgatcgcaag
gtagcgcagaagcttctcaatggccagcgccagctgcagccccggcgcgactcgctcacctgagcctg
ggaggaaaattctccaaggatgatctccactcagagctgaggaagcttttctactcagcagatgctgtg
tgttttgatgttgacagcaCGGTCATCAGTGAAGAAGGAATCGGATgctttcattggatttgaggaaatg
tgatcaGGCAACAAGTCAAGGATAACGCCAaatggtatatcactgattttgtagagctgctgggagaaccg
gaagaataacatccattgtcatacagctccaaacaacttcagatgaatttttacaagttacacagattgat
actgtttgcttacaattgcctattacaacttgctataaaaagttggtacagatgatctgcactgtcaagta
aactacagtttaggaatcctcaaagattgggtttgtttgttttaactgtagttccagttattatgatcact
atcgatttcctggagagttttgtaatctgaattctttatgtatattcctagctatatttcatacaaagtgt
ttaagagtggagagtcattaaacacctttactcttaggaatatagattcggcagccttcagtgaatatt
ggtttttttccctttggtatgtcaataaaagtttatccatgtgtcagaaaaaaaaaaaaa

BC065228 - human mRNA showing Exon 3

cgcaggcttctgcctggccgccgcgcctataagctaccaggaggagctttacgacttcccgtcctgcgg
gaagtggcgggcacgatcgcaaggtagcgcagaagcttctcaatggccagcgccagctgcagccccggcg
cgcactcgctcacctgagcctgggaggaaaattcttccaaggatgatctccactcagagctgaggaagc
ttttctactcagcagatgctgtgtgtttgtgtgttgacagcaCGGTCATCAGTGAAGAAGGAATCGGACgg
agtctcgctctgtcaccaggctggagtgcaatggtgcaatctcggctcactgcaacctccgcctcctgggt
tcaggcagtttctcctgcctccacctcctgagtagctgaaactacaggatgctttcattggctttggaggaa
atgtgatcaGGCAACAAGTCAAGGATAACGCCAaatggtatatcactgattttgtagagctgctgggagaa
ccggaagaataacatccattgtcatacagctccaaacaacttcagatgaatttttacaagttacacagatt
gatactgtttgcttacaattgcctattacaacttgctatagaaagtgggtacagatgatctgcactgtcaa
gtaactacagtttaggaatcctcaaagattgggtttgtttgttttaactgtagttccagttattatgatc
actattgatttcctggagagttttgtaatctgaattctttatgtatattcctagctatatttcatacaaag
tgttttaagagtggagagtcattaaacacctttactcttaggaaaaaaaaaaaaaaaaaaaaa

Chimp Genomic Seq showing Exons 1, 2 3 and 4

GCCACAGGCTCCCTGGCTGGCGGCAGCTAAAGTGGCTGTTGGGTGTCCGCAGGCTTCTGCCTGGCCGTCGC
 CGCCTATAAGCTACCAGGAGGAGCTTTACGACTTCCCGGCCCTGCGGGAAGTGGCGGGCACGATCGCAAGGT
 AGCGCAGAAGCTTCTCAATGGCCAGCGCCAGCTGCAGCCCCCTGCGGCGCACTCGCCTCACCTGAGCCTGGG
 GGAGGAAAATCTTCCAAGGATGATCTCCCACTCAGAGCTGAGGAAGCTTTTCTACTCAGCAGATGCTGTG
 TGTTTTGATGTTGACAGCACGGTCATCAGTGAAGAAGGAATCGGGGACGGAGTCTCGCTCTGTCACCAGGC
 TGGAGTGCAATGGTGCAATCTCGGCTCACTGCAACCTCCGCCCTCTGGGTTCAAGCAATTCTCCTGCCTCC
 ACCTCCTGAGTAGCTGAAACTACAGGATGCTTTTCAATTGGATTGGAGGAAATGTGATCAGGCAACAAGTCA
 AGGATAACGCCAAATGGTATATCACTGATTTTGTAGAGCTGTGGGAGAACCAGAAAGAATAACATCCATTG
 TCATACGGCTCCAAACAACCTTCAGATGAATTTTACAAGTTACACAGATTGATACTGTTTGCTTACAATTG
 CCTATTACAACCTTGATATAGAAAGTTGGTACAGATGATCTGCACTGTCAAGTAACTACAGTTAGGACTCC
 TCAAAGATTGGTTTGTGTTTGTGTTTTTAAGTGTAGTTCCAGTATTATATGATGACTATTGATTTTCTAGA
 GAGTTTGTAACTGAAATCTTTATGTATATTCTAGCTATATTTCATACAAAGTGTTTTAAGAGTGGAGA
 GTCAATTAAACACCTTTACTCTTAGGAATATAGATTGGGCAGCCTTCAATGAATATTGGTTTTTTTCCCTT
 TGGTATGTCAATAAAAGTTTATCCATGTGTCAGA

Sequence of the African American specific Insertion (AAIns)

```

1  aagccacaggctccctggctggcgtcagctaaagtggctgttgggtgtccgcaggcttct
61  gcctggccgccgcctataagctaccaggaggagctttacgacttcccgctcctgcggg
121  aagtggcgggcacgatcgcaaggtagcgcagagaagcttctcaatggccagcgcagctgca
181  gccccggcggcactcgcctcacctgagcctgggaggaataatctccaaggatgatct
241  cccactcagagctgaggaagcttttctactcagcagatgctgtgtgtttgatgtgaca
301  gcacgggtcatcagtgaaagaaggaatcggatgctttcattggatttggaggaatgtgac
361  aggcaacaagtcaaggataaacgcccaaatgggtatatcactgatttttagagctgctggga
421  gaaccggaagaataacatccattgtcatacagctccaaacttcagatgaatttttac
481  aagttacacagattgatactgtttgcttacaattgcctattacaacttgctataaaaagt
541  tggtaacagatgatctgcactgtcaagtaaaactacagtaggaatcctcaagattggttt
601  gtttggttttaactgtagtccagttatatgatcactatcgatttccctggagagtttt
661  gtaatctgaattctttatgtatatccctagctatatattcatacaaaagtgttttaagagtg
721  gagagtcaattaaaacaccttactcttaggaatatagattcggcagccttcagtgaaatat
781  tgggttttttccctttgggtatgtcaataaaaagtttatccatgtgtcagaaaaaa

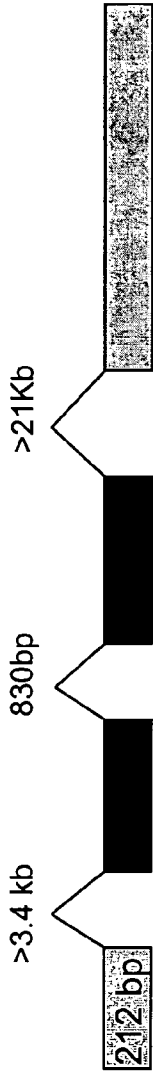
```

Figure 8: Nucleotide sequence of AAIns sequence in the PSPHL mRNA.

The nucleotide sequence of the PSPHL mRNA (GenBank accession number AJ001612) is shown with the nucleotide numbers on the left. The sequence of a 326 bp portion of the African American specific insertion (AAIns) that overlaps with the PSPHL mRNA sequence is shown in bold italics. The position of the P3 primer pair is shown heavy underlined and they amplify a product of 133 bp.

The Predicted PSPHL Gene Structure

A



B



Figure 9A: The predicted gene structure of PSPHL. Solid boxes represent for the Exons and the lines represents the introns. The size and the position of the introns, exons in PSPHL gene is predicted based on the assumption that the human PSPHL and Chimpanzee PSPHL genes are similar to each other.

9B: Schematic figures of the two spliced variants of PSPHL mRNA. Numbers above the bars represent the nucleotide position of the exon boundary in the respective clones. Note that the 5 prime of the exon 1 sequence of BC065228 was not complete, and the last exon in BC065228 was shortened when compared to AJ001612 possibly due to alternative cleavage during mRNA synthesis.

Expression of the PSPHL Gene is Determined by the Presence of AAlns.

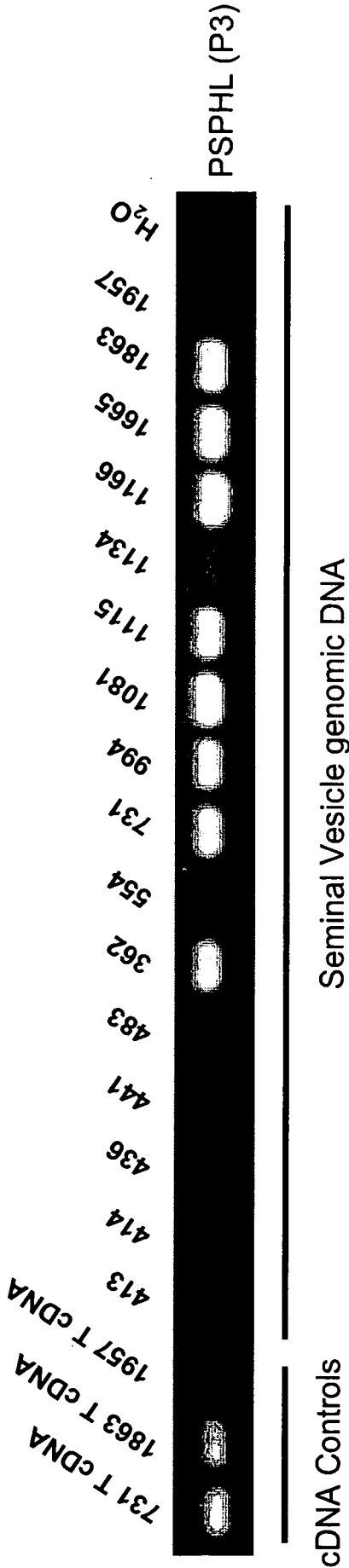


Figure 10. Expression of the PSPHL Gene is Determined by the Presence of AAlns.

A total of 11 cases for which expression status of PSPHL were known were examined for the presence of AAlns in genomic DNA isolated from seminal vesicles. The 8 PSPHL mRNA positive cases (362, 731, 994, 1081, 1115, 1166, 1665, 1863) were also positive for the presence of AAlns, as detected by the primer set P3. Three cases that were known to be negative for PSPHL expression (554, 1134, 1957), plus 5 European American genomic DNA samples randomly picked from the case-control cohort (413, 414, 436, 441, 483), were all negative for AAlns (with the exception of a faint band for 1134, which may be caused by contamination). Three cDNA samples were used as positive (731, 1863) and negative controls (1957).

The P3 primers:

TCAGCTAAAGTGGCTGTTGGGTGT (Forward) nt 25-48 in AJ001612 sequence
AAGCTTCTGCGCTACCTTGCGAT (Reverse) nt 135-157 in AJ001612 sequence.

Negative Samples
Positive Samples

AAInS is associated with Cancer Risk in European Americans

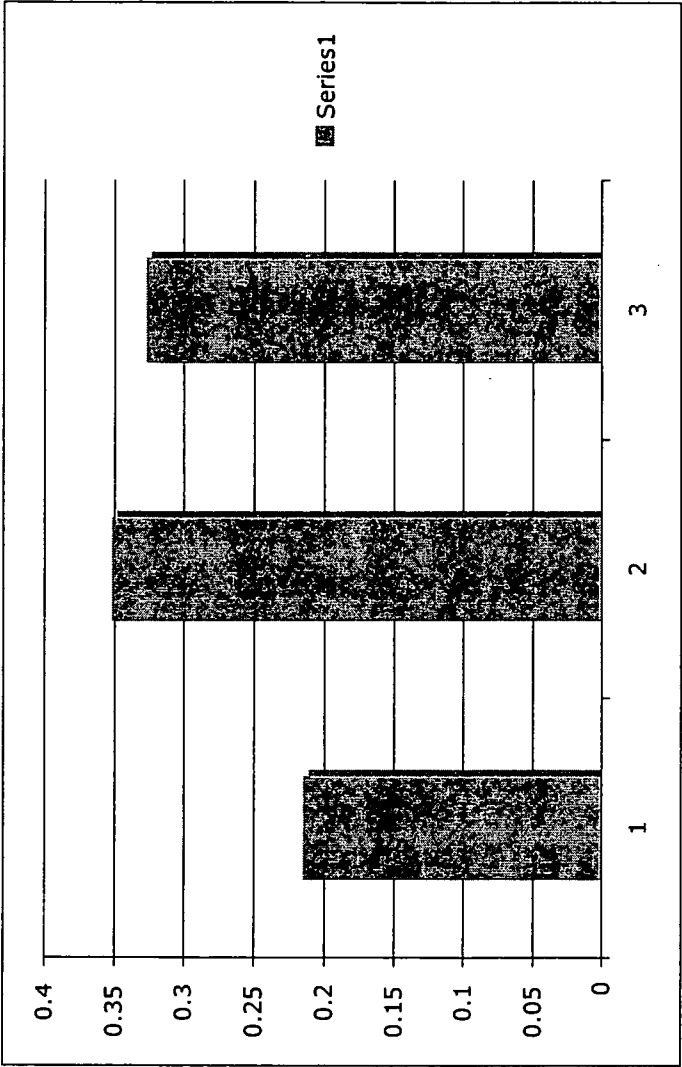


Figure 11:
Percentages of individuals (expressed as a fraction) with positive AAInS in European American controls (1), prostate cancer cases with Gleason 6 and below (2), and prostate cancer cases with Gleason 7 and above (3).



Figure 12: Multiplexed PCR using primer sets P3 and P4 were used to genotype the AAIns locus.

The P3 primer pair, as described in Figures 7 and 9, amplified the allele with the AAIns sequence, while P4 primer pair, as designed below, amplified the allele without the sequence. The validity of the assay is confirmed here by the use of mixed genomic DNA samples known to have two copies of the AAIns sequence (1665, 1863), and zero copies of the AAIns sequence (1957, 1704). The P4 primer sequence was designed against a sequence 5' of the AAIns sequence, as determined by alignment of the Chimpanzee human genome assemblies as of 2006, and based on the prediction of the breakpoints in chromosome 7 human reference sequence.

The P4 primers:

AGTCTTGCTATCTTGCCCAGGCTGAT (forward), nt 5419659-5419684 in chromosome 7 human reference assembly
GTAGAGACTGGGTTTCACCATGTTGG (Reverse), nt 5421321-5421346 in chromosome 7 human reference assembly

Lane	Sample
1.	Mixture of genomic DNA from sample 1665 and 1704
2.	Mixture of genomic DNA from sample 1665 and 1957
3.	Mixture of genomic DNA from sample 1863 and 1704
4.	Mixture of genomic DNA from sample 1863 and 1957
5.	Mixture of genomic DNA from sample 1704 and 1957
6.	Mixture of genomic DNA from sample 1665 and 1863
7.	No DNA template
8.	No DNA template

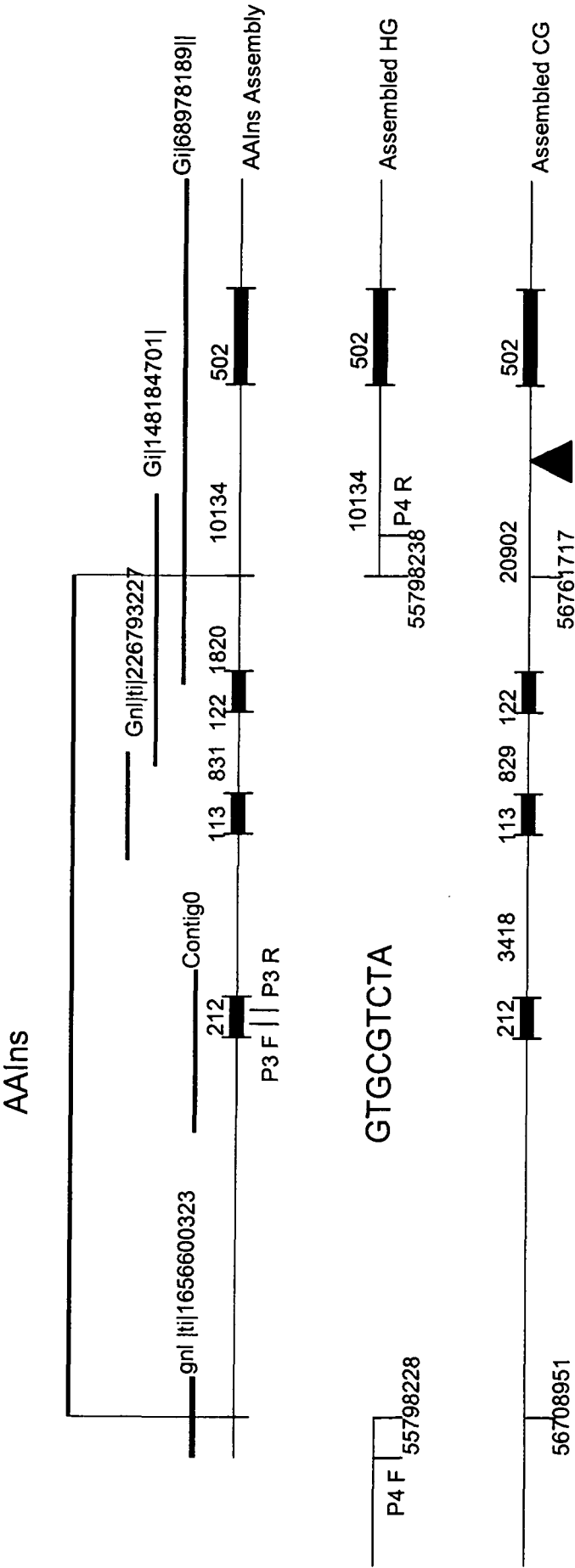


Figure 13: Schematic diagram of AAIns in relation to the assembled Human genome (HG) and Chimpanzee genome (CG) chromosome 7 sequences (not to scale).

The human WGS and Trace sequences presented in Exhibit F and H were used to obtain the partial assembly of AAIns and their positions and IDs marked. Coordinates for the assembled genome were marked below the line positions and sizes of the exons and introns in both genomes and AAIns were marked above their respective spaced-out positions. The 5' breakpoint is chr7 55798228 in the assembled human genome, and separated by 9bp (GTGCGTCTA) from the 3' breakpoint at Chr7 55798238 in the assembled human genome. The position of primer sets (P3 and P4) used to genotype the AAIns locus are marked with red vertical bars as shown. P3 primer set amplifies sequences in Exon1 which is part of AAIns, while P4 primer set amplifies the PSPHL-null allele. A Black triangle marks the site of a ~9kb viral insertion specific to the Chimpanzee genome, which resulted in size differences in intron 3 between the two genomes.

Figure 14

AJ001612 (1-839):

aagccacagg	ctccctggct	ggcgtcagct	aaagtggctg	ttgggtgtcc	50
gcaggcttct	gcctggccgc	cgccgcctat	aagctaccag	gaggagcttt	100
acgacttccc	gtcctgcggg	aagtggcggg	cacgatcgca	aggtagcgca	150
gaagcttctc	aatggccagc	gccagctgca	gccccggcgg	cgcactcgcc	200
tcacctgagc	ctgggaggaa	aattcttcca	aggatgatct	cccactcaga	250
gctgagggaag	cttttctact	cagcagatgc	tgtgtgtttt	gatgttgaca	300
gcacggtcat	cagtgaagaa	ggaatcGGAT	GCTTTCATTG	GATTTGGAGG	350
AAATGTGATC	AGGCAACAAG	TCAAGGATAA	CGCCAAATGG	TATATCACTG	400
ATTTTGTAGA	GCTGCTGGGA	GAACCGGAAG	AATAACATCC	ATTGTCATAC	450
AGCTCCAAAC	AACTTCAGAT	GAATTTTAC	AAGTTACACA	GATTGATACT	500
GTTTGCTTAC	AATTGCCTAT	TACAACCTGC	TATAaAAAAGT	TGGTACAGAT	550
GATCTGCACT	GTCAAGTAAA	CTACAGTTAG	GAATCCTCAA	AGATTGGTTT	600
GTTTGTTTTT	AACTGTAGTT	CCAGTATTAT	ATGATCACTA	TCGATTTCTT	650
GGAGAGTTTT	GTAATCTGAA	TTCTTTATGT	ATATTCCTAG	CTATATTTCA	700
TACAAAGTGT	TTTAAGAGTG	GAGAGTCAAT	TAAACACCTT	TACTCTTAGG	750
AATATAGATT	CGGCAGCCTT	CAGTGAATAT	TGGTTTTTTT	CCCTTTGGTA	800
TGTCAATAAA	AGTTTATCCA	TGTGTCAGAA	aaaaaaaa		

Human Genomic chr7p11.2 (55808267-55808970):

gagagcagag	atcacaccac	tgcgctccag	cctgggcaac	agagcctcca	55808316
aaaaaataaa	taaaaataaa	aatcatthtat	ctctcttttc	ttacttctta	55808366
GGATGCTTTC	ATTGGATTTG	GAGGAAATGT	GATCAGGCAA	CAAGTCAAGG	55808416
ATAACGCCAA	ATGGTATATC	ACTGATTTTG	TAGAGCTGCT	GGGAGAACCG	55808466
GAAGAATAAC	ATCCATTGTC	ATACAGCTCC	AAACAAC TTC	AGATGAATTT	55808516
TTACAAGTTA	CACAGATTGA	TACTGTTTGC	TTACAATTGC	CTATTACAAC	55808566
TTGCTATAgA	AAGTTGGTAC	AGATGATCTG	CACTGTCAAG	TAAACTACAG	55808616
TTAGGAATCC	TCAAAGATTG	GTTTGTTTGT	TTTAACTGT	AGTTCCAGTA	55808666
TTATATGATC	ACTATCGATT	TCCTGGAGAG	TTTTGTAATC	TGAATTCTTT	55808716
ATGTATATTC	CTAGCTATAT	TTCATACAAA	GTGTTTTAAG	AGTGGAGAGT	55808766
CAATTAAACA	CCTTTACTCT	TAGGAATATA	GATTCGGCAG	CCTTCAGTGA	55808816
ATATTGGTTT	TTTTCCCTTT	GGTATGTCAA	TAAAAGTTTA	TCCATGTGTC	55808866
AGAacagatt	tgtggaattt	gcagttctgc	ctcttggata	ttggaatatc	55808916
caatatthtca	ggtactgtgt	tgaagcataa	acatccccca	ggcaagaaac	55808966

aggc

Figure 15

AJ001612 (1-839):

aaGCCACAGG	CTCCCTGGCT	GGCGtCAGCT	AAAGTGGCTG	TTGGGTGTCC	50
GCAGGCTTCT	GCCTGGCCGc	CGCCGCCTAT	AAGCTACCAG	GAGGAGCTTT	100
ACGACTTCCC	GtCCTGCGGG	AAGTGGCGGG	CACGATCGCA	AGGTAGCGCA	150
GAAGCTTCTC	AATGGCCAGC	GCCAGCTGCA	GCCCCgGCGG	CGCACTCGCC	200
TCACCTGAGC	CTGGGAGGAA	AATTCTTCCA	AGGATGATCT	CCCACTCAGA	250
GCTGAGGAAG	CTTTTCTACT	CAGCAGATGC	TGTGTGTTTT	GATGTTGACA	300
GCACGGTCAT	CAGTGAAGAA	GGAATCGGAT	GCTTTCATTG	GATTTGGAGG	350
AAATGTGATC	AGGCAACAAG	TCAAGGATAA	CGCCAAATGG	TATATCACTG	400
ATTTTGTAGA	GCTGCTGGGA	GAACCGgAAG	AATAACATCC	ATTGTCATAC	450
aGCTCCAAAC	AACTTCAGAT	GAATTTTtAC	AAGTTACACA	GATTGATACT	500
GTTTGCTTAC	AATTGCCTAT	TACAACCTGc	TATAaAAAAGT	TGGTACAGAT	550
GATCTGCACT	GTCAAGTAAA	CTACAGTTAG	GAAtCCCTCAA	AGATTGGTTT	600
GTTTGTTTTT	AACTGTAGTT	CCAGTATTAT	ATGATcACTA	TcGATTTCCT	650
gGAGAGTTTT	GTAATCTGAA	TTCTTTATGT	ATATTCCTAG	CTATATTTCA	700
TACAAAGTGT	TTTAAGAGTG	GAGAGTCAAT	TAAACACCTT	TACTCTTAGG	750
AATATAGATT	CGGCAGCCTT	CAgTGAATAT	TGGTTTTTTTT	CCCTTTGGTA	800
TGTCAATAAA	AGTTTATCCA	TGTGTCAGAA	aaaaaaaa		

Chimpanzee Chr7 (56755112-56781414):

ggcgcgccct	tgtgacgtcg	cagaggcccg	cccctcacgc	ggagccaatc	56755162
ggaactcgag	gcggggctgt	tggtcttccg	ggagcgcgca	tgcgcggggg	56755212
CCACAGGCT	CCCTGGCTGG	CGgCAGCTAA	AGTGGCTGTT	GGGTGTCCGC	56755262
AGGCTTCTGC	CTGGCCGTcG	CCGCCTATAA	GCTACCAGGA	GGAGCTTTAC	56755312
GAATTCCCGg	CCTGCGGGAA	GTGGCGGGCA	CGATCGCAAG	GTAGCGCAGA	56755362
AGCTTCTCAA	TGGCCAGCGC	CAGCTGCAGC	CCCtGCGGCG	CACTCGCCTC	56755412
ACCTGAGCCT	Gggtacgtgc	gccccacaac	accttcccca	gccagggccc	56755462
ggggaccctcg	ggagcgctcc	cgccacctg	gcgccgctca	tcctgggag	56755512
ggtgggaccc	cactgaggcc	gccacgcatt	agggagctgc	actcccaggt	56755562
ttgacggcca	gltgtaatat	cattaagtct	ttgaaatctt	gtagcggggt	56755612
agaaggggct	aggaaaggaa	gaaaacatct	ttttaaaaaat	ataagcgatc	56755662
ggccggggcgc	ggtggctcac	gcctgtaatc	ccaacacttt	gggaggccga	56755712
ggcaggtgga	tcacgaggtc	aggagttcaa	gaccagcctg	gccagcatgg	56755762
tgaaaccctg	tctctactaa	aaatacaaaa	attaaccggg	cgtggtggcg	56755812
ggcgccctgta	atcccagcaa	ctcggggaggc	tgaggcagag	aactgtttga	56755862
acccagaagg	cggaggttgc	agtgagccga	gatcgctca	ctgcactcca	56755912
gcctggggcga	cagagcgaga	ctccgtctga	acaaacaaaat	atgtgtatgt	56755962
atataagtga	tcgaaccctg	gaggttgagt	ttacagtaag	ccgagattgc	56756012
gccactgcac	tccagcctgt	gtgacagagt	gagaaaatgt	ctctaaaaaa	56756062
ttatatgcaa	gtgagagctt	tttttccagc	gctcatgctc	agactgaaga	56756112
aagtaattgg	gccggggcgt	tagctcatgt	ccataatccc	agcacttttg	56756162
gaggccgagg	caggcggtc	gcttgagctc	gggagttcca	gactagcttg	56756212
ggcaacatgg	tgaaaccctg	tctctataaa	aatacagaaa	attattgggc	56756262
atggtggcat	gcgctttag	tccccgctac	ttgagcggtc	gaagccgaag	56756312
gatcgacctc	ctgggctcag	gcaatccatt	tcagcctccc	aaattgctgg	56756362
gattacagga	atgagccatc	gtgcctggcc	ttacactata	ttttaatgct	56756412
ttttttgaaa	atggaaactt	ttacaggcaa	ttcactttct	tcaaactaat	56756462
gataaggaaa	tgatgctgtt	ctgtttctgtt	ttgttttttg	gttttgtctg	56756512
tttctttttt	ttccttttatc	gagacagggt	cttgctctgt	tgcccaggct	56756562
ggagtgaggc	ggtgaaatca	tggctcactg	cagccttgac	ctccgggctc	56756612
aagggaatcgt	tcccggctcag	cctccccagt	agctaggact	acaggtgcat	56756662
gctaccatgc	ttgggctaatt	ctttttttga	aatggagctc	cactctgtct	56756712
cccaggctgg	agtgcagtgg	tgcaatctca	gctcactgca	ggctggtctc	56756762
aaactctgac	ttcagatggg	ccatccacct	cagcatccca	aagtgtctgg	56756812

attacaggtg	tgatccacca	cgcccaatga	ttttgctcat	tttagatact	56756862
acaacttttt	agtttaattt	ttttttttcc	tgagatagaa	tctcactttg	56756912
tcgccaggct	ggagtgcagt	ggcgtgatct	cagctcactg	cagcctccac	56756962
ctcctgggtt	caagcgattc	tcctgcctca	gtctcccgag	tagctgagac	56757012
tacaggtgct	tgccaccacg	cccaggagtt	caagactgag	cccaggagtt	56757062
caagcctgca	gtgagccatg	atcgtgccac	tgcaactccag	cctgggcaac	56757112
aagcaagacc	ctgtctccac	aaataagtaa	ataaacatcc	tatctgtacc	56757162
agggtacagg	caggcgctat	tattccctact	tttcatcctc	aactctgagt	56757212
tgagtcatac	atcaactctg	atgacctttg	tttccactcc	ccttaagcca	56757262
ttcacaggaa	cagctacttt	acatagaaaa	gtcaatcttg	gagggttgag	56757312
taatttgctt	aatcactgta	aaaggtagag	cttcaattat	cacccaggcc	56757362
catctgttgc	tggtattccc	tctccacccc	tccgtgtggc	ttccctttat	56757412
ttattttatt	taattttttt	tgagatggag	ttgcactttt	gttgcccagg	56757462
ctggagtgca	atggcatgat	ctcggtcac	tgcaacctcc	acctcccagg	56757512
ttcaagcaat	tctcctgcct	caggctccca	agtaggtggg	attacaggtg	56757562
cctgccacca	agcccagcta	attttttgta	cttttagtag	agatgggggt	56757612
tcaccatggt	gaccaggctg	gtcttgagct	cctgacctca	ggtgaccac	56757662
cctcctcagc	ctcccaaagt	gctgagatta	caggcgtgag	tcacggtgcc	56757712
tggccttggc	tttcctttta	taaatcacga	ctcaatttga	tgtggatgct	56757762
ctgttcattc	ccagaatata	atgaatggat	tatatatata	ttttttgttt	56757812
ttgagacaga	gtcttgcttt	gtcttgacag	ctggagtgca	gtggcgcgat	56757862
ctcagctcat	tgacagctct	gcctcccagg	ttccagcaat	tctcctgcct	56757912
cagctcctct	ggtagctggg	attacgggca	ctcaccacca	cgcccagcta	56757962
attttttttt	tttttttttt	tttttttagt	aaatttcacc	gtgtcgcca	56758012
ggctggtctc	aaactcctga	cctcaggtga	tctgcccagc	ttggcctccc	56758062
aaagtgtctg	gattataggg	atgaccacc	gccccagcc	tcaacatcaa	56758112
ataaatttaa	aagtgtgtaa	cagccaggcg	cagtggctca	cacctgtaat	56758162
cccagcactt	tggggggcca	aggcgggcag	atcacctgag	gtcnnnnnnn	56758212
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56758262
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56758312
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56758362
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56758412
tcaagaccag	cctggccaac	atggtgaaac	cccgtctcta	ataaaaaata	56758462
aaaataaaaa	aaacagttag	ctgggtatgg	tggcatgtgc	ctgtaatccc	56758512
agctactcag	gaggctgagg	cacgagaatc	actagaacct	aggagggtga	56758562
ggttgcagta	agccaagttc	gtgccactac	cctccagcct	gggcaacaga	56758612
gtaagactcc	atctaaaaaa	aaaaaaaaat	gaagaagaag	aaattagtgt	56758662
agtgtgggaa	gtgaaaaaaa	aaaaaaagaa	aaggaaaaga	aaaatgattg	56758712
aattcatgaa	tacactctta	tgtggcctgc	accgactttg	acacaaatta	56758762
gattggctta	gtaggcaagg	gtgggatctt	ttcataattt	tatttgatgt	56758812
ctaaaaatac	tttctctttt	tttcttatag	GAGGAAAATT	CTTCCAAGGA	56758862
TGATCTCCCA	CTCAGAGCTG	AGGAAGCTTT	TCTACTCAGC	AGATGCTGTG	56758912
TGTTTTGATG	TTGACAGCAC	GGTCATCAGT	GAAGAAGGAA	TCGgtgagct	56758962
agccaaaatc	tgtggtgtag	aggacgcggt	gtcagaaatg	tagggatagc	56759012
atttattcac	cttatgaaat	gataaagaat	ttcttttttt	tttttttttg	56759062
agacagagtc	tcactctatt	gcccagactg	gagtgcagtg	gcacaatctt	56759112
ggctcactgc	agcttcgacc	tcccaggctt	aagcaattat	cctgcctcag	56759162
cctcctgagt	acctgggatt	acaggcgtgt	gcaaccacat	ccggctaatt	56759212
tttttatttt	tagtagagat	ggagtttcac	catgttggtc	aggctggtct	56759262
caaactcctg	acctcgtgat	ccaccgcgct	cagcctccca	aagtgtctgg	56759312
attacaggca	tgagccacca	tgcccggcca	tgaaatgata	agaattttct	56759362
aaagagtggc	tgttttggat	gaagtgtctg	accttggtca	tggacatga	56759412
gcactaggct	ttttctttta	ccccttagag	ttgaatttga	tatattcatt	56759462
catttatata	aataatttatg	gcaaaaaaat	atggataaga	catgttctta	56759512
gcttgattag	gggaatgatg	accatcacta	accaaagaag	atggcctagg	56759562
gggcagggtc	caagtgtcca	gtggctttgt	gtcatggttg	aaagtgtggc	56759612
cttagtggtt	ccctggaaag	gttcagagtc	tcttggtgga	tcctgggtca	56759662
gagccctctt	ctccctccct	ccctcctgcc	ctcccacctc	ctgcccgtca	56759712
gctgggcacc	acctctctga	gccccagtc	cccagtcag	caccatgtca	56759762
ttttcttttt	tttttttttc	aggacggagt	ctcgctctgt	caccaggctg	56759812
gagtgtcaat	gtgcaatctc	ggctcactgc	aacctccgcc	tcctgggttc	56759862
aagcaattct	cctgcctcca	cctcctgagt	agctgaaact	acaggcacgc	56759912
accaccacac	ccggtttatt	tttgtatttt	tagtagagat	ggtgtttcac	56759962
catgttggcc	agcctggtct	tgaactcctg	acctcgtgat	ctgcccgcct	56760012

cggcctccca	aagtgtctggg	attacaggca	tgagccacca	cacccggcca	56760062
ccatttttatt	ttcaactccc	tttcatggaa	gaacgttttag	ccttttggctct	56760112
ctttcttgat	ttatgacagc	ttgcggcttt	caagaaaact	acctatgaat	56760162
aggctgtgta	actttttatt	atttattttat	ttttttgaga	tgaagtctct	56760212
gtcttctagg	ctggggtgca	gtggcatgat	ctcggtcac	tgcaacctcc	56760262
ccctcccagg	ttccagcaat	tctcctgcct	cagcttccca	agtagctggg	56760312
attacaggca	cgcaccacca	cacctggcta	atttttgcat	ttttttttat	56760362
tagagatggg	gtttttgtcat	gttggccagg	ctggtctcga	actcctggct	56760412
tcagatgagc	tgcccacctc	agcttcccaa	agttttggca	ttataggcat	56760462
gagtcactgc	acccagcttg	ttaatgttat	tttcaagcac	accttcaaaa	56760512
gtttattcca	aggccttgct	ctcatagcag	caaagcctgt	ctggtgcata	56760562
gtggggctct	tgtttagcatc	ttccttttgg	tggatgttgt	gtagaccg	56760612
ggcagtaagg	catgttaacc	tcgaggacat	tggacctggc	tggcctgact	56760662
gttgggtctc	cctcctagga	catggcgagc	catgggtggg	gcagtgtctt	56760712
tcaaagtgtc	tctcacggag	cacttagccc	caatccagcc	ctccagggag	56760762
caggtgcaga	gactcatagc	agagcacccc	ccacacctga	ccccagcat	56760812
aagtaagagg	agccctgct	ccaggtgtat	ttcagcacca	gtgttggggg	56760862
gcatgtcctc	ctgagagcat	ctaacgattg	cttttagagag	ccctctgggt	56760912
ggtttattta	ttttcatttt	tattttttaa	ttttattatt	ttaaaaaat	56760962
ttatttattt	attttatttt	ttattttattg	aggcagagtc	tcgctctgtt	56761012
gccagacta	gagtgaatg	ctacgatctt	ggctcactgc	aacttctgcc	56761062
tcccgggttc	aagtgattct	cctgcctcag	cctcctgagt	agttgggatt	56761112
tcaggcacc	accatgccag	gctagttttt	gtatttttag	tagagacagg	56761162
atttcaccat	gttggttagg	ctggtctcga	actccagacc	tcaggtagtc	56761212
tgccgcctc	agactcccaa	agtgtctggga	tcacaggcgt	gagccactgt	56761262
gcccggcaaa	atttttattta	cttttttttt	ttttgagaca	aagtctcact	56761312
ctgtcgccca	ggctggagtg	cagtggtgca	atctccactc	actgcaactt	56761362
ttgcctccca	ggttcacacg	attctcctgc	ctcagcctca	ggagtagctg	56761412
ggattacagg	cgcctgccac	cacacctggc	taatttttgt	attttttagta	56761462
tagatgggtt	tcgccatggt	ggccaggctg	gtctcgaact	cttcatctca	56761512
agtgatccac	ctgcctcggc	ctcccaacat	gctgagatta	caggtgtgag	56761562
tactgcacc	tggcctatta	ttattttttt	tagattcaga	gtctcattct	56761612
gttgcccagg	ttggagtaca	gtggtataat	catggcccac	tgcagcctcg	56761662
aactcctggg	gtcaagtagc	cctcccacct	tagcctcctt	agtagtaggg	56761712
actgcagggt	catgccacca	tgcccagcta	attaaagaat	tttttttttt	56761762
ttttctgaaa	gtcttgcctt	gtcaccagg	ctggagtga	gtggcatgaa	56761812
cacagctcgc	tgcagcatca	atcttctggg	ctcaagggat	cctcctgcct	56761862
ctgcttccaa	agtagctggg	actacagaca	taggccacca	tgccattttt	56761912
tttttttttt	ttttttgaga	cggagtcttg	ctctgttgcc	caggctggag	56761962
tgcagtggcg	ccatcttggc	tactgcaag	ctccacctcc	tgggttcacg	56762012
ccattctctt	ccctcagcct	cccagtagtc	tgggactaca	ggcgctgcc	56762062
accacagctg	actaattttt	tgtattttta	gtagagacgg	ggtttcacca	56762112
tgttagccag	gatggtctca	atctcctgac	cttgtgatcc	acctgcctca	56762162
gcctctcaaa	gtgctgggat	tacaggcggtg	agtcaccgcg	cccggcccgc	56762212
cgtgccta	taaaaaaaaa	aaattttttt	tatagacaca	gcattctcact	56762262
acgttgccca	ggctggtctt	gaactggtgg	gcttaagtga	ccaattttca	56762312
caactcagcc	tcccaaagt	ttggaattat	aggtttgagc	cactataccc	56762362
agcctttaag	aatcttcttg	tagagatgga	atcttgctgt	gttgcccagg	56762412
ctgttctcaa	gctcctggcc	tcaagtgatt	ttctaccttg	gcctcccaaa	56762462
gtgctggaat	taaaggcatg	ggctgtgtgg	ttcttttggc	acttacacgt	56762512
ggtcttgtct	ggccggttgt	ctggtcctgt	ctgtttctgc	ctttcctctt	56762562
tctccaggga	aaacctaagc	tttcttctgt	tgtcctcatc	ttgtgttttt	56762612
ctgggtccat	gggcagagta	gagttctaga	acggtttcct	aaagcagcca	56762662
agccctaccc	attgatttct	aagtacattt	agaaaaacac	atataaggct	56762712
gggcatgggtg	gctcgcacct	gtagtcccag	cactttggga	ggccgaagtg	56762762
gatggatcac	caggtcagga	gttcgagacc	agcctaacca	acatgatgaa	56762812
accttgtctc	tactaaaaat	acaaaaatta	agccggggcac	ggtagctcac	56762862
gtctgtaate	ccagcacttt	gggaggccaa	gatgggtggt	atcacctgag	56762912
gtcaggagtt	cgagaccagc	ctggcccaca	tggtgaaacc	nnnnnnnnnn	56762962
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763012
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763062
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763112
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763162
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763212

nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763262
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763312
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763362
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnctgtgagc	56763412
cgagatcatg	ccagtgcact	ccagctaggg	caacagagca	agattccatc	56763462
tcaaaaaaaa	aaaaaatcag	ccgggcattg	tggccccggc	ctgtaatact	56763512
caggaggctg	aggcaggaga	atcgcttgaa	tccgggaggt	ggagggtgca	56763562
gtgagccaag	atcgagcact	tgcactccag	cctgggtgac	aagagtgaaa	56763612
ctctgtctca	aaaaataaag	agaaaaacat	atagaaaaca	ttaacacccc	56763662
aggcagtata	ccttgtcaaa	catacctcag	gcaaattgat	tcaggagaag	56763712
aaaatacatc	ttatttccct	cttcatgttt	cgtttttttt	tttttctttt	56763762
gtattactca	gtgttgggta	ttttgtattt	tattttgcgg	ggagctggta	56763812
agtcgcctac	aggagcgaaa	tgttcagggt	ttcctaatac	ctgggtggctt	56763862
taggagtatt	gtagagcatg	ttgcttcaaa	gctcaatatc	ccagcaacca	56763912
atgtatttgc	cagtaggctg	aaattctact	ttaatggtaa	gatgttaacg	56763962
gtaacatgtt	ccctttctta	gcagttccat	tattcggtat	tctgggtaat	56764012
gtcttttggg	atgcaacttg	aacagtcaca	tcagagttaa	atattgagat	56764062
gaattggttct	cttaattgat	gtcttttgcc	ctttgggtaca	ctttgctcag	56764112
caaaacagac	ttagatcatt	acctgtctta	gctttattat	ctataacatc	56764162
acctgtctgt	acataaatgt	ctgcacttca	tgaacatta	attcctacat	56764212
agcccctatg	acttactttg	ctatgtgttg	tatttcaatt	ttatgggagc	56764262
ttctatgcat	ttgaagcaat	ttagactcaa	ggaggagact	tttggctggg	56764312
cacggtggct	cacgcctgta	atcccagcac	tttgggagcc	tgaggcgagga	56764362
ggatcacttg	aggtcaggag	ttcgagacca	ccctggccaa	cgtgggtgaag	56764412
ccccactttt	gctaaaaaaa	aaaaaaaaaa	aaaaaaaaat	aggaaaatta	56764462
gctgggcatg	gtggcacatg	cctgtagtcc	cagctacttg	ggaggctgag	56764512
gcaggagaat	cacttgaagc	tgggaggcgg	agggtgcagt	gcgccgagat	56764562
cgcgccactg	cacttacagc	ctgggtgaca	gagcgatact	ctgtctttaa	56764612
aaaaaaaaaga	cagagacttt	agaagggaga	cttttgttaa	cagattactg	56764662
aactacagct	gcccttgggg	aagcctccta	tcccctctct	ggcccatagc	56764712
ggtgccacag	ctgttatatt	atctcagatc	tccagttata	tcaaggagtg	56764762
actgagagtg	atattctttt	tttttttttt	tttttttgag	acagggtctc	56764812
actctgttgc	tccggctgga	gtgcagtggc	tccatctcgg	ctcactgcaa	56764862
cctctgccgc	ctgggttcaa	gcgattctcc	tgcctcagcc	tcccaagtag	56764912
ctgggattac	agggtgccc	caccacgcct	ggctaatttt	tgtacagtgc	56764962
ccaaccgtac	ttttgtttta	gacagctgat	ttgtcctata	atcccagcac	56765012
tttgggaggc	tgaggtagga	agatcgcttg	agcacaggag	tttgagacca	56765062
gcctgggcaa	catagtaaga	ccccatctct	acaaaataaa	taaataaaat	56765112
aaaataaaaa	cagctgattt	agaactatat	aagattttaa	acaggagtta	56765162
aacaaaaaca	actctcatta	caagataaaa	acaaaataat	aaaaacaaat	56765212
tcaatcttgt	tatgtcaact	gactttgtgc	aaacccaaga	aaagttatga	56765262
ataaggaact	aaaaaccttt	taaaactttt	aacaatgcct	tatccctgta	56765312
gcatgggcac	caccttttag	cctaggtgtg	tgtcaataaa	gagttgattt	56765362
catttcccca	aataatacca	tcaacaactc	ataaccgaga	ttttaacatc	56765412
attttatact	ttatatacta	gttcctcata	tgttaacctc	tttaatcctc	56765462
ataacagccc	tgtgaagtca	aatattatgt	ttggtttttg	tttttttaga	56765512
aagagtctca	ctctgtaccc	caggccggag	ttcagttgtg	tgatcgtagc	56765562
ttatcacaga	cctgaacttc	tgggctcaaa	cgatcctccc	acctcagcct	56765612
cctgggtagc	tgggattaca	ggtatgtgcc	actgcattcc	acaccaccac	56765662
agcattttat	tcctctgaac	taacgtcatt	ttctgagact	agcagagggt	56765712
ttatttactt	ttttattttt	atttttattt	ttttttgaga	cggagtctcg	56765762
ctctgtcacc	caggctggag	tgcagtggcg	tgatcttggc	tcactgcacc	56765812
ctccgcttcc	caggttcaag	cgattcttct	gcctccgcct	ctcaagtagc	56765862
taggactaca	agtgcacgcc	accacgcctg	cgatgggtat	tttgtatttt	56765912
tagtagagac	ggggtttcac	catattggcc	aggctgggtc	cgaactcctg	56765962
accttgtgat	ctgcccgcgc	tggcctccca	aagtgtctgg	attacaggca	56766012
tgagccacca	cgcccagccc	tggttttatt	tattttttga	gacggagtct	56766062
cgctctgtca	cccaggctgg	agtgcagtgg	cgcggtctca	gctcactgca	56766112
gcctcgacct	cccggactca	agcgatgttc	ccatctcagc	ctcctgagta	56766162
gtgtgtgaaa	agagctgggc	acattccttc	gccccgggct	caaaacaaac	56766212
aagcccagta	caaacacatc	ctatcctcct	atcctaccac	atctcaccat	56766262
atatctctta	aacttccccc	gggctcaaaa	caaacaagcc	cagtacaaac	56766312
accaccagga	aagtctccga	taagaggaca	gatgagggga	cagctgttca	56766362
aagttttact	gaaagagcgg	gaacccaaa	aattcctttg	ttccccctgta	56766412

acttttcaggc	tataaaaaag	caaacacttg	cattgttcag	ggccctcttg	56766462
tatgcggtgg	aatggaagga	ccaggttcga	acttgtagta	aagatccttg	56766512
ccgcttggt	ttgactctgg	actctgggtg	tcttctttga	agaacaaacg	56766562
gtctaggcat	aacatctggg	ggctcggtccg	ggattcccca	agcccaccag	56766612
acccctggtc	aatggatctg	ctaggatcga	tctactgata	ggtgagctgg	56766662
ctcgtctccg	tttgctccgtg	tctgttctga	atccgaatct	gtgactcgcg	56766712
aggtctgaaa	ctggagctgg	cacagtcctg	gcggacgcgc	tataggacag	56766762
ccagcggaga	ccagtaggag	acgtccctcg	gctctcatct	gatctatatt	56766812
gcgatctgag	ctgccccggt	ttggcgggcc	gcagccgtat	ctgagctgcc	56766862
cggttgatcg	ggcagcatct	gtctctgatc	tgagctgccc	cggtttgga	56766912
ggcagcagcc	gtacttgagc	tgccccggtg	atcgggcagc	atctgtctct	56766962
gatctgagct	gccccggttt	ggcaggcagc	agccgtatct	gagctgcccg	56767012
gttgggcagg	cagcatctgt	ctctgatctg	agcagccccg	gtttggcggg	56767062
cagcagccgt	atctgagctg	cccggttggg	cgggcagcag	ctgtctctga	56767112
tctgagctgc	cccggtttgg	cgggcagcat	ctgtctctga	tctgagctgc	56767162
cccggtttgg	cgggcagcag	ccgtatctga	gctgcccagt	tagggcgga	56767212
gcagctgtct	ctgatctgag	ctgccccggt	ttggcgggca	gcactgtct	56767262
ctgatatgag	ctgccccggt	ttggcaggca	gcagccgtat	ctgagctgcc	56767312
cggttaggcg	ggcagcagct	gtctctaatc	taggctgccc	cagcgcgacc	56767362
gcgatccagg	cagctaactc	tgacccgggc	ttccgggtgcg	cgcttgcgat	56767412
ctgaactgcc	ccggtttggc	gggcagcagc	tgtctctaata	ctaggctgcc	56767462
ccagcacgac	cgcgatccag	gcagctaact	ctgacccggg	cttccgggtgc	56767512
gcgcttgcca	tctgaactgc	cccggtttgg	cgggcagcag	ctgtctctaa	56767562
tctaggctgc	cccagcgcca	cgcgactctc	ggcagctaac	tctgacccgg	56767612
gcttccggtg	cgcgcttgcg	atctgaactg	ccccggtttg	gcgggcagca	56767662
gctgtctcta	atctaggctg	ccccagcgcg	accgcgatcc	aggcagctaa	56767712
ctctgacccg	ggcttccggt	gcgcgcttgc	gatctgaact	gccccggttt	56767762
ggcgggcagc	agctgtctct	aatctaggct	gccccagcgc	gaccgcgatc	56767812
caggcagcta	actctgaccc	gggcttccga	tgcgcgcttg	cgatctgaac	56767862
tgccccagtt	tggcaggcag	cagctgtctc	taatctaggc	tgccccagcg	56767912
cgaccgcgat	ccaggcagct	aactctgacc	cgggcttccg	gtgcgcgctt	56767962
gcgatctgaa	ctgccccagt	ttggcgggca	gcagctgcct	ctgaccaagg	56768012
ctgccaaaagt	gcgcctgcga	ttagatatca	ttaataagtc	agatcagatt	56768062
tcctttacaa	ggattcaaac	ccacattcct	ttacaggaaa	agactacaaa	56768112
ttattacagg	ataggttaata	cccagagcac	tcctctatct	ctccttacga	56768162
gtaatttcaa	agaagtttaga	gcaagaggcc	atgatcttgg	tatagaaatc	56768212
aggaaaagaa	agctaattac	tctgtgtcgc	tccgaatggc	ctgcctttga	56768262
tgtggggtgg	ccgcccgaaa	ggaccttccg	acttgctgtc	atcactaagg	56768312
taaagtccaa	gattttccta	cctaggcggtg	caggccactt	agatcatatc	56768362
ccatatatcc	tcatatggca	ggaccttggt	gagaacccgc	ctccttagct	56768412
gtcccccttc	caattggcct	ctgaacactg	taaggcacta	gttgctcgac	56768462
cactaaaatc	caggcaacca	actgcccccc	ccatcctggt	ttacctgaca	56768512
gcggggcccc	actgttcaca	gaacccccctc	cgtacccctc	cgggccccag	56768562
gccccagccc	ccctggctga	gctgcgggaa	ggagcaggcg	gatgggaggc	56768612
ggctggcaca	cacgggccccg	ctgaaaagga	aagtaacttt	gaaaggccgg	56768662
tggggaggac	gcgaaggcgc	acttcgcgga	ctagcccccc	tcagccgcct	56768712
gactccacgg	tggtcttacc	ccttcggcatt	ataggacccc	cggatgacac	56768762
aggaatcccc	aggctccagt	actggccgaa	ctccaccagt	gatctgtata	56768812
actggaaaac	tcagagtgtc	cggttttcag	acaaccccaa	agatttactg	56768862
gctttactag	atagtgtcat	gttcacccac	cagcccaactt	gggatgattg	56768912
tcagcagctc	ctccgaattt	tgttcaccac	ggaagagcga	gaaagaatac	56768962
agatagaagc	tagaaaagcta	gtccccgggag	acgacgggtca	accgattgcc	56769012
aaccccagacc	tcataaacgc	aaccttttct	ctgaccaggc	cagcgtagga	56769062
ctacaacacg	gcagaaggta	ggggacgggt	acacctttat	cgccagactc	56769112
taatggcggg	tctccgggca	gctgctcgca	agcccaactaa	tttggctaaa	56769162
gtatattcta	ttctgcaagg	aaagacagag	agcccagcta	cgtacttaga	56769212
aagattaatg	gaagctttta	gacagtacac	ccccatagat	ccagaggctc	56769262
caggaagtca	ggcagctggt	gtaatgtctt	tcgtaaatca	ggcagcccca	56769312
gatattaaga	gaaaactcca	gaaattagaa	gacttagaaa	gaaagcggat	56769362
tcaagacctc	cttcagatag	cccagcaggt	ttacaataac	agagatactc	56769412
cagaggaaaa	gcaatttaag	gccgctgaaa	aaaatgacca	aggtcctggc	56769462
agcagtggta	cagaaaaagc	atctacagcc	agagtacacc	caacctaggc	56769512
ggccccccccg	gcatgataat	ctgagcaaag	accaatgtgc	ctattgtaaa	56769562
agggctggcc	actaggtaag	agactgcccc	aaaaagaaac	cacgaggaca	56769612

aggacccgga	ccccctaagt	ctacacccgt	actagtcact	caagacgaag	56769662
actaaagaag	acgaggttcg	gacccccctcc	ccaaacctag	agtaactttg	56769712
caagtggagg	ggtccccagt	ccagttcttg	gtcgatatgg	gagcacagca	56769762
ctcggctcta	gttaaaacta	ataggaaatt	atcctccaaa	tcctcgtagg	56769812
tacaagaggc	cacaggagtt	aagaaatacc	catagacaac	acaaagaaca	56769862
gtaaacctcg	gagccaagaa	tgtaacccat	tccttcctag	tcacccctga	56769912
aagccccgtg	cccctattga	ggagagacct	gctaactaaa	ataggagcac	56769962
agatccattt	cctccctgag	aggcccgtcg	tgaccaaccc	ccacaataga	56770012
accgtgtcca	tcctgactaa	aaacctagaa	gataagtacc	ggctccacca	56770062
ggagaaagcg	gcccctgacc	aagacatagc	accttggtc	cagcagtatc	56770112
cagaagcgta	ggcggaacg	gggggcttaa	gtctagcaaa	acaccgtcct	56770162
gccttattta	ttgaacttaa	gcctaggaca	gaccctgtgc	aggatatgcca	56770212
ataccgatg	cccctagaag	ccaagagaag	gattgccccg	catatccgcc	56770262
ggctccttga	ccaagaggtc	ctacgccccat	gtcactcacc	ctggaatact	56770312
ccattgttgt	cggtacgaaa	acctaatagt	gaagaatata	gacctgtaca	56770362
agacttaaaa	gaaatcaaca	agaaagttgt	agacatacat	ccaagtgtac	56770412
ctaacccgta	taccctccta	agtaccttaa	accctaaaca	tcaatggtac	56770462
actgttttag	atttaaaaga	tgctttcctt	agtttgcctt	tagccccctca	56770512
gagccaaaag	ctcttcgcct	tcgagtagaa	tgaccctgat	aaaggcataa	56770562
gtggccaact	gacatggacc	aggcgccgca	aggattcaaa	aactctccta	56770612
ccctgcttaa	aagaaggcag	cgctggctaa	cagatgcccg	gaaacaaact	56770662
gttctgcaga	tccccaggcc	acaatccacc	cgacaagtga	gagagttcct	56770712
aaggtcgga	agattttgcc	gactatggat	acctaggttc	gcagaactag	56770762
ctaaaccctt	gtatcaggca	acacgggggc	aacagccatt	taattggaca	56770812
gacgaagccg	agttggcctt	ccaacagatt	aaaaccgccc	tactctccgc	56770862
gcctgacta	ggactacctg	atgttaccac	gcccttccac	ttatacgtgg	56770912
atgaaaataa	aggtgtcgcc	aaggcagtaa	taactcagaa	cttaggcccc	56770962
tggcggaggc	cagttgccta	cctgtcaaag	aagttagacc	cagtagttgc	56771012
cgggtggccc	ccttgtctcc	gaatgattgc	ggccacggct	ctgatggtgc	56771062
aagatgctga	taaacttgtc	ataggccaag	aattgcgggt	cgttactcca	56771112
catgccatcg	aaggtgtact	caaacagcca	cctaactgat	ggatgagtaa	56771162
cgcccggtc	acccactacc	aaggactact	actaaatcct	ctcaggataa	56771212
ttttcctgcc	cccaacgacc	ttaaaccctg	cctcgctgct	gcccaccccg	56771262
gacctggacg	ccccactcca	tgactgcacc	gagatactag	ctcaggtgca	56771312
cggagttcga	gaagacctgc	acgaccaccc	acttcttgac	gctgacctcg	56771362
tctagttcac	tgataagagc	agcttcatgc	atcaaggcca	gaggtacgct	56771412
ggggcggcag	taacttcaga	gactgaggta	atctgggcgg	aaccctgcc	56771462
cccggggaca	tcggcccaga	aggccgaact	gatagcgctc	acccaagctc	56771512
ttaccttaag	ggcggggaaa	aagctgacag	tatatacaga	cagccgatat	56771562
gcttttgcaa	cggcgcataat	acatgaggcc	atttacaagg	aacgaagggt	56771612
actgacggct	gaaggaaaaa	agataaaaaa	caagagatcc	tagccctgct	56771662
aacagcccta	tggaggccag	aaaaattagc	acttgtacat	tgcccaaggc	56771712
atcagaaact	aactactcca	actgctcaag	gcaactttct	ggcagaccag	56771762
actgcaagaa	atatggcgaa	ggctcccagc	caactccttg	cactccagct	56771812
ccctgaccgg	ggcccccagg	acttgccata	tttccctgaa	tattcagaac	56771862
aagatctcca	gtggattgac	aaacttcccc	tgaacaaaat	ccagaataag	56771912
tcggtggactg	atactaataa	ccaaaccac	ctaccagaaa	aattaagaca	56771962
acaggtgtta	gaacacatcc	accgaaccac	ccacctgggg	gcccggcgga	56772012
tgatagacct	gatcagatgc	tccaagctca	atatcagaca	tatagctgag	56772062
acggccagca	gtattgtgac	aagttgcaaa	gtctgccagc	ttaacaacgc	56772112
atacccccaa	tctcaagctg	cagcaggaac	aaggctcaga	ggaaccaggc	56772162
ccagtatcta	ctaagaagta	gatttttactg	aaataaagcc	aagaaagtac	56772212
aggtagccgt	actttactgt	ctttgttagat	actttttcaa	ggtggactga	56772262
agcattccca	accaaagag	aaactgtctca	ggtcgtagca	aagaaaattc	56772312
tggaagatat	ccttcccaag	tatggcttcc	ccatccagat	aagggtcaaat	56772362
aataggcccc	ctttcgctgc	taaagtaagt	caagacttgg	cttccatcct	56772412
taaggcaaat	tagaaactac	attgcgctta	caggccccag	agttcaggac	56772462
aggtagaaaa	gatgaatcgg	accttaaaaag	agaccttaac	taaattgact	56772512
atagagactg	gcgctaatta	ggtagtcctt	cttccctatg	ctctgttcca	56772562
ggcccgtaat	accctttaca	aactagccct	cacccttac	gaaatcatgt	56772612
atggcgagacc	tccacccctg	gttccttagct	taaaagatga	cctgcttaag	56772662
tctgaaacag	aaaatgtctc	tgaattctta	ttttccttac	aagccttaca	56772712
gaaaattcac	caagaaatct	ggcccaagct	gagagagcta	tatgagacca	56772762
gtccccccacc	gacaccccat	ccgtaccagc	caggagactg	agtccctggtt	56772812

aagcgacacc	gacaagagac	cctagagccc	aggtggaaag	gaccactcca	56772862
agtactcctg	accacaccca	ccgccctgaa	agtagaaggc	attgcgtcgt	56772912
ggatccacta	caccacacgtc	aagccagtg	acccaacctc	cgaccttcta	56772962
aggccaatca	tggcgggcgt	ggctaaagca	ccggacacgt	ggactgtgga	56773012
cagagctaag	aacaaccctt	taaaactcac	cctgcgccgg	cagcataact	56773062
cactgcaaac	atgcagttaa	gtagtctaac	tctaacatta	gtcgccctag	56773112
tggccgctga	ggaaaacata	aagccagctc	ctaattccctt	tgtctggaga	56773162
ttctggcttt	atgaaaacca	aaccacacct	aggcaacctc	ctaagcccgg	56773212
gaaattagtg	gccagtgca	attgccccctc	ctcaagggtgc	aatagcccaa	56773262
ttttactaaa	ttttaccgat	ttcccagtag	ccaaaccagt	ggcaccaata	56773312
atatgcttcg	agtatgatca	gactgaatac	aattgttaagc	actattgggtg	56773362
gcaccaaagt	gccggctgcc	cttataactg	ttgtaacatc	cataaatacc	56773412
aatggtaagg	tggagaagaa	cagatagatc	ccagatggcc	cttccatcgc	56773462
agacgagata	gagacctttc	atatacatgg	atagttaaag	acccctggaa	56773512
ctcccgtcgg	accacgcctc	aacacagggc	tgtatactac	tcctctgcct	56773562
ccacatggcc	tagcagtcac	ctctatctgt	ggcgggggtct	agtgcaggta	56773612
cggccctcgg	tccattggaaa	tatccagcga	caagaaaacc	gcctgacaca	56773662
agatttacgt	cttttttcct	ggttaaaatt	attgcaagaa	agattagaac	56773712
ttgccaacct	tacaggactt	cacagcctgt	ctggctgctt	tctattgcca	56773762
ctctaaggcg	tccaccgcta	accgctgtcc	ccctgccatg	aagatcatcc	56773812
acctctgccc	aagctaacaa	ccaccaaaac	ctctcatatg	cccctatccc	56773862
taacgtgcca	ctatacctaa	accccagtc	agagaagttt	ccctagtgtt	56773912
tctcaggaac	taattccagc	ctctgcaaca	tactgcaac	gccccctaac	56773962
atcaccctaa	aggctccgtc	aggcataattc	ttctgggtga	atagaacatt	56774012
atctaaaaac	ctatcaagcc	cctctgtttac	caacctactg	tgtcttcctg	56774062
tcacattagt	tccccgggta	actctcctta	ctgccgggtga	gttcctaaag	56774112
tataccagta	actggactag	tgtctgttatt	caccagacc	ctagaccaag	56774162
acctgcacga	gccatatttc	tccccctcat	tgcaggaatc	tccttcaccg	56774212
cataccttcat	ggcggccgga	ctggctgggg	gagccctagg	tcacaccctc	56774262
atagaaagta	acaagctgta	ccaacaattt	gccgttgcta	tagaagagtc	56774312
agctgagtc	cttgccctccc	tccagcggca	gctcacgtcc	ctagcacagg	56774362
taaccttgca	gaaccggagg	gccttagacc	tactcactgc	tgaaaaagga	56774412
ggaatgtgta	tgtttctaaa	ggaagactgt	tgtttctaca	taaataaatc	56774462
aggactcgta	gaagaccaag	tccaacagtt	atgcaagtta	agcacagaag	56774512
taagaacacg	gcagtttgct	tcagctgcag	accaatagtg	gaactcatct	56774562
atgttttcac	ttgtagcccc	cttccttgga	cccctgctga	gtctactatt	56774612
tctgcttacc	gtaggacctt	gtgttggttaa	cagaattttg	cggttcatta	56774662
aagaaaagtt	taacactgta	caactcatag	tcctcagagc	ccaataccaa	56774712
cctgtaaacy	ctgaaacaga	atcagactta	taagacccaa	gatttggtct	56774762
aaaaaatacc	tgaaaagaaa	ggggggggaat	gaaaagagcc	aggcacattc	56774812
ctcagccccg	ggctcaaaac	aaacaagccc	agtacaaaca	catcccattc	56774862
tcccattccc	ccacatatca	ccatatactc	cttaaaacttc	ccccgggctc	56774912
aaaacaaaca	agcccagtac	aaacaccacc	aggaaagtct	ccgataagag	56774962
gacagatgag	gggacagccg	ttcaaagttt	tactgaaaga	gcgggaacca	56775012
aaagaattcc	tttggtcccc	tgtaactttc	aggctataaa	aaagcaaaca	56775062
ctcgcatgtg	tcagggccct	cttggtatgtg	gtggaatgga	aggaccagg	56775112
tcgaactttg	agtaaagatc	cttgccgctt	ggctttgact	ctggactctg	56775162
gtggtcttct	ttgaaagata	aacggctctag	gcataaacact	gggactgccg	56775212
gcacatgcca	ccacaactat	ctaattttgt	ttattttttg	tagagatgag	56775262
gtctcactat	gttgcccagg	ctaattttga	actcctggct	caagtgatcc	56775312
tcctaccttt	gcctgccaga	gtgctgggat	tatagggtgtg	agccaacatg	56775362
cccagctggg	gattttattt	ctaagcattt	taaaatcaat	gcaggaaaca	56775412
aatgctatat	agcaggtgtc	gtaatatcaa	atgtgtccat	ttcacatagg	56775462
gcttagcctt	tatcatgtta	ttctaacata	agaaaaacta	cgccaaaaat	56775512
atagtgtagc	caggtgcagt	ggctcacacc	tgtaatccca	gcactttgag	56775562
aggctaaggc	gggcggtatca	cctgaggcca	ggagtttgag	acctgcctgg	56775612
gcaacatggg	gaaacctcat	atctacaaaa	aatacaaaaa	aattagctgg	56775662
aagtgggtgat	gggtgcctgt	cattgcagct	acttaggagg	ctgaggcatg	56775712
agaatcagtc	aaaccaggga	gggtggaggtt	gcagtgaacc	aagatcacgt	56775762
cactgtattc	cagcctgggt	aacagaacaa	gactgcatct	caaaacatat	56775812
atgaagtttt	taccacctag	tttcctgaga	tctaaaaata	gggattgtct	56775862
aaaaaataaa	aataaaaaaa	tgatttttatg	aatgatgaat	gaaaaagagt	56775912
taaaaggaac	ctttttggct	ggatgtgaag	gtgactcttg	cctgtaatcc	56775962
cagcactttg	ggaggcagag	acaggaggat	cacctgtggt	caggagtttg	56776012

agaccagcct	ggccaaaaag	gtttgagaac	ctttttggct	ggatgcgaag	56776062
gtgactcttg	cctgtaatcc	cagcactttg	ggaggcagag	ataggcagat	56776112
cacctatggg	caggagtttg	agaccaaact	ggagtttgag	atcagccccg	56776162
tctctactga	aaatacaaaa	attatccggg	tgtggtagcg	ggcgccctga	56776212
atcccagcta	cttgggaggg	tgaggcagga	gaatcacttg	aaccagggag	56776262
gcaggagggt	gcagtgaagg	gagatcgcg	cactgcactc	cagcctgggc	56776312
atgatggagc	aagactccat	ctcaaaaaag	aaaaagggtt	tagccaggct	56776362
ggtctcaaac	tcccaacctc	aggtaatctg	cctgccttgg	cctcccaaag	56776412
tgctgggatt	acaggcatga	gccactgtgg	cccgcctttg	aatcctttct	56776462
ttttttgaca	gagtctcact	ctgtcaccca	ggctggagtg	caatgggtgcg	56776512
atctcagctc	actgcaacct	ccgcctcccc	ggttcaagca	attctcttgc	56776562
ctcagcctcc	caagtagctg	ggattacagg	tgcttgccac	cacgcccagc	56776612
taattttttg	tatttttagt	agagatgggg	ttttgccatg	ttggccaggc	56776662
tggtctcgaa	ctcctgacct	caggtgatcc	accacctca	gcctcccaaa	56776712
gtgctaggat	tacaggcatg	agccaccaca	ctggctgctt	ttgaattggt	56776762
agcttagata	atttataacc	agtgttgggg	ttggctaagc	aatgaacatt	56776812
ggcctagctt	aatgagaaat	acataaaagt	aaaatgaaga	gttttctcag	56776862
agctcccaat	ctcattttca	agaaatgaca	aaatcagcca	ggagcggtgg	56776912
ctcacacctg	tgatcccagc	actttcggag	gccgagggtg	gaggattgct	56776962
taagcacaga	ggttcaagac	cagctgagca	acgtggcaaa	actctgtctc	56777012
taaaaaatta	ttcagggatt	gtggtgtgca	cctgtcatcc	tggtactca	56777062
ggaggctaag	gcaggatgat	cacctgagcc	cagaagttca	aggctgcagt	56777112
gagccatgac	tgtaccactg	cactccagcc	tggtgacag	ggtgagaccc	56777162
tgtctcgtaa	aaccaaaaag	gaatgacaag	attggttggg	tgcaagtggg	56777212
catacctgta	atcccagcac	tttgggaagt	gaggcaggag	gagtttgaga	56777262
gcagcctggc	aacatagcaa	gaccctacaa	aaaatttaaa	aagtagccag	56777312
gtgaggtggg	gcacacttgt	agtctcagct	actcgagagg	ctgaggcacg	56777362
agaatcactt	gaaccaggga	ggcagaggtt	gcagtgaagg	gagaatgtgc	56777412
cactgcactn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777462
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777512
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777562
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777612
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777662
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777712
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777762
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777812
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777862
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777912
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777962
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778012
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778062
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778112
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778162
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778212
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778262
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778312
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778362
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778412
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778462
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778512
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778562
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778612
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778662
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778712
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778762
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778812
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778862
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778912
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56778962
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56779012
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56779062
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56779112
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56779162
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56779212

taattttttt	tgtattttta	gtagagacag	ggtttcacca	cgttagccag	56779262
gatggtctcg	atctcctaac	ctcgtgatct	gcccgcctcg	gcctcccaaa	56779312
ctgctggaat	tacaggcatc	agccaccgtg	cccagcctaa	tttttgtatt	56779362
tttagtagat	acagggtttc	accatgttgg	ccaggctagt	cttgaactcc	56779412
tgcactcagg	tgatttgccc	acctcggcct	cccaaattac	tgggattaca	56779462
ggcatgagcc	accatgcctg	gcctgccata	tagttttatt	attattttatt	56779512
ttattttttg	agacagagtc	tcactctctc	acccaggctg	gagtgcagtg	56779562
gcgtaatcac	tgttcactgc	aacctccgcc	tcccgggtaca	agcagttctg	56779612
ctgcctcagc	ctcctgagta	gctgggatta	caggcacctg	ccaccacgcc	56779662
cagctatttt	ttgtattttt	agtagagaca	gggtttcacc	atgttggcca	56779712
ggctggctct	gatctcttga	cctcagatga	tcctcccgcc	tcagcctccc	56779762
aaagtactgg	gaatacaggt	gtgagccacc	atgctcagct	cctgccatct	56779812
agtttagaac	tcaaacattt	acctactgat	gtgattttta	tttttgtcgg	56779862
catttgtctt	ggttataaga	gggcagggac	cacatcttac	tctccagaag	56779912
ccccgacatt	gtccagctga	atgggtcccc	agcagtgggtg	catgtactga	56779962
tgcaggagag	tgcagtgagg	gataggagga	agccaggagg	cgaagatggg	56780012
aggaggttaag	gaggaagggtg	ctgtgtactc	agcaaaggca	ccctggggga	56780062
ggccaggctt	acctctcttt	gcaggaatag	tttatgtttg	aaacacttgg	56780112
ctgggtgcgg	tggctcacgc	ctgtaatccc	agcactttgg	gagcccgagg	56780162
cgggcgaatc	acctgagggtc	aggagtttga	gaccagcctg	gccaacatgg	56780212
tgaaccacca	tctctactaa	aaatacaaaa	attagccagg	tgtggtggca	56780262
ggcacctgta	atcccagcta	ctcgggaggc	tgaggcagga	gaatcacttg	56780312
aaccaggag	gcggagggtt	cagtgggcag	agatcacgcc	attgcactgc	56780362
agctggggca	acagagtggg	actccgtctc	aaaaaaaaaa	aaaaaaaaag	56780412
tttgaaacac	ttactgctct	ctattgctcc	tgcaaccttt	gttttctgat	56780462
tttctaaaaa	tcatttctcc	gccagggtgt	gtgggtcata	cctataatcc	56780512
cagcaccttg	ggaggccgaa	gtggggcaaat	caccagaggt	caggagtcca	56780562
agaccagcct	ggcccaacat	agtgaaaccc	catctctact	aaaaacacaa	56780612
aaattagcca	ggcgtggtgg	cagggtgctg	taatcccagc	tacttgggag	56780662
gctgaggcag	gagaaccact	tgaacccggg	aggcggaggt	tgcagagagc	56780712
agagatcaca	ccactgcact	ccagcctggg	caacagagcc	tccaaaaaaa	56780762
taaataaaaa	taaaaatcat	ttatctctct	tttcttactt	cttagGATGC	56780812
TTTCATTGGA	TTTGGAGGAA	ATGTGATCAG	GCAACAAGTC	AAGGATAACG	56780862
CCAAATGGTA	TATCACTGAT	TTTGTAGAGC	TGCTGGGAGA	ACCGaAAGAA	56780912
TAACATCCAT	TGTCATACgG	CTCCAAACAA	CTTCAGATGA	ATTTTACAA	56780962
GTTACACAGA	TTGATACTGT	TTGCTTACAA	TTGCCTATTA	CAACTTGaTA	56781012
TAgAAAGTTG	GTACAGATGA	TCTGCACTGT	CAAGTAAACT	ACAGTTAGGA	56781062
cTCCtCAAAG	ATTGgttttGT	TTGTTTGTtTT	TTAACTGTAG	TTCCAGTATT	56781112
ATATGATgAC	TAttGATTTC	CTaGAGAGTT	TTGTAATCTG	AATTCTTTAT	56781162
GTATATTCCT	AGCTATATTT	CATACAAAGT	GTTTTAAGAG	TGGAGAGTCA	56781212
ATTAAACACC	TTTACTCTTA	GGAATATAGA	TTCGGCAGCC	TTCAaTGAAT	56781262
ATTGGTTTTT	TTCCCTTTGG	TATGTCAATA	AAAGTTTATC	CATGTGTCAG	56781312
AAcagatttg	tgggaatttgc	agttctgcct	ctcggatatt	ggaatatcca	56781362
atatttcagg	tactgtgttg	aagcataaac	atcccccagg	caagaaacag	56781412

gc

Figure 16

BC065228 (1-841):

CCGCAGGCTT	CTGCCTGGCC	GcCGCCGCCT	ATAAGCTACC	AGGAGGAGCT	50
TTACGACTTC	CCGtCCTGCG	GGAAGTGGCG	GGCACGATCG	CAAGGTAGCG	100
CAGAAGCTTC	TCAATGGCCA	GCGCCAGCTG	CAGCCCCgGC	GGCGCACTCG	150
CCTCACCTGA	GCCTGGGAGG	AAAATTCTTC	CAAGGATGAT	CTCCCACTCA	200
GAGCTGAGGA	AGCTTTTCTA	CTCAGCAGAT	GCTGTGTGTT	TTGATGTTGA	250
CAGCACGGTC	ATCAGTGAAG	AAGGAATCGG	ACGGAGTCTC	GCTCTGTCAC	300
CAGGCTGGAG	TGCAATGGTG	CAATCTCGGC	TCACTGCAAC	CTCCGCCTCC	350
TGGGTTCaG	CAgTTCTCCT	GCCTCCACCT	CCTGAGTAGC	TGAAACTACA	400
GGATGCTTTC	ATTGGcTTTG	GAGGAAATGT	GATCAGGCAA	CAAGTCAAGG	450
ATAACGCCAA	ATGGTATATC	ACTGATTTTG	TAGAGCTGCT	GGGAGAACCG	500
gAAGAATAAC	ATCCATTGTC	ATACaGCTCC	AAACAAC TTC	AGATGAATTT	550
TTACAAGTTA	CACAGATTGA	TACTGTTTGC	TTACAATTGC	CTATTACAAC	600
TTGcTATAGA	AAGTTGGTAC	AGATGATCTG	CACTGTCAAG	TAAACTACAG	650
TTAGGAaTCC	TCAAAGATTG	GTTTGTTTGT	TTTTAACTGT	AGTTCCAGTA	700
TTATATGATc	ACTATTGATT	TCCTgGAGAG	TTTGTGAATC	TGAATTCTTT	750
ATGTATATTC	CTAGCTATAT	TTCATACAAA	GTGTTTTAAG	AGTGGAGAGT	800
CAATTAAACA	CCTTTACTCT	TAGGAaaaaa	aaaaaaaaaa	a	

Chimp Genomic chr7 (56755158-56781338):

aatcggaaact	cgaggcgggg	ctgttggtct	tccgggagcg	cgcattgcgcg	56755208
ggggggccaca	ggctccctgg	ctggcgggcag	ctaaagtggc	tgttggtgt	56755258
CCGCAGGCTT	CTGCCTGGCC	GtCGCCGCCT	ATAAGCTACC	AGGAGGAGCT	56755308
TTACGACTTC	CCGgCCTGCG	GGAAGTGGCG	GGCACGATCG	CAAGGTAGCG	56755358
CAGAAGCTTC	TCAATGGCCA	GCGCCAGCTG	CAGCCCCtGC	GGCGCACTCG	56755408
CCTCACCTGA	GCCTGGgtac	gtgcgccccca	caacaccttc	cccagccagg	56755458
gccccggggac	ccccgggagcg	tcccccgcca	cctggcgccg	ctcatcctgg	56755508
gcaggggtggg	accccaactga	ggccgcccacg	cattagggag	ctgcactccc	56755558
gagtttgacg	cccagttgta	atagcattaa	gtctttgaaa	ttttgtagcg	56755608
gggtagaagg	ggctaggaaa	ggaagaaaac	atctttttta	aaatataagc	56755658
gatcgggccg	gcgcggtggc	tcacgcctgt	aatcccaaca	ctttgggagg	56755708
ccgagggcagg	tggatcacga	ggtcaggagt	tcaagaccag	cctggccagc	56755758
atggtgaaac	cctgtctcta	ctaaaaatac	aaaaattaac	cgggcgtggt	56755808
ggcggggcgcc	tgtaatccca	gcaactcggg	aggctgaggc	agagaactgt	56755858
ttgaaccagg	aaggcggagg	ttgcagtgag	ccgagatcgc	gtcactgcac	56755908
tccagcctgg	gcgacagagc	gagactccgt	ctgaacaaac	aaatatgtgt	56755958
atgtatataa	gtgatcgaac	ccgggaggtt	gagtttacag	taagccgaga	56756008
ttgcgccact	gcactccagc	ctgtgtgaca	gagtgagaaa	atgtctctaa	56756058
aaaattatat	gcaagtgaga	gctttttttc	cagcgctcat	gctcagactg	56756108
aaagaaagtaa	ttgggcccgg	gcgttagctc	atgtccataa	tcccagcact	56756158
ttgggaggcc	gaggcaggcg	gatcgcttga	gctcgggagt	tccagactag	56756208
cttgggcaac	atggtgaaac	cctgtctcta	taaaaatata	gaaaattatt	56756258
gggcatggtg	gcatgcgctt	gtagtccccg	ctacttgagc	ggctgaagcc	56756308
gaaggatcga	cctcctgggc	tcaggcaatc	catttcagcc	tcccaaattg	56756358
ctgggattac	aggaatgagc	catcgtgcct	ggccttacac	tatattttaa	56756408
tgcttttttt	gaaaatggaa	actttttacg	gcaattcact	ttcttcaaac	56756458
taatgataag	gaaatgatgc	tgttctgttc	tgttttgttt	tttggttttg	56756508
tctgtttctt	ttttttcctt	tatcgagaca	gggtcttgct	ctgttgccca	56756558
ggctggagtg	aggcggtgaa	atcatggctc	actgcagcct	tgacctccgg	56756608
gctcaaggaa	tcgttcccgg	tcagcctccc	cagtagctag	gactacaggt	56756658
gcattgctacc	atgcttggtc	aattcttttt	ttgaaatgga	gtctcactct	56756708
gtctcccagg	ctggagtgc	gtggtgca	ctcagctcac	tgcaggctgg	56756758
tctcaaaactc	tgacttcaga	tggtccatcc	acctcagcat	cccaaagtgc	56756808
tgggattaca	ggtgtgatcc	accacgcccc	atgattttgc	tcatttttaga	56756858
tactacaact	tttttagttta	attttttttt	ttcctgagat	agaatctcac	56756908
tttgtcgcca	ggctggagtg	cagtggcgtg	atctcagctc	actgcagcct	56756958

ccacctcctg	ggttcaagcg	attctcctgc	ctcagttctcc	cgagtagctg	56757008
agactacagg	tgcttgccac	cacgcccagg	agttcaagac	tgagcccagg	56757058
agttcaagcc	tgacgtgagc	catgatcgtg	ccactgcact	ccagcctggg	56757108
caacaagcaa	gaccctgtct	ccacaaataa	gtaaataaac	atcctatctg	56757158
taccagggta	caggcaggcg	ctattattcc	cacttttcat	cctcaactct	56757208
gagttgagtc	atacatcaac	tctgatgacc	tttgtttcca	ctccccttaa	56757258
gccattcaca	ggaacagcta	ctttacatag	aaaagtcaat	cttggagggt	56757308
tgagtaattt	gcctaatac	tgtaaaaggt	agagcttcaa	ttatcaccca	56757358
ggcccatctg	ttgctgggtat	tccctctcca	cccctccgtg	tggtctccct	56757408
ttattttatt	atttttaattt	tttttgagat	ggagttgcac	ttttgttgcc	56757458
caggctggag	tgcaatggca	tgatctcggc	tcactgcaac	ctccacctcc	56757508
caggttcaag	caattctcct	gcctcaggct	cccaagtagg	tggtgattaca	56757558
ggtgcctgcc	accaagccca	gctaattttt	tgtactttta	gtagagatgg	56757608
ggtttcacca	tggtgaccag	gctgggtcttg	gactcctgac	ctcaggtgac	56757658
ccacctcct	cagcctccca	aagtgcctgag	attacaggcg	tgagtcacgg	56757708
tgccctggcct	tggttttctt	ttaataaaatc	acgactcaat	ttgatgtgga	56757758
tgctctgttc	attcccagaa	tataatgaat	ggattatata	tatatttttt	56757808
gtttttgaga	cagagtcttg	ctttgtcttg	caggctggag	tgagtggtcg	56757858
cgatctcagc	tcattgcagc	ctctgcctcc	cagggtccag	caattctcct	56757908
gcctcagttc	cctgggttagc	tggtgattacg	ggcactcacc	accacgcca	56757958
gctaattttt	tttttttttt	tttttttttt	tagtaaat	caccgtgtcg	56758008
tccaggctgg	tctcaaactc	ctgacctcag	gtgatctgcc	cgacttggtc	56758058
tcccaagtg	ctgggattat	aggcatgacc	caccgcccc	agcctcaaca	56758108
tcaataaat	ttaaaagttg	gtaacagcca	ggcgagtg	ctcacacctg	56758158
taatcccagc	actttggggg	gccaaggcgg	gcagatcacc	tgaggtcnnn	56758208
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56758258
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56758308
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56758358
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56758408
gagttcaaga	ccagcctggc	caacatgggt	aaaccccgct	tctaataaaa	56758458
ataaaaaata	aaaaaaacag	ttagctgggt	atgggtggcat	gtgcctgtaa	56758508
tcccagctac	tcaggaggct	gaggcacgag	aatcactaga	acctaggagg	56758558
tggtggttgc	agtaagccaa	gttcgtgcca	ctaccctcca	gcctgggcaa	56758608
cagagtaaga	ctccatctaa	aaaaaaaaaa	aaatgaagaa	gaagaaatta	56758658
gtgtagtgtg	ggaagtga	aaaaaaaaaa	agaaaaggaa	aagaaaaatg	56758708
attgaattca	tgaatacact	cttatgtggc	ctgcaccgac	tttgacacaa	56758758
attagattgg	cttagtaggc	aagggtggga	tcttttcata	attttatttg	56758808
atgtctaaaa	tacatttatc	tttttttctt	atagGAGGAA	AATTCTTCCA	56758858
AGGATGATCT	CCCACTCAGA	GCTGAGGAAG	CTTTTCTACT	CAGCAGATGC	56758908
TGTGTGTTT	GATGTTGACA	GCACGGTCAT	CAGTGAAGAA	GGAATCGgtg	56758958
agtagccaa	aatctgtggg	gtagaggacg	cggtgtcaga	aatgtaggga	56759008
tagcatttat	tcaccttatg	aaatgataaa	gaatttcttt	tttttttttt	56759058
tttgagacag	agtctcactc	tattgcccag	actggagtgc	agtggcaca	56759108
tcttggtcca	ctgcaacctc	cacctcccag	gttcaagcaa	ttatcctgcc	56759158
tcagcctcct	gagtagctgg	gattacaggc	gtgtgcaacc	acatccggct	56759208
aattttttta	tttttagtag	agatggagtt	tcaccatggt	ggtcaggctg	56759258
gtctcaact	cctgacctgc	tgatccaccg	gccacagcct	cccaaagtgc	56759308
tgggattaca	ggcatgagcc	accatgcccc	gcatgaaat	gataaagaat	56759358
ttctaaagag	tggtctgttt	ggatgaagtg	ctggaccttg	gctatggaac	56759408
atgagcacta	ggctttttct	ttcaccctct	agagttgaat	ttgatataat	56759458
cattcattta	tataaatatt	tatggcaaaa	aaatatggat	aagacatggt	56759508
cttagcttga	ttaggggaat	gatgaccatc	actaaccaaa	gaagatggcc	56759558
tagggggcag	ggctcaagtg	tccagtggct	ttgtgtcatg	gttgaaagtg	56759608
tggccttagt	ggttccctgg	aaaggttcag	agtctcttgg	tggatcctgg	56759658
gtcagagccc	ctctctccct	ccctccctcc	tgccctccca	cctcctgcc	56759708
tgacgtggg	caccacctc	tgacgccc	gtccccag	catgcaccat	56759758
gtcattttct	tttttttttt	tttcagGACG	GAGTCTCGCT	CTGTCACCAG	56759808
GCTGGAGTGC	AATGGTGCAA	TCTCGGCTCA	CTGCAACCTC	CGCTCCTGG	56759858
GTTCAAGCAa	TTCTCCTGCC	TCCACCTCCT	GAGTAGCTGA	AACTACAGgc	56759908
acgcaccacc	acaccgggtc	tatttttgta	tttttagtag	agatgggtgt	56759958
tcaccatggt	ggccagcctg	gtcttgaact	cctgacctcg	tgatctgccc	56760008
gcctcggcct	cccaaagtgc	tggtgattaca	ggcatgagcc	accacaccgg	56760058
gccaccattt	tattttcaac	tccctttcat	ggaagaacgt	ttagcctttg	56760108
gtctctttct	tgatttatga	cagcttgccg	ctttcaagaa	aactacctat	56760158

gaataggctg	tgtaactttt	atattatatt	ttattttttt	gagatgaagt	56760208
ctctgtcttc	taggctgggg	tgcagtggca	tgatctcggc	tcactgcaac	56760258
ctccccctcc	cagggttccag	caattctcct	gcctcagctt	cccaagtagc	56760308
tgggattaca	ggcacgcacc	accacacctg	gctaattttt	gcattttttt	56760358
ttattagaga	tgggggtttg	tcatgttggc	caggctgggtc	tcgaactcct	56760408
ggcttcagat	gagctgcccc	cctcagcttc	ccaaagtttt	ggcattatag	56760458
gcatgagtca	ctgcacccag	cttggttaatg	ttattttcaa	gcacaccttc	56760508
aaaagtttat	tccaaggcct	tgctctcata	gcagcaaagc	ctgtctgttg	56760558
catagtgggg	ctcttggttag	catcttcctt	ttggtgggatg	ttgtgtagac	56760608
ccggggcagt	aaggcatggt	aacctcgagg	acattggacc	tggctggcct	56760658
gactgttggg	tctccctcct	aggacatggc	gagccatggg	tggggcagtg	56760708
tctttcaaag	ttgctctcac	ggagcactta	gccccaatcc	agccctccag	56760758
ggagcagggtg	cagagactca	tagcagagca	ccccccacac	ctgaccccca	56760808
gcataagtaa	gaggagccgc	tgctccaggt	gtatttcagc	accagtgttg	56760858
gggggcatgt	cctcctgaga	gcatctaacg	attgcttttag	agagccctct	56760908
gggtgggttta	tttatttttca	tttttatattt	ttaattttat	tatttttaaaa	56760958
aaattttatt	atattatatt	ttatttatatt	attgaggcag	agtctcgtct	56761008
tggtgcccag	actagtagtc	aatgctacga	tcttggtctca	ctgcaacttc	56761058
tgcttcccgg	gttcaagtga	ttctcctgcc	tcagcctcct	gagtagttgg	56761108
gatttcaggc	accaccatg	ccaggctagt	ttttgtattt	ttagtagaga	56761158
caggatttca	ccatgttggg	taggctgggtc	tcgaactcca	gacctcaggt	56761208
agtctgcccg	cctcagactc	ccaaagtgtc	gggatcacag	gcgtgagcca	56761258
ctgtgcccgg	caaaatttta	tttacttttt	ttttttttga	gacaaagtct	56761308
cactctgtgc	cccaggctgg	agtgcagtgc	tgcaatctcc	actcactgca	56761358
actttttgcct	cccagggttca	cacgattctc	ctgcctcagc	ctcaggagta	56761408
gctgggatta	caggcgctctg	ccaccacacc	tggctaattt	ttgtattttt	56761458
agtatagatg	ggtttcgcca	tggtggccag	gctgggtctcg	aactcttcat	56761508
ctcaagtgat	ccacctgcct	cggcctccca	acatgctgag	attacagggtg	56761558
tgagtcactg	cacctggcct	attattattt	tttttagatt	cagagtctca	56761608
ttctgttgcc	cagggtggag	tacagtggta	taatcatggc	ccactgcagc	56761658
ctcgaactcc	tgggggtcaag	tagccctccc	accttagcct	ccttagtagt	56761708
agggactgca	ggtgcatgcc	accatgcccc	gctaattaaa	gaattttttt	56761758
ttttttttct	gaaagtcttg	ctttgtcacc	caggctggag	tgcagtggca	56761808
tgaacacagc	tcgctgcagc	atcaatcttc	tgggctcaag	ggatcctcct	56761858
gcctctgctt	ccaaagtagc	tgggactaca	gacataggcc	accatgccat	56761908
tttttttttt	tttttttttt	gagacggagt	cttgctctgt	tgcccaggct	56761958
ggagtgcagt	ggcgccatct	tggtcactgt	caagctccac	ctcctggggt	56762008
cacgccattc	tcttccctca	gcctcccag	tagctgggac	tacaggcgcc	56762058
tgccaccaca	gctgactaat	tttttgatt	tttagtagag	acgggggttt	56762108
accatgttag	ccaggatggt	ctcaatctcc	tgacctgtg	atccacctgc	56762158
ctcagcctct	caaagtgtctg	ggattacagg	cgtgagtcac	cgcgcccggc	56762208
ccgcggtgcc	taattttaaaa	aaaaaaattt	tttttataga	cacagcatct	56762258
cactacgttg	cccaggtctg	tcttgaactg	gtgggcttaa	gtgaccaatt	56762308
ttcacaactc	agcctcccaa	agtgttgga	ttatagggtt	gagccactat	56762358
accagcctt	taagaatctt	ctttagagag	tggaaatctg	ctgtgttgcc	56762408
caggctgttc	tcaagctcct	ggcctcaagt	gattttctac	cttggcctcc	56762458
caaagtgtct	gaattaaagg	catgggctgt	gtggttcttt	tggcacttac	56762508
ccgtggtctt	gtctggccgg	ttgtctggtt	ctgtctgttt	ctgcctttcc	56762558
tctttctcca	gggaaaacct	aagctttcct	tgtttgtcct	catcttgtgt	56762608
ttttctgggt	ccatgggcag	agtagagttc	tagaacggtt	tcctaaagca	56762658
gccaagccct	accattgat	ttctaagtac	atttagaaaa	acacatataa	56762708
ggctgggcat	ggtggctcgc	acctgtagtc	ccagcacttt	gggaggccga	56762758
agtggatgga	tcaccaggtc	aggagttcga	gaccagccta	accaacatga	56762808
tgaaaccttg	tctctactaa	aaatacaaaa	attaagccgg	gcacggtagc	56762858
tcacgtctgt	aatcccagca	ctttgggagg	ccaagatggg	tggtatcacc	56762908
tgaggtcagg	agttcgagac	cagcctggcc	cacatgggtga	aaccnnnnnn	56762958
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763008
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763058
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763108
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763158
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763208
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763258
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763308
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56763358

nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnctgt	56763408
gagccgagat	catgccagtg	cactccagct	agggcaacag	agcaagattc	56763458
catctcaaaa	aaaaaaaaaa	tcagccgggc	attgtggccc	cggcctgtaa	56763508
tactcaggag	gctgaggcag	gagaatcgct	tgaatccggg	aggtggagggt	56763558
tgagtgagc	caagatcgca	gccttgcact	ccagcctggg	tgacaagagt	56763608
gaaactctgt	ctcaaaaaat	aaagagaaaa	acatatagaa	aacattaaca	56763658
ccccaggcag	tataccttgt	caaacatacc	tcaggcaaat	gcattcagga	56763708
gaagaaaata	catctttattt	ccctcttcat	gtttcgtttt	tttttttttc	56763758
ttttgtatta	ctcagtggtg	ggtattttgt	atttttattt	gcgggggagct	56763808
ggtaagtgcg	ctacaggagc	gaaatgttca	ggttttccta	atatctgggtg	56763858
gctttaggag	tattgtagag	catgttgctt	caaagctcaa	tatcccagca	56763908
accaatgtat	ttgccagtag	gctgaaattc	tactttaatg	gtaagatgtt	56763958
aacggtaaca	tgttcccttt	cttagcagtt	ccattattcg	gtattctggg	56764008
taatgtcttt	tggaatgcaa	cttgaacagt	cacatcagag	ttaaatattg	56764058
agatgaatgg	ttctcttaaat	tgatgtcttt	tgcccttttg	tacactttgc	56764108
tcagcaaaac	agacttagat	catcacctgt	cttagcttta	ttatctataa	56764158
catcacctgt	ctgtacataa	atgtctgcac	ttcatgaaac	attaattcct	56764208
acatagcccc	tatgactttac	tttgctatgt	gttgattttc	aatttttatgg	56764258
gagcttctat	gcatttgaag	caatttagag	tcaaggagga	gacttttggc	56764308
tgggcacggg	ggctcacgcc	tgtaatccca	gcactttggg	agcctgaggc	56764358
aggaggatca	cttgagggtca	ggagttcgag	accaccctgg	ccaacgtggg	56764408
gaagccccaa	ctttgctaaa	aaaaaaaaaa	aaaaaaaaaa	aaataggaaa	56764458
attagctggg	catggtggca	catgcctgta	gtcccagcta	cttgggaggc	56764508
tgaggcagga	gaatcacttg	aagctgggag	gcggagggtg	cagtgcgcgg	56764558
agatgcgcgc	actgcactta	cagcctgggt	gacagagcga	tactctgtct	56764608
ttaaaaaaaa	aagacagaga	ctttagaagg	gagacttttg	ttaacagatt	56764658
actgaactac	agctgccctt	ggggaagcct	cctatcccct	ctctgggtcca	56764708
tagcggtgcc	acagctgtta	tattatctca	gatctccagt	tatatcaagg	56764758
agtgaactag	agtgatattc	tttttttttt	tttttttttt	tgagacaggg	56764808
tctcactctg	ttgctccggc	tggagtgcag	tggctcgatc	tcggctcact	56764858
gcaacctctg	ccgcctgggt	tcaagcgatt	ctcctgcctc	agcctcccaa	56764908
gtagctggga	ttacagggtgc	ccatcaccac	gcctggctaa	tttttgtaca	56764958
gtgcccacc	gtacttttgt	tttagacagc	tgattttgtc	tataatccca	56765008
gcactttggg	aggctgagggt	aggaagatcg	cttgagcaca	ggagtttgag	56765058
accagcctgg	gcaacatagt	aagaccccc	ctctacaaaa	tataataaata	56765108
aaataaaaaa	aaaacagctg	atttagaact	atataagatt	taaaacagga	56765158
gttaaacaaa	aacaactctc	attacaagat	aaaaacaaaa	taataaaaaac	56765208
aaattcaatc	ttgttatgtc	aactgacttt	gtgcaaacc	aagaaaagtt	56765258
atgaataagg	aactaaaaac	cttttaaaac	ttttaacaat	gccttatccc	56765308
tgtagcatgg	gcaccacctt	ttagcctagg	tgtgtgctca	ataagagttg	56765358
atttcatttc	cccaataaat	accatcaaca	actcataacc	gagattttta	56765408
catcatttta	tactttatat	actagttcct	catatgttac	ctcatttaata	56765458
cctcataaca	gccctgtgaa	gtcaaatatt	atgtttgggt	ttgggttttt	56765508
tagaaagagt	ctcactctgt	caccagggcc	ggagttcagt	tgtgtgatcg	56765558
tagcttatca	cagccctgaa	cttctgggct	caaacgatcc	tcccacctca	56765608
gcctcctggg	tagctgggat	tacagggtatg	tgccactgca	tcccacacca	56765658
ccacagcatt	ttattctctt	gaactaacgt	cattttctga	gactagcaga	56765708
ggtttttatt	ttttttttat	tttttaagtt	attttttttt	gagacggagt	56765758
ctcgctctgt	caccagggct	ggagtgcagt	ggcgtgatct	tggtctactg	56765808
caccctccgc	ttcccagggt	caagcgattc	ttctgcctcc	gcctctcaag	56765858
tagctaggac	tacaagtgca	cgccaccacg	cctggcatgg	tatttttgta	56765908
tttttagtag	agacgggggt	tcaccatatt	ggccaggctg	gtctcgaact	56765958
cctgaccttg	tgatctgccc	gccgtggcct	cccaaagtgc	tgggattaca	56766008
ggcatgagcc	accacgcccc	gccctgggtt	tatttatttt	ttgagacgga	56766058
gtctcgctct	gtcaccaggg	ctggagtgcg	gtggcgcggt	ctcagctcac	56766108
tgcagcctcg	acctcccgga	ctcaagcgat	gttcccatct	cagcctcctg	56766158
agtagctggg	gaaaagagct	gggcacattc	ctcagccccg	ggctcaaaac	56766208
aaacaagccc	agtacaaaac	catectatcc	tcctatccta	ccacatctca	56766258
ccatatatct	cttaaaattc	ccccggggct	aaaacaaaac	agcccagtag	56766308
aaacaccacc	cggataagct	ccgataagag	gacagatgag	gggacagctg	56766358
ttcaaagttt	tactgaaaga	gcgggaacca	aaagaattcc	tttgttcccc	56766408
tgtaactttc	aggctataaa	aaagcaaaac	cttgcatgtg	tcaggggcct	56766458
cttgatgctg	gtggaatgga	aggaccaggt	tcgaacttgt	agtaaagatc	56766508
cttgccgctt	ggctttgact	ctggactctg	gtggtcttct	ttgaagaaca	56766558

aacggtctag	gcataacatc	tgggggctcg	tccgggattc	cccaagccca	56766608
ccagaccctt	ggtcaatgga	tctgctagga	tcgatctact	gataggtgag	56766658
ctggctcgtc	tccgtttgtc	cgtgtctgtt	ctgaatccga	atctgtgact	56766708
cgcgaggtct	gaaactggag	ctggcacagt	cctggcggac	gcgctatagg	56766758
acagccagcg	gagaccagta	ggagacgtcc	cctggctctc	atctgatcta	56766808
tattgcgatc	tgagctgccc	cggtttggcg	ggccgcagcc	gtatctgagc	56766858
tgcccgggtt	atcgggcagc	atctgtctct	gatctgagct	gccccggttt	56766908
ggcaggcagc	agccgtatct	gagctgcccg	gttgatcggg	cagcatctgt	56766958
ctctgatctg	agctgccccg	gtttggcagg	cagcagccgt	atctgagctg	56767008
cccgtttggg	caggcagcat	ctgtctctga	tctgagcagc	cccgttttgg	56767058
cgggcagcag	ccgtatctga	gctgcccggg	tgggcgggca	gcagctgtct	56767108
ctgatctgag	ctgccccggg	ttggcgggca	gcatctgtct	ctgatctgag	56767158
ctgccccggg	ttggcgggca	gcagccgtat	ctgagctgcc	cagttaggcg	56767208
ggcagcagct	gtctctgata	tgagctgccc	cggtttggcg	ggcagcatct	56767258
gtctctgata	tgagctgccc	cggtttggca	ggcagcagcc	gtatctgagc	56767308
tgcccgggta	ggcgggcagc	agctgtctct	aactctaggc	gccccagcgc	56767358
gaccgcgata	caggcagcta	actctgaccc	gggcttccgg	tgccgcgctt	56767408
cgatctgaac	tgccccgggt	tggcgggcag	cagctgtctc	taatctaggc	56767458
tgcccagca	cgaccgcgat	ccaggcagct	aactctgacc	cgggcttccg	56767508
gtgcgcgctt	gcgatctgaa	ctgccccggg	ttggcgggca	gcagctgtct	56767558
ctaacttagg	ctgccccagc	gcgaccgcga	tccaggcagc	taactctgac	56767608
ccgggcttcc	ggtgcgcgct	tgcgatctga	actgccccgg	tttggcgggc	56767658
agcagctgtc	tctaacttag	gctgccccag	cgcgaccgcg	atccaggcag	56767708
ctaactctga	cccgggcttc	cgggtgcgcg	tgcgatctg	aactgccccg	56767758
gtttggcggg	cagcagctgt	ctctaactta	ggctgcccc	gcgcgaccgc	56767808
gatccaggca	gctaactctg	acccgggctt	ccgatgcgcg	cttgcgatct	56767858
gaactgcccc	agtttggcag	gcagcagctg	tctctaactt	aggctgcccc	56767908
agcgcgaccg	cgatccaggc	agctaactct	gacccgggct	tccgggtgcg	56767958
gcttgcgata	tgaactgccc	cagtttggcg	ggcagcagct	gcctctgacc	56768008
aaggctgcca	aagtgcgcct	gcgattagat	atcattaata	agtcagatca	56768058
gatttccctt	acaaggattc	aaacccacat	tcctttacag	gaaaagacta	56768108
caaattatta	caggatagg	aataccca	gcactcctct	atctctcctt	56768158
acgagtaatt	tcaaagaagt	tagagcaaga	ggccatgatc	ttggtataga	56768208
aatcaggaaa	agaaagctaa	ttactctgtg	tcgctccgaa	tggcctgcct	56768258
ttgatgtggg	gtggccgccc	gaaaggacct	tccgacttgc	tgtcatcact	56768308
aaggtaaaat	ccaagatttt	cctacctagg	cgtgcaggcc	acttagatca	56768358
tatcccatat	atcctcatat	ggcaggacct	tgttgagaac	ccgcctcctt	56768408
agctgtcccc	tttccaattg	gcctctgaac	actgtaaggc	actagttgct	56768458
cgaccactaa	aatccaggca	accaactgcc	ccccccatcc	tgttttacct	56768508
gacagcgggg	ccccactggt	cacagaaccc	cctccgtacc	cctccgggcc	56768558
ccaggcccca	gcccccttgg	ctgagctgcg	ggaaggagca	ggcggatggg	56768608
aggcggtctg	cacacacggg	cccgttgaaa	aggaaagtaa	ctttgaaagg	56768658
ccggtgggga	ggacgcgaag	gcgcacttcg	cggactagcc	ccccctcagc	56768708
gcctgactcc	acggtggctt	taccccttcg	ggaaatagga	cccccgatg	56768758
acacaggaat	ccccaggctc	cagtactggc	cattctccac	cagtgatctg	56768808
tataactgga	aaactcagag	tgctcgggtt	tcagacaacc	ccaaagattt	56768858
actggcttta	ctagattctg	tcattgttca	ccaccagccc	acttgggatg	56768908
attgtcagca	gtcctcctga	attttgttca	ccacggaaga	gcgagaaaga	56768958
atacagatag	aagctagaaa	gctagtcccg	ggagacgacg	gtcaaccgat	56769008
tgccaacccc	gacctcataa	acgcaacctt	tcctctgacc	aggccagcgt	56769058
aggactacaa	cacggcagaa	ggtaggggac	ggctacacct	ttatcgccag	56769108
actctaattg	cgggtctccg	ggcagctgct	cgcaagccca	ctaatttggc	56769158
taaagtatat	tctattctgc	aaggaaagac	agagagccca	gctacgtact	56769208
tagaaagatt	aatggaagct	tttagacagt	acacccccat	agatccagag	56769258
gctccaggaa	gtcaggcagc	tgttgtaatg	tctttcgtaa	atcaggcagc	56769308
cccagatatt	aagagaaaac	tccagaaatt	agaagactta	gaaagaaagc	56769358
ggattcaaga	cctccttcag	atagcccagc	aggtttacaa	taacagagat	56769408
actccagagg	aaaagcaatt	taaggccgct	gaaaaaaatg	accaaggtcc	56769458
tggcagcagt	ggtacagaaa	aagcatctac	agccagagta	cacccaacct	56769508
aggcgggccc	ccgggcagta	taatctgagc	aaagaccaat	gtgcctattg	56769558
taaaagggct	ggccactagg	taagagactg	ccccaaaaag	aaaccacgag	56769608
gacaaggacc	cggaccccc	aagtctacac	ccgtactagt	cactcaagac	56769658
gaagactaaa	gaagacgagg	tccggacccc	ctccccaac	ctagagtaac	56769708
tttgcaagtg	gaggggtccc	cagtcagtt	cttggtcgtat	atgggagcac	56769758

agcactcgggt	cttagttaa	actaatagga	aattatcctc	caaatectcg	56769808
taggtacaag	aggccacagg	agttaagaaa	tacccataga	caacacaaag	56769858
aacagtaaac	ctcggagcca	agaatgtaac	ccattctttc	ctagtcatcc	56769908
ctgaaagccc	ctgtccccta	ttgaggagag	acctgctaac	taaaatagga	56769958
gcacagatcc	atcttctccc	tgagaggccc	gtcgtgacca	acccccacaa	56770008
tagaaccgtg	tccatcctga	ctaaaaacct	agaagataag	taccggctcc	56770058
accaggagaa	agcggccctt	gaccaagaca	tagcaccttg	gctccagcag	56770108
tatccagaag	cgtaggcgga	aacggggggc	ttaagtctag	caaaacaccg	56770158
tcctgcctta	tttattgaac	ttaagcctag	gacagaccct	gtgcagggtat	56770208
gccaataccc	gatgccccta	gaagccaaga	gaaggattgc	cccgcataatc	56770258
cgccggctcc	ttgaccaaga	ggtcctacgc	ccatgtcact	caccctggaa	56770308
tactccattg	ttgtcggtag	gaaaacctaa	tagtgaagaa	tacagacctg	56770358
tacaagactt	aaaagaaatc	aacaagaaag	ttgtagacat	acatccaagt	56770408
gtacctaacc	cgtataccct	cctaagtacc	ttaaacccta	aacatcaatg	56770458
gtacactgtt	ttagatttaa	aagatgcttt	ctttagtttg	cctttagccc	56770508
ctcagagcca	aaagctcttc	gccttcgagt	agaatgacc	tgataaaggc	56770558
ataagtggc	aactgacatg	gaccaggcgc	cgcaaggatt	caaaaactct	56770608
cctaccctgc	ttaaaagaag	gcagcgctgg	ctaacagatg	cccggaaaca	56770658
aactgttctg	cagatcccca	ggccacaatc	cacccgacaa	gtgagagagt	56770708
tcctaaggte	ggcaagattt	tgccgactat	ggatacctag	gttcgcagaa	56770758
ctagctaaac	ccttgatatca	ggcaacacgg	gggcaacagc	cattttaattg	56770808
gacagacgaa	gccgagttgg	ccttccaaca	gattaaaacc	gccctactct	56770858
ccgcgcctgc	actaggacta	cctgatgtta	ccaagccctt	ccacttatac	56770908
gtggtatgaa	ataaagggtg	cgccaaggca	gtaataactc	agaacttagg	56770958
cccctggcgg	aggccagttg	cctacctgtc	aaagaagtta	gacccagtag	56771008
ttgccgggtg	gcccccttgt	ctccgaatga	ttgccggccac	ggctctgatg	56771058
gtgcaagatg	ctgataaaact	tgtcataggg	caagaattgc	gggtcggttac	56771108
tccacatgcc	atcgaagggtg	tactcaaaac	gccacctaata	cgatggatga	56771158
gtaacgcccc	gctcacccac	taccaaggag	tactactaaa	tcctctcagg	56771208
ataatttttc	tgcccccaac	gaccttaaac	cctgcctcgc	tgctgccccaa	56771258
cccggacctg	gacgccccac	tccatgactg	caccgagata	ctagctcagg	56771308
tgacaggagt	tcgagaagac	ctgcacgacc	acccacttcc	tgacgctgac	56771358
ctcgtctagt	tactgataaa	gagcagcttc	atgcatcaag	gccagaggta	56771408
cgctggggcg	gcagtaactt	cagagactga	ggtaatctgg	gcggaacccc	56771458
tgcccccg	gacatcgccc	cagaaggccg	aactgatagc	gctcacccaa	56771508
gtctttacct	taaggcgggg	gaaaaagctg	acagtatata	cagacagccg	56771558
atatgtcttt	gcaacggcgc	atatacatga	ggccattttac	aagggaacgaa	56771608
ggttactgac	ggctgaagga	aaaaagataa	aaaacaagag	atcctagccc	56771658
tgctaacagc	cctatggagg	ccagaaaaat	tagccattgt	acattgcccc	56771708
aggcatcaga	aactaactac	tccaactgct	caaggcaact	ttctggcaga	56771758
ccagactgca	agaaatattg	cgaaggctcc	cagccaactc	cttgacttcc	56771808
agctccctga	cccgggcccc	caggacttgc	catatttccc	tgaatattca	56771858
gaacaagatc	tccagtggat	tgacaaaactt	cccctgaaac	aaatccagaa	56771908
taagtgggtg	actgatacta	atgaccaaac	catcctacca	gaaaaattaa	56771958
gacaacagggt	gttagaacac	atccaccgaa	ccaccacact	ggggggccccg	56772008
cggatgatag	acctgatcag	atgctccaag	ctcaatatca	gacatatagc	56772058
tcgacagccc	accagtattg	tgacaagttg	caaagtctgc	cagcttaaca	56772108
acgcataccc	ccaatctcaa	gctgcagcag	gaacaaggct	cagaggaacc	56772158
aggcccagta	tctactaaga	agtagatttt	actgaaataa	agccaagaaa	56772208
gtacagggtac	cggtacttac	ttgtctttgt	agatactttt	tcaagggtgga	56772258
ctgaagcatt	cccaacccaa	agagaaaactg	ctcagggtcgt	agcaaagaaa	56772308
attctggaag	atatccttcc	caagtatggc	ttccccatcc	agataaggte	56772358
aaataatagg	cccgttttcg	tcgctaaagt	aagtcaagac	ttggcttcca	56772408
tccttaaggc	aaattagaaa	ctacattgcg	cttacaggcc	ccagagttca	56772458
ggacaggtag	aaaagatgaa	tcggacctta	aaagagacct	taactaaatt	56772508
gactatagag	actggcgcta	attaggtagt	ccttcttccc	tatgctctgt	56772558
tccaggccccg	taataccctt	tacaaactag	gcctcacccc	ttacgaaatc	56772608
atgtatggca	gacctccacc	cctggttcct	agcttaaaaag	atgacctgct	56772658
taagtctgaa	acagaaaatg	tctctgaatt	cttattttcc	ttacaagcct	56772708
tacagaaaat	tcaccaagaa	atctggoccc	agctgagaga	gctatatgag	56772758
accagtcccc	caccgacacc	ccatccgtac	cagccaggag	actgagtcct	56772808
ggttaagcga	caccgacaag	agaccctaga	gcccagggtg	aaaggaccac	56772858
tccaagtact	cctgaccaca	cccaccgccc	tgaaagtaga	aggcattgctg	56772908
tcgtggatcc	actacacca	cgtcaagcca	gtggacccaa	cctccgacct	56772958

tctaaggcca	atcatggcgg	cgggtggctaa	agcaccggac	acgtggactg	56773008
tggacagagc	taagaacaac	cccttaaaac	tcaccctgcg	ccggcagcat	56773058
aactcactgc	aaacatgcag	ttaagtagtc	taactctaac	attagtcgcc	56773108
ctagtggccg	ctgaggaaaa	cataaagcca	gctcctaate	cctttgtctg	56773158
gagattctgg	ctttatgaaa	accaaaccac	ccctaggcaa	cctcctaagc	56773208
ccgggaaatt	agtggccagt	gcagattgcc	cctcctcaag	gtgcaatagc	56773258
ccaattttac	taaattttac	cgatttccca	gtagccaaac	cagtggcacc	56773308
aataatatgc	ttcgagtatg	atcagactga	atacaattgt	aagcactatt	56773358
ggtggcacca	aagtgccggc	tgcccttata	actgttgtaa	catccataaa	56773408
taccaatggt	aaggtggaga	agaacagata	gatcccagat	ggcccttcca	56773458
tcgcagacga	gatagagacc	tttcatatac	atggatagtt	aaagaccctt	56773508
ggaactcccg	ctggaccacg	cctcaacaca	gggctgtata	ctactcctct	56773558
gcctccacat	ggcctagcag	tcacctctat	ctgtggcggg	gtctagtgcg	56773608
ggtacggccc	ctgggtccatg	gaaatatcca	gcgacaagaa	aaccgcctga	56773658
cacaagattt	acgtcctttt	tcctgggttaa	aattattgca	agaaagatta	56773708
gaacttgcca	accttacagg	acttcacagc	ctgtctggct	gctttctatt	56773758
gccactctaa	ggcgtccacc	gctaaccgct	gtccccctgc	catgaagatc	56773808
atccacctct	gcccagctga	acaaccacca	aaacctctca	tatgccctta	56773858
tccttaacgt	gccactatac	ctaaacccca	gtcaagagaa	gtttccctag	56773908
tgtttctcag	gaactaatte	cagcctctgc	aacatcactg	caacgcccc	56773958
taacatcacc	ttaaaggctc	cgtcaggcat	attcttctgg	tgtaatagaa	56774008
cattatctaa	aaacctatca	agcccctctg	ttaccaacct	actgtgtctt	56774058
cctgtcacat	tagttccccg	gttaactctc	cttactgccg	gtgagttcct	56774108
aaagtatacc	agtaactgga	ctagtgtctg	tattcaccca	gacctagac	56774158
caagacctgc	acgagccata	tttctcccc	tcattgcagg	aatctccctc	56774208
accgcctcct	tcattggcggc	cggactggct	gggggagccc	taggtcacac	56774258
cctcatagaa	agtaacaagc	tgtaccaaca	atgtgccgtt	gctatagaag	56774308
agtcagctga	gtcccttgcc	tcctccagc	ggcagctcac	gtccctagca	56774358
caggtaacct	tgcagaaccg	gagggcctta	gacctactca	ctgctgaaaa	56774408
aggaggaatg	tgtatgtttc	taaagggaaga	ctgttggttc	tacataaatg	56774458
aatcaggact	cgtagaagac	caagtcacac	agttatgcaa	gttaagcaca	56774508
gaagtaagaa	cacggcagtt	tgcttcagct	gcagaccaat	agtggaaactc	56774558
atctatgttt	tcactgttag	cccccttctt	tggacccttg	ctgagttctac	56774608
tatttctgct	taccgtagga	ccttgtgttg	ttaacagaat	tttgccggttc	56774658
attaaagaaa	agtttaacac	tgtacaactc	atagtcctca	gagcccaata	56774708
ccaacctgtg	aacgctgaaa	cagaatcaga	cttataagac	ccaagattgg	56774758
ctctaaaaaa	tacctgaaaa	gaaagggggg	gaatgaaaag	agccaggcac	56774808
attcctcagc	ccggggctca	aaacaaacaa	gccagtcaca	aacacatccc	56774858
atcctcccat	cccaccacat	atcaccatat	atctcttaaa	cttcccccg	56774908
gctcaaaaac	aacaagccca	gtacaaacac	caccaggaaa	gtctccgata	56774958
agaggacaga	tgaggggaca	gccgttcaaa	gttttactga	aagagcggga	56775008
accaaaaagaa	ttccttttgt	ccccgttaac	tttcaggcta	taaaaaagca	56775058
aacactcgca	ttgttcaggg	ccctcttgta	tgtggtggaa	tggaaggacc	56775108
aggttcgaac	ttgtagttaa	gacccctgcc	gcttggtctt	gactctggac	56775158
tctggtggtc	ttctttgaag	aacaaacggt	ctaggcataa	cactgggact	56775208
gccggcacat	gccaccacaa	ctatctaatt	ttgtttattt	tttgtagaga	56775258
tgaggtctca	ctatgttgcc	caggctaatt	ttgaactcct	ggctcaagtg	56775308
atcctcctac	ctttgcctgc	cagagtgaat	ggattatagg	tgtgagccaa	56775358
catgcccagc	tggggatatt	atcttctaagc	atctttaaatt	caatgcagga	56775408
aacaaatgct	atatagcagg	tgtcgtaata	tcaaatgtgt	ccatttcaca	56775458
taggggttag	cctttatcat	gttattctaa	cataagaaaa	actacgcaa	56775508
aaatatagtg	tagccagggtg	cagtggctca	cacctgtaat	cccagcactt	56775558
tgagaggcta	aggcgggcgg	atcacctgag	gccaggagtt	tgagacctgc	56775608
ctggggcaaca	tggtgaaacc	tcatatctac	aaaaaatata	aaaaaattag	56775658
ctggaagtgg	tgatgggtgc	ctgtcattgc	agctacttag	gaggctgagg	56775708
catgagaatc	agtcaaacc	aggaggtgga	ggttgacagt	agccaagatc	56775758
acgtcactgt	attccagcct	gggtaacaga	acaagactgc	atctcaaaac	56775808
atatatgaag	tttttaccac	ctagtttcct	gagatctaaa	aatagggatt	56775858
gtctaaaaaa	taaaaaataa	aaaatgattt	tatgaatgat	gaatgaaaaa	56775908
gagttaaaag	gaaccttttt	ggctggatgt	gaagggtgact	cttgccctgta	56775958
atcccagcac	tttgggaggc	agagacagga	ggatcacctg	tggtcaggag	56776008
tttgagacca	gcctggccaa	aaagggttga	gaaccttttt	ggctggatgc	56776058
gaaggtgact	cttgccctgta	atcccagcac	tttgggaggc	agagataggc	56776108
agatcaccta	tggtcaggag	tttgagacca	aactggagtt	tgagatcagc	56776158

cccgctctcta	ctgaaaatac	aaaaattatc	cgggtgtggt	agcgggccc	56776208
tgtaatccca	gctacttggg	aggctgaggg	aggagaatca	cttgaaccca	56776258
ggaggcagga	ggttgcaagt	agccgagatc	gcgccactgc	actccagcct	56776308
gggcatgatg	gagcaagact	ccatctcaaa	aaagaaaaag	gttctagcca	56776358
ggctggctct	aaactcccaa	cctcaggtaa	tctgcctgcc	ttggcctccc	56776408
aaagtgctgg	gattacaggc	atgagccact	gtggcccgcg	tttgaatcct	56776458
ttcttttttt	gacagagtct	cactctgtca	cccaggctgg	agtgaatgg	56776508
tgcgatctca	gctcactgca	acctccgcct	cccagggttca	agcaattctc	56776558
ttgcctcagc	ctcccaagta	gctgggatta	cagggtgcctg	ccaccacgcc	56776608
cagctaatttt	tttgtatttt	tagtagagat	ggggttttgc	catgttggcc	56776658
aggctggctc	cgaactcctg	acctcagggtg	atccaccacac	ctcagcctcc	56776708
caaagtgtca	ggattacagg	catgagccac	cacactggct	gcttttgaat	56776758
tgttagctta	gataatttat	aaccagtgtt	ggggttggct	aagcaatgaa	56776808
cattggccta	gcttaatgag	aaatacataa	aagtaaaatg	aagagttttc	56776858
tcagagctcc	caatctcatt	ttcaagaaat	gacaaaatca	gccaggagcg	56776908
gtggctcaca	cctgtgatcc	cagcactttc	ggaggccgag	gtgggaggat	56776958
tgcctaaagca	cagaggttca	agaccagctg	agcaacgtgg	caaaactctg	56777008
tctctaaaaa	attattcagg	gattgtggtg	tgcacctgtc	atcctggcta	56777058
ctcaggaggc	taaggcagga	tgatcacctg	agcccagaag	ttcaaggctg	56777108
cagtgtgcca	tgactgtacc	actgcactcc	agcctgggtg	acagggtgag	56777158
accctgtctc	gtaaaaccaa	aaaggaatga	caagattggg	tgggtgcagt	56777208
gggacatacc	tgtaatccca	gcacttttga	agctgaggca	ggaggagttt	56777258
gagagcagcc	tggcaacata	gcaagaccct	acaaaaaatt	taaaaagtag	56777308
ccagggtgag	tggggcacac	ttgtagtctc	agctactcga	gaggctgagg	56777358
cacgagaatc	acttgaaccc	aggaggcaga	ggttgcagtg	agccgagaat	56777408
gtgccactgc	actnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777458
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777508
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777558
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777608
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777658
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56777708
nnnnnnnnnn	nnnnnnnnnn	nnnnnnntgcc	tcctagggtc	acgccattct	56777758
cctgcctcag	cctcccaggt	agctgggact	acagacgccc	gccaccacgc	56777808
ccggctaatt	ttttgtattt	ttagtagaga	cggggtttca	ccatgttagt	56777858
caggatggtc	tcaatctcct	gacctcgtga	tccaccgcgc	ttggcctccc	56777908
aaagtgtctg	gattacatgc	gtgagccacc	gcacctggcc	aagactgtct	56777958
taaaaaaaaaa	aaaaagtga	aggaaaagac	aagtaacaga	attttgtact	56778008
ttcaaatgag	tatatgcagt	attgtgtcta	tatcctgcca	gaaaaaagga	56778058
aatactattt	tgaaaaatca	tgcacaaaag	gagggtatca	gttagtacia	56778108
ctaatagaat	aataataact	ataaacaatt	tttattttat	tatttttttc	56778158
ttgagacagg	gtcttactct	gtcaccgcgg	ctagagggca	gtggcacaa	56778208
cacagctcac	tgaacctcgc	aactcccagg	ctcaagtgat	acttctgcct	56778258
cagcgtcctg	agtagctagg	actataggca	tgagccacta	ccccagata	56778308
attttttaatt	ttttttatag	agatgggggc	tcactatatt	gcccattgct	56778358
gtctcaaaact	ctggcctcaa	gccatcctcc	ctcctcagcc	tcccaacgtg	56778408
ctggggattag	aggaataagc	tactgcgcac	ggcctataaa	caatttttaac	56778458
tatgaaaaaaa	taagagtgtt	tcaatcaatg	acatgaatca	atcaatcaag	56778508
aaaaaaaaaaaa	ggtgtagatc	gggggacatg	gtcacacact	gtaatcccag	56778558
cactttggga	ggccgaggtg	ggcagatcac	ctgagctcag	gaattcagac	56778608
cagtctagtc	aacatagcga	aacccccatc	ctacaaaaaa	ctacaagaat	56778658
tagctgggta	tgggtggcata	tgctgttagt	cccagctact	tgggaggctg	56778708
agggtgggagg	atggccttag	actaggaggt	ggagggttgc	gtgagccgag	56778758
attgtaccac	tgcactccag	cctgggtgac	agagccagac	cctgtctcaa	56778808
aaaaaaaaaac	aaaaaaaacg	tacagcactg	ccaacagttc	tccctcatcc	56778858
cagaatagaa	aatccccac	cattttcttat	gcacacactg	cgactcttgg	56778908
cctcccaggc	agggtgggtgc	atgttcttga	ccagatccca	ggtccaaggg	56778958
gatttctctc	acaccacaga	cttctctttg	atcttccctt	ctttgaaatt	56779008
acagggcagg	gatgcgattt	ccatccaggt	agttttgttt	gtttgtttgt	56779058
ttgtttgttt	gttgcggttt	cgctctgttt	cccaggctgg	agtgcagtga	56779108
tgcgatctat	acctccacgt	cccagggttca	agcgattctc	agcgattctc	56779158
ctgcccagga	ctcccagata	gctgggacta	caagtgcagt	ccaccgcact	56779208
cggctaattt	tttttgtatt	tttagtagag	acagggtttc	accacgttag	56779258
ccaggatggg	ctcgatctcc	taacctcggt	atctgcccgc	ctcggcctcc	56779308
caaactgctg	gaattacagg	catcagccac	cgtgccccag	ctaatttttg	56779358

tatttttttagt	agatacaggg	tttcaccatg	ttggccaggc	tagtcttgaa	56779408
ctcctgcact	caggtgattt	gcccacctcg	gcctcccaaa	ttactgggat	56779458
tacaggcatg	agccaccatg	cctggcctgc	catatagttt	atattattatt	56779508
tatttttattt	tttgagacag	agtctcactc	tctcaccag	gctggagtgc	56779558
agtggcgtaa	tcactgttca	ctgcaacctc	cgcctcccgg	tacaagcagt	56779608
tctgctgcct	cagcctcctg	agtagctggg	attacaggca	cctgccacca	56779658
cgccagcta	ttttttgtat	ttttagtaga	gacagggttt	caccatgttg	56779708
gccaggctgg	tctcgatctc	ttgacctcag	atgatcctcc	cgcctcagcc	56779758
tcccaaagta	ctgggaatac	aggtgtgagc	caccatgctc	agctcctgcc	56779808
atctagttta	gaactcaaac	atttacctac	tgatgtgatt	tttattttttg	56779858
tcggcatttg	tcttggttat	aagagggcag	ggaccacatc	ttactctcca	56779908
gaagcccccga	cattgtccag	ctgaatgggt	ccccagcagt	ggtgcatgta	56779958
ctgatgcagg	agagtgcagt	gagggatagg	aggaagccag	gaggcgaaga	56780008
tgggaggagg	taaggaggaa	ggtgctgtgt	actcagcaaa	ggcaccctgg	56780058
gggaggccag	gcttacctct	ctttgcagga	atagtttatg	tttgaaacac	56780108
ttggctgggt	gcggtggctc	acgcctgtaa	tcccagcact	ttgggagccc	56780158
gaggcgggcg	aatcacctga	ggtcaggagt	ttgagaccag	cctggccaac	56780208
atggtgaaac	cccattctcta	ctaaaaatac	aaaaattagc	caggtgtggt	56780258
ggcaggcacc	tgtaatccca	gctactcggg	aggctgaggc	aggagaatca	56780308
cttgaaccca	ggaggcggag	gttgcagtgg	gcagagatca	cgccattgca	56780358
ctgcagcctg	ggcaacagag	tgagactccg	tctcaaaaaa	aaaaaaaaaa	56780408
aaagtttgaa	acacttactg	ctctctattg	ctcctgcaac	ctttgttttc	56780458
tgatttttcta	aaaatcattt	ctccgccagg	tgctgtggct	catacctata	56780508
atcccagcac	cttgggaggc	cgaagtgggc	aaatcaccag	aggtcaggag	56780558
ttcaagacca	gcctggccca	acatagtga	accccatctc	tactaaaaac	56780608
acaaaaatta	gccaggcgtg	gtggcagggtg	cctgtaatcc	cagctacttg	56780658
ggaggctgag	gcaggagaac	cacttgaacc	cgaggaggcg	aggttgcaga	56780708
gagcagagat	cacaccactg	cactccagcc	tgggcaacag	agcctccaaa	56780758
aaaataaaata	aaaataaaaa	tcatttatct	ctcttttctt	acttcttagc	56780808
ATGCTTTCAT	TGGaTTTGGA	GGAAATGTGA	TCAGGCAACA	AGTCAAGGAT	56780858
AACGCCAAAT	GGTATATCAC	TGATTTTGTA	GAGCTGCTGG	GAGAACCGaA	56780908
AGAATAACAT	CCATTGTCAT	ACgGCTCCAA	ACAAC TTCAG	ATGAATTTTT	56780958
ACAAGTTACA	CAGATTGATA	CTGTTTGCTT	ACAATTGCCT	ATTACAAC TT	56781008
GaTATAGAAA	GTTGGTACAG	ATGATCTGCA	CTGTCAAGTA	AACTACAGTT	56781058
AGGAcTCCTC	AAAGATTGgt	ttGTTTGTTT	GTTT TTAAC T	GTAGTTCCAG	56781108
TATTATATGA	TgACTATTGA	TTTCCTaGAG	AGTTTGTAA	TCTGAATTCT	56781158
TTATGTATAT	TCCTAGCTAT	ATTCATACA	AAGTGTTTTA	AGAGTGGAGA	56781208
GTCAATTAAA	CACCTTTACT	CTTAGGAAta	tagattcggc	agccttcaat	56781258
gaatattggt	ttttttccct	ttggtatgtc	aataaaagtt	tatccatgtg	56781308
tcagaacaga	tttgtggaat	ttgcagtt			

Figure 17

Contig0

– this contains exon 1 and is assembled from sequences gnl|ti|945553014, gnl|ti|949228349, gi|41830605|gb|AADD01081234.1, and gnl|ti|954837275

CAGTGCAGCTGAGATCAGACTTCCATGTGTAACCTCCACTACCCTACCAGGATGCCTTTTCATAAAGGTAAGAAATG
TAAATTTGGCCTTAATATACAAAGTTGCCAGGGCAGCACTGGGTCACTTCTACATATAGTACTTCTACGTTTCATCAG
CGGAAACTTTTAAGGGAAGGTGAAAATGCTTCTAGAAGGCGACTGGACACCAGCGCCTTTGGGCTCTTCTTCTAAGGC
CAATAGTGACCTAAATTATTGACTGACTGCTCCAATCAAGTGGGCAAAAGGGTACCAAGGCCGCAACATCAGACAA
ATTCACCTTGAGGGCCTACCTATGTGCTTTGAAAGACAAAACCTGCTGTTGTGAAGGACACTGTATTTTCAGAAAAACAT
AATCATATTAACAACCTAGTAACAATGTAAAATGCTGATGTGTTGAATGCTACTTTAGAAAAACATGTAAAATCTAC
AAAAAAAATTTATGATACAAAACCTACGTTATCAATCATCTAGCTAGCTAACTATCTACAGACATGGTTTTTCATTTT
TGTTGCTCAGGATGGAAAGCAGTGGGATGATCATAGCTCACTGCAGCCTTGAGCTCCTGGCCTCAAGTGATCCTCCT
GCCTCCTAAGTAGCTGGGGCCACAGGTGGACACAGTGACACCTGGGTTTTTTCTTTTGTAGAGACAGGGTTTTCTACTA
CACTGCCCAGGCTGGTGTCAAACCTTTGGAGTCTCGCTGTGTCAACCAGGCTGGGGTGCAGTGGTGGGGATCTCGGCT
CACTGCAACCTCTGCCACCTGGGTTCAAGCGATTACTTCCCTGTATCAGCCTCCCGAGTAGCTGGGACTACAGGCATG
TGCCACCACACCCGGCTAATTTTTTGTATTTTTTGTAGAGAACTGGGTTTTACCATGTTGGCCAGGCTGGTCTCAAAC
TCCTGACCTCAGGTGATCCACCTGCCTCGGCTCCCAAAGTGCTGGGATTACAGTCATGAGCCGATGAACACTTTCT
TATGCTATTAAATGGCCTAACCCAGGCGGGTCGTGGTGGCTCACGCCTGTAATCCCGAAAACCTCGGATGGCCAAGGT
AAGAAGATCACCTTGAACCTAGGAATTCCAAACCTGGCCTGGGCAACATAGCGAGACTCCCATCCCTACAAAAGATACA
AAAAATTAGGCTGGCGCACACCACGCTCGGCTAGTTTTTGTACCTTTTGTAGAGACAGGGTTGCGTCATGTTGCCTA
AGCCGGTCTCGAACTCCTGAACCCAAGCCATCCATCCTCCCGCCTCGGCCTCCCAAAGTGCTGGAGATTACAGGGGC
CCAGCCAGCCTCATGTTTTCTTTTAAGCAGTCCCTTCCCTGTTGCACACTTGGGTTAGTTTTCTTTTTTAATTTTTT
TAAACAGGGGGTTACCTCAATCTCGCAGCCTGGAGTGCTGTGGTGGGATCACAGCTCATTTGGAGCCTTGAACCTTG
GGTTCAAGTAGCTGGGGGGCTGAGGTAGGACTACAGAGATGGGGTCCCGCATGTTGCCAGGCTGCTCTTGGCCTG
AAGGGCATCTCCCGCTGGCCGTGCCCGGACATAGTTTTCTATTTTTGACCGACATAAACACATGTGCTGAGTCGGTGT
TTGTCAACACACAGGACCTGGCGGGGAGGTTCGCGGTTACCAGGCTCCACTCTAAGTAGAAGACTGCCAGCTCCAAG
CACTGTACCTCCCGGTGACGTGCGCGAACGCCCGCCCTGTGACGATACCTAAGGCCACCTTCATGACGCCGCCGAA
GGCCCGCCCTGTGACGCCCGCGAGGCCCGCCCTCACGCGGAGCCAATCGGAACTCGAGGCGGGGCTGTTGGGTC
TTCGGGAGCGCGCATGCGCGGGGGGCCACAGGCTCCCTGGCTGGCGTCAGCTAAAGTGGCTGTTGGGTGTCCGCAGG
CTTCTGCCTGGCCGCCGCCGCTATAAGCTACCAGGAGGAGCTTTACGACTTCCCGTCTGCGGGAAGTGGCGGGCA
CGATCGCAAGGTAGCGCAGAAGCTTCTCAATGGCCAGCGCCAGCTGCAGCCCCGGCGGCGCACTCGCCTCACCTGAG
CCTGGgtACGTGCAGCCCCACAACACCTTCCCCAGCCAGGGCCCGGGACCCCGGAGCGTCCCCCGCCACCTGGCG
CCGCTCATACCTGGGCAAGGGTGGGACCCCACTGAGGCCCGCCACGCATTAGGGAGCTTGCACCTCCCGAGTTTTGA
CCTCTGACGGGAGTTGTAATAGCATTAAAGTTTTTGAATTTTGTAGCGGGGGTAGAAGGGGCTTGGAAAGGGAAG
AAAACATCTTTTAAATATAACGTTCCGGCCGGGCCCGGGGTTAACCCTTTAATCCACATTTGGGAGGCCAGGCAG
GGGATTACGAGGTAGGAGTTAAGACCACCTGGCCGCATGGGGAACCCTGTTTTACTAAA

NO MORE MATCHES WITH THE ASSEMBLED HUMAN GENOME AFTER THIS 3' END

>gnl|ti|226793227 name:55249978 mate:226799032

- this Trace sequence has exon 2 (red underlined), and the underlined 3' sequence overlaps with gi|148184701|gb|ABBA01007462.1| Homo sapiens CTG_1103276812568, whole genome shotgun sequence (see below).

NO MORE MATCHES BEFORE THIS 5' END

GAGATATAGATGGGGAGATTTTTAATTTAGTTTTTTATTTAAAATTGTGTTTATAAGGAAAGAGATTATTATGTNTT
TTGATAGAGAATTTGGGAGAGTTTGTTTAATTTAGAGAGAGATGGGTTAGNAAGATTTATAGGGTTGGANTGTGATT
TANGAAAAGATTTAAGGTGTTTGGAGAATAATTTTTTGGGGAGGAATTTTTTGAATAAAAAGAAAANTTTGGTAAAAA

NGGTGGATTTGGTTTTANATGAAATAAAAAAATTAACCGGATGTGGTTGCACACGCCCTGTATNCCCAGCTACTCAG
 GAGGCTGAGGCACGAGAATCACTAGAACCAGGAGGTGGAGGTTGCAGTAAGCCAAGTTCGTGCCACTACCCTCCAG
 CCTGGGCAACAGAGTAAGACTCCATCTAAAAAATAAAGAAAGAAATAGTGTAGTGTGGGAAGTGAAAAA
 AAAAAAGAAAAGGAAAAGAAAAATGATTGAATTCATGAATACACTCTTATGTGGCCTGCACCGACTTTGACACAAA
 TTAGATTGGCTTAGTAGGCAAGGGTGGGATCTTTTCATAATTTTATTTGATGTCTAAATACATTTATCTTTTTTTC
TTATAGGAGGAAATTCCTTCCAAGGATGATCTCCCACTCAGAGCTGAGGAAGCTTTTCTACTCAGCAGATGCTGTGT
GTTTTGATGTTGACAGCACGGTCATCAGTGAAGAAGGAATCGGTGAGCTAGCCAAAATCTGTGGTGTAGAGGACGCG
 GTGTCAGAAATGTAGGGATAGCATTATTTACTTTATGAAATGATAAAGAATTTTTTTTTTTTTTTTTTTTGTAGACA
 GAGTCTCACTCTATTGCCCAGACTGGAGTGCAGTGGCACAATCTTGGCTTACTGCAACCTCCACCTCCCAGGTTCAA
 GCGATTATCCTGCCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCGTGTGCAACCACATCCGGCTAATTTTTTTTATTT
 TTAGTAGAGATGGAGTTTACCATTGTTGGTCAGGCTGGTCTCAAACCTCTGACCTCGTGATCCACCCACCTCAGCCT
CCCAAAGTGCTGGGATTACAGGCGTGAGCCACCATGCCCGGCCATGAAATGATAAAGAATTTCTAAAGGGTGGCTGT
TTTGGATGAAGTGCTGGACCTGGCTATAAAAGATGAGCACTAGGCTTTTTCTCTCACCCCTTAGAGTTGAATTTGA
TATATTGAGAACTNTGCTATCGCTCCNN

>gi|148184701|gb|ABBA01007462.1| Homo sapiens CTG_1103276812568, whole genome shotgun sequence

- contains exon 3 (nucleotides 504-625) and 3' sequence. The underlined seq overlaps with the 3' end of gnl|ti|226793227 (above).

CACCCACCTCAGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCATGCCCGGCCATGAAATGATAAAGAATTT
CTAAAGAGTGGCTGTTTTGGATGAAGTGCTGGACCTGGCTATGGAACATGAGCACTAGGCTTTTTCTCTCACCCCT
TAGAGTTGAATTTGATATATTATTTCATTTATATAAATATTTATGGCAAAAAAATATGGATAAGACATGTTCTTAGC
 TTGATTAGGGGAATGATGACCATCACTAACCAAGAAGATGGCCTAGGGGGCAGGGCTCAAGTGTCCAGTGGCTTTG
 TGTATAGTTGAAGTGTCGCTTAGTGGTTCCCTGGAAAGGTTTCAGAGTCTCTTGGTGGATCCTGGGTGAGAGCCC
 CTCTCTCCCTCCCTCCCTCCCTGCCCTCCCACTCCTGCCCTGCAGCTGGGCACCACCCTCTGCAGCCCCAGTCCCC
 AGTCATGCACCATGTCATTTCTTTTTTTTTTTTTTTTTTTCagGACGGAGTCTCGCTCTGTCAACCAGGCTGGAGTGCAA
TGGTGCAATCTCGGCTCACTGCAACCTCCGCTCCTGGGTTCAAGCAATTCTCCTGCCTCCACCTCCTGAGTAGCTG
AAACTACAGgcACGCACCACCACACCCGGCTTATTTTTGTATTTTTTAGTAGAGATGGTGTTCACCATGTTGGCCAG
 CCTGGTCTTGAATCCTGACCTCGTGATCTGCCGGCTCGGCCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACC
 ACACCTCGGCCACCATTTTTATTTTCAACTCCCTTTTCATGGAAGAAGCTTTAGCCTTTGGTCTCTTCTTGATTTATGA
 CAGCTCGCGGCTTTCAAGAAAACCTACCTATGAATAGGCTGTGTAACCTTTATTTATTTATTTATTTTGTAGATGGA
 GTCTCTGTCTTCTAGGCTGGGGTGCAGTGGCATGATCTCGGCTCACTGCAACCTCCCCCTCCAGGTTCCAGCAATT
 CTCCTGCCTCAGCTTCCCAAGTAGCTGGGATTACAGGCATGCACCACCACACCTGGCTAATTTTTTGCATTTTTTTTTT
 TATTAGAGATGGGGTTTTGTCTATGTTGGCCAGGCTGGTCTCGAACTCCTGGCTTCAGATGAGCTGCCACCTCAGCT
 TCCCAAAGTTTTGGCATTATAGGCATGAGTCACTGCACCCAGCTTGTTAATGTTATTTTTCAAGCACACCTTCAAAG
 TTTATTCCAAGGCTTGCTCTCATAGCAGCAAAGCCTGTCTGTTGCATAGTGGGGCTCTTGTAGCATCTTCCTTTT
 GGTGGATGTTGTGTAGACCGGGGAGCAGTAAGGCATGTTAACCTCGAGGACATTGGACCTGGCTGGCCTGACTGTTGG
 GTCTCCCTCCTAGGACATGGCGAGCCATGGGTGGGGCAGTGTCTTTCAAAGCTGCTCTCACGGAGCACTTAGCCCCA
 ATCCAGCCCTCCAGGGAGCAGGTGCAGAGACTCATAGCAGAGCACCACCTGACCCCAAGTAAGAG
 GAGCCGCTGCTCCAGGTGATTTTCAAGCAAGTGTGGGGAGCATGTCCTCCTGAGAGCATCTAACGATTGCTTTAG
 AGAGCCCTCTGGGTGGTTTATTTTATTTTCAATTTTATTTTAAATTTATTTTAAATTTATTTTATTTTATTTTATTT
 TATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT
 TATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT
 TCGGCTCACTGCAACTTCTGCCTCCCGGGTTCAAGTGATTCCTGCCTCAGCCTCCTGAGTAGTTGGGATTTTCAAG
 CACCCACCATGCCAGGCTAGTTTTTTGTATTTTTTAGTAGAGACAGGATTTTACCATTGTTGTTTAGGCTGGTCTCGAAC
 TCCTGACCTCAGGTGGTCTGCCCCGCTCAGACTCCCAAAGTGCTGGGATCACAGGCGTGAGCCACTGCGCCTGGCAA
 AATTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT
 TCACTGCAACTTTTGCTCCAGGTTACACGATTCTTCTGCCTCAGCCTCAGGAGTAGCTGGGATTACAGGCGCCT
 GTCACCACACCTGGCTAATTTTTGTATTTTTTAGTAGAGATGGGTTTCGCCATGTTGGCCAGGCTGGTCTCGAACTCT
 TCATCTCAAGTGATCCACCTGCCCTCGGCCTCCCAACGTGCTGAGATTACAGGTGTGAGTCAACACACCTGGCTTATT
 ATTTTTTTTTTTAGATTGAGAGTCTATTCTGTTGCCAGGTTGGAGTACAGTGGTATAATCATGGCCCACTGCAGC
 CTCGAACTCCTGGGGTCAAGTAGCCCTCCACCTTAGCCTCCTTAGTAGTAGGGACTGCAGGTGAATGCCACCGTGC

CCAGCTAATTAAAGAATTTTTTTTTTTTTCTGAAAGTCTTGCTTTGTACCCAGGCTGGAGTGCAGTGGCATGAACA
CAGCTCACTGCAGCATCAATCTTCTGGGCTCAAGGGATCCCTCTGCCTCTGCTTCCTAAGTAGCTGGGACTACAGAC
ATAGGCCACCATGCCAATTTTTTTTTTTTTTTTTTGTAGATGGAGTCTCGCTCTGTTGCCAGGCTGGAGTGCAGTGG
CGCCATCTTGGCTCACTGCAAGCTCCACCTCCTGGGTTACAGCCATTCTCTTCCCTCAGCCTCCCGAGTAGCTGGGA
CTACAGGCGCCTGCCACCACACCTGACTAATTTTTTTGTATTTTTTAATAGAGACGGGGTTTACCATTGTTAGCCAGGA
TGGTCTCAATCTCCTGACCTTGTGATCCACCTGCCTCAGCCTCTCAAAGTGTGGGATTACAGGCGTGAGTCACCGC
GCCTGGCCCCACCATGCCAATTTTTTAAAAAAAATTTTTTTATAGACATAGCATCTCACTACATTGCCCAGGCTGGT
CTTGAACCTTGTGGGCTTAAGTGACCAATTTTCACAACTCAGCCTCCCAAAGTGTGGAATTATAGGTTTAAAGACACT
ATACCCAGCCTTTAAGAATCTTCTTGTAGAGATGGAATCTTGCTGTGTGTTGCCAGGCTGTTCTCAAGCTCCTTGCCCT
CAAGTGATTTTCTACCTTGGCCTCCCAAAGTGTGGAATTAAAGGCATGGACTGTGTGGTCTTCTTGGCACTTACAC
GTGGTCTTGTCTGGCCGGTGTCTGGTCTGTCTGTTTCTGCCCTTTCCTCTTCTTCCAGGGAAAACTAAGCTTTCC
TTTTTGTCTCATCTTGTGTTTTTCTGGGTCCATGGGCAGAGTAGAGTTCTAGAATGGTTTCTTAAAGCAGCCAAGC
CCTACCTTTGATTTCTAAGTACATTTAGAAAAACACATATAAGGCTGGGCATGGTGGCTCACGCCTGTAATCCCAG
CACTTTGGGAGGCCGAAGTGGATGGATCACCAGGTCAGGAGTTTCGAGACCAGGCTAACCAACATGATGAAACCTGT
CTCTACTAAAAATAAAAAATAAGCCGGGCACGGTAGCTCAGCTCTGTAATCCAGCACTTTGGGAGGCCAAGATGG
GTGTATCACCTGAGGTGAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAACCCAGTCTCTACTAAAAATACAAAA
AAAGTAGCTGGGCGTGGTGATAGGCGCTGTAATCCAGCTACCCAGGAGGCTGAGGGAGGAAAAATCACTGGAACCC
GGGAGGCAGAGGTTGCTGTGAGCCGAGATCATGCCAGTGCACCTCCAGCTAGGGCAACAGAGCAAGATTCCATCTCAA
AAAAAAAAAAAAAAAAAATTAGCCGGGCATTGTGGCATGGGCCCTATAATACTCAGGAGGCTGAGGCAGGAGAAGCGC
TTGAACCCGGGAGGTAGAGGTTGCAGTGAGCCAAGATTGCAGCCTTGCACCTCCAGCCTGGGTGACAAGAGTGAACT
CTGTCTCAAAAAATAAAGAGAAAAACATATAGAAAACATTAACACCCAGGAGTATACCTTGTCAAACATACCTCA
GGCAAATGCATTCAGGAGAAGAAAATACATCTTATTTCCCTCTTTCATGTTTCGTTTTTTTTTTTTTGTTCCTTTGT
ATTACTCAGTGTTGGGTATTTTTGTATTTTATTTTGCAGGGAGCTGGTAAGTCGCCTACAGGAGCGAAATGTTTCAGGT
TTTCCTAATATCTGGTGGCTTTAGGAGTATTATAGAGCATGTTGCTTCAAAGCTCAATATCCCAGCAACCAATGTAT
TTGCCAGTAGGCTGAAATTTCTACCTTAATGGTAAGATGTTAACGGTAACATGTTCCCTTTCTTAGCAGTTCCATTAT
TCAGTATTCTGGGTAATGTCTTTTGAATGCAACTTGAACAGTCACATCAGAGTTAAATATTGAGATGAATGGTTCT
CTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGACGGAGTCTCGCTCTGTGCGCCAGGCTGGGGTGCAGTGGCGCGATCTCGG
CTCACTGCAAGCTCCCCCTCCCGGGTTCACGCCATTCTCCTGCCCTCAGCCTCCCCAGTAGCTGGGACTACAGGCGCC
CGCTACCATGTCAGGATAATTTTTTGTATTTTTAGTAGAGACGGGGTTTCTCCGTGTTAGCCAGGATGGTGTGATC
TCCTGACCTCGTGATCCGCCCCGCTCGGCCTCACAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGCCGAG
ATGAATGGTTCTCTTAATTGATGTCTTTTGCCTTTTGGTACCTTTGCTCAGCAAAACAGACTTAGATCATCACCTG
TCTTAGCTTTATTATCTATAACATCACCTGTCTGTACATAAATGTCTGCATTTCG

>gi|68978189|gb|AADB02010625.1| Homo sapiens chromosome 7
CRA_219000002701389, whole genome shotgun sequence

- this contains partial exon 3, exon 4, and 3' sequence

CAATGGTGCATCTCGGCTCACTGCAACCTCCGCCTCCTGGGTTCAAGCAATTCTCCTGCCTCCACCTCCTGAGTAG
CTGAAACTACAG*gc*ACGCACCACCACACCCGGCTTATTTTTTGTATTTTTTAGTAGAGATGGTGTTCACCATGTTGGC
CAGCCTGGTCTTGAACCTCCTGACCTCGTGATCTGCCGGCCTCGGCCTCCCAAAGTGTGGGATTACAGGCATGAGCC
ACCACACTCGGCCACCATTTTATTTTCAACTCCCTTTCATGGAAGAACGTTTAGCCTTTGGTCTCTTCTTGTATTTA
TGACAGCTCGCGGCTTTCAAGAAAACATCTATGAATAGGCTGTGTAACCTTTTATTTATTTATTTATTTTGTAGAT
GGAGTCTCTGTCTTCTAGGCTGGGGTGCAGTGGCATGATCTCGGCTCACTGCAACCTCCCCCTCCAGGTTCCAGCA
ATTCTCCTGCCTCAGCTTCCCAAGTAGCTGGGATTACAGGCATGCACCACCACACCTGGCTAATTTTTTGCATTTTTT
TTTTATTAGAGATGGGGTTTTGTCTGTTGGCCAGGCTGGTCTCGAACCTCTGGCTTCAGATGAGCTGCCACCTCA
GCTTCCCAAAGTTTTTGGCATTATAGGCATGAGTCACTGCACCCAGCTTGTTAATGTTATTTTCAAGCACACCTTCAA
AAGTTTATTCCAAGGCCCTGTCTCATAGCAGCAAAGCCTGTCTGTTGCATAGTGGGGCTCTTGTAGCATCTTCCT
TTTGGTGGATGTTGTGTAGACCCGGGGCAGTAAGGCATGTTAACCTCGAGGACATTGGACCTGGCTGGCCTGACTGT
TGGGTCTCCCTCCTAGGACATGGCGAGCCATGGGTGGGGCAGTGTCTTCAAAGCTGCTCTCACGGAGCACTTAGCC
CCAATCCAGCCCTCCAGGGAGCAGGTGCAGAGACTCATAGCAGAGCACCCCCACACCTGACCCCCAGCATAAGTAA
GAGGAGCCGCTGCTCCAGGTGTATTTTCAGCACCAGTGTGGGGAGCATGTCTCTGAGAGCATCTAACGATTGCTT
TAGAGAGCCCTCTGGGTGGTTTATTTATTTTCAATTTTTTAAATTTTATTATTTTAAAAATATTTTATTTATTT
ATTTATTTATTTATTTATTTATTTATTTATTTATTTAGGAGCAGAGTCTCGCTCTGTGCGCCAGACTGGAGTGCATGGTACG

ATCTCGGCTCACTGCAACTTCTGCCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCTCCTGAGTAGTTGGGATTTTC
AGGCACCCACCATGCCAGGCTAGTTTTTGTATTTTTTAGTAGAGACAGGATTTTACCATGTTGTTTAGGCTGGTCTCG
AACTCCTGACCTCAGGTGGTCTGCCCGCCTCAGACTCCCAAAGTGCTGGGATCACAGGCGTGAGCCACTGCGCCTGG
CAAAATTTTATTTATTTATTTTTTTTTTGAGACAAAGTCTCACTCTGTGCGCCAGGCTGGAGTGACAGTGGTGAATCTC
AACTCACTGCAACTTTTGCCTCCCAGGTTTACACGATTCTCCTGCCCTCAGCCTCAGGAGTAGCTGGGATTACAGGCG
CCTGTCACCACACCTGGCTAATTTTTTGTATTTTTTAGTAGAGATGGGTTTTCGCCATGTTGGCCAGGCTGGTCTCGAAC
TCTTCATCTCAAGTGATCCACCTGCCTCGGCCTCCCAACGTGCTGAGATTACAGGTGTGAGTCACCGCACCTGGCTT
ATTATTTTTTTTTTTTAGATTTCAGAGTCTCATTTCTGTTGCCAGGTTGGAGTACAGTGGTATAATCATGGCCCACTGC
AGCCTCGAACTCCTGGGGTCAAGTAGCCCTCCCACTTAGCCTCCTTAGTAGTAGGGACTGCAGGTGAATGCCACCG
TGCCCAGCTAATTAAGAATTTTTTTTTTTTTTCTGAAAGTCTTGCTTTTGTACCCAGGCTGGAGTGACGTGGCATGA
ACACAGCTCACTGCAGCATCAATCTTCTGGGCTCAAGGGATCCTCCTGCCTCTGCTTCTTAAGTAGCTGGGACTACA
GACATAGGCCACCATGCCAATTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTTGCCAGGCTGGAGTGACGT
GGCGCCATCTTGGCTCACTGCAAGCTCCACCTCCTGGGTTACAGCCATTCTCTTCCCTCAGCCTCCCGAGTAGCTGG
GACTACAGGCGCCTGCCACCACACCTGACTAATTTTTTGTATTTTTTAATAGAGACGGGGTTTACCATTGTTAGCCAG
GATGGTCTCAATCTCCTGACCTTGTGATCCACCTGCCCTCAGCCTCTCAAAGTGCTGGGATTACAGGCATGAGTCACC
GCGCCTGGCCCACTAATTTTTTAAAAAATAATTTTTTTATAGACATAGCATCTCAGCTTGCCTCAGGCT
GGTCTTGAACCTTGTGGGCTTAAGTGACCAATTTTCAAACTCAGCCTCCCAAAGTGTTGGAATTATAGGTTTGAGAC
ACTATACCCAGCCTTTAAGAATCTTCTTGTAGAGATGGAATCTTGCTGTGTTGCCAGGCTGTTCTCAAGCTCCTTG
CCTCAAGTGATTTTTCTACCTTGGCCTCCCAAAGTGCTGGAATTAAGGCATGGACTGTGTGGTTCTTCTGGCACTTA
CACGTGGTCTTGTCTGGCCGGTGTGTCTGGTCTGTCTGTTTCTGCCTTCTCTCTTCTCCAGGGAAAACCTAAGCTT
TCCTTTTTTGTCTCATCTTGTGTTTTTCTGGGTCCATGGGCAGAGTAGAGTTCTAGAATGGTTTCTTAAAGCAGCCA
AGCCCTACCCATTGATTTCTAAGTACATTTAGAAAAACACATATAAGGCTGGGCATGGTGGCTCACGCCTGTAATCC
CAGCACTTTGGGAGGCCGAAGTGGATGGATCACCAGGTCAGGAGTTCGAGACCAGGCTAACCAACATGATGAAACCC
TGTCTCTACTAAAAATACAAAAATTAAGCCGGGCACGGTAGCTCACGTCTGTAATCCAGCACTTTGGGAGGCCAAG
ATGGGTGTATCACCCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCAGTCTCTACTAAAAATAC
AAAAAAGTAGCTGGGCGTGGTGATAGGCGCTGTAATCCAGCTACCCAGGAGGCTGAGGGAGGAAAATCACTGGA
ACCCGGGAGGCAGAGGTGCTGTGAGCCGAGATCATGCCAGTGCACCTCCAGCTAGGGCAACAGAGCAAGATTTCCATC
TCAAAAAAAAAAAAAAAAAAAAAAATTAGCCGGGCATTGTGGCATGGGCCTATAATACTCAGGAGGCTGAGGCAGGAGA
AGCGCTTGAACCCGGGAGGTAGAGGTTGCAGTGAGCCAAGATTGCAGCCTTGCACCTCCAGCCTGGGTGACAAGAGTG
AAACTCTGTCTCAAAAAATAAGAGAAAAACATATAGAAAACATTAACACCCAGGCAGTATACCTTGTCAAACATA
CCTCAGGCAAATGCATTACAGGAGAAGAAAATACATCTTATTTCCCTCTTCATGTTTCGTTTTTTTTTTTTTGTCTTCT
TTTGTATTACTCAGTGTGGGTATTTTGTATTTTATTTTGCAGGGAGCTGGTAAGTCGCCTACAGGAGCGAAATGTT
CAGGTTTTTCCATAATCTGGTGGCTTTAGGAGTATTATAGAGCATGTTGCTTCAAAGCTCAATATCCAGCAACCAA
TGATTTTGCCAGTAGGCTGAAATTTACCTTAATGGTAAGATGTTAACGGTAACATGTTCCCTTTCTTAGCAGTTCC
ATTATTCAGTATTTCTGGGTAAATGTCTTTTGGAAATGCAACTTGAACAGTCACATCAGAGTTAAATATTGAGATGAATG
GTTCTCTTTTTTTTTTTTTTTTTTTTTTTTGGAGACGGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGACGTGGCGCGAT
CTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTACAGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAG
GCGCCCGCTACCATGCCCGGCTAATTTTTTGTATTTTTTAGTAGAGACGGGGTTTACCCTGTTAGCCAGGATGGTCT
CGATCTCCTGACCTCGTGATCCGCCCGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCGG
CCGAGATGAATGGTTCTCTTAATTGATGTCTTTTGGCCTTTGGTACCCTTTGCTCAGCAAAACAGACTTAGATCATC
ACCTGTCTTAGCTTTATTATCTATAACATCACCTGTCTGTACATAAATGTCTGCATTTCTGTAACATTAATTCTTA
CATAGCCCCCTATGACTTACTTTGCTATGTGTTGTATTTCAATTTTTATGGGAGCTTCTATGCATTTGAAGCAATTTAG
ACTCAAGAAGGAGACTTTTGGCTGGGCACGGTGGCTCAGCCTGTAATCCAGCACCTTTGGGAGCCTAAGGCAGGAG
GATCACTTGAGGTCAGGAGTTCGAGACCACCTGGCCAAAGTGGTGAAGCCCCAACTCTGCTAAAAAATAAATAA
AAAAAATAGGAAAATTAGCTGGGCATGGTGGCACATGCCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA
TCACTTGAAGCTGGGAGGCGGAGGTTGCAGTGCGCCGAGATCGCGCCACTGCACCTTAACAGCCTGGGTGACAGAGCG
ATACTCTGTCTTTAAAAAATAAGACAGAGACTTTAGAAAGGAGACTTTTGTTAACAGATTACTGAACCTACAGCTGC
CCTTGGGGAAGCCTCTATCCCCCTCTGTGGTCCATAGCGGTGCCACAGCTGTTATATTATCTCAGATCTCCAGTTAT
ATCAAGGAGTGACTGAGAGTGATATTCTTTTTTTTTTTTTTTTTTTTTTTTGGAGACAGGGTCTCACTCTGTGCTCCTGC
TGGAGTGACGTGGCTTGATCTCGGCTCACTGCAACCTCTGCCGCTGGGTCAAGTGATCTCCTGCCCTCAGCCTCC
CAAGTAGCTGCGATTACAGGTGCCCATCACCAAGCTGGCTAATTTTTTGTACAGTGCCCAACCGTACTTTTGTTTTA
GACAGCTGGTTTGTCTCTATAATCCAGCACTTTGGGAGGCTGAGGTAGGAAGATCGCTTGAGCACAGGAGTTTGAGA
CCAGCCTGGGCAATATAGTAAGACCCCATCTCTACAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
CTATATAAGATTTAAACAGGAGTTAAACAAAAACAACCTCTCATTAACAAGATAAATAAATAAATAAATAAATAAATAA
CAATCTTGTATGTCAACTGACTTTGTGCAAAACCAAGAAAAGTTATGAATAAGGAACATAAATAAATAAATAAATAA
TTAACAATGCCCTATCCCTGTAGCATGGGCACCACCTTTTAGCCTAGGTGTGTGCTCAATAAGAGTTGATTTTCAATTT
CCCCAATAATACCATCAACAACCTATAACCGAGATTTTAACATCTTTTTTATACTTTATATACTAGTTCTCATATG

[illegible]

TATACTCACTGCAACCTCCACCTCCCAGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAAG
TGCATGCCACTGCACCTCGGCTAATTTTTTTTGTATTTTTTAGTAGAGACAGGGTTTCACCACGTTAGCCAGGATGGTC
TCGATCTCCTAACCTCGTGATCTGCCCCTCGGCTCCCAAAGTGCTGGAATTACAGGCATCAGCCACCATGCCCA
GCCTAATTTTTTGTATTTTTTAGTAGATACAGGGTTTCACCATGTTGCCAGGCTAGTCTTGAACCTCTGCGCTCAGGT
GATTTGCCACCTCGGCTCCCAAAGTACTGGGATTACAGGCATGAGCCAACATGCCTGGCCTGCCATATAGTTTAT
TTCTTATTTATTTTATTTTTTGGAGACAGAGTCTCACTCTCTCACCAGGCTGGAGTGCAGTGGCATAATCACTGTTT
ACTGCAACCTCCGCTCCCGGTACAAGCAGTTCTGCTGCCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCACCTGCC
ACCACGCCCAGCTATTTTTTGTATTTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGATCTCTTG
ACCTCAGATGATCTCCCGCTCAGCCTCCCAAAGTACTGGGAATACAGGTGTGAGCCACCATGCTCAGCTCCTGCC
ATCTAGTTTGAAGCTCAAACATTTACCTACTGATGTGATTTTTTATTTTTTGTAGGCATTTGTCTTGGTTATAAGAGGG
CAGGGACCACATCTTACTCTCCAGAAGCCCTGACATTTGTCCAGCTGAATGGGTCCCCAGCAGTGGTGCATGTACTGA
TGCAGGAGAGTGCAGTGAGGGATAGGAGGAAGCCAGGAGGCGAAGACGGGAGGATGTAAGGAGGAAGGTGCTGTGTA
CTCAGCAAAGGCACCTTGGGGGAGGCCAGGCTTACCTCTCTTTGCAGGAATAGTTTATGTTTGAACACCTGGCTGG
GTGCGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGCCCGAGGTGGGCGAATCACCTGAGGTGAGGAGTTTGGAG
ACCAGCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATGGCCGGGCGCGGTGGCTCACGCCCTG
TAATCCCAGCACTTTGGGAGGCCGAGGCGGTGGATCAGGTGAGGATCGAGAGATCCTGGCTAACAGGTG
AAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCGCGGTGGCGGGCGCTGTAGTCCCAGCTACTGGGGAG
GCTGAGGCAGGAGAATGGCGTGAACCCGGGAAGCGGAGCTTGCAGTGAGCCGAGATTGCCCCACTGCAGTCCGCAGT
CCGGCTTGGGCGACAGAGCGAGACTCCGTCTCAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATA
CAAAAAATTAGCCAGGTGTGGTGGCAGGCACCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCATTTGAA
CCCAGGAGGCGGAGGTTGCAGTGGGCAGAGATCACGCCATTGCAGTGCAGCCTGGGCAACAGAGTGAGACTCCGTCT
CAAAAAAATAAAAAAAGTTTGAACACTTACTGCTCTCTATTGCTCCTGCAACCTTTGTTTTCTGATTTTCTA
AAAATCATTTCTCTCCGGGTGCTGTGGCTCATACCTATAATCCCAGCACCTTTGGGAGGCCGAAGTGGGCAAATCAC
CAGAGGTGAGGAGTTCAAGACCAGCCTGGCCCAATATAGTGAAACCCCATCTCTACTAAAAACAAAAAATTAGCCA
GGCGTGGTGGCAGGTGCCTGTAATCCCAGCTACTTTGGGAGGCTGAGGCAGGAGAACCCTTGAACCCGGGAGGCGGA
GGTTGCAGAGAGCAGAGATCACACCACTGCGCTCCAGCTGGGCAACAGAGCCTCCAAAAAATAAATAAAAAATAA
AATCATTTATCTCTCTTTTCTTACTTCTTAgGATGCTTTTCATTGGATTGGAGGAAATGTGATCAGGCAACAAGTCA
AGGATAACGCCAAATGGTATATCACTGATTTTGTAGAGCTGCTGGGAGAACCAGGAAGATAACATCCATTGTCTATAC
AGCTCCAAACAACCTCAGATGAATTTTTTACAAGTTACACAGATTGATACTGTTTGCTTACAATTGCCTATTACAAC
TGCTATAGAAAGTTGGTACAGATGATCTGCACTGTCAAGTAACTACAGTTAGGAATCCTCAAAGATTGGTTTGT
GTTTTTAACCTGTAGTTCCAGTATTATATGATCACTATCGATTTCTTGGAGAGTTTTGTAATCTGAATCTTTATGTA
TATTCCTAGCTATATTTTATACAAAGTGTTTTAAAGAGTGGAGAGTCAATTAACACCTTTACTCTTAGGAATATAGA
TTCGGCAGCCTTTCAGTGAATATTGGTTTTTTTCCCTTTTGGTATGTCAATAAAAGTTTATCCATGTGTGAGCATAA
ACAGATTTGTGGAATTTGCAGTTCTGCCCTTGGATATTGGAATATCCAAATTTTTCAGGTACTGTGTGAAGCATAA
ACATCC
CCCAGGCAAGAAACAGGCTTAAGACTTTGCTTAGTGCTGGGCAACAATGTCTCACCCCATGAGCCCATCCCACTTC
CTAAAGCCCTGGGACCTGGAAAGTCTTTTTTTTACCTCTTGGGATTTGTCTATATGAAATGTTACAGAGATTGGTT
TTTTTTTTTTCGCCCTCAATATAACAATAGAGTCTTAAACAAAAGTACTGATTGAAAAATGGATATTTCGGGCTCACGC
CTGTAATCTCAGCACTCTGGGAGGACAAGGTAAAGAGGATCAACTTGAGGCCTGGAGTTTGGAGACCAGCTGGGCAAC
ACAGCAAGACCTGTCTCTACAAAAAATAAAAAATAGCTGGATGTGGTGGCACCCACCTGTAGTCCAGCTACTT
GGTGGGCTGAGATGGGAGGATCCCTTGAACCTCAGGAGTTCAAGGCTGCAGTGAGCTGTGACTCCACCATTGTAATCC
AGCCTGGGTGACAGAGTGAGACCCTGTCTCTTAAAAAGAAAAAAGGATATTTGAAATCACATAAAGTAAAGAAAT
AAAAACAAGTTTAATCACAAAGGCCAGCATTTTCCCTTAAATGAATTAATATTAATAACTCCTAACAGCCTATATA
AACACATTTCTTTTTTATAAATGACGTGGAAATACCTTGAAGCCAGGTTTGTGTTTGTTTTGGAGCAGAGCCAAGCAA
GGGAAATTTCAGAACACGCTTTGAACGTGTGAGATTCTATACCTCCCTGGAGGTATGGGGGAGGGGAAAGCCACGG
GGTTAGGAAGTGGATACGTGTTGGATAATTCAGTAGAAAAAGCCATAAGGTTTTTAATATTGAGAAATTAGTTGAAAA
CTTACTTAACCAAATTTGGCTATATAATTAATAAATAATTAATTCAGCCACATAAGCAAAATCAAATTAATCAAATTT
CCTTCCAAATCAGTTTCTGAAGGAGAGACATGAAATAGGTTTGGCACCATTAAATGAGATTCCAAACAATGAGGCC
GTTGTTCCCAACTGTGAATGCCTCCTGGTCGCTCTCCTAGAAACCGTTTTTGTGGTGGTTTGCATTTTTAGTAGTGCC
AAGCTCTGGGCAGGACCAGGCATGCCAGGACCTTGAAGTGCTGCAACGGCATCAGCCTCTATGCTGCCCGGTGT
GCATGGCAGGAGCAGGCCATTGGTGGGCGGGTGGGAGGACCGATCCCCACTGGTTTTCTCATGCTGACCGTACAG
CTGCTACTGCCCCACTAGTTCTTGGCCGCGGCAGATAGAGCCCATTTCTCAAGATGGGAATTGCAATAGAGAGCT
TAATCTGTCGACAACCGGCCAAACAGGAGGCCGAGCTTTTATTTTACATCAATCTCCCCAAAAATTCGGAGGCTAG
ACTTCTTCAAGGATACTTTGGCAGGCCAGGGGATTCGCTTCTGGGTGGAGCCACAGAATTGGCCATGGGGTGGGGTCT
GGGTGGGGGCGGGGGCGGGGTAGTTAGGTCCGGGTGGAGCCACTGATGGTCAGAAACACAAAAACCTGAGCCAGGC
ACGGTGGCTCACGCCGTGAATCCCAGCATTTGGGAGGCCAAGATGGGCAGATCACTTGAAGTCAGGAGTTGGAGACC
CACCTGGCCAACATGATGAAACCCCATCTCTACTAAAAATACAAAATGTGGTGGGGCACGCCTGTATCCAGCTAC
TCGGGTGGCTGAGGCAGGAGAATTGCTTGAATCCAGGAGGCAGAGGTTGCAGTGAGCAGAGATTGTGCCACTGCACT

CAGAGTGTGGGATTACAGGTATGAAGAACTGTGCCAGCCTCATTTCATCTTCTTTAGCTAACCTAACTGTCTC
TGTGTTCTCCAGCATTCATTCCCTTGAATACTCAATCTCCTGTGTGGCTGCTGCAAGTCTGGGATCTTTTTAGGA
TTCCCTCTCTAAGGGCCATTTATTTATTTATTTATCTATTTATCTATCTATTTATTTATTTATTTATTTGAGACGGA
GTCTCGCTCTGTGCGCCAGGCTAGAGTGCAGTGGCGTGATCTCGGCTCACTGCAAGCTCTGCCCTCCGGGTTTCATGC
CATTCGCTCAGCCTCCCGAGTAGCTGGGAATACAGGCGCCACCACCACGCCCCGGCTAATTTTTTGTATTTTTAG
TAGAGACAGGGTTTCACTGTGTAGCCAGGATGGTCTCAATCTCCTGACCTCTCCACCTGCCCTCGGCCCCCAAAG
TGCTGGGATTACAAGCAGAAGCCACCGCGCCCCGGCCGGGCCATTTCATAAAGTGGTGCCTTAATTTGCCACAGTCCG
CAGTCCACATTCAGCCAACAGCAAAGCCAAACGCCAGCTGTGCCCTGGATCTCCCTACATACCACATTCACCCCTTAT
CTCGCATACATTTCTTCACTATGCTAGGACAAGAACACTTTTGACCTAATCAATGCACATGTTCTCAGTTCAAAGT
AAAACACATAATAAATAACCCCTCCATGATGCTCACCTCCTGGTGAGTCTCTGTGTTAAAAAGCAATTGTCTTCAA
TCTAACTCCTGTGAATGTTTATACTTCTGTTTATTTCTAAACCTCCCTAATTAGAGACTTCCAGTATCTTGCAATTC
CTCAACAAGTCATAGGGAGCCAGGAATAAATTATTAATAAATAAATTAGCAATAGGCTTTACCTTGTTTTATCATTT
AGTTGGTGATCTTTTTCCAATTTAATTAGTCTTGCTGATATAGAAAACATTTTATAGAAGAGGCTGTTGGGTAAAGAA
CAAGGCTCTTTCTAGCTGTTCTACAGAAGAAGGAGGAGGAGGAGGAGGAAGGAGAAGGAGAAAGAAAAAGAAAA
AGAAACAAGGCCCATTTTAGGAAACTTTTAGCCATGAGTAGAACACAGCCGAGGAAAACCTGTTTTACATAAGAAA
TATATTTTGGCTAGGCACGGTGGCTCCTGCCTGTAATCCTAGCACTTTGGGAGGACAAGGCAGGCAGATCACCTGAG
GTCAGGAGTTTGAGATCAGCCTGGCCAACATGGTGAAACCTGTCTCTACTAAAAATACAAAAATTAGCGGGCGTGG
TGGTGGGTGCCGTGAATCGCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGAACCCGGGAGGTGGAGGTGTGTG
CCACTGCACCTCCAGCCTGGGTGACAGAGCGAGACTCCATCTCAAAAAAGAAAAAGAAACAGATTTACAGAAAGCAC
ATTTAATGCAGGCAAATTTTGAGCATCAGAGAAATTTGGTCTGGATTAAGTCCAGAGATACAGCATTGCTTTGGTTAC
AAGAATGGTTTGAATCCAGCTCTACTGCTTTAATAACTGCCTGTGCCCTAGCAAGGTACTTAACCCCTCAGTATCT
TCTCAAAAAATGGGGATGGCAGTAGTACAGTGGTCCATGCCCTCAGAGTGTGTTGACGAGAATCAATTACATGGTGGC
GGGCACATAATTAGTGCTCAATAAATGCTGTAGCTATTATCTTCATTCTTGCTCTATGGGTGAAAGAAAAATGGCC
GGCGCAGTGGCTCATGCCGTGAATCCCAGCACTTTGGGAGGCTGAGGTGGGTGGATCATGAGGTGAGGAGATCAAG
ACCATCCTGGCTAACATGGTGAAACCTGTCTCTACTAAAAATACAAAAAAATTAGTTGGGCGTGGTGGCGGACGC
CTGCAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGGCAGGAGAATGGCGTGAACCCGGGAGGTGGAGCTTGCACCTG
AGCTGAGATCGCGCCACTGCACTACAGCCTGGGCGACAGAGCAAGACTCCGTCTCAAAAAAAAAAAAAAAAAAGAAAG
AAAAAGAAAAAGAAAAATGCAATACTGGATCAGATTAATTTCTTCCAGAAGCTGGCCGGGTGTGGTGGCTCACA
CTTCTAATCCCAGAACTTTGGGAGGCCAAGGGGGCGGATCACCTGAGTTCCAGGAGTTCAATACCAGCCTGGCCAGC
GTGGTGAAACCTGTCTCTACTATAAATACAAAAATTAACGGGCATGTTGGCGGGTACCTGTAATCTCAGCTACTC
AGGAGGCTGAGGCCAGTAGAATTGCTTGAACAGGAGGCGCATGTTGCAGTAAGCTGAGATCAGGCCACTGCATCC
AGCCTGGGCGACAGAGAGACCTATCTCAAGAAAAGCAAAACATATAAAAAATTTTTTAATACAGATTTTTTGGG
CTTTGTTTCATAGGTGATTCCACCAGCAGAGAGACCTAAGAATGTGTATTTTTATCAAGCCCTCTAGTGACCTTTA
TGATCAGGCAAACCTCAGGTAATAACATTGATTGTAGGGTTGGTGAGTGCTTATTTTCCAGGTTAAAAATGATACGGGA
CATAGAGCAGCAGTTCTCAAATTTTAATGTGCTTAAGAATCACCAGGAAAGTTGTTAAAAATTAGATATCCAGAGAT
TTAGACATCCAGAGATTTAGATTACAGAGCTCTGGGAGACACTCAAAAATTTGCATTTTAAACCAGTTTCTTCTCTCT
AGGGTAAGTCATTCAGAGGCACTTGTCTTCATCCATTCTGTGCTGCTATAACAGGTAATTTATAATGAACAGAAAT
TATTTGGTTCAGGGTCTGGAGGTTGATGAAGTCCAAGTAGAGGGCTACATCTGCTGAGGGCCTTCTTGCTGTGGC
AGACAATGGCAGAAAGTATCATATCATGAGAGAGAGCAGAAAGGGGACCGAGCTCATCCTTCATAAGGAACCCACTC
ACATCATAATGAACCCCACTCCACAATAACGGCATTCATCCATTCTGAAGGCAGAGTCTTCATGACCAATCACC
TCTTAAAGGTCTTACCTGTCAACATTATTGCATTGGGGATTAGGTTTCCAACACGTGAACCTTTGGGGGATACATTCA
AGCAATAGCATTCGCCCCGTGACCACCAAAATTCATGTACTGCTCACATGCAAAATACATTCATTCTATCCCAATAG
CTCCAAAAGTCTTAACTTGTTTCAGGAGGACCACCTCAAAAAGTTTAAAGTTTCAGAGTCTCATCTAAATCAGATATGG
GTGAGAACCAGGGCACAATTCATCCTGAGGCAAATTTTCTCCAGCTGTGAGCATGTGAAATTAAAAAAAAAAAAAA
AAGTGATGTGCTTCCAAAATACAATGGTAGGACAGATATAGGGTAGACGTTTTTTCTTTTCTTCTCTCTCTTTT
CTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
GGCTCACTGCAACCTCCACCTCCAGGTTCAAGTGATTCTCTGCTCACCCTCCTGAGTAGCTGGGATTACAGGCA
TGAGCCACCGTGCCAGCAGACGTTTTTCAATTTCAAAAAGGAGAAATAGGCCCAAAGAAGTAACAGGCCCAAAGTAAG
TCCAAACCAACACGGCAGACAAATTAATCTTAAGACTCCAGAATAATCTTTTATTTCTGCACTGCTACCTGGAT
ACATCAGGAGGGTTGAGTCCCCAAAGCCTAGAGAGCCTGCCCCCTCAGAGCTAGCATAAGGCTACTATACATG
CTGTGGCCACTGGGCTCAGTTACACAGCAGCTCTCATGGGTTGAAGTATCATGCCTTCAGCTCTCCAGGCTGGAG
TTGCACATTATTCATCTGTAGGTGAGCCAGGTGAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGG
CAGACGGATCACCTGAGGTGAGGAGTTTGAAGACAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACA
AAAATTAGCTGGGCATGGTGGCGCACACCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACC
CAGGAGGTGGGGTTGCAAGTGAAGATTGTGCCACTTCATTCAGCCTGGGTGACAGAGCAAGACTCTGTCTCA
AAAAACAAGCAAACAAACAAAAAATCTACAGTTCTTGATCTCAGAGGTGGCCCCACTCTCACAGCTCCATTAGGC
ATTGCCCTGGTGGGGCTCTCTGCAGTAGCTCCACCTCTGTGTCAAACCTTCTGCCTGGGGCTCCCGACTGCTCAATA

[illegible]

AAAGTCACCTGGAGCATAGAAAATGTTTCTGCAATTATTAATCTACTTTTTTTTTTTTTTTTGGAGACAGGGTCTCAC
TCTGTCACCCAAGCTGGAGTGCAGTGGTATGATCATGGGTCACTAAAGCCTCACCTCTTGGGCTCAAGCAAGATCT
TCCTGCCTCAGCCTTCCAAGAGCTGGGATTACATGTATGCACCACCACGCCAGCTAATTTTTAGTTCTTGTAGAGA
CAGGGGTCTCACTATGTTGCCCAGACTGGTCTGAACCTCTTGGGCTTGAGAGATCCTCCACCTTGGCTTCCCCAAA
TGCTGGGATTACAGGCATGAGCCACCACACCTGGCCTCACATATTTTGACGTCTAAAACACCAAAGATATTTTAAAG
GAGCTGAACAAACCAGTAATTTTTTAAAGAAAAAGTAGTGGGAGCATCTAGTACAACCAAAACCAGGCATCAAAGTC
AGCAGTTAAACATTCAGGCTGGGCACAGTGGCTCATGCCTGTAATCTCAACACTTTGGGAGGCCAGGTGGGAGGA
TTGCTTGAGCCCTGGAAAGCGGAGGTTGCAGTAAGCCATGATCATGCCACTGTATTCCAGCCTGGAAGACACAGTGA
AGCCCTGTCTCAAAAATAAAAGAACCAATAAAATAATCCAGAACCAGTGCTCACTTCGGCAGCACATATACTAAAAT
TGGCATGACATGGAGAAGATGAGCATGGTACCTGGGAAAGTATGACACACAAATTCATTAAGCATTTTCATATTTTAA
AAAAACATTTTCGGAACCAAGAAAATAAATGAATCAAAAGAAAGGGAAGACCCTGGTGATTCCTTTGATATTAGAAAA
AATCAGCCAGGCGTGGTGGCTCATGCCTGTAATCCCAGCAATTTGGGAGGCCGAAGCGGGCGGATCACTTTAGGTCA
GGAGTTCAAGACAGCCTGGCCAACATGGTGAACCTCCGTATCTACTAAAAGTACAAAAATTAGGCTGGGTGTGGTG
GCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGGGGTGGATCACCTGAGGTGAGGAGTCCAAGACCAGCCT
GGCCAACATGGCACCCTGCACCTCCAGCCTGGGCAACAGAGTGAGACTCTGCCTTGAAAAATAATAATAATAATAA
ATAATCAATTGCAGATAGACCATAGACCTAAATGTGAAGAGGTAAAAACAATAAATCTTTTAGAATTGCTCTGTTCA
TGTATGGCTATTGAGCACTTGAGATGTGCTATTAAAGTGTAATTTGCACTGGATTTTAAAGCACTTAGTAGGAAAAA
AAACTAAAAATATCTCAATTTTAAACATTACAATTTTGAATAATAATTTGGATATACTGAGTTAAACACTACTAAAA
TTAGTTTTATCTGATTATTTTGTCTTTTAAATGTGGATACTAGATCATTTTAAATTCATATGTGGCTTACTTCCT
ATTTCTATTGGACAGCACTAATGTAGAAGGTAATAAAAGATATATACTTTTCAAGACGTCAAATTAGGAAAAGGTGTG
TTAAATTGTATTTCAAAGCATGAATCAGGCCAGGCGTGGTGGCTCACACCTATAATCCCAGCACTTTGAGAGGCTG
AGGCAGGCAGATCACTTGAGGTTGGGAGTTCGAGATCACCTGACCAACATAGTGAAACCTGTCTCTACTAAAAAT
ACAAAATTATTTGGGCATGGTGGCGCATGCCTGTAATCCCAGCTATTTCAGAAGGCTGAGGCAGGAGAATCGCTTGAA
CCCAGGAGACAGAGGTTGCCGTGAGCCGAGATCCCAGCCTGGGCAACAAGAGTGAAACTCCATCTCAAAAAA
AATCCATGAATCATAAAGGAAAAGACTAATAGATTAATATTTGACCACACTGAAATTTTGTCTCAAAAAGACGTAA
TTGAAAGACTAAAAAGCCACAGACAGGAAAAATACATTTGTAGCGCCTACAAGTCTCAAAAGCACTTATGCCAATAATA
TATAAATCCAGCCAACCTGATAAAGAAAAGACAACCCAATAGAAACATAGGAAAACATCCGGAAAAGATGCTTCACA
AAAGAGAATCGTTAAATGACCAGTAAACATAAAAAAGTGTTCAGCCTCATTAATAAATCAGAAAATGCACATTA
CCATAATGAGATACCACTACAAACCCACCAGAATTACTATAATTAAAAGAGTCGACAAGACTGTGGAGCAACAAAGA
TATATATTTATCTTGTTTTTAGATGTCTCTCTCTATATAAATGTGCACCCAGAGACATCTAAAAACAATGGTCA
TGACAGTATTATTTGTAATAGCCCAGGTTGAAACAACCCAGTGTCCATTAGCAGTTGAAGGACTAAATAAATGGTGG
TATATCACAGAAGGGCACAGTACATAGTAAGAAAAAGAAACACAGCTATATATAATAGCATAAATGAATTTTC
ACAAGCATAATGTTGAGCAAAAGAAGCCAAACACAAAAGCATGCAAACTGCATAATTCCATTTATATAAGGTA
CAAAAAACAACCCAGCTAAGGTGATCTACAGTTTTTGAAGTCAGGATGGGGTTCCCTTGAGGAGGAAGAGAAGAC
AATGTTGGCCGGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGATGGATTACGAGGTCA
GGAGATCGAGACCAGCCTGGCTAACACAGTGAAACCTGTCTCTACTAAAAATGCAAAAAATTAGCTGGAAGTGGTG
GCGGGCACCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATGGCATGAACCCAGGAGGCGCAGCTTGCAGT
GAGCCGAGATCGCCACTGCTCTCCAGCCTGGGTGACAGAGCAATAATCCATCTCAAAAAAAGAGAAGACAATGTT
TGGGAAGGGGTATGGGGGATTCTGGGGTGTCTGTTAATGTTCTATTTTCATGACTTGTGTGTTAAAAATGTGTATTTT
GGGCTGGGCGCAGTGGCTCACACCTGTAACCCAGCACTTTGGGAGGCCAAGGCAGGTAGATCACCTGAGGTTAGGA
GTTCAAAACCAGCCTGGCCAACATGGTGAAACCTGTCTCTACTAAAAATACAAAAAATTAGCTGGGCATGGTGGCA
GTGCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAAAATCACTTGAACCCAAGGGTGGGAGGTTGCAGTGAGC
CGAGATCACACCCTGCACCTCTAACCTGGGCAACAAGAGCAAACTCTGTCTCAAAAAAAGTGATTTTGTGATGAA
TTTGTATGATCTATATATTTTTTGGAGAAAGCGCTTCCGTGATGCACAATAAAAAAGGTTTTGTTTTATTTTGTAT
TATTTCAAAATATTTTTTAAATTTAATATCTAAATATACTATTTTAGATATTCAAAATTTCTGATGATGGCTTCACT
TTTGCATCATAACATCTTTGCAATTAAATCAATTAATGTGATTAAATCAATCCATTTTGTAGTTGGGTGGTCTG
TGATTGCGTCATTGTGTGCCCCGACAGTCTTTAAGAGTGCCAATAAGTATATATATGTGTAATTTTAAAAACAATT
TTATAATTCAACTTTTATTTTAGATTACAGAGATGACATGTGCAGGTTTGTACGTGAGTATTTTCGCATGATGCTGA
GGTTGGGGGTATGAACGATCCCATCCAGATAGTAGAGTACCTGACCAATAGGTGGCTTTTCTGCCCTTGCCCTCC
CCCTCTCTTCCCTCTACCCCTCTAGTAGACCCTGTGTCTGTGCCCATCTTTATGTCCATATGTACCCAAAGTA
TGTCCATATGTCCCCTTGTAAAGTAAGAATGCAGCATTTGGTTTTCTGTTTTCTATGTTAATTTCTTAGGATCAG
GGCCTCTGGTTGCATCCATGTTGCTGCAAAAGACATGATCTCATTTCTTTTTATGGCTGTGTAGTATTTTACAGTGT
ATATGTACCACATTTTCTTTATCCAGTCCACCCTGATGGGCATATAAGTTGACTTCATGTCTTTACCATTGTGAAT
AGTGCTGCGATAAACATACAAGTGCATGTGCCTTTTTGGTAGAATGATTTATTTTCTTTTGGGTCTATACCAAGTAA
TGAGATTGCTGGGTAAATGGTAGTCTGAGCAGACTGTTTGAGGCCAGGAGTTGTGACCAGCCTGGGCAACATGGC
AAAACCCCATCTCTACTAAAAATACAAAGAAATTAGCTGGGCATAGTGGCACATGCCTGTAGCCCCAGCTACTTGGG
AGGCTGAGATGGGAGAAATCACTTGAACCCAAGAGACGGAGGTTTCAAGTGAGCTGAGATCACACCCTGCACCTCAGC

CTGGGTGACAGAGCAGGACTCTGTCTCAAAAAACAAAAAAGACAATGTGAGAATCTTAAAAGCTCAGAGAGAAGA
GTAACCTCTCACAGACAAGGGATCCGGTTAACATTAGCAGCTGATTTCTCAGCAGAAAACCTTGAAGGCCAGTAGATA
GTGGATTATATATTTAAAATAATGAAGAAACATGTCAATTGAGAAGTCTATAGCTGGAAAACCTGTCCTTCAAAAAT
GAGGGAGAAATTAAGACATTTCCAGCTTTTTTTTAAAAGCTGAGAAATCCATTTATCTCTAAATTTGCCATACAAGA
AGTGTAAAGTCCCTCAGGTTGAAATAAATGAACCTTAGGCAGTAATTATATAAGTAAATAAGCAAGTTATATGAAT
ATATAAAGCCCTCTGGTAAAGGTAAATACATAAAACAAACATAAAAAACAGTATTATTGTAATTTTGGTTTGTAACTCC
GCTTTTTATTTTCTACAGGATTTAAAAGACAAATGCATAAAATGTAATTATAAATCTGTAGGTGGTGTACAATGAA
TAAAGATACAATTTGTACATCAATAACCTAAAAAGAGTAGACCTATATAGCAGTAGAATTTTGGTATGTGATTGAA
GTTAAGTTGAAATAAATTCAAATCAGAATGTTATAACTCCAGGATGTTATATGTAATTTTCATAGTAACCAAAAACG
AAATCTACATAGAATATACACAAAAGGAAATGAGAATAGAAACAAAACCTTGTCCTACAAAAAAAATCAACTAAAGA
TAAAAGAAGTAAATAATTGAGAAAATGACTGGCAAAAATCAGTAACCTGACGTATTAACCTTTCCATGCTACATAAA
TCTGGAACTCTATTTTCACATGAACTGGAGCTGAAAGAGACAAATATTTACTTATAAAGTTAAAAGTTATATGGAA
AACCAACACTAATTTTTTTTTAGAAAAAATATGGAATAAAAAATATGCCTTATAGTATCATATAATATCATGTTTT
ATAGCTCTTGGAAAATAGAAAATAAAATGTTCTGTTAGCATGAATCCTTCTGTGCCCTCAAAAACCTTATGGATTAT
ACGATTATTACCTAAAAGTCTATTCTCAAATGCAGCACAATGATAATTTTTTATAAGGTAGATGTTATTTTTAAATA
TGGAATAATATTGGTGATTTTCAGTTTCAGTTTAAATAACATTGGATTAAAGATGAAAGAATGAGAAGATAAAGGTCCC
TCAGCAATATAACTCACAAACATGTTCAAGAAGCAGTAAGAAGTTACATTAATTATCTTTTTGAAAGTCAATAATCTAC
ATTTTTTAATGTATGCATATGGCATAGCTAATGTACTATTGCTGGGTCCATTTATTCAATGAATAATTGCTGCTATGC
GTCAGACATTTTTCTAGGCCTAGGAATGGATACATAAGTGAACAAAGCAAAGATTCTGGTCTTGTAGAGTTTCCAT
TAAAAGATCATTTAGTAAAATTTTTCTTCCCCCAAATATAAAATCTGTAAGATGATTTAACAATATGTATAAAAGT
CATTTGTGGGTGAGGCACAGTGGCTTATACCAGGTGTGGTGACTCAGCGCTCTGTCACCCAGGCTGGAGTGCAGTGGC
GCAATCTCTGCTCACTACAACCTCTGCCTCCTGGCTACAAGCGATTCTCCTGCCTCAGCTTTCCGAGTAGCAGGGAC
TACAGGTGCACAACACCACACCCGGCTAACTTTTGTATTATTAGTAGAGATGGGGTTTCGTATGTTGGCCGGGCTG
TTCTCAAGCTCCTGACCGCAAGTGATCCGCCACCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCAC
ACCTGGCCTTATTTTCTACAACCTTTGATAACTTTGGCATATACCCCAAATCTGTAACACATAAATTATCATTTCAA
TGCAACTCTTTCTCTAGCTTTTTGAATTTCTTTATTTCTAATATCAGTTTTAAGTCTGACACAAAAGCATGGGAGTTCTT
ATCAAATTTCAACCTTTTATCATAAAAACTGTCAAGAAATATCAGTAGTAGAACTTAAAAAGGAAAATAGGCCTAT
TAATTAGATTTGTATAACTTAACATTTTATGCCATATGAGTCCCCAACAAAGCCTCCAGCTTCTATTTGGATATAAAA
GTTAAAAGTCACTACTGGATCCAAAAGCAGGACTATGGTAAATAAATTTCTCCACCTAACAGCTCCTTTTACATGA
TGCTACGTGTTTCTTTTGTTTTTTCATTTTGGCAAATATTGATTGAAAGTCATATTTGTGTTAGTGTACGTCTTAAG
TGCCAAGATACAGAATCTGAAAAGAGGGACACAGGATCTGCCCTTGGGAGTTTACTTTTTTTTTTTTTTTTTTTT
TTTGAGGCGGAGTTTGTCTTGTACCCAGGCTGGAGTACAATGGTGAGATCTTGGCTCACTGCAACCTCCACCTC
CGGGTTCAGTGAATTTCTCTGCTCAGCTCCCAAGTGGCTGGGATTACAGGTACCGGCCACCAAACCCAGCTAAT
TTTTGTATTTTTTAGTAGAGGCAGGGTTTCATCATTTTTGGTCAAGGCTGGTTTTGAACTCCTGACCTCAGGTGGTCCAC
CCGCTCAGCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACACCCCGCTAGGAGTTTACACATTTAGTTGGG
GAAAAATATACAATAAGCAAGCCAATTTTTAAAATGAGAATTCAGAGTTAAATGCTACAAAGACAATCTCACAGGAA
GTTGCGATGTAGAATGATAGGGCTCTCAGAATAGTAAGAGAACTATTATTTCTTACGATGTTTGTCTTTCTTTGTA
TCGGTGCTCAGCTGAGTCTGCAGTGCTTCGGAGGCAGCTTTTCAATTTTATAAAAATCTATGATTTCTCCTTCCAGTTG
TTTTTTCTCTTCTCGAGCTTCTTATCTCCTCCTGTTGAATCATTTTTAAGATGCTCGAATTTGTCTGCAGCTGTG
AAACCAATGTGCAGTTGTGACACCAAAGCACAGTGTGGCTGAACACCCAAAAGAATATGCTTTTTTCTAAATATCAA
ACAAACCCAAATCATCACAGTAGAGCATGATCTTAATAACAACCTCAAAAACCTCAGGAGTAAACACTGAGATATGGA
ATTTTCTTTTTCTTTCTTTTTTCTTTTGTAGATGGAGTCTCACTCTGCTGCCAGGCTGGAGTGCAGTGGCGCGA
TCTCAGCTCACTGCAACCTCCATCTCCTGGTTCGAGTGATTCTCCTGCCTTAGCCTCCCAGAGCTGGGACTACAG
GCATGCACCACCATGCCAGCTAATTTTTGTATTTTTTAGTAGAGACGGGGTTTACCATGTTGGCCAGGCTGGTCTT
GAACTCCTAACCTCAGGTGATCCTCCCGCTTTGGCTCCCAAAGACTTTTTTTTTTAATATAGAAGCAAGGTCTCACTA
TGTTGCCAGGCTGGTCTCAAACTCCTGAGCTCAAGTGATCCTCCACCTCAGCTTCCCAAAGTGCTGGGACTGACT
GGATGCAGTGGCTCATGCTTGTAACTCAGCACTCTGGGAGGCCAAGGCAGAAGGATCGCTTGAGTCCAGGAGTTCA
AGACCAGCTGGGCAATGTAACACAATAGTAACTTCAACAGGAGAGAGCATCTGTAAAATGAAATATAGATCATCT
GAAATTATCCAGTGAGAGGACAAAGAAAAAAGAATAAAAAAGAGAAAGAGGCTGGGCGTGGTGGCTCACACCTG
TAATCCCAGCACTTTGGTAGGCCGAGGCAGGCAGTCAAGGTGAGGATTCAGAGGTTCAAGACCAGCTGGCCAACATAGTG
AAACCCGTTCTTACTAAAAATACAAAAAGTAGCTGGGTGTGGTGGTACATGCCTGTAGTCCAAGCTATTTGGGAGG
CTGAGGCAGGAGAATCACTTGAACCCAGGAGGCGGAGGTTGCAGTAAGCCGAGACCGTGCCATTGCACCTCTAGCCTG
GGCAACAGAGAGAGACTCCATCTCAAAAAAAGAAAAAAGAGAGAGAGAAAAGAAAGCCAATGAGACACCATTAG
GCAAAACCACTGTCAGGTTATGGGAGTCTGAGAAAGGAAAGTAGAGAAAGGAGAAGAAAGCTTATATAAAGAAAGAAATG
GCTGAAAACCTCTCTAAATCATGGGAAAGATTTAGACATCCAAATCCATGAAGCTTAAAGATTCTTAAAGAGGTTCAA
ACCAAATAGATACACCAAGTCACAATATAATCAAATAGTCAAAGTCAAAGATTTTTTAAAATGTTGAAGAGGCA
AGAGAAAAGTGGCATGACATAGACAAGAGAATCTTCATTAGACTATCAGCAGATTTCTCAGAAGGAACCTTGCAAGT

CAGGAGAGAATCAAATAATACATTCAAAGTGCTGAAAGAAAAAACTGCCAGCAACTAATACTATGTCTGACAAAGCT
AACCTTCAGAAAGAAAGAAGAAATAACATGTTTCTCGACAAACAAAGCTGAGGGAATTCAGGACCACTAGGTGTAC
CTTAAAAGAAATGCTTAAGGGAGTTTTTCAAGTAAAAATGAAAGAATGCAGCTGGGCATGGTGGCTCATGCCTGTAA
TCCCAGCATTTTTGGGAGGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGGTCAAGACCAGCCTGGCCAACATGGCGA
AACCTTATCTCTACTAAAAATACAAAAAATTAGCCAGGTGTGGTGGCAGCTGCTGAGATTGCACCACTGCACTCCAG
CCTGGGTGACAAGAGCCAAACTGCATCTCAAACAACAACAACAACAAAAAACTTGAGGCCTGGCCTCCTACTCCCTT
CCCCTCCCTTCTCCAGGCCCAAGCCACCTTGCTGAGGAGGGGGCTGAGGAAGTGTGAGCCCCCTGCCAGGAACCC
CTGCCCAGACCATGTACTCGGCCACAGGCCCTGATGCTTGCATTTCAGCAAGGCCTCCCGTGGCCTCGGCATGTTT
AGGTAGAGGAACGTGGAACCTCGCTCTTTGGCTGTGTTCCCCTGGTACTCCATCTCCTTCTGACCCCTCCCTGCAG
CCACGCAAAGCCCAGCAACCTGCCAGACACTCAGTGGCCTCCAACCTAGACAAAAGAACCTGCTGAGTTGGCAGCTGT
TGCTCATGAGCGTCCACCAGGTGGGACAGGGAGTGTGACCTTGGGCGGGCCCCCTGAAGCCACCTGCCCTGAAAGCC
CAGGGCCCAGAACCCACACACTTTGGGGGTGGTGGAACTGGTAAGGGCCACCTTCCACCATGGAGGAGGAGCCC
TGGGCCCCCTCAGGGGAGCCCCCTGCTGGACAGTGAAGACAGAGAATGACCACAATGATGCTTCTCCTCCTCATGTCT
CCTGACACCCAGTTGCCTCTACCAACCAATGATGTCAAGGCCAGTCTCTCAGTGCCCTGCCCAAGGAACAGAACTC
ATCTTCTGAGAAGGATGGACGCAGCCCCAACAAATGGGACAAAGGACACATCCGGCAGCCCCGTGAGTGGCAGTATG
ATCTTCAGCAAGCGGCACCAGGCTCTGGCAGGGCCACCAGTGTACCCCAACCAGGACAACCTGGACCGTCAGCCAG
ATCTTGGGCGAGTGGTGGTACACTCTGGGGCCCAATGAGAGGCAGAAGTACCATGACCTGACCTCCAGGTGAAGGT
GGCCCACTTGCAACAAGGACCGAAAGAGTCCAGGCCCGGAGACTCACACCTGTAATCCCAGCACCTTTGGGAGGCTG
GGGCGGGCGGATCACCTGAGGTGAGAAGTTGGCTGGCCAACATGGTGAAATCCCGTCTCTACTAAAAATACAAAA
TTAGCCGGGAGTGGTGTGGGCAACAAATCACAAAGGAAGACAGAAAGAGAGGAACAAAAGAACTACAAAACAGTCA
GAAAACAATTAATAAAATGATAATAGTAAGTCTTACCTATAAAAATTACTTTAAAAGTAAATAGATTAAAGCCACT
AATCACAAAGACACAAAGTAGCAGAATGAATAAAGAAACAATTCTCAATAATATGTGGCTTACAAGAGACTCACTTTT
GCCTTAAGGACACACTTAGACTAAAAGTGAAGGAATAGAAAACATGGCTGGGCGCAGTGGCTCATGCCTGTAATCC
CAGCACTTTGGGAGGCCAAGGCAAATGGATCACTTGAGGTGAGGAGTTCAAGACCAGTCTGGCCAACATGGTGAAAC
GCCATCTCTACTAAAAATACAAAAAATAGCCACGCCTTGTAGTGGGCACCTGTAGTCCCAGCTAC
TCAGGAGGCTGAGGCAGGAGAATCTCTTGAGTCCAGGAGGTAGAGGTTGCAGTGAGCTGAGATCGTGCCACTGTACT
CCAGCCTAGGTGACAGAGCAAGACTCTTGCTCAAAAAAAAAAAAAAAAAAGGAAAAAGAAAAAGATATTC
CATGCAAATTAAGGCAAAAGAGGACAGGGCAGGTATACTTATTGCAGACAATTAGACTTTTGGTCAAATGTTGTCA
TAAGAGACAATATCACTATATAATAATAAAGAAATCAATTTATAAAGAGGTTATAACAATCATCAAATATATGACAA
CATTGAGACACCTAATATATAAAGCAAATATTAACAGAAATGAAGGAAGAACTTACAGCAATATGAGAACAGTAGG
GGACTACTGTATCCCACCTTTAGCACCAGATAGTCCATCCAACAGAAAGTAAGTAGGGAAAGAGTAGATGTGAATA
ACTATAGCCCAATAGACCAATAGGCATAGATAGATTATTTCACTGAACAGTAGAAGAATACATTTCTTCAAGGG
CACAGGAACATTTCTCAACATAGATCATATATTGGGCCACAAATGACTCTTAACAAATTTTAAAAGTTTGAATCA
TATTATGTATCTTTTTATCATAATAGTATAATATTAGAAATCAATAACAGAAAGATTTTATTATTGAAGCACTATTC
ATAATAGCCAAAATTTGAAGCAACCTAACTGTCCACAACAGATGAGTGGATGAAAAAAATGTCACATATACACAG
TGGACTACTATTCAGCCATAAAAAAGAAATGAGATCCTGTCACTTGCAACAACATGGATGGAACCTGGAGGACCTCACG
TTAAGTGAAATAAGCCAGGCACAGAATGATAAACTTCACACGTTCTCACTTACTTGTGGGAGCTAAAAATTAATA
ATTGAACTCAGTCAGGCGCAGTAGCTCACACCTGTAATCCCAGCACTTCAGGAGGCTGAGGCAGGTGGATCACCTGA
GGTCAGGAGTTTGAAGACCAGCCTGGCCAACATGGTAAACTCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGT
GGTGGCACATGCCTGTAATCCCAACCCTTGGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCAGAGGTTG
CAGTGAGCTGAGATCCACAGCCGCACACCAGCCTGGGTGACAGAGCGAGACTTTGTCTCAAAAAAAAAAAAAATTA
AACTCATGGAGATAGAGAATGAAATGATGATTACCAGAGGCTGGGAAGGGTGGTAGGGGTGGGGGGAAGTGAGGACA
TTTCTTGGTACAAAAATATAGTTAGAAAGAATAAATAAGACGCAGTATTTGATAGCACAAACAGGGTCACATACACA
ATAATAATTTAATTGTACCTTTTAAAATAACTGAGTATCATTTGGATTTCTTGTAACACAAAGGACAAATGCTTGAAG
GGATAGATATCCCATTAACTTGATGTGATTATTACACATTTGTATGCCTATATCAAAATATCTATCGTACCCCATAA
ACAACATATCAATAGTTATGGGATACAGCAAAAGCAGTTTAAAGAGTGAAGTTTATTGCAATAAATGCCTATATTAAG
AAAAAATAGATTAAATAATGTAATTTACACCTCAAAAGAAAAAGAAAAACAACCAAGCGCAAAGTCAGTAAAAGC
AAGAAAAATAGTAAAAATCAGCAGAAATAAATAGATATAGACTAGAGAAACAATAGAAGAAAAATCAATGAACATAA
GTGTTTTTTAAAAGATAAATAAATGAACAACCTTAGAAAAAACTAAGAAAAAAGAGAGAAAACTCAACAAATAA
AATCAGAAATGAAAGGTGACACATAACCCTGACACCAGAAAGAAATACAAATGATCATGACTACTTGAATAATTA
TAAGCCAACAAAATTTGATAACCTAAAAAGAAATTTGGATATCCTTGGTTAATAAATTTCTTATAACCACACGATCTA
CGAAGACTGGATCATGAAGAACTGAAAATCTGAAAAGACTACTAATAGGTAAGGAGATTTAATCAGTAACAAAAA
TGCTACCAAAAGAAAAGGCCGTACTAGGTGACTTTACTGGGGAAGGCTACAAAACATTTAAAATGGAAATAATAGGCC
AGGTGCAGTGGCTCATGCCTGCAATCCCAGCAGTTTAGGAGGCTGAGGTGGGCAGATCACTTGAACCTCAGGAGTTTCG
AGACCAGCCTGGGCAGCATAGTGAGACCCCATCTCTACTAAAAGTACAAAAAATAGCCAGGCACGGTGGTACATGC
CTGCGGTCCCAGCTACTCGAGAAGCTGAGGTGGGAGGATCACTTGAGCCTGGGAGGTGGAAGTTGCAATGAGCTGAG

ATCGTGCCACTTCACTCCAGCCTGGGTGACACAGCGAGATCTTGTCTTAAATAGATAAAATAAAATGGAAAAATACAA
AAATACTTCCCAAAGTCTTCCAAAGAATGGAAAGAGGAGAGTATGGTTTCTTTTTTTTTTTTTTCTATGAGACGGAG
TCTTGCTCTGTCTCCAAGGCTGGAGTGCAGTGCAGTGGTGCATCTCAGCTCACTGCAACCTTCACCTCCTGGGTTC
AAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGCCAGGATTACAGCCGCACACCACCACGCCCGGCTAATTTTTATAT
TTTTAGTAGAGATGGGGTTTTCGCCAAGTTGGCCAGGCTGGTCTTGAACCTCCTGACCTGACGTGATCTGCCTGCCTCG
GCCCTCCAGAGTGCTGGGATTACAGGCATGAGCCACCTCGCCTGGCCAGGAGAATACACTTCTAAATGCATTCTGAAG
AAGACATTATTACTCTGATACTACAGCCAGAAGAGGACACTATGAGAAAAAAAATTTATAGGCCAACATTTGTGATG
AGCCAACCTAGGAAATCCTCAACAAAATACTAGCAAACTGAATTAGATCGCACATTTAAATATCAGATGGCCAAGTGGGA
ATTTATTCTTGGAAATGCAAGGATGATTCAAATACACAAATCAATAAATGTGACCACATGCAGAAAAAAAATTTAGAT
CTTTATCTCAAAACATATGTAAATCTCAACTCAAAATGGATGAAAGACTTAACCGCTATAAGATCTAAAGCTGTACA
ACTACTTTAAAAAATTAGGGATAAGTTCCCTTAACCTTGGGTCTGACAATGATTTTTTTGGATACAAACTCCAAAGCAT
AGACAACATAGGCCAAAAATGGACAAATAAAATTTATCAAACTAAAAAGCTCCTGCACAGCCAAAAAAAATCAA
CAGACTAAAGAAACAACTATGAAATGTGAGAAAAATTTTGCAACTGTACATCTGATAAGGGGTAAATGCCCCAAA
TATACAAGAACCTCAATAATTCAATAGTATGAAAACCTCATTCCGAAAATGGGCAATGAACTGAACTTTTTTTTTT
TTTTTGAGATGGAGTCTCACTGCTGCCAGGCTAGAGTGCAGTGGTGCAATCTGGGCTCACTGAACCTCTACCT
CCCCGGTTCAAGTGATTCTCTACCTCAGCCTCCCGATTAGCTGGGACTACAGGCGTGTGCCACCACATCTGGCTAA
TTTTTGATTTTTTAATAGAGACAAGGTTTCACCATGTTGGTTCAGGCTGGTCTTTAACTCCTCACCTCAAATGATCCG
CCCACCTCCACCTCCCAAAGTGCTGGGATTATAGGCGAGAGCCACTGCACCCGGCTGAACCTGAACATTTTTTAAAA
GAAGACATAAAAGGCAATAGGGATATGAAAAAATATTCACCTCATTAATCATCAGAGAAATGCCAATTTAAACC
ACAATTAGATGGCACCTCATATCTGTTAGAAAAACAGAAAGATAAGTGTTTACAAGAATGATTTTTATAGAAAGATAA
GTATTTACAAAAAAGGTAAGTGTTTATAAGAATAATTTTATAAAACAATGATAAAAAAGTAAATCAGTACAGCTAT
TATGAAAAATATGGGTATTCATTAAGCAACTGAAAATAGAATAACCATATGGTTCAGCAACCCCACTAATAGTTATG
TGTTCAAAGGAAATAAAATCAGTATGTCAAAGAGATATCTACACTTTCATGTTTATTGCAGCATTAATTGACAATAA
TCAAGACATGTAATCAACCCTAAGAGAACACATAGTTGAGATATGGAAAAGAACTAATCTCCAGAGAGAAAGAAAA
AAGAAGCAGACGTACAGGAATAAGGTTTCATAACTAAGGGAAGTGAGGTGGGAGAGAAAGAGAAAAATTTAAATGGTTG
CTTCGGTTTTCTGGTCACTTCCTAGTTTTGGTTTAAAGTCTTCATGAGGCCAGGCGCGATGGCTCATGTCTGTAATCCC
AGCACTTTGGGAGGCCGAGGCAGGCGGATCACGAGGTTCAGGAGATCGAGATCATCCTGGCTAACACGGTGAAACCCC
ATCTCTACTAAAAATACAAAAAATTTTTAGTTGGGCACGGTGGTGAGCGCCTGTAGACCCAGCTACTC
GGGAGGCTGGGGCAGGAGAATGGCGTGAACCCAGGAGGCGGAGCTTGCGGTGAGCCCAGATCGCGCCACTGCACTCC
AACCTGGGTAAACAGAGCAAGACTCTGTCTAAAAAATTTTTTTTTTTTTTTTTTTTCTCATGAGATCCAGCTATGT
CTCTTGCTCTTAGATGTCTGACGTACCTTGTATACATCCAACTTTTTCTTTTATACCTAAGCTAATGGTTTGGT
TTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCACTCTCGCCAGGCTGGAGGCAGTGGCACCATCTCGGCTCACTG
CAACCTCCACCTACTGGGTTCAAGCGATTCTCCTACCTCAGCCTCCCGAGTAGCTGGAATTATAGGCATGCAGACC
ATGCCCAGCTAATTTTTATATTTTTAGTTGAGACGGGTTTACCACGTTGGCCAGGCTGGTCTTGAACCTCTGGCT
TTAGGGGATCCGCCCCCTCCATCTCCCAAAGTGCTGGGATTATAGGTGTGAGCCACCCGACCCGGCTGGAATTTTT
ATTACTTATAATTCAACTGTCTTTGAATGAGACATCTGGCCTATAGCACCTAATTACTACTATTCTGTCTTATGAA
AAAGGAAGTTTATGTTCCAATTTACCTAATTAACTAAATGTTCTTAACTAAAGCCACCTACACTTGAGAGAAAGA
AATTATGGTTTTCAAAAATATGTCTTCAACCAAAACGTAATTTGTGGCAATAATTAAACGATATTAATAGCTTTTTT
CCCCATTTTATTAGAAATGGCATGGAATAACATAATTTACCAACTTAAATTTTTCTCATTAATCAAGGCCAGCATT
GTCAACATTTGCTTTTTTATAAAAGTGAAGACAATTTTCTAGTGCATTAATAGCATTAGACAGTACAAGACTCTTCTC
AGAGTAAATAGCTTGGGTTTTGTACAAATTAACCTTTATGCATGAAAATTTAAATTATAAATAAAACATAATTTTTTA
TAAATATATCTTCAGGGTTATTTAAATGGGAAAATTAAGTTCCTCTAGAAAAAACACTACAGACTATAATTTTGTA
GGCTATATTCAAGTCTACATGGTAAAAAGCTACTTTTTGAAAAATAGCAAGTATCCTCCAAAAACATCAAGTAACT
CTTGAGATTAAACCCCAATTAATAGGGCACCTAAAAATTAATACAAATTATATCAAACTGTCAGTACTAACCTC
TTTTTCAGCTTCTTTAAATGTTGCTTCTTCTCCTTGACTCGCTGCATAAATCTCTGTTTCAACTCTTCTTCTCCC
TCTGACATTGATCATAGAATCTTGTCTTTTGGCTTCAAAGATTCTTGAAGAACTAAAGAGAAATGTTTCTTAGTC
TTATATTAAGATGAGGTAGGAGGATCGCTTGAACCCAGGAGGCGGAGTTTGAGAGAAATCAAAATCGTGCCACTGCA
CGTCAGCCTGGGAGACAGAGTAAGACTCCATCATAAAAAATAAAATAAATGAAAAATAAAACACACTAGTATAA
ACATGTAAAAAATATATATTGTATGTTACATCCTATATAAGGAAATTTACCAACATTCTCTCATGTTTAGCAGCACA
AAATGGAAATTTCCATATCATAGATGTAACCAAGGACAACTTAAGATCAGAGTATATAGACAGTATATGAT
AAAGAATACTAGAAATTTTCAACACTTATATAGAGTGAAATTTTATCCCAACCACTTTCCAGGATATCATACATTTT
ATATCTTTTTGACAATAACATTCAGGGAAACCTCTAATGTATGAGACGATTTTCAAAAAATCCTCTAAAAATAAGCAAA
AAGTCTTTCAAAAATATCATCATTCATTTAAGATAGTTAAATTTTGGGTGTCTACAAAGAAACAATGAATGATATACT
AGATTATATACACTACAGTGCCATGTTTTTGAACACAAAGTCTATCTTCTTTTTTTTTTTGAGACAGAGTCTCTCT
TTGTGCCCCAGACTGGAGTGCAGTGGCACGATCTCAGCTCACTGCAATCTCTGCTGCCAGGTTCAAGCGATTCTCC
TGCCCTCAGCCTCCCGAGTTAACTGGGATTACGGGCACCCGCCACCACGCCAGCTAATTTTTTTTTTGATTTTTTAGTA
GGGACGGAGCTTTACCATGTTGGCCAGGCTGATCTTAACTCCTGACCTCAGGTAATCTGCCTGTCTCGGTTTTCCCA

AAGTCTGGGATTACAGGCGGGAGGCCACTGCACCCGGCTCTATTCTTCAAGTTAGAGCAAAACACCTGCTCTCAAACCT
 TTTCTCATTTAACTTCCAGGATGCTGCACTACTGCTTTTTCAGCTGGTTCCCTCCATCAAACCTCTTCTTAATGTCTTTT
 GTCAAACCTTTGCTAATATTTTTTCCATCTTTTTCTTTATCCTTTCTTTTTTTTTTCTTTTTTTTTTTTAAAGATGGAGTCT
 TGCTCAGTCCGCCACGCTGGAAGTGCAGTGAGGTGATCTCAGTCACTGCAACCTCCGCCTCCAGGGTTCAAGCGAT
 TCTCCTGCCTCAGCCTCCCAAGTGGCTGGGATACAGGCGTGTGCCACCATGCCCGGCTACTTTTTTGTATTTTTTAGTA
 GACACAGGGTTTTACCATGTTGGCCAGGCTGCTCTCAAATACCTGACCTTAGATGATCCGCCCGCCTCAGCCTCCCA
 AAGTGTCTGGGATTACAGGCATGAGCCACAGCGCCCCGGCTATCCTTTCTTTTACATAATACAGATACATTTTCATATA
 AATACAAAAAATGGAACAAAATACTGGGTTTTTTCATCAATATACATTTTCAGCATATATACATATTGAATACACATA
 TTTTCAGCATAAAAAATGAGATCTCGCATTATATTGTTCCCTACTATCTCTACCTTCGTCAAAAGATCATAAAAACACCT
 AACATTTTCAGGTGTTTTCTTTATGATGGCTCAAGATGGGAGAAGTATAATTATTTCTCATATCCAAATTATCATACCTT
 GTGTTTGTAGGGACTCTTAAATATTCCATTTTGGTTGTGCTAGCAATATTTTTTACTCACCTAACTGGCTGGTTGTTTGTG
 GACCCACATCTGTAAGGCCATTTTCTGCAGTTTGTGGTACCTATAAACATTACATAGTGTGAGTGTGGGTTTTTCTT
 TTTAGATTTTTCCATATTGGTACAAAGAAGCATATCTCGGAGCTTAAACGAAGTCACAGTGATTTTCATTTTCCACTAG
 TAAAAAATAAATAATGAATCAACAACATTCAGTATCAATCCCTGACAATCAGTAAGTGTTTCATGAATGAATACATGA
 ATAAAAAACACATTTTAAATAAACACACAACCTGAACAATTTGATTAAATAGCTATAAATTC'TTTTAAGCATGTAAGTGTG
 GTTGTACTATACATACCAACCATAGGCAGAGATTATTTATAGGTTAGCAAATTTGCTTTTAAACTTATTACCTGAAGA
 ATAAGCAGGAAATAGTGAAATAATATTTTATGTGCATATATGCCATAGGTTTTTATTAACAATTGTGTAGTATGTTA
 TAAACTACCTCTGGGTATTTTGGAGTGCTAACAGAATATGTAATATAACATCACTAGTGTAATAATTTTATCAGAG
 TAAATTTCTGCTAGCAAAAAGGCCTTACAATTGAAAATAC'TTTAGTTACATATTTCCATTTATATATCATATTTAGGA
 GACACCGTTAGTCAATATTTCAAATCAATCAACAAATTCACATAGCTGTTTTTAAATGATACATACAGACTGACAGTGA
 AGACACGGAAAGATATTTTCATGCAATTAGAAACTAAAAGAGAAGGCACGGAAAGCTATACTCCTATCACAGAAGATA
 GACTTTAAGTCAAAAACCTGTAAAATGATATAAAAGAAGGTCATCATATCATGATGAAGAGGTCAGTTTCATTAAAAAAA
 TAACAATTGTAAATATATATGACCCCAATATCACAACTTCTAAATATATAAGGCAAATATTAGCAGACCTGAGGGGG
 AGATAGACTGTATTACAATAATAGTAAGGGACTTAGATACCCTACTTTTCAGCAATGGATAGATTATCCAGACAGAAT
 ATCAGTAAGAAATATCAGATTTTAAACTACACTTTAGACCAAATAGACCTAATGGACATATGTGGAATATTTTCATCC
 AACAGCAAGATAATACCCATTTTCTCAAACAACACAAGACATCTTCAGGATAGATCATATGTTTAGGCCACAAAA
 TAAGTCTTTAACATATTTTAAAGACTGAAATAAAAATGTGAGTGCACATGTCTTTTTTGACAGATGCAATTTTCATCCAC
 TAGGGTACATACCCAGTAGTGGTACTGTGGATCATACAGTAATACTATTTTTTAGTTTTTGGAGAACCTTCATATTA
 TTTCCCAAATGGCTGTACCAACTTAAATTTCCCAACCAATGATGTTTAAGAATTCCCTTTTTTCTTTAGAAATATATAT
 ACAC'TGCTGGTGGGAATATAAATTAGTTTCAGCCATTGTGAAAAGTAGTTTGATAATTTCTTTTTTTCTTTTTTTGAG
 ACAGAGTTTCGCTTTTGTCAACCAGGCTGGAGTGCAGCAGTGCAGCTCAGCTCACTGCAACCTCCGCCTCCTGGGT
 TCAAGTGATTCTCCAACCTCAGCCTCTCAAGTAGCTGGGATTACAGAAGTCCACCACCAGGCCAGCTAATTTTTGT
 ATTTTTTAGTAAAGATGGAGTTTACCATGTTGACCAAGATGGTTTGAACCTCCTGACCTCAGGTGATCCACTCTCCT
 CAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCACGCTCGGCCGACAATTTCTTAAAGAAGTTAAACAG
 AACTACCATTTGACCCAGTAATCCCATTACCAGGTTTATATCCAAAGGAATATATATTGTTCTATCATAAAGACACA
 GGCACGTGTATGTTTCATTGTCAGCACTATTTCACAATAACAAAGACATAGAATCAATCTAAATGCCCATCAATGGTAGA
 CTGGATAGAGAAAATGAACATATACACTATGGAATACTACACAGCCATAAAAAAGAACAAGATCGTGTCTTTTGCA
 GCAACATGAATGGAGCTGAAGCCATTATCCTAAGTGATGTAATGCAGGAACAGAAAACCAAATACCACATACTCTCA
 TTTATAAGTAGGAGATAAATATTGAGAACACATAGACACAAAGGGGAGAACACAGACACTGGGGCCTCCTTGAGGG
 TGGAGGTTGGGAGGAGATGAGGACAGAAAAGCTACTATGCTTACTTTTTTTTTTTTTTTTTTTTTTTTGGTTAGTATA
 TGTGCTGCTAAAGTGAGCAATTTTTTTTTTTTGTAGATGGAATCTCGCTCTGTTGCTGAGGCTGGAGTGCATGGCAC
 GATCTCGGCTCACTGCAACCTGCACCTTCTGGGTTCAGTGAATCTCCTGCTTTCAGCCTCTTAAAGTAGCTGGGATTA
 CAGGAGCGCGCCACCACACCTGGCTAATTTTTTGTATTTTTTTAGTAGAGACGGGGTTTTTGCCATGTTGGTCAAGGCTAG
 TC'TTGAAC'TCCTGAAC'TCAGGTGATCCACGTGCCTCAGCCTGCCAAAGTGCTGGGATTACAGACTCAAGCCACTGCT
 CCAGCCAGTACTATGCTAATTATCTCTGTGATGAAATAATCTTTTACACCAAACCCCCATGACATGCAATTTATCTA
 TATAACAAATTTGCACATGTACCCCTGAACCTAAAAATAAAAAATTAAAAAAAAAAAAAAAAAAGAATTCTCTCTTCTCCA
 TATGCTTGCCAGCACTTGTAATGGCTTGTCTTTTTTTTTTTTTTTTTTTTGGAGACAGAGTCTCACTCTGTTGCCAGGCT
 GGAGTGCAGTGGCGCAATCTCGGCTCACTGCAAGCTCTGCCTCCTGGGTTACAGCCATTCTCCTGCCCCAGCCTCCC
 AAGTAGCTGGGACTACAGGCGCCCGTCACTACGCCAGGCTGATTTTGTGTATTTTTTAGTAGAGACAGGGTTTCACCG
 CGTTAGCCAGGATGGTCTCCATCTCCTGACCTCGTGATCCACCTGCCCTCAGCCTCCCAAAGTGCTGGGATTACAGGT
 GTGAGCCACTGCGCCCGGCTGTCTTTTTTGATAATAGTCATTCTAAAAAGACATGAAGTGATACTTTCATCATGTGTTT
 AATTCGCATTTTCCCTGATGTTTTAGAGATGCTATTTTTTTTCATATTTCTTTTGCCATTTGTATTTCTTCTTTTGAGA
 AATGCCCTGTTTCAGATCCTTTGCCCATTTTTCCTTGTGTTATTTGTTTTCTTGCTATTGAGTTGTTAGAGTAATATCA
 AAAATATTATATATTTTGGACATTTTGCCCCCTTATCAGATGTATAAGTTGCAAAATATTTTTTCCAATCTCTAGGTTGT
 ATCTTTATGCTATTTAAATGTTTTCTTTGGTTTTTGTGTGCTGGTGTTTTGAAGTCTTTTTTTTTTTTTTTTTTTGAGG
 CAGAGTCTTGCTCTGTCAACCAGGCTGGAGTTCAATGGCAGCATCTTGGCTCACTGCAACCTCTGCCTCCCGGGTTT
 AAGCAATTTCTCCTGCCTCAGCCTCCCTCCCGAGTAGCTGGGATTACTGGCACCCACCAACATGCCAGCTAATTTTTT

GTATTTT TAGTAGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTCTGAAACTCCTGACCTCAGGTGATCACCCGCC
TCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGTCACTGCACCCTGCCAGCTTTTGTATATAATCCGCTCTCCTG
TGGCAGGAGTGCACAGACCAGGGATCATGAAGACCCCTCGAGTCAAACCTGGGTGTGAGGCTGGAGTTGGCTTCACT
ATGGACACATCCAATAAGCCCAAACTAGGGTAACAGTGATAGCACAGCTTGACCTTTCCCTTCTGCCTTTTCTCTTG
TACAGACTGAGTTCTGTTCTTGGATGCAGGACATCTTTTGACTGTTGCAATTTGAAGATTCTTATGTTTATTTTCTT
CTTCTTCTCTTCCACTATCATTTTCATCTACTTGTATCACCTTTTCGTGTTTCAGCAGATGTCAGGGGTGTTTGAATA
GCTCCCGCGGTGATATCCAGGTGAGCTGTTGCTGCCCTTCCAAGCCCTGTTTGTTCCTCCAAGGGAGAAGTAGTT
GAGCCATTCATGCCATTATCTTTGCGCAGTGCCCTGGGGAGCCGCAGGATCTGGATGTTGTAGAGTTTATCCACCT
CCTTGAAGAGGTTCTGCCGTGTCAGACTCAATTGGCTTGATTTCGATTTCCACTTTGCAGTCGAAGTCTTTAAGAAAG
GAGGCGAGCTTCCACCTCCCTAAGGAGTTGGTCTTGGCCACCCAACCTACTGCCCTTTCTAGGAGCCATGGTGTGGG
TGGAGGCTGCAGGGAGCGATGGAGATGGGGAAGGAAGGCAAACCTCTTCTCCAATTTTAAATCAAGTTTACAGTTCT
TTTTTTTTTTTTTTTTTTTTTTTTTGTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGTAGTGGTGTGATCTCAGCT
CACTGCAACCTCCGCTTCCCAGGTTCAAGTGATCTCCTGCCCTCAGCCCCCAGGTAGCTGAGACTATAGGCATGCA
CCACCACGCTTGTTTTTCTGTTTTTTAGTAGAGAAGGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAATCCTGA
CTTCAGGTGATCTGCCACCTCAGCCTCCTAAAGCACTGGGATTACAGGTGTGAGCCACCGCACCCAGCATTTTTTT
TTTTTTTGTAGACAGAGTCTCGCTCTGTCACTCAGGCTGGAGTGAGGTGCAATCTCGGCTCACTGCCACCTCCGCC
TCCCAGGTTCAAGTGATTCTCTGCCTCAGCCTCCTGAGAAGCTGGGATTACAGGCGCCCGCCACCACACATGGCTA
AGTTTTTGCATTTTTTAGTAGAAATGGGGTTTTGCCATGTGGGCCACGCTGGTCTCAAGCTCCTGGCCTCAAGTGATC
CACCCGCCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGTGTCTGGCCAAGTTTTACAGTTCTTAGT
GTGCAGGTATTTTTACTTCTTGTTTAGATTTATTCCTAAGCATTTTATTTTTTACAGCATTATTCACAATAGCTAAGAT
ATGGAAGCAATCTAATCTAAATGCCCATCAACAAGATACACAATAAAATACTATAGAGCTTTTAAAAAGAATGAAAT
CCTGTCATGTGTAAAACCTGAATGAACTGTAGGACATTATGCTAAGAGAAATAAACCAGGCACAGAAGGACAAATA
CTGCATAATCTCACTTATATGTGGTAACTAAAAATGTCAATCTCACAGAAATAGAGAGTAGACAGGTGGTGAAGTGA
GGAAAGTGGGGAGACTGAGGGATGAGGAAATGAAAAATATTGATCAAACGGTACGAAGTTTTAGTTAGACTGGAGGA
ATAAGCTTCAGTGATCTACGGCACTGTATGGTGAAGTACAGTTAATAATAAGCTACTGTATATTTCAAATTTGCTAAA
ATGGGGAGTGGGGTGGAGCAAGGTAGCTGAATAGAACCTCTCATCAGTCATCCCCCTACATAAACACCAAATTGAAC
AAATATCCATGCAAGAAAACACCTTCATAAGGATGAAAAATCAGGTGAGCAATCATAGTACCTGATTTTAAATATTAT
ATCAAGAAAAGAGGCACCTGACGATTATAGTAAAGAGAGTCTTGCATAGCCTATGCCATCTGACATTCATCCCCCAGC
AGCACCATACTAAACAGAGAGAGAATCTGTGTGTTTTGGGGAGGGAGAGTGAAGTGATTGTGGGACTTTGCATTGGA
ACTCAGTGACTGCCCCGCCACAGTGCTACACAACATAGGGCAGAATCTGTCTAGTGCCACGACAGGAGCCTTTTTT
TTTTTTTTTAGATGGAGTCTCGCTCTTTCACGCAGGCTGGAGTGCAGTGGCGCAATCTCGGCTTACTGCAACCTCCAC
TTCGACAGGTCAAGTGATTCTCCGGCCTCAGCCTCCCAAGTAGCTGGGACTACAGCTGCACACCCTATGCCAGCT
AATTTTGTATTTTTTAGTAGAGACAGGGTTTACCATTATGGTGGTGGTCTTGAATCCTGACCTCAGGTGAGTGC
CACCCGCTTGGCCTCCCTAAAGTGCTGGGACTATAGGTGTAAGCCACCCTCCTGGCCAAAGGGAGCATTTAACCTA
GCTCCCACTAAAGAGGAATACCCCATGCCAGAGATAGATAGGAACCTTAAGTTCCAGCTACCTCCACCCTAGTTAAT
TAAAGTGCCCTGGGGTCTTGAATAAATCTGAAAGGCAATCAGGCCCAAGGTCTGCAGTCTTGGGCAAGTCCAGTT
ACTGAGCTTGCTCAGAGCCAGTGAATAGGAATGCGTGTGACCCAGCAAGACACCAGTTTGGTGGCCAAGGAGTGC
TTGTATCACACCTCCCCCACTCCAGGCAGTGCAGCCCGAAAAGAGACTCCTTCCACTTGTGGGAAAGAAAAGGAAG
ATTAGAAGAGAACTTTATCTTACAACCTTGAAGTACAGCTCAGCCACAGTAAAACACACAAGAGAGTCCCTTCCACTTG
TGGGAAAGAAAAGGAAGATTAGAAGAGAACTTTATCTTACAACCTTGAAGTACAGCTCAGCCACAGTAAAACACACCA
CTAAGCCCATCTCTGAAGCCCCCAATTCCAGGCCCTACTTCTTGGACATTTCTAGACCTACTCTGGGACAGAAGGGA
ACTTGCTGCACTGAAGAGATGGACCCAGCCAGCAGGACTCACCACCTGCTGACTAAAGAGCCCTTGGGCCTTAAAC
ATCAGTAGTAGCAAGGCAGTAGTTGCCATGGGCCTTGGGCAGGACCCAGTAATTTGCTGGTTTCATATGTGACCCAG
CACAGTACCAGCTGTGATGGCCACAGGAGTGATGATGTCACCTTCCCTCAACTCCAAGCAACCTAGCACAGACAGA
GAGACTCCTTCTGTTTGGGAGAAAGTGAGAAAAGAGTTTGGAGAGATTCTGCCTCATAATCCAGGGAAATCTCCCAGA
TATTATCCAAGACCACTGAGGCAGTACGAGTCTGCTGGCGTCACAGTGTACTGGGCTTGGGGTACCACCTAATACA
TTCCCTTTGAATACTTAGAAAACCTTCTCAAATAGAATAGGCACAATTGAGCCAGACTGCAAAGATTACAATAAAT
ATCTAACTCTTCAATGCCAGACATAGATGAACATTGACAAGCATCAAGACCATCCCAGAAAACATGACCTCACCAA
ATGAACCAAATGAGGCACATATGACAAATCCCAATCACATATCTCAGGACAGATATATGTAACCTTTCAGAGATCTC
AAAATAGCTGTTTTTAAAAGAGCTCAGTGAAATTCAAGATAACACAGAGAAGAAATTCAGAATTCATTTAGCATTATT
ATATGCCAACAGTGAACAATCTGAAAAAGAAATAAAAACTAATCCAATTTATGACAGCTATAAAATAAAATACCTAG
AAATAAACATAACCAAGAAATGAAAAATCTCTACAATGAAAACTATAAAATACTGATGAAAAGAAATGAAAAAGAAAT
GCAAAACAATGAAAACACAGTTTCATGTTTCATGGATTGGAAGAATCAATAATATTGTTAAAATGTTTCATACTGCTCAA
AGCAATCTAAAAGATTCAATGCGATCCCTATCAAAAATACCAATGACATCTTTCATAGAAAAAGAAAAAGCAATTTCAA
GATGTATATGGAACCAAAAAATACTTAAAGCAATTTCTGGGCAAAAATAACACTGGAGGAATCGCATTACCTGACTT
CGAATTATACTACAAAGCTATAGTAAGTGAAACAGCATGGTACCGGCATAAAAACAGACACAAAGACCAATGGAACA
GAATAGCGAACCAGAAATAAATCCACACAGCCACAGAGAACTCATTTTTCAACAAAGGTGCCAAGCACATACATTGG

GGAAAGGATAAGCGCTTCAATAAATGGTGTGTGAAAACGGATATCCATTTGCAGAAGAAAGAACTAGACTTTTAT
TCTCTCAACACATATAAAAAATCAAATCAAAATGGATTAAAGACTTAAATCTAAGACCTCAAATTTATGAAACGACTAC
AAAAAACATTGGGGAAGCTCTTTAGGACATTTAGTGAGTAATACCCCACAAGCACAGGCAACCAAAGCAAAAATAGA
CAAATGGGATCACCTCAAGTTAAAAAGCTTCTGCAAAACAAACGAAACAACAAAGTGAAGAAACAATACACACAATG
AGAAAATATTTGCAAACTATTTATCTGACAAAAGATTAATAGCCAGAATATATTTAAGAAGCTCAAACAACCTCAATA
GGAAAAAATCTAATAATTTAATTTAAAGTGGGCAAAAGATCCGAATAGACATTTCTCCAAAGAAGACATACAAATG
TCTTCTTTTGTCTTATTGATAATAACTGTTTAAACATGTGTGAAGTAACTGCTCATTTGTGGGTTTTTTTTTAAATGCCA
AACAGCAAATGAAAAGGTGCTCAATATCACTGATTACCAGAGAAAGGCAAATAAAAACTACAATGAGATATAATCT
CACCCCAGTTAAAAAGCTTTTTATCCAAAAGACAGGCAATAACAAATGCTGGTGAGGATGTGGAGAAAAAGGAATCC
TTGTACACTGTTGATAGGAATGTCAATCAGTACAGTATCTATGGAAAACAGTATGGAGTTTCCCTCAAAAACTAAAA
ATAGAATACCATATGATCCAGCAATCCCACTAGTAGGTGTATACCCAGAAGAAAGGAAAGTAGTATATCAAAGAGA
TAGCTCCACTCCCATGTTTGTACAGCACTATTCACAATAGTCAAGATTTGGAAGCAACCTAAGTATCCATCATCAG
AGGAATGAATAAAGTAAATGTGGTACATATACATGATGGAGTACTATTACGCCATAAAAAATAATGAGATCCTGTCA
TTTGCAACAACATGGATGGAACCTGAGGACATTTCTGTTAAGTGAAATAAGCAAGACACAGGAAGACAACTTCACATA
TTCTCACTCATTTGTGGAGCTAAAAATTAACAATTAAGCCATGGGGATAAAGAGTGGAAATGTTGATTACCAAG
GCTAGGAAAGGATTTGGGGAGGAGAGAAAGTGGGGATAGTTGATGGGTACGAATATATAGTTACATGAATGAATAAG
ATCTAGTATTTGATAGCACAAAAGGGTGACTACAGTCAACAATAATTTATTGTACATTTTAGAATAACTAAAAGAGT
ATAGTTGGATTGTTTGTGACACAAAGAAAGGATAAATGCTTGAGTTGATGGGTACTCTATCCTGATGTGATTACTAC
ACACTGAATGCCAGTATCAAATATCTCATGTGCCCTATAAATATACACACCTACTAGGTATCCACAAAAATTA
TAAAAAACACAGGCCAGTGTGGTGGCTCATGCCGTGTAATCCTAGCATTTTGGGAGGCTGAGGCAGGCAGATCACC
TGAGGTCAGGAGTTTGAGACTATCCTGACCAACAAGCCAAAACCTGTCTCTACTAAAAATACAAAAA
ATTACCCAGGCTTGGTGGCGCATGCCGTGTAATCCCAGCTACTGAGGAAGGAGAATTGCTTGAACCCAGGAGGCAAG
GTTGCAGTGAGCCAAGATCACACCATTGCACTCCAGCCTGGGCAACAAGAGCTAACTCCATCTCAAAAAA
AAAAACACAACAAGCTGTAACCTTACCCATGCTAAAATGGCTATTATCAATAAGACAAACAAGTGTGGTAAGGATG
TGGAAAAAAGGGAACCTTGTACACTGTTGGTGGCAATGTAAATTAGGATAACCATTTATGAAAAACGGTATGGAGTT
TCTTCAAAAAATTAATAAATAAACAACCTACCATATGATCCAGCCATCTCACTACTGGATACATTACCTCCCA
GAAGTCAGTATGTGAGAGAGATACTTACACTACCATTATTCACAAGAACCAAGATATGGAAGCAATCTAAGTGTCCA
TGAATAGATGCATGAAGAAAAAAGTGTGGCATATATACGCAATGGAATTCAGCCTTAAAAATCAGGAAATCTATC
ATTTGCAGCAACATGGATGGAACCTGGATGACTTTACACTAAGTGAATACGCCAAACACAGAAAGACAAATGGTAAAT
GATCTCACTAATAGATCAAATTTTTTAAAGTTAACTTGTAGAAACAGAGAGTAGAATGGTGGTTACGTGAGTTAGG
TAGAGGGTATGTAGGAGAGATGTTGGTCAAAGGATACAAAATTTTAGTTAGGTAAGAACAAGTTAAAGAGATCTATT
GTACAACATAGTGACCATTGTTAATAACAATATTTTATATATATATATATATTTTTTTTTTTTGAAAAGGAGTCTCGC
TCTGTTGCCAGCTGGAGTGCAGTGGCGTGATCTCAGCTCAGTCAAGCTCCACCTCCCGGGTTACATCATCTCC
TGCCTCAGCCTCCCAAGTAGCTGGGACTACAGGTGCCCGCCACCATGCCTGGCTAATTTTTTTTGTATTTTAGTAGA
GATTGGGTTTCACTGTGTTAGCCAGGCTGGTCTCAATCTCCTGACCTCATGATCCGCCCGCTTCAGCCTCCCAAGT
GCTGGGATTACAGGCGTGAGCCACTGCGCCCGGCCAATATTTTATATTCTTGAAAAATTGCTAAGATAGATGTTAAGT
TTTCATCACAAAAATAATTGAGATAACACATATGTTAATTATATCAATTTAGCCATCCCACAATGTATATGTATTTT
AAAACATCATGTTGTACATAAGTACATAAAAATTTTATTTGTCAATTAAGATTAATTAATGAAAAATAATTTGTATGG
CTCTCTCCTGTGTTTGTATGTCCAGAAGATTGACCACAATATATGGTTTATTGTGATTAATACTGTGACCAATTT
TATGTATCAACATATAGTATAGATATATATTTTATAACATATACTAACATGTAATACAACATAATAAATATTTAAAC
ATTAAGTTTTAAGTATGCTTAACAAATTTTAAATTTTCTAATATGTCAATTTTTTTAACAACATCATTTTGAGGCTGGA
CACGGTGGCTCATGCCGTGTAATCCCAGCACTTTGGGAGGCCGAGGCAGGTGGATCACCGGAGGTCAGGAGTTCAAGA
TCAGCCTCACTAATATGGTGAACCCCATCTCTACTAAAAATACAAAAATTAGCCCGGTGTGGCGGCAAGCGCATGT
AATCCCAGCTACTCGGAAGGCTTAGGCAGGAGAATAGCTGGAACCTGGGAGACAGAGGTTGCAAGTGAAGCAAGATCA
TGCCACTGCACTCCAGCCTGGGCAACACGAGTGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAATCATTTGGAA
ACCAGTACAAGTAGAAATCTTCCATTCCCCAATAATTATATGAACCACAAGTATATGAACCAGACAATAAATTA
CTTACTACACTATTTCTAATCTATTATGATGCTAAATAATAATATAAATGTCAAAGGTATAATCTCTAGCCAAAGGTA
AATTGTACTTTCTTGTTTTTTAAAGTTTATAACTCTTTTTTCTAGTGACCAGTTTAAATGGTTTCTACATTAGGTA
GTAATGAGTTTATGTAACATGTCTACACACTAGTATTTTGTCTCCATTTTGTGAAAGTGCTTTCCAGCACTCTCA
GTGTCATGATTTATACAGAGAGATAATATGATACATGGCACATTTGGGAGCGTTGGGATAGACAAAGAGTTCTTCCCA
GCCTCTGTAGATTTTGGCTTAAAGCAGCGTTACGCAAACTTCTATGAGAATGAATCATAGTCAAAGTAAAGCAATAGT
TTAAATAAAACAGAGTTTGTGTTCCCATTAGTAACGTGTTAACCTTACAGATTAGCAAGAGTTACCAATATTAGTTCA
AGTTAAAAATGTTATTTTGTGAGACTATGAGTGATACACATATATTTAACCTAATCATTTCCAAAGCCTGCCCTCTGT
TTTTACCTTGCAAACTCCCCAAGGGTAGTGACGGCCTCTGACCATCCTTTTTTCAACTTTCACTTCATCTGTACTC
CCTACCACAGCAAAGGGTAACAGCCCTTAGAAAAACAAGATGACATGAATTGAATTCTCTAGCACCATTAGAGCTTA
CTCTATTACTCTCATTTCTGTTATAAAAAATGACAGGCCAGCGCGGTGGCTCATGCCGTGTAATCCCAGCACTTTGGGA
GGCCGAGGTGGCGGATCATGAGGTGAGGAGATCGAGACCATCCTGGCTATCATGGCGGAACCCCATCTCTACTAAA

TATACAAAAAATTAGCCGGGCATGGTGGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGC
GTGAACCCGGAAGGTGGAGCTTGCAGTGAGCCGAGATCATACCCTGCACCCAGCCTGGGCAACAGAGCAAGACTC
CGCCACACACACAAAAATGACAGAGATGATGATATTGTAATAAATTTAATTAGACAGGGATCTACCCCTGCATCTC
ATTTTAAGACTTGACAGTATGCACTAATTGGTTTCTGTGTCAGCATAAGTCTCAGGCTGTGCACACAATTGGAACTGT
CTACTGAGCCAGAGCCTAAAGCAAGGTATCCTGCTGTGGGAAACACATAGAAAAAGATTTTCAGTAGGCATCATTT
TTCTCACACTATGTTTACCCTGAACCATCTTTAACCTGAATGTTGAAGAAATAACATGCAAATAACTTTTAACTCTG
GTATAAAAGTTAAAAGACAAAAATATTA AAAACAAC TACAAC TATTATAATAATTAATGGATACACAATATAAAATA
CTTTAATAAGCAATACTGTATATGTCTCAATAAAATATGTTTGT TTTTTT GAGACAGAGCCTGGCTCTGT CACCCAG
GCTGTAGTGCAGTGGCATGATTTTGGCTCACTGCAACCTCCACCTTCCAGGTTCAAGCAATTCTCCTGCCCTCAGCCT
CCCGAGTAGCTGGGATTACAGGTGCCTGCCACCATGCCCTGGCTAATTTTTGTATTGTTTAGTAGAGACGGAGTTTCA
CCATGTTGGCCACGCTGGTCTTGAACCTCCTGACCTCAGGTGTCTGTCCACCTCAGCCTCCCAAAGTGCTGAGATTA
CGGGCGTGAGCCACCACGCCCATCCCAAATAATTTTTTACATTTTACATTACTTAGCATTTCAATTCAGAAATCTAT
TAGTGTTTGAGGTATCTTAAATAAAGGAGGTGTAAATTATAAGAAAAGAGAAATGGAAGACATGGAATTAACAGTT
ATCAGTTACAGTTACCTGTTACATTCTAGACTCTGTGTTTCAGATCCTTCACAAAGATTTTGGCACTTACACCTCAG
AAAAAAGCCTTAATAGAAAGTGTCAATTGATCCCATAGCATAGATAACAGTCTCAGGAAGATGAGTGAATTTGTGCAG
GATCTTTTGGCTTTTAAAGTAGAGAAGCTAGGATGGAAATCTAAATCTGTATACTTTCAAAATGTATCATCTTTTCGGC
TGGGCATGGTGGCTCATGCCATAATCCTAGCACTTTGGGAGGCCGAGGTGGGCGGATCACCTGAGGTGAGGAGTTT
GAGACCAGCCTTGCCAACATGGTAAAACCCGCTCTCTACTAAAAATACAGAAATTAGCGGGCATGGTGGCAGGCAC
ATGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTGAATCCAGGAGGTGGAGGTTGCAGTGAGCTGAG
ATCGCGCCATTACATCTCGGTCTGAGAAACAAGAGTGAGACTCTGTCTCAAAAGAAAAAAGAAAAAGAAATTTATCA
TCTTTCAAAATTAAGCTAATATTTTAGGTATTTACTAGCCTAAAATATTTACACAGATAAAGAGTTTCATCATCTTT
ATTACTGCCCTGATTCCAGAACCAGAGACAACACAGCATTCGCATCCAACCCCATCTACCTGTTTTTGGAGGCTAAG
TCCTTGGCAAGTCTATTTGAGAGTTGGAGGCTCCCTTTTACCACCTCAAGACCTACTTGTGAAATAGGGGCCCTGCCT
TGGATGCAGCAAACAGAGAAGATGGGGGCCCTGACCATCCTTGCCCTGGTTTCATGAGGCAGTGGCTCATTGTGTGGGG
AAGTAATCCAAAGAACCTCAGACTGGTGCTCCTCCTCACTGAGCTCTCAGTTACCAGTGGGTGCC TGAAGCTTCAGG
GCAGGGTGGGTAACCTGCCATTGTTTTCCACCCACAATTCAGAACCTGCCTCAGAGATTTTGCCCTTGGAGGAGAAA
CAGGCCATGAAATGGAGAAAACCCATTCAAAAAATATATGTAATCCCAAGGGACCTGAATAGCCAAAAAATATTTG
AAAAGAAAGACCAAAGTTAGAGGTCTCACACTTCTAGAATTCAAAACATATTACAAAGCTACAGTAATCAAAATAGT
TTGGAAGTAGCATAAAGACAGACATGTAGACCAATGGAATCAAAACAGAGTCCAGGGTTAATCCCTCCTTACCTAT
TCTTCTGTTGATGGATATTTGGGTTGCTTCCATCTCTCAGCTACTGTGGATAATAATGCAATTTATATGGCTTTTACA
AATATCTAAGTAAGACCCCTGCTTTCAATCTTTTACAGAGAAAGAGGATTGTTAGATCATATGGTAATTTTATTTTCAA
CTTTTGGAGAAACCACCTATATTAGTCTGTTTTAACGCTACTGATAAAAGATATACCCAAGACTGGGAAGAAAAAGAG
GCTTAATTTGGACTTTACAGTTCCACATGGCAGGGGAGGCCTTAGAAGCATGGTGGGAAGTGAAGGCATTTCTCATAT
GGCAGCGGCAAGAGAAAAATTAGGAAGAAGCAAAAGCAGAAACCCCTGATAAACCTGTGAGATCTCGTGAGACTTATT
CACTATCACAAAGATAGCACAGGAAGACCAGCCCCATAATTCAATTACCTTCCCCTGGGTCCCCCGAAACATG
TGGGAATTCTGGGAGATATGATTCAAGTTGAGATTTGGGTGGAGACACAGCCAAACCATATAACCACCATACTGTTT
TTCATAGTGGTTCTAGATGAGAAAAAGATAAAACTCTATTGAAAAAATCAGGGACTATCTAAATAAATGGAGAGACA
TTCCATGTACATGGATAGAAAGACTCAATATTGTCAAGATGTCAATGTTCCATAACTTAACCTATGTATTCACCACA
ACCCCAATCAAAATTTTACCACAAATTATTTTGTGAATAATAACAAATTAACCTCAAAATTTATATGAAAAGGCAAAAG
AACTAGAATAGCCAAATGCAATATTGAAGAAGAACAAAATTTGGAAGACATACTATTCAACCTCAGGATTCAGTGCAAA
GCCATAGTAATCAAAACAGCGTGGCATGGGCAAAATAATAGACAAACAGATCAATAGAGCAGAGCAGAGAGCCCAGA
AATATGCCTACACAAATAAAGTTAACTAATATTTGACAAAAGCAATGGGAATTTAATGGAGACAAGATATTTCTTT
TCAATAAACAGTGCTGGAACCTGAATACTCACATGTAAAAATAAGAATCTACAAACAGACCTTACATATTTTACAAAA
TTAACTCAAATTTGGATCATAGGCCATAATGTAAAGCATAAAACAATAAAATTCCTAGAAGATAATGTAGAGTCTTAA
TTAGAGAAAGGGAGTCAGGCTGGTGGGAGCAGGGGAAAAAGCAAAAAAAAAAAAAAAAAAATCAGATAAGCTACAAGT
CTGCCTTTCTTACCATCCAGAACACACGGCCCTCCTGCACAAATAACTCACAATCTTCTGTGCCTAGCTATCAGC
AGACCTCGGCTGATACAAAAAATTGCAAGTTAGCTCACTGCCACTGATAGTGGCAGGAGGCGGACAAATGCCTAGG
CAGATAGGGGCAAGTCCCCAGGGAAACCCCACTTCCAACCCAAGGATAGTTTAAAGCCTGAAAGCCAAGTTACAAGT
CAAATCCATGGACTGGATTGAGAACCCTGTCTTCTGCTTGGTGCACCTTCTCTGGCTGATCTCCACCTTTCACCTG
TTTTACATATACTACACTTCCATAATTGTTTTTTTACATATGTTGTGCCCCACCTTTGGGTGGTGCCCTTGTTTTTAAC
CTTTTTGCATACTCACAAACCAATCAGCATGCCTATCTTGAGCCATAAAAGCCCTGGGCTTAGCCACACA
AGGAGAGAAACAGCCCCAGGTCCCCCTCTCCACTGAGAGCTGTTCCATCACTCAATAAAATTTCTTCTGTCTTCTTCA
TCCTTCAATTGTCAGGGTATCCTCATTTCTTCTTGGTTGTGGGACAAGAGCTCATGAAGTGTGATCTCAGATAGAAG
CTTTAACACAGGCAGGCTGGGGCATGCTGGCCAGCCACAGGCTGAGCCACTGTGCAAGCCAGAATGGGCCAAATGG
ATGGGGCTACCTGTGTGGCAGGTAGCGCGCCCAAGCAAGGCCAGGTGGGGGTGTCACTGGCTGGAGGTCCCCTGGCT
GGCAAAATGATTGAGAGGAATCCTGCATCACAACTTGGCATTATCAGCACTACACAAAGCCCTCTTCAGCATACAG
CCTAAGAAGTAGCCTATAAAATCTGCAGCAATCCTTTGTTTGTGTTGTCAGTCAGCTTCTTTCTTGAAGCTGCCTGTT

[illegible]

AGAATGTGAATGAAGTACAGATATCTTAAAAAAAAAACAATTGAATTTCTAGAAATGAAAAACACACTGAAGGGTTG
GGTTACAGATGAAAGCTTTTTTTTTTTTTTTTTTCTTTTTTGAGACAGAGTCTTGCTCTGTGCCCCAGGCTGCAGTGCA
GTGGCGCCATCTCAGCTCACTGCAAGCTCCACCTCCCGGGTTTCACACCATTCTCCTGCCTCAGCCTCCTGAGTAGCT
GGGACTACAGGCACCCACTACCACGCCCCGGCTAATTTTTTTGTATTTTTTAGTAGAGGTGGGGTTTCACCGTGTTAGCC
AGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCCATCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA
CTGCACCCGCGCCAGATGAAAGCTTTAACAATAGGCTAGACTAAGCAGAAGACAGACAGAAAAATAGGATCCCAGTGCC
TTTTGGCTTGTAACCTTCTTCTGAGAAAGTCGGCTGTTAGTCTAATGTGATTTCTTTTATAGGTGATTTGAAAATTG
TCTCTTGCTTCTTTTCAGAGTATTTTCTTCAAGTTGGCTTTGGATAGTCTGGTGAAGTTCATCTTGCAAAATATTTT
CAGGTATTTTCCGAGCTTCTCGTATCTGGATGTGTACCTCTCTAGCAAGGCCAAGACAATTTTCTGATTTATCCCT
CAAATATGTTTCCAATCTTTTTACCTTTTCTCTCTCAGAAATGCCTACAACCTATAAGTTTGGTTGCTTTA
CATAATTTCCGATTTATTGAACCTTTGATTATTTATTTAAATTTACTTTTTTTTTTGTCTGTGTTAGTTCAAAAAAC
CTGTTTTGTTTTTTGAAAAGATAAGAAAATTGATATGGTGCTAGGCCGGGCACGGTGGCTCACGCCTATAATCCTAG
CACTTTGGGAGGCCAAGACAGGTGGATTGCCTGAGCTCAGGAGTTCAGATCAGCCTGGGCAAGATGGTGAAACCCC
GTATCTACTAAAATACAAAAATCAGCTGTGCATGGTGGCAGGTGCTGTAATCTCAGCTACTCAGGAGGCTGAGGC
ATGAGAAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCAAGATCGTGCCACTGAACCCCAGCCTGGACAACA
AAGCAAAACTCTGTCTCAAAAAAAAAAAAAAGAAAGAAAAGAAAAGAAAATTTGATATGGTGCTAGCTAGATTAACC
AAGAAAAAGACAGAAGATTCAAGTAAGCACAAAGCACAAATCAGAAGTGATAAAAGTGACATTACAACCTGATACCACA
GAAAACAAAAGATCATCAGAGACTAATATGAATATCTCTATGCATACAACTAGAAAACCTAGATAAAATGGTTAAA
TTCATCCCTCCCTGCTCGGATGTGGCTGGAGCTGGAGGTGCAGAGGGCCAGAGATGCTGCAGACCTTCCACCCAGAG
CAGCTTGGAGGCAGTGAATAACAGCTCTTCAAGTCTGCAACAAAAATGGCCTCTTCCAGAGAATTGTTTTGAACCTG
GGAGGCAGAGGTTGCAGTGAGCCGAGATAGCGCCACTGTGCTCCAGCCTGGGTGACAGAACGAGACTCCATCTCAAA
AAAAAAAAAAAAAAAAAAAAAGCCTCTTCATAGCTTAAATTACAAAAGTAAACTAAAAGACTTACATGCTTTTGGCC
AGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCGGATCACGAGGTGAGGAGATCGAG
ACTATCCTGGCTAACACGGTGAAACCCCGTCTCTACTGAAAATACAAAAATTAGCCGGGCGTGGTGGCGGGCGCCT
ATAGTCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATGGCGTGAACCCGGGAGGCGGAGCTTGCAGTGAGCAGAGAT
CGCACCGCTGCACCTCCAGCCTGGGCGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAGGACTTGCATGTTTTTTT
CTTTACTCCAACCTCATTCCTTACATGTAGGCTCAATCTTTTCAGTATTTGCTTTACTGGTTCAGCAAAAGCCAGGAA
ATACAACTTTGTGGTAATCAGAATGTTATCCAACGTATATATATATATATATATATATATATATATATATTTTTTTTTT
TTGAGACGGAGTCTCGCTCTGTCTACCAGGCTGGAGTGCAGTGGCGCGATCTCCGCTCACTGCAAGCTCCGCCTCTC
GGGTTACAGCCATTCTGCCTCAGCCTCCCGAGTTGCTGGGACTGGAGGCGCCTGCCACCACACCTGGCTAATTTTTT
GTATTTTTTAGTACAGAAGGGTTTCATGGTGTGGCCAGGATAGTCTCGATTTCTGACCTCATGATCCGCCCGCCT
TGGCCTCCCAAAGTGCTGAGATTACAGGGGTGAGCCACCAGCCAGGCCCAACTGTATATTTGTTTACCTTATTGT
AATATTTGGTGAACAGTGGTCAATAAATAGTTTTATATTCTTTTAAAAAAAAGAAAATGGGTAAATACCTGGAACAT
ACAACCTCCTAAGACTAAACCAGGAAAAAATAGATATCCTAAATAGACCAATAACAAGTAATGAAATTGACTCACTA
ATAAAATAATTTCAAACAAATAAAAAACAAGACAGATAGGTTTACAGCCAAATTTTGCCACAAGTATAATGAAGAGC
TGGTACCAGTCTTACTGAAACTATTCCAAAAATTTATGGAATTTTTGGATTGGAAGAATCAACATCATTTACAAAGAC
CACACTAACATAAGCAATGTACAAATTAAATGCAATTCCTATCAAATTACCAATATCATTTTTTTCACAGAATTAGAA
AAAACGATCCTAAATATCATATGGAACCAAAAAAGCATTCAAATAGCCAAAGTACTCCTAAGCAAAAAGAACTAAGT
CAGAGGCATCATGTGTCTGACTTCAAATTATACTATAATGCTATACTAAAAAAACAGTATGGCATTGGTACAAAAA
CAGACAGATTAATGGAATGGAATAAAGAGTCCATTGCTAGGAAGGATACAGCTTAAAAAGGGAGAAAAAATTGTGA
AAGAAAAGAGAATAAAGAACCAGTAATAAAATCACATGCCTACAACCAAGTGATGTTGACAAAGTTGGCAAAAATA
AACAGTGGGGAAGAGCATTTCTATTCAATAAAATGGTGCTGGGAAAATTGGCTACCCATAGGCCAAAAGAAATGAAACTG
AATTTCCCATTTCTCATCATATACAAACTTTAACTCAAGACGTATTAATGACTTAAATGAAAAGACCAGAAGCTATAAA
AATCCTAGAAGAAAATTTAGGAAAAACTCTTCTAGATATTGGCCTAGGAAAAATAATTCATAACTTAGACCTGAAAAG
GAAATGCAACAAAACCAATATACACAAATGAGACTTAACTAAAGAGCACAGCAAAAAGAAATAATCAACAGAGTAA
ACAGACAACCTACACAATAAGAACAAAATATTTGCAAACTTTGCATCTGACAAAGAACTAATATCCAGAATCTAAAA
AGAACACAAACAACTCAACAACAACAAAGAATAAACAACTCATTAATAAAGACTGCAAAAGGATATGAACAGACATT
TCTCAAAAAGAGACATACAAGTGGCCACCAACACATGAAAAATTGCTCAACATCACCAATCATCAGAGAAATGCAA
ATTAACCACAAATAAAGGCAGGGCGCAGTGGCTCACACTGTAATCCAGCATTGGGAGGCCAAGGCGGGTGGGA
TCACCTGAGGTGAGGTTCAAGACAGCTGACCAATATGGTTAAACCCCATCTCTACCAAAAATACAAAAATTAG
CTGGGCATGGTGGCAGGCACCTGTAATCCCAGCTACTCGGAGGCTGAGGCAGGGGAATCTCTTGAACCCAGGAGGCA
GAGGTTGCAGTGAGATGAGATCATACCACTGAACCTCAGCCTGAGTGACAGGGCGAGACTCTGTCTCAACACACAC
ACACACACACACACACACACACACACACACACAAAAATAAGATATAAACACAACAGATGTTGGTGTGAATGCAGAG
AATAATGAATGCTTATTTACTGTTGGTGGGACCATAACAACCTCTATGAAAAACAGCATAGAGATTTCTCGAAGCAC
TAAAAATAGAGCTACCATTTGGCCAGTGATTATACTACTAGGCATCTGAAAGAAAAGAAAATAATTATATAAAAAAG
TCACTTTCACTCATGTGTATTACATTCCTACTCATAATAGTAAAGTCATGGAATTAACCTAAATGTCCATTAACAGT
GGACTGGATTTATTTTAAATATGGTATATATACACCATGGAATACATATGCAGCCATAAAAAATGAAATCATGCTCTGTG

CAGTAACATGGATGGAGCTGGGGGCCATTATCCTAAGTGAAATAACTCAGAAACAGAAAATCAAATACCACATGTTCTCATAAGCAGGAGCTAAACAATGAGTACACAAGGACATAAAAATGGAACAATAGATAGTGAGACTCCAAAAGGGAGAGACTGGGAGGAGGCCGATGGTCAAAATATTATCTATTGGTAGTGAGATGTGATCAGCCACTGCACACCAGCCAGGCAACAGAGTGAGATCATGTCTCAAAAAAAAAATGATCTGTTGGGTACAATATTTACTATTTTTCAGGATAGTTACTAGAAAGCCAAACCTCACCATTATGCAATATATCTATTTTGACAAATATGCACCTTATACCCTCTGAATCTAAAATAAAAAACAAAGGAATTCTATAGTAGTAAACACCTACATCTTTTTTTTAAAAAAGATCTCGCCTGTGTGGTGGCTCACGCCTGTAATCCCAGCACTTTGTGAGGCGGAGGCAGGCAGATTGCCCAAGGTCAGGAGTTTCGAGAAAGCCTGGCTAACATGGTGAAACCCCATCTTACTAAAAATACAAACAAATTAGCCGGATGTGGTGGCGCACGCCTGTAGTCCCAGCTACTCGGGAAGCTGAGGGAGGAGAATCACTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCTGAGATTGCGCCACTGCCTCCAGCCGGGCAACAGAGAGAGACTCCGTCTCAAAAAAAAAAGATCTCAAATAAACAACTAACTTTACACACACACAAACACACACACAAAATAGCAGAAGGAAAGAAATAACAAAGATTAGAGCATAAATAAATGAAATAGAGAATAAAAAAATTTTAATCGGCCAGGCACGGTGGCTCACACCTGTAATCCCAGCAGATCTCTTGAGGCCAGGAGTTTGAGACCAGCCTGGCTCAACATGGTGAAACCCCTGTCTGTACTAAAAACACAAAAATTAGCCAAGTGTGGTGGCATGCTCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATCACTTGAAGCCGGGAGATAGATCTTACAGTGAGCCGAGATTGTGCCACTGCACCTCCAGCTTGGGCAACAGAGTGAGACTCTGTCTCAAAAAAGAAAAACAAAAAATAATCAATGAAGCTAAGAGTTGATTTGGGGAAAAAACTCAAATCAAAGCCCTTACTTAGACTAAGAACAAAAAGAGAGAATATGTTAATAAAATAAGAAATGAAAGAGAAGACATTACAACCTATTACAGAAGTTTATTTCTTACAGAAATAAACAGAATTATAAGGGACTACTATGAACAATTATACACCAACCAACTTGACAACCTAGAAGAAATGGTGAAATTACTAGAAAATATAAACTACCAAGATGGAATGAAGAATAAATAGAGAGCTTAAAGAGACCAATAACAAATAGAGAACTAAAGAAGTAATTTTTTCAAACTCCAAAATATAGCCTAGGACCAGATTACTTCATGGGTAAATTTCTACCAAACCTTCAAAGAAAAATTAATATAAAACATCCTCAAATTTCTTTGAAAAATAGAAAAAGAGGTAACACCTTCAGAGTCATTTTATGAGGCAGCATCCCTTGATATCAAAGCAAGAAAACACACTTTCAAGAAGAAAATTATAGGCCAATATCCTTGAGGAACATGCTGTATGAATTAAATCTAAATAAAATCTGTTTACAAAGAATCGATTCTCAGCAGAAGCTTTACAAAATTAAAAAATTAATAATTTCTCACACTGGATTCTGGATAAACACTCATTTCTAGCCTCTTTGTGTCAGATATGTTACTGAAACTTACACTAACTGAGGAGTTTCGCTTGAGCAGCAGTTTCTTTCATCTGTTGGGAGCTGATATATCTGGATGCCATTGCTAATCAATTCACTCATTATCTTATTCTTAAACGCTCTGTAAATCATTTTTTAGAAATAGTGTCTGCTTTGGCAATCAGTGGTATAATATTCACCTATGAAGAAAGAGACAAAACAAATCTCATCATTTGTTTCATCTCACAGTTTTTTTTGTTTTTTGTTTTTTGTTTTTGAGACGGAGTCTCGCTTTGTGCGCCAGGCTGGAGTGCAGGGGCACGATCTCGGCTCACGCAAGCTCCGCCTCCTGGGTTACGCCACTCTCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCACCACCATGCCCGGCTACTGCTGCTGAGGATTCACGGCTGAGCCACTGCGCCAGCCATCCTATAGTTTATAGTAAAGCCTTACAATCTCTGACAACAAATTGCATAACTAGGTAATCACTGGCTGTACGCTTTTCTTAAAAATTATTTTCCAAATATCAATACTTTGAAGATTGATCTTTTATCCAAATTCAGGTTTAGAAGTCTCAACCTTGTTTCTTCTTAAAAATCCCTATACCTCTACACATATGGAGTTAGTAAATGTAAATAGTAGGACAATGTAAATATTTTTATTTATTTATTTATTTATTTATTTATTTATTTTTCAGACAGAGTCTGGCTCTGTTGCCAGGCTCTGGAGTGCAGTAGCGGATCTTGCTCACTGCAAGCTCCGCTCCCAGGTTTCATGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCACCACCATGCCCGGCTACTTTTTTGGTATTTTTTAGTAGAGACGGGGTTTACCCTGTTAGCCAGGATATTTCTCGATCTCCTGACCTCGTGATCCGCTCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGCCTGAGCCACCGTGCTGCTATTTAATTTATTTTCAATTTATTTCTTATTAAGCTGAGTGATAAAAGACAAATTCAAAGAATTTCTTCAAATTTGTGGAAGGGATTTGTCTCTTACTTTTTTACATCACTTTTTTCCATCAATCAATCTTAATTAATGTCCCATTTTTCACTCTCTTGCCCCGCCTAGATGTGTTTGAGATAGATTTGCTATTGTTTTTCAAAGACCACCTAAAAAGAATATAGATGTTTTTGATTACCATATTTCTTTCTTTTTTTGAGGCGGAATCTTGCTCTGTCGCCCAGGCTGGAGTGCAATGGCGCAATCTAGGCTCAGGGCAACGTCTGCCTCCCGGGTTCAAGCAAGTCTCCTGCTTCAGCCTCCTGAGTAGCTGGGATTACAGGTGCCCGCCACCATACCCGGTTGATTTTTGCATTTTTTAGTAGAGACAGGGTTTACCATGTTGGCCAGGCTAGTCTCGAATCCTGACCTCAGGATACAACCTTCAGCCTCCCAAAGTGTTGGGATTATAGGCATGAGCCACCACACCTGGCCTGGTTACCATATTTCTAAAGCCGTGTGTATGTATGTGTGTGTGTGTGTATAAGAACACCTTGCAACAAATATGTTTCAATTTTTCTTTGTAAAAAGACATAATTTCCACATTGAAAGTTAAATATTTGCATATTTGAGGTTGAATTTTTTACCCACAAAGTTTTCTTCTAGTGCTTCAATATGAGCACTCTGCTTTTTTGCAATTTTATTAATTAAGTCCACAAAATATAGCTCCACAAAATATCTGTAAATATAGTAGAAATACATAAAATATAAATTTTAGGCAGACTGCTTTTTGTTTTTAGATTGCTCAAAACACGTTTTTAAAAAATCCATGGACAGACCGGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCAAGACTGGCAGATCACCTGGGGCCAGGAGTTCAAGACCAGCCTGACCAACATGAACAAACCCGCTCTTACTAAAAATGCAAAATTAGCCAGGCGTGTTGGCTCATGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCCGAGATTGCGCCATTGCACCTCCAGCCTGGGCAACAAGAGCGAACTCCGTCTCAACAACAACAACAACAACAAAAGAGTGATATTTCTGACTATTGAAGGTACTACTAATTAAGAACTTGAAGCAGCTCTTAATTTAACTAGAAATCTAATTTAATTTCAACTACTTAATGACATCTGAGATTAAGTACAAAGTCTCTTTTCAGTTTGAACAATATATAGTAAAAGTTGAGATTCTCGGATAAAAGGAACCAACTTAGAACTTTTTCTCTCTTCAACTAAATAAAGATGTCTTCAA

[illegible]

CCCTTCCCCTTTTATAAAGGGCCAGCGTCTATTATTTGTCTAATACTGAAACTTTTTTTATGTTTAGATACACAGAC
AAGGCTGGGCGCAGTGGTTTCACGCTTTGGGAGGCTGAGGCGGGCAGATCACTCGAGGTCAGGAGTTCAAGACCAGCC
TGGCCAACATGGCAAGACTCCGTCTCTATTAAAAATACAAAAATTAGGCCGAGCACAGTGGCTCACGCCTGTAATCT
CAGCATTTTTGAGAGGCTGAGGTGGGTGGATCACCCGAGGTCGGGAGTTTCGAGACCAGCTTGACCAACATGGAGAAAC
CCCATCTCCACTAAAAATACAAAAATTAGCCAGGCATGGTGGCACATGCCTGTAATCTCATCTACTCAGAAGGCTGAG
GCAGGAGAATCGCTTGAACCCAGGAGGTGGAGGTTGCAGTGAGCCAGATCACGCCATTGCACTCTGGCCTGGGCAA
CAAGAATGAACTCCGCCAAAAAAGAGGCTGGGAGGCTGGTGGCACAGGCCTGTCAGGCCTA
TAGTCCCAGCTACTTGGGAGGCTGAGGCCAGAGAATCACTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCCAAGATT
GTGCCACTGCACTCCAGCCTGGGAGACAGTAAGACTCCATCTCAAAATAAAATAAAATAGATACACAGATAC
TTACCATTGTGTTATAATTCTACAGTATTCCGTACAGTACCATGCTGTCCAGGTTTGTAGCCTAGGAGCAATAGGCC
ATACCATAAAAGGCTATAGCCTAGATGTGTAGTAGGCTATATCATGTAGGTATGTGTAGATACACTCTAAGATGTTT
GCATAATGATGAAATCACCTAAAGATATGTAATACTACACTTGCATGCAAGTGCCTTCAAGATGTCTTATTATCTCAA
ATTAATATTACATCATGAAAGTTGAGACCAGGAACAATGGCTCATGCCCTGTAATCCCAACAACCTCTGGGAGGGCAA
GGTGGGTGGATCACTTGGAGCCAGAAGTTCAAGTTCAGCCTGGGCAACATAGTGACACCCCTGTCTCTACAAAAATAC
AAAAATTAGCCAATCTTGGTAGCACACTCCTGTCCAGTTGCTCGGATGGCTAAGGTGGGAAGATTGCTTGGAGCCCA
GGAGGTTGAGGCTGCAGTGAGCTATGATGACGCCACTGCACCTAGCCTGGGTGACAGAGCAAGACCCTGCCCTCAA
TACAAAAAAGAGAAAGAGGCTGGGCACGGTGGCTCACACCTGTAATCCCAGCACTCTGGGAGGCCAAGG
CGGGAGGATCACGAGGTCAGGAGATCGAGACCATCTGGCTAACATGGTGAAACCCCATCTCTACTAAAAATACAA
AAAATTAGCCGGGCGTGGTGGCAGGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAAGCTGGAGAATGGCGTGAACC
TGGGAGGCGGAGCTTGCAGTGCACCACTGCACCTCCAGCCTGGGCGACAGAGGGAGACTCCGTCTCAAAAAA
AAAAAGATGGTAGATTTACCAAAACATACCTTACTGTCAAGGTTCTTCAATTGTTAATAGATCAAGAGACTTCAGGGAA
TGTCTGTAGGTGAAATGAAGTAAAGACACACGTGGACGCGAGAATCATGGTACTCAAACAAGGAACGTTTAAATCTT
CAGTTCTTCTTGAAGATAGGCCTCAAATTGGGCATCTATGTAGTCAACTATTGGTTGGTAGCTAAAAAATTTAT
ACATTTAGCATAACAGACTAATAAAATCTAAAGCCTGCTTTTTGACCCTTAATGAGCAGTTAACATGTATGTAAGC
AGTTTGTCTATAGTGATGAAGAAAAAGTTCAATACATGTCCATACAAAGATATTAATTTCTAAACCAATAATAAGGCTC
ATGCTTTACTCTATGGTTATGTAAAATGTAAACATAGAGGAAAGCTGGGTGCATCCGACAGGGGACTAGTATCCAGA
ATATACAAAGAATGCAAGCAACTAAACAACAACAAGAAGAACAATAATCCCATTAAAAAGTGGGCAAAGGACA
TGAATAGACATTTTTTCAAGAACACCAAAATACAAGGCAAACCACAATGAGATATCATCTTATCTCAGTCAGAATGGC
TATTGTTAAAAAGATAAAAAATGGCCGGGCGCGGTGGCTCACGCCCTGTAATCACAGCACTTTGGGAGGCCAAGGTGG
GTGGATCACGAGGTCAGGAGTTCAAGACCATCCTGGCTAACACGGTGAAACCCCTGTCTCTGCTAAAAATACAAAAAT
TATCTGGGTGTGGTGGCAGGCCCTGTAGTCCCAGGTACTCAGGAGGCTGAGGTAGGAGAATGACATGAACCCAGGA
GACAAAGCTTACAGTGAGCCGAGATCGCGCCACTGCGCTTACGCCCTGGGCAACAGAGCAAGACTCCATCTCAAAAAA
AGACAAAAAATAACAGGCCAAGTAGAGTGGCTCATGTCTATAATCCCAGCACTTTGGAAGGCCAAGGTAGGAGGATC
ACTTGAGGCCAGGAGTTTGACACCAGCCTGGGCAACATAGCAAGACCCCATCTATACAAACAATTTTTTAAATTAGC
CAGGCATGGTGGCATGTGCCCTGTAGTCTCAGTTACTCAAGAGGCTGAGGTGGAAGGATTCAGTTGAGCCTGGGAGGT
CGAAGCTGCAGTGAGCCAAGCTTGTGCCACTGCACCTCCAGCCTGGGTGATGGAATGAAAACTGTCTCAAAAAATAA
TAATAAAAAATAATTAAATAACATAATTTGGCAAGGGTACACAGAAATGGGAACCCCTAGACACTGTTGGTGGGAATG
TAAATTAGTACAACGCTTACGGAATAATTTGTACAAAATAAAGTTAAATAAAGAAATAAGGAAATAAGCTAAACTAAT
AGAAGCTATATTTAAATTTTTTGTCTATAACTGGCTAAATAAAATCTTGTAAAGAAAGTCAATTATATACAAAAGTA
TTAATCTTATAAGAAAAGAACATTACTTAGGAAAATATTTAAATCTGGAAGAAAAGGTATTACTTACACAAATTAGT
CTATGGCAAAGGAAGTCAAATTCCTTGAAATAAGAAAACCTTTTTTTAATTTAAATAAGAAAACACTCACCTGGCTTC
TTTGTCTATTTGATCACCATAACCCTACTGTCTCCACAACAGTCAATTTCAACTGAACATTGCTTTCTGAAAGTTCAAT
ATGTCTGAATTTGAAGTCCAACATTTGAGTAAAAATGTGAGGATTTGTTATCTTTCAAGTTAGTATTAAACAATGTG
TCTATCAGTGTGATTTTCCAATTCAGTCTCCCTGTAAATAGACATAGAGTCATTGTTCATAGGGACATAAAGGGGC
TAAGAAAAAGCAACACTTTGAATTCCTACGGTGGTTTTTAAATCTTAGAAAGTTGAACTCATAGAAGTAGAGAGTA
GTATTCTGGCTACCAGAGGCTGGGGGAGCAGGAGAAAGGAAAGGTGGAGTAGTTGGTCAAAGGTACAAAGTTTCTTT
TTTTTTTTTTTACTTTTATTTTAGGTTGGGAGTACACGTGCAGGTTTGTATATCAGTAAACTCATGACATGAGGGT
TTGTTGTACAGATTATTTATCACCAGGTAGTACATAGCATGGTAACCAATAGTTATTTTATGTATCTCATCTCCTCCTCCT
CCAAACCTCCACCTCAAGTAGACACCAGTGTCTGTATTCCCTTTTTGTGTCCATGTGTCTCATCTTGTAGCTC
CCACTTACAAGTGAGAACATGGCACATGTTCTCAGCCATAAAAAAGAGGAGATCATGTCTTTGCAAGAACATGGA
TGGAGCTGGAGGCCATCATCTTAGCAAAACCATCAGAGCTCATTGTAGCGCTATTACAGTAGCAAAAGACATGGAAT
CAACCTAAATGTTTCATCAATGGTAGACTGGATAAAGAAAAATAGGTACATATACACCATGGAATACTACACAGCCAT
GAAAAAGAATGAGATCATGTCTTTGCAGGAACATGGATGGAGCTAGAGGCCATCATCTTAGCAAAAGTTTCAA
CAGGAGTAGAAATAAGTTTTGAGATCTATTGCACAGCAGAGTGACTATAGTCAATGTATTGTATATTTCAAATAAT
CAAGAGAGTTAATTTCAAATGTCTCATATAAAAAAATGATAAGTAAACCGGTAATAGATACATTAATTCACTTGA
TTGAATCATTCCAAATGTAAACATATAGCAAAACCTCACATTGTATCCATACATGTAATAACTATGATTTTTCAA
TAAAAATAATATTAATTTTAAAAATAAATAAAGGGTCTCAAATACTTCAAAGACTGGGATGGCCTTGAATAATAC

[illegible]

TGTATTTTATGAAACCAGATATAATGGCAATATAAACACCTGACAGGCCAGGGGCGGTGGCTCATGCCTGTATCCTA
GCATTCTGTGAGGTTGAGGTGGGTGGATCACCTGAGGTCAGGAGTTCAAGACAACCTGACCAACATGGTGAAACTC
CATCTCTACTAAACACAAAAAATTAGCCGGGCATGGTGGCACGTGCCTGTAATCCCAGCTACCTGGGAGGCTAATGC
AGGAGAATCCCTTGAACCCAGGAGGCAGAGGTTGCAGTGAGCTGAGATTACACCATGCCTCCAGCCTGGGCAACAG
AGTGAGACTCCATCTCAAAAAAACAAACAAACGAACAAAAACACCTGGCAAAGACATCACAAAAAGAGAAAAATTAT
ACATTAATTTCTGTGCATGTGCATAGCTGCAAAAAATCTTTTTTACTACTATAATTTTAATTGTACAACCTAGTAAGTG
TGCAAATAAATTTTCTTAACAACCTGAAATATTGAAATGGAAGATAATATTCATTTAACTATTCCCCACATCCTAT
CTCCTTCTCTTTTTCCCTATTTTTTATGATTTTTTAGGATTCTAGACATTAAAAAATACTTTTATATACACAAATATAG
AGGTATTAAATTTTAAAAACACATACATGGCCAGGTACAGTGGCTCACGCCGTGAATCGCAGCACTTTGGGAGGTTG
AGGTGGGCAGATCACCAGAGGTGAGGAGTTCAAGACCAGTCTGGCCAGCGTGGTGAACCCGGTCTCTACTAAAAAT
ACAAAAAATAAATTAGCCAGTCATGGTGGCGGAGGCCTGTAACCCAGCTACTCAGGAGGCTAAGAGAGGAGAATT
GATTGAACCCGGGAGGCAGAAGTTGCAGTAAGCAAGATCGCGCCATTGCACGCCAGCCTGGGCCACAAGAGTGAAA
CTCTTGCTCTCAAAAAAAGCAAAACATTCTGCAACTTGCTTTTTTCTACTAAAATTCTATACATGTTAATCAGCT
AGCTTTATTGAGGAATAATTTATATGTAATAAAATTCACCAACTTTATGTGCACAATTTGATAAAATTTTAACAAATG
TGTATAGTAATGTAATCACCATCATGATCACCATATAGAACATTTCTGACATCCCTAAAACCTTTCCCTATTAGCCCC
TATTCAGTCAATTTGTAACCCCTCATCCCTCAGCCCTGGCAACCTCTCATCTGTTTTTCTAGTTGCTATAGTTTTACTTTT
TTCTAGAATTTCAATTTAGAAGAAATCATACTGCATGGAGTTTTTTTTTTTTTATCTTAACAAAAACATTACCAATCAATT
CACTTGATATATAAAAAAGATTATAGATTAAAACAATTGGCATTTATTTCTAGAAATGCAAGGTTAATTCAGCACTCAA
AAATCAATAAATGATAAACTATAAAATACTGATGCAAGAAATGAAAAGGACACAAAAAATGAAAAGATATTCCAC
GTGTATGGACTGAAAAGAAATCAATATCATTAAATGTTTCATACCACCGAAAGCAATCTACATATTTAACTCAATCCCT
ATCAAAATACCAATGACATTCTCCACAGAAATAGAAAAAATAGGCCGGGCGCGGTGGCTCACACCTGTAATCCCAGC
ACTTTGGGAGGCCGAGGTGGGTGGATCATGAGGTCAGGAGATCAAGACCAACCTGGCTAACACGGTGAAACCTGTCT
TCTACTAAAAATACAAAAAAGAGGCTGGGCGCAGTGGCTCACGCCGTGAATCCCAGCACTTTGGGAGGCTGAGGCGG
GTGGATCATGAGGTCAGGAGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATACAAAAA
TTAGCCAGGTGTGGTGGTGGGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGACGTGAACCTGAG
AGGCAGAGGTTGCAGTGAGCCGAGATTGCGCCACTGCACTCCAGCCTGGGCAACAGAGTGAGACTCTGTCAAAAAA
AAAAAAGAAGAAGAAAAAGAAAAAGAAATAGAAAAAATAATCCTAAAATTTATACGTACCCACAAAAGACACAGAA
TATTCAAAGCAATGCTGAGTACTCTGGGTACACTGAAGGAATCACATTACTTGACTTCAAATTATAATACAGAACTA
CAATAACCAAAACAGCATGGTACTGGCATGAAAACAGACATATAGACCAATGGAACAGAATACAGAACTCAGAAATA
AAATTCAGACATGTACAGTGAACTCATTTTTGACGAAGGTGCTAATAGCATAACAGTGGGGAAAGAAGAGTCTCTTCAA
TAAATTGTGCTGGGAAAATTTGGATATCCATATGTAAAAGAATGAAAACAGACTCCTATCTTTATCATATATAAAAA
TCAATAAAAAATGAATTAAGACATAAATCTAAGCTGTGAAACTATGAACCTACTAAAAGAAAAAGATTGGGGAAACT
TTCCAGGACATTGACCTGGGCAATATTTCTTGGATAATGCCCCACAAGCACAGGAAACCAAGCAAAAAATGGACAA
ATGGGATCATATGAAATTAAGGCTTCTACACAGCAAAAGAAAGTGATCAACAAAGTGAGGAGACAACCCACTGAAT
GGGAGAAAATATTTGCAAACTATCCCCCGATAAGGGATTAATAACCAGAATATGTAAGAAGCTCAAAATACTCAATA
GGAAAAATCTAATAATCCAATTTAAAAAGTGGCAAGTGATTGGAATACACTTGTCTCAAAAGAAGACATACAAATG
GCAACAGGTATATAAAAAAGGTGCTCAACATCATTGACCATCAGAGAAATTTCCAATCAAACCTACAATGATATATCA
TCTCACCTGAGTTAAAGTGGTTTTTACC AAAAGACTGGCAATAACAAATGCTGGCGAAAATGTGGAAAAAAGAGA
ACCCTCTTACACTGTTGGTGCAAATGTGAATTAGTATGGAGAACGTATGGAGGTACCTCAAAAACTAAAAATAGAA
CTACCTATGATCCAGCAATCCCAGTCTAGGTATATACCCAAAAGAAAGAAATTAGTACATTGAAGAGATATCTACA
CTCCCATGCTTATTGTCAGCACTATTACACTAGGCAAGATTTGGAAGCAACCTAAGTGTCCATCAACAGATGAATGG
ACCAAGAAAATGTGGTACATATACACAATGGAAAACCTATTCAATCATCAAAAAGAAATGAGATCTTGTTATTTCCAAC
AACATGGATGGAACCTAACATTATGTTAAGTGAATAAGCAAGTCATGGAAAGACAACTTCACATGTTCTCACTTA
CTTGTGGGAGCTAAAAATTAACAATTTGAACCTCATAGAGATAGAGAGTGGAAGGATGGTTACCAGAGGCTAAGAAG
GGTAGTGGGGGTGGTGAGGGGATGGTTAATGGGTACAAGGATATAGTTAGATAGAATGAATAGGGCCAGTATTTGG
TAGCACAACAGGGCGACTATAGTCAACAATAAATTGTACATTTTAGAAGAATGAAGAGTACAATTGAAATGTTCTA
ACACAAAGAAATGAAAAAATACTTAAGGTAGTGGATACTCCATTATCTTGATATGATTATTACACATTGTGCGCCTA
TAACAAAATATCTCATATGCCCCATAAATATATAGACCTGCTATGTACCCATAAAAAATTAAAAAGTAAGGGCTGAGC
GTGGTGGCTCATGCCTGTAAATCCCAGAAATTTGGGAGGCTGAGGCTGGAAGATCGCGTGAGCTCAGGAGTTTCGAGAC
CAATCTGGGCAACATAGTGAAACCCGTGTCTCTACTGAAAATAAAAACTAAAAAATAGCAGAGTATGGTGGTACATG
CTCGTCTGCTCCAGCTACTCTGGAGGCTGAGGCGGGAGAATCACTTAAGCCCAGGAGTTAGAGTTTGCAGTGAGCTAT
GATTGTGCCACTGCGCTGCAGCCTGGGTAACCGAGTAAACCCGTGTCTCAAAAAAATAAATTTAAAAATCAATAAA
CATATTTACCATATTAATAGTATAAAGTAAATCACCATATGACTGACTCAATAAGTGAATAATATTATATTTGAT
GAATTGAATTGAATACCCATTTGTGATTTTTTAAAGTCTATTAGTAAAAATAAGAAATTAAGAGAACTTCATCAATCTG
ATAAAACTGTATTAGAAAAAATTTAAATCTAACATCATACTTAATAGTAAACATCTGAGTGCTTTCTTCACAAGA
TTGAAAATAAAGCAGAGATGTCAGCTGTTGCCATTTTATACATTGTACTGGAGGTTCTTTTTTTTTTTGAGACGGAGT
CTCGCTCTGTGCGCCAGGTTGGAGTGCAGTAGTGCATCTCGGCTCACTGCAAGCTCCGCTCCCGGGTTACAGCTA

TTCTCCTGCGCTCAGCCTCCTGAGCAGCTGGGACCATAGGCACCCGCCACCACGCCCCGGCTAATTTTTTGTATTTTTTA
GTAGAGATGGGTTCACCATGTTAGCCAGGATGGTCTCAATCTCCTGACCTTGTGATCCGCCCCGCTCGGCCCTCCCA
AAGTGCTGGGATTACAGGCGTGAGCTACCACACCCGGCCTGTACTGGAGGTTCTTACCAGTCAATAAGGCAAACAAA
AAGCATAAATATTCAAAGAGTTTGTCAATTTATATTGAAAAAGGTAGAAAATATACAAGAATCTACAGAGCATCT
ACCAGAACTACTAAGAGAACATATCAAAATTACAAGATACAAAGTCAATATACAAAAATCAGTTATATTTCATTTCTA
GCTTCTGCTCAGAGATGTGGAGAGCAGAAAAAGACATCACTACTAATACTATAGGAATAAAATAATGAACATAAGGC
AAGATCACAACTTTTTTTTTTTTTTTTTTTTGTGAGACAGAGTCTCGCTCTTTTCGCCCAGGCCGGAGTGCAGTGGCACTA
TCTCGGCTCACTGCAAGCTCTGCCTCCCAGGTTACGCTATTCTCCTCTCTCAGCCTCCCAAGTAGCTGGGACTACA
GGCGCCCCGCCACCGTGCCAGCTAATTTTTTGTATTTTTTAGTAGAGGTGGGGTTTCACCGTGTTAGCCAGGATGGTCT
TCGAGCTCCTGACCTCATGATCCACCCGCTTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACGGCGCCCCG
GTCAGATCACAACTTTTTAAAACTCATCCTAGTACTGACGTTTCAGGTCGACCAATTGTTCCAGAATCGAAGGAGA
GACAGAAACCCACAGGAAAAGACAAGATGGGAGCATCAACTTTTCTAGGCAAGAAAGACGCAGACATCAGTGAGAAG
AGTCTAAAGCAAGGAGCAAATTGCTAGAACATGCCCTTGCCCTGACAAGTAAAAAGTATGAGCCATAAAGAGGAATCA
TCTGTAGGTGCTTTTTTCTTTTCTTTTCTTTTCTTTTGTAGACAGAGTGTGCTCTATTGACCAGGCTGGAGTGCAG
TGGTGCAGTCTCGGCTCACTGCAACTTCCGCTCCCTGGTTCAAGCAATTCTCCTGACTCAACGTCCTGAGTAGCTG
GGTTATAGCCGTGCCACCACGCCCACTAATTTTTTGTATTTTTTAGTAGAGACGGAGTTTCACTATGTTGGCCAGGT
TGGTCTTGAACGCCCTCACCTCGTGATCCACCCACCTCGCCCTCCCAAAGTGCTGGGATTACAGGCGTGAACATAACAC
GCCCAGGCCTGTAGCTGCTTTTTTTTTTGTGTTGTTTCTGGTTTGGTTTTTATTTGTCCCTTATTTGGAAATCTGTCATT
CTAAGTAGAGAATACAAGGTATATTAATTAATTTTTTAAATTTATTAAGGTTTGATTTAATGTATGTATTTAATTTTT
TATTTGTAATTTCAACTTTTAGATTTCAGAGAGTACATGTGCAGGTTTGTATGAGGAAATACTGAATGTGCTGAGTT
GGGGGTGTGAATGATCCCATCACCCAGGTAAGTACTGAGCATAGTACTCAATAGGTAGATTTTCTGCCCTCACCCCTCCC
TCCCTCACCTCTAGTAGTCTCAGTGTCCATTGTTGCCATCTATGGCATGAGTCTCATACTCATATGTCTATGAGT
ACCCAATGTTTACCCCCACTTATAAATGAGAACATGTGGTATTTGGTTTTCTGTTTCAATGATTAATTAACCTTAGGAT
AATGGCCTCCAGCAGCACCCATGTTGCTGCAAAGGACACATTTCAATTTTTTATAGCTGCATAGTATTCATGGTGTA
TATGTACCATTTTTTCGTTTTTTTTTTTGATATGGAATCTTGCTCTGTCAACCCAGGCTGGAGTGCAGTGGTGTGATTT
CAGCTCACTGCAATCTCCGCCTCCCAGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAATAGCTGGGATCACAGGT
GCCCACCACCACCTGGCTAATTTTTTGTATTTTTTAGTAGAGACAAGATTTTACCATGTTGGCCAGCCTGGTCTCA
AACTCCTGACCTCAGGTGATCCGCCTGTCTTGGCCTCCGAAAGTGCTGGGGTTACAGGCATGAGCCACCGCGCTTGG
CCATGTACCACATTTTTTAATCCAATCTGCCACTGATGGACACCTAGGTTGATTCCATGTCTCAGCTATTGTGAATA
GCGCTGAGATGAAAACAAGTGACAGGCCAGGTGTGGTGGCTCACACCCATAATCCCAACACTTCAGGAGGCTGAGGCA
GGCAGATCACTTGAGGTGAGGAGTTTGAGAGCAGCTTGGCCAACATGGTGAAACCCCATCCCTACTAAAATTACAAA
CATTAGCTGGGCGTGGTGCAATCTCAGCTACTCAGGAGGCTGAGGCACGAGAATCGGTTGAACCCGGGA
AGTGGAGGTTCCAGTAAGCCAAGATTGTGCCACTGCACCTCCGCCTGGGTGACAGAGCGAGACTCTATCTCAAATTT
AAAAAAAAGAAAACAGGATGGGCGTGGTGGCTTACACCTGTATTCCAGCACTTTGAGAGGCTGAGGCAGGCGGATC
ACCTGAGGTGCGGAGTTTGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTAATAAAAAATACAAAATTAGCCA
GGTGTGGTGGCGCATGCCTGTAATCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCAGGAGGCAGA
GGCTGCAGTGAGCCAAGATTGAGCCATTGAACTCCAGCCTGGGCAACAAGAGGAAAACTCTGTCTCGAAAAGAAAAG
AAAAGGAAAAAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGGAAAGGAAGGGAAGGGG
AGGGGAGGGGAGAGGAAGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGG
GGAAGAGAGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGG
AGTCTCGCTGTGCGCCAGGCTGGAGTGCAGTGGCACGATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTACAGCC
ATTCTCCTGCCTCAGCCTCCCGAGCAGCTGGGACTACAGGCGCCCGCCACCACGCCCCGGCTAATTTTTTGCATTTTT
AGTAGAGACGGGGTTTACCGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTCGTGATCCGCCCCGCTCGGCCCTCC
CAAAGTGCTGGGATTACAGGCGCGAGCCACCGCGCCCGGCCACCTGTGTCTTTTTTGTAGAACAAAAATAATCTGT
AGCTCCTT

Figure 18

>gi|68978189|gb|AADB02010625.1| Homo sapiens chromosome 7 CRA_219000002701389.
whole genome shotgun sequence

```

caatggtgca atctcggctc actgcaacct ccgcctcctg ggttcaagca 50
attctcctgc ctccacctcc tgagtagctg aaactacagg cacgcaccac 100
cacacccggc ttatttttgt attttttagta gagatgggtg ttcaccatgt 150
tggccagcct ggtcttgaac tcctgacctc gtgatctgcc ggcctcggcc 200
tcccaaagtg ctgggattac aggcattgagc caccacactc ggccaccatt 250
ttattttcaa ctccctttca tggaagaacg tttagccttt ggtctctttc 300
ttgatttatg acagctcgcg gctttcaaga aaactaccta tgaataggct 350
gtgtaacttt tatttattta tttatttttt gagatggagt ctctgtcttc 400
taggctgggg tgacgtggca tgatctcggc tcaactgcaac ctccccctcc 450
caggttccag caattctcct gcctcagctt cccaagtagc tgggattaca 500
ggcatgcacc accacacctg gctaattttt gcattttttt tttattagag 550
atgggggttt gtcatgttgg ccaggctggg ctggaactcc tggcttcaga 600
tgagctgccc acctcagctt cccaaagttt tggcattata ggcatgagtc 650
actgcaccca gcttggtaat gttattttca agcacacctt caaaagttaa 700
ttccaaggcc ttgctctcat agcagcaaag cctgtctgtt gcatagtggg 750
gctcttggtt gcatcttctt tttgggtggg gttgtgtaga cccggggcag 800
taaggcatgt taacctcgag gacattggag ctggctggcc tgactgttgg 850
gtctccctcc taggacatgg cgagccatgg gtggggcagt gtctttcaaa 900
gctgctctca cggagcactt agccccaatc cagccctcca gggagcaggt 950
gcagagactc atagcagagc accccccaca cctgaccccc agcataagta 1000
agaggagccg ctgctccagg tgtatttcag caccagtgtt ggggagcatg 1050
tcctcctgag agcatctaac gattgcttta gagagccctc tgggtgggtt 1100
atattatttt attttattt ttttaatttt ttattttaaa aatatttatt 1150
tatttatttt tttttattt tttattttta tttattgagg cagagtctcg 1200
ctctgtcgcc cagactggag tgcaatggta cgatctcggc tcaactgcaac 1250
ttctgcctcc cgggttcaag tgattctcct gcctcagcct cctgagtagt 1300
tgggatttca ggcacccacc atgccaggct agtttttgta tttttagtag 1350
agacaggatt tcaccatggt gtttaggctg gtctcgaact cctgacctca 1400
ggtggtctgc cgcctcaga ctcccaaagt gctgggatca caggcgtgag 1450
ccattgcgcc tggcaaaatt ttatttattt attttttttg agacaaagtc 1500
tcaactctgt gccagggctg gagtgcagtg gtgcaatctc aactcactgc 1550
aacttttgcc tcccagggtc acacgattct cctgcctcag cctcaggagt 1600
agctgggatt acaggcgcct gtcaccacac ctggctaatt tttgtatttt 1650
tagtagagat gggtttcgcc atggttgcca ggctggcttc gaactcttca 1700
tctcaagtga tccacctgcc tcggcctccc aacgtgctga gattacaggt 1750
gtgagtcacc gcacctggct tattattttt ttttttagat tcagagtctc 1800
attctgttgc ccaggttgga gtacagtggg ataactcatgg cccactgcag 1850
cctcgaactc ctgggggtcaa gtagccctcc caccttagcc tccttagtag 1900
tagggactgC AGGTGAATGC CACCGTGCCC AGCTAATTAA AGAATTTTTT 1950
TTtttttCTG AAAGTCTTGC TTTGTACCCC AGGCTGGAGT GCAGTGGCAT 2000
GAACACAGCT CACTGCAGCA TCAATCTTCT GGGCTCAAGG GATCCTCTG 2050
CCTCTGCTTC CTAAGTAGCT GGGACTACAG ACATAGGCCA CCATGCCAAT 2100
TTTTTTTTTT TTTTTTGAGA cGGAGTCTCG CTCTGTTGCC CAGGCTGGAG 2150
TGCAGTGGCG CCATCTTGCG TCACTGCAAG CTCCACCTCC TGGGTTCACG 2200
CCATTCTCTT CCTCAGCCT CCCGAGTAGC TGGGACTACA GGCGCCTGCC 2250
ACCACACCTG ACTAATTTTT TGTATTTTTA ATAGAGACGG GGTTCACCA 2300
TGTTAGCCAG GATGGTCTCA ATCTCCTGAC CTGTGATCC ACCTGCCTCA 2350
GCCTCTCAAA GTGCTGGGAT TACAGGCATG AGTCACCGCG CCTGGCCAC 2400
CATGCCTAAT TTTAAAAAAA AAAATTTTTT TATAGACATA GCATCTCACT 2450
ACGTTGCCCA GGCTGGTCTT GAACTTGTGG GCTTAAGTGA CCAATTTTCA 2500
CAACTCAGCC TCCCAAAGTG TTGGAATTAT AGGTTTGAGA CACTATACCC 2550
AGCCTTTAAG AATCTTCTTG TAGAGATGGA ATCTTGCTGT GTTGCCCAGG 2600
CTGTTCTCAA GCTCCTTGCC TCAAGTGATT TTCTACCTTG GCCTCCCAA 2650
GTGCTGGAAT TAAAGGCATG GACTGTGTGG TTCTTCTGGC ACTTACACGT 2700
GGTCTTGCTT GGCCGGTTGT CTGGTCCTGT CTGTTTCTGC CTTTCCTCTT 2750
TCTCCAGGGA AAACCTAAGC TTTTCCTTTT GTCTCATCT TGTGTTTTTC 2800
TGGGTCCATG GGCAGAGTAG AGTTCTAGAA TGGTTTCTTA AAGCAGCCAA 2850
GCCCTACCCA TTGATTCTA AGTACATTTA GAAAAACACA TATAAGGCTG 2900

```

GGCATGGTGG	CTCACGCCTG	TAATCCCAGC	ACTTTGGGAG	GCCGAAGTGG	2950
ATGGATCACC	AGGTCAGGAG	TTCGAGACCA	GGCTAACCAA	CATGATGAAA	3000
CCCTGTCTCT	ACTAAAAATA	CAAAAAATTAA	GCCGGGCACG	GTAGCTCACG	3050
TCTGTAATCC	CAGCACTTTG	GGAGGCCAAG	ATGGGTGTAT	CACCTGAGGT	3100
CAGGAGTTTC	AGACCAGCCT	GGCCAACATG	GTGAAACCCA	GTCTCTACTA	3150
AAAAATACAAA	AAAAGTAGCT	GGGCGTGGTG	ATAGGCGCCT	GTAATCCCAG	3200
CTACCCAGGA	GGCTGAGGGA	GGAAAATCAC	TGGAACCCGG	GAGGCAGAGG	3250
TTGCTGTGAG	CCGAGATCAT	GCCAGTGCAC	TCCAGCTAGG	GCAACAGAGC	3300
AAGATTCCAT	CTCAAAAAAA	AAAAAAAAAAAA	AAAATTAGCC	GGGCATTGTG	3350
GCATGGGCCT	ATAATACTCA	GGAGGCTGAG	GCAGGAGAAG	CGCTTGAACC	3400
CGGGAGGTAG	AGGTTGCAGT	GAGCCAAGAT	TGCAGCCTTG	CACTCCAGCC	3450
TGGGTGACAA	GAGTGAAACT	CTGTCTCAAA	AAATAAAGAG	AAAAACATAT	3500
AGAAAACATT	AACACCCAG	GCAGTATACC	TTGTCAAACA	TACCTCAGGC	3550
AAATGCATT	AGGAGAAGAA	AATACATCTT	ATTTCCCTCT	TCATGTTTCG	3600
TTTTTTTTTT	TTGTTTTCTT	TTGTATTACT	CAGTGTTGGG	TATTTTGTAT	3650
TTTATTTTGC	AGGGAGCTGG	TAAGTCGCCT	ACAGGAGCGA	AATGTTTCAGG	3700
TTTTTCCTAAT	ATCTGGTGGC	TTTAGGAGTA	TTATAGAGCA	TGTTGCTTCA	3750
AAGCTCAATA	TCCCAGCAAC	CAATGTATTT	GCCAGTAGGC	TGAAATTCTA	3800
CCTTAATGGT	AAGATGTTAA	CGGTAACATG	TTCCCTTTCT	TAGCAGTTCC	3850
ATTATTCAGT	ATTCTGGGTA	ATGTCTTTTG	GAATGCAACT	TGAACAGTCA	3900
CATCAGAGTT	AAATATTGAG	ATGAATGGTT	CTCTTTTTTT	TTTTTTTTTT	3950
TTTTTTTGAGA	CGGAGTCTCG	CTCTGTGCGC	CAGGCTGGAG	TGCAGTGGCG	4000

Human chromosome 7 match.

attatttcagc	cacagaagaa	atgaaatact	gatacacact	atgacataaa	55798187
ggaacttttga	aaacattgtg	ctaagaggga	aaaaaaaagcc	agtgcgtcta	55798237
CAGGTGAATG	CCACCGTGCC	CAGCTAATTA	AAGAATTTTT	TTTCTGAAAG	55798287
TCTTGCTTTG	TCACCCAGGC	TGGAGTGCAG	TGGCATGAAC	ACAGCTCACT	55798337
GCAGCATCAA	TCTTCTGGGC	TCAAGGGATC	CTCCTGCCTC	TGCTTCCTAA	55798387
GTAGCTGGGA	CTACAGACAT	AGGCCACCAT	GCCAATTTTT	TTTTTTTTTT	55798437

Figure 19

(1)

ATTATTCAGC CACAGAAGAA ATGAAATACT GATACACACT ATGACATAAA 55798187
 GGAAC~~TTT~~GA AAACATTGTG CTAAGAGGGa AAAAAAGCC Agtgcgtcta 55798237

(2)

>gnl|ti|1656600323 name:104440205 mate:1656600324

GTGGGCTCAGCCATCCTCCCAACTGAGCCTCCTGAGTAGCTGGGACTACAGGTGTGAGCCATCACACTCAACTGGTGT
 AGCTATTTTAGAAGACAACTGGCAGTTTCTCAAAAGGC~~TAAACATACAGTCATCATATAATGCAACAATTTCACTCC~~
 TAGGCATATATCCCAGAGAAATAGAAATATATGTCCACACAAAACTTGTACAGCAATCTTCATAGCAGCATTGTTCA
 TAATAGCCAAATACGTGGAAACAACCCAAATGTCCATCAACTGATGAACAGATAAA~~CAAAATGCAGTGTGTCTCTACCA~~
 TGGAAATATTATTCAGCCACAGAAGAAATGAAATACTGATACACACTATGACATAAAGGAAC~~TTT~~GAAAAACATTGTGCT
 AAGAGGGAAAAAAGCCA~~CAAAAGATCACATATTGTACAATTCTACTTGTCCAGATTAGGAAATTCTATAGTGACAA~~
 AAAATGAATCAATGGTTGCCTTAGGCTGGGGGGCCAAGGCAGGTAGGGGGGAGTAGGAGGTAGTGGCTAAGAGGTGCG
 GATTTCTCTATAGGGTAATGAAAGGTTCTAA~~AACTGACTGTGGTGATCGATGCACAGCTCTATGAATATTCTAA~~AAC
 TACTGAATTTGCAGATTTTCAATAAAATAAATCAATTGAAGGGTATGTGAATCATATTTTA

(3)

GTGGGCTCaA	GCCATCCTCC	CAACTGAGCC	TCCTGAGTAG	CTGGGACTAC	55797868
AGGTGTGAGC	CATCACACTC	AACTGGTGTA	GCTATTTTAG	AAGACAAACT	55797918
GGCAGTTTCT	CAAAAGGCTA	AACATACAGT	CATCATATAA	TGCAACAATT	55797968
TCACTCCTAG	GCATATATCC	CAGAGAAATA	GAAATATATG	TCCACACAAA	55798018
AACTTGATACA	GCAATCTTCA	TAGCAGCATT	GTTCgTAATA	GCCAATACGT	55798068
GGAAACAACC	CAAATGTCCA	TCAACTGATG	AACAGATAAA	CAAAATGCAG	55798118
TGTGTCTCTA	CCATGGAATA	TTATTCAGCC	ACAGAAGAAA	TGAAATACTG	55798168
ATACACACTA	TGACATAAAG	GAAC TTT GAA	AACATTGTGC	TAAGAGGGGA	55798218
AAAAAAGCCA	gtgcgtctac	aggtgaatgc	caccgtgccc	agctaattaa	55798268
agaatttttt	ttctgaaagt	cttgctttgt	caccaggct	ggagtgcagt	55798318
ggcatgaaca					

(4)

GTGGGCTCaA	GCCATCCTCC	CAACTGAGCC	TCCTGAGTAG	CTGGGACTAC	56708591
AGGTGTGAGC	CATCACACcC	AACTGGTGTA	GCTATTTTAG	AAGACAAACT	56708641
GGCAGTTTCT	CAAAAGGCTA	AACATACAGT	CATCATATAA	TGCAACAATT	56708691
TCACTCCTAG	GCATATATCC	CAGAGAAATA	GAAATATATG	TCCACACAAA	56708741
AACTTaTACA	GCAATCTTCA	TAGCAGCATT	aTTCATAATA	GCCAATACGT	56708791
GGAAACAACC	CAAATGTCCA	TCAACTGATG	AACAGATAAA	CAAAATGCAG	56708841
TGTGTCTCTA	CCATGGAATA	TTATTCAGCC	ACAGAAGAAA	TGAAATACTG	56708891
ATACACACTA	TGACATAAAG	GAAC TTT GAA	AACATTGTGC	TAAGAGGGgA	56708941
AAAAAAGCCA	CAAAAGATCA	CATATTGTAC	AATTCTACTT	GTCCAGATTA	56708991
GGAAATTCTA	TAGTGACAAA	AAATGAATCA	ATGGTTGCCT	aAGGCTGGGG	56709041
GGCCAAGGCA	GGTAGGGGGG	AGTAGGAGGT	AGTGGCTAAG	AGGTGCGGAT	56709091
TTCTCTATAG	GGTAATGAAA	GGTTCTAAAA	CTGACTGTGG	TGATCGATGC	56709141
ACAGCTCTAT	GAATATTCTA	AAACCTACTG	AATTGCAGAT	TTCAATAAAT	56709191
AAATCAATTG	AAGGGTATGT	GAATCATATT	TTA		

Figure 20

Assembled human genome sequence (query sequence):

ATTATTCAGC	CACAGAAGAA	ATGAAATACT	GATACACACT	ATGACATAAA	55798187
GGAAC TTTGA	AAACATTGTG	CTAAGAGGGa	AAAAAAAGCC	Agtgcgtcta	55798237
CAGGTGaATG	CCACCgTGCC	CAGCTAATTA	AAGAATTTT	TTTCTGAAAG	55798287
TCTTGCTTTG	TCACCCAGGC	TGGAGTGCAG	TGGCATGAAC	ACAGCTCaCT	55798337

Chimpanzee chr7 (56708861-56761823):

ATTATTCAGC	CACAGAAGAA	ATGAAATACT	GATACACACT	ATGACATAAA	56708910
GGAAC TTTGA	AAACATTGTG	CTAAGAGGGg	AAAAAAAGCC	Acaaaagatc	56708960
acataattgta	caattctact	tgtccagatt	aggaaattct	atagtgacaa	56709010
aaaatgaatc	aatggttgcc	taaggctggg	gggccaaggc	aggtaggggg	56709060
gagtaggagg	tagtggctaa	gaggtgcgga	tttctctata	gggtaatgaa	56709110
aggttctaaa	actgactgtg	gtgatcgatg	cacagctcta	tgaatattct	56709160
aaaacctact	gaattgcaga	tttcaataaa	taaatcaatt	gaagggtatg	56709210
tgaatcatat	tttaataaag	ctattattta	aaataattag	tggctgggca	56709260
cagatgcctg	cctgtaatcc	cagcactttg	ggaggctgag	gcaggaggat	56709310
cacttgaggt	caggagtttg	agcctagtct	gagcaacctc	gcaagatccc	56709360
atctctatga	taaaaaataa	aaaaattagc	tgggcatggg	ggcacacgtc	56709410
catagtccca	actacttggg	agactgagga	ttgcttaagc	ccaggagttt	56709460
gaggctacag	tgaaccatga	tcattgtcact	gtactgcagc	ctgggtgaca	56709510
gagcaagacc	ctgtctctaa	aaaggaaaga	aaatacaagt	ttttaatcac	56709560
tttgtgagtg	tagccaagtt	tgaggagaaa	tagacaagag	taaaagagca	56709610
ctgaataacg	agagtaagtg	gctgggttagg	cccagttgct	agctaagtgg	56709660
cttctgaaac	attcattagt	aaagttacag	ctctggggaca	gtcatgcagt	56709710
caaagaatga	atgctaaatt	cattacaaat	gctgatggtc	tttctttaca	56709760
tgccttctag	tgaaaaattc	ctaagtgcct	aaatagcaag	tctgcaatga	56709810
tagcagctgt	ttattaacga	ctacaaaaaa	gaaatggagg	ccgggcatgg	56709860
tggctcatgc	ctgtaatccc	tgcacttttg	gaggctgagg	tgggcagatc	56709910
acctgaggtc	aggagtttga	gaccagcctg	gccaacacgg	tgaacccta	56709960
tctctactaa	aaatacaaaa	attagatggg	tgtggtggcg	ggcgccctgta	56710010
atcccagcta	ctcaggaggc	tgaggcagga	gaattgcttg	aaccagagaag	56710060
gcagaggttg	cagtggagcca	aaatcacgcc	attgcaccag	cctggggcgac	56710110
aaaagcaaga	caccatcttt	aaaaaataaa	aataaaaaaag	gcatcttctt	56710160
caagaatgac	atagtgtttc	atgataagcc	ctaattttgc	atttgttcaa	56710210
gtattgattt	gatttagcaa	atatgacacc	aatcttggat	aaagtgtgaa	56710260
caacacaatt	tcatcttctc	cttaatcaaa	actgattagg	tagtctaata	56710310
cctattctga	tcttattaata	aactgataag	ataaaaaaat	tagggaatga	56710360
tggagccagt	aagatgttac	aacctgttct	aaggagaatt	ccaaaaccca	56710410
cacatatttc	agaccattaa	gcatgatgaa	atatactgta	ttactatatt	56710460
taaaaataaa	ctgattatat	agccaacacc	aattgggcag	gggtctcctc	56710510
atccacagcc	acacaatcct	gatcatgcag	ctatgtgggt	gcaaggcctg	56710560
catagcctag	aagggactgg	tctgacctga	gatttcactt	cattcatact	56710610
tgtattttga	gacagggtcc	tgctctgtca	cccaggcttg	agtgcagtgg	56710660
tataatcata	gctcactgca	gccttgacca	gctgggctca	agagatcatc	56710710
ctgcctcagc	tgtaccagta	cctgggaata	caggcaagca	ccaccatata	56710760
aggctttttt	tttttttaat	ttttgtagac	agagaagtct	tactatgttg	56710810
tccaagatgg	tctcaaactc	ctagcctcaa	gagatctgcc	cacctcagcc	56710860
tctctagtaa	cggggccata	ggacacacca	ccnnnnnnnn	nnnnnnnnnn	56710910
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56710960
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56711010
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56711060
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56711110
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56711160
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56711210
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56711260
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56711310
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56711360

nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711410
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711460
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711510
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711560
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711610
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711660
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711710
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711760
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711810
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711860
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711910
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56711960
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56712010
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56712060
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56712110
nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	nnnnnnnnnnnn	56712160
nnnnnnnnnnnn	nnaccagta	ggtctagtaa	ggctaaagtc	aggtgtcaga	56712210
ggggctgcat	tcttttggag	gctctagaga	gaatgtgctc	ccttgccctt	56712260
tccagtttct	agaagccacc	cccatcctt	gactcacctt	gaagatggcc	56712310
cctgactcca	tcttcgaagt	cagaagtgca	gcactctcaa	atcttctctt	56712360
ctgacctctt	cttccatcac	cacatctcct	tctctaattc	tgactctcct	56712410
acctcctttc	tcttataaag	atccttgtga	ttgggtgggca	tgggggctcc	56712460
catctataat	cccaacattt	tgggaggcca	aagaggaagg	aatgcttgag	56712510
gccaaagatt	taagatcagc	cttggcatca	tagcaagacc	ctgcctttac	56712560
aaaattaaaa	ttaaaatcag	ctgggcatgg	tgatgaacac	ctgtagtccc	56712610
agctactgga	gaggttaatt	tgggaggatt	gcttttagctt	aggaggtcaa	56712660
ggctacagtg	agctacgatc	acatcactgc	actccagcct	gggtaagaga	56712710
gtgagactct	gtctcaaata	taagaaaaga	aatatatatt	tggtatttgc	56712760
ccctggttcc	tggcacaaaag	cttctaaagc	tcttataaag	tccttagtga	56712810
tagaggtaat	aggagcacct	tctgttttaa	tatttgggtc	tagtcccagg	56712860
ttcctgacac	aagggcctct	aagttctttc	agatctgcag	catggtaaga	56712910
aagcatgtgg	gatgctgttg	agatgcctgg	gtggctggaa	gttcctagac	56712960
tgcttcagaa	gggctgggtg	ccagagggaac	aaaccacatt	ttttttttaa	56713010
acggagtttg	gctcttgtag	cccaggctgg	agtgcattgg	gacaatctcg	56713060
gctcactaaa	acctccgcct	cccacgttca	agctattctc	ctgcctcggc	56713110
ctcctgagta	gctgggatta	caggcatgtg	ctatcatgcc	cagctaattt	56713160
ttgtattttt	agtagagacg	gggttttcac	atgttgggtc	ggctgggtct	56713210
gaactcctga	catcagggtga	tctgcccacc	tcggcctccc	aaagtgttgg	56713260
gattacaggc	gtaagccacc	gtgcccaccc	ccaaccacat	tttttgaggc	56713310
ttggaacttt	cagccttact	ggctgaactc	caggaggcaa	aaggggctgg	56713360
agattgactt	aaccaccaat	ggctaattgat	tttatcaatc	atgcctccat	56713410
aaaaatccta	aacaacaggg	tttggagagc	ttccatgttg	ctgaacacaa	56713460
ggagggtgctg	ggagggtagc	gtgcccacaa	gagggcatgg	aagctctgtg	56713510
cccctcccca	cttaccttgt	cctgtgcac	tctttcattg	gctgttctctg	56713560
agatgcagcc	tttacctgta	gccagtaata	gaaaataaac	tggccagatg	56713610
tggtggctca	tgctgttaat	cccagcactt	ttggaggcag	aggtggggga	56713660
atcacttgag	cccaggaatt	tgagaccagc	ctgggcaatg	taacgagacc	56713710
ccatctacac	aaaaaattta	aaaaatttag	cagatatggt	ggtgggaaac	56713760
ttgtaattcc	agctacttga	caggctgaag	caggaaaatc	acttgagccc	56713810
ccggagggttg	aggcttcaat	gagctatgat	tgcaccagtg	cactccagcc	56713860
tgggcaacag	aacaagaccc	tgtctcttaa	aaaaaaacaa	aaaaaaacct	56713910
gtttttctaa	gttctgtgag	tttgtctagc	naataatcaa	cccnagaaa	56713960
gaggctcatgg	gaacccttga	tttctaaact	gttgggtcaa	atacagggtga	56714010
caacctanga	cttngcactg	gcactctaaag	tnnnnnnnnn	ntgccacatc	56714060
ttgttcaggg	ggttaatttt	ttatttatat	aaaaagctca	aagcaaataa	56714110
tagaagaacc	ctaaatccct	acagcagctg	actggtaaac	tttgatacat	56714160
ccatataata	aatattatc	caaccatgaa	aaggattaag	aaaaatcaat	56714210
aggtattggc	atagatgtcc	atgaaatatg	aaaatgtgaa	gtgatttcca	56714260
attagcctgt	atcttgaaaa	atatagcaca	ttttctcaat	agcaattatt	56714310
tcttgagata	agattatgga	tctaaagagt	aaaggaaaat	tttcaactct	56714360
ttcttaaaaa	tatttactat	ttttataaatt	taaagagatt	aaataaatag	56714410
atcatcaaat	tagtaaaaaga	cacattaact	ctataaataa	atggaaaagt	56714460
cacagacagc	ccaggcagag	tggctcacac	ttataatacc	agcacttttg	56714510
gagagggtgg	gggaggattg	cttgagggca	ggagttccag	accagcctgg	56714560

aaaacatagc	aagacatcgt	ctgtacttaa	aaaaaaaaaa	aaaaaaaaaa	56714610
tagccaggca	cggtggcatg	tgcctatagt	cccagctact	gaggtggaag	56714660
gatcacctga	gcccaggcgg	tcaagactgt	ggtaagttga	gactgtgctt	56714710
gtgccactac	actcaagcct	ggaagacaga	gcgagactta	gtctcaaaaa	56714760
aaaaaagaca	ataaagaaat	aaagctaacg	agctaacata	agaaaagata	56714810
ctaaaaatgtg	acaaatagga	tgggcacggt	ggctcacatc	tgtaatccca	56714860
gcactttggg	aggccgaggt	gggtggatca	tgagttcagg	agttcaagac	56714910
cagcctgacg	aacatggtga	aaccctgtct	ctactaaaaa	tgcagaaatt	56714960
agccaggcat	ggtggcgtgt	gcctgtaatc	ccagctactc	aggaggctga	56715010
ggcaggagaa	tcgcttgaac	ctgggaggcg	gaggctacag	tgagccaagg	56715060
tcacgccact	gcactccagc	ctggttaagag	agcaagactc	tgtctcaaaa	56715110
aaaaaaaaaaaa	aaaaaaaaagt	gacaaatgaa	aataaagtaa	tatgaggtct	56715160
tttattttatt	aaaaataatc	aaacagaaaa	taacttggtta	aattataata	56715210
ccctgtgctg	gcaaaggtgc	agtgaatgg	gcattttctt	atactattag	56715260
tgatgttttaa	attgtaaata	aaccttccag	ggtaaagctt	ggcaattttt	56715310
aaaaaataat	actgacagtg	actcactata	ctgcctcctc	caactcctgg	56715360
cctcaagcag	tcctcccacc	tcaacctcca	aaagtgcctg	gattataggt	56715410
cgacagccac	catgcctgaa	agcttggcaa	cttacctcaa	gcataataag	56715460
aatgctcatg	ccctgtgact	cacagtaatc	tcacttctgg	aaatatcatc	56715510
tttggtatata	attcaaccta	aacaaaaggt	catatgcaca	aacacagtga	56715560
aaatctcggg	gtattttttt	cttttttttaa	aaaaatatgg	aatgcttcac	56715610
aaatttgcat	gtcattcttg	cacagaggcc	atgccaatct	ctctattggt	56715660
ccaactttag	tatatgagct	accgaagcaa	gcatgagtaa	tttaagataa	56715710
agtgggttaag	tgaagtaagg	aagagttatg	gggaatttaa	aaatctatgg	56715760
tattttatagg	gatctcatta	acagctcagt	aaatattagc	tgctactttt	56715810
attatttttta	tggtaatcc	actcaattaa	aaattgtcgc	taaaaattac	56715860
cattgtcatg	gaacataatg	tatcctacag	tataactgta	gaaacagaca	56715910
caattttgtcc	cttggtatat	gggggggattg	gttccagctc	ctccatttct	56715960
atgtataacca	aaatccacac	atactcaagt	tttccagtc	gtcctgtggt	56716010
attgagatag	atagatagat	agatagatga	tagatagata	gatagataga	56716060
tagatagata	gatagatata	atgagaaaat	tagcgagggtg	tggtgacaag	56716110
cacctgtagt	cccagctact	tgtgaggctg	agtcaggagg	attgcttgag	56716160
cctacagtga	gccataattg	caccactgca	ctccagctctg	ggcaacagag	56716210
tgagacagaa	ggttgactgt	ttaatagaat	ttttctgttg	acttgaagat	56716260
atggtcagga	ctgtggatt	cttcataaga	taactatcta	atccaattaa	56716310
tgctgggaatt	gggaacagaa	aatgttttgg	tgaactattta	ttccttcact	56716360
tgctgtttaca	taaactattg	ccagtgggta	cttaacaacc	aagaatgtca	56716410
tctctgagct	actcacaatg	aaaggtgatg	tctgggaccc	aggtgtgttg	56716460
aggtcccat	gtcttggcta	tgggtgctga	gtgggactta	cttatccaac	56716510
catttttctat	attccggcac	tgggaaactt	gggtttatcc	atcttgataa	56716560
aaagtcattt	aaattccact	tggcaagaac	cacaaatgga	agaaaggtca	56716610
cgaaatcaca	ggctagccct	tgttctcaag	ggaatcttca	gcttaggtgg	56716660
ttctgtaaaa	gaggaatcat	attgttgaaa	aatcatcgca	ggtcagctga	56716710
ggtggcccat	acctataatc	ccagcccat	gggagactaa	ggcaggagga	56716760
tcccatgagg	ccaggagttc	aagaccagcc	taggcaacac	agtgaacct	56716810
catctctaca	aaaaattaga	aaatgagctg	ggtgcggtgg	cacattccta	56716860
tagtcccagc	tactctgcag	gctcaagtgg	gaggatgaaa	gtggaagctg	56716910
cagaagtggg	agctgcagtg	agctctgatg	tcaccactgc	actccagcct	56716960
gggtgacaga	gacctgtct	caacatacac	atacatacac	acacacgcac	56717010
acacacacac	acccatgcac	gcacgcagac	acatgcagac	acacacacac	56717060
accctcatct	caatctgtcc	agccttgact	aaagaaaagg	atctgctggt	56717110
tacagaagag	gtatgctctt	ttgtaggaca	gggaggggacc	agcaaacttg	56717160
ttcacagtct	tttctctcat	ctctgcttag	ttttccaaga	accctcatag	56717210
tggaaacaga	gtccctggga	aaaggaccta	aatctttata	ttaccagggg	56717260
agaaatatgc	ctcctttgtc	aattaataaa	tagaatacct	gccttaaaat	56717310
cgagggaggt	ctgctagaat	gaatcactcc	ctacgacccc	gacctatgca	56717360
gggaacatga	aaaactggag	tttggccagg	cgcagtggct	cacacctgta	56717410
atcccagcac	tttgggaggc	cgaggcggtg	gaccatctga	ggtcaggagt	56717460
tcgagaccag	cccgggccaa	catggtgaaa	ccccatctct	actaaaaata	56717510
caaaaattag	ctgggtgtgg	tggtggacgt	gtgtaatccc	agctactcag	56717560
gaggctgagg	caggaaaatc	agttgaactc	gggaggtgga	ggttgagctg	56717610
agccgagatc	ggaccactgc	gctccagcat	aggtgacaga	gcgagactct	56717660
gtcttaaaaa	aaaaaaaaaaaa	aaaaaaacca	aaactggagt	ttgggggcct	56717710
ataatattgt	agggcacatt	ccagattgta	cagatcttac	caaagtacgg	56717760

cctttccaca	aagcactgcc	aaacaccagc	tggctgtctc	catggggcaa	56717810
caccatatct	ctggctcttg	tttccccagg	agatgctcta	ttccagcctt	56717860
cccttacctt	ttgtatcccg	cacatttcct	agcataaggc	ctcaaaacat	56717910
agcaggctgc	caagtcactg	ctgatctact	gcatactga	cttcagattc	56717960
cagcagggca	ctttgtgttt	gagaaattaa	agagaagaga	atggatccag	56718010
actgacgctg	gccaaacttag	cggtccaaag	atataagctg	ctcaaactga	56718060
ctacagtgcc	tctgattttc	tgagtctgcc	ctgctcctag	ggctctttgg	56718110
acgaaattaa	tttctccttg	ttgaagaaat	ggtttgctcg	aatataagaa	56718160
atttggccag	gcacagtggc	tcacgcctct	aatcccagca	ctttgggggg	56718210
gccgaggcgg	atggataacg	aggtcaggag	ttcaagacta	gcctggccaa	56718260
catggagaaa	ccccgtctct	actaaaaata	caaacattag	ccgggtatgg	56718310
tggcaggggc	ctggagtccc	agctactcag	gaggctgagg	caggagaatc	56718360
gcttgaaccc	aggaggtgga	ggttgcagtg	agctgagact	gtgccattgt	56718410
actccagtcc	gggtgacaga	gcaaaactcc	atctcaaaaa	aaaaaaaaaa	56718460
aaaaaagaaa	agaaaagaga	atataaaaaa	tttaaagaag	aaaaaaatgt	56718510
ttaaataaaa	ttttctccta	aagagcactc	aaaagcacac	tcccatcaat	56718560
ggtggatcta	tttctctctc	tcttgctttt	tttttctgag	atagggctct	56718610
tcactttatc	acccaggctg	gagtagtagt	gcaagatcat	ggctcactga	56718660
ggctttcaat	tcctggactc	acgctatcct	ctcacttcag	cctcctgagt	56718710
agctgggatt	gtaggtgcat	gccaccatgc	ctggctaatt	ttttaatttt	56718760
ttgtagagag	gaggttctca	ctatgtggcc	caggctggtc	ttgaactcct	56718810
gggtcaaact	aatcctcccc	ccttggcctc	ccaaagtgtc	gggatatagg	56718860
ctttagccac	tgtgcccagt	ctatttatct	ttctataact	catatattaa	56718910
atcttctctc	ttaaattgcc	tcctaccac	ccatctgcag	ggatagcaaa	56718960
cacaggtatc	agcaggtcag	gcaacctacg	tgagggaaag	aggctggagt	56719010
aatgtgggag	accgggggaa	actgtagcaa	actgcaaggt	attcacttca	56719060
acagttcttc	aacactgtga	tagtcaaacc	aaacacagaa	acaggaagcc	56719110
cagttcagtc	accagcagcc	aacgtgggat	atttcatcta	catgctacaa	56719160
ggcaaaactgc	ttaggatgaa	tgacctggtc	tccaccaacc	tactcttggg	56719210
gtccagccgc	tcctgcccaa	gtccctacac	tctactgaca	tgaccacat	56719260
tcacagtact	ctcctgacac	cttttcttgt	tcactttttt	ttttcccaa	56719310
caaatttttt	ttcttttttt	tttttttttt	tttttacttt	ttttgagaca	56719360
gagtccccct	ctgtcacaca	ggctgaagta	cagtgggggtg	attatggctc	56719410
actgcagctt	ccatcttctg	ggctcaagtc	atcttcccac	ctcagtctcc	56719460
caagtactcg	ggattacagg	catgtactac	cacacctggt	taagtatatt	56719510
atttttcttt	tttctttctt	tctttttttt	gttggttgaga	caaagtctca	56719560
ccctgtcacc	catgctggag	tgcagtgggg	tgatctcggc	ttactgcaac	56719610
ctccgactcc	caggttcaag	caattctcct	gcctcagcct	tccaggtagc	56719660
taggattaca	ggtgtatgcc	accatgcctg	gctaattttt	gtatttttag	56719710
tagagacagg	gtttcaccat	gttggccagg	ccggtcttga	gctcctgacc	56719760
tcaggtgatc	caccacacac	ggcctcccaa	actgctggga	tcacaggcac	56719810
gagccactgc	acccaaccag	tttattattt	tttctaggga	cagggctctca	56719860
ctatgttgcc	caggttggtc	ttcaactcct	gggcacaagc	aatcccctag	56719910
ccccctcctt	ccaaagtgtc	ggaattacag	gcttgagcca	acacatcagg	56719960
ctctaacaaa	tatttcttga	gcacctacta	cataccaggc	actatattag	56720010
gtgcccagga	aacagcagtg	aacatgggat	aagagtttct	gctctcacag	56720060
agcttatagg	gtagacagaa	attgatatag	tttgaatgtc	tgtctcctcc	56720110
aattctcata	ttgaaattaa	atccctaata	ttggagggtg	tgcttggcgg	56720160
gaagtgtgtg	ggtcatgggg	gtggatccct	catgaatggc	ttagtgtccg	56720210
ccccatggta	atgagtgagt	tctctctcta	ctagtctcat	tgagagctga	56720260
ctgtttaaag	agcctggcac	ctcctccctc	ctcccttgct	ccgtcttctc	56720310
accaagtgc	gtgctggctc	ctccttcacc	ttctaccatg	agtggaaacc	56720360
tcctgaggtc	ttcaccagca	gacgtggca	ctatgctttt	tttttcttga	56720410
gacagggctc	cactctgaca	ctcaggttag	agagcagtg	tgtaattatg	56720460
gctcactgca	accctgacct	cctgggctaa	agagatcttt	ttaccttagc	56720510
ctcccaagta	gctgggacca	taggcatgca	tcagtgcacc	atcacgcctg	56720560
gctgattttt	ttttattttt	agtagagaca	atgtctcacc	atgtttccca	56720610
ggctgggttc	aaacttcttg	gctcaagcaa	tcctcctgcc	tcagcgtccc	56720660
aaagtgtctg	gattacaggt	gtgggccact	gtgcccagcc	tggcaccatg	56720710
ctacttgtat	agtctgcaca	actgtgagcc	aaataaaact	tttttcttta	56720760
taaagtaccc	agcctcaggt	gtttccttat	agcaatgcaa	aatggactaa	56720810
tctaggaatc	gtatgccagc	acttaccaga	ctgtgatgag	cacatgacag	56720860
aagtacagag	tgggactgaa	acgttcccca	ggggcttctt	gttgcacatg	56720910
agagtggggg	gagacatctc	ggctgaggct	aaagatgagt	aaagatgaga	56720960

gtgaaagagt	gatgatgggg	gtaaggagag	ggtttcctga	agaaggacca	56721010
gtaaagtccc	ccaagaaaac	tgtgaggtca	cttactagaa	atcaccagca	56721060
tcccatttcc	cattggcaag	gagctcagca	cagtcccttg	gataacccaa	56721110
cctatgcccc	aatcccatct	gtgtgagtct	atctgcccgg	acccttctta	56721160
gcatattaga	gtccaatcag	gagacaaaa	ccactcaaaa	gtttaaagtg	56721210
gtaaaattta	atacagagaa	ttattcatta	taacagggca	acagcataat	56721260
aagaaattgg	ctacccaaaa	gtaaaaaagaa	ctctagagaa	cacaggacta	56721310
tcagaggcca	ggcatgggtg	ctcatgctta	taattccagt	aatatgaaaa	56721360
gccaaaggcag	gaggactgct	tcaggccagg	agtttgagac	cagcctggga	56721410
aacacagtga	gaccctgtct	ctaccaaaga	aaaaaaaaat	agccatgtgt	56721460
ggtggtgcac	acctgtagtc	ctagctactt	gggaggctga	ggtgggagtg	56721510
ttgctttgagc	cagggaggtc	atggctgcag	tgggctacga	tttcatcact	56721560
atgcttttagc	ctgggcaaca	gagggagacc	ctgtctcaaa	aaaaaaaaaa	56721610
aaaaaaaaagt	ataggaccag	ctgatacaat	gaacggtaac	tgcccctact	56721660
gtccctacga	tgagatggag	taaccaagga	agggagtctt	cactagggct	56721710
gagatccagt	catccttggg	agagggcaca	gtcatgggta	attcagtggc	56721760
agaattgtctg	cattaccaca	tggtggaaca	tgctagcaat	ctgccgtctg	56721810
gaatttgctg	aaaatccgcc	ttctatgggtg	ctagattaag	ctattaatga	56721860
acagctgtct	cactacaaaa	ccaccacgc	aggtgcagag	aaactgctgg	56721910
tcactgggtg	ctgctgagca	ccgtgcaggg	caggagccca	gtgctggaca	56721960
agctcagtgg	ctgtggagat	tgacacacta	caatcagaaa	gaacagcctc	56722010
ttacaaagt	cctctagcac	aatgtactga	caacatttaa	taatgcagca	56722060
gctgatagag	gaagcatata	aggaacccac	gtctatgttt	gcagagcagg	56722110
cagtgaagg	tgaatttggg	actgatgggc	agtcattgga	tggacagctg	56722160
gtacatgaca	catttcaatc	taccaaggaa	aaaaggaaga	gagaaaagag	56722210
atagaggacc	aagcatgggtg	gttactcct	ataatcccag	cagtttgagg	56722260
ggccaaggca	ggtggatcac	ttgggtccag	gagttcgaaa	tccagcctga	56722310
gtaacatagt	gaaaccctga	ctctactaaa	aataaaaaaa	ttagctgagt	56722360
gtggtggctt	gtgactgtag	tcccaactgc	tcaagaggct	gaggtaaaag	56722410
gatcccaagt	gtctgagagt	tcaaggctgc	agtgcagccat	gatcacacta	56722460
ctgcactcct	gcctgggcga	cacagcaaga	tcctgtctca	aaaaaagaaa	56722510
tgctggtgaa	aatgtgaaga	aattggaacc	cacatacatt	attggtggga	56722560
acaaaaaaaa	tggtttaacc	actttgtttg	ggtgtttctt	tcttgctcatt	56722610
ttaatttgat	tttttaaaaa	atcgagatag	gatctcacta	tcttgcccag	56722660
gctggtctct	tttaacttgc	aggctcaagt	gatcctccca	cctcagcctc	56722710
tcattgtagct	gggattacag	gtgtgagtca	ctacacctga	ctggtgtaac	56722760
cactttggaa	aacagtttct	caaaaggcta	aatgtacagt	atcatagaat	56722810
gcaacaattt	ctctcctagg	tatatataccc	agagaaataa	aaatatatgt	56722860
ccacacaaaa	acttctaaat	gaatcttcac	agcagcatta	ttcacagtgg	56722910
ccaatacatg	gaaacaaccc	aaatgttcac	caactgatga	ataaacaaaa	56722960
gtggtgtgt	ctctaccatg	gaatattatt	cagccataga	aggaacgcaa	56723010
tagtgatgac	tgctatgaca	ggaaggaaact	ctgaaaacat	tgtgctaaga	56723060
gggaaaaaaa	agccagccac	aaacaatcac	atattgcaca	atcttattta	56723110
tataaaagg	ccaagccggg	tatggtggct	cacacctgta	atcccacttt	56723160
gggaggccga	ggcaggcaga	tcattgaggtc	aggagtttga	gaccagcctg	56723210
accaacatgg	tgaaagcctg	tctctactaa	aaatacaatg	attagccagg	56723260
tggtggtggca	ggcgcccata	atcccagcta	ctcagaagac	tgaggcagaa	56723310
ttgcttgaac	ccaggaggca	gagattgcag	tgggctgaga	ttgtaccatt	56723360
gcactccagc	ctagatgaca	gagcaagact	ctgtctcaaa	aaaaagaaga	56723410
aaaaaaagg	ccagattagg	caaatctata	gtgacagaaa	atagatcagt	56723460
ggttgccctat	gggaacacag	gggatgagg	ggagtagaaa	tggctaagaa	56723510
gtgtggattt	ctctacaggg	taatgaaaa	ttctaaaatg	gatttgtgatg	56723560
atagatgcac	agctccgtga	atattataaa	aaccactgaa	ttgcatactt	56723610
tgataaatga	atttatggta	tgtgaaccat	atttcaataa	agttgttatt	56723660
taaaaaaagg	aaataggggg	ccaggcacag	gtggctcata	cctgcctata	56723710
atcccagcaa	tttgggaggc	tgaggcagga	ggatcacttg	aggctaggag	56723760
tttgagacca	tcctgggaaa	cacagcacta	tcccattctt	acaacaaaaa	56723810
ataaaaaatt	aagctttgca	tggtggtgta	cgtccgtagt	cccagctaata	56723860
tgggaggctg	aagtgggagt	attgcatgag	cccaggagtt	tgaagctaca	56723910
gcgagccatg	atcacaccac	tgcactacag	cttgagtgac	agagcaagac	56723960
cctgtctcta	aaaaggaaag	aaaagaaatg	caagttttta	tcactctgtg	56724010
agggtaaaca	agtttgaggg	gaaacagaag	aacaaaagag	cactgaatgg	56724060
tgagagtggg	aggctggtta	ggctcattgc	tagctaagag	acttctgaaa	56724110
aattcatgag	taaaatcaca	gctctggggg	tcagtcaggc	agtcaaatga	56724160

tgaatgttaa	atccattaca	aatgcccac	gtctttcttt	acattccttc	56724210
tagtaaaaaa	ttcctaactg	cctaaatagc	aagtgttctg	aaatgatatc	56724260
agctgtttat	taaagaagta	acatctcaaa	ttttaaaaaa	catgagtgca	56724310
atattatttt	cctcctccca	ttttacaggt	gtgcaacgta	aacaacagag	56724360
aattttaagta	attttcctaa	gtggcagagc	acagatttaa	atcctatcac	56724410
tttggcctca	gagtttgtgc	tttattttat	tttattattt	taaaattcat	56724460
ctttttttaa	accaagaggc	ttatgcaaaa	ttctgctagg	attcaaaaaa	56724510
agcctcctgc	ttttcctgac	catgggtgca	ggcagggctc	atgctgcatt	56724560
ttctttactt	ataattggat	ggagacatga	tgacaggagc	tctcttgggc	56724610
atcgtgcatc	agaaaggaga	gtgaacagaa	gcacagggaa	gccagtctta	56724660
cccatgtttg	agcaactgaa	ccaatgtcag	cctcagccta	cctgcacaat	56724710
tactttttatg	tgagaaaaac	aggcacctcc	ctatttggaa	aatcacagct	56724760
agctacagta	gcaaacatc	ctcaagttat	tcctggtaac	attattccta	56724810
cagccccaat	ccttttaact	cattacatta	gctctacatt	tctctattgt	56724860
tcctaacatc	tagttcataa	ccaagagctt	atgttattaa	ataagctttg	56724910
agttaacagg	attaatgaag	agtttaaaga	taccaaaaaa	atctcataga	56724960
agtacatgct	gaatgtgcca	actgactact	atgaactaat	ttacttttga	56725010
gaaacatgct	atttcttata	tgctactgac	agtgcctgta	aaaatgggtgc	56725060
actttaactg	gctgttttat	tttctcccca	atcttaatat	ctaatatcca	56725110
cagactgaga	tcagagcacc	agcatatgca	atgtcacttt	atactagagt	56725160
tagaaggcga	gacaaagagt	agaaatcaga	ctcactttgc	tagctgctct	56725210
cctctgtcca	gaatatttta	aaggagactg	tattccccc	gggctcgctg	56725260
acctggcact	ctctccttgc	ggaggctcac	accactgtct	tgtgcagctc	56725310
tcagtgcctt	tgcagaatgt	gtcaccatgc	tataaaacac	cttctcctcc	56725360
acctctcaac	ctaagcatgt	gtcacttggc	ttcaggacag	cttgtcttct	56725410
cactgtcctg	cctttccttc	tgaacgtctt	ggctcttctg	gtggctctct	56725460
atctcactat	cctgggtgaag	ctttgggggtc	attacttctt	cagttgcttc	56725510
cctcgccaat	gtatcaactg	aagacagtat	ggacgagttg	cttcagagta	56725560
caactactgt	caactctggtc	ccagctgggtc	ccaggacttc	ttctgactct	56725610
aaagttttctg	agacaccaga	tatagttgga	ataaaatatg	ctccctagaa	56725660
tgagattgtc	atcattacct	ggacttttta	atctgcactt	ttgctcatgc	56725710
tattccttct	atcaggagta	gctttgtctc	tccttggtga	cattctcctc	56725760
ttccagcaag	gcctaattta	agtgtaat	acttcataaa	gttccctgtca	56725810
ccttggaatt	aacctccctt	ctacatagtg	tcattgttca	ccttggtcaa	56725860
atctttataa	atccaggctc	aggctcatgt	ggtcacggag	tgtgtctttt	56725910
taccaaaagg	aacaccttgg	ccaggcacag	tggtctcatgc	ctgtaatccc	56725960
agcactttgg	gaggccgagg	cagatggatc	acctgaggtc	aggagtccga	56726010
gaccagcctg	accaacatgg	agaaaccctg	tctctactaa	aaatacaaaa	56726060
ttagctgggt	gtgggtgggtg	gcacctgtaa	tcccagctag	tcgggagggt	56726110
gaggcaaaa	aatcgcttga	acgcaggagg	cagaggttgc	cgtgagccaa	56726160
gatagcgcca	ctgtactcca	gcctagacaa	caagagtga	actctgtctc	56726210
aaaaaaaaaa	aaaaaaaaaa	aaaaaggaac	acctcatgca	ctttacaagg	56726260
agctccttca	ggaaaggatg	gcacctcggtg	tagaggagag	atctctgatt	56726310
aattaattaa	agacattcag	caattaaatt	aattaaagac	attcagaaat	56726360
ttagctttat	cacaggcagc	tgttcacacg	taaaactaat	ttggttacct	56726410
tcactctgtat	acacatgtag	ctagtgaatt	gacataacta	ttaaactgga	56726460
gaatcaacag	aaacatgcca	aggactggcc	cttaccttcc	acagtgcact	56726510
caatttcagg	aacctttcca	ttactctcag	ggcttcatgg	tctttttggg	56726560
gactgcaaag	ctgaaaaata	agcacagaat	tacttttagta	tattgttagga	56726610
aaaaatacag	aaggagatgc	tttataagag	cagagtatta	aaattttattc	56726660
agtttggctc	ttttcaaagg	aacctgcaga	attaaacttg	tagtgtaatc	56726710
ctccacagga	ggctcaacaa	agccatggaa	acatgaatga	agtgggattg	56726760
caaagtgtgac	atttccaata	aaattttttt	tttaaaaaaa	ggaaaaagag	56726810
gtccattaaa	taaaacacct	taaaaagcca	agaaactaac	tgcaagttcc	56726860
acaaagagag	tggcccatga	ttctgtgtga	tttaagggaat	gctgagaaag	56726910
gatgtcttaa	tgcagtga	tgtaatagga	agtacaggta	atctcttacc	56726960
tttagtggtt	attttattag	tcattccagg	aggcgagtag	gaaataacta	56727010
gttcagggtc	agaaagaaaa	ccaagaaatg	ttacggtagc	cagtacaatg	56727060
gtgggtctcac	agctagttaa	acagtcgtga	tgtgaccagt	gtttttccct	56727110
aaaggcggtc	gcaaaatcat	cctgtaattgt	aacaataatt	tctatcattg	56727160
tggacaaaaa	ccctgacttt	tccttcccat	gactggcaag	taattatatg	56727210
aagagtaaat	tacttctttc	atgaaaaaga	aaattaacac	tagtatactt	56727260
tcttcagcca	ctaaaattct	acagaaatgt	agatataata	agaaactcaa	56727310
ccatcccac	agggcagggtg	cagtgtctca	tgctgtgaat	cccagcactt	56727360

ttggaggctg	aggtgggtgg	gtcacccgag	gtcaggagtt	cgagaccagc	56727410
ctgaccaact	tggtgaaacc	ctgtctctac	tagaaacaga	aaaatcagct	56727460
gtctctacta	aaaacacaaa	aattagccag	gcatgggtgg	acatgcctgt	56727510
aatcccagct	acttggggagg	ctgaggcaag	agaatcgctt	gaaccaggga	56727560
ggcaaagttt	gcaatgagtc	gagattacgc	cattgcattc	cagcctgggc	56727610
aacaagagtg	aaactctgtc	tcaaaaaaac	aaaaaagaaa	agaaactcaa	56727660
tcatcaatat	ttagtcgctc	acaagattga	cccagggaata	acagagcagc	56727710
taaaaccag	caaattagac	ttcattcaag	tacttaggtg	tactgtgagc	56727760
caccacggtg	cctctcaaca	ccatggaggc	cgtggtcact	attccttgac	56727810
tgacgaagac	agctgagtag	aggatggggg	actgcagggg	caatggagag	56727860
tcgcaggact	gcagctcaga	gccaggcctt	tccatcaagg	gcttacagcc	56727910
acgcctgtca	cctctagcta	gaaaaacatt	ttagtagaaa	tatcatttta	56727960
ttgtttta	tccagaaatg	actacagtgg	caaatagaac	ttactttttt	56728010
aatattgttg	ttgttgttgt	tgaggcagag	tctcactctg	tcgcccaggc	56728060
tagagtacag	tgctgtgatc	ttggctcact	gcagcctctg	cctcccagct	56728110
tcaagcgatt	ctcctgcttc	aggctcccaa	gtagctggga	ttacaggcac	56728160
cagccaccat	acctggctaa	tttttaacat	tttttagtaga	gatgttggtc	56728210
aggctggctc	tgaactcctg	acctcaggtg	attcacctgc	cttggcctcc	56728260
caaagtgtctg	ggattacagg	cgtgagccac	catacccggc	ctggaactta	56728310
ctttttaaca	acaaacttgg	tttttattact	ccacagatga	tcctttcaag	56728360
tcatggaatt	ccaaaccagg	tagggccttg	aaagggaatt	ctgtctggca	56728410
agccttctac	atacaggaac	tcgggggttg	attcaaacac	agaatatggt	56728460
acttttactg	cctctgagtc	caagagcttg	agctgaaagt	gcaagagaac	56728510
ataatttgta	gaagataaaa	tttgacacag	acattctttt	atttgtatct	56728560
gcattctcaa	gtacattatg	aacaagttcc	ctacaatcaa	gcaatattca	56728610
cttcataattg	aaggcatttt	gcttaggata	ttgggcaaat	ctgaagaaca	56728660
agcctaaaag	ataagtactt	tcaggaggaa	tatttttaaca	cttttctttt	56728710
tttttttttt	tctttgagat	ggagtcttgc	tctcttgccc	aggctagagt	56728760
gcagtgggtg	gatctcagct	cacagtaatc	tccacctccc	gggttcaagt	56728810
gattctcttg	cctcagcctc	ccaagtagct	gggactacag	ttacacgtca	56728860
ccacaccag	ctaatttttt	gtatttttag	tagagatgag	gtttcaccat	56728910
gttagccaag	atggctctcaa	tctcctgacc	tcaagtgatc	catctgcctc	56728960
agcctcccaa	agtgtctggg	ttacaggcat	gagccatcgt	gcccggccca	56729010
ttttaacact	ttcaattaaa	aggagagtta	tgttatcagt	caaaatgacc	56729060
tctgtgttct	aagcagagat	gaccgaaggc	cattccgtac	agcagtggcc	56729110
cctagcagcc	tttgcattct	catcaaggag	cagggtagct	gttcaaagac	56729160
aggtgggtgtc	tatgacgcga	ctaagcagta	ctcagtaagt	tcaagtcatc	56729210
agacacagag	tagtcaggat	tcctcataag	ggacaggaaa	cgtcagcatc	56729260
aggatttgag	ataatggaca	ggaatgccc	ggaatgctgc	aagtgtctcc	56729310
cactccttac	ttccttctcc	atgacagata	cacccaatca	actgaaacac	56729360
actttcttgc	taaacctcct	cagaatcggt	caagacaaca	ttccaggaaa	56729410
ctgtctgtcaa	caaatataag	tattttgcac	acgtgtgtta	gccagattag	56729460
aagtacagac	ggtggaatga	cacctatccc	aatgcctccc	ttctgtctcc	56729510
attctgacac	atgagaaagc	tgaatccaag	agaaatgaag	ttgtctgtct	56729560
ggtttcaata	gcaagtgtag	tagtagaaat	aaaccaaagc	ccggtttctg	56729610
ggaccttcct	tttttaggtc	ctgaacttct	ataattcatt	cagacatgac	56729660
ccttcatgtg	ctatgtatta	aagacatact	gtagggaaag	gactatgagt	56729710
cccttcattgt	ctatatcaaa	aagcaagact	cttgtctcca	aaaaataaaa	56729760
aatacaaaaa	tgtttcaatt	aaaaaagtgt	gatggctttt	ggcaatgttt	56729810
aaataaaaaa	taagagtatc	actttcaa	aagctgacac	tttctgaaca	56729860
tgtaataaacc	agggtcacag	aatcacagag	gccagtaaga	agattaaaaa	56729910
gtaccctgtg	gagccccagt	gacacatata	ccaccaggaa	aatttctctgt	56729960
tctgtgcgtg	ccaacgaatc	cacagcaaca	taatgtcggc	aagtggcacg	56730010
acctgcctcc	cctacctcca	ggcagccaca	gggaaagtgg	agagcaaagg	56730060
gaccattccc	tgtgtcactg	ataagtgagg	ctcaatggga	aacatgaggc	56730110
atgagcggca	tccaaaggta	ggtgacaagt	tcagtgtctc	ccctttgtct	56730160
tctgtcttca	tgtattatct	gcattctctg	ctcttttcat	gataaggaag	56730210
gaggcagcaa	tagcgtgtct	actgtggaaa	tgtgggacta	ctcatcgggc	56730260
atcctggaac	ttaatgtaga	agtttatata	gccctgggaa	gctttcagaa	56730310
cctggctctcc	taaggtctcc	ggctcttgaa	aggtcttgaa	caggcagtag	56730360
actgtgaacc	cccaggagtt	cacacaggtt	ttctaggggtc	ttcacactgc	56730410
taaccgcagt	catggtaact	ccttcatttc	ctttttaaac	aattgttttc	56730460
ccaattatcc	ctttaaaatt	gggcagttta	ggctgggcat	ggtggctcac	56730510
acctataatc	ccagcacttt	gggaggctga	ggcgggcaga	tcacttgagg	56730560

tcaggagttt	gagaccagcc	tggctgacaa	tgtaaaaccc	catctctact	56730610
aaaaaataca	ataattagct	gggtgtgggtg	gtggatacct	gtaatcccag	56730660
ctgcttggga	agttgaggca	tgagaatccc	ttgaaccag	gaggcagagg	56730710
ttgcagttag	cggagactgt	gccactgtac	tccagcttgg	gcaatagagc	56730760
aaaactcaat	cttaaaaaaa	aaaaaaaaat	gggtagttaa	gagcctctga	56730810
tatagcgagt	gtgcacaatg	attcgtgggtg	atgagctatt	aaacatcatt	56730860
aactacgcta	tgggtggtttt	ttaaaatcgt	gattggctag	gcgaggtggc	56730910
tcattgtctgc	aatcccagca	ctttggggagg	cggaggcgga	cggatcacct	56730960
gagattgaga	gttagagacc	agtctgacca	acacggagaa	acactgtctc	56731010
tagaaaaaat	acaaaattag	ctgggtgtgg	tggcgcatgc	ttgtaatccc	56731060
agctactcag	gaggctgagg	caggagaatc	acttgaatcc	aggaggtgga	56731110
ggttgcggtg	agccaagatt	gtgccattgc	actccagcct	aagcaacaag	56731160
agtgaaactc	tgtcttaaaa	aaaaaaaaagaa	aaagaaaaaa	aagaaaaatc	56731210
aatacccagc	agacggtgag	acatccataa	tgtgaccagg	tgccaaggaa	56731260
tgtccataag	gatggaactg	cagtcaagtc	tcctgaattt	ccttccttgg	56731310
gccccctccc	tgggcccacc	ccctacccat	ataaccacat	cctcactttt	56731360
ctcccctaaa	cacaaatcaa	gggacactca	aggcaagcat	agtactcact	56731410
tcttaaccaa	tccatccctc	cttccaaacc	caacacatgg	agagaactta	56731460
tgttagcctt	gtaaataaag	aaataactta	catcccacag	ttatttttcc	56731510
agtctgcctt	tttttaacca	ctcacttggc	agaaacttat	tctcctggcc	56731560
agtagctatt	tgtaaaactg	gtctgcctcc	ctattttcac	accaaagata	56731610
ctaattcctt	agtagaaagg	ggagtattct	aaatcagaca	attagcagat	56731660
cacctagaat	ttccaccccc	tccacagaga	cacacaaaca	agagaattaa	56731710
cccaatttgt	tgtctataat	aaataggatg	ttctactcac	cactgagccc	56731760
cagaggttgt	tactagatga	ctaatacccta	gcagtaatgg	ctacaatggc	56731810
ccagttccct	atattattac	tgtgtgcaga	aatctgaaag	gcgctcttct	56731860
cccaccccca	tcttccctta	catacttttg	cacatcaatc	cattaagaat	56731910
ctgaagtgtg	atattatctta	ttcactatta	tcctgttaaa	ttgttctaca	56731960
gctgccatgg	gagtaattgc	agtgatcagt	gattcagact	tcttgagcaa	56732010
agggccttaa	gtaggtttca	ttattaaaac	ttccaaaggt	tttattattg	56732060
aaataaacct	gagagagagc	agtatgcctc	catagggcag	tcataactaat	56732110
gattcttaag	ttacgaaggg	atgtcagact	catcatcaat	gttgcaactg	56732160
tgagtatgac	taactgggaa	tttatcaacg	aataaaaacca	actggcattc	56732210
taagaatgaa	gaataataat	ctcattttaa	gtagcatatg	cagtgtctgtg	56732260
caacccttta	aaaagctgcc	tttggtgggg	cgcagtggct	catgtctgta	56732310
atcccgacac	tttgggagcc	tgaggcaggc	agatcacctg	aggtcagcaa	56732360
atagagacca	tcctgactaa	cacggtgaaa	ctccgtttct	actaaaaata	56732410
caaaaaatta	gccagggtgtg	atggcacatg	accgtaatcc	cagctactcg	56732460
ggaggccgag	gcaggagaat	cacttgaacc	caggaggcgg	aggttgcagt	56732510
gagccgagat	catgccattg	cactccagcc	tgggcaacaa	aagcgaaact	56732560
ccatctaaaa	aaacaaaaac	aaaaacaaaa	aaacctgcct	ttggccaggc	56732610
acagtggctc	agggctgtaa	ttccagcatt	ttgggaggcc	gaggcagggtg	56732660
gatcatattt	acctcttaaa	tatctctcaa	aacctcgctt	ccccttcata	56732710
ttcactggca	cctcctattc	agcatctcat	cactcattca	acaatattta	56732760
atgagtatga	agcatagcta	ggtgttttta	aacagaataa	aaacagtgtg	56732810
atctttgttt	tctggtggct	caacatctag	ttattttctt	attcttttct	56732860
ttccaagaca	gagtcttgtt	ctgtcaccca	ggctgcagta	cagcggcacg	56732910
atctcggtc	actgcaactt	ccgcctcctg	gattcaagca	attcttctac	56732960
ctcagcctcc	caaggaaactg	ggaatacagg	tgtgcaccac	cacgcccagc	56733010
taatttttct	attttttagta	gagatgggat	ttcatcatgt	tggccaggct	56733060
ggtctccatc	tcctgacctc	gtgatccacc	cacctcagcc	tcccaaagtg	56733110
ctaggattgc	aggcgtgagc	caccgcgccc	agcaaacatc	tagttatttc	56733160
ttatctgaac	tattcaggta	tactcttgtc	tcttaccatt	cattcacaga	56733210
gtagtaccga	gttaagagac	ggtaagagac	aaaagtatac	ctgggtacac	56733260
catgccataa	gtacccagag	tgtctcctcc	tcaaaactta	gattttattc	56733310
aggggtcagt	aaactttttc	tataaaaacc	agacagtaaa	tatttttaggc	56733360
ttctgggcca	tatgttctct	gttgaaacta	ctcaactctg	ctcttgtaac	56733410
attaaaaaca	gccatggaca	atatgtaaac	ccatgggtga	gactgtgttc	56733460
caatatgggt	tattcataaa	aaaaggcatc	aggccagatt	tgacccatgg	56733510
gacacagttt	gccaaccctc	gatttaattg	caatgctcct	tgctccattg	56733560
gaattatcca	gattcctttc	catagcctag	aagggtgctc	atcctcatct	56733610
gagtattact	tgttctctcca	ggatccacta	tgtctccagcc	accctgacca	56733660
tctttatggg	gcttaaacac	cgtgccaaagt	ccattcacat	ctcagggtccc	56733710
tgcacttgat	gttctaccga	tcacaaccac	tcttgcccag	tgagctgcct	56733760

cagctcaa	at	gtcacctc	cagagaggtc	ttcactgccc	accccaacag	56733810
gtcagtg	at	atcccaac	ccaaactttc	taactcccaa	agctctat	56733860
at	at	tacttt	caccgtctga	aaacatatca	tttgttgcat	56733910
gtacatag	tg	cttctcca	ccaccacacc	accactagaa	cggagagctc	56733960
ccaagaac	ag	actgtctt	actcattctc	cccttagttc	ccagcaccta	56734010
gatctgtgc	tg	gcacttgg	caggggctca	aaaaatgtct	gttatcggt	56734060
gggcgcagt	gt	tcatgcct	gtaatcccag	cactttggga	agctgaggta	56734110
ggcagatca	ct	gaggtcag	gggttcaaga	ccaacctggc	caacatggta	56734160
aaaccccat	ct	tacttaaa	aaatacaaaa	attagccggg	cgtgggtggta	56734210
cacacctgt	tt	cccagcta	ctcgggaggt	tgaggaagga	gaatcacttg	56734260
aatccgggg	gc	agaggttg	tggtgagcca	agatcacacc	attgcactcc	56734310
agcatgggc	aca	agagtaa	aactccatct	aaaagaaaaa	aagaaaagaa	56734360
aaaaagaa	ga	aggaagga	aagaaatgaa	agaaaaggaa	aggaaagaaa	56734410
aggaaaggaa	agg	aaaggaa	aggaaaggaa	aggaaaggaa	aggaaaggaa	56734460
aggaaaggaa	agg	gaaaaaga	atatgaacag	ctggcgagat	caagacttct	56734510
ggcccaaagg	gc	ctgtagct	ggtctgacca	ttcatcaaaa	aatctgaatt	56734560
taaatggaga	ca	cactggac	tgtcacatgc	aataaaaagaa	ctgaaacagc	56734610
aaatggtct	ga	tttgggtt	tccttccacc	ctacacggct	ccagcaaattg	56734660
gaggagtga	cg	tgcctgcg	gtgctcctga	cacatctagg	gtagttccta	56734710
ttcacagtgg	tc	aacctttg	ggattttcca	cacctcacag	ggtcatcgct	56734760
catctaggct	aa	cctagaac	gtccagataa	aagctcctgt	ttagagactg	56734810
gagagaaaat	tg	tcaaacct	gaaagcccat	gaactgtctc	agccagggga	56734860
gaacacacgt	ag	gaaagact	aaagagtgtat	gtcagaagcc	taattttcaga	56734910
ttgacctgta	ca	gcttttaca	attgtgagta	attaacttac	ttttgctagg	56734960
ctctgttatt	ct	cacatcaa	taatgagaaa	gcgaactgaa	tagtctctca	56735010
aatccctttc	aa	ttttaaca	tcctctgact	ttatgattaa	agagtgatac	56735060
taataaggct	aa	aactgtgg	cttaacccca	ataggagcct	gctgggatta	56735110
tgctacacaa	aa	ctattcca	cacagacaga	gccattctga	ccacgtccct	56735160
ccctacacca	gc	taacatat	agatatgcac	cactggatatg	agaagtccct	56735210
cctacagagg	at	tcaataat	aggctgattt	aggtcgattt	agtcattcaa	56735260
ccaacaaata	tt	tactgagc	actttctagg	tatttaggga	tacaacagaa	56735310
caaaacagac	aa	aatccctt	tccttatgga	gcttacgttt	tagtgaagaa	56735360
gagtggggga	gt	cagacatc	aaataggaaa	tgagttgagg	atatatttta	56735410
cgcagagcac	ct	ggaggcca	cttttgataa	ggtgattttt	gagcctcaat	56735460
ccaaagcagt	ga	gacagctga	gccatgcaga	cactagggga	agaatgttcc	56735510
aggcagagga	aa	cagcagg	cccagagaca	ggacagggtc	aagcatgcct	56735560
ggcatgttaa	ag	gaaacgcc	atgagctcag	tgttttataca	gaggagtaag	56735610
agcaggagaa	tag	caggaag	cacagtgaga	caggtagtag	ggggcaagat	56735660
catgtagaaa	ac	agaaagcc	actggagagt	ttgaacagag	gggtaaacat	56735710
gatctaaatt	cc	actgaaaa	aggactgttc	cagctgttgt	actgacagta	56735760
cagaggctgg	gt	gcgggtggc	tcacacctat	aatcccaatg	ctttggggagg	56735810
ccgaggcagg	cg	gatcactt	gaggccagga	gttcaagacc	agtctggcca	56735860
acatggcgaa	ac	cctgtctc	tactaaaaat	gcaaaaaaat	tagccgggaa	56735910
tggtgattcg	tg	cctgtaat	cccagctact	agggaggctg	aggcaggaga	56735960
atcgcttgaa	ct	gggaggc	ggaggttaca	atgagcagag	atcacaccat	56736010
tgtactccag	ct	tggcgag	agagcaagac	tctgtctcaa	aaaagaaaaa	56736060
aaaaaagaa	tag	accgaag	gaggcaagag	tggaacaag	aaaaccagtt	56736110
aggagactac	tg	agataaat	caggtgagag	atgatggcag	atgggtccag	56736160
ggtggcgagca	gt	gaggaatg	aagagggaca	gactcgggga	tatatctctga	56736210
agttgaatca	ac	aggacact	ggaggacagg	atgagcttgg	aaaatcttgg	56736260
gactactgca	ga	attgaatg	aagctaaatg	gcactaaatt	ccaacatcca	56736310
ctgtgaaatc	ct	gggatcag	ggatttgttc	atctctgtca	cctagtgtcc	56736360
caccagctac	tt	ggcacata	atgaacatat	aaaaagtcct	cttgaataaa	56736410
tgtggaattg	aa	atattac	aaagccaaaa	aacctgcagt	ttgataggct	56736460
attaggagac	ca	aaaagcat	ctggtactag	acatatttta	ctaacagagt	56736510
cttgccctgt	cg	cccaggct	ggcaatggca	tgatgtcagc	tcactgcaac	56736560
ctccgcctcc	ca	ggtcagg	caattctcat	tcctcagcct	ccccagtagc	56736610
tgggactaca	gg	cactcacc	actatgcccc	gctaattttt	ttgttttgtt	56736660
ttgttttttt	gt	atttttagt	agagacgggg	tttcaccatg	ttgccgaggg	56736710
tggtctcaaa	ct	ctctgagct	caggcaatcc	acccaccttg	gcctcccaaa	56736760
gtgctaggat	ta	caggtgtg	agccactgag	ctcagcctaa	atctattcct	56736810
ttttataaaa	aa	tttccctt	actatggccg	gacacggtag	ctcacgcctg	56736860
taatcccagc	ac	tttgggag	gccgaagcag	gcatatagtg	aggtcaggag	56736910
atcgagatca	tc	cttggtcaa	cacggtgaaa	ccccgtctct	actaaaaata	56736960

aaaaaataaa	aataaaaaatt	tcccttttata	actcattttca	cagtagtaac	56737010
caatgaagct	gacaaacaac	ttcacattat	acaggtgaga	tgtgaacaac	56737060
ttagccatct	ccctgagagg	ccagatgtag	tcaccaaacc	tcagcaactc	56737110
ccatctaaaa	tacagggata	agtgttacac	aagtagtaag	cccatgttct	56737160
tataaatagt	aaatcatggg	ccgggcacgg	tggctcacgc	ctgtaatccc	56737210
agcacttttg	gaggctgagg	tgggtggatc	acgaggtcaa	gagttcaaga	56737260
ccagcctggc	caacatgggtg	aaaccccatc	tctactaaaa	atacaaaaat	56737310
tagccaggtg	tgggtgggtgg	tgcctgtaat	cccagctact	caggaggctg	56737360
aggcagagaa	atgctggaac	ccgggagggtg	gaggttgtag	taagccaaga	56737410
ttgtgccact	gcactccagc	cagggcaaca	gagggagact	ccatctcaaa	56737460
ataaaataaa	ataaaataaa	ataaaataaa	ataaaataaa	tattaaaata	56737510
aagacataaa	ggatgtcact	tacatggaca	cgtgttcagg	gcttggtggtg	56737560
aggagttgtg	gatgaggacg	aaagtcgact	tggagacaat	ggagtccgac	56737610
tggagttact	ggatgaggaa	aggagagaaa	gagagagaaa	aggaaaagggtg	56737660
taacggaagt	ggaagattgc	ctcagtgtag	aggtcatctc	tgattgctga	56737710
caaaggcagt	ggaaaatggt	ggagaagggtc	tgaaaatatt	gctaaagggtt	56737760
ggtggtttca	gaaaatcttg	taacattttc	tttactttta	aacttagcat	56737810
ctcaagggtt	tttagtatga	ggtttttttt	ttttgttttt	tgttttttgt	56737860
tttgcccat	ggatttcaga	gtagttgaac	cattcagcca	cagaaaataa	56737910
ctccagtatc	aaccacattt	gaatattacc	aataaagaca	tgggactctc	56737960
ttcctctgag	gtcgagcgcg	tctattcact	tgctgctcag	cctcctcttc	56738010
tgcccaccct	cacctccttt	tctggcatct	gctcctccca	gggaggactt	56738060
cacaccctcc	tcacagagcc	agggtgccacc	tcccactctt	cgaggctcca	56738110
tcacgtgcct	ggctccacca	ccttctcgcg	tgtgacatct	acctttcacc	56738160
aagctgagga	cacagcacac	ccatggtgtg	tttaaggaga	cagttctgat	56738210
ggccacaaaa	ttcttttgtt	atgtggtctg	tgatttgaat	tgccctgtcc	56738260
aaagactaat	gcatggttca	tggtcaggcc	aagtaagggtt	atggttggat	56738310
gagaaaccac	caagcacctt	taattccttc	caagatcctt	taggagtggtg	56738360
acaaagtcaa	gtgctgttag	tgatggcaag	agaggagact	ctctctgctg	56738410
cactctggtg	atgatggcag	gtccagcact	gctacagcac	caatccttgc	56738460
agtgtcttct	atggcgcttg	cacattcctt	atataataca	ggctttgcag	56738510
tctcagtttc	tgcaatcact	gtcacaacta	cccctaacaa	aagcagtgac	56738560
tgtattcctc	acaaccttga	attttaagctt	tgtcttgaga	aatgaagggtt	56738610
tctacaggtt	tcttcatggt	cacattaatc	caaattgaag	atttacctcc	56738660
aggaaatatt	attcttctct	ttagcaaaaa	atgtgtaact	atgtgtaact	56738710
tccaatcttc	aattcttcac	caatgctttg	attcctttggc	taaatcaaaa	56738760
gcatgctgag	ggccaggcac	agtggttcac	acctgtaatc	tcagcacttc	56738810
gggaagctaa	ggcaggctga	tcacaagggtc	aggagttcaa	gaccagccgg	56738860
gccaacatgg	tgaaccccca	tctcttctaa	aaacacaaaa	attagccagc	56738910
atggtggtgt	gcacctgtaa	ttccagggtac	tgggaggctg	aggcaggaga	56738960
actgcttgaa	cccagagggt	gaagggtata	gtgagcaaaa	atcatgccac	56739010
tgcactccag	cctgggcaac	aaagcaagac	tctgtctcaa	aaaaaaaaaag	56739060
aaagaaaagaa	agaaaaaaca	tgcttaacct	atgagttaag	cactgagttt	56739110
tgtcttctcc	actgcaggac	tcctggacaa	ggattacttt	cttaacctac	56739160
tacagaattt	ctcaaccact	ttccatgcca	taatccacac	aaaatctatt	56739210
tggcacactg	actgaccttc	actaattgcc	ccaagggtact	gtgagagctg	56739260
aaggagcagt	attttaggac	acacccatta	acccaacttg	agaacattta	56739310
ttcactgtag	tattaatgag	aatgattaac	tgcacaaatt	tgtaaacacc	56739360
taatatgcac	acataattca	tacttgattc	ccaacactta	tttcttgaca	56739410
atacaaaact	tcatactaaa	tggctagtta	ttggttcttg	gcaacttaag	56739460
ttcctaagggt	ggtagataat	ctacaaatgt	ggtcacttac	tctgacagtg	56739510
caaaatggga	aacaaaaaga	tgggaaagaa	agaaagctct	aataaagagc	56739560
tctaacaaaag	aaaataggct	gggcacgggtg	gttcacacct	gtaatcctaa	56739610
cactttggga	ggccgagggtg	ggcagatcac	ctgagggtcag	gggtccgaga	56739660
tcagcccggc	taacatggcg	aaaccctgcc	tctactaaaa	atacaaaaat	56739710
tagttaggca	tgggtggcaca	ggcctgtaat	cccagggtact	tgggaggctg	56739760
aggcaggaga	attgcttgaa	cccaggaggc	ggagggttgcg	gtgagccaag	56739810
atcgcaccat	tgtactccag	cctgggaaac	aaagagcaaaa	cttcgtctca	56739860
aaaaaaagag	aaaaaaaaaa	aaagaacaga	acattatttg	ttcaatgggg	56739910
acttcttaga	tccttcctat	taggctgtg	tattcgtgtg	caggccagag	56739960
ttacagtatt	tgccatttta	ggaacagggtc	tcatatgaat	aaacattaag	56740010
gtccaaatgc	ggacagtaac	tgctatccct	gaacagttgt	gtggctactt	56740060
gaaactgaaa	tatgctgtga	gtatgaaaga	cacatcaatt	ctcaacaaaa	56740110
agaacatatg	caaatgaagt	aaagaatgtg	aaatacttca	ttgtgtattt	56740160

atttataatg	atttcaggtt	gaactcgaaa	atgtaaaatt	atgtatgtgg	56740210
gttgcatttg	cagcacacat	tgtatttctc	ttggatacca	ctggccaaag	56740260
ttcaaagctt	aatgccgggt	gcggtgactc	acatctataa	ttccagcact	56740310
ttgggaggct	caggcagaag	gatcacctga	gcctaggagt	ttgagaccag	56740360
cctggccaaa	atggtgaaac	cttgtctcta	cacaaaatac	aaaacttagc	56740410
caggcatggt	ggcgcacacc	tgtaatccca	gctactcagg	aggctgagga	56740460
aggagaattg	cttgagcctg	gttgacagtga	gctgagatca	caccactgca	56740510
ctccagcctg	ggtgacagag	tgagactctg	tatcaaaaaa	aggaagaaag	56740560
aaaaaaagct	tcagttatca	ctacactgca	tatgtgcata	ttgataaatc	56740610
aaggagtgcc	tgtattgagc	ttctgcaaaa	gatttttttag	ggaggtgaag	56740660
gtgcgatctt	ttccatcatt	ttaaaagatt	aattttttaa	agcttcaaga	56740710
ttaatctgtg	cttttaaaaa	ggatttcaca	actattttac	ccgttgtatt	56740760
tgttcttggc	tgagtaacta	cagttgtatt	gtgcataaga	agttctggac	56740810
cataagtaca	tagttgatat	taatattata	ataaagcatt	tagaagaaat	56740860
ttacaggcct	ttctcacaca	gctgtgcact	gctaattatt	attattatta	56740910
ttattattag	agacagagtc	tcgtcctggt	gtccaggctg	gagtacagtg	56740960
gtgcgatctt	gactcactgc	aacttccgcc	ttccatgttc	aagcaattat	56741010
cctgcctcag	cctcccaagt	agctgaaacc	acaagtgcgc	accaccacac	56741060
ctggctaatt	tttgtatttt	tagtagacac	ggggttttgc	aatgttggca	56741110
acactggtct	cgaactcctg	acctcagggtg	accacccgcc	tcagcctccc	56741160
aaagtgctgg	gattacaagt	gtaagccact	gtgcctggcc	aattttttaa	56741210
cataatccat	tctgccttca	ttttatagtg	aaaataatca	ttttggtgat	56741260
ttgaattttg	aaaatagcaa	ttttcagaca	ttcaggtgag	accagctgga	56741310
tgtctccaaa	gtcatctttt	tttaagataa	atatacaaaa	gtcacttttc	56741360
ccccaaacca	gtcgaattct	gtttctgaat	cttagatgaa	ctgtctcaca	56741410
gaagagccag	aagcagcttg	ctgtatcatc	ttaaaccttt	cctctccata	56741460
tataactcag	ttccaatgaa	tatacacaga	tgtacaacct	caatcagtaa	56741510
gcaagactca	accacttca	acattccctt	catgagtaga	gaggccttct	56741560
gttcctaccc	ccaacactca	aactacccat	gtattaccct	tcaaaattta	56741610
gccagaataa	cctttctaaa	gggcagatcc	aaattatttt	aagagggaac	56741660
tgctatgaga	ccatttaaatg	actgccactg	cttttgtagg	atctaaggcc	56741710
ttcctccatt	tggccaccca	agtctatcta	tgcaatccca	ccatgcccac	56741760
atggcactcc	tcgtccacaa	aagaatgggt	gaggtatgtg	aatatgtggg	56741810
catatttgac	atctccctga	ctttactcat	atttggaaca	cgttccctcc	56741860
ctctctggatg	cacaaatgac	tgatctggga	tctgggtatg	ccatccttgc	56741910
cttctccctt	aaaaatgctc	aataactctt	cattgcattt	ttcataatat	56741960
taagcttccc	aaggacaggg	ttcttgctag	acccaatttt	gcacatctga	56742010
tacttatact	aaatacatgt	ttgttcaata	agtgaatctg	aatgttaatt	56742060
gtaaatataa	ggaagtcata	atgagagggg	aacaaaaggc	ccttcctcat	56742110
cgaattaatc	tctccttagg	gaagacactg	aatttggttc	tgcaagattat	56742160
aatcttacct	tttaataaga	aatttaattc	gattgcccac	ttgaatgatg	56742210
ttttccagcc	ttgggattcc	ataccacgag	ggacttctga	aaagaatggt	56742260
ttctggtagt	ccttccacat	acagatcatt	cgcattggcct	tcaaattttt	56742310
ggtacagtag	agccttggct	tcagtgtctc	ccaaggcttc	agctgaacaa	56742360
aaccaaagca	attcagaaga	ctatgcttgt	gttgaaatca	gatttttagta	56742410
caatccaaag	caaggtgcag	gtcaattttt	ttaaattgca	aaaactctag	56742460
aaacatatatt	actaactact	gaagcacgct	taaagcaaag	gcaagtataa	56742510
atgacccatt	attagacaga	tcaatgagat	agaaaattaa	caagggtatt	56742560
aaggacttga	agtcagttct	ggaccaagcg	tacttaagag	acatctacag	56742610
aattctccac	cccaaatcca	cagaatatac	attcttctca	gcaccacatc	56742660
acacttattc	taaaattgac	cacataatta	gaagtaaaac	actcctcagc	56742710
aatatgcaaaa	gaacggaaat	aataacaaac	aggctctcat	tgcatagtgc	56742760
aattaaatta	gaactcaaaa	ttaagaaaact	cactcaaaaac	cgcacaaaact	56742810
acatggaaac	tgaacaacct	gctcctgaat	gagaactggg	taaataacga	56742860
aattaaggca	gaaataaaga	tgctctttga	aaccagtaag	aacaaagaca	56742910
caacatatca	gaatctctgg	gacacgttta	aagcagtgtt	tagagggaaa	56742960
tttatagcac	taaatgacta	atgcactaaa	tgtccacaag	agaaagcagg	56743010
aataatctaa	aatataacat	ttatgcagct	aacaaacatg	aaaaaaagct	56743060
catcactggg	cattagagaa	atgcaaatca	aaaacaatga	gataccatct	56743110
cacgtcagtt	aaaatagcaa	acaataaaaa	gtcaggaaac	aacagatgct	56743160
ggagaggatg	tggagaaata	ggaacgcttt	tatacttttg	gtgggagtgt	56743210
aaattagttc	aatcattgtg	gaagacagtg	taatgattcc	tcaaggatct	56743260
agaaccagaa	ataacatttg	accagcaat	cccattactg	ggtatatacc	56743310
caaaggatta	taaatcattc	taccataaag	acacatgcac	atgtatgttt	56743360

actgcggcac	tattcacatt	agcaaaagct	tggaaccaac	ctaaatgccc	56743410
aacaatcaca	gactggataa	agaaaaatgtg	gcacatatatac	actatgaaat	56743460
actatgcaga	cataaaaaag	gatgagttca	tgtcctttgt	aggaaacatgg	56743510
atgaagctgg	aaaccatcat	tctcagcaaa	ctaacacaag	aacagaaaaac	56743560
caaaaactgc	atgttctcac	tcataagtgg	gagttgaaca	atgagaacac	56743610
atggacacgg	aggggaacat	cacacaccag	ggcctgtcgg	ggaatgcggg	56743660
gttaggggag	ggacagcatt	aggagaaata	cctaattgtag	atgatgggct	56743710
gatgggtgca	gcaaaccacc	atggcacgtg	taacaaacct	gcacgttctg	56743760
aacatgtacc	ccacaactta	aagtacaata	aaaaataata	accaaactct	56743810
gcccccaatc	caaggccaca	aataatttaa	ttaaagccta	taaaatagga	56743860
cacacagaaa	agtcactctg	tttccttgac	tccacaactt	gtatgtgtgg	56743910
cttataatgc	atttacttta	ttatcagtct	caataaaactc	tgggggagaa	56743960
agccgggggt	aacacaccta	atcctttcat	gaaatgtcta	acatggccgc	56744010
gagtggtgtg	gatcttttct	caagccccc	tgcctgagtg	ccatttgctg	56744060
atggcccttc	cccattgattg	aggaaagctgg	gtcactgggt	gaacactaag	56744110
tacaagacag	catgagaatc	atgcttgacg	ctgcagttag	ctggagggct	56744160
ggtagaaggc	atgtgtccat	gattgtatat	accgagttct	gctatcttaa	56744210
tgatatctgt	gtatgttttc	tatctttgaa	gatacttgca	tttttacata	56744260
tcagaatccc	aaggctgaga	aataaaaatg	ttaaaaattt	aaaaaaaata	56744310
ataaaaataa	atgacccatc	attaccaacc	agattttggc	agattttaagt	56744360
gtcctgatcc	taagacaatc	ttaataactc	taagatctta	aaagattaaa	56744410
ctttgttttt	caatattctt	aggaaagaaga	gaaaaaaact	atagacaata	56744460
aagaacacaa	cctgaatact	atttgttttt	ttttgttgtt	gttttgtttt	56744510
gttttttttc	gagacggagt	ctcgctctgt	caccagggt	ggagtgcagt	56744560
ggcatgatct	tgggttcattg	caagctccgc	ctcctgggtt	caagtgatcc	56744610
tcttgccctca	gcctcctgag	tagctaagat	tacaggtgcc	cgccaccacg	56744660
cctggctaatt	ttttgaactt	ttcagtagag	acgggggttt	accatggttg	56744710
ccaggctggc	ctcaaactcc	tgacctcaag	caatctgctc	accttgacct	56744760
cccaaagtgc	tgggattaca	ggtgtgagcc	accacagctg	gcctgcaaata	56744810
actatttgca	aataactattt	ggaagcagcc	ttttgccact	aactgcttgt	56744860
gaaattaaca	accctcctcc	actggctgct	gcaaacaggc	acatttcagg	56744910
ctgaattttc	atccacactg	caaccacgag	atcaatgttc	tttaccctt	56744960
caacctcacg	agtatgctat	tttaaaagaa	ataactttcc	acatagaatt	56745010
catgaaatgg	gaaataaaaa	tagtaagcag	aggccagttt	agattaagag	56745060
gcataacaac	caaagttagc	acatgaagat	tgtttgacc	cagatttaaa	56745110
gcaaaatttt	tcagataatg	aggaaatgag	atactgttca	ggtaagagta	56745160
aggaaacttg	gccaggcacg	gtggctcacg	cctgtaatca	cagcacttgg	56745210
ggaggacgag	gtgggaggat	tacctgaggt	caggagtttg	agaccagcct	56745260
ggtcaacatg	gtgaaacctt	gtctctacta	aaaatacaaa	aatttagccg	56745310
gcatggtggc	gtgcacctgt	aatcccagct	accaggagg	ctgagttagg	56745360
agaattgtctg	gagcccgagg	ggcagaggct	gctgtgagct	gagactgcac	56745410
tccagcctag	ggtgacacag	caagactttg	tctcaaaaaa	aaaaaagagt	56745460
gaggaaactc	aactaagggt	tacagatgag	ctgacaccaa	aaaatcattg	56745510
taattttgtc	agatgtgatg	agatagtagc	tgtataagaa	aatgttaatg	56745560
ctgttcaggg	atgcacactg	aacagataaa	atgatgtgat	gactgatttg	56745610
ctttacaata	ttcctgcaaa	gggagggtaaa	ggaatagaag	aacaaaatgt	56745660
aggacaatct	taatgatcac	agaaattgag	gaatgggtat	tttagtggtc	56745710
actgtactat	tttctctact	tttgaatagg	tatgaaaact	gtaattttaa	56745760
aatgtaaaaa	aaggggcagg	agcgatggct	catgcctgta	atcccaacac	56745810
tttgggaggc	tgaggcaggt	ggatcacttg	ggtcaggagt	tcaagaccag	56745860
cctggccaac	atggtgaaag	cctatctcta	ccaaaaaaaa	atcagctggg	56745910
tgttgtggca	ggcacctgta	gtcccagcta	ctcagaaggc	tgaggcaaga	56745960
gaattgcttg	aacctgggag	gcagaggctg	cagtgaagca	agatcgtgac	56746010
actgcactcc	agcctgggag	acagagttag	aatccaccaa	aaaaaaaaaa	56746060
aagtaaaaaat	gaattgacta	atccctgctc	tgaagccaca	tccaaacaaa	56746110
actctaaaaa	gaaagattca	agcaatttat	atcgtgctat	ttatatatag	56746160
atctcccttg	cacctctgta	tgtaacattt	aaatatattt	caacctgata	56746210
gttgggagtg	gctcaaagct	gggcatgggt	gcatgtgcct	gtagccccag	56746260
ctaattagga	gactgaggca	ggaggataac	ttggggccag	gagttctagt	56746310
ccatcccaga	caacataaga	ccctgtttct	ttgtttcttt	tcttttcttt	56746360
ttcttttttt	ctttgttttt	tttttttgaga	tggagtttca	ctcttgttgt	56746410
ccaggctgga	gtgcaatggt	gcaatctcgg	ctcactgcaa	cctccgtctc	56746460
ctgggttcaa	gtgattctcc	tgcctcagcc	tcccaagtag	ttgggattac	56746510
aggcgacacg	caccacgccc	agctaattga	gtatttttag	tagagatgcg	56746560

gtttctccat	gttggtcagg	ctggtctcga	acccccact	tcaggtgatc	56746610
caccgcctc	agcctcccaa	aatgctggga	ttacaggtgt	gagccaccgc	56746660
atccagtgga	ccctgtttct	ttaaaaagaa	aagtcgggga	gctgtcaggg	56746710
agctcattga	tgaattatct	tgggtgatatc	aacaattgct	atcaaaacac	56746760
aactataaaa	ttcaaaatta	tagggcgggc	acgggtggctc	acacctgtaa	56746810
tcccagcact	ttgggaggcc	gaggcaggcg	gatcatgagg	tcgggagatc	56746860
gagaccatct	tggctaacac	ggtgaaaccc	cgtctctact	aaaaatacaa	56746910
aaaaaaattc	gccgggcatg	tgggtgggcac	ctatagtccc	agctacttgg	56746960
gaggctgagg	caggagaatg	gcatgaaccc	aggaggcgga	acttgcaagt	56747010
agccgagatc	gcaccactgc	actccagcct	gggtgacaga	gagagactcc	56747060
gtctcaaaaa	aaaaaacaaa	acaaaaaaac	aaaatccaaa	attatatcta	56747110
ttgccagatc	tttttcctta	cttccagact	tatttatcca	actgtctcct	56747160
tcacctccac	acctcaactt	aaatgtcaac	ttgaaattaa	ttaataatct	56747210
aattatctat	tactaattaa	tctttgttct	tagattggct	ttcctctaac	56747260
atctcaagat	ttgtcttaaa	ttcttacttc	atgatattag	caccaaactc	56747310
gaagggttct	agtacaggaa	ctctctgaga	tcggaaacac	atgtatcttt	56747360
gagaaaccac	acaaaatgaa	ataattatca	tagctgtaaa	aagaaatatg	56747410
ccataaaca	aataatagta	aaggattaga	ggttttcagt	cacactagag	56747460
agtgatgtta	aataaaatta	ttctgttaat	tcaacaagga	aatatataaa	56747510
cctaattcta	acaaagtttt	tataaaataa	aaaccagttg	gagttgagag	56747560
tgaaaagtgg	aagtttagaa	agaggcgagg	ccaactacag	aacgtgctca	56747610
tcacacatca	tcaaggctgc	cccaatcaca	ccccctgcct	ccccattccc	56747660
cttattccac	ttaatctttt	cccagacctt	ataccttctt	acagagtaaa	56747710
catctcctac	ttactacacg	tttccccacc	agattgaatg	ctccaaaatg	56747760
cacggggcag	gccactgtct	tctgcaactgc	ccagctcccc	ccaccagca	56747810
cctgactcac	aggtgccaaa	tgactgcagg	attatgaggt	tggaagtagt	56747860
gaatatcttc	tgaataactg	cagtggttcc	cagcaatgac	aaatctaaga	56747910
tttcctctct	ccctttcagg	aatcttagag	aaagcagcaa	tactgtcaga	56747960
gggtgttgaa	cgagaacaac	tcgatcttaa	atagaagctg	ggtaaaatga	56748010
ggctgagacc	tactgagctg	cgttcctaga	tggttaaggc	attctaagtc	56748060
acagggtgag	agaggagatc	agcacaagat	acaggtcata	aagaccttgt	56748110
tgataaaatg	ggttgcagta	aagaattcag	ccaaaaccca	ccgaaaccaa	56748160
gatggtgaca	tgagtgcact	ctggttgtcc	tcaatgctac	actcccacca	56748210
acgccatgac	agtttacaaa	tgccatggca	acatcaggaa	gttaccctat	56748260
atggtctaaa	aaggggaaac	ataaaaataa	tccactcctg	gccaggtgca	56748310
gtgggtcaca	cctgtaattt	cagcattctg	ggatgccaa	gtaggtggat	56748360
cacctgaggt	caggagtttg	agaccagcct	ggcaacgtgg	taaaaccccg	56748410
tctctactaa	aaatacaaaa	acaaattagt	tgggtgtggt	ggtatgcacc	56748460
tgtgatccca	gctacaaggg	aggctgaggc	aggagaattg	cttgaaccca	56748510
ggaggcgagg	gttgcaagtga	gccaagatca	caccactgca	ctccggcctg	56748560
gagtcgaatc	caaaaaaaaa	aaaaaaataa	gtaagagagg	ctgggtgttg	56748610
tggtctcaca	ctgtaatccc	agcacatggg	aggctgaggt	gggtggatca	56748660
cctgaggtca	ggaattcgag	accagcttgg	cccacaaggt	gaaactacat	56748710
ctctactaaa	aatacaaaaa	ttagctgggc	atggtggcag	gtgccagtaa	56748760
tcccggctac	ttgggagact	gaggcaggag	aatcgcttga	atctgggagg	56748810
cggaggttgc	agtgaccgca	gactgcgcca	ctgcacccgc	ctgggtaaca	56748860
tgagcgaaac	tccgtctcaa	aaacaaaaca	gtaagaaaaa	actggagctg	56748910
agtcctgctc	attatattta	ctttcctaaa	ataaaacaac	attataaaaa	56748960
acattactca	gtaataatac	tgagttgaaa	ggggaaaaaa	gagacttacc	56749010
aaatcttttt	gcagaaaagc	tgatccacca	tctttcttca	tttagtgatt	56749060
ctggcatacc	attcttcctt	tactgtcaaa	atcacaggaa	tacaagtttt	56749110
ctcaatttaa	tgggtttaat	agcttggggt	attaaatatt	attattataa	56749160
aataggcggt	attatttttt	cttttcctgt	gagtaaaatt	atttctaaga	56749210
tccagaatgc	atttactact	gaaacaatgc	tgacacacac	tgacaaaaaa	56749260
aggaagaacc	agacattcat	taaagctggg	aaaaagcaga	actcactgtg	56749310
ctcaagaacc	actagcaggc	tgggtgcagt	ggctcacacc	tgtaatccca	56749360
gcactttggg	aggctgaggc	gggcagatca	cctgaggttg	ggagttcaag	56749410
aacagcctgg	ccaacatggt	gacacccgca	ctctactaca	aatacaaaaa	56749460
tcaggcaggt	gtggtggcag	gtgcctgtaa	tccagctac	ttgggagcct	56749510
gaggcaggaa	gaagtgcctg	aacctgggag	gtggagggtg	caatgaattg	56749560
agatcacacc	actgcactcc	agcatgggca	acaaagcaaa	actccatctc	56749610
aaacaaacaa	acaaacagct	agccacttaa	cagatgcctc	aataaatcag	56749660
agtctaattc	tcttaaaaaa	aaaaaaaaaa	agacataaga	gatatacagt	56749710
tgcaagttag	aatgctaaag	aacataaaca	tagaacaact	taatgtacat	56749760

ataaattcaa	tatatatcca	atcattgtga	ctgacacagt	aaaatattaa	56749810
aatactat	tcaaaatgta	tacaagctta	atgttctatt	tattcaaact	56749860
atttatttaa	aatacaaatc	atcaacataa	tttgccacta	aaatctcatg	56749910
cagtccttc	acaggacaca	tgatccactg	ggagttaata	aattagcagc	56749960
tggcaggcag	tgacacacag	caaaaatgaa	aaccaagagg	tgaaacagtt	56750010
ttgaaataaaa	gaaggtttta	aagctaagag	aaattattga	attactaagt	56750060
cattagcact	aattttgagc	caactaatta	atatgagatg	atacaacgtc	56750110
ctatgctttg	gtaaattcaa	actatgttta	aacactgtct	gtaatgtgac	56750160
tttcaaaatg	ctcctggctt	tacaaagatg	agattataat	gtagtaatac	56750210
atgctaaagc	atttccccct	gcagaacatg	ttctaacttt	catcagtcac	56750260
tattgagagtc	cagaagatga	aggaaaaggt	catggatttc	gctgagaact	56750310
taccagagct	gaactccctc	actttccgtt	ccccagcatt	ggcaggttcc	56750360
aggactggtg	gctgtggtgg	ctcggttggtc	tttgttctct	tagaagggtg	56750410
agaataatca	tcatcttgaa	aaagaaaatg	gtgattattg	aaggaaccat	56750460
cttagagtta	cagccacctg	cgggtcaaca	acattcaaaa	gctgaacggg	56750510
gctttaaagc	tattttgagt	attaataatt	atctctgtat	tgtgatcttc	56750560
agcatacttt	tttctagttt	catttgaaat	tttattcttt	tgggatgtgc	56750610
tcaagcgaat	actgcttttt	cctgtcttgc	ttcattattt	tttagtttgc	56750660
ttcatttgag	tcatcattat	aaatctcccc	ttctctctca	ataactttca	56750710
aattgatgcc	agaactata	ttctattttc	aggctgttga	gaaaaaactt	56750760
tcaatgaaga	cagccaccta	aagctataca	aatatagaag	aaaggggata	56750810
aaataaagct	tagattggaa	acaaatttaa	gattatacaa	aattcaagca	56750860
tacataaaca	agggagctg	gataattata	tgttcaaata	cttttaacac	56750910
gtgcaaaatg	tgtagactta	aagaaatagg	gctggccagg	cgtgggtggc	56750960
cacgcctata	attccaacac	tttgggaggc	caaggcaggc	ggagcacttg	56751010
aggtcaggaa	ttcaagacca	gcctggccaa	catggtgaaa	ccccatctct	56751060
actaaaaata	caaatattag	gctgggctgt	gtggcttatg	cctgtgatcc	56751110
cagcactttg	ggaggccaag	aagggtagat	cacctgaggt	cgggagtttg	56751160
agaccagcct	gaccaaata	gagaaaccca	atctctacta	aaaatacaac	56751210
attagccagg	catggtggtg	catgccttga	atcccagcta	ctcgggaggc	56751260
tgaggcagga	gaattgcttg	aacccgaaag	gcaaagggtg	tggtgagccg	56751310
agatcacacc	attgttgccc	agcctgggca	acaagagtga	cactctgtct	56751360
caaaaaaaaa	aagaaaacaa	aaattagcca	ggagtgtctg	cacatgtaat	56751410
cccagctact	cgggaggctg	aggcaggaga	atcgcttgaa	ccaggggaggc	56751460
agaagttgcg	gttagccgag	atcaacacca	ctgcactcaa	gcctgggtaa	56751510
cagagcaaga	ccttgtcgaa	agaaaaagaa	aaagaaaagg	aaagaaagga	56751560
aagaaaagaa	atagggctac	attatattatg	aaacagatac	tggtgactca	56751610
gtcactagaa	agtctgtgta	taaatgagca	gtgagatatt	caagaacagc	56751660
acacacacac	tactcaggac	agctattgtg	agagttccat	gctcatttcc	56751710
ttgtggatac	accagcaact	cactctgtcta	tgatcctgca	atacatctca	56751760
tgttagcatt	agagacactc	gggccaggcg	cgatggctca	cacctgtaat	56751810
cccaacactt	tgagaggccg	aggcaggcac	atcacctgag	gtcaggagtt	56751860
ccaggccaga	ccggccaaca	tggtgaaacg	ccgtctctac	caaaaataca	56751910
aaaattagcc	aagcgtggtg	gcgcgtgcct	gtaatcccag	ctactcagga	56751960
gccagaggca	ggagaattgc	ttgaacccag	gaggtagagg	ttgcagtgag	56752010
ccgagatcgt	gccactgcac	tccagcctgg	gcaacacagc	aaggatccat	56752060
ctcaaaaaaa	caaaaaataa	aaaagaaagaa	aaagaaagaa	aaagaattag	56752110
agacatctgg	atcaaatcag	ctgccactgt	catcttccag	tcctgcaaag	56752160
agtcagataa	cgtcctatta	agattgtctg	tcacacatca	tctataaaat	56752210
actgaaaata	tcattttaag	aaatcttttt	ttaattttga	gagggagttt	56752260
cgctcttggt	cccaggetga	gtgcgannnn	nnnnnnnnnn	nnnnnnnnnn	56752310
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56752360
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56752410
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56752460
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56752510
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56752560
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56752610
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56752660
tagtagagac	agggtttctc	catggttggtc	aggctggtct	tgaactccca	56752710
acttcagggtg	atccaccac	ctcagcctcc	caaagtgtctg	ggattacagg	56752760
cgtgagccac	cgcgcttggtc	caagaaatcc	ttctttaaag	aaaaggccag	56752810
gctcggtggt	tcacgactat	aatttcagca	ctttgggagg	ccaagggtggg	56752860
tggtacactt	gaggtcagta	gtttgagacc	agcctgggca	acatgatgta	56752910
accccatctc	tactaaaaat	aaaaaaaaca	aacaaaaaaa	attagccggg	56752960

tgtagtgggtg	ggcacctgta	atcccagctt	ctcaggaggc	tgaggcagga	56753010
gaatcgctta	aacctgggag	gcggagggtg	cagtgagtgg	agatcatgcc	56753060
actgcactcc	agcctgggca	acaacagcaa	gactccatct	caaaaaacac	56753110
aaaaacaaaa	acagactcca	tctcaaaaaa	aatcaaaaaa	caaaacaaaa	56753160
aacaaatgtg	gtatggaaat	gaaaatactg	tgtaaagac	agtttcatgc	56753210
aaaataaaaag	accactcaaa	tacaataaag	tgtcttttta	gatgggtata	56753260
attgttattc	ttattttaaca	gctaaagaaa	caggctcaga	gaatgttatt	56753310
tgattggact	gtgttgcatc	tctgaacagt	gcagctgaga	tcagacttca	56753360
tgtgtaactc	cactacccta	ccaggatgcc	tctcataaag	gtaagaaatg	56753410
taaatttggc	ctaataatac	aagttgccag	ggcagcactg	ggtcagttct	56753460
acataatagta	cttctatgct	catcaacgga	aactttaagg	gaagggtgaa	56753510
atgcttctag	aaggcgactg	gacaccagga	cctttgggct	cttttctaag	56753560
gccaatagtg	acctaaatta	ttgactgact	gcttcaatca	agtggacaaa	56753610
aggggtaccaa	ggccgccaac	atcagacaaa	ttcacttgag	ggcctatcta	56753660
tgtgctttga	aagacaaaacc	tgctgtttgtg	aaggacaccg	tatttcagaa	56753710
aaacataatc	atattaacaa	ctagtaacaa	tgtaaaatgc	tgatgtgttg	56753760
aatgctactt	tagaaaaacg	tgtaaaatgc	tacgaaaaaa	atztatgata	56753810
caaaaactacg	ttatcaatca	tctagctagc	taactatcta	cagacatggg	56753860
ttcatttctgt	tgctcaggat	ggaaagcagt	gggatgatca	tagctcactg	56753910
cagccttgag	ctcctggcct	caagtgatcc	tcctgcctcc	taagtagctg	56753960
gggccacagg	tggacacagt	gacacctggg	ttttttcctt	tgtagagaca	56754010
gggtttcact	acattgcccc	ggctgggtgc	aaactttgga	gtctcgctgt	56754060
gtcaccagg	ctgggggtgca	gtgggtggat	ctcgggtcac	tgcaacctct	56754110
gcctctcggg	ttcaagagat	tctcctgcat	cagcctccca	agtggctggg	56754160
actacaggca	tgtgccacca	cacccggtca	atttttgtat	ttttttaga	56754210
gactgggttt	caccatgttg	gccaggctgg	tctcaaactc	ctgacctcat	56754260
gtgatccacc	cgctcgggcc	tcccaaagtg	ctgggattac	agtcatgagc	56754310
cgatgaacac	tttcttatgc	tattaaatgg	cctaaccag	gcgggtcgtg	56754360
gtggctcacg	ctgtaaatcc	cgaaaactcg	gatggccaag	gtaagaagat	56754410
cacttgaacc	taggaattcg	aaactggcct	gggcaacata	gcgagactcc	56754460
catccctaca	aaagatacaa	aaattaggcc	aggcgcacac	cacgctcggc	56754510
tagtttttgt	atcttttgtta	gagacggggg	tgcgtcatgt	tgcttagggc	56754560
gggtctcgaa	tcctgaaccc	aagccatcca	tcctcccgc	tcggcctccc	56754610
aaagtgtgtg	gattacaggg	cccagccagc	ctcatgtttt	ctttaagcac	56754660
tcctccctc	ttgcacactt	ggttagtttt	ctttttaatt	ttttagacag	56754710
ggtttacctg	agtctcgcag	gctggagtgc	tgtgggtggga	tcacagctca	56754760
ttggagcctt	gaaccttggg	gttcaagtag	ctgggggggt	gaggtaggac	56754810
tacagagatg	gggtcccgc	atgttgccag	gctgctcttg	gcctgaaggg	56754860
cttctcccgc	tggccgcgc	cggacatagt	tttctatatt	tgaccgacat	56754910
aaacactgtg	ctgagtcggc	gtttgtcaac	acacaggacc	tggcggggag	56754960
gtcgcgggta	cccggctcca	ctctaagtag	aagactgccc	agctccaagc	56755010
actgtacctc	ccctgtgacg	tcgcggaacg	ccgctctgt	gacgatactg	56755060
aaggcccacc	ttcatgacgc	cgcggaaggc	ccgcccctgt	gacgcccgcg	56755110
gaggcgcgcc	cttgtgacgt	cgcagaggcc	cgcccctcac	gcggagccaa	56755160
tcggaactcg	aggcggggct	gttgggtctt	cgggagcgcg	catgcgcggg	56755210
gggccacagg	ctccctggct	ggcggcagct	aaagtggctg	ttgggtgtcc	56755260
gcaggcttct	cctgcgcgt	cgcgccttat	aagctaccag	gaaggagcttt	56755310
<u>acgacttccc</u>	<u>ggcctgcggg</u>	<u>aaagtgcggg</u>	<u>cacgatccga</u>	<u>aggtagccga</u>	56755360
<u>gaagcttctc</u>	<u>aatggccagc</u>	<u>gccagctgca</u>	<u>qcccctgcgg</u>	<u>cgcactcgc</u>	56755410
<u>tcacctgagc</u>	<u>ctgggtacgt</u>	<u>gcgccccaca</u>	<u>acaccttccc</u>	<u>cagccagggc</u>	56755460
<u>ccggggaccc</u>	<u>cgggagcgtc</u>	<u>ccccgccacc</u>	<u>tggcgccgct</u>	<u>catcctgggc</u>	56755510
<u>aggggtgggac</u>	<u>cccactgagg</u>	<u>ccgccacgca</u>	<u>ttagggagct</u>	<u>gcactcccga</u>	56755560
<u>gtttgacggc</u>	<u>cagttgtaat</u>	<u>agcattaaat</u>	<u>ctttgaaatt</u>	<u>ttgtagcggg</u>	56755610
<u>gtagaagggg</u>	<u>ctaggaaagg</u>	<u>aagaaaacat</u>	<u>cttttttaaaa</u>	<u>atataagcga</u>	56755660
<u>tcggccgggg</u>	<u>gcgggtggctc</u>	<u>acgcctgtaa</u>	<u>tcccaacact</u>	<u>ttgggaggcc</u>	56755710
<u>gaggcagggtg</u>	<u>gatcacgagg</u>	<u>tcaggagttc</u>	<u>aagaccagcc</u>	<u>tggccagcat</u>	56755760
<u>ggtgaaaccc</u>	<u>tgtctctact</u>	<u>aaaaatacaa</u>	<u>aaattaaccg</u>	<u>ggcgtgggtg</u>	56755810
<u>cgggcgcctg</u>	<u>taatcccagc</u>	<u>aactcgggag</u>	<u>gctgaggcag</u>	<u>agaactgttt</u>	56755860
<u>gaacccagaa</u>	<u>ggcggagggt</u>	<u>gcagtgcgtc</u>	<u>gagatcgctg</u>	<u>cactgcactc</u>	56755910
<u>cagcctgggc</u>	<u>gacagagcga</u>	<u>gactccgtct</u>	<u>gaacaaacaa</u>	<u>atatgtgtat</u>	56755960
<u>gtatataagt</u>	<u>gatcgaaccc</u>	<u>gggaggttga</u>	<u>gtttacagta</u>	<u>agccgagatt</u>	56756010
<u>gcgccactgc</u>	<u>actccagcct</u>	<u>gtgtgacaga</u>	<u>gtgagaaaat</u>	<u>gtctctaaaa</u>	56756060
<u>aattatatgc</u>	<u>aagtgcagagc</u>	<u>tttttttcca</u>	<u>gcgctcatgc</u>	<u>tcagactgaa</u>	56756110
<u>gaaagtaatt</u>	<u>gggccggggc</u>	<u>gttagctcat</u>	<u>gtccataatc</u>	<u>ccagcacttt</u>	56756160

gggaggccga	ggcaggcgga	tcgcttgagc	tcgggagttc	cagactagct	56756210
tgggcaacat	ggtgaaaccc	tgtctctata	aaaatacaga	aaattattgg	56756260
gcatggtggc	atgcgcttgt	agtccccgct	acttgagcgg	ctgaagccga	56756310
aggatcgacc	tcctgggctc	aggcaatcca	tttcagcctc	ccaaattgct	56756360
gggattacag	gaatgagcca	tcgtgcctgg	ccttacacta	tattttaatg	56756410
ctttttttga	aaatggaaac	ttttacaggc	aattcacttt	cttcaaacta	56756460
atgataagga	aatgatgctg	ttctgttctg	ttttgttttt	tggttttgtc	56756510
tgtttctttt	ttttccttta	tcgagacagg	gtcttgctct	gttgcccagg	56756560
ctggagttag	gcggtgaaat	catggctcac	tgcagccttg	acctccgggc	56756610
tcaaggaatc	gttcccggtc	agcctcccca	gtagctagga	ctacagggtg	56756660
atgctaccat	gcttggttaa	ttcttttttt	gaaatggagt	ctcactctgt	56756710
ctcccaggct	ggagtgcagt	gggtgcaatct	cagctcactg	caggctggtc	56756760
tcaaactctg	acttcagatg	gtccatccac	ctcagcatcc	caaagtgtct	56756810
ggattacagg	tgtgatccac	cacgcccac	gattttgtct	attttagata	56756860
ctacaacttt	ttagttta	tttttttttt	cctgagatag	aatctcactt	56756910
tgctcgccagg	ctggagtgc	gtggcgtgat	ctcagctcac	tgcagcctcc	56756960
acctcctggg	ttcaaggat	tctcctgcct	cagctcctcg	agtagctgag	56757010
actacagggtg	cttgccacca	cgcccaggag	ttcaagactg	agcccaggag	56757060
ttcaagcctg	cagttagcca	tgatcgtgcc	actgcactcc	agcctgggca	56757110
acaagcaaga	ccctgtctcc	acaaataagt	aaataaacat	cctatctgta	56757160
ccagggtaca	ggcaggcgct	attattccca	cttttcatcc	tcaactctga	56757210
ggtgagtc	acatcaactc	tgatgacctt	tgtttccact	ccccttaagc	56757260
cattcacagg	aacagctact	ttacatagaa	aagtcaatct	tggagggttg	56757310
agtaattttg	ctaactcactg	taaaaggtag	agcttcaatt	atcaccagg	56757360
cccatctgtt	gctgggtattc	cctctccacc	cctccgtgtg	gcttcccttt	56757410
atttatttat	tttaattttt	tttgagatgg	agttgcactt	ttgttgccca	56757460
ggctggagtg	caatggcatg	atctcggtc	actgcaacct	ccacctccca	56757510
ggttcaagca	attctcctgc	ctcaggctcc	caagtaggtg	ggattacagg	56757560
tgctgccac	caagcccagc	taattttttg	tacttttagt	agagatgggg	56757610
tttcaccatg	ttgaccaggc	tggtcttgga	ctcctgacct	caggtgacct	56757660
acctcctca	gcctcccaaa	gtgctgagat	tacaggcgtg	agtcacgggtg	56757710
cctggccttg	gctttccttt	aataaatcac	gactcaattt	gatgtggatg	56757760
ctctgttcat	tcccagaata	taatgaatgg	attatatata	tattttttgt	56757810
ttttgagaca	gagtcttgct	ttgtcttgca	ggctggagtg	cagtggcgcg	56757860
atctcagctc	attgcagctc	ctgectccca	ggttccagca	attctcctgc	56757910
ctcagctctc	tgggtagctg	ggattacggg	cactcaccac	cacgcccagc	56757960
taattttttt	tttttttttt	ttttttttta	gtaaatttca	ccgtgtcgtc	56758010
caggctgggtc	tcaaactcct	gacctcaggt	gatctgccc	acttggcctc	56758060
ccaaagtgtc	gggattatag	gcatgaccca	ccgccccag	cctcaacatc	56758110
aaataaattt	aaaagtgggt	aacagccagg	cgcagtgggt	cacacctgta	56758160
atcccagcac	tttggggggc	caaggcgggc	agatcacctg	aggtcnnnnn	56758210
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56758260
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56758310
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	56758360
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnagtcagga	56758410
gttcaagacc	agcctggcca	acatggtgaa	accccgctct	taataaaaa	56758460
aaaaaataaa	aaaaacagtt	agctgggtat	ggtggcatgt	gcctgtaatc	56758510
ccagctactc	aggaggctga	ggcacgagaa	tcactagaac	ctaggagggtg	56758560
gaggttgcag	taagccaagt	tcgtgccact	acctccagc	ctgggcaaca	56758610
gagtaagact	ccatctaaaa	aaaaaaaaaa	atgaagaaga	agaaattagt	56758660
gtagtgtggg	aagtgaaaaa	aaaaaaaaaa	aaaaggaaaa	gaaaaatgat	56758710
tgaattcctg	aatacactct	tatgtggcct	gcaccgactt	tgacacaaat	56758760
tagattggct	tagtagggcaa	gggtgggac	ttttcataat	tttatttgat	56758810
gtctaaaata	catttatctt	tttttcttat	aggaggaaaa	ttcttccaag	56758860
gatgatctcc	cactcagagc	tgaggaaqct	tttctactca	gcagatgctg	56758910
tggtgtttga	tggtgacagc	acggtcatca	gtgaagaagg	aatcggtgag	56758960
ctagccaaaa	tctgtggtgt	agaggacgcg	gtgtcagaaa	tgtagggata	56759010
gcatttatct	accttatgaa	atgataaaga	atctcttttt	tttttttttt	56759060
tgagacagag	ttgccagac	tggagtgcag	tggagcaatc	tggcacaatc	56759110
ttggctcact	gcaacctcca	cctcccagg	tcaagcaatt	atcctgcctc	56759160
agcctcctga	gtagctggga	ttacaggcgt	gtgcaaccac	atccggctaa	56759210
tttttttatt	tttagtagag	atggagtttc	accatgttgg	tcaggctggg	56759260
ctcaaactcc	tgacctcgtg	atccacccgc	ctcagcctcc	caaagtgtct	56759310
ggattacagg	catgagccac	catgcccggc	catgaaatga	taaagaattt	56759360

ctaaagagtg	gctgtttttg	atgaagtgtc	ggaccttggc	tatggaacat	56759410
gagcactagg	ctttttcttt	caccccttag	agttgaatth	gatataattca	56759460
ttcattttata	taaatatttta	tggcaaaaaa	atatggataa	gacatgtttct	56759510
tagcttgatt	aggggaatga	tgaccatcac	taaccaaaga	agatggccta	56759560
gggggcaggg	ctcaagtgtc	cagtggcttt	gtgtcatggt	tgaaagtgtg	56759610
gccttagtgg	ttccctggaa	agggttcagag	tctcttgggtg	gatcctgggt	56759660
cagagccct	ctctccctcc	ctccctcctg	ccctcccacc	tctgcccctg	56759710
cagctgggca	ccaccctctg	cagccccagt	cccccagtea	tgaccatgt	56759760
cattttcttt	tttttttttt	tcaggacgga	gtctcgtctt	gtcaccaggc	56759810
tggagtqcaa	tggtqcaatc	tcggctcact	gcaacctccg	cctcctgggt	56759860
tcaagcaatt	ctcctgcctc	cacctcctga	gtagctgaaa	ctacaggcac	56759910
gcaccaccac	acccgqctta	tttttqtatt	tttagtagag	atggtgtttc	56759960
accatgttgg	ccagcctggt	cttgaactcc	tgacctcgtg	atctgcccgc	56760010
ctcggcctcc	caaagtgtctg	ggattacagg	catgagccac	cacaccggc	56760060
caccatttta	ttttcaactc	cctttcatgg	aagaacgttt	agcctttggt	56760110
ctctttcttg	atltatgaca	gcttgccggt	ttcaagaaaa	ctacctatga	56760160
ataggtctgtg	taacttttat	ttattttatt	atlttttttga	gatgaagtct	56760210
ctgtcttcta	ggctgggggtg	cagtggcatg	atctcggctc	actgcaacct	56760260
ccccctccca	ggttccagca	attctcctgc	ctcagcttcc	caagtagctg	56760310
ggattacagg	cacgcaccac	cacacctggc	taatttttgc	atlttttttt	56760360
attagagatg	gggttttgtc	atgttggtcca	ggctgggtctc	gaactcctgg	56760410
cttcagatga	gctgcccacc	tcagcttccc	aaagtltttg	cattataggc	56760460
atgagtcact	gcacccagct	tgtaaatgtt	atltttcaagc	acaccttcaa	56760510
aagtlttatc	caaggccttg	ctctcatagc	agcaaagcct	gtctgttgca	56760560
tagtggggct	cttgtttagca	tcttctcttt	ggtggatgtt	gtgtagaccc	56760610
ggggcagtaa	ggcatgttaa	cctcgaggac	attggacctg	gctggcctga	56760660
ctgttgggct	tccctcctag	gacatggcga	gccatgggtg	gggcagtgtc	56760710
tttcaaaagt	gctctcacgg	agcacttagc	cccaatccag	ccctccagg	56760760
agcagtgcca	gcagactata	gcagagcacc	ccccacacct	gacccccagc	56760810
ataagtaaga	ggagccgctg	ctccagggtg	atlttcagcac	cagtgttggtg	56760860
gggcatgtcc	tcttgagagc	atctaacgat	tgcttttagag	agccctctgg	56760910
gtgggtttatt	tatttttcatt	tttatttttt	aattttatta	ttttaaaaaa	56760960
atlttatlttat	ttattttatt	atlttatlttat	tgaggcagag	tctcgctctg	56761010
ttgcccagac	tagagtgcaa	tgctacgatc	ttggctcact	gcaacttctg	56761060
cctcccgggt	tcaagtgatt	ctcctgcctc	agcctcctga	gtagttggga	56761110
tttcaggcac	ccaccatgcc	aggctagttt	ttgtatlttt	agtagagaca	56761160
ggatttcacc	atgttggtta	ggctgggtctc	gaactccaga	cctcaggtag	56761210
tctgcccgcc	tcagactccc	aaagtgtctg	gatcacaggc	gtgagccact	56761260
gtgcccggca	aaatltttatt	tactlttttt	tttttttgaga	caaagtctca	56761310
ctctgtcgcc	caggctggag	tgcagtgggtg	caatctccac	tactgcaac	56761360
ttttgcctcc	caggttcaca	cgattctcct	gcctcagcct	caggagtagc	56761410
tgggattaca	gggcctgccc	accacacctg	gctaattlttt	gtatlttttag	56761460
tatagatggg	tttcgccatg	ttggccaggc	tggtctcgaa	ctcttcatct	56761510
caagtgatcc	acctgcctcg	gcctcccac	atgctgagat	tacagggtgtg	56761560
agtcactgca	cctggcctat	tattatlttt	tttagattca	gagtctcatt	56761610
ctgttgccca	ggttgagta	cagtgggtata	atcatggccc	actgcagcct	56761660
cgaactcctg	gggtcaagta	gccctcccac	cttagcctcc	ttagtagtag	56761710
ggactgCAGG	TGcATGCCAC	CaTGCCCAGC	TAATTAAAGA	AtltttttttT	56761760
TTTTTTCTGA	AAGTCTTGCT	TTGTCACCCA	GGCTGGAGTG	CAGTGGCATG	56761810
AACACAGCTC	gCT				

Figure 21

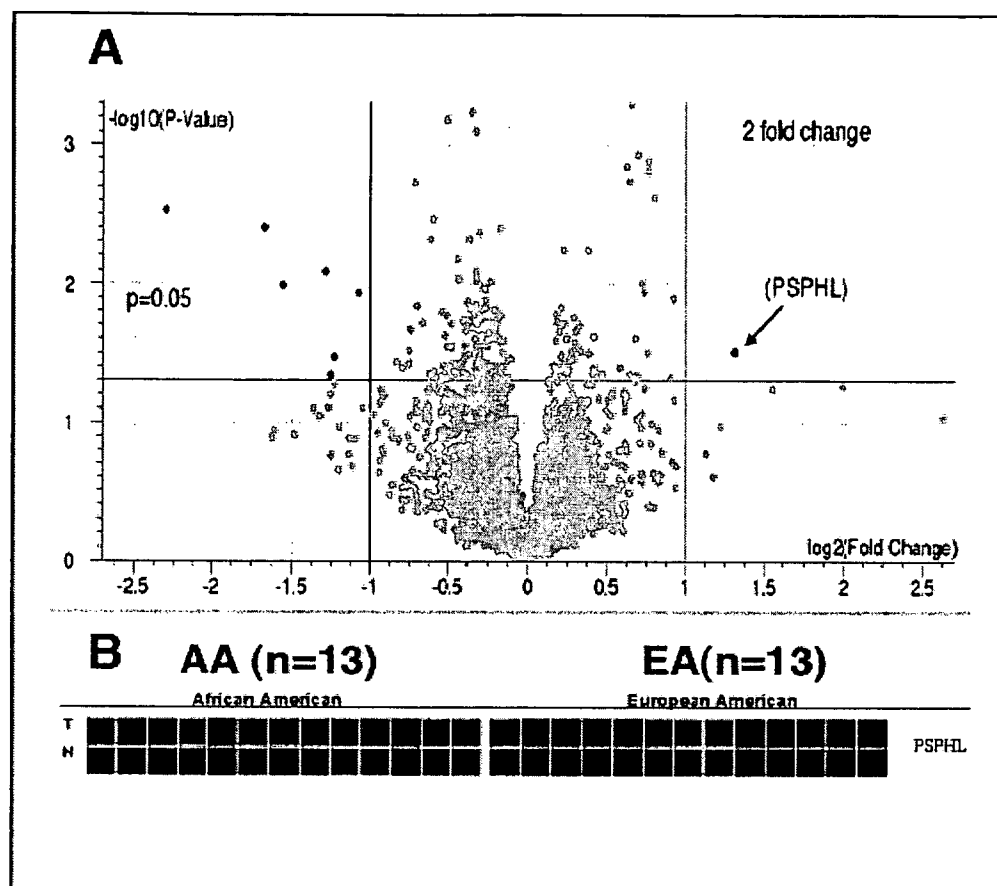


Figure 22

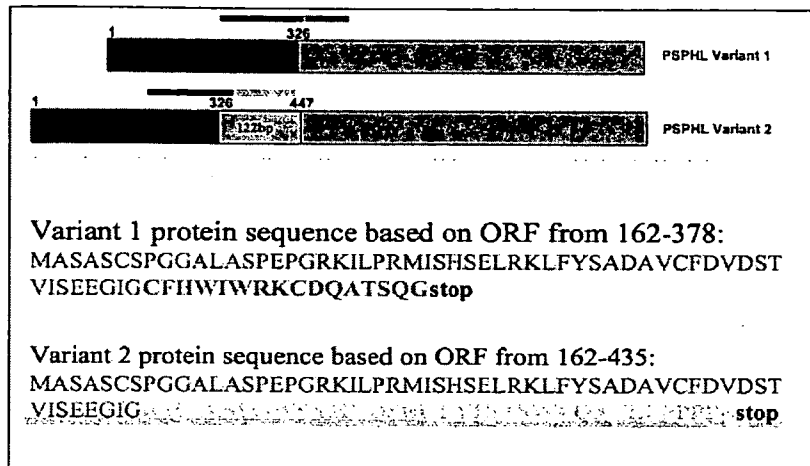


Figure 23

A

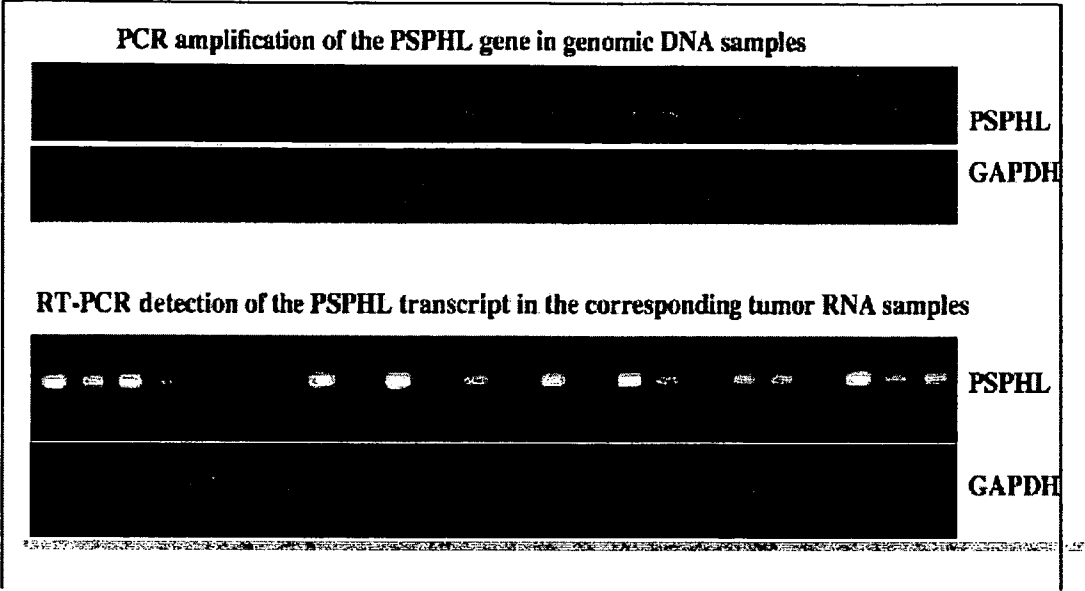


Figure 23

B

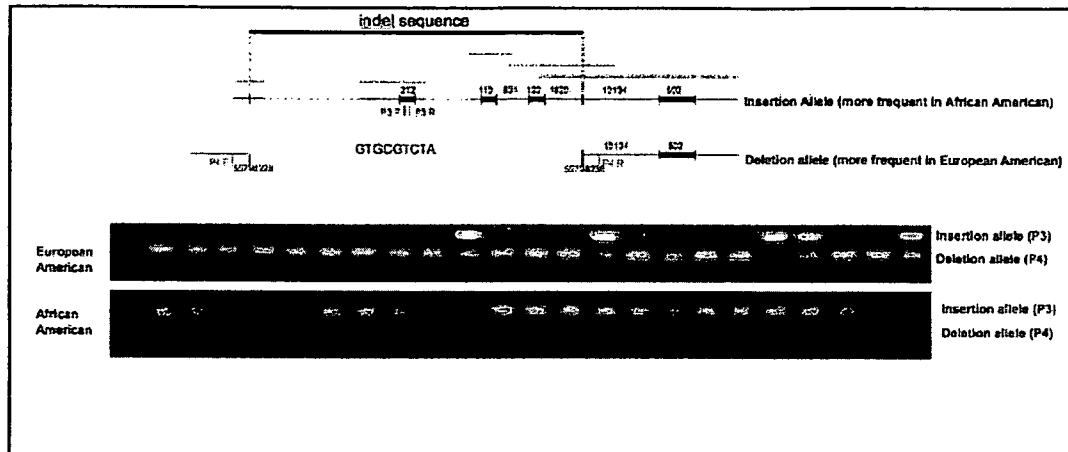


Figure 24

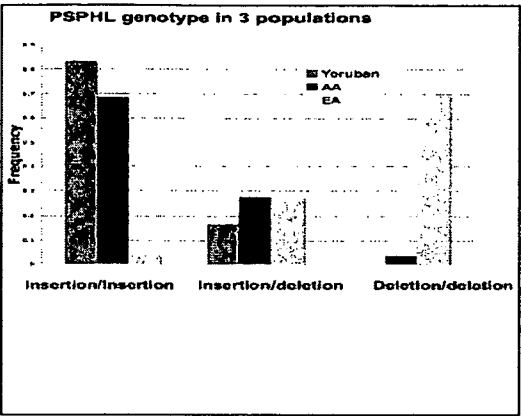


Figure 25

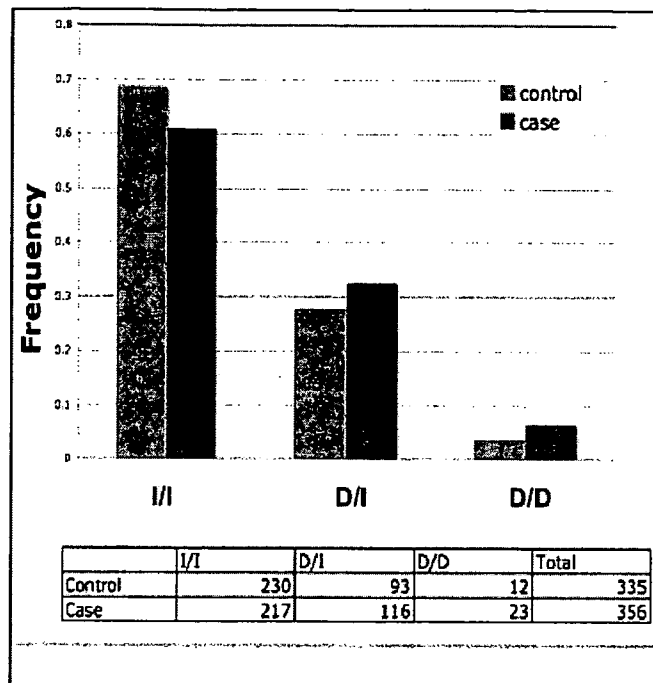


Figure 26

