(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 336 368 A1

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
22.06.2011 Bulletin 2011/25

(21) Application number: **10176402.5**

(22) Date of filing: **07.12.2001**

(51) Int Cl.:
C12Q 1/70 *(2006.01)* ***G01N 33/574*** *(2006.01)*
C07K 14/15 *(2006.01)* ***A61K 48/00*** *(2006.01)*
A61P 35/00 *(2006.01)* ***C12Q 1/68*** *(2006.01)*
A61K 39/21 *(2006.01)*

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **07.12.2000 US 251830 P**
07.12.2001 US 16604

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
01996222.4 / 1 415 005

(71) Applicant: **Novartis Vaccines and Diagnostics, Inc.**
Emeryville, CA 94608 (US)

(72) Inventors:
- **Garcia, Pablo**
  San Francisco, CA 94131 (US)
- **Hardy, Stephen F.**
  San Francisco, CA 94121 (US)
- **Williams, Lewis T.**
  Mill Valley, CA 94902 (US)
- **Escobedo, Jaime**
  Alamo, CA 94507 (US)

(74) Representative: **Marshall, Cameron John**
Carpmaels & Ransford
One Southampton Row
London
WC1B 5HA (GB)

Remarks:
This application was filed on 13-09-2010 as a divisional application to the application mentioned under INID code 62.

(54) **Endogenous retroviruses up-regulated in prostate cancer**

(57) Human endogenous retroviruses of the HML-2 family show up-regulated expression in prostate tumors. This finding can be used in prostate cancer screening, diagnosis and therapy.

EP 2 336 368 A1

Description

[0001] All documents cited herein are incorporated by reference in their entirety.

TECHNICAL FIELD

[0002] The present invention relates to the diagnosis of cancer, particularly prostate cancer. In particular, it relates to a subgroup of human endogenous retroviruses (HERVs) which show up-regulated expression in tumors, particularly prostate tumors.

BACKGROUND ART

[0003] Prostate cancer is the most common type of cancer in men in the USA. Benign prostatic hyperplasia (BPH) is the abnormal growth of benign prostate cells in which the prostate grows and pushes against the urethra and bladder, blocking the normal flow of urine. More than half of the men in the USA between the ages of 60 and 70 and as many as 90 percent between the ages of 70 and 90 have symptoms of BPH. Although this condition is seldom a threat to life, it may require treatment to relieve symptoms.

[0004] Cancer that begins in the prostate is called primary prostate cancer (or prostatic cancer). Prostate cancer may remain in the prostate gland, or it may spread to nearby lymph nodes and may also spread to the bones, bladder, rectum, and other organs. Prostate cancer is diagnosed by measuring the levels of prostate-specific antigen (PSA) and prostatic acid phosphatase (PAP) in the blood. The level of PSA in blood may rise in men who have prostate cancer, BPH, or an infection in the prostate. The level of PAP rises above normal in many prostate cancer patients, especially if the cancer has spread beyond the prostate. However, one cannot diagnose prostate cancer with these tests alone because elevated PSA or PAP levels may also indicate other, non-cancerous problems.

[0005] In order to help determine whether conditions of the prostate are benign or malignant further tests such as transrectal ultrasonography, intravenous pyelogram, and cystoscopy are usually performed. If these test results suggest that cancer may be present, the patient must undergo a biopsy as the only sure way to diagnose prostate cancer. Consequently, it is desirable to provide a simple and direct test for the early detection and diagnosis of prostate cancer without having to undergo multiple rounds of cumbersome testing procedures. It is also desirable and necessary to provide compositions and methods for the prevention and/or treatment of prostate cancer.

[0006] It is an object of the invention to provide materials that can be used in the prevention, treatment and diagnosis of prostate cancer. It is a further object to provide improvements in the prevention, treatment and diagnosis of prostate cancer.

DISCLOSURE OF THE INVENTION

[0007] It has been found that human endogenous retroviruses (HERVs) of the HML-2 subgroup of the HERV-K family show up-regulated expression in prostate tumors. This finding can be used in prostate cancer screening, diagnosis and therapy.

[0008] The invention provides a method for diagnosing cancer, especially prostate cancer, the method comprising the step of detecting the presence or absence of an expression product of a HML-2 endogenous retrovirus in a patient sample. Higher levels of expression product relative to normal tissue indicate that the patient from whom the sample was taken has cancer.

[0009] The HML-2 expression product which is detected is either a mRNA transcript or a polypeptide translated from such a transcript. These expression products may be detected directly or indirectly. A direct test uses an assay which detects HML-2 RNA or polypeptide in a patient sample. An indirect test uses an assay which detects biomolecules which are not directly expressed *in vivo* from HML-2 *e.g.* an assay to detect cDNA which has been reverse-transcribed from a HML-2 mRNA, or an assay to detect an antibody which has been raised in response to a HML-2 polypeptide.

A - THE PATIENT SAMPLE

[0010] Where the diagnostic method of the invention is based on HML-2 mRNA, the patient sample will generally comprise cells, preferably, prostate cells. These may be present in a sample of tissue, preferably, prostate tissue, or may be cells, preferably, prostate cells which have escaped into circulation (*e.g.* during metastasis). Instead of or as well as comprising prostate cells, the sample may comprise virions which contain mRNA from HML-2.

[0011] Where the diagnostic method of the invention is based on HML-2 polypeptides, the patient sample may comprise cells, preferably, prostate cells and/or virions (as described above for mRNA), or may comprise antibodies which recognize HML-2 polypeptides. Such antibodies will typically be present in circulation.

[0012] In general, therefore, the patient sample is tissue sample (*e.g.* a biopsy), preferably, a prostate sample (*e.g.* a biopsy) or a blood sample.

[0013] The patient is generally a human, preferably human male, and more preferably an adult human male.

[0014] Expression products may be detected in the patient sample itself, or it may be detected in material derived from the sample (*e.g.* the supernatant of a cell lysate, or a RNA extract, or cDNA generated from a RNA extract, or polypeptides translated from a RNA extract, or cells derived from culture of cells extracted from a patient *etc.*). These are still considered to be "patient samples" within the meaning of the invention.

[0015] Methods of the invention can be conducted *in vitro* or *in vivo.*

[0016] Other possible sources of patient samples include isolated cells, whole tissues, or bodily fluids (*e.g.* blood, plasma, serum, urine, pleural effusions, cerebro-spinal fluid, *etc.*)

B - THE mRNA EXPRESSION PRODUCT

[0017] Where the diagnostic method of the invention is based on mRNA detection, it typically involves detecting a RNA comprising six basic regions. From 5' to 3', these are:

1. A sequence which has at least 75% identity to SEQ ID 155 (*e.g.* 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity); or a sequence which has at least 50% identity to SEQ ID 155 (*e.g.* 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) and is expressed at least 1.5 fold (*e.g.* 2, 2.5, 5, 10, 20, 50, etc., fold) higher level relative to expression in a normal (*i.e.,* non cancerous) cell with at least a 95% confidence level; or a sequence which has at least 80% identity (*e.g.* 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) to at least a 20 contiguous nucleotide fragment (*e.g.* 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 115, 120, 125, 130, 135, 140, 145, etc., contiguous nucleotides) of SEQ ID 155; or a sequence which has at least 80% identity (*e.g.* 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) to at least a 20 contiguous nucleotide fragment (*e.g.* 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 115, 120, 125, 130, 135, 140, 145, etc., contiguous nucleotides) of SEQ ID 155 and is expressed at least 1.5 fold (*e.g.* 2, 2.5, 5, 10, 20, 50, etc., fold) higher level relative to expression in a normal (*i.e.*, non cancerous) cell with at least a 95% confidence level. This sequence will typically be at the 5' end of the RNA. SEQ ID 155 is the nucleotide sequence of the start of R region in the LTR of the 'ERVK6' HML-2 virus [ref. 1]. This portion of the R region is found in all full-length HML-2 transcripts.

2. A downstream region comprising a sequence which has at least 75% sequence identity to SEQ ID 156 (*e.g.* 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity); or a sequence which has at least 50% identity to SEQ ID 156 (*e.g.* 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) and is expressed at least 1.5 fold (*e.g.* 2, 2.5, 5, 10, 20, 50, etc., fold) higher level relative to expression in a normal (*i.e.*, non cancerous) cell with at least a 95% confidence level; or a sequence which has at least 80% identity (*e.g.* 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) to at least a 20 contiguous nucleotide fragment (*e.g.* 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, etc., continuous nucleotides) of SEQ ID 156; or a sequence which has at least 80% identity (*e.g.* 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) to at least a 20 contiguous nucleotide fragment (*e.g.* 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, etc., contiguous nucleotides) of SEQ ID 156 and is expressed at least 1.5 fold *(e.g.* 2, 2.5, 5,10,20, 50, etc., fold) higher level relative to expression in a normal (*i.e.*, non cancerous) cell with at least a 95% confidence level. SEQ ID 156 is the nucleotide sequence of the RU_5 region downstream of SEQ ID 155 in the ERVK6 LTR. This region is found in full-length HML-2 transcripts, but may not be present in all mRNAs transcribed from a HML-2 LTR promoter.

3. A downstream region comprising a sequence which has at least 75% sequence identity to SEQ ID 6 (*e.g.* 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,

97%, 98%, 99%, 99.5%, 99.9%, 100% identity); or a sequence which has at least 50% identity to SEQ ID 6 (*e.g.* 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) and is expressed at least 1.5 fold (*e.g.* 2, 2.5, 5, 10, 20, 50, etc., fold) higher level relative to expression in a normal (*i.e.*, non cancerous) cell with at least a 95% confidence level; or a sequence which has at least 80% identity (*e.g.* 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) to at least a 20 contiguous nucleotide fragment (*e.g.* 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, etc., contiguous nucleotides) of SEQ ID 6; or a sequence which has at least 80% identity (*e.g.* 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) to at least a 20 contiguous nucleotide fragment (*e.g.* 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, etc., contiguous nucleotides) of SEQ ID 6 and is expressed at least 1.5 fold (*e.g.* 2, 2.5, 5, 10, 20, 50, etc., fold) higher level relative to expression in a normal (*i.e.*, non cancerous) cell with at least a 95% confidence level. SEQ ID 6 is the nucleotide sequence of the region of the ERVK6 virus between the U_5 region and the first 5' splice site. This region is found in full-length HML-2 transcripts, but has been lost by some variants and, like region 2 above, may not be present in all mRNAs transcribed from a HML-2 LTR promoter.

4. A downstream region comprising any RNA sequence. This region will typically comprise the coding sequence of one or more HML-2 polypeptides, but may alternatively comprise: a mutant viral coding sequence; a viral or non-viral non-coding sequence; or a non-viral coding sequence. Transcription of any of these sequences can come under the control of a HML-2 LTR.

5. A downstream region comprising a sequence which has at least 75% sequence identity to SEQ ID 5 (*e.g.* 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity); or a sequence which has at least 50% identity to SEQ ID 5 (*e.g.* 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 68%, 69%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) and is expressed at least 1.5 fold (*e.g.* 2, 2.5, 5, 10, 20, 50, etc., fold) higher level relative to expression in a normal (*i.e.*, non cancerous) cell with at least a 95% confidence level; or a sequence which has at least 80% identity (*e.g.* 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) to at least a 20 contiguous nucleotide fragment (*e.g.* 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, 275, 280, 285, 290, 295, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, etc., contiguous nucleotides) of SEQ ID 5; or a sequence which has at least 80% identity (*e.g.* 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, 99.9%, 100% identity) to at least a 20 contiguous nucleotide fragment (*e.g.*. 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, 275, 280, 285, 290, 295, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, etc., contiguous nucleotides) of SEQ ID 5 and is expressed at least 1.5 fold (*e.g.* 2, 2.5, 5, 10, 20, 50, etc., fold) higher level relative to expression in a normal (*i.e.*, non cancerous) cell with at least a 95% confidence level. SEQ ID 5 is the nucleotide sequence of the U_3R region in the 3' end of ERVK6. This sequence will typically be near the 3' end of the RNA, immediately preceding any polyA tail.

6. A 3' polyA tail.

[0018] The percent identity of the sequences described above are determined by the Smith-Waterman algorithm using the default parameters: open gap penalty = -20 and extension penalty = -5.

[0019] These mRNA molecules are referred to below as "PCA-mRNA" molecules ("prostate cancer associated mRNA"), and endogenous viruses which express these PCA-mRNAs are referred to as PCAVs ("prostate cancer associated viruses"). Nevertheless, said PCAVs may also be associated with other types of cancer.

[0020] Although some PCA-mRNAs include all six of these regions, most HERVs are defective in that they have accumulated multiple stop codons, frameshifts, or larger deletions *etc.* This means that many PCA-mRNAs do not include all six regions. As all PCA-mRNAs are transcribed under the control of the same group of LTRs, however, transcription of all PCA-mRNAs is up-regulated in prostate tumors even though the mRNA may not encode functional polypeptides.

[0021] Where a mRNA to be detected is driven by 5' LTR of HML-2 in genomic DNA, the first of these regions will always be present, but the remaining five are optional. Conversely, where a mRNA to be detected is controlled by 3' LTR of HML-2, the fifth of these regions will always be present, but the remaining five are optional.

[0022] In general, therefore, the mRNA to be detected has the formula N_1—N_2—N_3—N_4—N_5— polyA, wherein:

- N_1 has at least 75% sequence identity to SEQ ID 155; or has at least 50% identity to SEQ ID 155 and is expressed at least 1.5 fold higher relative to expression in a normal (i.e., non cancerous) cell with at least a 95% confidence level; or has at least 80% identity to at least a 20 contiguous nucleotide fragment of SEQ ID 155; or has at least 80% identity to at least a 20 contiguous nucleotide fragment of SEQ ID 155 and is expressed at least 1.5 fold higher relative to expression in a normal (i.e., non cancerous) cell with at least a 95% confidence level;
- N_2 has at least 75% sequence identity to SEQ ID 156; or has at least 50% identity to SEQ ID 156 and is expressed at least 1.5 fold higher relative to expression in a normal (i.e., non cancerous) cell with at least a 95% confidence level; or has at least 80% identity to at least a 20 contiguous nucleotide fragment of SEQ ID 156; or has at least 80% identity to at least a 20 contiguous nucleotide fragment of SEQ ID 156 and is expressed at least 1.5 fold higher relative to expression in a normal (i.e., non cancerous) cell with at least a 95% confidence level;
- N_3 has at least 75% sequence identity to SEQ ID 6; or has at least 50% identity to SEQ ID 6 and is expressed at least 1.5 fold higher relative to expression in a normal (i.e., non cancerous) cell with at least a 95% confidence level; or has at least 80% identity to at least a 20 contiguous nucleotide fragment of SEQ ID 6; or has at least 80% identity to at least a 20 contiguous nucleotide fragment of SEQ ID 6 and is expressed at least 1.5 fold higher relative to expression in a normal (i.e., non cancerous) cell with at least a 95% confidence level;
- N_4 comprises any RNA sequence;
- N_5 has at least 75% sequence identity to SEQ ID 5; or has at least 50% identity to SEQ ID 5 and is expressed at least 1.5 fold higher relative to expression in a normal (i.e., non cancerous) cell with at least a 95% confidence level; or has at least 80% identity to at least a 20 contiguous nucleotide fragment of SEQ ID 5; or has at least 80% identity to at least a 20 contiguous nucleotide fragment of SEQ ID 5 and is expressed at least 1.5 fold higher relative to expression in a normal (i.e., non cancerous) cell with at least a 95% confidence level; and
- at least one of N_1, N_2, N_3, N_4 or N_5 is present, but polyA is optional.

[0023] Although only at least one of N_1, N_2, N_3, N_4 or N_5 needs to be present, it is preferred that two, three, four or five of these regions are present. It is preferred that at least one of N_1 and/or N_5 is present.

[0024] N_1 is preferably present in the mRNA to be detected (*i.e.* the invention is preferably based on the detection of mRNA driven by a 5' LTR). More preferably, at least N_1-N_2 is present.

[0025] Where N_1 is present, it is preferably at the 5' end of the mRNA *(i.e.* 5'- N_1—...).

[0026] Where N_5 is present, it is preferably immediately before a 3' polyA tail (*i.e.* ... —N_5-polyA-3').

[0027] Where N_4 is present, it preferably comprises a polypeptide-coding sequence (*e.g.* encoding a HML-2 polypeptide). Examples of HML-2 polypeptide-coding sequences are described below.

[0028] The RNA will generally have a 5' cap.

B.1 - Enriching RNA in a sample

[0029] Where diagnosis is based on mRNA detection, the method of the invention preferably comprises an initial step of: (a) extracting RNA (*e.g.* mRNA) from a patient sample; (b) removing DNA from a patient sample without removing mRNA; and/or (c) removing or disrupting DNA which comprises SEQ ID 4, but not RNA which comprises SEQ ID 4, from a patient sample. This is necessary because the genomes of both normal and cancerous prostate cells contain multiple PCAV DNA templates, whereas increased PCA-mRNA levels are only found in cancerous cells. As an alternative, a RNA-specific assay can be used which is not affected by the presence of homologous DNA.

[0030] Methods for extracting RNA from biological samples are well known [*e.g.* refs. 2 & 8] and include methods based on guanidinium buffers, lithium chloride, SDS/potassium acetate *etc.* After total cellular RNA has been extracted, mRNA may be enriched *e.g.* using oligo-dT techniques.

[0031] Methods for removing DNA from biological samples without removing mRNA are well known [*e.g.* appendix C of ref. 2] and include DNase digestion.

[0032] Methods for removing DNA, but not RNA, comprising PCA-mRNA sequences will use a reagent which is specific to a sequence within a PCA-mRNA *e.g.* a restriction enzyme which recognizes a DNA sequence within SEQ ID 4, but which does not cleave the corresponding RNA sequence.

[0033] Methods for specifically purifying PCA-mRNAs from a sample may also be used. One such method uses an affinity support which binds to PCA-mRNAs. The affinity support may include a polypeptide sequence which binds to the PCAV-mRNA *e.g.* the cORF polypeptide, which binds to the LTR of HERV-K mRNAs in a sequence-specific manner, or HIV Rev protein, which has been shown to recognize the HERV-K LTR [3].

B.2 - Direct detection of RNA

[0034] Various techniques are available for detecting the presence or absence of a particular RNA sequence in a sample [*e.g.* refs. 2 & 8]. If a sample contains genomic PCAV DNA, the detection technique will generally be RNA-specific; if the sample contains no PCAV DNA, the detection technique may or may not be RNA-specific.

[0035] Hybridization-based detection techniques may be used, in which a polynucleotide probe complementary to a region of PCA-mRNA is contacted with a RNA-containing sample under hybridizing conditions. Detection of hybridization indicates that nucleic acid complementary to the probe is present. Hybridization techniques for use with RNA include Northern blots, *in situ* hybridization and arrays.

[0036] Sequencing may also be used, in which the sequence(s) of RNA molecules in a sample are obtained. These techniques reveal directly whether a sequence of interest is present in a sample. Sequence determination of the 5' end of a RNA corresponding to N_1 will generally be adequate.

[0037] Amplification-based techniques may also be used. These include PCR, SDA, SSSR, LCR, TMA, NASBA, T7 amplification *etc.* The technique preferably gives exponential amplification. A preferred technique for use with RNA is RT-PCR [*e.g.* see chapter 15 of ref. 2]. RT-PCR ofmRNA from prostate cells is reported in references 4, 5, 6 & 7.

B.3 - Indirect detection of RNA

[0038] Rather than detect RNA directly, it may be preferred to detect molecules which are derived from RNA (*i.e.* indirect detection of RNA). A typical indirect method of detecting mRNA is to prepare cDNA by reverse transcription and then to directly detect the cDNA. Direct detection of cDNA will generally use the same techniques as described above for direct detection of RNA (but it will be appreciated that methods such as RT-PCR are not suitable for DNA detection and that cDNA is double-stranded, so detection techniques can be based on a sequence, on its complement, or on the double-stranded molecule).

B.4 - Polynucleotide material

[0039] The invention provides polynucleotide materials for use in the detection of PCAV nucleic acids.

[0040] The invention provides an isolated polynucleotide comprising: (a) the nucleotide sequence N_1-N_2-N_3-N_4-N_5-polyA as defined above; (b) a fragment of at least x nucleotides of nucleotide sequence N_1—N_2—N_3—N_4—N_5 as defined above; (c) a nucleotide sequence having at least *s*% identity to nucleotide sequence N_1—N_2—N_3—N_4—N_5 as defined above; or (d) the complement of (a), (b) or (c). These polynucleotides include variants of nucleotide sequence N_1—N_2—N_3—N_4—N_5—polyA (*e.g.* degenerate variants, allelic variants, homologs, orthologs, mutants *etc.*).

[0041] Fragment (b) is preferably a fragment of N_1.

[0042] The value of x is at least 7 (*e.g.* at least 8,9,10,11,12,13,14,15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 75, 80, 90, 100 *etc.*). The value of *x* may be less than 2000 (*e.g.* less than 1000, 500, 100, or 50).

[0043] The value of *s* is preferably at least 50 (*e.g.* at least 55, 60, 65, 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.5, 99.9 *etc.*).

[0044] The invention also provides an isolated polynucleotide having formula 5'-A-B-C-3', wherein: -A- is a nucleotide sequence consisting of a nucleotides; -C- is a nucleotide sequence consisting of *c* nucleotides; -B- is a nucleotide sequence consisting of either (a) a fragment of *b* nucleotides of nucleotide sequence N_1—N_2—N_3—N_4—N_5 as defined above or (b) the complement of a fragment of *b* nucleotides of nucleotide sequence N_1—N_2—N_3—N_4—N_5 as defined above; and said polynucleotide is neither (a) a fragment of nucleotide sequence N_1— N_2—N_3—N_4—N_5 or (b) the complement of a fragment of nucleotide sequence N_1—N_2—N_3— N_4—N_5.

[0045] The -B- moiety is preferably a fragment of N_1—N_2, and more preferably a fragment of N_1. The -A- and/or -C- moieties may comprise a promoter sequence (or its complement) *e.g.* for use in TMA.

[0046] The value of *a+c* is at least 1 (*e.g.* at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27,28, 29, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.*). The value of *b* is at least 7 *(e.g.* at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.*). It is preferred that the value of a+b+c is at least 9 (*e.g.* at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.*). It is preferred that the value of *a+b+c* is at most 500 (*e.g.* at most 450, 400, 350, 300, 250, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9).

[0047] Where -B- is a fragment of N_1—N_2—N_3—N_4—N_5, the nucleotide sequence of -Atypically shares less than n% sequence identity to the *a* nucleotides which are 5' of sequence -Bin N_1—N_2—N_3—N_4—N_5 and/or the nucleotide sequence of -C- typically shares less than n% sequence identity to the c nucleotides which are 3' of sequence -C- in N_1—N_2—N_3—N_4—N_5. Similarly, where -B- is the complement of a fragment of N_1—N_2—N_3—N_4—N_5, the nucleotide

sequence of -A- typically shares less than *n*% sequence identity to the complement of the *a* nucleotides which are 5' of the complement of sequence -B- in N_1—N_2—N_3—N_4—N_5 and/or the nucleotide sequence of -C- typically shares less than n% sequence identity to the complement of the c nucleotides which are 3' of the complement of sequence -C- in N_1—N_2—N_3—N_4—N_5. The value of *n* is generally 60 or less (*e.g.* 50, 40, 30, 20, 10 or less).

[0048] The invention also provides an isolated polynucleotide which selectively hybridizes to a nucleic acid having nucleotide sequence N_1—N_2—N_3—N_4—N_5 as defined above or to a nucleic acid having the complement of nucleotide sequence N_1—N_2—N_3—N_4—N_5 as defined above. The polynucleotide preferably hybridizes to at least N_1.

[0049] Hybridization reactions can be performed under conditions of different "stringency". Conditions that increase stringency of a hybridization reaction of widely known and published in the art [*e.g.* page 7.52 of reference 8]. Examples of relevant conditions include (in order of increasing stringency): incubation temperatures of 25°C, 37°C, 50°C, 55°C and 68°C; buffer concentrations of 10 X SSC, 6 X SSC, 1 X SSC, 0.1 X SSC (where SSC is 0.15 M NaCl and 15 mM citrate buffer) and their equivalents using other buffer systems; formamide concentrations of 0%, 25%, 50%, and 75%; incubation times from 5 minutes to 24 hours; 1, 2, or more washing steps; wash incubation times of 1, 2, or 15 minutes; and wash solutions of 6 x SSC, 1 x SSC, 0.1 x SSC, or de-ionized water. Hybridization techniques are well known in the art [*e.g.* see references 2, 8, 9, 10, 11 *etc.*]. Depending upon the particular polynucleotide sequence and the particular domain encoded by that polynucleotide sequence, hybridization conditions upon which to compare a polynucleotide of the invention to a known polynucleotide may differ, as will be understood by the skilled artisan.

[0050] In some embodiments, the isolated polynucleotide of the invention selectively hybridizes under low stringency conditions; in other embodiments it selectively hybridizes under intermediate stringency conditions; in other embodiments, it selectively hybridizes under high stringency conditions. An exemplary set of low stringency hybridization conditions is 50°C and 10xSSC. An exemplary set of intermediate stringency hybridization conditions is 55°C and 1xSSC. An exemplary set of high stringent hybridization conditions is 68°C and 0.1 x SSC.

[0051] The polynucleotides of the invention are particularly useful as probes and/or as primers for use in hybridization and/or amplification reactions.

[0052] More than one polynucleotide of the invention can hybridize to the same nucleic acid target (*e.g.* more than one can hybridize to a single RNA).

[0053] References to a percentage sequence identity between two nucleic acid sequences mean that, when aligned, that percentage of bases are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art, for example those described in section 7.7.18 of reference 11. A preferred alignment program is GCG Gap (Genetics Computer Group, Wisconsin, Suite Version 10.1), preferably using default parameters, which are as follows: open gap = 3; extend gap = 1.

[0054] Polynucleotides of the invention may take various forms *e.g.* single-stranded, double-stranded, linear, circular, vectors, primers, probes *etc.*

[0055] Polynucleotides of the invention can be prepared in many ways *e.g.* by chemical synthesis (at least in part), by digesting longer polynucleotides using restriction enzymes, from genomic or cDNA libraries, from the organism itself *etc.*

[0056] Polynucleotides of the invention may be attached to a solid support (*e.g.* a bead, plate, filter, film, slide, resin, *etc.*)

[0057] Polynucleotides of the invention may include a detectable label (*e.g.* a radioactive or fluorescent label, or a biotin label). This is particularly useful where the polynucleotide is to be used in nucleic acid detection techniques *e.g.* where the nucleic acid is a primer or as a probe for use in techniques such as PCR, LCR, TMA, NASBA, bDNA *etc.*

[0058] The term "polynucleotide" in general means a polymeric form of nucleotides of any length, which contain deoxyribonucleotides, ribonucleotides, and/or their analogs. It includes DNA, RNA, DNA/RNA hybrids, and DNA or RNA analogs, such as those containing modified backbones or bases, and also peptide nucleic acids (PNA) *etc.* The term "polynucleotide" is not intended to be limiting as to the length or structure of a nucleic acid unless specifically indicated, and the following are non-limiting examples of polynucleotides: a gene or gene fragment, exons, introns, mRNA, tRNA, rRNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, any isolated DNA from any source, any isolated RNA from any sequence, nucleic acid probes, and primers. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. Unless otherwise specified or required, any embodiment of the invention that includes a polynucleotide encompasses both the double-stranded form and each of two complementary single-stranded forms known or predicted to make up the double stranded form.

[0059] Polynucleotides of the invention may be isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the polynucleotides will be obtained substantially free of other naturally-occurring nucleic acid sequences, generally being at least about 50% (by weight) pure, usually at least about 90% pure.

[0060] Polynucleotides of the invention (particularly DNA) are typically "recombinant" *e.g.* flanked by one or more nucleotides with which it is not normally associated on a naturally-occurring chromosome.

[0061] The polynucleotides can be used, for example: to produce polypeptides; as probes for the detection of nucleic acid in biological samples; to generate additional copies of the polynucleotides; to generate ribozymes or antisense oligonucleotides; and as single-stranded DNA probes or as triple-strand forming oligonucleotides. The polynucleotides

are preferably uses to detect PCA-mRNAs.

[0062] A "vector" is a polynucleotide construct designed for transduction/transfection of one or more cell types. Vectors may be, for example, "cloning vectors" which are designed for isolation, propagation and replication of inserted nucleotides, "expression vectors" which are designed for expression of a nucleotide sequence in a host cell, "viral vectors" which is designed to result in the production of a recombinant virus or virus-like particle, or "shuttle vectors", which comprise the attributes of more than one type of vector.

[0063] A "host cell" includes an individual cell or cell culture which can be or has been a recipient of exogenous polynucleotides. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation and/or change. A host cell includes cells transfected or infected *in vivo* or *in vitro* with a polynucleotide of this invention.

B.5 - Nucleic acid detection kits

[0064] The invention provides a kit comprising primers (*e.g.* PCR primers) for amplifying a template sequence contained within a PCAV nucleic acid, the kit comprising a first primer and a second primer, wherein the first primer is substantially complementary to said template sequence and the second primer is substantially complementary to a complement of said template sequence, wherein the parts of said primers which have substantial complementarity define the termini of the template sequence to be amplified. The first primer and/or the second primer may include a detectable label.

[0065] The invention also provides a kit comprising first and second single-stranded oligonucleotides which allow amplification of a PCAV template nucleic acid sequence contained in a single- or double-stranded nucleic acid (or mixture thereof), wherein: (a) the first oligonucleotide comprises a primer sequence which is substantially complementary to said template nucleic acid sequence; (b) the second oligonucleotide comprises a primer sequence which is substantially complementary to the complement of said template nucleic acid sequence; (c) the first oligonucleotide and/or the second oligonucleotide comprise(s) sequence which is not complementary to said template nucleic acid; and (d) said primer sequences define the termini of the template sequence to be amplified. The non-complementary sequence(s) of feature (c) are preferably upstream of (*i.e.* 5' to) the primer sequences. One or both of the (c) sequences may comprise a restriction site [12] or promoter sequence [13]. The first and/or the second oligonucleotide may include a detectable label.

[0066] The kit of the invention may also comprise a labeled polynucleotide which comprises a fragment of the template sequence (or its complement). This can be used in a hybridization technique to detect amplified template.

[0067] The primers and probes used in these kits are preferably polynucleotides as described in section B.4.

[0068] The template is preferably a sequence as defined in section B.1 above.

C-POLYPEPTIDE EXPRESSION PRODUCT

[0069] Where the method is based on polypeptide detection, it will involve detecting expression of a polypeptide encoded by a PCAV-mRNA. This will typically involve detecting one or more of the following HML-2 polypeptides: gag, prt, pol, env, cORF. Although some PCA-mRNAs encode all of these polypeptides (*e.g.* ERVK6 [1]), the polypeptide-coding regions of most HERVs (including PCAVs) contain mutations which mean that one or more coding-regions in the mRNA transcript are either mutated or absent. Thus not all PCAVs have the ability to encode all HML-2 polypeptides.

[0070] The transcripts which encode HML-2 polypeptides are generated by alternative splicing of the full-length mRNA copy of the endogenous genome [*e.g.* Figure 4 of ref. 143].

[0071] HML-2 gag polypeptide is encoded by the first long ORF in a complete HML-2 genome [140]. Full-length gag polypeptide is proteolytically cleaved.

[0072] Examples of gag nucleotide sequences are: SEQ IDs 7, 8, 9 & 11 [HERV-K(CH)]; SEQ ID 85 [HERV-K108]; SEQ ID 91 [HERV-K(C7)]; SEQ ID 97 [HERV-K(II)]; SEQ ID 102 [HERV-K10].

[0073] Examples of gag polypeptide sequences are: SEQ IDs 46, 47, 48, 49, 56 & 57 [HERV-K(CH)]; SEQ ID 92 [HERV-K(C7)]; SEQ ID 98 [HERV-K(II)]; SEQ IDs 103 & 104 [HERV-K10]; SEQ ID 146 ['ERVK6'].

[0074] An alignment of gag polypeptide sequences is shown in Figure 7.

[0075] HML-2 prt polypeptide is encoded by the second long ORF in a complete HML-2 genome. It is translated as a gag-prt fusion polypeptide. The fusion polypeptide is proteolytically cleaved to give a protease.

[0076] Examples of prt nucleotide sequences are: SEQ ID 86 [HERV-K(108)]; SEQ ID 99 [HERV-K(II)]; SEQ ID 105 [HERV-K10].

[0077] Examples of prt polypeptide sequences are: SEQ ID 106 [HERV-K10]; SEQ ID 147 ['ERVK6'].

[0078] HML-2 pol polypeptide is encoded by the third long ORF in a complete HML-2 genome. It is translated as a gag-prt-pol fusion polypeptide. The fusion polypeptide is proteolytically cleaved to give three pol products - reverse transcriptase, endonuclease and integrase [14].

[0079] Examples of pol nucleotide sequences are: SEQ ID 87 [HERV-K(108)]; SEQ ID 93 [HERV-K(C7)]; SEQ ID 100

[HERV-K(II)]; SEQ ID 107 [HERV-K10].

[0080] Examples of pol polypeptide sequences are: SEQ ID 94 [HERV-K(C7)]; SEQ ID 108 [HERV-K10]; SEQ ID 148 ['ERVK6'].

[0081] An alignment of pol polypeptide sequences is shown in Figure 8.

[0082] HML-2 env polypeptide is encoded by the fourth long ORF in a complete HML-2 genome. The translated polypeptide is proteolytically cleaved.

[0083] Examples of env nucleotide sequences are: SEQ ID 88 [HERV-K(108)]; SEQ ID 95 [HERV-K(C7)]; SEQ ID 101 [HERV-K(II)]; SEQ ID 107 [HERV-K10].

[0084] Examples of env polypeptide sequences are: SEQ ID 96 [HERV-K(C7)]; SEQ ID 108 [HERV-K10] ; SEQ ID 149 ['ERVK6'].

[0085] Alignments of env polynucleotide and polypeptide sequences are shown in Figures 6 and 9.

[0086] HML-2 cORF polypeptide is encoded by an ORF which shares the same 5' region and start codon as env. After amino acid 87, a splicing event removes env-coding sequences and the cORF-coding sequence continues in the reading frame +1 relative to that of env [15, 16; see below]. cORF has also been called Rec [17].

[0087] Examples of cORF nucleotide sequences are: SEQ ID 89 and SEQ ID 90 [HERVK(108)]

[0088] Examples of cORF polypeptide sequences are SEQ ID 109.

C.1 - Direct detection of HML-2 polypeptides

[0089] Various techniques are available for detecting the presence or absence of a particular polypeptides in a sample. These are generally immunoassay techniques which are based on the specific interaction between an antibody and an antigenic amino acid sequence in the polypeptide. Suitable techniques include standard immunohistological methods, immunoprecipitation, immunofluorescence, ELISA, RIA, FIA, *etc.*

[0090] In general, therefore, the invention provides a method for detecting the presence of and/or measuring a level of a polypeptide of the invention in a biological sample, wherein the method uses an antibody specific for the polypeptide. The method generally comprises the steps of: a) contacting the sample with an antibody specific for the polypeptide; and b) detecting binding between the antibody and polypeptides in the sample.

[0091] Polypeptides of the invention can also be detected by functional assays *e.g.* assays to detect binding activity or enzymatic activity. For instance, a functional assay for cORF is disclosed in references 16, 129 & 130. A functional assay for the protease is disclosed in reference 140.

[0092] Another way for detecting polypeptides of the invention is to use standard proteomics techniques *e.g.* purify or separate polypeptides and then use peptide sequencing. For example, polypeptides can be separated using 2D-PAGE and polypeptide spots can be sequenced (*e.g.* by mass spectroscopy) in order to identify if a sequence is present in a target polypeptide.

[0093] Detection methods may be adapted for *use in vivo* (*e.g.* to locate or identify sites where cancer cells are present). In these embodiments, an antibody specific for a target polypeptide is administered to an individual (e.g. by injection) and the antibody is located using standard imaging techniques (*e.g.* magnetic resonance imaging, computed tomography scanning, *etc.*). Appropriate labels (*e.g.* spin labels *etc.*) will be used. Using these techniques, cancer cells are differentially labeled.

[0094] An immunofluorescence assay can be easily performed on cells without the need for purification of the target polypeptide. The cells are first fixed onto a solid support, such as a microscope slide or microtiter well. The membranes of the cells are then permeablized in order to permit entry of polypeptide-specific antibody (NB: fixing and permeabilization can be achieved together). Next, the fixed cells are exposed to an antibody which is specific for the encoded polypeptideand which is fluorescently labeled. The presence of this label (*e.g.* visualized under a microscope) identifies cells which express the target PCAV polypeptide. To increase the sensitivity of the assay, it is possible to use a second antibody to bind to the anti-PCAV antibody, with the label being carried by the second antibody. [18]

C.2 - Indirect detection of HML-2 polypeptides

[0095] Rather than detect polypeptides directly, it may be preferred to detect molecules which are produced by the body in response to a polypeptide *(i.e.* indirect detection of a polypeptide). This will typically involve the detection of antibodies, so the patient sample will generally be a blood sample. Antibodies can be detected by conventional immunoassay techniques *e.g.* using PCAV polypeptides of the invention, which will typically be immobilized.

[0096] Antibodies against HERV-K polypeptides have been detected in humans [143].

C.3 - Polypeptide materials

[0097] The invention provides polypeptides for use in the detection methods of the invention. In general, these polypep-

tides will be encoded by PCA-mRNAs *e.g.* by sequence(s) in the -N_4-region.

[0098] The invention provides an isolated polypeptide comprising: (a) an amino acid sequence selected from the group consisting of SEQ IDs 109 (cORF), 146 (gag), 147 (prt), 148 (pol), 149 (env); (b) a fragment of at least *x* amino acids of (a); or (c) a polypeptide sequence having at least *s*% identity to (a). These polypeptides include variants (*e.g.* allelic variants, homologs, orthologs, mutants *etc.*).

[0099] The value of *x* is at least 5 (*e.g.* at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 75, 80, 90, 100 *etc.*). The value of *x* may be less than 2000 (*e.g.* less than 1000, 500, 100, or 50).

[0100] The value of *s* is preferably at least 50 (*e.g.* at least 55, 60, 65, 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.5, 99.9 etc.).

[0101] The invention also provides an isolated polypeptide having formula NH_2-A-B-C-COOH, wherein: A is a polypeptide sequence consisting of a amino acids; C is a polypeptide sequence consisting of c amino acids; B is a polypeptide sequence consisting of a fragment of b amino acids of an amino acid sequence selected from the group consisting of SEQ IDs 109, 146, 147, 148, 149; and said polypeptide is not a fragment of polypeptide sequence SEQ ID 109, 146, 147, 148 or 149.

[0102] The value of *a+c* is at least 1 (*e.g.* at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.*). The value of b is at least 5 (*e.g.* at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.*). It is preferred that the value of a+b+c is at least 9 (*e.g.* at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50,60, 70, 80, 90,100 *etc.*). It is preferred that the value of *a+b+c* is at most 500 (*e.g.* at most 450, 400, 350, 300, 250, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9).

[0103] The amino acid sequence of -A- typically shares less than *n*% sequence identity to the *a* amino acids which are N-terminal of sequence -B- in SEQ ID 109, 146, 147, 148 or 149 and the amino acid sequence of -C- typically shares less than *n*% sequence identity to the c amino acids which are C-terminal of sequence -B- in SEQ ID 109, 146, 147, 148 or 149. The value of *n* is generally 60 or less (*e.g.* 50, 40, 30, 20, 10 or less).

[0104] The fragment of (b) may comprise a T-cell or, preferably, a B-cell epitope of SEQ ID 109, 146, 147, 148 or 149. T- and B-cell epitopes can be identified empirically (*e.g.* using the PEPSCAN method [19, 20] or similar methods), or they can be predicted (*e.g.* using the Jameson-Wolf antigenic index [21], matrix-based approaches [22], TEPITOPE [23], neural networks [24], OptiMer & EpiMer [25, 26], ADEPT [27], Tsites [28], hydrophilicity [29], antigenic index [30] or the methods disclosed in reference 31 *etc.*).

[0105] References to a percentage sequence identity between two amino acid sequences means that, when aligned, that percentage of amino acids are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art, for example those described in section 7.7.18 of reference 11. A preferred alignment is determined by the Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 2, BLOSUM matrix of 62. The Smith-Waterman homology search algorithm is taught in reference 32.

[0106] Polypeptides of the invention can be prepared in many ways *e.g.* by chemical synthesis (at least in part), by digesting longer polypeptides using proteases, by translation from RNA, by purification from cell culture (*e.g.* from recombinant expression), from the organism itself (*e.g.* isolation from prostate tissue), from a cell line source *etc.*

[0107] Polypeptides of the invention can be prepared in various forms (*e.g.* native, fusions, glycosylated, non-glycosylated *etc.*).

[0108] Polypeptides of the invention may be attached to a solid support.

[0109] Polypeptides of the invention may comprise a detectable label (*e.g.* a radioactive or fluorescent label, or a biotin label).

[0110] In general, the polypeptides of the subject invention are provided in a non-naturally occurring environment *e.g.* they are separated from their naturally-occurring environment. In certain embodiments, the subject polypeptide is present in a composition that is enriched for the polypeptide as compared to a control. As such, purified polypeptide is provided, whereby purified is meant that the polypeptide is present in a composition that is substantially free of other expressed polypeptides, where by substantially free is meant that less than 90%, usually less than 60% and more usually less than 50% of the composition is made up of other expressed polypeptides.

[0111] The term "polypeptide" refers to amino acid polymers of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified naturally or by intervention; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation or modification, such as conjugation with a labeling component. Also included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, *etc.*), as well as other modifications known in the art. Polypeptides can occur as single chains or associated chains. Polypeptides of the invention can be naturally or non-naturally glycosylated (*i.e.* the polypeptide has a glycosylation pattern that differs from the glycosylation pattern found

in the corresponding naturally occurring polypeptide).

[0112] Mutants can include amino acid substitutions, additions or deletions. The amino acid substitutions can be conservative amino acid substitutions or substitutions to eliminate non-essential amino acids, such as to alter a glycosylation site, a phosphorylation site or an acetylation site, or to minimize misfolding by substitution or deletion of one or more cysteine residues that are not necessary for function. Conservative amino acid substitutions are those that preserve the general charge, hydrophobicity/hydrophilicity, and/or steric bulk of the amino acid substituted. Variants can be designed so as to retain or have enhanced biological activity of a particular region of the polypeptide (*e.g.* a functional domain and/or, where the polypeptide is a member of a polypeptide family, a region associated with a consensus sequence). Selection of amino acid alterations for production of variants can be based upon the accessibility (interior vs. exterior) of the amino acid (*e.g.* ref. 33), the thermostability of the variant polypeptide (*e.g.* ref. 34), desired glycosylation sites (*e.g.* ref. 35), desired disulfide bridges (*e.g.* refs. 36 & 37), desired metal binding sites (*e.g.* refs.38 & 39), and desired substitutions with in proline loops (*e.g.* ref. 40). Cysteine-depleted muteins can be produced as disclosed in reference 41.

C.4 - Antibody materials

[0113] The invention also provides isolated antibodies, or antigen-binding fragments thereof, that bind to a polypeptide of the invention. The invention also provides isolated antibodies or antigen binding fragments thereof, that bind to a polypeptide encoded by a polynucleotide of the invention.

[0114] Antibodies of the invention may be polyclonal or monoclonal and may be produced by any suitable means (*e.g.* by recombinant expression).

[0115] Antibodies of the invention may include a label. The label may be detectable directly, such as a radioactive or fluorescent label. Alternatively, the label may be detectable indirectly, such as an enzyme whose products are detectable (*e.g.* luciferase, ß-galactosidase, peroxidase *etc.*).

[0116] Antibodies of the invention may be attached to a solid support.

[0117] Antibodies of the invention may be prepared by administering (*e.g.* injecting) a polypeptide of the invention to an appropriate animal (*e.g.* a rabbit, hamster, mouse or other rodent).

[0118] Antigen-binding fragments of antibodies include Fv, scFv, Fc, Fab, $F(ab')_2$ *etc.*

[0119] To increase compatibility with the human immune system, the antibodies may be chimeric or humanized [*e.g.* refs. 42 & 43], or fully human antibodies may be used. Because humanized antibodies are far less immunogenic in humans than the original non-human monoclonal antibodies, they can be used for the treatment of humans with far less risk of anaphylaxis. Thus, these antibodies may be preferred in therapeutic applications that involve *in vivo* administration to a human such as, use as radiation sensitizers for the treatment of neoplastic disease or use in methods to reduce the side effects of cancer therapy.

[0120] Humanized antibodies may be achieved by a variety of methods including, for example: (1) grafting non-human complementarity determining regions (CDRs) onto a human framework and constant region ("humanizing"), with the optional transfer of one or more framework residues from the non-human antibody; (2) transplanting entire non-human variable domains, but "cloaking" them with a human-like surface by replacement of surface residues ("veneering"). In the present invention, humanized antibodies will include both "humanized" and "veneered" antibodies. [44, 45, 46, 47, 48, 49, 50].

[0121] CDRs are amino acid sequences which together define the binding affinity and specificity of a Fv region of a native immunoglobulin binding site [*e.g.* refs. 51 & 52].

[0122] The phrase "constant region" refers to the portion of the antibody molecule that confers effector functions. In chimeric antibodies, mouse constant regions are substituted by human constant regions. The constant regions of humanized antibodies are derived from human immunoglobulins. The heavy chain constant region can be selected from any of the 5 isotypes: alpha, delta, epsilon, gamma or mu.

[0123] One method of humanizing antibodies comprises aligning the heavy and light chain sequences of a non-human antibody to human heavy and light chain sequences, replacing the non-human framework residues with human framework residues based on such alignment, molecular modeling of the conformation of the humanized sequence in comparison to the conformation of the non-human parent antibody, and repeated back mutation of residues in the framework region which disturb the structure of the non-human CDRs until the predicted conformation of the CDRs in the humanized sequence model closely approximates the conformation of the non-human CDRs of the parent non-human antibody. Such humanized antibodies may be further derivatized to facilitate uptake and clearance *e.g*, via Ashwell receptors. [refs. 53 & 54]

[0124] Humanized or fully-human antibodies can also be produced using transgenic animals that are engineered to contain human immunoglobulin loci. For example, ref. 55 discloses transgenic animals having a human Ig locus wherein the animals do not produce functional endogenous immunoglobulins due to the inactivation of endogenous heavy and light chain loci. Ref. 56 also discloses transgenic non-primate mammalian hosts capable of mounting an immune response

to an immunogen, wherein the antibodies have primate constant and/or variable regions, and wherein the endogenous immunoglobulin-encoding loci are substituted or inactivated. Ref. 57 discloses the use of the Cre/Lox system to modify the immunoglobulin locus in a mammal, such as to replace all or a portion of the constant or variable region to form a modified antibody molecule. Ref. 58 discloses non-human mammalian hosts having inactivated endogenous Ig loci and functional human Ig loci. Ref. 59 discloses methods of making transgenic mice in which the mice lack endogenous heavy claims, and express an exogenous immunoglobulin locus comprising one or more xenogeneic constant regions.

[0125] Using a transgenic animal described above, an immune response can be produced to a PCAV polypeptide, and antibody-producing cells can be removed from the animal and used to produce hybridomas that secrete human monoclonal antibodies. Immunization protocols, adjuvants, and the like are known in the art, and are used in immunization of, for example, a transgenic mouse as described in ref. 60. The monoclonal antibodies can be tested for the ability to inhibit or neutralize the biological activity or physiological effect of the corresponding polypeptide.

D - COMPARISON WITH CONTROL SAMPLES

D.1 - The control

[0126] HML-2 transcripts are up-regulated in tumors, including prostate tumors. To detect such up-regulation, a reference point is needed *i.e.* a control. Analysis of the control sample gives a standard level of RNA and/or protein expression against which a patient sample can be compared.

[0127] A negative control gives a background or basal level of expression against which a patient sample can be compared. Higher levels of expression product relative to a negative control indicate that the patient from whom the sample was taken has, for example, prostate cancer. Typically, for prostate cancer, for example, negative controls would include lifetime baseline levels of expression or the expression level observed in pooled normals. Conversely, equivalent levels of expression product indicate that the patient does not have a HML-2-related cancer such as prostate cancer.

[0128] A positive control gives a level of expression against which a patient sample can be compared. Equivalent or higher levels of expression product relative to a positive control indicate that the patient from whom the sample was taken has cancer such as prostate cancer. Conversely, lower levels of expression product indicate that the patient does not have a HML-2 related cancer such as prostate cancer.

[0129] For direct or indirect RNA measurement, or for direct polypeptide measurement, a negative control will generally comprise cells which are not from a tumor cell, *e.g.* a prostate tumor cell. For indirect polypeptide measurement, a negative control will generally be a blood sample from a patient who does not have a prostate tumor. The negative control could be a sample from the same patient as the patient sample, but from a tissue in which HML-2 expression is not up-regulated *e.g.* a non-tumor non-prostate cell. The negative control could be a prostate cell from the same patient as the patient sample, but taken at an earlier stage in the patient's life. The negative control could be a cell from a patient without a prostate tumor. This cell may or may not be a prostate cell. The negative control cell could be a prostate cell from a patient with BPH.

[0130] For direct or indirect RNA measurement, or for direct polypeptide measurement, a positive control will generally comprise cells from a tumor cell *e.g.* a prostate tumor. For indirect polypeptide measurement, a negative control will generally be a blood sample from a patient who has a prostate tumor. The positive control could be a prostate tumor cell from the same patient as the patient sample, but taken at an earlier stage in the patient's life (*e.g.* to monitor remission). The positive control could be a cell from another patient with a prostate tumor. The positive control could be a prostate cell line.

[0131] Other suitable positive and negative controls will be apparent to the skilled person.

[0132] HML-2 expression in the control can be assessed at the same time as expression in the patient sample. Alternatively, HML-2 expression in the control can be assessed separately (earlier or later).

[0133] Rather than actually compare two samples, however, the control may be an absolute value *i.e.* a level of expression which has been empirically determined from samples taken from prostate tumor patients (*e.g.* under standard conditions).

D.2 - Degree of up-regulation

[0134] The up-regulation relative to the control (100%) will usually be at least 150% (*e.g.* 200%, 250%, 300%, 400%, 500%, 600% or more).

D.3 -Diagnosis

[0135] The invention provides a method for diagnosing prostate cancer. It will be appreciated that "diagnosis" according to the invention can range from a definite clinical diagnosis of disease to an indication that the patient should undergo

further testing which may lead to a definite diagnosis. For example, the method of the invention can be used as part of a screening process, with positive samples being subjected to further analysis.

[0136] Furthermore, diagnosis includes monitoring the progress of cancer in a patient already known to have the cancer. Cancer can also be staged by the methods of the invention. Preferably, the cancer is prostate cancer.

[0137] The efficacy of a treatment regimen (therametrics) of a cancer associated can also monitored by the method of the invention *e.g.* to determine its efficacy.

[0138] Susceptibility to a cancer can also be detected *e.g.* where up-regulation of expression has occurred, but before cancer has developed. Prognostic methods are also encompassed.

[0139] All of these techniques fall within the general meaning of "diagnosis" in the present invention.

E - PHARMACEUTICAL COMPOSITIONS

[0140] The invention provides a pharmaceutical composition comprising polynucleotide, polypeptide, or antibody as defined above. The invention also provides their use as medicaments, and their use in the manufacture of medicaments for treating prostate cancer. The invention also provides a method for raising an immune response, comprising administering an immunogenic dose of polynucleotide or polypeptide of the invention to an animal.

[0141] Pharmaceutical compositions encompassed by the present invention include as active agent, the polynucleotides, polypeptides, or antibodies of the invention disclosed herein in a therapeutically effective amount. An "effective amount" is an amount sufficient to effect beneficial or desired results, including clinical results. An effective amount can be administered in one or more administrations. For purposes of this invention, an effective amount is an amount that is sufficient to palliate, ameliorate, stabilize, reverse, slow or delay the symptoms and/or progression of prostate cancer.

[0142] The compositions can be used to treat cancer as well as metastases of primary cancer. In addition, the pharmaceutical compositions can be used in conjunction with conventional methods of cancer treatment, *e.g.* to sensitize tumors to radiation or conventional chemotherapy. The terms "treatment", "treating", "treat" and the like are used herein to generally refer to obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete stabilization or cure for a disease and/or adverse effect attributable to the disease. "Treatment" as used herein covers any treatment of a disease in a mammal, particularly a human, and includes: (a) preventing the disease or symptom from occurring in a subject which may be predisposed to the disease or symptom but has not yet been diagnosed as having it; (b) inhibiting the disease symptom, i. e. arresting its development; or (c) relieving the disease symptom, *i.e.* causing regression of the disease or symptom.

[0143] Where the pharmaceutical composition comprises an antibody that specifically binds to a gene product encoded by a differentially expressed polynucleotide, the antibody can be coupled to a drug for delivery to a treatment site or coupled to a detectable label to facilitate imaging of a site comprising cancer cells, such as prostate cancer cells. Methods for coupling antibodies to drugs and detectable labels are well known in the art, as are methods for imaging using detectable labels.

[0144] The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. The effective amount for a given situation is determined by routine experimentation and is within the judgment of the clinician. For purposes of the present invention, an effective dose will generally be from about 0.01mg/kg to about 5 mg/kg, or about 0.01 mg/ kg to about 50 mg/kg or about 0.05 mg/kg to about 10 mg/kg of the compositions of the present invention in the individual to which it is administered.

[0145] A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which can be administered without undue toxicity. Suitable carriers can be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Pharmaceutically acceptable carriers in therapeutic compositions can include liquids such as water, saline, glycerol and ethanol. Auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, can also be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier. Pharmaceutically acceptable salts can also be present in the pharmaceutical composition, e.g. mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such

as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington: The Science and Practice of Pharmacy (1995) Alfonso Gennaro, Lippincott, Williams, & Wilkins.

[0146] The composition is preferably sterile and/or pyrogen-free. It will typically be buffered around pH 7.

[0147] Once formulated, the compositions contemplated by the invention can be (1) administered directly to the subject (*e.g.* as polynucleotide, polypeptides, small molecule agonists or antagonists, and the like); or (2) delivered *ex vivo,* to cells derived from the subject (*e.g.* as in *ex vivo* gene therapy). Direct delivery of the compositions will generally be accomplished by parenteral injection, e.g. subcutaneously, intraperitoneally, intravenously or intramuscularly, intratumoral or to the interstitial space of a tissue. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal applications, needles, and gene guns or hyposprays. Dosage treatment can be a single dose schedule or a multiple dose schedule.

[0148] Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art [*e.g.* ref. 61]. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoetic, lymph cells, macrophages, dendritic cells, or tumor cells. Generally, delivery of nucleic acids for both ex vivo and in vitro applications can be accomplished by, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

[0149] Differential expression PCAV polynucleotides has been found to correlate with prostate tumors. The tumor can be amenable to treatment by administration of a therapeutic agent based on the provided polynucleotide, corresponding polypeptide or other corresponding molecule (*e.g.* antisense, ribozyme, *etc.*). In other embodiments, the disorder can be amenable to treatment by administration of a small molecule drug that, for example, serves as an inhibitor (antagonist) of the function of the encoded gene product of a gene having increased expression in cancerous cells relative to normal cells or as an agonist for gene products that are decreased in expression in cancerous cells (*e.g.* to promote the activity of gene products that act as tumor suppressors).

[0150] The dose and the means of administration of the inventive pharmaceutical compositions are determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the progression of the disease, and other relevant factors. For example, administration of polynucleotide therapeutic compositions agents includes local or systemic administration, including injection, oral administration, particle gun or catheterized administration, and topical administration. Preferably, the therapeutic polynucleotide composition contains an expression construct comprising a promoter operably linked to a polynucleotide of the invention. Various methods can be used to administer the therapeutic composition directly to a specific site in the body. For example, a small metastatic lesion is located and the therapeutic composition injected several times in several different locations within the body of tumor. Alternatively, arteries which serve a tumor are identified, and the therapeutic composition injected into such an artery, in order to deliver the composition directly into the tumor. A tumor that has a necrotic center is aspirated and the composition injected directly into the now empty center of the tumor. An antisense composition is directly administered to the surface of the tumor, for example, by topical application of the composition. X-ray imaging is used to assist in certain of the above delivery methods.

[0151] Targeted delivery of therapeutic compositions containing an antisense polynucleotide, subgenomic polynucleotides, or antibodies to specific tissues can also be used. Receptor-mediated DNA delivery techniques are described in, for example, references 62 to 67. Therapeutic compositions containing a polynucleotide are administered in a range of about 100 ng to about 200 mg of DNA for local administration in a gene therapy protocol. Concentration ranges of about 500 ng to about 50 mg, about 1 μg to about 2 mg, about 5 μg to about 500 μg, and about 20 μg to about 100 μg of DNA can also be used during a gene therapy protocol. Factors such as method of action (*e.g.* for enhancing or inhibiting levels of the encoded gene product) and efficacy of transformation and expression are considerations which will affect the dosage required for ultimate efficacy of the antisense subgenomic polynucleotides. Where greater expression is desired over a larger area of tissue, larger amounts of antisense subgenomic polynucleotides or the same amounts re-administered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of, for example, a tumor site, may be required to effect a positive therapeutic outcome. In all cases, routine experimentation in clinical trials will determine specific ranges for optimal therapeutic effect.

[0152] The therapeutic polynucleotides and polypeptides of the present invention can be delivered using gene delivery vehicles. The gene delivery vehicle can be of viral or non-viral origin (see generally references 68, 69, 70 and 71). Expression of such coding sequences can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence can be either constitutive or regulated.

[0153] Viral-based vectors for delivery of a desired polynucleotide and expression in a desired cell are well known in the art. Exemplary viral-based vehicles include, but are not limited to, recombinant retroviruses (*e.g.* references 72 to 82), alphavirus-based vectors (e.g. Sindbis virus vectors, Semliki forest virus (ATCC VR-67; ATCC VR-1247), Ross River virus (ATCC VR-373; ATCC VR-1246) and Venezuelan equine encephalitis virus (ATCC VR-923; ATCC VR-1250; ATCC VR 1249; ATCC VR-532)), adenovirus vectors, and adeno-associated virus (AAV) vectors (*e.g.* see refs. 83 to

88). Administration of DNA linked to killed adenovirus [89] can also be employed.

[0154] Non-viral delivery vehicles and methods can also be employed, including, but not limited to, polycationic condensed DNA linked or unlinked to killed adenovirus alone [*e.g.* 89], ligand-linked DNA [90], eukaryotic cell delivery vehicles cells [*e.g.* refs. 91 to 95] and nucleic charge neutralization or fusion with cell membranes. Naked DNA can also be employed. Exemplary naked DNA introduction methods are described in refs. 96 and 97. Liposomes that can act as gene delivery vehicles are described in refs. 98 to 102. Additional approaches are described in refs. 103 & 104.

[0155] Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in ref. 104. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials or use of ionizing radiation [*e.g.* refs. 105 & 106]. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun [107] or use of ionizing radiation for activating transferred gene [108 & 109].

Vaccine compositions

[0156] The invention provides a composition comprising a polypeptide or polynucleotide of the invention and a pharmaceutically acceptable carrier.

[0157] The composition may additionally comprise an adjuvant. For example, the composition may comprise one or more of the following adjuvants: (1) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ [110; Chapter 10 in ref. 111], containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing MTP-PE) formulated into submicron particles using a microfluidizer, (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (DetoxTM); (2) saponin adjuvants, such as QS21 or StimulonTM (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes), which ISCOMS may be devoid of additional detergent [112]; (3) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (4) cytokines, such as interleukins (e.g. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12 *etc.*), interferons (e.g. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc.; (5) monophosphoryl lipid A (MPL) or 3-O-deacylated MPL (3dMPL) [*e.g.* 113, 114]; (6) combinations of 3dMPL with, for example, QS21 and/or oil-in-water emulsions [*e.g.* 115, 116, 117]; (7) oligonucleotides comprising CpG motifs *i.e.* containing at least one CG dinucleotide, with 5-methylcytosine optionally being used in place of cytosine; (8) a polyoxyethylene ether or a polyoxyethylene ester [118]; (9) a polyoxyethylene sorbitan ester surfactant in combination with an octoxynol [119] or a polyoxyethylene alkyl ether or ester surfactant in combination with at least one additional non-ionic surfactant such as an octoxynol [120]; (10) an immunostimulatory oligonucleotide (*e.g.* a CpG oligonucleotide) and a saponin [121]; (11) an immunostimulant and a particle of metal salt [122]; (12) a saponin and an oil-in-water emulsion [123]; (13) a saponin (*e.g.* QS21) + 3dMPL + IL-12 (optionally + a sterol) [124]; (14) aluminium salts, preferably hydroxide or phosphate, but any other suitable salt may also be used (*e.g.* hydroxyphosphate, oxyhydroxide, orthophosphate, sulphate etc. [chapters 8 & 9 of ref. 111]). Mixtures of different aluminium salts may also be used. The salt may take any suitable form (*e.g.* gel, crystalline, amorphous etc.); (15) chitosan; (16) cholera toxin *or E.coli* heat labile toxin, or detoxified mutants thereof [125]; (17) microparticles of poly(α-hydroxy)acids, such as PLG; (18) other substances that act as immunostimulating agents to enhance the efficacy of the composition. Aluminium salts and/or MF59™ are preferred.

[0158] The composition is preferably sterile and/or pyrogen-free. It will typically be buffered around pH 7.

[0159] The composition is preferably an immunogenic composition and is more preferably a vaccine composition. The composition can be used to raise antibodies in a mammal (*e.g.* a human).

[0160] Vaccines of the invention may be prophylactic (*i.e.* to prevent disease) or therapeutic (*i.e.* to reduce or eliminate the symptoms of a disease).

[0161] Efficacy can be tested by monitoring expression of polynucleotides and/or polypeptides of the invention after administration of the composition of the invention.

F - SCREENING METHODS AND DRUG DESIGN

[0162] The invention provides methods of screening for compounds with activity against cancer, comprising: contacting a test compound with a tissue sample derived from a cell in which HML-2 expression is up-regulated; or a cell line; and monitoring HML-2 expression in the sample. A decrease in expression indicates potential anti-cancer efficacy of the test compound.

[0163] The invention also provides methods of screening for compounds with activity against prostate cancer, com-

prising: contacting a test compound with a polynucleotide or polypeptide of the invention; and detecting a binding interaction between the test compound and the polynucleotide/polypeptide. A binding interaction indicates potential anti-cancer efficacy of the test compound.

[0164] The invention also provides methods of screening for compounds with activity against prostate cancer, comprising: contacting a test compound with a polypeptide of the invention; and assaying the function of the polypeptide. Inhibition of the polypeptide's function (*e.g.* loss of protease activity, loss of RNA export, loss of reverse transcriptase activity, loss of endonuclease activity, loss of integrase activity *etc.*) indicates potential anti-cancer efficacy of the test compound.

[0165] Typical test compounds include, but are not restricted to, peptides, peptoids, proteins, lipids, metals, nucleotides, nucleosides, small organic molecules, antibiotics, polyamines, and combinations and derivatives thereof. Small organic molecules have a molecular weight of more than 50 and less than about 2,500 daltons, and most preferably between about 300 and about 800 daltons. Complex mixtures of substances, such as extracts containing natural products, or the products of mixed combinatorial syntheses, can also be tested and the component that binds to the target RNA can be purified from the mixture in a subsequent step.

[0166] Test compounds may be derived from large libraries of synthetic or natural compounds. For instance, synthetic compound libraries are commercially available from Maybridge Chemical Co. (Trevillet, Cornwall, UK) or Aldrich (Milwaukee, WI). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts may be used. Additionally, test compounds may be synthetically produced using combinatorial chemistry either as individual compounds or as mixtures.

[0167] Agonists or antagonists of the polypeptides of the invention can be screened using any available method known in the art, such as signal transduction, antibody binding, receptor binding, mitogenic assays, chemotaxis assays, *etc.*. The assay conditions ideally should resemble the conditions under which the native activity is exhibited *in vivo,* that is, under physiologic pH, temperature, and ionic strength. Suitable agonists or antagonists will exhibit strong inhibition or enhancement of the native activity at concentrations that do not cause toxic side effects in the subject. Agonists or antagonists that compete for binding to the native polypeptide can require concentrations equal to or greater than the native concentration, while inhibitors capable of binding irreversibly to the polypeptide can be added in concentrations on the order of the native concentration.

[0168] Such screening and experimentation can lead to identification of an agonist or antagonist of a HML-2 polypeptide. Such agonists and antagonists can be used to modulate, enhance, or inhibit HML-2 expression and/or function. [126]

[0169] The present invention relates to methods of using the polypeptides of the invention (e.g. recombinantly produced HML-2 polypeptides) to screen compounds for their ability to bind or otherwise modulate, such as, inhibit, the activity of HML-2 polypeptides, and thus to identify compounds that can serve, for example, as agonists or antagonists of the HML-2 polypeptides. In one screening assay, the HML-2 polypeptide is incubated with cells susceptible to the growth stimulatory activity of HML-2, in the presence and absence of a test compound. The HML-2 activity altering or binding potential of the test compound is measured. Growth of the cells is then determined. A reduction in cell growth in the test sample indicates that the test compound binds to and thereby inactivates the HML-2 polypeptide, or otherwise inhibits the HML-2 polypeptide activity.

[0170] Transgenic animals (e.g. rodents) that have been transformed to over-express HML-2 genes can be used to screen compounds *in vivo* for the ability to inhibit development of tumors resulting from HML-2 over-expression or to treat such tumors once developed. Transgenic animals that have prostate tumors of increased invasive or malignant potential can be used to screen compounds, including antibodies or peptides, for their ability to inhibit the effect of HML-2 polypeptides. Such animals can be produced, for example, as described in the examples herein.

[0171] Screening procedures such as those described above are useful for identifying agents for their potential use in pharmacological intervention strategies in prostate cancer treatment. Additionally, polynucleotide sequences corresponding to HML-2, including LTRs, may be used to assay for inhibitors of elevated gene expression.

[0172] Potent inhibitors of HERV-K protease are already known [127]. Inhibition of HERV-K protease by HIV-1 protease inhibitors has also been reported [128]. These compounds can be studied for use in prostate cancer therapy, and are also useful lead compounds for drug design.

[0173] Transdominant negative mutants of cORF have also been reported [129,130]. Transdominant cORF mutants can be studied for use in prostate cancer therapy.

[0174] Antisense oligonucleotides complementary to HML-2 mRNA can be used to selectively diminish or oblate the expression of the polypeptide. More specifically, antisense constructs or antisense oligonucleotides can be used to inhibit the production of HML-2 polypeptide(s) in prostate tumor cells. Antisense mRNA can be produced by transfecting into target cancer cells an expression vector with a HML-2 polynucleotide of the invention oriented in an antisense direction relative to the direction of PCAV-mRNA transcription. Appropriate vectors include viral vectors, including retroviral vectors, as well as non-viral vectors. Alternately, antisense oligonucleotides can be introduced directly into target cells to achieve the same goal. Oligonucleotides can be selected/designed to achieve the highest level of specificity and, for example, to bind to a PCAV-mRNA at the initiator ATG.

[0175] Monoclonal antibodies to HML-2 polypeptides can be used to block the action of the polypeptides and thereby control growth of cancer cells. This can be accomplished by infusion of antibodies that bind to HML-2 polypeptides and block their action.

[0176] The invention also provides high-throughput screening methods for identifying compounds that bind to a polynucleotide or polypeptide of the invention. Preferably, all the biochemical steps for this assay are performed in a single solution in, for instance, a test tube or microtitre plate, and the test compounds are analyzed initially at a single compound concentration. for the purposes of high throughput screening, the experimental conditions are adjusted to achieve a proportion of test compounds identified as "positive" compounds from amongst the total compounds screened. The assay is preferably set to identify compounds with an appreciable affinity towards the target *e.g.*, when 0.1% to 1% of the total test compounds from a large compound library are shown to bind to a given target with a K_i of 10μM or less (*e.g.* 1μM, 100nM, 10nM, or less)

G - THE HML-2 FAMILY OF HUMAN ENDOGENOUS RETROVIRUSES

[0177] Genomes of all eukaryotes contain multiple copies of sequences related to infectious retroviruses. These endogenous retroviruses have been well studied in mice where both true infectious forms and thousands of defective retrovirus-like elements (*e.g.* the IAP and Etn sequence families) exist. Some members of the IAP and Etn families are "active" retrotransposons since insertions of these elements have been documented which cause germ line mutations or oncogenic transformation.

[0178] Endogenous retroviruses were identified in human genomic DNA by their homology to retroviruses of other vertebrates [131, 132]. It is believed that the human genome probably contains numerous copies of endogenous proviral DNAs, but little is known about their function. Most HERV families have relatively few members (1-50) but one family (HERV-H) consists of ~1000 copies per haploid genome distributed on all chromosomes. The large numbers and general transcriptional activity of HERVs in embryonic and tumor cell lines suggest that they could act as disease-causing insertional mutagens or affect adjacent gene expression in a neutral or beneficial way.

[0179] The K family of human endogenous retroviruses (HERV-K) is well known [133]. It is related to the mouse mammary tumor virus (MMTV) and is present in the genomes of humans, apes and old world monkeys, but several human HERV-K proviruses are unique to humans [134]. The HERV-K family is present at 30-50 full-length copies per haploid human genome and possesses long open reading frames that potentially are translated into viral proteins [135, 136]. Two types of proviral genomes are known, which differ by the presence (type 2) or absence (type 1) of a stretch of 292 nucleotides in the overlapping boundary of the pol and env genes [137]. Some members of the HERV-K family are known to code for the gag protein and retroviral particles, which are both detectable in germ cell tumors and derived cell lines [138]. Analysis of the RNA expression pattern of full-length HERV-K has also identified a doubly-spliced RNA that encodes a 105 amino acid protein termed central ORF ('cORF') which is a sequence-specific nuclear RNA export factor that is functionally equivalent to the Rev protein of HIV [139]. HERV-K10 has been shown to encode a full-length gag homologous 73 kDa protein and a functional protease [140].

[0180] Patients suffering from germ cell tumors show high antibody titers against HERV-K gag and env proteins at the time of tumor detection [141]. In normal testis and testicular tumors the HERV-K transmembrane envelope protein has been detected both in germ cells and tumor cells, but not in the surrounding tissue. In the case of testicular tumor, correlations between the expression of the env-specific mRNA, the presence of the transmembrane env, cORF and gag proteins and antibodies against HERV-K specific peptides in the serum of the patients, have been reported. Reference 142 reports that HERV-K10 gag and/or env proteins are synthesized in seminoma cells and that patients with those tumors exhibit relatively high antibody titers against gag and/or env.

[0181] Gag proteins released in form of particles from HERV-K have been identified in the cell culture supernatant of the teratocarcinoma derived cell line Tera 1. These retrovirus-like particles (termed "human teratocarcinoma derived virus" or HTDV) have been shown to have a 90% sequence homology to the HERV-K10 genome [138, 143].

[0182] While the HERV-K family is present in the genome of every human cell, a high level of expression of mRNAs, proteins and particles is observed only in human teratocarcinoma cell lines [144]. In other tissues and cell lines, only a basal level of expression of mRNA has been demonstrated even using very sensitive methods. The expression of retroviral proviruses is generally regulated by elements of the 5' long terminal repeat (LTR). Furthermore, the activation of expression of an endogenous retrovirus may trigger the expression of a downstream gene that triggers a neoplastic effect.

[0183] The sequence of HERV-K(II), which locates to chromosome 3, has been disclosed [145].

[0184] HML-2 is a subgroup of the HERV-K family [146]. HERV isolates which are members of the HML-2 subgroup include HERV-K10 [137,142], the 27 HML-2 viruses shown in Figure 4 of reference 147, HERV-K(C7) [148], HERV-K (II) [145], HERV-K(CH) Table 11 provides a list of all known members of the HML-2 subgroup of the HERV-K family as determined by searching the DoubleTwist database containing all genomic contigs with the sequence AF074086 using the Smith-Waterman algorithm with the default parameters: open gap penalty = -20 and extension penalty = -5.

[0185] The invention is based on the finding that HML-2 mRNA expression is up-regulated in prostate tumors. Because HML-2 is a well-recognized family, the skilled person will be able to determine without difficulty whether any particular endogenous retroviruses is or is not a HML-2. Preferred members of the HML-2 family for use in accordance with the present invention are those whose proviral genome has an LTR which has at least 75% sequence identity to SEQ ID 150 (the LTR sequence from HML-2.HOM [1]). Example LTRs include SEQ IDs 151-154.

H-HERV-K(CH)

[0186] The present invention is based on the discovery of elevated levels of multiple HML-2 polynucleotides in prostate tumor samples as compared to normal prostate tissue. One particular HML-2 whose mRNA was found to be up-regulated is designated herein as 'HERV-K(CH)'.

[0187] Sequences from HERV-K(CH) are shown in SEQ IDs 14-39 and have been deposited with the ATCC (see Table 7). The skilled person will be able to classify any further HERV as HERV-K(CH) or not based on sequence identity to these HERV-K(CH) polynucleotides. Preferably such a comparison is to one or more, or all, of the polynucleotide sequences disclosed herein or of the polynucleotide inserts in the ATCC-deposited isolates. Alternatively, the skilled artisan can determine the sequence identity based on a comparison to any one or more, or all, of the sequences in SEQ IDs 7-10 and SEQ IDs 14-39 taking into consideration the spontaneous mutation rate associated with retroviral replication. Thus, it will be apparent when the differences in the sequences are consistent with a HERV-K(CH) isolate or consistent with another HERV.

[0188] HERV-K(CH) is therefore a specific member of the HML-2 subgroup which can be used in the invention as described above. It can also be used in methods previously described in relation to HERV-K *e.g.* the diagnosis of testicular cancer [142], autoimmune diseases, multiple sclerosis [149], insulin-dependent diabetes mellitus (IDDM) [150] *etc.*

H.1- HERV-K(CH) Nucleic acids

H.1.1- HERV-K(CH) genomic sequences

[0189] The invention provides an isolated polynucleotide comprising: (a) the nucleotide sequence of any of SEQ IDs 7-10; (b) the nucleotide sequence of any of SEQ IDs 27-39; (c) the complement of a nucleotide sequence of any of SEQ IDs 7-10; or (d) the complement of the nucleotide sequence of any of SEQ IDs 27-39.

H.1.2 - HERV-K(CH) fragments

[0190] The invention also provides an isolated polynucleotide comprising a fragment of: (a) a nucleotide sequence shown in SEQ IDs 7-10; (b) the nucleotide sequence shown in any of SEQ IDs 27-39; (c) the complement of a nucleotide sequence shown in SEQ IDs 7-10; or (d) the complement of the nucleotide sequence shown in any of SEQ IDs 27-39.

[0191] The fragment is preferably at least x nucleotides in length, wherein x is at least 7 *(e.g.* at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 75, 80, 90, 100 *etc.*). The value ofx may be between about 150 and about 200 or be between about 250 and about 300. The value of x may be about 350, about 400, about 450, about 500, about 550, about 600, about 650, about 700, or about 750. The value of x may be less than 2000 (*e.g.* less than 1000, 500, 100, or 50).

[0192] The fragment is preferably neither one of the following sequences nor a fragment of one of the following sequences: (i) the nucleotide sequence shown in SEQ ID 42; (ii) the nucleotide sequence shown in SEQ ID 43; (iii) the nucleotide sequence shown in SEQ ID 44; (iv) the nucleotide sequence shown in SEQ ID 45; (v) a known polynucleotide; or (vi) a polynucleotide known as of 7th December 2000 (*e.g.* a polynucleotide available in a public database such as GenBank of GeneSeq before 7th December 2000).

[0193] The fragment is preferably a contiguous sequence of one of polynucleotides of (a), (b), (c) or (d) that remains unmasked following application of a masking program for masking low complexity (*e.g.* XBLAST) to the sequence (*i.e.* one would select an unmasked region, as indicated by the polynucleotides outside the poly-n stretches of the masked sequence produced by the masking program).

[0194] These polynucleotides are particularly useful as probes. In general, a probe in which x=15 represents sufficient sequence for unique identification. Probes can be used, for example, to determine the presence or absence of a polynucleotide of the invention (or variants thereof) in a sample. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes can be any species *e.g.* primate species, particularly human; rodents, such as rats and mice; canines; felines; bovines; ovines; equines; yeast; nematodes; *etc.*

[0195] Probes from more than one polynucleotide sequence of the invention can hybridize with the same nucleic acid if the nucleic acid from which they were derived corresponds to a single sequence (*e.g.* more than one can hybridize to

a single cDNA derived from the same mRNA).

[0196] Preferred fragments (*e.g.* for the identification of HERV-K(CH) polynucleotides associated with cancer) which do not correspond identically in their entirety to any portion of the sequence(s) shown in SEQ IDs 42-45 are: SEQ ID 59 (from gag region), SEQ IDs 60-70 (from pol region) and SEQ IDs 71-82 (from 3' pol region).

[0197] Preferred fragments (*e.g.* for the simultaneous identification of HERV-K(CH) polynucleotides, HERV-KII polynucleotides and/or HERV-K10 polynucleotides) which do correspond identically in their entirety to any portion of the sequence(s) shown in SEQ IDs 44 & 45 are SEQ IDs 83 & 84 (from gag region).

[0198] Polynucleotide probes unique to HERV-K(CH), HERV-KII and HERV-K10 gag regions are provided in Table 1; polynucleotide probes unique to HERV-K(CH), HERV-KII, and HERV-K10 protease 3' and polymerase 5' regions are provided in Table 2; polynucleotide probes unique to HERV-K(CH), HERV-KII, and HERV-K10 3' pol only regions are provided in Table 3.

H.1.3 - HERV-K(CH) fragments plus heterologous sequences

[0199] The invention also provides an isolated polynucleotide comprising (a) a segment that is a fragment of the sequence shown in SEQ IDs 7-10 or SEQ IDs 27-39, wherein (i) said fragment is at least 10 nucleotides in length and (ii) corresponds identically in its entirety to a portion of SEQ ID 44 and/or 45; and, optionally, (b) one or more segments flanking the segment defined in (a), wherein the presence of said optional segment(s) causes said polynucleotide to not correspond identically to any portion of a sequence shown in SEQ IDs 7- 10 or SEQ IDs 27-39. In some embodiments, the optional flanking segments share less than 40% sequence identity to the nucleic acid sequences shown in SEQ IDs 7-10, SEQ ID 44 and/or SEQ ID 45. In other embodiments, the optional flanking segments have no contiguous sequence of 10, 12, 15 or 20 nucleotides in common with SEQ IDs 7-10, SEQ ID 44 and/or SEQ ID 45. In yet other embodiments, the optional flanking segment is not present. In further embodiments, a fragment of the polynucleotide sequence is up to at least 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1000, or 1500 nucleotides in length.

[0200] The invention also provides an isolated polynucleotide having formula 5'-A-B-C-3', wherein: A is a nucleotide sequence consisting of a nucleotides; B is a nucleotide sequence consisting of a fragment of b nucleotides from (i) the nucleotide sequence shown in SEQ IDs 7-10, (ii) the nucleotide sequence shown in any of SEQ IDs 27-39, (iii) the complement of the nucleotide sequence shown in SEQ IDs 7-10, or (iv) the complement of the nucleotide sequence shown in any of SEQ IDs 27-39; C is a nucleotide sequence consisting of c nucleotides; and wherein said polynucleotide is not a fragment of (i) the nucleotide sequence shown in SEQ IDs 7-10, (ii) the nucleotide sequence shown in any of SEQ IDs 27-39, (iii) the complement of the nucleotide sequence shown in SEQ IDs 7-10, or (iv) the complement of the nucleotide sequence shown in any of SEQ IDs 27-39.

[0201] In this polynucleotide, a+c is at least 1 (*e.g.* at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.)* and *b* is at least 7 (*e.g.* at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.*). It is preferred that the value of a+b+c is at least 9 (*e.g.* at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.).* It is preferred that the value of *a+b+c* is at most 200 (*e.g.* at most 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9).

[0202] A and/or C may comprise a promoter sequence (or its complement).

H.1.4 - Homologous sequences

[0203] The invention provides a polynucleotide having at least *s*% identity to: (a) SEQ IDs 7-10; (b) a fragment of *x* nucleotides of SEQ IDs 7-10; (c) SEQ IDs 11-13; (b) a fragment ofx nucleotides of SEQ IDs 11-13. The value of s is at least 50 (*e.g.* at least 55, 60, 65, 70, 75, 80, 85, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.5, 99.9 *etc.).* The value ofx is at least 7 *(e.g.* 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.).*

[0204] These polynucleotides include naturally-occurring variants (*e.g.* degenerate variants, allelic variants, *etc.*), homologs, orthologs, and functional mutants.

[0205] Variants can be identified by hybridization of putative variants with the polynucleotide sequences disclosed in SEQ IDs 14-39 herein, preferably by hybridization under stringent conditions. For example, by using appropriate wash conditions, variants can be identified where the allelic variant exhibits at most about 25-30% base pair (bp) mismatches relative to the selected polynucleotide probe. In general, allelic variants contain 15-25% bp mismatches, and can contain as little as even 5-15%, or 2-5%, or 1-2% bp mismatches, as well as a single bp mismatch.

[0206] The invention also encompasses homologs corresponding to any one of the polynucleotide sequences provided herein, where the source of homologous genes can be any mammalian species (*e.g.* primate species, particularly human; rodents, such as rats, *etc.).* Between mammalian species (*e.g.* human and primate), homologs generally have substantial sequence similarity (*e.g.* at least 75% sequence identity, usually at least 90%, more usually at least 95%) between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a

larger sequence, such as a conserved motif, coding region, flanking region, domain, *etc.* A reference sequence will usually be at least about 18 contiguous nt long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art.

[0207] A preferred HERV-K(CH) isolate is an isolate sequence which is shown in SEQ IDs 7-10. Another preferred class of HERV-K(CH) isolates are those having a nucleotide sequence identity of at least 90%, preferably at least 95% to the 3' polymerase region shown in SEQ ID 13 which relates to integrase, as measured by the alignment program GCG Gap (Suite Version 10.1) using the default parameters: open gap = 3 and extend gap = 1. Another preferred class of HERV-K(CH) isolates are those having a nucleotide sequence identity of at least 98%, more preferably at least 99% to the 5' polymerase region shown in SEQ ID 12 which relates to reverse transcriptase, as measured by the alignment program GCG Gap (Suite Version 10.1) using the default parameters: open gap = 3 and extend gap =1. Another typical classification of the relationship of retroviruses is based on the amino acid sequence similarities in the reverse transcriptase protein. Thus, an even more preferred class of HERV-K(CH) isolates are those having an amino acid sequence identity of at least 90%, more preferably 95% to the 5' polymerase region encoded by the nucleotide sequence shown in SEQ ID 12, as determined by the Smith-Waterman homology search algorithm using an affine gap search with a gap open penalty of 12 and a gap extension penalty of 2, BLOSUM matrix of 62. Thus, these prostate cancer-associated polynucleotide sequences define a class of human endogenous retroviruses, designated herein as HERV-K(CH), whose members comprise variations which, without wanted to be bound by theory, may be due to the presence of polymorphisms or allelic variations.

H.1.5 - HERV-K(CH) hybridizable sequences

[0208] The invention provides an isolated polynucleotide comprising a polynucleotide that selectively hybridizes, relative to a known polynucleotide, to: (a) the nucleotide sequence shown in SEQ IDs 7-10; (b) the nucleotide sequence shown in any of SEQ IDs 27-39; (c) the complement of the nucleotide sequence shown in SEQ IDs 7-10; (d) the complement of the nucleotide sequence shown in any of SEQ IDs 27-39; (e) a fragment of the nucleotide sequence shown in SEQ IDs 7-10; (f) a fragment of the nucleotide sequence shown in any of SEQ IDs 27-39; (g) the complement of a fragment of the nucleotide sequence shown in SEQ IDs 7-10; (h) the complement of a fragment of the nucleotide sequence shown in any of SEQ IDs 27-39; (j) a nucleotide sequence shown in SEQ IDs 14-39; or (k) polynucleotides found in ATCC deposits having ATCC accession numbers given in Table 7. The fragment of (e), (f), (g) or (h) is preferably at least x nucleotides in length, wherein x is as defined in H.1.2 above, and is preferably not one of the sequences (i), (ii), (iii), (iv), (v) or (vi) as defined H.1.2 above.

[0209] Hybridization reactions can be performed under conditions of different "stringency", as described in B.4 above. In some embodiments, the polynucleotide hybridizes under low stringency conditions; in other embodiments it hybridizes under intermediate stringency conditions; in other embodiments, it hybridizes under high stringency conditions.

H.1.6 - Deposited HERV-K sequences

[0210] The invention also provides an isolated polynucleotide comprising: (a) a HERV-K(CH) cDNA insert as deposited at the ATCC and having an ATCC accession number given in Table 7; (b) a HERV-K(CH) sequence as shown in any one of SEQ IDs 14-26; (c) a HERV-K(CH) sequence as shown in any one of SEQ IDs 27-39; or (d) a fragment of (a), (b) or (c). The fragment of (d) is preferably at least x nucleotides in length, wherein x is at least 7 *(e.g.* at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.*).

H.1.7 - Preferred HERV-K(CH) sequences

[0211] Preferred polynucleotides of the invention are those having a sequence set forth in any one of the polynucleotide sequences SEQ IDs 7-10 and SEQ IDs 14-39 provided herein; polynucleotides obtained from the biological materials described herein, in particular, polynucleotide sequences present in the isolates deposited with the ATCC and having ATCC accession numbers given in Table 7 or other biological sources (particularly human sources) or by hybridization to the above mentioned sequences under stringent conditions (particularly conditions of high stringency); genes corresponding to the provided polynucleotides; variants of the provided polynucleotides and their corresponding genes particularly those variants that retain a biological activity of the encoded gene product (*e.g.* a biological activity ascribed to a gene product corresponding to the provided polynucleotides as a result of the assignment of the gene product to a protein family(ies) and/or identification of a functional domain present in the gene product). Other polynucleotides and polynucleotide compositions contemplated by and within the scope of the present invention will be readily apparent to one of ordinary skill in the art when provided with the disclosure here.

H.1.8 - General features of polynucleotides of the invention

[0212] General features of the polynucleotides described in this section H.1 are the same as those described in section B.4 above.

[0213] The isolated polynucleotides preferably comprise a polynucleotide having a HERVK(CH) sequence.

[0214] A polynucleotide of the invention can encode all or a part of a polypeptide, such as the gag region, 5' pol region or 3' pol region of a human endogenous retrovirus. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.*

[0215] Polynucleotides of the invention can be cDNAs or genomic DNAs, as well as fragments thereof, particularly fragments that encode a biologically active gene product and/or are useful in the methods disclosed herein (*e.g.* in diagnosis, as a unique identifier of a differentially expressed gene of interest, *etc.).* The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding a polypeptide. mRNA species can also exist with both exons and introns, where the introns may be removed by alternative splicing. Furthermore it should be noted that different species of mRNAs encoded by the same genomic sequence can exist at varying levels in a cell, and detection of these various levels of mRNA species can be indicative of differential expression of the encoded gene product in the cell.

[0216] A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, *etc.*, including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' and 3' end of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' and 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue, stage-specific, or disease-state specific expression.

[0217] Polynucleotides of the invention can be provided as linear molecules or within circular molecules, and can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. Expression of the polynucleotides can be regulated by their own or by other regulatory sequences known in the art. The polynucleotides can be introduced into suitable host cells using a variety of techniques available in the art, such as transferrin polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated DNA transfer, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, gene gun, calcium phosphate-mediated transfection, and the like.

[0218] A polynucleotide sequence that is "shown in" or "depicted in" a SEQ ID NO or Figure means that the sequence is present as an identical contiguous sequence in the SEQ ID NO or Figure. The term encompasses portions, or regions of the SEQ ID NO or Figure as well as the entire sequence contained within the SEQ ID NO or Figure.

H.2 - HERV-K(CH) polypeptides

H.2.1 - HERV-K(CH) open reading frames

[0219] The invention provides an isolated polypeptide: (a) encoded within a HERV-K(CH) open reading frame; (b) encoded by a polynucleotide shown in SEQ ID 11, 12 or 13; or (c) comprising an amino acid sequence as shown in any one of SEQ IDs 46-49, 50-55, 56-57 or 58.

[0220] Deduced polypeptides encoded by the HERV-K(CH) polynucleotides of the invention include the gag translations shown in SEQ IDS 46-49 and the 3' pol translations shown in SEQ IDs 50-55. A polypeptide sequence encoded by the polynucleotide having the sequence shown in SEQ ID 15 is provided in SEQ ID 56; a polypeptide sequence encoded by the polynucleotide having the sequence shown in SEQ ID 14, is shown in SEQ ID 57. A consensus 3' pol polypeptide sequence encoded by the polynucleotides having the sequence shown in SEQ IDs 21-27, inclusive, is provided in SEQ ID 58.

[0221] The polypeptides encompassed by the present invention include those encoded by polynucleotides of the invention, *e.g.* SEQ IDs 7-10 and SEQ IDs 14-39, as well as polynucleotides deposited with the ATCC as disclosed herein, as well as nucleic acids that, by virtue of the degeneracy of the genetic code, are not identical in sequence to the disclosed polynucleotides and encode the polypeptides. Thus, the invention includes within its scope a polypeptide encoded by a polynucleotide having the sequence of any one of the polynucleotide sequences provided herein, or a variant thereof.

[0222] While the over-expression of the polynucleotides associated with prostate tumor is observed, elevated levels

of expression of the polypeptides encoded by these polynucleotides may likely play a role in prostate tumors.

[0223] Typically, in retroviruses, a single large gag polypeptide is synthesized (*e.g.* a 73 kDa gag protein in HERV-K10) which is subsequently cleaved into multiple functional peptides by a functional protease encoded by the pol or protease region of the genome. Overexpression of sequences corresponding to both gag and pol domains of the HERV-K(CH) suggest such a mechanism. Sequences corresponding to the env and the nuclear RNA transport protein cORF region of the HERV-K(CH) genome may also be overexpressed. The polypeptides encoded by the open reading frames within the over-expressed polynucleotide sequences may play a significant role in the progression of prostate tumors.

[0224] The detection of these polypeptides by antibodies or other reagents that specifically recognize them may aid in the early diagnosis of prostate tumor or any other cancers associated with the overexpression of these HERV-K(CH) sequences.

[0225] Furthemore, inhibition of the function of these polypeptides may suggest means for therapy and treatment of prostatic or other HERV-K(CH) sequence related cancers. One method of accomplishing such inhibition is by administration of vaccines as a preventative therapy or antibody-mediated drug therapy as a post-neoplasia regimen for treatment of such cancers.

H.2.2 - HERV-K(CH) fragments

[0226] The invention provides an isolated polypeptide comprising a fragment of: (a) a polypeptide sequence encoded within a HERV-K(CH) open reading frame; (b) a polypeptide sequence encoded by a polynucleotide shown in SEQ ID 11, 12 or 13; or (c) an amino acid sequence as shown in any one of SEQ IDs 46-49, 50-55, 56-57 or 58.

[0227] The fragment is preferably at least x amino acids in length, wherein x is at least 5 *(e.g.* at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 75, 80, 90, 100, 125, 150, 200, 300, 400, 500 or more *etc.*). The value of x will typically not exceed 1000.

[0228] The fragment may include an epitope *e.g.* an epitope of the amino acid sequence shown in SEQ IDs 56, 57 or 58.

[0229] SEQ IDs 46-49 provide a translation of the HERV-K(CH) polynucleotides having a sequence shown in SEQ IDs 14, 15, 16 and 40 (the sequence of SEQ ID 40 is from a polynucleotide found in a normal prostate library) corresponding to polynucleotides encoding the gag region. SEQ IDs 50-55 provide a translation of the HERV-K(CH) polynucleotides having a sequence shown in SEQ IDs 21-26, inclusive, corresponding to the 3' region of pol. SEQ IDs 56 & 57 provide translations of the HERV-K(CH) polynucleotide of SEQ ID 15 and SEQ ID 14, respectively. SEQ ID 58 provides a consensus translation of the polynucleotide from the 3' pol region (SEQ IDs 21-26, inclusive). Encompassed with the present invention are polypeptide fragments, such as, epitopes, of at least 5 amino acids, at least 6 amino acids, at least 8 amino acids, at least 10 amino acids, at least 11 amino acids, at least 12 amino acids, at least 13 amino acids, at least 14 amino acids and at least 15 amino acids of the translations shown in SEQ IDs 46-49 and 50-55. In a preferred embodiment, the HERV-K(CH) epitopes of the amino acid sequence as shown in SEQ IDs 56-58 were determined by the Jameson-Wolf antigenic index [21].

[0230] The following regions in 3' pol (SEQ ID 58) were determined to be antigenic by Jameson-Wolf algorithm: amino acids: 1-10; 15-35; 45-55; 60-85; 100-115; 125-140; 170-190; 195-215; 230-268. Additional epitope-containing fragments include amino acids 1-8; 2-10; 1-15; 5-15; 7-15; 10-20; 12-20; 15-23; 20-28; 28-35; 15-30; 15-40; 20-30; 45-52; 48-55; 60-68; 60-70; 65-73; 70-78; 75-83; 70-80; 65-75; 68-75; 75-85; 78-85; 65-85; 60-75; 100-108; 103-110; 105-113; 108-115; 125-133; 128-135; 132-140; 170-178; 175-182; 180-187; 182-190; 195-202; 200-208; 205-212; 208-215; 230-237; 235-242; 240-247; 245-252; 250-257; 255-262; 260-268; 230-250; 235-255; 240-260; 245-268; 230-245; 235-245; 235-250; 240-255; 245-260; 250-268; 15-55; 170-215; 45-85.

[0231] The following regions in gag (SEQ ID 56) were determined to be antigenic by Jameson-Wolf algorithm: amino acids: 1-40; 45-60; 80-105; 130-145; 147-183; 186-220; 245-253; 255-288. Additional epitope-containing fragments include amino acids 1-8; 2-10; 1-15; 5-15; 7-15; 10-20; 12-20; 15-23; 20-28; 28-35; 30-37; 33-40; 1-20; 20-40; 1-15; 15-30; 15-40; 45-52; 50-57; 55-62; 50-60; 1-60; 80-87; 85-92; 80-90; 90-97; 95-102; 98-105; 85-100; 90-105; 80-100; 85-105; 130-137; 135-142; 140-147; 145-152; 150-157; 155-162; 160-167; 165-172; 170-177; 175-183; 180-187; 185-192; 190-197; 195-202; 200-207; 205-212; 210-217; 213-220; 185-220; 190-220; 195-220; 200-220; 205-220; 255-262; 260-267; 265-272; 270-277; 275-282; 280-288; 245-288; 250-288; 260-288; 265-288; 270-288.

[0232] The following regions in gag (SEQ ID 57) were determined to be antigenic by Jameson-Wolf algorithm: amino acids: 1-40; 80-105; 145-180; 185-225; 240-335. Additional epitope-containing fragments include amino acids 1-8; 2-10; 1'-15; 5-15; 7-15; 10-20; 12-20; 15-23; 20-28; 28-35; 30-37; 33-40; 1-20; 20-40; 1-15; 15-30; 15-40; 80-87; 85-92; 80-90; 90-97; 95-102; 98-105; 85-100; 90-105; 80-100; 85-105; 145-152; 150-157; 155-162; 160-167; 165-172; 170-177; 175-182; 180-187; 185-192; 190-197; 195-202; 200-207; 205-212; 210-217; 215-212; 218-225; 145-160; 150-165; 155-170; 160-175; 170-185; 180-225; 185-225; 190-225; 195-225; 200-225; 205-225; 210-225; 215-225; 240-247; 245-252; 250-257; 255-262; 260-267; 265-272; 270-277; 275-282; 280-287; 285-292; 290-297; 295-302; 300-307; 305-312; 310-317; 315-322; 320-327; 325-332; 328-335; 245-285; 250-285; 260-285; 265-285; 270-295; 275-300; 280-305; 285-310; 295-315; 300-320; 305-325; 325-335; 245-335; 250-335; 255-335; 260-335; 270-335; 275-335;

280-335; 285-335; 290-335; 295-335; 305-335; 310-335; 315-335; 320-335.

H.2.3 - HERV-K(CH) fragments plus heterologous sequences

[0233] The invention also provides an isolated polypeptide having formula 5'-A-B-C-3', wherein: A is an amino acid sequence consisting of a amino acids; B is an amino acid sequence consisting of a fragment of b amino acids from (i) the amino acid sequence encoded by a polynucleotide shown in SEQ ID 11, 12 or 13; (ii) any one of SEQ IDs 46-49, 50-55, 56-57 or 58; C is an amino acid sequence consisting of c amino acids; and wherein said polypeptide is not a fragment of the amino acid sequence defined in (i) or (ii).

[0234] In this polypeptide, a+c is at least 1 (*e.g.* at least 2, 3, 4, 5, 6, 7, 8, 9,10,11,12,13,14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.)* and b is at least 7 (*e.g.* at least 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.*). It is preferred that the value of a+b+c is at least 9 (*e.g.* at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 *etc.).* It is preferred that the value of *a+b+c* is at most 200 *(e.g.* at most 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 40, 30, 25, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9).

H.2.4 - Homologous sequences

[0235] The invention provides a polypeptide having at least *s*% identity to: (a) the polypeptide sequences encoded by SEQ IDs 7-45; (b) a fragment of *x* amino acids of the polypeptide sequences encoded by SEQ IDs 7-45; (c) the polypeptide sequences SEQ IDs 46-58; (d) a fragment of x amino acids of the polypeptide sequences SEQ IDs 46-58. The value of s is at least 35 (*e.g.* at least 40, 45, 50, 55, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.5, 99.9 *etc.).* The value ofx is at least 7 (*e.g.* 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100.

[0236] These polypeptides include naturally-occurring variants (*e.g.* allelic variants, *etc.*), homologs, orthologs, and functional mutants.

[0237] The invention thus encompasses variants of the naturally-occurring polypeptides, wherein such variants are homologous or substantially similar to the naturally occurring polypeptide, and can be of an origin of the same or different species as the naturally occurring polypeptide (*e.g.* human, murine, or some other species that naturally expresses the recited polypeptide, usually a mammalian species). These polypeptide variants are encoded by polynucleotides that are within the scope of the invention, and the genetic code can be used to select appropriate codons to construct the corresponding variants.

H.2.5 - Preferred HERV-K(CH) sequences

[0238] The invention provides polypeptides, such as those shown in SEQ IDs 46-58, encoded by HERV-K(CH) polynucleotides that are differentially expressed in prostate cancer cells. Such polypeptides are referred to herein as "polypeptides associated with prostate cancer" or "HERVK(CH) polypeptides". The polypeptides can be used to generate antibodies specific for a polypeptide associated with prostate cancer, which antibodies are in turn useful in diagnostic methods, prognostic methods, therametric methods, and the like as discussed in more detail herein. Polypeptides are also useful as targets for therapeutic intervention, as discussed in more detail herein.

[0239] Preferred polypeptides are encoded by polynucleotides of the invention.

H.2.6 - General features of polypeptides of the invention

[0240] General features of the polypeptides described in this section H.2 are the same as those described in section C.3 above.

[0241] The isolated polypeptides of the invention preferably comprise a polypeptide having a HERV-K(CH) sequence.

[0242] Polypeptides, such as polypeptides of the gag regions or polypeptides of the pol regions, encoded by the polynucleotides disclosed herein, such as polynucleotides having the sequences as shown in SEQ IDs 7-10 and SEQ IDs 14-39, and in isolates deposited with the ATCC and having ATCC accession numbers given in Table 7 and/or their corresponding full length genes, can be used to screen peptide libraries to identify binding partners, such as receptors, from among the encoded polypeptides. Peptide libraries can be synthesized according to methods known in the art (*e.g.* see refs. 151 & 152).

[0243] In general, the term "polypeptide" as used herein refers to both the full length polypeptide encoded by the recited polynucleotide, the polypeptide encoded by the gene represented by the recited polynucleotide, as well as portions or fragments thereof.

[0244] A polypeptide sequence that is "shown in" or "depicted in" a SEQ ID NO or Figure means that the sequence

is present as an identical contiguous sequence in the SEQ ID NO or Figure. The term encompasses portions, or regions of the SEQ ID NO or Figure as well as the entire sequence contained within the SEQ ID NO or Figure.

H.3 - Anti-HERV-K(CH) antibodies

[0245] The present invention also provides isolated antibodies or antigen binding fragments thereof, that bind to a polypeptide of the present invention. The present invention also provides isolated antibodies or antigen binding fragments thereof, that bind to a polypeptide encoded by a polynucleotide of the present invention. The present invention also provides isolated antibodies that bind to a polypeptide of the invention, or antigen binding fragment thereof, encoded by a polynucleotide made by the method comprising the following steps i) immunizing a host animal with a composition comprising said polypeptide of the present invention, or antigen binding fragment thereof, and ii) collecting cells from said host expressing antibodies against the antigen or antigen binding fragment thereof. The present invention also provides isolated antibodies that bind to a polypeptide, or antigen binding fragment thereof, encoded by a polynucleotide of the present invention made by the method comprising the following steps: providing a cell line producing an antibody, wherein said antibody binds to a polypeptide of the present invention, or antigen binding fragment thereof, encoded by a polynucleotide of the present invention and culturing said cell line under conditions wherein said antibodies are produced. In additional embodiments, the antibodies are collected and monoclonal antibodies are produced using the collected host cells or genetic material derived from the collected host cells. In additional embodiments, the antibody is a polyclonal antibody. In a further embodiment, the antibody is attached to a solid surface or further comprises a detectable label.

[0246] The present invention further provides antibodies, which may be isolated antibodies, that bind a polypeptide encoded by a polynucleotide described herein. Antibodies can be provided in a composition comprising the antibody and a buffer and/or a pharmaceutically acceptable excipient. Antibodies specific for a polypeptide associated with cancer are useful in a variety of diagnostic and therapeutic methods, as discussed in detail herein.

[0247] Expression products of a polynucleotide described herein, as well as the corresponding mRNA (particularly mRNAs having distinct secondary and/or tertiary structures), cDNA, or complete gene, or fragments of said expression products can be prepared and used for raising antibodies for experimental, diagnostic, and therapeutic purposes. For polynucleotides to which a corresponding gene has not been assigned, this provides an additional method of identifying the corresponding gene. The polynucleotide or related cDNA is expressed as described above, and antibodies are prepared. These antibodies are specific to an epitope on the polypeptide encoded by the polynucleotide, and can precipitate or bind to the corresponding native polypeptide in a cell or tissue preparation or in a cell-free extract of an *in vitro* expression system.

[0248] Polyclonal or monoclonal antibodies to the HERV-K(CH) polypeptides or an epitope thereof can be made for use in immunoassays by any of a number of methods known in the art. By epitope reference is made to an antigenic determinant of a polypeptide. The presence of an epitope is demonstrated by the ability of an antibody to bind a polypeptide with specificity. Two antibodies are considered to be directed to the same epitope if they cross block each others binding to the same polypeptide.

[0249] One approach for preparing antibodies to a polypeptide is the selection and preparation of an amino acid sequence of all or part of the polypeptide, chemically synthesizing the sequence and injecting it into an appropriate animal, typically a rabbit, hamster or a mouse.

[0250] Oligopeptides can be selected as candidates for the production of an antibody to the HERV-K(CH) polypeptide based upon the oligopeptides lying in hydrophilic regions, which are thus likely to be exposed in the mature polypeptide. Additional oligopeptides can be determined using, for example, the Antigenicity Index [30].

[0251] In other embodiments of the present invention, humanized monoclonal antibodies are provided, wherein the antibodies are specific for HERV-K(CH) polypeptides and do not appreciably bind other HERV polypeptides. The phrase "humanized antibody" refers to an antibody derived from a non-human antibody, typically a mouse monoclonal antibody. Alternatively, a humanized antibody may be derived from a chimeric antibody that retains or substantially retains the antigen-binding properties of the parental, non-human, antibody but which exhibits diminished immunogenicity in humans as compared to the parental antibody. The phrase "chimeric antibody," as used herein, refers to an antibody containing sequence derived from two different antibodies (see, *e.g.* ref. 153) which typically originate from different species. Most typically, chimeric antibodies comprise human and murine antibody fragments, generally human constant and mouse variable regions.

[0252] In the present invention, HERV-K(CH) polypeptides of the invention and variants thereof are used to immunize a transgenic animal as described above. Monoclonal antibodies are made using methods known in the art, and the specificity of the antibodies is tested using isolated HERV-K(CH) polypeptides.

[0253] Methods for preparation of the human or primate HERV-K(CH) or an epitope thereof include, but are not limited to chemical synthesis, recombinant DNA techniques or isolation from biological samples. Chemical synthesis of a peptide can be performed, for example, by the classical Merrifeld method of solid phase peptide synthesis [154] or the FMOC

strategy on a Rapid Automated Multiple Peptide Synthesis system (E. I. du Pont de Nemours Company, Wilmington, DE) [155].

[0254] Polyclonal antibodies can be prepared by immunizing rabbits or other animals by injecting antigen followed by subsequent boosts at appropriate intervals. The animals are bled and sera assayed against purified HERV-K(CH) usually by ELISA or by bioassay based upon the ability to block the action of HERV-K(CH). When using avian species, e.g. chicken, turkey and the like, the antibody can be isolated from the yolk of the egg. Monoclonal antibodies can be prepared after the method of Milstein and Kohler by fusing splenocytes from immunized mice with continuously replicating tumor cells such as myeloma or lymphoma cells. [156, 157, 158]. The hybridoma cells so formed are then cloned by limiting dilution methods and supemates assayed for antibody production by ELISA, RIA or bioassay.

[0255] The unique ability of antibodies to recognize and specifically bind to target polypeptides provides an approach for treating an overexpression of the polypeptide. Thus, another aspect of the present invention provides for a method for preventing or treating diseases involving overexpression of a HERV-K(CH) polypeptide by treatment of a patient with specific antibodies to the HERV-K(CH) polypeptide.

[0256] Specific antibodies, either polyclonal or monoclonal, to the HERV-K(CH) polypeptides can be produced by any suitable method known in the art as discussed above. For example, murine or human monoclonal antibodies can be produced by hybridoma technology or, alternatively, the HERV-K(CH) polypeptides, or an immunologically active fragment thereof, or an anti-idiotypic antibody, or fragment thereof can be administered to an animal to elicit the production of antibodies capable of recognizing and binding to the HERV-K(CH) polypeptides. Such antibodies can be from any class of antibodies including, but not limited to IgG, IgA, IgM, IgD, and IgE or in the case of avian species, IgY and from any subclass of antibodies.

H.4 - HERV-K(CH) vectors and host cells

[0257] The present invention also encompasses vectors and host cells comprising an isolated polynucleotide of the present invention.

H.5 - HERV-K(CH) kits, libraries and arrays

[0258] The invention provides kits, electronic libraries and arrays comprising polynucleotides of the invention, for use in diagnosing the presence of cancer in a test sample.

[0259] In general, a library of polynucleotides is a collection of sequence information, which information is provided in either biochemical form *(e.g.* as a collection of polynucleotide molecules), or in electronic form (*e.g.* as a collection of polynucleotide sequences stored in a computer-readable form, as in a computer system and/or as part of a computer program). The sequence information of the polynucleotides can be used in a variety of ways, *e.g.* as a resource for gene discovery, as a representation of sequences expressed in a selected cell type (*e.g.* cell type markers), and/or as markers of a given disease or disease state. In general, a disease marker is a representation of a gene product that is present in all cells affected by disease either at an increased or decreased level relative to a normal cell (*e.g.* a cell of the same or similar type that is not substantially affected by disease). For example, a polynucleotide sequence in a library can be a polynucleotide that represents an mRNA, polypeptide, or other gene product encoded by the polynucleotide, that is either over-expressed or under-expressed in a tissue affected by cancer, such as prostate cancer relative to a normal (*i.e.* substantially disease-free) tissue, such as normal prostate tissue.

[0260] The nucleotide sequence information of the library can be embodied in any suitable form, *e.g.* electronic or biochemical forms. For example, a library of sequence information embodied in electronic form comprises an accessible computer data file (or, in biochemical form, a collection of nucleic acid molecules) that contains the representative nucleotide sequences of genes that are differentially expressed (*e.g.* over-expressed or under-expressed) as between, for example, i) a cancerous cell and a normal cell; ii) a cancerous cell and a dysplastic cell; iii) a cancerous cell and a cell affected by a disease or condition other than cancer; iv) a metastatic cancerous cell and a normal cell and/or non-metastatic cancerous cell; v) a malignant cancerous cell and a non-malignant cancerous cell (or a normal cell) and/or vi) a dysplastic cell relative to a normal cell. Other combinations and comparisons of cells affected by various diseases or stages of disease will be readily apparent to the ordinarily skilled artisan. Biochemical embodiments of the library include a collection of nucleic acids that have the sequences of the genes in the library, where the nucleic acids can correspond to the entire gene in the library or to a fragment thereof, as described in greater detail below.

[0261] The polynucleotide libraries of the subject invention generally comprise sequence information of a plurality of polynucleotide sequences, where at least one of the polynucleotides has a sequence of any of sequence described herein. By plurality is meant at least 2, usually at least 3 and can include up to all of the sequences described herein. The length and number of polynucleotides in the library will vary with the nature of the library, *e.g.* if the library is an oligonucleotide array, a cDNA array, a computer database of the sequence information, etc.

[0262] Where the library is an electronic library, the nucleic acid sequence information can be present in a variety of

media. "Media" refers to a manufacture, other than an isolated nucleic acid molecule, that contains the sequence information of the present invention. Such a manufacture provides the genome sequence or a subset thereof in a form that can be examined by means not directly applicable to the sequence as it exists in a nucleic acid. For example, the nucleotide sequence of the present invention, *e.g.* the nucleic acid sequences of any of the polynucleotides of the sequences described herein, can be recorded on computer readable media, *e.g.* any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as a floppy disc, a hard disc storage medium, and a magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. One of skill in the art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present sequence information. "Recorded" refers to a process for storing information on computer readable medium, using any such methods as known in the art. Any convenient data storage structure can be chosen, based on the means used to access the stored information. A variety of data processor programs and formats can be used for storage, *e.g.* word processing text file, database format, *etc.* In addition to the sequence information, electronic versions of libraries comprising one or more sequence described herein can be provided in conjunction or connection with other computer-readable information and/or other types of computer-readable files (*e.g.* searchable files, executable files, *etc*, including, but not limited to, for example, search program software, *etc.*).

[0263] By providing the nucleotide sequence in computer readable form, the information can be accessed for a variety of purposes. Computer software to access sequence information is publicly available. For example, the gapped BLAST [159] and BLAZE [160] search algorithms on a Sybase system can be used to identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms.

[0264] As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention. The data storage means can comprise any manufacture comprising a recording of the present sequence information as described above, or a memory access means that can access such a manufacture.

[0265] "Search means" refers to one or more programs implemented on the computer-based system, to compare a target sequence or target structural motif, or expression levels of a polynucleotide in a sample, with the stored sequence information. Search means can be used to identify fragments or regions of the genome that match a particular target sequence or target motif. A variety of known algorithms are publicly known and commercially available, *e.g.* MacPattern (EMBL), BLASTN and BLASTX (NCBI). A "target sequence" can be any polynucleotide or amino acid sequence of six or more contiguous nucleotides or two or more amino acids, preferably from about 10 to 100 amino acids or from about 30 to 300 nt A variety of comparing means can be used to accomplish comparison of sequence information from a sample (e.g. to analyze target sequences, target motifs, or relative expression levels) with the data storage means. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer based systems of the present invention to accomplish comparison of target sequences and motifs. Computer programs to analyze expression levels in a sample and in controls are also known in the art.

[0266] A "target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration that is formed upon the folding of the target motif, or on consensus sequences of regulatory or active sites. There are a variety of target motifs known in the art. Protein target motifs include, but arc not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, hairpin structures, promoter sequences and other expression elements such as binding sites for transcription factors.

[0267] A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. One format for an output means ranks the relative expression levels of different polynucleotides. Such presentation provides a skilled artisan with a ranking of relative expression levels to determine a gene expression profile.

[0268] As discussed above, the "library" as used herein also encompasses biochemical libraries of the polynucleotides of the sequences described herein, *e.g.* collections of nucleic acids representing the provided polynucleotides. The biochemical libraries can take a variety of forms, *e.g.* a solution of cDNAs, a pattern of probe nucleic acids stably associated with a surface of a solid support (*i.e.* an array) and the like. Of particular interest are nucleic acid arrays in which one or more of the genes described herein is represented by a sequence on the array. By array is meant an article of manufacture that has at least a substrate with at least two distinct nucleic acid targets on one of its surfaces, where the number of distinct nucleic acids can be considerably higher, typically being at least 10 nt, usually at least 20 nt and often at least 25 nt. A variety of different array formats have been developed and are known to those of skill in the art. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis and the like, as disclosed in the above-listed exemplary patent documents.

[0269] In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the where the polypeptides of the library will represent at least a portion of the polypeptides encoded by a gene corresponding to a sequence described herein.

[0270] Polynucleotide arrays provide a high throughput technique that can assay a large number of polynucleotides or polypeptides in a sample. This technology can be used as a tool to test for differential expression. A variety of methods of producing arrays, as well as variations of these methods, are known in the art and contemplated for use in the invention. For example, arrays can be created by spotting polynucleotide probes onto a substrate (*e.g.* glass, nitrocellulose, *etc.*) in a two-dimensional matrix or array having bound probes. The probes can be bound to the substrate by either covalent bonds or by non-specific interactions, such as hydrophobic interactions. Samples of polynucleotides can be detectably labeled (*e.g.* using radioactive or fluorescent labels) and then hybridized to the probes. Double stranded polynucleotides, comprising the labeled sample polynucleotides bound to probe polynucleotides, can be detected once the unbound portion of the sample is washed away. Alternatively, the polynucleotides of the test sample can be immobilized on the array, and the probes detectably labeled. Techniques for constructing arrays and methods of using these arrays are described in, for example, references 161 to 177.

[0271] Arrays can be used to, for example, examine differential expression of genes and can be used to determine gene function. For example, arrays can be used to detect differential expression of a gene corresponding to a polynucleotide described herein, where expression is compared between a test cell and control cell (*e.g.* cancer cells and normal cells). For example, high expression of a particular message in a cancer cell, which is not observed in a corresponding normal cell, can indicate a cancer specific gene product. Exemplary uses of arrays are further described in, for example, references 178 and 179. Furthermore, many variations on methods of detection using arrays are well within the skill in the art and within the scope of the present invention. For example, rather than immobilizing the probe to a solid support, the test sample can be immobilized on a solid support which is then contacted with the probe.

[0272] A gene or polynucleotide that is differentially expressed in a cancer cell when the polynucleotide is detected at higher or lower levels in cancer compared with a cell of the same cell type that is not cancerous. Typically, screening for polynucleotides differentially expressed focuses on a polynucleotide that is expressed such that, for example, mRNA is found at levels at least about 25%, at least about 50% to about 75%, at least about 90%, preferably at least about 2-fold, more preferably at least about 5-fold, at least about 10-fold, or at least about 50-fold or more, higher (e.g. overexpressed) or lower (e.g. underexpressed) in a cancer cell when compared with a cell of the same cell type that is not cancerous. The comparison can be made between two tissues, for example, if one is using *in situ* hybridization or another assay method that allows some degree of discrimination among cell types in the tissue. The comparison may also be made between cells removed from their tissue source. Thus, a polypeptide encoded by a polynucleotide that is differentially expressed in a cancer cell would be of clinical significance with respect to cancer.

[0273] In one preferred embodiment of the present invention, an array comprises at least two polynucleotides, each having a sequence selected from the group consisting of SEQ IDs 14-39 and polynucleotides present in isolates deposited with the ATCC and having ATCC accession numbers PTA-2561, PTA-2572, PTA-2566, PTA-2571, PTA-2562, PTA-2573, PTA-2560, PTA-2565, PTA-2568, PTA-2564, PTA-2569, PTA-2567, PTA-2559, PTA-2563, PTA-2570. In another preferred embodiment, an array comprises at least one polynucleotide having a sequence selected from the group consisting of SEQ IDs 14-39 and polynucleotides present in isolates deposited with the ATCC and having ATCC accession numbers PTA-2561, PTA-2572, PTA-2566, PTA-2571, PTA-2562, PTA-2573, PTA-2560, PTA-2565, PTA-2568, PTA-2564, PTA-2569, PTA-2567, PTA-2559, PTA-2563, PTA-2570 and at least one of a polynucleotide having a sequence shown in SEQ ID 42 or 43.

[0274] The polynucleotides described herein, as well as their gene products, are of particular interest as genetic or biochemical markers *(e.g.* in blood or tissues) that will detect the earliest changes along the carcinogenesis pathway and/or to monitor the efficacy of various therapies and preventive interventions. For example, the level of expression of certain polynucleotides can be indicative of a poorer prognosis, and therefore warrant more aggressive chemo- or radiotherapy for a patient or vice versa. The correlation of novel surrogate tumor specific features with response to treatment and outcome in patients can define prognostic indicators that allow the design of tailored therapy based on the molecular profile of the tumor. These therapies include antibody targeting, antagonists (*e.g.* small molecules), and gene therapy. Determining expression of certain polynucleotides and comparison of a patients profile with known expression in normal tissue and variants of the disease allows a determination of the best possible treatment for a patient, both in terms of specificity of treatment and in terms of comfort level of the patient. Polynucleotide expression can also be used to better classify, and thus diagnose and treat, different forms and disease states of cancer. Two classifications widely used in oncology that can benefit from identification of the expression levels of the genes corresponding to the polynucleotides described herein are staging of the cancerous disorder, and grading the nature of the cancerous tissue.

[0275] The polynucleotides that correspond to differentially expressed genes, as well as their encoded gene products, can be useful to monitor patients having or susceptible to cancer to detect potentially malignant events at a molecular level before they are detectable at a gross morphological level. In addition, the polynucleotides described herein, as well as the genes corresponding to such polynucleotides, can be useful as therametrics, *e.g.* to assess the effectiveness of

therapy by using the polynucleotides or their encoded gene products, to assess, for example, tumor burden in the patient before, during, and after therapy.

[0276] Furthermore, a polynucleotide identified as corresponding to a gene that is differentially expressed in, and thus is important for, one type of cancer can also have implications for development or risk of development of other types of cancer, e.g. where a polynucleotide represents a gene differentially expressed across various cancer types.

[0277] In another embodiment, the diagnostic and/or prognostic methods of the invention involve detection of expression of a selected set of genes in a test sample to produce a test expression pattern (TEP). The TEP is compared to a reference expression pattern (REP), which is generated by detection of expression of the selected set of genes in a reference sample (*e.g.* a positive or negative control sample). The selected set of genes includes at least one of the genes of the invention, which genes correspond to the polynucleotide sequences described herein. Of particular interest is a selected set of genes that includes gene differentially expressed in the disease for which the test sample is to be screened.

[0278] "Reference sequences" or "reference polynucleotides" as used herein in the context of differential gene expression analysis and diagnosis/prognosis refers to a selected set of polynucleotides, which selected set includes at least one or more of the differentially expressed polynucleotides described herein. A plurality of reference sequences, preferably comprising positive and negative control sequences, can be included as reference sequences. Additional suitable reference sequences are found in GenBank, Unigene, and other nucleotide sequence databases (including, *e.g.* expressed sequence tag (EST), partial, and full-length sequences).

[0279] "Reference array" means an array having reference sequences for use in hybridization with a sample, where the reference sequences include all, at least one of, or any subset of the differentially expressed polynucleotides described herein. Usually such an array will include at least 2 different reference sequences, and can include any one or all of the provided differentially expressed sequences. Arrays of interest can further comprise sequences, including polymorphisms, of other genetic sequences, particularly other sequences of interest for screening for a disease or disorder (*e.g.* cancer, dysplasia, or other related or unrelated diseases, disorders, or conditions). The oligonucleotide sequence on the array will usually be at least about 12 nt in length, and can be of about the length of the provided sequences, or can extend into the flanking regions to generate fragments of 100 nt to 200 nt in length or more. Reference arrays can be produced according to any suitable methods known in the art. For example, methods of producing large arrays of oligonucleotides are described in references 180 & 181 using light-directed synthesis techniques. Using a computer controlled system, a heterogeneous array of monomers is converted, through simultaneous coupling at a number of reaction sites, into a heterogeneous array of polymers. Alternatively, microarrays are generated by deposition of pre-synthesized oligonucleotides onto a solid substrate, for example as described in reference 182.

[0280] A "reference expression pattern"' or "REP" as used herein refers to the relative levels of expression of a selected set of genes, particularly of differentially expressed genes, that is associated with a selected cell type, *e.g.* a normal cell, a cancerous cell, a cell exposed to an environmental stimulus, and the like. A "test expression pattern" or "TEP" refers to relative levels of expression of a selected set of genes, particularly of differentially expressed genes, in a test sample (*e.g.* a cell of unknown or suspected disease state, from which mRNA is isolated).

[0281] REPs can be generated in a variety of ways according to methods well known in the art. For example, REPs can be generated by hybridizing a control sample to an array having a selected set of polynucleotides (particularly a selected set of differentially expressed polynucleotides), acquiring the hybridization data from the array, and storing the data in a format that allows for ready comparison of the REP with a TEP. Alternatively, all expressed sequences in a control sample can be isolated and sequenced, *e.g.* by isolating mRNA from a control sample, converting the mRNA into cDNA, and sequencing the cDNA. The resulting sequence information roughly or precisely reflects the identity and relative number of expressed sequences in the sample. The sequence information can then be stored in a format (*e.g.* a computer-readable format) that allows for ready comparison of the REP with a TEP. The REP can be normalized prior to or after data storage, and/or can be processed to selectively remove sequences of expressed genes that are of less interest or that might complicate analysis (*e.g.* some or all of the sequences associated with housekeeping genes can be eliminated from REP data).

[0282] TEPs can be generated in a manner similar to REPs, *e.g.* by hybridizing a test sample to an array having a selected set of polynucleotides, particularly a selected set of differentially expressed polynucleotides, acquiring the hybridization data from the array, and storing the data in a format that allows for ready comparison of the TEP with a REP. The REP and TEP to be used in a comparison can be generated simultaneously, or the TEP can be compared to previously generated and stored REPs.

[0283] In one embodiment of the invention, comparison of a TEP with a REP involves hybridizing a test sample with an array, where the reference array has one or more reference sequences for use in hybridization with a sample. The reference sequences include all, at least one of, or any subset of the differentially expressed polynucleotides described herein. Hybridization data for the test sample is acquired, the data normalized, and the produced TEP compared with a REP generated using an array having the same or similar selected set of differentially expressed polynucleotides. Probes that correspond to sequences differentially expressed between the two samples will show decreased or increased

hybridization efficiency for one of the samples relative to the other.

[0284] Methods for collection of data from hybridization of samples with a reference arrays are well known in the art. For example, the polynucleotides of the reference and test samples can be generated using a detectable fluorescent label, and hybridization of the polynucleotides in the samples detected by scanning the microarrays for the presence of the detectable label using, for example, a microscope and light source for directing light at a substrate. A photon counter detects fluorescence from the substrate, while an x-y translation stage varies the location of the substrate. A confocal detection device that can be used in the subject methods is described in reference 183. A scanning laser microscope is described in reference 163. A scan, using the appropriate excitation line, is performed for each fluorophore used. The digital images generated from the scan are then combined for subsequent analysis. For any particular array element, the ratio of the fluorescent signal from one sample (*e.g.* a test sample) is compared to the fluorescent signal from another sample (*e.g.* a reference sample), and the relative signal intensity determined.

[0285] Methods for analyzing the data collected from hybridization to arrays are well known in the art. For example, where detection of hybridization involves a fluorescent label, data analysis can include the steps of determining fluorescent intensity as a function of substrate position from the data collected, removing outliers, *i.e.* data deviating from a predetermined statistical distribution, and calculating the relative binding affinity of the targets from the remaining data. The resulting data can be displayed as an image with the intensity in each region varying according to the binding affinity between targets and probes.

[0286] In general, the test sample is classified as having a gene expression profile corresponding to that associated with a disease or non-disease state by comparing the TEP generated from the test sample to one or more REPs generated from reference samples (*e.g.* from samples associated with cancer or specific stages of cancer, dysplasia, samples affected by a disease other than cancer, normal samples, *etc.).* The criteria for a match or a substantial match between a TEP and a REP include expression of the same or substantially the same set of reference genes, as well as expression of these reference genes at substantially the same levels (*e.g.* no significant difference between the samples for a signal associated with a selected reference sequence after normalization of the samples, or at least no greater than about 25% to about 40% difference in signal strength for a given reference sequence. In general, a pattern match between a TEP and a REP includes a match in expression, preferably a match in qualitative or quantitative expression level, of at least one of, all or any subset of the differentially expressed genes of the invention.

[0287] Pattern matching can be performed manually, or can be performed using a computer program. Methods for preparation of substrate matrices (*e.g.* arrays), design of oligonucleotides for use with such matrices, labeling of probes, hybridization conditions, scanning of hybridized matrices, and analysis of patterns generated, including comparison analysis, are described *e.g.* in reference 184.

H.6 - HERV-K(CH)-based diagnostic methods

[0288] The invention provides methods for diagnosing the presence of cancer in a test sample associated with expression of a polynucleotide in a test cell sample, comprising the steps of: i) detecting a level of expression of at least one polynucleotide of the invention, or a fragment thereof, or at least one polynucleotide found in an isolate selected from the group consisting of ATCC accession numbers given in Table 7, or a fragment thereof; and ii) comparing said level of expression of the polynucleotide in the test sample with a level of expression of polynucleotide in the control cell sample, wherein differential expression of the polynucleotide in the test cell sample relative to the level of polynucleotide expression in the control cell sample is indicative of the presence of cancer in the test cell sample.

[0289] In some embodiments of the present invention, the cancer is prostate cancer. In other embodiments of the present invention, the cancer is testicular cancer.

[0290] In yet other embodiments of the present invention, the detecting is measuring the level of an RNA transcript; measuring the level of a polynucleotide; or measuring by a method including PCR, TMA, bDNA, NAT or Nasba. In further embodiments, the polynucleotide is attached to a solid support.

[0291] The present invention also provides compositions comprising a test cell sample and an isolated polynucleotide of the present invention. The present invention further provides methods for detecting cancer associated with expression of a polypeptide in a test cell sample, comprising the steps of: i) detecting a level of expression of at least one polypeptide of the invention, or a fragment thereof and ii) comparing said level of expression of the polypeptide in the test sample with a level of expression of polypeptide in the control cell sample, wherein an altered level of expression of the polypeptide in the test cell sample relative to the level of expression of the polypeptide in the control cell sample is indicative of the presence of cancer in the test cell sample. The present invention also provides methods for detecting cancer associated with the presence of an antibody in a test cell sample, comprising the steps of: i) detecting a level of an antibody of the present invention, and ii) comparing said level of said antibody in the test sample with a level of said antibody in the control cell sample, wherein an altered level of antibody in said test cell sample relative to the level of antibody in the control cell sample is indicative of the presence of cancer in the test cell sample. In some embodiments, the cancer is prostate cancer and in other embodiments, the cancer is testicular cancer.

[0292] This invention also provides methods for detecting cancer associated with elevated levels of HERV-K(CH) polynucleotides, in particular in prostate cancer, by means of (i) detecting polynucleotides having at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or at least 100% identity to the polynucleotide shown in SEQ IDs 7-10 or to polynucleotides in isolates deposited with the ATCC and having ATCC deposit accession numbers PTA-2561, PTA-2572, PTA-2566, PTA-2571, PTA-2562, PTA-2573, PTA-2560, PTA-2565, PTA-2568, PTA-2564, PTA-2569, PTA-2567, PTA-2559, PTA-2563, PTA-2570, as measured by the alignment program GCG Gap (Suite Version 10.1) using the default parameters: open gap = 3 and extend gap = 1 or polynucleotides hybridizing under high stringency conditions to the polynucleotide shown in SEQ IDs 7-10; (ii) detecting polypeptides, or fragments thereof encoded by the sequences of (i); and (iii) detecting antibodies specific for one or more of the polypeptides. Furthermore, (iv) detecting particles associated with overexpression of HERV-K(CH) polynucleotides may also be used in the diagnosis of cancer, in particular, prostate cancer, and monitoring its progression.

[0293] The treatment regimen of a prostate or other cancer associated with elevated levels of HERV-K(CH) polynucleotides may also monitored by detecting levels of the polynucleotides and polypeptides in order to assess the staging of the cancer and/or efficacy of particular cancer therapies.

[0294] The present invention provides methods of using the polynucleotides described herein for detecting cancer cells, in particular prostate cancer cells, facilitating diagnosis of cancer and the severity of a cancer (*e.g.* tumor grade, tumor burden, and the like) in a subject, facilitating a determination of the prognosis of a subject, and assessing the responsiveness of the subject to therapy (*e.g.* by providing a measure of therapeutic effect through, for example, assessing tumor burden during or following a chemotherapeutic regimen). Detection can be based on detection of a polynucleotide that is differentially expressed in a cancer cell, and/or detection of a polypeptide encoded by a polynucleotide that is differentially expressed in a cancer cell. The detection methods of the invention can be conducted *in vitro* or *in vivo,* on isolated cells, or in whole tissues or a bodily fluid *e.g.* blood, plasma, serum, urine, and the like).

[0295] The detection methods can be provided as part of a kit. Thus, the invention further provides kits for detecting the presence and/or a level of a polynucleotide that is differentially expressed in a cancer cell (*e.g.* by detection of an mRNA encoded by the differentially expressed gene of interest), and/or a polypeptide encoded thereby, in a biological sample. Procedures using these kits can be performed by clinical laboratories, experimental laboratories, medical practitioners, or private individuals. The kits of the invention for detecting a polypeptide encoded by a polynucleotide that is differentially expressed in a cancer cell may comprise a moiety that specifically binds the polypeptide, which may be an antibody that binds the polypeptide or fragment thereof. The kits of the invention used for detecting a polynucleotide that is differentially expressed in a prostate cancer cell may comprise a moiety that specifically hybridizes to such a polynucleotide. The kit may optionally provide additional components that are useful in the procedure, including, but not limited to, buffers, developing reagents, labels, reacting surfaces, means for detection, control samples, standards, instructions, and interpretive information.

[0296] Accordingly, the present invention provides kits for detecting prostate cancer comprising at least one of polynucleotides having the sequence as shown in SEQ IDs 7-10, SEQ IDs 14-39, or fragments thereof, or having the sequence found in an isolate deposited with the ATCC and having ATCC accession numbers PTA-2561, PTA-2572, PTA-2566, PTA-2571, PTA-2562, PTA-2573, PTA-2560, PTA-2565, PTA-2568, PTA-2564, PTA-2569, PTA-2567, PTA-2559, PTA-2563, PTA-2570 or fragments thereof

[0297] In some embodiments, methods are provided for detecting a polypeptide encoded by a gene differentially expressed in a prostate cancer cell. Any of a variety of known methods can be used for detection, including, but not limited to, immunoassay, using antibody that binds the polypeptide, e.g. by enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and the like; and functional assays for the encoded polypeptide, e.g. binding activity or enzymatic activity.

[0298] As will be readily apparent to the ordinarily skilled artisan upon reading the present specification, the detection methods and other methods described herein can be readily varied. Such variations are within the intended scope of the invention. For example, in the above detection scheme, the probe for use in detection can be immobilized on a solid support, and the test sample contacted with the immobilized probe. Binding of the test sample to the probe can then be detected in a variety of ways, *e.g.* by detecting a detectable label bound to the test sample to facilitate detected of test sample-immobilized probe complexes.

[0299] The present invention further provides methods for detecting the presence of and/or measuring a level of a polypeptide in a biological sample, which polypeptide is encoded by a polynucleotide that is differentially expressed in a prostate cancer cell, using an antibody specific for the encoded polypeptide. The methods generally comprise: a) contacting the sample with an antibody specific for a polypeptide encoded by a polynucleotide that is differentially expressed in a prostate cancer cell; and b) detecting binding between the antibody and molecules of the sample.

[0300] Detection of specific binding of the antibody specific for the encoded prostate cancer-associated polypeptide, when compared to a suitable control is an indication that encoded polypeptide is present in the sample. Suitable controls include a sample known not to contain the encoded polypeptide or known not to contain elevated levels of the polypeptide;

such as normal prostate tissue, and a sample contacted with an antibody not specific for the encoded polypeptide, e.g. an anti-idiotype antibody. A variety of methods to detect specific antibody-antigen interactions are known in the art and can be used in the method, including, but not limited to, standard immunohistological methods, immunoprecipitation, an enzyme immunoassay, and a radioimmunoassay. In general, the specific antibody will be detectably labeled, either directly or indirectly. Direct labels include radioisotopes; enzymes whose products are detectable (e.g. luciferase, ß-galactosidase, and the like); fluorescent labels (e.g. fluorescein isothiocyanate, rhodamine, phycoerythrin, and the like); fluorescence emitting metals, e.g. ^{152}Eu, or others of the lanthanide series, attached to the antibody through metal chelating groups such as EDTA; chemiluminescent compounds, e.g. luminol, isoluminol, acridinium salts, and the like; bioluminescent compounds, e.g. luciferin, aequorin (green fluorescent protein), and the like. The antibody may be attached (coupled) to an insoluble support, such as a polystyrene plate or a bead. Indirect labels include second antibodies specific for antibodies specific for the encoded polypeptide ("first specific antibody"), wherein the second antibody is labeled as described above; and members of specific binding pairs, e.g. biotin-avidin, and the like. The biological sample may be brought into contact with and immobilized on a solid support or carrier, such as nitrocellulose, that is capable of immobilizing cells, cell particles, or soluble proteins. The support may then be washed with suitable buffers, followed by contacting with a detectably-labeled first specific antibody. Detection methods are known in the art and will be chosen as appropriate to the signal emitted by the detectable label. Detection is generally accomplished in comparison to suitable controls, and to appropriate standards.

[0301] In some embodiments, the methods are adapted for use *in vivo,* e.g. to locate or identify sites where cancer cells, such as prostate cancer cells, are present.

[0302] In some embodiments, methods are provided for detecting a cancer cell by detecting expression in the cell of a transcript that is differentially expressed in a cancer cell. Any of a variety of known methods can be used for detection, including, but not limited to, detection of a transcript by hybridization with a polynucleotide that hybridizes to a polynucleotide that is differentially expressed in a prostate cancer cell; detection of a transcript by a polymerase chain reaction using specific oligonucleotide primers; *in situ* hybridization of a cell using as a probe a polynucleotide that hybridizes to a gene that is differentially expressed in a prostate cancer cell. The methods can be used to detect and/or measure mRNA levels of a gene that is differentially expressed in a prostate cancer cell. In some embodiments, the methods comprise: a) contacting a sample with a polynucleotide that corresponds to a differentially expressed gene described herein under conditions that allow hybridization; and b) detecting hybridization, if any.

[0303] Detection of differential hybridization, when compared to a suitable control, is an indication of the presence in the sample of a polynucleotide that is differentially expressed in a cancer cell. Appropriate controls include, for example, a sample which is known not to contain a polynucleotide that is differentially expressed in a cancer cell, and use of a labeled polynucleotide of the same "sense" as the polynucleotide that is differentially expressed in the cancer cell. In a preferred embodiment, the cancer cell is a prostate cancer cell. Conditions that allow hybridization are known in the art, and have been described in more detail above. Detection can also be accomplished by any known method, including, but not limited to, *in situ* hybridization, PCR (polymerase chain reaction), RT-PCR (reverse transcription-PCR), TMA, bDNA, and Nasba and "Northern" or RNA blotting, or combinations of such techniques, using a suitably labeled polynucleotide. A variety of labels and labeling methods for polynucleotides are know in the art and can be used in the assay methods of the invention. Specific hybridization can be determined by comparison to appropriate controls.

[0304] Polynucleotide generally comprising at least 10 nt, at least 12nt or at least 15 contiguous nucleotides of a polynucleotide provided herein, such as, for example, those having the sequence as depicted in SEQ IDs 7-10, and 3-28, are used for a variety of purposes, such as probes for detection of and/or measurement of, transcription levels of a polynucleotide that is differentially expressed in a prostate cancer cell. A probe that hybridizes specifically to a polynucleotide disclosed herein should provide a detection signal at least 5-, 10-, or 20-fold higher than the background hybridization provided with other unrelated sequences. It should be noted that "probe" as used herein is meant to refer to a polynucleotide sequence used to detect a differentially expressed gene product in a test sample. As will be readily appreciated by the ordinarily skilled artisan, the probe can be detectably labeled and contacted with, for example, an array comprising immobilized polynucleotides obtained from a test sample (*e.g.* mRNA). Alternatively, the probe can be immobilized on an array and the test sample detectably labeled. These and other variations of the methods of the invention are well within the skill in the art and are within the scope of the invention.

[0305] Nucleotide probes are used to detect expression of a gene corresponding to the provided polynucleotide. In Northern blots, mRNA is separated electrophoretically and contacted with a probe. A probe is detected as hybridizing to an mRNA species of a particular size. The amount of hybridization can be quantitated to determine relative amounts of expression, for example under a particular condition. Probes are used for in situ hybridization to cells to detect expression. Probes can also be used *in vivo* for diagnostic detection of hybridizing sequences. Probes are typically labeled with a radioactive isotope. Other types of detectable labels can be used such as chromophores, fluorophores, and enzymes. Other examples of nucleotide hybridization assays are described in refs. 185 and 186.

[0306] PCR is another means for detecting small amounts of target nucleic acids (see, *e.g.* refs. 187, 188 & 189). Two primer polynucleotides nucleotides that hybridize with the target nucleic acids are used to prime the reaction. The primers

can be composed of sequence within or 3' and 5' to the HERV-K(CH) polynucleotides disclosed herein. Alternatively, if the primers are 3' and 5' to these polynucleotides, they need not hybridize to them or the complements. After amplification of the target with a thermostable polymerase, the amplified target nucleic acids can be detected by methods known in the art (*e.g.* Southern blot). mRNA or cDNA can also be detected by traditional blotting techniques (*e.g.* Southern blot, Northern blot, etc.) described in ref. 8 (*e.g.* without PCR amplification). In general, mRNA or cDNA generated from mRNA using a polymerase enzyme can be purified and separated using gel electrophoresis, and transferred to a solid support, such as nitrocellulose. The solid support is exposed to a labeled probe, washed to remove any unhybridized probe, and duplexes containing the labeled probe are detected.

[0307] Methods using PCR amplification can be performed on the DNA from a single cell, although it is convenient to use at least about 10^5 cells. The use of the polymerase chain reaction is described in ref. 190, and a review of techniques may be found in pages 14.2 to 14.33 of reference 8. A detectable label may be included in the amplification reaction. Suitable detectable labels include fluorochromes,(*e.g.* fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 6-carboxy-X-rhodamine (ROX), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein, 5-carboxyfluorescein (5-FAM), N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), or 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX)), radioactive labels, (*e.g.* ^{32}P, ^{35}S, ^{3}H, *etc.),* and the like. The label may be a two stage system, where the polynucleotides is conjugated to biotin, haptens, *etc.* having a high affinity binding partner, *e.g.* avidin, specific antibodies, *etc.*, where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

[0308] The present invention further relates to methods of detecting/diagnosing a neoplastic or preneoplastic condition in a mammal (for example, a human).

[0309] Examples of conditions that can be detected/diagnosed in accordance with these methods include, but are not limited to prostate cancers. Polynucleotides corresponding to genes that exhibit the appropriate expression pattern can be used to detect prostate cancer in a subject. Reference 191 reviews markers of cancer.

[0310] One detection/diagnostic method comprises: (a) obtaining from a mammal (eg a human) a biological sample, (b) detecting the presence in the sample of a HERV-K(CH) polypeptide and (c) comparing the amount of product present with that in a control sample. In accordance with this method, the presence in the sample of elevated levels of a HERV-K(CH) gene product indicates that the subject has a neoplastic or preneoplastic condition.

[0311] The compound is preferably a binding protein, *e.g.* an antibody, polyclonal or monoclonal, or antigen binding fragment thereof, which can be labeled with a detectable marker (eg fluorophore, chromophore or isotope, etc). Where appropriate, the compound can be attached to a solid support. Determination of formation of the complex can be effected by contacting the complex with a further compound (eg an antibody) that specifically binds to the first compound (or complex). Like the first compound, the further compound can be attached to a solid support and/or can be labeled with a detectable marker.

[0312] The identification of elevated levels of HERV-K(CH) polypeptide in accordance with the present invention makes possible the identification of subjects (patients) that are likely to benefit from adjuvant therapy. For example, a biological sample from a post-primary therapy subject (*e.g.* subject having undergone surgery) can be screened for the presence of circulating HERVK(CH) polypeptide, the presence of elevated levels of the polypeptide, determined by studies of normal populations, being indicative of residual tumor tissue. Similarly, tissue from the cut site of a surgically removed tumor can be examined (*e.g.* by immunofluorescence), the presence of elevated levels of product (relative to the surrounding tissue) being indicative of incomplete removal of the tumor. The ability to identify such subjects makes it possible to tailor therapy to the needs of the particular subject. Subjects undergoing non-surgical therapy (*e.g.* chemotherapy or radiation therapy) can also be monitored, the presence in samples from such subjects of elevated levels of HERV-K (CH) polypeptide being indicative of the need for continued treatment. Staging of the disease (for example, for purposes of optimizing treatment regimens) can also be effected, for example, by prostate biopsy *e.g.* with antibody specific for a HERVK(CH) polypeptide.

[0313] The present invention also relates to a kit that can be used in the detection of a HERVK(CH) polypeptide. The kit can comprise a compound that specifically binds a HERV-K(CH) polypeptide, such as, for example, binding proteins including antibodies or binding fragments thereof (e.g. $F(ab')_2$ fragments) disposed within a container means. The kit can further comprise ancillary reagents, for processing the binding assay.

DEFINITIONS

[0314] The term "comprising" means "including" as well as "consisting" *e.g.* a composition "comprising" X may consist exclusively of X or may include something additional *e.g.* X + Y.

[0315] The term "about" in relation to a numerical value x means, for example, $x \pm 10\%$.

[0316] The terms "neoplastic cells", "neoplasia", "tumor", "tumor cells", "cancer" and "cancer cells", (used interchangeably) refer to cells which exhibit relatively autonomous growth, so that they exhibit an aberrant growth phenotype char-

acterized by a significant loss of control of cell proliferation (i.e. de-regulated cell division). Neoplastic cells can be malignant or benign and include prostate cancer derived tissue.

BRIEF DESCRIPTION OF DRAWINGS

[0317]

Figure 1 is a schematic representation of a human endogenous retrovirus with a depiction of the HERV-K(CH) polynucleotides and their position relative to the retrovirus.

Figure 2 is a schematic representation of open reading frames within the HERV-K(HML-2.HOM) (also known as 'ERVK6') genome [1].

Figure 3 shows splicing events described in the prior art [16] for HERV-K mRNAs.

Figure 4 shows splice sites identified near the 5' and 3' ends of the env ORF. The three reading frames are shaded differently.

Figure 5 shows northern blot analysis of PCAV transcripts in cancer cell lines. The top arrow on the left shows the position of the genomic mRNA transcript. The next arrow shows the position of the env transcript. The bottom two arrows show the positions of other ORFs. The lanes contain RNA from the following cell lines: (1) Tera 1; (2) DU145; (3) PC3; (4) MDA Pca-2b; (5) LNCaP. Tera 1 is a teratocarcinoma cell line; the others are prostatic carcinoma cell lines.

Figure 6 shows an alignment of env genomic DNA sequences from 27 HERV-K viruses. A consensus sequence (SEQ ID 157) is shown on the bottom line.

Figures 7-9 show alignments of inferred polypeptide sequences for gag (7), pol (8) and env (9) from various HERV-K viruses, together with consensus sequences (SEQ IDs 158-160).

MODES FOR CARRYING OUT THE INVENTION

[0318] Certain aspects of the present invention are described in greater detail in the non-limiting examples that follow. The examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all and only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (*e.g.* amounts, temperature, *etc.)* but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric.

Source of human prostate cell samples and isolation of polynucleotides expressed by them

[0319] Candidate polynucleotides that may represent genes differentially expressed in cancer were obtained from both publicly-available sources and from cDNA libraries generated from selected cell lines and patient tissues. A normalized cDNA library was prepared from one patient tumor tissue and cloned polynucleotides for spotting on microarrays were isolated from the library. Normal and tumor tissues from 13 patients were processed to generate T7 RNA polymerase transcribed polynucleotides, which were, in turn, assessed for expression in the microarrays. The tissues that served as sources for these libraries and polynucleotides are summarized in Table 4.

[0320] Normalization: The objective of normalization is to generate a cDNA library in which all transcripts expressed in a particular cell type or tissue are equally represented [refs. 192 & 193], and therefore isolation of as few as 30,000 recombinant clones in an optimally normalized library may represent the entire gene expression repertoire of a cell, estimated to number 10,000 per cell. The source materials for generating the normalized prostate libraries were cryopreserved prostate tumor tissue from a patient with Gleason grade 3+3 adenocarcinoma and normal prostate biopsies from a pool of at-risk subjects under medical surveillance. Prostate epithelia were harvested directly from frozen sections of tissue by laser capture microdissection (LCM, Arcturus Engineering Inc., Mountain View, CA), carried out according to methods well known in the art (*e.g.* ref. 194), to provide substantially homogenous cell samples.

[0321] Total RNA was extracted from LCM-harvested cells using RNeasy™ Protect Kit (Qiagen, Valencia, CA), following manufacturer's recommended procedures. RNA was quantified using RiboGreen™ RNA quantification kit (Mo-

lecular Probes, Inc. Eugene, OR). One μg of total RNA was reverse transcribed and PCR amplified using SMART™ PCR cDNA synthesis kit (ClonTech, Palo Alto, CA). The cDNA products were size-selected by agarose gel electrophoresis using standard procedures (ref. 8). The cDNA was extracted using Bio 101Geneclcan® II kit (Qbiogene, Carlsbad, CA). Normalization of the cDNA was carried out using kinetics of hybridization principles: 1.0 μg of cDNA was denatured by heat at 100° C for 10 minutes, then incubated at 42°C for 42 hours in the presence of 120 mM NaCl, 10 mM Tris.HCl (pH=8.0), 5 mM EDTA.Na^+ and 50% formamide. Single-stranded cDNA ("normalized" cDNA) was purified by hydroxyapatite chromatography (#130-0520, BioRad, Hercules, CA) following the manufacturer's recommended procedures, amplified and converted to double-stranded cDNA by three cycles of PCR amplification, and cloned into plasmid vectors using standard procedures (ref. 8). All primers/adaptors used in the normalization and cloning process are provided by the manufacturer in the SMART™ PCR cDNA synthesis kit (ClonTech, Palo Alto, CA). Supercompetent cells (XL-2 Blue Ultracompetent Cells, Stratagene, California) were transfected with the normalized cDNA libraries, plated on plated on solid media and grown overnight at 36°C.

[0322] Characterization of normalized libraries: The sequences of 10,000 recombinants per library were analyzed by capillary sequencing using the ABI PRISM 3700 DNA Analyzer (Applied Biosystems, California). To determine the representation of transcripts in a library, BLAST analysis was performed on the clone sequences to assign transcript identity to each isolated clone, i.e. the sequences of the isolated polynucleotides were first masked to eliminate low complexity sequences using the XBLAST masking program (refs. 195, 196 and 197). Generally, masking does not influence the final search results, except to eliminate sequences of relative little interest due to their low complexity, and to eliminate multiple "hits" based on similarity to repetitive regions common to multiple sequences *e.g.* Alu repeats. The remaining sequences were then used in a BLASTN *vs.* GenBank search. The sequences were also used as query sequence in a BLASTX vs. NRP (non-redundant proteins) database search.

[0323] Automated sequencing reactions were performed using a Perkin-Elmer PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit containing AmpliTaq DNA Polymerase, FS, according to the manufacturer's directions. The reactions were cycled on a GeneAmp PCR System 9600 as per manufacturer's instructions, except that they were annealed at 20° C. or 30° C. for one minute. Sequencing reactions were ethanol precipitated, pellets were resuspended in 8 microliters of loading buffer, 1.5 microliters was loaded on a sequencing gel, and the data was collected by an ABI PRISM 3700 DNA Sequencer. (Applied Biosystems, Foster City, CA).

[0324] The number of times a sequence is represented in a library is determined by performing sequence identity analysis on cloned cDNA sequences and assigning transcript identity to each isolated clone. First, each sequence was checked to see if it was a mitochondrial, bacterial or ribosomal contaminant. Such sequences were excluded from the subsequent analysis. Second, sequence artifacts (*e.g.* vector and repetitive elements) were masked and/or removed from each sequence.

[0325] The remaining sequences were compared via BLAST [198] to GenBank and EST databases for gene identification and were compared with each other via FastA [199] to calculate the frequency of cDNA appearance in the normalized cDNA library. The sequences were also searched against the GenBank and GeneSeq nucleotide databases using the BLASTN program (BLASTN 1.3MP [198]). Fourth, the sequences were analyzed against a non-redundant protein (NRP) database with the BLASTX program (BLASTX 1.3MP [198]). This protein database is a combination of the Swiss-Prot, PIR, and NCBI GenPept protein databases. The BLASTX program was run using the default BLOSUM-62 substitution matrix with the filter parameter: "xnu+seg". The score cutoff utilized was 75.

[0326] Assembly of overlapping clones into contigs was done using the program Sequencher (Gene Codes Corp.; Ann Arbor, Mich.). The assembled contigs were analyzed using the programs in the GCG package (Genetic Computer Group, University Research Park, 575 Science Drive, Madison, Wis. 53711) Suite Version 10.1.

[0327] Summary of polynucleotides described herein: Table 6 provides a summary of polynucleotides isolated as described above and identified as corresponding to a differentially expressed gene (see below). Specifically, Table 6 provides: 1) the HERVK ORF for each clone ID; 2) the clone ID assigned to each sequence; 3) the % patients having the expression ratio of >/= 2X; >/= 2-5X; >/= 5X; and less than 1/2 X; and the Tumor/Normal mRNA Expression Ratio per patient "Pat", eg, patient 93, patient 95, patient 96, *etc.*

Detection of elevated levels of cDNA associated with prostate cancer using arrays

[0328] cDNA sequences representing a variety of candidate genes to be screened for differential expression in prostate cancer were assayed by hybridization on polynucleotide arrays. The cDNA sequences included cDNA clones isolated from cell lines or tissues as described above. The cDNA sequences analyzed also included polynucleotides comprising sequence overlap with sequences in the Unigene database, and which encode a variety gene products of various origins, functionality, and levels of characterization. cDNAs were spotted onto reflective slides (Amersham) according to methods well known in the art at a density of 9,216 spots per slide representing 4608 sequences (including controls) spotted in duplicate, with approximately 0.8 μl of an approximately 200ng/μl solution of cDNA.

[0329] PCR products of selected cDNA clones corresponding to the gene products of interest were prepared in a 50%

DMSO solution. These PCR products were spotted onto Amersham aluminum microarray slides at a density of 9216 clones per array using a Molecular Dynamics Generation III spotting robot. Clones were spotted in duplicate, for a total of 4608 different sequences per chip.

[0330] cDNA probes were prepared from total RNA obtained by laser capture microdissection (LCM, Arcturus Engineering Inc., Mountain View, CA) of tumor tissue samples and normal tissue samples isolated from the patients described above.

[0331] Total RNA was first reverse transcribed into cDNA using a primer containing a T7 RNA polymerase promoter, followed by second strand DNA synthesis. cDNA was then transcribed *in vitro* to produce antisense RNA using the T7 promoter-mediated expression (*e.g.* ref. 200), and the antisense RNA was then converted into cDNA. The second set of cDNAs were again transcribed *in vitro,* using the T7 promoter, to provide antisense RNA. This antisense RNA was then fluorescently labeled, or the RNA was again converted into cDNA, allowing for third round of T7-mediated amplification to produce more antisense RNA. Thus the procedure provided for two or three rounds of *in vitro* transcription to produce the final RNA used for fluorescent labeling. Probes were labeled by making fluorescently labeled cDNA from the RNA starting material. Fluorescently-labeled cDNAs prepared from the tumor RNA sample were compared to fluorescently labeled cDNAs prepared from normal cell RNA sample. For example, the cDNA probes from the normal cells were labeled with Cy3 fluorescent dye (green) and cDNA probes prepared from the tumor cells were labeled with Cy5 fluorescent dye (red).

[0332] The differential expression assay was performed by mixing equal amounts of probes from tumor cells and normal cells of the same patient. The arrays were pre-hybridized by incubation for about 2 hrs at 60°C in 5X SSC/0.2% SDS/1 mM EDTA, and then washed three times in water and twice in isopropanol. Following pre-hybridization of the array, the probe mixture was then hybridized to the array under conditions of high stringency (overnight at 42°C in 50% formanide, 5X SSC, and 0.2% SDS. After hybridization, the array was washed at 55°C three times as follows: 1) first wash in 1X SSC/0.2% SDS; 2) second wash in 0.1X SSC/0.2% SDS; and 3) third wash in 0.1X SSC.

[0333] The arrays were then scanned for green and red fluorescence using a Molecular Dynamics Generation III dual color laser-scanner/detector. The images were processed using BioDiscovery Autogene software, and the data from each scan set normalized. The experiment was repeated, this time labeling the two probes with the opposite color in order to perform the assay in both "color directions." Each experiment was sometimes repeated with two more slides (one in each color direction). The data from each scan was normalized, and the level fluorescence for each sequence on the array expressed as a ratio of the geometric mean of 8 replicate spots/genes from the four arrays or 4 replicate spots/gene from 2 arrays or some other permutation.

[0334] Table 6 summarizes the results for gene products differentially expressed in the prostate tumor samples relative to normal cells. The ratio of differential expression is expressed as the normalized hybridization signal associated with the tumor probe divided by the normalized hybridization signal with the normal probe; thus, a ratio greater than 1 indicates that the gene product is increased in expression in cancerous cells relative to normal cells, while a ratio of less than 1 indicates the opposite. The results from each patient are identified by "Pat" with the corresponding patient identification number. "Concordance" indicates the % of patients in which differential expression of the selected gene product in tumor cells was at least a two-fold different from normal cells.

[0335] In at least 79% of prostate patients assayed, 8 out of 10 genes, whose expression was elevated by at least 500%, were represented in HERV-K(CH) sequences.

[0336] Table 6 provides those gene products that were differentially expressed and were classified as gag, 5'-pol (reverse transcriptase) and 3'-pol (integrase) related sequences. It may be possible to examine the function of these gene products in development of cancer and metastasis through use of small molecule inhibitors known to affect the activity of such enzymes.

Analysis of the Prostate Cancer Associated Sequences

[0337] In order to determine whether there was homology to any known sequences, the PCR products of 16 different clones from one prostate tumor patient were sequenced. PCR products from these and other clones from the same library were spotted on DNA microarrays. RNA from 13 prostate tumor patients were assayed on the microaways and then the full inserts of some of the 16 clones were sequenced (Table 6).

[0338] The 16 isolates were initially determined in a first pass sequencing reaction to have the sequences as shown in SEQ IDs 27-39, inclusive. The isolate from the normal prostate tissue was initially determined in a first pass sequencing reaction to have the sequence as shown in SEQ ID 41. A first pass sequencing reaction refers to a high-throughput process, where PCR reactions generate the sequencing template then sequencing is performed with one of the PCR primers, in a single direction. A search of public databases revealed that these 16 isolates have some degree of identity to regions of the human endogenous retrovirus HERV-K(II) sequence disclosed in Genbank accession number AB047240 and shown in SEQ ID 44, and also to HERV-K(10), but are nonetheless unique.

[0339] The isolates were subjected to a second round of nucleic acid sequencing and were found to have the sequences

as shown in SEQ IDs 14-26, inclusive. The isolate from the normal prostate tissue was subjected to a second round of nucleic acid sequencing and found to have the sequence as shown in SEQ ID 40. This second round of sequencing is a customized process, where sequencing is performed on purified dsDNA template in a DNA vector. Sequencing is done from both ends of the template, forward and reverse, with primers designed from the flanking regions of the vector, and new primers are synthesized for every additional reaction needed to span the entire insert.

[0340] The Genbank disclosure of HERV-K(II) provides only an incomplete characterization of its genetic features and no association with any disease. The Genbank disclosure characterizes HERV-KII as having a gag gene located at nucleotide 2113-4116 and an env gene located at nucleotide 7437-8174. Detailed analysis of the reported HERV-K (II) sequence indicates that the HERV-K(II) genome includes regions related to gag, protease, 5'-end of pol (reverse transcriptase) and 3'-end of pol (integrase) domains of a retrovirus. Specifically, the location of the protease gene is from about nucleotide 3917 to about 4920 and the location of the polymerase domain is from about nucleotide 4797 to about 7468.

[0341] Composite HERV-K(CH) polynucleotide sequences are shown in SEQ IDs 7, 8, 9 and 10 and Figure 1 provides a schematic illustration of a human endogenous retrovirus and the HERVK(CH) species within the schematic illustration. SEQ ID 7 is a composite sequence of the polynucleotides SEQ IDs 14-16, inclusive, and has a consensus sequence as shown in SEQ ID 11. This region corresponds to the gag region of a human endogenous retrovirus. SEQ IDs 8 and 9 are composites sequence of the polynucleotides having a sequence as shown in SEQ IDs 17-20, inclusive, and has a consensus sequence as shown in SEQ ID 12. This region corresponds to the 5' pol region of a human endogenous retrovirus. SEQ ID 10 is a composite sequence of the polynucleotides having a sequence as shown in SEQ IDs 21-26, inclusive, and has a consensus sequence as shown in SEQ ID 13. This region corresponds to the 3' pol region of a human endogenous retrovirus

[0342] Homology to HERV-K(II) gag region varied from 87% to 99%. Homology to HERVK(II) 5'-pol (reverse transcriptase) region varied from 87% to 97%. Homology to HERV-K(II) 3'-pol (integrase) region was approximately 89%. When compared to the human endogenous provirus HERV-K10, the homology of the gag region clones was approximately 79%, the 5'-pol region between 81 % and 89% and the 3'-pol region was approximately 89%. Table 5 illustrates the homology of the sequences of the individual clones with the corresponding HERV-K(II) and HERV-K(10) regions. Because the presence of polyA stretches in the HERV-K(CH) sequences (and deposited isolates) may be an artifact of cloning, the % identity shown in Table 5 was determined with alignments performed with polynucleotides excluding the terminal polyA stretch.

[0343] Consensus polynucleotide sequences SEQ IDs 11-13 were generated with Multiple Sequence Alignment (MSA), a web implementation of the GCG Pileup and Pretty programs. The program uses a clustering algorithm similar to the Clustal program described in reference 201. The default values for the alignments and consensus extraction were 8 for gap open and 2 for gap extension. The poling plurality or minimum number of like sequences specified to assign a residue to the consensus sequence was 2.

[0344] The polynucleotide sequences shown in SEQ IDs 14-16, inclusive, were used for the consensus polynucleotide sequence shown in SEQ ID 11. The polynucleotide sequences shown in SEQ IDs 17-20, inclusive, were used for the consensus polynucleotide sequence shown in SEQ ID 12. The polynucleotide sequences shown in SEQ IDs 21-26, inclusive, were used for the consensus polynucleotide shown in SEQ ID 13. The "N" represents where there is no qualifying minimum representative base. i.e. at least two sequences with the same base at that site.

[0345] Northern blotting of prostate cancer cell lines using nucleotides 243-end of SEQ ID 150 labeled as a probe indicates that they express PCAV transcripts of several sizes, corresponding to both full-length viral genomic sequences and to sub-genomic spliced transcripts (Figure 5). Expression of such transcripts have also been observed in teratocarcinoma cell lines [15], as shown in lane 1 of figure 14.

Investigation of other human endogenous retroviruses

[0346] HERV-K(CH) is a member of the HML-2 subgroup of the HERV-K family. HERV-K(II) and HERV-K(10) are also members of this sub-group.

[0347] The same microarray techniques as described above were used to study the expression of members of the HERV-K family in the HML-2 and HML-6 subgroups in prostate tumor tissue. The expression of HERV-H viruses was also studied.

[0348] The results in table 9 show that HERV-H is not up-regulated in prostate tumors. The HML-6 subgroup of HERV-K is also not up-regulated. The only endogenous retroviruses that are up-regulated in prostate tumors are in the HML-2 subgroup.

Investigation of tumors other than prostate tumors

[0349] HML-2 endogenous retroviruses are up-regulated in prostate tumors. Tumor samples taken from patients with

breast and colon cancer were investigated for up-regulation of HML-2 and HML-6 HERV-K viruses using the microarray techniques described above.

[0350] The results in table 10 show that the HML-2 viruses are up-regulated in tissue from prostate tumors, but not from colon or breast tumors. HML-6 expression is not up-regulated in any of the tumors.

[0351] *Detection of HERV-K(CH) sequences in human prostate cancer cells and tissues.*

[0352] DNA from prostate cancer tissue and other human cancer tissues, human colon, normal human tissues including non-cancerous prostate, and from other human cell lines are extracted following the procedure of ref. 202. The DNA is re-suspended in a solution containing 0.05 M Tris HCl buffer, pH 7.8, and 0.1 mM EDTA, and the amount of DNA recovered is determined by microfluorometry using Hoechst 33258 dye [ref. 203].

[0353] Polymerase chain reaction (PCR) is performed using Taq polymerase following the conditions recommended by the manufacturer (Perkin Elmer Cetus) with regard to buffer, Mg^{2+}, and nucleotide concentrations. Thermocycling is performed in a DNA cycler by denaturation at 94° C. for 3 min. followed by either 35 or 50 cycles of 94° C. for 1.5 min., 50° C. for 2 min. and 72° C. for 3 min. The ability of the PCR to amplify the selected regions of the HERV-K(CH) gene is tested by using a cloned HERV-K(CH) polynucleotide(s) as a positive template(s). Optimal Mg^{2+}, primer concentrations and requirements for the different cycling temperatures are determined with these templates. The master mix recommended by the manufacturer is used. To detect possible contamination of the master mix components, reactions without template are routinely tested.

[0354] Southern blotting and hybridization are performed as described in reference 204, using the cloned sequences labeled by the random primer procedure [205]. Prehybridization and hybridization are performed in a solution containing 6xSSPE, 5% Denhardt's, 0.5% SDS, 50% formamide, 100 μg/ml denaturated salmon testis DNA, incubated for 18 hrs at 42° C., followed by washings with 2xSSC and 0.5% SDS at room temperature and at 37° C. and finally in 0.1xSSC with 0.5% SDS at 68° C. for 30 min (ref. 8). For paraffin-embedded tissue sections the conditions described in ref. 206 are followed using primers designed to detect a 250 bp sequence.

Expression of cloned polynucleotides in host cells.

[0355] To study the polypeptide products of HERV-K(CH) cDNA, restriction fragments from the HERV-K(CH) cDNA are cloned into the expression vector pMT2 (pages 16.17-16.22 of ref. 8) and transfected into COS cells grown in DMEM supplemented with 10% FCS. Transfections are performed employing calcium phosphate techniques (pages 16.32-16.40 of ref. 8) and cell lysates are prepared forty-eight hours after transfection from both transfected and untransfected COS cells. Lysates are subjected to analysis by immunoblotting using anti-peptide antibody.

[0356] In immunoblotting experiments, preparation of cell lysates and electrophoresis are performed according to standard procedures. Protein concentration is determined using BioRad protein assay solutions. After semi-dry electrophoretic transfer to nitro-cellulose, the membranes are blocked in 500 mM NaCl, 20 mM Tris, pH 7.5, 0.05% Tween-20 (TTBS) with 5% dry milk. After washing in TTBS and incubation with secondary antibodies (Amersham), enhanced chemiluminescence (ECL) protocols (Amersham) are performed as described by the manufacturer to facilitate detection.

Generation of antibodies against polypeptides.

[0357] Polypeptides, unique to HERV-K(CH) are synthesized or isolated from bacterial or other (e.g. yeast, baculovirus) expression systems and conjugated to rabbit serum albumin (RSA) with m-maleimido benzoic acid N-hydroxysuccinimide ester (MBS) (Pierce, Rockford, Ill.). Immunization protocols with these peptides are performed according to standard methods. Initially, a pre-bleed of the rabbits is performed prior to immunization. The first immunization includes Freund's complete adjuvant and 500 μg conjugated peptide or 100 μg purified peptide. All subsequent immunizations, performed four weeks after the previous injection, include Freund's incomplete adjuvant with the same amount of protein. Bleeds are conducted seven to ten days after the immunizations.

[0358] For affinity purification of the antibodies, the corresponding HERV-K(CH) polypeptide is conjugated to RSA with MBS, and coupled to CNBr-activated Sepharose (Pharmacia, Sweden). Antiserum is diluted 10-fold in 10 mM Tris-HCl, pH 7.5, and incubated overnight with the affinity matrix. After washing, bound antibodies are eluted from the resin with 100 mM glycine, pH 2.5.

ELISA assay for Detecting HERV-KCH) Gag and/or Pol related sequences.

[0359] To test blood samples for antibodies that bind specifically to recombinantly produced HERV-K(CH) antigens, the following procedure is employed. After the recombinant HERVK(CH) pol or gag or env related polypeptides are purified, the recombinant polypeptide is diluted in PBS to a concentration of 5 μg/ml (500 ng/100 μ). 100 microliters of the diluted antigen solution is added to each well of a 96-well Immulon 1 plate (Dynatech Laboratories, Chantilly, Va.), and the plate is then incubated for 1 hour at room temperature, or overnight at 4° C., and washed 3 times with 0.05%

Tween 20 in PBS. Blocking to reduce nonspecific binding of antibodies is accomplished by adding to each well 200 μl of a 1% solution of bovine serum albumin in PBS/Tween 20 and incubation for 1 hour. After aspiration of the blocking solution, 100 μl of the primary antibody solution (anticoagulated whole blood, plasma, or serum), diluted in the range of 1/16 to 1/2048 in blocking solution, is added and incubated for 1 hour at room temperature or overnight at 4° C. The wells are then washed 3 times, and 100μl goat anti-human IgG antibody conjugated to horseradish peroxidase (organon Teknika, Durham, N.C.), diluted 1/500 or 1/1000 in PBS/Tween 20, 100 μl of o-phenylenediamine dihydrochloride (OPD, Sigma) solution is added to each well and incubated for 5-15 minutes. The OPD solution is prepared by dissolving a 5 mg OPD tablet in 50 ml 1% methanol in H_2O and adding 50 μl 30% H_2O_2 immediately before use. The reaction is stopped by adding 25 1 of 4M H_2SO_4 Absorbance are read at 490 nm in a microplate reader (Bio-Rad).

Preparation of vaccines.

[0360] The present invention also relates to a method of stimulating an immune response against cells that express HERV-K(CH) polypeptides in a patient using HERV-K(CH) gag, and/or pol polypeptides of the invention that acts as an antigen produced by or associated with a malignant cell. This aspect of the invention provides a method of stimulating an immune response in a human against prostate cells or cells that express a HERV-K(CH) pol or gag polynucleotides and polypeptides. The method comprises the step of administering to a human an immunogenic amount of a polypeptide comprising: (a) the amino acid sequence of a human endogenous retrovirus HERV-K(CH) polypeptide or (b) a mutein or variant of a polypeptide comprising the amino acid sequence of a human endogenous retrovirus HERV-K(CH) polypeptide.

Generation of transgenic animals expressing polypeptides as a means for testing therapeutics.

[0361] HERV-K(CH) nucleic acids are used to generate genetically modified non-human animals, or site specific gene modifications thereof, in cell lines, for the study of function or regulation of prostate tumor-related genes, or to create animal models of diseases, including prostate cancer. The term "transgenic" is intended to encompass genetically modified animals having an exogenous HERV-K(CH) gene(s) that is stably transmitted in the host cells where the gene (s) may be altered in sequence to produce a modified polypeptide, or having an exogenous HERV-K(CH) LTR promoter operably linked to a reporter gene. Transgenic animals may be made through a nucleic acid construct randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like. Of interest are transgenic mammals, e.g. cows, pigs, goats, horses, *etc.,* and particularly rodents, e.g. rats, mice, *etc.*

[0362] The modified cells or animals are useful in the study of HERV-K(CH) gene function and regulation. For example, a series of small deletions and/or substitutions may be made in the HERV-K(CH) gene to determine the role of different domains in prostate tumorigenesis. Specific constructs of interest include, but are not limited to, anti-sense constructs to block HERV-K(CH) gene expression, expression of dominant negative HERV-K(CH) gene mutations, and over-expression of a HERV-K(CH) gene. Expression of a HERV-K(CH) gene or variants thereof in cells or tissues where it is not normally expressed or at abnormal times of development is provided. In addition, by providing expression of polypeptides derived from HERV-K(CH) in cells in which it is otherwise not normally produced, changes in cellular behavior can be induced.

[0363] DNA constructs for random integration need not include regions of homology to mediate recombination. Conveniently, markers for positive and negative selection are included. For various techniques for transfecting mammalian cells, see ref. 207.

[0364] For embryonic stem (ES) cells, an ES cell line is employed, or embryonic cells is obtained freshly from a host, e.g. mouse, rat, guinea pig, etc. Such cells are grown on an appropriate fibroblast-feeder layer or grown in the presence of appropriate growth factors, such as leukemia inhibiting factor (LIF). When ES cells are transformed, they may be used to produce transgenic animals. After transformation, the cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be detected by employing a selective medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of integration of the construct. Those colonies that are positive may then be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old superovulated females. The ES cells are trypsinized, and the modified cells are injected into the blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant females. Females are then allowed to go to term and the resulting chimeric animals screened for cells bearing the construct. By providing for a different phenotype of the blastocyst and the ES cells, chimeric progeny can be readily detected.

[0365] The chimeric animals are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. If the gene alterations cause lethality at some point in development, tissues or organs are maintained as allogeneic or congenic grafts or transplants, or in in vitro culture. The transgenic animals may be any non-human mammal, such as laboratory animals, domestic animals, etc. The transgenic animals are used in functional studies, drug screening, etc., e.g. to determine the effect of a candidate drug on prostate

cancer, to test potential therapeutics or treatment regimens, etc.

Diagnostic Imaging Using HERV-K(CH) Specific Antibodies

[0366] The present invention encompasses the use of antibodies to HERV-K(CH) polypeptides to accurately stage prostate cancer patients at initial presentation and for early detection of metastatic spread of prostate cancer. Radioimmunoscintigraphy using monoclonal antibodies specific for HERV-K(CH) gag or HERV-K(CH) pol or portions thereof or other HERV-K(CH) polypeptides can provide an additional tumor-specific diagnostic test. The monoclonal antibodies of the instant invention are used for histopathological diagnosis of prostate carcinomas.

[0367] Subcutaneous human xenografts of prostate cancer cells in nude mice is used to test whether a technetium-99m (^{99m}Tc)-labeled monoclonal antibody of the invention can successfully image the xenografted prostate cancer by external gamma scintography as described for seminoma cells in ref. 208. Each monoclonal antibody specific for a HERV-K(CH) polypeptide is purified from ascitic fluid of BALB/c mice bearing hybridoma tumors by affinity chromatography on polypeptide A-Sepharose. Purified antibodies, including control monoclonal antibodies such as an avidin-specific monoclonal antibody [209] are labeled with ^{99m}Tc following reduction, using the methods of refs. 210 and 211. Nude mice bearing human prostate cancer cells are injected intraperitoneally with 200-500 μCi of ^{99m}Tc-labeled antibody. Twenty-four hours after injection, images of the mice are obtained using a Siemens ZLC3700 gamma camera equipped with a 6 mm pinhole collimator set approximately 8 cm from the animal. To determine monoclonal antibody biodistribution following imaging, the normal organs and tumors are removed, weighed, and the radioactivity of the tissues and a sample of the injectate are measured. Additionally, HERV-K(CH) -specific antibodies conjugated to antitumor compounds are used as prostate cancer-specific chemotherapy.

DEPOSITS

[0368] The materials listed in Table 7 were deposited with the American Type Culture Collection.

[0369] All publications and patent applications mentioned in this specification are incorporated herein by reference to the same extent as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference.

[0370] The foregoing description of preferred embodiments of the invention has been presented by way of illustration and example for purposes of clarity and understanding. It is not intended to be exhaustive or to limit the invention to the precise forms disclosed. It will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that many changes and modifications may be made thereto without departing from the spirit of the invention. It is intended that the scope of the invention be defined by the appended claims and their equivalents.

TABLE 1- GAG protease (5') probes, isolate specific

Isolate	Nucleotides	SEQ ID	Isolate	Nucleotides	SEQ ID
K(CH)	1224-1238	161		1490-1510	188
KII	2098-2114	162		1502-1520	189
	874-890	163		1522-1538	190
	894-908	164		1561-1576	191
	910-927	165		1586-1605	192
	927-944	166		1620-1635	193
	989-1004	167		1653-1669	194
	1019-1036	168		1698-1723	195
	1046-1063	169		1722-1743	196
	1063-1078	170		1748-1762	197
	1084-1103	171		1773-1788	198
	1131-1145	172		1820-1834	199
	1148-1163	173		1872-1887	200
	1164-1185	174	K10	1917-1935	201
K10	1206-1223	175		1940-1955	202
	1216-1235	176		1955-1969	203
	1243-1260	177		1973-1995	204
	1258-2375	178		2008-2042	205
	1277-1295	179		2049-2064	206
	1300-1329	180		2076-2093	207
	1347-1361	181		2097-2113	208
	1367-1382	182		2122-2139	209
	1392-1410	183		2148-2118	210
	1412-1428	184		2176-2196	211
	1426-1442	185		2198-2212	212
	1445-1461	186		2219-2235	213
	1463-1477	187		2246-2261	214

TABLE 2 - Protease (3'seq) Polymerase (5'seq) Probes

Isolate	Nucleotides	SEQ ID	Isolate	Nucleotides	SEQ ID
K(CH) consensus	170-188	215	K10	11-38	113
	205-221	216		37-54	114
	253-268	217		70-90	115
	316-336	218		226-243	116
	401-417	219		249-264	117
	490-504	220		308-324	118
	538-552	221		327-342	119
	872-886	222		381-397	120
K(CH)	109-125	223		440-454	121
	1374-1388	224		541-557	122
	1402-1416	225		678-698	123
KII	140-159	110		722-741	124
	410-426	111		753-767	125
	1127-1141	112		771-785	126
				854-869	127
				872-890	128
				1195-1209	129
				1308-1323	130
				1335-1349	131
				1349-1365	132

TABLE 3 - 3' POL probes only

Isolate	Nucleotides	SEQ ID
K(CH) consensus	3-17	133
	25-39	134
	82-104	135
	136-151	136
	154-169	137
	189-203	138
	322-337	139
	461-475	140
	630-645	141
	712-727	142
	757-771	143
	818-833	144
KII	1636-1651	145

TABLE 4 - ORFS and sources of initial isolates/clones from prostate cDNA libraries

HERVK ORF	Chiron Clone ID	Source of Clone
gag	035JN002.E02	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
gag	035JN013.H09	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
gag	035JN023.F12	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
gag	037XN001.D10	Normal Prostate Tissue, Pooled from 10 individuals
pol5'	035JN001.F06	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pol5'	035JN003.E06	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pol5'	035JN013.C11	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pol5'	035JN013.F03	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pol3'	035JN003.G09	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pol3'	035JN010.A09	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pol3'	035JN015.F06	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pol3'	035JN020.B12	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pol3'	035JN020.D07	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pol3'	035JN022.G09	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pol3'	035JN015.H02	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pol3'	035JN016.H02	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3

TABLE 5 - Identity of HERV-K(CH) polynucleotides with HERV-K(II) and HERV-K(10)

Clone ID	Region	% Identity HERV-K(II)	% Identity HERV-K(10)
035JN003.G09	3'-pol	89.423	89.423
035JN010.A09	3'-pol	89.663	89.663
035JN015.F06	3'-pol	89.423	89.423
035JN020.B1	3'-pol	89.303	89.303
035JN020.D07	3'-pol	89.614	89.614

(continued)

Clone ID	Region	% Identity HERV-K(II)	% Identity HERV-K(10)
035JN022.G09	3'-pol	89.354	89.354
035JN002.E02	gag	99.524	79.881
035JN013.H09	gag	99.017	79.975
035JN023.F12	gag	98.849	79.335
035XN001.D10	gag	87.383	79.947
035JN001.F0	5'-pol	97.211	88.844
035JN003.E06	5'-pol	97.450	86.723
035JN013.C11	5'-pol	97.156	85.444
035JN013.F03	5'-pol	87.962	81.521

Table 6

DNA microarray results: 13 patients tumor vs. normal prostate, expression of HERV-K RNA																		
		Percent Patient with Express on Ratio				Tumor/NormalmRNA Expression Ratio												
HERVK ORF	Chiron Clone ID	>=2 x	>=2-5 x	>=5 x	<=half x	Pat 93	Pat 95	Pat 96	Pat 97	Pat 151	Pat 155	Pat 231	Pat 232	Pat 251	Pat 282	Pat 286	Pat 294	Pat 351
gag	035JN002.E02	57.1	42.9	7.1	0.0	4.8	3.0	2.1	1.0 ’	2.3	2.5	1.9	1.7	6.9	1.5	0.6	2.6	2.9
gag	035JN013.H09	78.6	78.6	50.0	0.0	9.3	4.5	5.2	1.4	5.5	13.8	4.2	3.5	31.2	4.5	1.0	12.1	8.6
gag	035JN023.F12	78.6	78.6	57.1	0.0	9.1	4.1	5.1	1.6	5.5	17.0	4.5	3.2	28.2	5.2	1.0	12.7 7	7.3
gag	037XN001.D10	64.3	64.3	14.3	0.0	5.4	3.4	2.5	1.5	3.6	4.6	2.9	1.8	10.0	1.7	1.0	3.5	4.3
pol5prime	035JN001.F06	42.9	21.4	7.1	0.0	2.0	2.6	1.8	1.5	2.7	1.8	2.0	1.8	7.8	1.2	1.0	1.9	2.3
pol5prime	035JN003.E06	42.9	21.4	7.1	0.0	2.1	2.6	1.8	1.4	2.6	1.9	2.0	1.7	7.7	1.2	1.0	1.8	2.1
pol5prime	035JN013.C11	85.7	78.6	57.1	0.0	6.9	5.6	6.9	2.0	7.4	24.0	4.8	4.3	37.4	4.4	1.0	13.1	8.8
pol5prime	035JN013.F03	85.7	71.4	21.4	0.0	4.6	3.4	3.7	22	4.6	8.4	4.1	3.4	21.8	2.3	1.0	5.0	5.8
pol3prime	035JN003.G09	71.4	57.1	7.1	0.0	4.1	3.3	3.3	1.6	4.9	3.3	2.2	3.5	14.9	1.5	1.0	2.5	3.9
pol3prime	035JN010.A09	85.7	78.6	71.4	0.0	8.0	4.4	12.6	2.1	12.4	55.9	5.1	9.5	70.0	5.8	1.0	26.3	9.7
pol3prime	035JN015.F06	85.7	78.6	71.4	0.0	7.6	4.0	12.8	2.2	11.9	53.4	5.1	8.0	69 7	5.9	1.0	25.3	9.1
pol3prime	03JN020.B12	85.7	78.6	64.3	0.0	7.0	4.0	10.5	2.2	11.9	34.9	5.0	6.8	44.5	5.2	1.0	15.2	8.1
pol3prime	035JN020.D07	85.7	78.6	57.1	0.0	6.0	3.2	8.7	2.0	13.7	22.9	4.6	8.6	58.2	3.8	1.0	15.0	7.6
pol3prime	035JN022.G09	78.6	78.6	57.1	0.0	6.6	4.2	6.6	2.0	8.8	12.7	4.5	5.3	28.0	2.6	1.0	5.9	7.8
pol3prime	035JN015.H02	85.7	78.6	57.1	0.0	7.9	4.2	9.0	2.1	10.7	35.3	4.7	7.5	49.5	4.8	1.0	18.2	8.7
pol3prime	035JN016.H02	71.4	71.4	14.3	0.0	3.8	3.0	3.4	1.9	4.3	5.0	3.0	3.1	14.1	1.7	1.0	2.6	5.0

TABLE 7 - DEPOSITS

ATCC = American Type Culture Collection
CMCC = Chiron Master Culture Collection
All deposits made 10th April 2000

Cell Line	CMCC Accession No.	ATCC Accession No.
035JN003G09	5400	PTA 2561
035JN010A09	5401	PTA 2572
035JN015F06	5402	PTA 2566
035JN015H02	5403	PTA 2571
035JN020B12	5405	PTA 2562
035JN020D07	5406	PTA 2573
035JN022G09	5413	PTA 2560
035JN002E02	5404	PTA 2565
035JN013H09	5408	PTA 2568
035JN023F12	5409	PTA 2564
035XN001D10	5410	PTA 2569
035JN001F06	5411	PTA 2567
035JN003E06	5412	PTA 2559
035JN013C11	5407	PTA 2563
035JN013F03	5415	PTA 2570

TABLE 8 - Sequence listing

SEQ ID	DESCRIPTION
1	U5 region of herv-k(hml-2.hom) [GenBank AF074086]
2	U3 region of herv-k(hml-2.hom)
3	R region of herv-k(hml-2.hom)
4	RU5 region of herv-k(hml-2.hom)
5	U3R region of herv-k(hml-2.hom)
6	Non-coding region between U5 and first 5' splice site of herv-k(hml-2.hom)
7	Composite of three HERV-K(CH) polynucleotides [SEQ IDs 14-16] positioned in the gag region.
8&9	Composite of four HERV-K(CH) polynucleotides [SEQ IDs 17-20] positioned in the 5' pol region
10	Composite of six HERV-K(CH) polynucleotides [SEQ IDs 21-26] positioned in the 3' pol region
11	Consensus sequence of HERV-K(CH) gag region
12	Consensus sequence of HERV-K(CH) 5' pol region
13	Consensus sequence of HERV-K(CH) 3' pol region
14	Sequence for clone 035JN002.E02.
15	Sequence for clone 035JN023.F12.
16	Sequence for clone 035JN013.H09.
17	Sequence for clone 035JN013.C11

(continued)

SEQ ID	DESCRIPTION
18	Sequence for clone 035JN003.E06.
19	Sequence for clone 35JN001.F06.
20	Sequence for clone 035JN013.F03.
21	Sequence for clone 035JN020.D07.
22	Sequence for clone 035JN015.F06.
23	Sequence for clone 035JN003.G09.
24	Sequence for clone 035JN020.B12.
25	Sequence for clone 035JN022.G09.
26	Sequence for clone 035JN010.A09.
27	Sequence for clone 035JN002.E02.
28	Sequence for clone 035JN023.F12.
29	Sequence for clone 035JN013.1-H09.
30	Sequence for clone 035JN013.C11.
31	Sequence for clone 035JN003.E06.
32	Sequence for clone 035JN001.F06.
33	Sequence for clone 035JN013.F03.
34	Sequence for clone 035JN020.D07.
35	Sequence for clone 035JN015.F06.
36	Sequence for clone 035JN003.G09.
37	Sequence for clone 035JN020.B12.
38	Sequence for clone 035JN022.G09.
39	Sequence for clone 035JN010.A09.
40	Sequence for clone 037XN001.D10 and isolated from normal prostate tissue.
41	Sequence for clone 037XN001.D10 and isolated from normal prostate tissue.
42	EST polynucleotide sequence shown in GenBank accession number Q60732.
43	EST polynucleotide sequence SEQ ID 407 of WO 00/04149
44	Polynucleotide sequence for HERV-KII
45	Polynucleotide sequence for HERV-K10
46-49	Amino acid translations of SEQ IDs 11, 14, 15, 16
50-55	Amino acid translations of SEQ IDs 21-26 (note PSFGK motifs)
56-57	Amino acid translations of SEQ IDs 27 & 28
58	Consensus polypeptide sequence inferred from SEQ IDs 21-26
59-82	Polynucleotide probes not in SEQ IDs 42-45
83 & 84	Polynucleotide probes shared with SEQ IDs 42-45
85	HERV-K108 gag CDS
86	HERV-K108 prt CDS
87	HERV-K108 pol CDS
88	HERV-K108 env CDS

(continued)

SEQ ID	DESCRIPTION
89	HERV-K108 cORF 5' CDS
90	HERV-K108 cORF 3' CDS
91	HERV-K(C7) gag CDS
92	HERV-K(C7) gag amino acid sequence
93	HERV-K(C7) pol CDS
94	HERV-K(C7) pol amino acid sequence
95	HERV-K(C7) env CDS
96	HERV-K(C7) env amino acid sequence
97	HERV-K(II) gag CDS
98	HERV-K(II) gag amino acid sequence
99	HERV-K(II) prt CDS
100	HERV-K(II) pol CDS
101	HERV-K(II) env CDS
102	HERV-K10 gag CDS
103	HERV-K10 gag(i)
104	HERV-K10 gag(ii)
105	HERV-K10 prt CDS
106	HERV-K10 prt amino acid sequence
107	HERV-K10 pol/env CDS
108	HERV-K10 pol/env amino acid sequence
109	cORF amino acid sequence
110-132	Table 2 probes (cont^{d} at SEQ IDs 215-225)
133-145	Table 3 probes
146	HML-2.HOM ('ERVK6') gag amino acid sequence
147	HML-2.HOM ('ERVK6') prt amino acid sequence
148	HML-2.HOM ('ERVK6') pol amino acid sequence
149	HML-2.HOM ('ERVK6') env amino acid sequence
150	LTR of herv-k(hml-2.hom)
151-154	HML-2 LTR sequences
155 & 156	herv-k(hml-2.hom) RU5 region (5' and 3' regions, respectively)
157	Env consensus nucleic acid sequence (Figure 6)
158	Gag consensus sequence (Figure 7)
159	Pol consensus sequence (Figure 8)
160	Env consensus sequence (Figure 9)
161-214	Table 1 probes
215-225	Table 2 probes (cont^{d} from SEQ IDs 110-132)

TABLE 9 - Expression of HERV-H and HERV-K in prostate tumors

The "Result" column gives the % of patient samples which showed up-regulation of the GenBank sequence given in the first column in tumor tissue relative to non-tumor tissue.

GenBank ID	HERV	HML Subgroup	Result
AB047240	K	HML-2	65
AF164611	K	HML-2	63
AF164612	K	HML-2	63
AF079797	K	HML-6	3
BC005351	H	-	0
XM_054932	H	-	0

TABLE 10 - Expression of HERV-K viruses in colon and breast tumors

The "Result" columns give the % of patient samples which showed up-regulation of the Genbank sequence given in the first column in tumor tissue relative to non-tumor tissue.

GenBank ID	HERV	HML Subgroup	Result		
			Prostate	Breast	Colon
AB047240	K	HML-2	65	0	2
AF079797	K	HML-6		3 6	0
AF164611	K	HML-2	63	0	2
AF164612	K	HML-2	63	6	2

TABLE 11 - HML-2 subgroup of HERV-K Family

Query	Query Length	Target Locus	Target Description	Target Length	Score	Pscore	Matches	Similarities	Percent Alignment	Percent Query	Query Start	Query end	Target Start	Target End	Open Gap Penalty	Extension Penalty
N4	7428	NT_022283S1.2	/contig_orient=none /start=1 /end=160119 /chrom=2 Homo	102399	72570	3.9E-47	7334	7334	98	98	7428	1	13506	20899	-20	-5
N4	7428	NT_007386S1.3	/contig_orient=complement /start=1 /end=250001 /chrom=6	102399	72570	3.9E-47	7334	7334	98	98	1	7428	29800	37193	-20	-5
N4	7428	NT_009509S2.3	/contig_orient=forward /start=250002 /end=500002 /chrom=12	102399	72379	5.3E-47	7329	7329	98	98	7428	1	136539	143951	-20	-5
N4	7428	NT_009151S32.3	/contig_orient=complement /start=7623180 /end=7873180	102399	72707	3.1E-47	7345	7345	98	98	1	7428	114716	122137	-20	-5
N4	7428	NT_023901S1.2	/contig_orient=none /start=1 /end=166310 /chrom=8	102399	70366	1.3E-45	7222	7222	97	97	7428	1	94194	101616	-20	-5
N4	7428	NT_025820S3.2	/contig_orient=complement /start=455636 /end=651270	102399	67986	5.9E-44	7112	7112	95	95	1	7428	164603	172033	-20	-5
N4	7428	NT_024249S1.2	/contig_orient=none /start=1 /end=167403 /chrom=11 Homo	102399	67986	5.9E-44	7112	7112	95	95	1	7428	18873	26303	-20	-5
N4	7428	NT_011519S9.5	/contig_orient=forward /start=2016320 /end=2266320	102399	68342	3.4E-44	7058	7058	95	95	1	7428	62776	69910	-20	-5
N4	7428	NT_006788S1.3	/contig_orient=complement /start=1 /end=250000 /chrom=5	102399	68510	2.6E-44	7066	7066	95	95	1	7428	144115	151250	-20	-5
N4	7428	NT_004858S5.3	/contig_orient=complement /start=999278 /end=1248551	102399	68624	2.1E-44	7072	7072	95	95	7428	1	23642	30777	-20	-5
N4	7428	NT_005795S3.3	/contig_orient=forward /start=405779 /end=655779 /chrom=3	102399	67968	6.1E-44	7040	7040	94	94	1	7428	122036	129165	-20	-5
N4	7428	NT_025140S3.3	/contig_orient=none /start=449919 /end=649836 /chrom=19	97618	.68168	4.4E-44	7049	7049	94	94	7428	1	174979	182103	-20	-5
N4	7428	NT_009334S8.3	/contig_orient=complement /start=1574760 /end=1824759	102399	65447	3.4E-42	6913	6913	93	93	7428	1	116705	123823	-20	-5
N4	7428	NT_004406S4.3	/contig_orient=forward /start=797371 /end=985557 /chrom=1	85887	65099	6E-42	6910	6910	92	93	1	7428	103675	110860	-20	-5
N4	7428	NT_011192S4.3	/contig_orient=forward /start=750004 /end=949429 /chrom=19	102399	62351	4.9E-40	6844	6844	90	92	7428	1	17828	25313	-20	-5
N4	7428	NT_007592S14.3	/contig_orient=forward /start=3276099 /end=3526099 /chrom=6	102399	56493	5.7E-36	6795	6795	79	91	1	7428	141741	150230	-20	-5
N4	7428	NT_011512S23.3	/contig_orient=forward /start=5505084 /end=5755084	102399	64096	3E-41	6818	6818	92	91	7428	38	93282	100383	-20	-5
N4	7428	NT_019638S1.3	/contig_orient=none /start=1 /end=250001 /chrom=19	102399	57114	2.1E-36	6723	6723	82	90	7427	1	179318	187384	-20	-5
N4	7428	NT_022411S1.3	/contig_orient=none /start=1 /end=163648 /chrom=3	61348	65630	2.6E-42	6734	6734	95	90	7024	1	140614	147637	-20	-5
N4	7428	NT_005632S1.3	/contig_orient=complement /start=1 /end=214350 /chrom=3	102399	62739	2.6E-40	6630	5630	91	89	7428	236	48116	55040	-20	-5
N4	7428	NT_022504S1.3	/contig_orient=forward /start=1 /end=271641 /chrom=3	102399	56001	1.3E-35	6420	5420	86	86	7428	1	176629	183705	-20	-5
N4	7428	NT_023397S2.3	/contig_orient=complement /start=250002 /end=455242	102399	49492	4.1E-31	6275	6275	79	84	1	7428	25289	33146	-20	-5
N4	7428	NT_011520S13.5	/contig_orient=forward /start=3068083 /end=3318083	102399	47530	9.6E-30	6166	6166	77	83	1	7425	146733	154470	-20	-5
N4	7428	NT_019483S4.3	/contig_orient=forward /start=750003 /end=1000003 /chrom=8	102399	50179	1.4E-31	6184	6184	82	83	7428	1	13951	21821	-20	-5
N4	7428	NT_019483S2.3	/contig_orient=forward /start=250001 /end=500001 /chrom=8	102399	50122	1.5E-31	6177	6177	82	83	7428	1	131375	138737	-20	-5
N4	7428	NT_024033S5.3	/contig_orient=forward /start=1000005 /end=1250005	102399	57370	1.4E-36	5859	5859	97	78	1	5981	41571	47546	-20	-5
N4	7428	NT_023628S1.3	/contig_orient=complement /start=1 /end=151365 /chrom=7	102399	56440	6.2E-36	5651	5651	99	76	1772	7428	1	5656	-20	-5
N4	7428	NT_023323S1.2	/contig_orient=none /start=1 /end=103061 /chrom=5 Homo	102399	45124	4.5E-28	5600	5600	82	75	6704	1	1	6758	-20	-5

REFERENCES (the contents of which are hereby incorporated in full by reference)

[0371]

1 Mayer et al. (1999) Nat. Genet. 21 (3), 257-258 (1999)
2 Farrell (1998) RNA Methodologies (Academic Press; ISBN 0-12-249695-7).
3 Yang et al. (1999) Proc Natl Acad Sci USA 96(23):13404-8
4 Robbins et al. (1997) Clin Lab Sci 10(5):265-71.
5 Ylikoski et al. (1999) Clin Chem 45(9):1397-407
6 Ylikoski et al. (2001) Biotechniques 30:832-840
7 Shirahata & Pegg (1986) J Biol. Chem. 261(29):13833-7.
8 Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual. NY, Cold Spring Harbor Laboratory
9 Short protocols in molecular biology (4th edition, 1999) Ausubel et al. eds. ISBN 0-471-32938-X.
10 US patent 5,707,829
11 Current Protocols in Molecular Biology (F.M. Ausubel et al., eds., 1987) Supplement 30.
12 EP-B-0509612
13 EP-B-0505012
14 Berkhout et al. (1999) J. Virol. 73:2365-2375.
15 Löwer et al. (1995) J. Virol. 69:141-149.
16 Magin et al. (1999) J. Virol. 73:9496-9507.
17 Magin-Lachmann (2001) J Virol. 75(21):10359-71.
18 Hashido et al. (1992) Biochem. Biophys. Res. Comm. 187:1241-1248.
19 Geysen et al. (1984) PNAS USA 81:3998-4002.
20 Carter (1994) Methods Mol Biol 36:207-23.
21 Jameson, BA et al., 1988, CABIOS 4(1):181-186.
22 Raddrizzani & Hammer (2000) Brief Bioinform 1(2):179-89.
23 De Lalla et al. (1999) J. Immunol. 163:1725-29.
24 Brusic et al. (1998) Bioinformatics 14(2):121-30
25 Meister et al. (1995) Vaccine 13(6):581-91.
26 Roberts et al. (1996) AIDS Res Hum Retroviruses 12(7):593-610.
27 Maksyutov & Zagrebelnaya (1993) Comput Appl Biosci 9(3):291-7.
28 Feller & de la Cruz (1991) Nature 349(6311):720-1.
29 Hopp (1993) Peptide Research 6:183-190.
30 Welling et al. (1985) FEBS Lett. 188:215-218.
31 Davenport et al. (1995) Immunogenetics 42:392-297.
32 Smith and Waterman, Adv. Appl. Math. (1981) 2: 482-489.
33 Go et al, Int. J. Peptide Protein Res. (1980) 15:211
34 Querol et al., Prot. Eng. (1996) 9:265
35 Olsen and Thomsen, J. Gen. Microbiol. (1991) 137:579
36 Clarke et al., Biochemistry (1993) 32:4322
37 Wakarchuk et al., Protein Eng. (1994) 7:1379
38 Toma et al., Biochemistry (1991) 30:97
39 Haezerbrouck et al., Protein Eng. (1993) 6:643
40 Masul et al., Appl. Env. Microbiol. (1994) 60:3579
41 US patent 4,959,314
42 Breedveld (2000) Lancet 355(9205):735-740.
43 Gorman & Clark (1990) Semin. Immunol. 2:457-466
44 Jones et al., Nature 321:522-525 (1986)
45 Morrison et al., Proc. Natl. Acad. Sci, US.A., 81:6851-6855 (1984)
46 Morrison and Oi, Adv. Immunol., 44:65-92 (1988)
47 Verhoeyer et al., Science 239:1534-1536 (1988)
48 Padlan, Molec. Immun. 28:489-498 (1991)
49 Padlan, Molec. Immunol. 31(3):169-217 (1994).
50 Kettleborough, C.A. et al., Protein Eng. 4(7):773-83 (1991).
51 Chothia et al., J. Mol. Biol. 196:901-917 (1987)
52 Kabat et al., U.S. Dept. of Health and Human Services NIH Publication No. 91-3242 (1991)
53 US patent 5,530,101.
54 US patent 5,585,089.

55 WO 98/24893
56 WO 91/10741
57 WO 96/30498
58 WO 94/02602
59 US Patent 5,939,598.
60 WO 96/33735
61 WO 93/14778
62 Findeis et al., Trends Biotechnol. (1993) 11:202
63 Chiou et al. (1994) Gene Therapeutics: Methods And Applications Of Direct Gene Transfer. ed. Wolff
64 Wu et al., J. Biol. Chem. (1988) 263:621
65 Wu et al., J. Biol. Chem. (1994) 269:542
66 Zenke et al., Proc. Natl. Acad. Sci. (USA) (1990) 87:3655
67 Wu et al., J. Biol. Chem. (1991) 266:338
68 Jolly, Cancer Gene Therapy (1994) 1:51
69 Kimura, Human Gene Therapy (1994) 5:845
70 Connelly, Human Gene Therapy (1995) 1:185
71 Kaplitt, Nature Genetics (1994) 6:148
72 WO 90/07936
73 WO 94/03622
74 WO 93/25698
75 WO 93/25234
76 US patent 5,219,740
77 WO 93/11230
78 WO 93/10218
79 US patent 4,777,127
80 GB Patent No. 2,200,651
81 EP-A- 0 345 242
82 WO 91/02805
83 WO 94/12649
84 WO 93/03769
85 WO 93/19191
86 WO 94/28938
87 WO 95/11984
88 WO 95/00655
89 Curiel, Hum. Gene Ther. (1992) 3:147
90 Wu, J. Biol. Chem. (1989) 264:16985
91 US patent 5,814,482
92 WO 95/07994
93 WO 96/17072
94 WO 95/30763
95 WO 97/42338
96 WO 90/11092
97 US patent 5,580,859
98 US patent 5,422,120
99 WO 95/13796
100 WO 94/23697
101 WO 91/14445
102 EP 0524968
103 Philip, Mol. Cell Biol. (1994) 14:2411
104 Woffendin, Proc. Natl. Acad. Sci. (1994) 91:11581
105 US patent 5,206,152
106 WO 92/11033
107 US patent 5,149,655
108 US patent 5,206,152
109 WO 92/11033
110 WO90/14837
111 Vaccine Design - the subunit and adjuvant approach (1995) ed. Powell & Newman
112 WO00/07621

113 GB-2220221
114 EP-A-0689454
115 EP-A-0835318
116 EP-A-0735898
117 EP-A-0761231
118 WO99/52549
119 WO01/21207
120 WO01/21152
121 WO00/62800
122 WO00/23105
123 WO99/11241
124 WO98/57659
125 WO93/13202.
126 McSharry (1999) Antiviral Res 43(1):1-21.
127 Kuhelj et al. (2001) J Biol Chem 276(20):16674-82.
128 Schommer et al. (1996) J Gen Virol 77:375-379.
129 Magin et al. (2000) Virology 274:11-16.
130 Boese et al. (2001) FEBS Lett 493(2-3):117-21.
131 Larsson, E., et al., Current Topics in Microbiology and Immunology 148:115 (1989)
132 Mariani-Costantini, et al., J. Virol. 63:4982 (1989) and Shih, et al., Virology 182:495 (1991)
133 Tönjes et al. (1996) J. AIDS Hum. Retrovir. 13(Suppl 1):S261-S267.
134 Barbulescu et al., Curr. Biol. 9:861 (1999)
135 Ono, et al., J. Virol. 58:937 (1986)
136 Löwer et al., Proc. Natl. Acad. Sci USA 90:4480 (1993)
137 Ono et al., (1986) J. Virol. 60:589
138 Boller, et al., Virol. 196:349 (1993)
139 Yang et al., Proc. Natl. Acad. Sci USA 96:13404 (1999)
140 Mueller- Lantzsch et al., AIDS Research and Human Retroviruses 9:343-350 (1993)
141 Herbst et al., Amer. J. Pathol. 149:1727 (1996)
142 US patent 5,858,723
143 Löwer et al., Proc. Natl. Acad. Sci USA 93:5177 (1996)
144 Löwer et al., Virology 192:501 (1993)
145 Genbank accession number AB047240
146 Andersson et al. (1999) J. Gen. Virol. 80:255-260.
147 Zsíros et al. (1998) J. Gen. Virol. 79:61-70.
148 Tönjes et al. (1999) J. Virol. 73:9187-9195.
149 Johnston et al. (2001) Ann Neurol 50(4):434-42.
150 Medstrand et al. (1998) J Virol 72(12):9782-7.
151 US patent 5,010,175
152 International patent application WO 91/17823.
153 US Patent 4,816,567.
154 Merrifeld, J. Am. Chem. Soc. 85:2149, 1963
155 Caprino and Han, J. Org. Chem. 37:3404, 1972
156 Milstein and Kohler, Nature 256:495-497, 1975
157 Gulfre and Milstein, *Methods in Enzymology: Immunochemical Techniques 73*:1-46
158 Langone and Banatis eds., Academic Press, 1981
159 Altschul et al. Nucleic Acids Res.(1997) 25:3389-3402
160 Brutlag et al. Comp. Chem.(1993) 17:203
161 Schena et al. (1996) Proc Natl Acad Sci U S A. 93(20):10614-9
162 Schena et al. (1995) Science 270(5235):467-70
163 Shalon et al. (1996) Genome Res. 6(7):639-45
164 US patent 5,807,522
165 European patent application 0799897
166 WO 97/29212
167 WO 97/27317
168 European patent application 0785280
169 WO 97/02357
170 US patent 5,593,839

171 US patent 5,578,832
172 European patent application 0728520
173 US patent 5,599,695
174 European patent application 0721016.
175 US patent 5,556,752
176 WO 95/22058
177 US patent 5,631,734
178 Pappalarado et al., Sem. Radiation Oncol. (1998) 8:217
179 Ramsay Nature Biotechnol. (1998) 16:40
180 US patent 5,134,854
181 US patent 5,445,934
182 WO 95/35505
183 US patent 5,631,734
184 US patent 5,800,992
185 WO92/02526.
186 US patent 5,124,246.
187 Mullis et al., Meth. Enzymol. (1987) 155:335
188 US patent 4,683,195
189 US patent 4,683,202
190 Saiki et al. (1985) Science 239:487
191 Hanahan et al. Cell 100:57-70 (2000)
192 Weissman SM Mol Biol. Med. 4(3),133-143 (1987
193 Patanjali, et al. Proc. Natl. Acad. Sci. USA 88 (1991)
194 Simone et al. Am J Pathol. 156(2):445-52 (2000)
195 Claverie (1996) Meth. Enzymol. 266:212-227.
196 Automated DNA Sequencing and Analysis Techniques Adams et al., eds., Chap. 36, p. 267 Academic Press, San Diego, 1994
197 Claverie et al. Comput. Chem. (1993) 17:191
198 Altschul et. al, J. Mol. Biol., 215:403-410, 1990
199 Pearson & Lipman, PNAS, 85:2444, 1988
200 Luo et al. (1999) Nature Med 5:117-122
201 Higgins & Sharp CABIOS 5; 151-153 (1989)
202 Delli Bovi et al. (1986, Cancer Res. 46:6333-6338)
203 Cesarone, C. et al., Anal Biochem 100:188-197 (1979)
204 Southern, E. M., J. Mol. Biol. 95:503-517 (1975)
205 Feinberg, A. P., et al., 1983, Anal. Biochem. 132:6-13
206 Wright and Manos (1990, in "PCR Protocols", Innis et al., eds., Academic Press, pp. 153-158)
207 Keown et al., Methods in Enzymology 185:527-537 (1990)
208 Marks, et al., Brit. J. Urol. 75:225 (1995)
209 Skea, et al., J. Immunol. 151:3557 (1993)
210 Mather, et al., J. Nucl. Med. 31:692 (1990)
211 Zhang et al., Nucl. Med. Biol. 19:607 (1992)

SEQUENCE LISTING

SEQ ID 1:
CTTTGTCTCTGTGTCTTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCC
ACCCCTACA
SEQ ID 2:
TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTAT
GTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTG
CTGTGTCCACTCAGGGTTAAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTC
CTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAG
GAAAGCCAGGTATTGTCCAACGTTTCTCCCCATGTGATAGCCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGT
CCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA
AGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACT
SEQ ID 3:
GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTA
TGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCACATGTTTGTCT
GCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCACAGATGATCAGTAAATA
CTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTCTTTCTTTCTCTATA
SEQ ID 4:
GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTA
TGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCACATGTTTGTCT
GCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCACAGATGATCAGTAAATA
CTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTCTTTCTTTCTCTATAC
TTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA
CCCCTACA
SEQ ID 5:
TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTAT
GTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTG
CTGTGTCCACTCAGGGTTAAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTC
CTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAG
GAAAGCCAGGTATTGTCCAACGTTTCTCCCCATGTGATAGCCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGT
CCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA
AGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACT
GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTA
TGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCACATGTTTGTCT
GCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCACAGATGATCAGTAAATA
CTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTCTTTCTTTCTCTATA
SEQ ID 6:
TCTGGTGCCCAACGTGGAGGCTTTTCTCTAGGGTGAAGGTACGCTCGAGCGTGGTCATTGAGGACAAGTCGACGAGAGAT
CCCGAGTACATCTACAGTCAGCCTTACG
SEQ ID 7:
GGGAAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATA
AAGGAAGGAGTTAAACAATATGGRTCCAACTCCCCTTATATAAGAACAKTATTAGATTCCATTGCYCATGGAAATAGACT
TACTCCTTATGACTGGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTG
ATGGRGTACARGAACAGGTACGAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACA
GGTCCAAATTGGAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAG
GGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTG
ACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATG
GCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAAAAGGAAAAGTTCCAGCAGGAGTTGATGTAAT
TACMGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATGCTAATGGCTCAAGCAATGAGGGGGC
TCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATCGGTCATCKGAAAAGGAGTTGC
CCAGGCTTAAAYAARCAGAATATAATAAATCAAGCTATTACAGCAAAAAATAAAAAGCCATCTGGCCTGTGTCCAAAATG
TGGAAAAGCAAAACATTGGGCCAATCAATGTCATTCTAAATTTGATAAAGATGGGCAACCATTGTCTGGAAACAGGAAGA
GGGGCCAGCCTCAGGCCCCCCAACAAACTGGGGCATTCCCAGTTAAACTGTTTGTTCCTCAGGGTTTTCAAGGACAACAA
CCCCTACAGAAAATACCACCACTTCAGGGAGTCAGCCAATTACAACAATCCAACAGCTGTCCCGCGCCACAGCAGGCAGC
ACCGCAGTAGATTTATGTTCCACCCAAATGGTCTTTTTACTCCCTGGAAAGCCCCCACAAAAGATTCCTAGAGGGGTATA
TGGCCCGCTGCCAGAAGGGAGGGTAGGCCTTTGAGGGAGATCAAGTCTAAATTTGAAGGGAGTCCAAATTCATACTGGGG
TAATTTACTCAGATTATAAAGGGGGAATTCAGTTAGTGATCAGCTCCACTGTTCCCCGGAGTGCCAATCCAGGTGATAGA
ATTGCTCAATTACTGCTTTTGCCTTATGCA
SEQ ID 8:
ACAACAATGGCATGCAGAGATTACTATCCCAGCCTCCCTATACAGCCCCAGGAATCAAAAAATCATGACTAAAATGGGAT
AGCTCCCTAAAAAGGGACTAGGAAAGAAAGAAGTCCCAATTGAGGCTGAAAAAAATYAAAAAAGAAAAGGAATAGGGCAT
CCTTTTTAGGAGCGGTCACTGTAGAGCCTCCAAAACCCATTCCATTAACTTGGGAAAAAAAMAACTGTATGGTAAATCAG
CAGCCGCTTCCAAAACAAAARCTGGAGGCYTTACAYTTATTAGCAAAGAAACMATTAGAAAAAGGACATTGAGCCTTCAT
TTTCGCCTTGGAATTCTGTTTGTRATTCAGAAAAAATCCGGCAGATGGCGTATGCTAACTGAGCCATTAATGCCGTAATT
CAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCT
TTTTTACCATTCCTCTGGCAAAACAGGATTTTGARAAATTTGCTTTYACCACACCAGCCTAAATAATAAAGAACCAGCCA
CCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAACTATTTGTCAGCTCAAGCTCTGCAACCAGTTA
GAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGAC

CGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGGGRCTGACAATAACATCTGATAAGATTCAARCCTCTACTCCTTT
CCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCMCAAAAAAATAGAAATAAGAAAAGACACATTAAAAGCA
TTAAATGAGTTTCAAAAGTTGCTAGGAGATACTAATTGGATTTGGAGATATTAATTGGATTTGGCCAACTCTAGGCATTC
CTACTTATGCCATGTCAAATTTGTWCTCTTTCTTAAGAGGGGACTCGGAATTAAATAGTGAAAGAACGTTAACTCCAGAG
GCAACTAAAGAAATTAAATTAATTGAAGAAAAAATTCGGTCAGCACAAGTAAATAGAATAGATCACTTGGCCCCACTCCA
AATTTTGATTTTTACTACTGCACATTCCCTAACAGGCATCATTGTTCAAAACACAGATCTTGTGGAGTGGTCCTTCCTTC
CTCACAGTACAATTAAGACTTTTACATTGTACTTGGATCAAATGGCTACATTAATTGGTCAGGGAAGATTATGAATAATA
ACATTGTGTGGAAATGACCCAGATAAAATCACTGTTCCTTTCAACAAGCAACAGGTTAGACAAGCCTTTATCAATTCTGG
TGCATGGCAGATTGGTCTTGCCGATTTTGTGGGAATTATTGACAATCGTTACCACA

SEQ ID 9:

ACAACAATGGCATGCAGAGATTACTATCCCAGCCTCCCTATACAGCCCCAGGAATCAAAAAATCATGACTAAAATGGGAT
AGCTCCCTAAAAAGGGACTAGGAAAGAAAGAAGTCCCAATTGAGGCTGAAAAAAATYAAAAAAGAAAAGGAATAGGGCAT
CCTTTTTAGGAGCGGTCACTGTAGAGCCTCCAAAACCCATTCCATTAACTTGGGGGAAAAAAAAAAMAACTGTATGGTAAA
TCAGCAGCCGCTTCCAAAACAAAARCTGGAGGCYTTACAYTTATTAGCAAAGAAACMATTAGAAAAAGGACATTGAGCCT
TCATTTTCGCCTTGGAATTCTGTTTGTRATTCAGAAAAAATCCGGCAGATGGCGTATGCTAACTGAGCCATTAATGCCGT
AATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGAT
TGCTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGARAAATTTGCTTTYACCACACCAGCCTAAATAATAAAGAACCA
GCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAACTATTTGTCAGCTCAAGCTCTGCAACCA
GTTAGAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAAT
TGACCGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGGGRCTGACAATAACATCTGATAAGATTCAARCCTCTACTC
CTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCMCAAAAAAATAGAAATAAGAAAAGACACATTAAA
AGCATTAAATGAGTTTCAAAAGTTGCTAGGAGATACTAATTGGATTTGGAGATATTAATTGGATTTGGCCAACTCTAGGC
ATTCCTACTTATGCCATGTCAAATTTGTWCTCTTTCTTAAGAGGGGACTCGGAATTAAATAGTGAAAGAACGTTAACTCC
AGAGGCAACTAAAGAAATTAAATTAATTGAAGAAAAAATTCGGTCAGCACAAGTAAATAGAATAGATCACTTGGCCCCAC
TCCAAATTTTGATTTTTACTACTGCACATTCCCTAACAGGCATCATTGTTCAAAACACAGATCTTGTGGAGTGGTCCTTC
CTTCCTCACAGTACAATTAAGACTTTTACATTGTACTTGGATCAAATGGCTACATTAATTGGTCAGGGAAGATTATGAAT
AATAACATTGTGTGGAAATGACCCAGATAAAATCACTGTTCCTTTCAACAAGCAACAGGTTAGACAAGCCTTTATCAATT
CTGGTGCATGGCAGATTGGTCTTGCCGATTTTGTGGGAATTATTGACAATCGTTACCACA

SEQ ID 10:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTYGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAG
ATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATC
AAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAWGAAATTTCCCATTTTA
TATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCAT
CTGCATTCATKGARGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATARATTTGATATCACA
TGGAAACAGACAAAAAATATTGTACAACATTGCRCCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGYAAGAGTTAA
TCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCYTCATTTGGAAAATTGTCATTTGTCC
AYGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACTTCCCATGTYAAGAGACAT
TTATTATYTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAARACAGACAATGGGCCAGGTTACTGTAGTAAAGCAGT
TCAARAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGGACAGGCCATAATTGAAA
GAACTAATAGAACACTCAAAGCTCAATTGGTTAAACA

SEQ ID 11:

GGGAAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATA
AAGGAAGGAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACAGTATTAGATTCCATTGCTCATGGAAATAGACT
TACTCCTTATGACTGGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTG
ATGGAGTACAAGAACAGGTACGnAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAAC
AGGTCCAAATTGGAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCA
GGGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCT
GACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAAT
GGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAAAAGGAAAAGTTCCAGCAGGAGTTGATGTAA
TTACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATGCTAATGGCTCAAGCAATGAGGGGG
CTCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATCGGTCATCTGAAAAGGAGTTG
CCCAGGCTTAAATAAACAGAATATAATAAATCAAGCTATTA

SEQ ID 12:

NCTGAAAAAAATNAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGGAGCGGTCACTGTAGAGCCTCCAAAACCCATTCCA
TTAACTTGGGNAAAAAAANAACTGTATGGTAAATCAGCAGNCGCTTCCAAAACAAAANCTGGAGGCNTTACANTTATTAG
CAAAGAAACCATTAGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTAATTCAGAAAAAATCCGGCA
GATGGCGTATGCTAACTGAGCCATTAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGCCA
TGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGAAAAATTTGC
TTTTACCACACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAG
TTCAACTATTTGTCAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATAT
TTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGACCGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGGGACTGA
CAATAACATCTGATAAGATTCAAACCTCTACTCCTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCA
CAAAAAAATAGAAATAAGAAAAGACACATTAAAAGCATTAAATGAGTTTCAAAAGTTGCTAGGAGATACTAATTGGATTT
GGAGATATTAATTGGATTTGGCCAACTCTAGGCATTCCTACTTATGCCATGTCAAATTTGTNCTCTTTCTTAAGAGGGGA
CTCGGAATTAAATAGTGAAAGAACGTTAACTCCAGAGGCAACTA

SEQ ID 13:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAG
ATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATC
AAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTA

TATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCAT
CTGCATTCATGGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACA
TGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAA
TCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCC
ATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACTTCCCATGTTAAGAGACAT
TTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGGGCCAGGTTACTGTAGTAAAGCAGT
TCAAAAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGGACAGGCCATAATTGAAA
GAACTAATAGAACACTCAAAGCTCAATTGGTTA

SEQ ID 14:

GGGAAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATA
AAGGAAGGAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACAGTATTAGATTCCATTGCTCATGGAAATAGACT
TACTCCTTATGACTGGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTG
ATGGAGTACAGGAACAGGTACGAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACA
GGTCCAAATTGGAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAG
GGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTG
ACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATG
GCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAAAAGGAAAAGTTCCAGCAGGAGTTGATGTAAT
TACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATGCTAATGGCTCAAGCAATGAGGGGGC
TCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATCGGTCATCTGAAAAGGAGTTGC
CCAGGCTTAAATAAACAGAATATAATAAATCAAGCTATTACAGAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 15:

GGGAAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATA
AAGGAAGGAGTTAAACAATATGGGTCCAACTCCCCTTATATAAGAACATTATTAGATTCCATTGCTCATGGAAATAGACT
TACTCCTTATGACTGGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTG
ATGGAGTACAAGAACAGGTACGAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACA
GGTCCAAATTGGAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAG
GGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTG
ACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATG
GCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAAAAGGAAAAGTTCCAGCAGGAGTTGATGTAAT
TACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATGCTAATGGCTCAAGCAATGAGGGGGC
TCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATCGGTCATCGGAAAAGGAGTTGC
CCAGGCTTAAATAAACAGAATATAATAAATCAAGCTATTACAGCAAAAAATAAAAAGCCATCTGGCCTGTGTCCAAAATG
TGGAAAAGCAAAACATTGGGCCAATCAATGTCATTCTAAATTTGATAAAGATGGGCAACCATTGTCTGGAAACAGGAAGA
GGGGCCAGCCTCAGGCCCCCCAACAAACTGGGGCATTCCCAGTTAAACTGTTTGTTCCTCAGGGTTTTCAAGGACAACAA
CCCCTACAGAAAATACCACCACTTCAGGGAGTCAGCCAATTACAACAATCCAACAGCTGTCCCGCGCCACAGCAGGCAGC
ACCGCAGTAGATTTATGTTCCACCCAAATGGTCTTTTTACTCCCTGGAAAGCCCCCACAAAAGATTCCTAGAGGGGTATA
TGGCCCGCTGCCAGAAGGGAGGGTAGGCCTTTGAGGGAGATCAAGTCTAAATTTGAAGGGAGTCCAAATTCATACTGGGG
TAATTTACTCAGATTATAAAGGGGGAATTCAGTTAGTGATCAGCTCCACTGTTCCCCGGAGTGCCAATCCAGGTGATAGA
ATTGCTCAATTACTGCTTTTGCCTTATGCAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 16:

AAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATAAAG
GAAGGAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACAGTATTAGATTCCATTGCCCATGGAAATAGACTTAC
TCCTTATGACTGGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTGATG
GGGTACAAGAACAGGTACGAAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACAGG
TCCAAATTGGAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGG
CCTGGGGAAAAATTCAGGACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGAC
TTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGC
CTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAAAAGGAAAAGTTCCAGCAGGAGTTGATGTAATTA
CCGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATGCTAATGGCTCAAGCAATGAGGGGGCTC
ACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATCGGTCATCTGAAAAGGAGTTGCCC
AGGCTTAAACAAGCAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 17:

ACAACAATGGCATGCAGAGATTACTATCCCAGCCTCCCTATACAGCCCCAGGAATCAAAAAATCATGACTAAAATGGGAT
AGCTCCCTAAAAAGGGACTAGGAAAGAAAGAAGTCCCAATTGAGGCTGAAAAAAATTAAAAAAGAAAAGGAATAGGGCAT
CCTTTTTAGGAGCGGTCACTGTAGAGCCTCCAAAACCCATTCCATTAACTTGGGAAAAAAAAAAACTGTATGGTAAATCAG
CAGCCGCTTCCAAAACAAAAGCTGGAGGCCTTACACTTATTAGCAAAGAAACCATTAGAAAAAGGACATTGAGCCTTCAT
TTTCGCCTTGGAATTCTGTTTGTGATTCAGAAAAAATCCGGCAGATGGCGTATGCTAACTGAGCCATTAATGCCGTAATT
CAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCT
TTTTTACCATTCCTCTGGCAAAACAGGATTTTGAAAAATTTGCTTTTACCACACCAGCCTAAATAATAAAGAACCAGCCA
CCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAACTATTTGTCAGCTCAAGCTCTGCAACCAGTTA
GAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGAC
CGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGGGGCTGACAATAACATCTGATAAGATTCAAACCTCTACTCCTTT
CCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCCCAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 18:

CTGAAAAAAATCAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGGAGCGGTCACTGTAGAGCCTCCAAAACCCATTCCAT
TAACTTGGGGGAAAAAAAAACAACTGTATGGTAAATCAGCAGCGCTTCCAAAACAAAAACTGGAGGCTTTACATTTATTA
GCAAAGAAACAATTAGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTAATTCAGAAAAAATCCGGC
AGATGGCGTATAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTAAT
TATAATTGATCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGAGAAATTTGCTTTTACCACACCAG

CCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAACTATTTGTC
AGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGCTGCAG
AAACGAGAGACAAATTAATTGACCGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGGGACTGACAATAACATCTGAT
AAGATTCAAACCTCTACTCCTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCACAAAAAAAAAAAAA
AAAAAAAAAAAAA
SEQ ID 19:
CATTAGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTAATTCAGAAAAAATCCGGCAGATGGCGTA
TGCTAACTGAGCCATTAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGCCATGGTCCCCT
TTAATTATAATTGATCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGAAAAATTTGCTTTTACCAC
ACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAACTAT
TTGTCAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGC
TGCAGAAACGAGAGACAAATTAATTGACCGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGGGACTGACAATAACAT
CTGATAAGATTCAAACCTCTACTCCTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCACAAAAAATA
GAAATAAGAAAAGACACATTAAAAGCATTAAATGAGTTTCAAAAGTTGCTAGGAGATACTAATTGGATTTGGAGATATTA
ATTGGATTTGGCCAACTCTAGGCATTCCTACTTATGCCATGTCAAATTTGTACTCTTTCTTAAGAGGGGACTCGGAATTA
AATAGTGAAAGAACGTTAACTCCAGAGGCAACTAAAGAAAAAAAAAAAAAAAAAAAAAAAAAA
SEQ ID 20:
ATCTTTACCCTGTATAAACATCTTTCTCTTCCCAGTATTTCTAAGCATGTGACAATGAATATGCAAAGGAAGCGCAGCAG
TCCACCAGGTGTGGGATATGTGTGGCACAATTCAAGACAATGATTAAACCTCCACTTGATGTTGCAAAAGAGATTTTGAA
AAATTTGCTTTCACCACACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATG
CTTAATAGTTCAACTATTTGTCAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTACATCGTTCACTAT
GTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGACCGTTACACATTTCTGCAGACAGAGGTTGCCAACGC
GGGACTGACAATAACATCTGATAAGATTCAAGCCTCTACTCCTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAA
TTAAACCACAAAAAAATAGAAATAAGAAAAGACACATTAAAAGCATTAAATGAGTTTCAAAAGTTGCTAGGAGATACTAA
TTGGATTTGGAGATATTAATTGGATTTGGCCAACTCTAGGCATTCCTACTTATGCCATGTCAAATTTGTTCTCTTTCTTA
AGAGGGGACTCGGAATTAAATAGTGAAAGAACGTTAACTCCAGAGGCAACTAAAGAAATTAAATTAATTGAAGAAAAAAT
TCGGTCAGCACAAGTAAATAGAATAGATCACTTGGCCCCACTCCAAATTTTGATTTTTACTACTGCACATTCCCTAACAG
GCATCATTGTTCAAAACACAGATCTTGTGGAGTGGTCCTTCCTTCCTCACAGTACAATTAAGACTTTTACATTGTACTTG
GATCAAATGGCTACATTAATTGGTCAGGGAAGATTATGAATAATAACATTGTGTGGAAATGACCCAGATAAAATCACTGT
TCCTTTCAACAAGCAACAGGTTAGACAAGCCTTTATCAATTCTGGTGCATGGCAGATTGGTCTTGCCGATTTTGTGGGAA
TTATTGACAATCGTTACCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
SEQ ID 21:
CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAG
ATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATC
AAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTA
TATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCAT
CTGCATTCATGGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACA
TGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAA
TCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCC
ATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACTTCCCATGTCAAGAGACAT
TTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAAAAAAAAAAAAAAAAAAAAAAAA
SEQ ID 22:
CCAAAAGAATGAGTCATCAAAACTCAGTATCACTCGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAG
ATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATC
AAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTA
TATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCAT
CTGCATTCATGGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACA
TGGAAACAGACAAAAAATATTGTACAACATTGCGCCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGTAAGAGTTAA
TCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCCTCATTTGGAAAATTGTCATTTGTCC
ATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACTTCCCATGTTAAGAGACAT
TTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGGGCCAGGTTACTGTAGTAAAGCAGT
TCAAAAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGGACAGGCCATAATTGAAA
GAACTAATAGAACACTCAAAGCTCAATTGGTTAAACAAAAAAAAAAAAAAAAAAAAAAAAAA
SEQ ID 23:
CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAG
ATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATC
AAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTA
TATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCAT
CTGCATTCATGGAGGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAGATTTGATATCACA
TGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAA
TCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCC
ATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACTTCCCATGTTAAGAGACAT
TTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGGGCCAGGTTACTGTAGTAAAGCAGT
TCAAAAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGGACAGGCCATAATTGAAA
GAACTAATAGAACACTCAAAGCTCAATTGGTTAAACAAAAAAAAAAAAAAAAAAAAAAAAAA
SEQ ID 24:
CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAG
ATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATC
AAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAATGAAATTTCCCATTTTA

TATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCAT
CTGCATTCATGGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACA
TGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAA
TCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCC
ATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACTTCCCATGTTAAGAGACAT
TTATTATTTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGGGCCAGGTTACTGTAGTAAAGCAGT
TCAAGAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGGACAGGCCATAATTGAAA
GAACTAATAGAACACTCAAAGCTCAATTGGTTAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 25:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAG
ATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATC
AAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTA
TATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCAT
CTGCATTCATTGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACA
TGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAA
TCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCC
ACGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACTTCCCATGTTAAGAGACAT
TTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAGACAGACAATGGGCCAGGTTACTGTAGTAAAGCAGT
TCAAAAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGGACAGGCCATAATTGAAA
GAACTAATAGAACACTCAAAGCTCAATTGGTTAAGCAAAAAAAAAAAAAAAAAAAAAAAAAACATGTCGGCCGCCTCGG
CC

SEQ ID 26:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAG
ATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATC
AAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTA
TATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCAT
CTGCATTCATGGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACA
TGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAA
TCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCC
ATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACTTCCCATGTTAAGAGACAT
TTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGGGCCAGGTTACTGTAGTAAAGCAGT
TCAAAAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGGACAGGCCATAATTGAAA
GAACTAATAGAACACTCAAAGCTCAATTGGTTAAACAAAAAGAAAAAAAAAAAAAAAAA

SEQ ID 27:

ACCGGCCTTACGGCCGGGGAAGAGNTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAA
ATGTTAAAAGATATAAAGGAAGGAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACAGTATTAGATTCCATTGC
TCATGGAAATAGACTTACTCCTTATGACTGGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTA
AAACCTGGTGGATTGATGGAGTACAGGAACAGGTACGAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGAC
CAATTGTTAGGAACAGGTCCAAATTGGAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAG
GGCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTA
AAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTT
ATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAAAAGGAAAAGTTCCAGC
AGGAGTTGATGTAATTACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCNTAAGGCAATGCTAATGGCTC
AAGCAATGAGGGGGCTCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTT

SEQ ID 28:

TTACGGCCTTACGGCCGGGGAAGNNTNTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGA
AAATGTTAAAAGATATAAAGGAAGGAGTTAAACAATATGGGTCCAACTCCCCTTATATAAGAACATTATTAGATTCCATT
GCTCATGGAAATAGACTTACTCCTTATGACTGGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTT
TAAAACCTGGTGGATTGATGGAGTACAAGAACAGGTACGAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAG
ACCAATTGTTAGGAACAGGTCCAAATTGGAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTA
AGGGCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTC
TAAAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTACAGATGACAATGCCCGAAAAG
TTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAAAAGGAAAAGTTCCA
GCAGGAGTTGATGTAATTACAGAATATN

SEQ ID 29:

CGGCCTTACGGCCCGGGGAGANNTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAAT
GTTAAAAGATATAAAGGAAGGAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACAGTATTAGATTCCATTGCCC
ATGGAAATAGACTTACTCCTTATGACTGGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAA
ACCTGGTGGATTGATGGGGTACAAGAACAGGTACGAAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACC
AATTGTTAGGAACAGGTCCAAATTGGAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGG
GCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTAA
AGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTTA
TTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAAAAGGAAAAGTTCCAGCA
GGAGTTGATGTAATTACCG

SEQ ID 30:

NCCGGCCTTACGGCCGGGNCAACAATGGCATGCAGAGNTTACTATCCCAGCCTCCCTATACAGCCCCAGGAATCAAAAAA
TCATGACTAAAATGGGATAGCTCCCTAAAAAGGGACTAGGAAAGAAAGAAGTCCCAATTGAGGCTGAAAAAAATTAAAAA
AGAAAAGGAATAGGGCATCCTTTTTAGGAGCGGTCACTGTAGAGCCTCCAAAACCCATTCCATTAACTTGGGAAAAAAAA
AACTGTNTGGTAAATCAGCAGCCGNTTCCAAAACAAAAGCTGGAGGCCTTACACTTATTANCAAAGAANCCATTANAAAA

AGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTGATTCAAAAAAAATCCGGCANATGGCGTATGCTAACTGA
NCCATTAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTNTCCAGCCATGGTCCCCTTTAATTATAA
TTGATCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGAAAAATTTGCTTTTACCACACCAGCCTAA
ATAATAAANAACCANCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAACTATTNGTCAGCTC
AAGCTCTGCAACCAGTTAGAGACN

SEQ ID 31:

NCCTGGCCTTACGGCCGGGGCTGAAAAAAATCAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGGAGCGGTCACTGTAGA
GCCTCCAAAACCCATTCCATTAACTTGGGGGAAAAAAAAACAACTGTATGGTAAATCAGCAGCGCTTCCAAAACAAAAAC
TGGAGGCTTTACATTTATTAGCAAAGAAACAATTAGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTG
TAATTCAGAAAAAATCCGGCAGATGGCGTATAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTC
CAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGAGAA
ATTTGCTTTTACCACACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCT
TAATAGTTCAACTATTTGTCAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTACATCGTTCACTATGT
TGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGACCGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGG
GACTGACAATAACATCTGATAAGATTCAAACCTCTACTCCTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATT
AAACCACAAAAAAAAAAAAAAN

SEQ ID 32:

NNNNNCNCGGGCATTAGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTAATTCAGAAAAAATCCGG
CAGATGGCGTATGCTAACTGAGCCATTAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGC
CATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGAAAAATTT
GCTTTTACCACACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCANGGAATGCTTAAT
AGTTCAACTATTTGTCAGCTCAAAGCTCTGCACCCAGNTAGAGACAAGTTTCAGACTGGTTCATCGTCCTATGTGATATT
TTGTGTGCTGCAGAACGAGAGACAAATTATTGGCCGTTCACATTTTTGCAGACAGAGGTTGCCAACGCGGGACTGACAAT
AACATCTGATAAGATTAAACCTCTACTCCTTCCGTACTTGGGAATGCAGGTGGAGGAAAGGAAAATTAACCCCCNNAAAA
TTGAATTANGAAAAGACCCNTTAAAGCCTTAAATGAGTTCAAAAAGTTGCTAGGAGAAACTAATTGGATTTGGAGANATT
AATTGGATTTGGCAACTNTAGGCATTCCTACTTATGCCN

SEQ ID 33:

TCCGGCCTTACGGCCGGGNTCTTTACCCTGTATAAACATCTTTCTCTTCCCAGTATTTCTAAGCATGTGACAATGAATAT
GCAAAGGAAGCGCAGCAGTCCACCAGGTGTGGGATATGTGTGGCACAATTCAAGACAATGATTAAACCTCCACTTGATGT
TGCAAAAGAGATTTTGAAAAATTTGCTTTCACCACACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAA
GTATTGCCTCAGGGAATGCTTAATAGTTCAACTATTTGTCAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGAC
TGTTACATCGTTCACTATGTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGACCGTTACACATTTCTGCA
GACAGAGGTTGCCAACGCGGGACTGACAATAACATCTGATAAGATTCAAGCCTCTACTCCTTTCCGTTACTTGGGAATGC
AGGTAGAGGAAAGGAAAATTAAACCACAAAAAAATAGAAATAAGAAAAGACACATTAAAAGCATTAAATGAGTTTCAAAA
GTTGCTAGGAGATACTAATTGGATTTGGAGATATTAATTGGATTTGGCCAACTCTAGGCATTCCTACTTATGCCATGTCA
AATTTGTTCTCTTTCT

SEQ ID 34:

TTNCGGCCTTACGGCCGGGCCAAGATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCA
TTACAGTGTTAACAAGATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATAT
TGAGAGAGCCCTAATCAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAA
GAAATTTCCCATTTTATATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCT
GACTTGCTAGTATCATCTGCATTCATGGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAA
TAAATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTC
AGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATTGCACGTACCTTCATTTGG
AAAATTGTCATTTGTCCATGTGACAGNTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACTT
NCCATGTCAAGAGACATTTATTATCTTGGTTTCCTGGNTGGGGAGNTCCCNNNNNNNANNNNNNNNAAAAAAAANANNC
NNN

SEQ ID 35:

TTACGGCCTTACGGCCGGGCCAAAGATGAGTCATCAAAACTCAGTATCACTCGACTCAAAGAGCAGAGTTGGTTGCCGTC
ATTACAGTGTTAACAAGATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATA
TTGAGAGAGCCCTAATCAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAA
AGAAATTTCCCATTTTATATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGC
TGACTTGCTAGTATCATCTGCATTCATGGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAA
ATAAATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCGCCCAGTGTCAGATTCTACACCTGGCCACT
CAGGAGGTAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCCTCATTTGG
AAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACTT
CCCATGTTAAGAGACA

SEQ ID 36:

ATTTGCCTTACGGCCGGGCCAAAAGTATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGT
CATTACAGTGTTAACAAGATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGAT
ATTGAGAGAGCCCTAATCAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAA
AAGAAATTTCCCATTTTATATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAG
CTGACTTGCTAGTATCATCTGCATTCATGGAGGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAA
AATAGATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCAC
TCAGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCTTCATTTG
GAAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACT
TCCCATGTTAAGAGACATTTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGGGCCANG
TTACTGTAGTAAAGCAGTTCAAAAATTCTTAAATCAGTGGAAAATTACACATN

SEQ ID 37:

CGGCCTTACGGCCGGGCCAAANATGAAGGGNNNAANGNCGGTTCCCAGGGACNNAGGCGCNTTNCATGGTTGCNGTNGTT
ACACCTGTTAACAAGATTNTAATCAGTCTATTAACATTGTATCAAATTCTGCATATGTAGNACAGGCTACAAAGGATATT
GAGAGAGCCCTAATCAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAATG
AAATTTCCCATTTTATATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTNACTAAAGCAAATGAACAAGCTG
ACTTGCTNGTATCATCTGCATTCATGGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAAT
AAATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCA
GGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAA
AATTGTCATTTGTCCATGTGACAGNTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACANGAGAAAGTNCTTCC
CATGTTAAGAGACATTTATTATTTTGNTNTCCTGNCATTGGGAGTTCCANAAAAAGTAAAACAGACANTGGGCCAGGTTA
C

SEQ ID 38:

TACGGCCTTACGGCCGGGCCAAGATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCNT
TACAGTGTTAACAAGATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATT
GAGAGAGCCCTAATCAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAG
AAATTTCCCATTTTATATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTG
ACTTGCTAGTATCATCTGCATTCATTGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCNATAGGATTAAAAAAT
AAATTTGATATCACCTGGAAACAGACAAAAAATATTGTACAACATTGCACCCNNNGTCAGATTCTACACCTGGCCNCTCN
NGAGGCAAGAGTTAATCCCNCANGGCTATGTCCTNATGTGTTATGGCAAANGGATGTNATGCNCCNNCCTTCCTTTNGAA
AANNNNNNTTTGTNCCCCNNACANNNGATACTTATTCACNNTTNNTATNGGNNACCCCCCCCACNNGANAAANAACCTNC
CCNNTNNANANAAANTNNTTATTTTTNTTTTN

SEQ ID 39:

NCCGGCCTTACGGCCGGGCCAAGATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCAT
TACAGTGTTAACAAGATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATT
GAGAGAGCCCTAATCAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAG
AAATTTCCCATTTTATATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTG
ACTTGCTAGTATCATCTGCATTCATGGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAAT
AAATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCA
GGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAA
AATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTACTTCC
CATGTTAAGAGACATTTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGGGCCAGGTTA
CTGGAGTAAAGCAGTTCAAAAATTCTTAAATCAGTGG

SEQ ID 40:

AAGGCAGTCAAGCAGGAGTTAAACAATATGGACCTAACTCTCCTTATATTAGAATATTATTAAATTCCATTGCTCATGGA
AATAGACTTATTTCTTATGATTGGGAAATTCTGGCTATATCTTCCCTTTCACCCTCTCAGTATCTCCAGTTTAAAACCTG
GTGGATTGATGGGGTACAAGAACAGGTACGAAAAAATCAGGCTACTAATCCTGTTGCTTATATAGATGAAGACCAATTGC
TAGGAAGAGGTCCAAACTGGGACACTATTAACCAACAATCAGTAATGAAAATGAGGCTATTGAACAACTATAAGGGCTAT
TTGCCTCAGGGCCTGGGAAAACATTCAGGACCCAGGAACCTCATGCCCTTCTTTTAGTTCAATCAGACAAGGCTCTAAAG
AGCCATATCCAGACTTTGTGGCAAGGTTGCAAGATGCAGCTCAAAAATCCATTGCAGGTAACGCCCGAAAAGTTATTGTA
GAAATAATGGCTTATCAAAACGCAAATTCAGAGTGTCAATCAGCCATAAAGCCATTAAGAGGAAATGTTTCAGCAGGAGT
TGATGTAATTACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATGCCATTGGCTCAAGCAA
TTACAGGGGTTGCTATAGGAGGACAAGTTAAAACATTTGGGGGAAAATGTTATAATTGTGGTCAAATCGGTCATCTAAAA
AAGAATTGCCCGAGCTTAAATTACCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 41:

NCCGGCCTTACGGCCGGGAAAGGCAGTCAAGCAGGAGTTAAACAATATGGACCTAACTCTCCTTATATTAGAATATTATT
AAATTCCATTGCTCATGGAAATAGACTTATTTCTTATGATTGGGAAATTCTGGCTATATCTTCCCTTTCACCCTCTCAGT
ATCTCCAGTTTAAAACCTGGTGGATTGATGGGGTACAAGAACAGGTACCGAAAAAATCAGGCTACTAATCCTGTTGCTTA
TATAGATGAAGACCAATTGCTAGGAAGAGGTCCAAACTGGGACACTATTAACCAACAATCAGTAATGAAAATGAGGCTAT
TGAACAACTATAAGGGCTATTTGCCTCAGGGCCTGGGAAAACATTCAGGACCCAGGAACCTCATGCCCTTCTTTTAGTTC
AATCAGACAAGGCTCTAAAGAGCCATATCCAGACTTTGTGGCAAGGTTGCAAGATGCAGCTCAAAAATCCATTGCAGGTA
ACGCCCGAAAAGTTATTGTAGAAATAATGGCTTATCAAAACGCAAATTCAGAGTGTCAATCAGCCATAAAGCCATTAAGA
GGAAATGTTTCAGCAGGAGTTGATGTAATTACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGC
AATGCCATTGGCTCAAGCAATTACAGGGGT

SEQ ID 42:

AAAGGCAGTCAAGCAGGAGTTAAACAATATGGACCTAACTCTCCTTATATGAGAACATTATTAAATTCCATTGCTCATGG
AAATAGACTTATTTCTTATGATTGGGAAATTCTGGCTAAATCTTCCCTTTCACCCTCTCAGTATCTCCAGTTTAAAACCT
GGTGGATTGATGGGGTACAAGAACAGGTACGAAAAAATCAGGCTACTAATCCTGTTGCTTATATAGATGAAGACCAATTG
CTAGGAAGAGGTCCAAACTGGGACACTATTAACCAACAATCAGTAATGAAAATGAGGCTATTGAACAACTATAAGGGCTA
TTTGCCTCAGGGGCCTGGGAAAACATTCAGGACCCAGGGAACCTCATGCCCTTCTTTTAGGTTCAATCAGACAAGGT

SEQ ID 43:

GCTGACTTGCTAGTATCATCTGCATTCATTGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAA
AAATAAATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCA
CTCAGGAAGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCTTCATTT
GGAAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTCT
TCCCATGTTAAAAGACATTTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGGGCCAGG
TTCTGTAGTAAAG

SEQ ID 44:

GCCAAGGTGGGAGGATTGCTTGAGCACAGGAGTTTGAGGCTGAAGTGAGCTATGATCGCACCACTGCAATCAATCAATCA
ATAAACTTCAGTCAACCCTGCCAGGAGCTATGGAACAATTATTGTTTGTTGGAGTGTTCTGTGTTGGGCTAAATGTGAAG
CCTCTTTATACTTCTACCTTACTCAGTCACCATATGGGGGCTGCCCCAGAGAGGTCATGACCTCAAGTGAGGAAGTACTC

AGCAGCTGAGCCAGGCCCTACTGATAGCTGGAGGATGCTGCTGCCCATGCTGCCCACTGTGAGGCAGCAAGCCCTTGCTT
GAAGGGGGATCTGGATAGTATGTTTCTGTGTCTACCACCCCTAGAAATGGTGCCTAGAGTGAGTCATCACAAAAAGAATC
AGGATAGCTTGGTGTAGTGGCAGGTGCCTATAATCCCAGCTACTCAGGAGACTGTGGCAGGAGAATGACTTAAACCAGGG
AGTTGGAGGTTGCAGTGAGGTGAGGTCACACAACTGCACTCCAGACTGGGTGACAGAGTGAGACTCCATCTCAAAAAAAA
AAAAAAAAAGAAAAGAAAAGAAAAAAGAAAAAGAATCAGGAAATACTAATATTTAAAGGATAGGTGAATGGAGGAAAATA
ATCAATTGAAGGAGGCTGAGCAGATGAGGTCAAAGAAGATAGAGATCCATAACAGTAACCTCATAGAAGCTTATGGAAGC
ATTTTGACAGTGCTAAAAGCCACATAAAGTTCAAGTAAGACAGTTTCAGAAATGTATAAACATGAATGCCTTTGCAGTGA
CTTAAGTGTGATTCTGGTGTTTCCTTCTAAAAATACTGCCTTCTCAGGTGTGGGAAGGATTCTATCTTTTTAGGCTTTAC
CACCATAGTTCTCTGCAGGCTTGCAATCCTGAATCAGGCTTGACTTCAGAAAGTGCTTTAAAAGGGAGGCTGGGCGCGGT
GGCTCATGCCTGTAATCCCAGCACTCTGAGAGGCTGAGGTTGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCT
GTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTCTGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATC
TATGACCTTACCCCCAACCCCGTGCTCTCTGAAACAGGTGCTGTGTCAAACTCAGGGTTAAATGGATTAAGGGTTGTGCA
AGATGTGCTTTGTTAAACAAATGCTTGAAGGCAGCATGGTCCTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAG
GGACACAAACACTGCGGAAGGCCGCAGAGACCTCTGCCTAGGAAAGCAAGGTATTGTCCAAGGTTTCTCCCCATGTGATA
GTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCTGACACCCGTAAAGGGTCTGTGCTGAGGAG
GATTAGTGTAAGAGGAAGGCATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCCGTCCCTGGGCAATG
GAATGTCTCGGTATAAAACCCGATTGATTGTACGTTCCATCTACTGAGATAGGAAGAAAACGCCTTAGGGCTGGAGGTGT
GGGACAAGCCGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTGAT
TCTTTACCTTGTCTGTGATGCAAAGACCTTTGTTCACGTGTTTGTCTGCTGACCCTCTCCCCACTATTGTCTTGTGACCA
TGACACATCCCCCTCTCAGAGAAACACCCACGAATGATCAATAAATACTAAGGGAACTCAGAGACGGCGCGGATCCTCCA
TATGCTGAACGCTGGTTCCCTGGGTCCCCTTATTTCTTTCTCTATACTTTGTGTCTTTTTCTTTTCCAAGTCTCTCGTTC
CACCTTACGAGAAACACCCACAGGTGTGGAGGGGCAACCCACCCCTTCATCTGGTGCCCAACGTGGAGGCTTTTCTCTAG
GGTGAAGGTACGCTCGAGCGTGGTCATTGAGGACAAGTTGACGAGAGATCCCGAGTACATCTACAGTCAGCCTTGCGGTA
AGTTTGTGCGCTCGGAAGAAGCTAGGGTGATAATGGGGCAAACTAAAAGTAAAACTAAAAGTAAATATGCCTCTTATCTC
AGCTTTATTAAAATTCTTTTAAAAAGAGGGGGAGTTAGAGTATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGA
ACAATTTTGCCCATGGTTTCCAGAACAAGGAACTTTAGATCTAAAAGATTGGAAAAGAATTGGCGAGGAACTAAAACAAG
CAGGTAGAAAGGGTAATATCATTCCACTTACAGTATGGAATGATTGGGCCATTATTAAAGCAGCTTTAGAACCATTTCAA
ACAAAAGAAGATAGCGTTTCAGTTTCTGATGCCCCTGGAAGCTGTGTAATAGATTGTAATGAAAAGACAGGGAGAAAATC
CCAGAAAGAAACAGAAAGTTTACATTGCGAATATGTAACAGAGCCAGTAATGGCTCAGTCAACGCAAAATGTTGACTATA
ATCAATTACAGGGGGTGATATATCCTGAAACGTTAAAATTAGAAGGAAAAGGTCCAGAATTAGTGGGGCCATCAGAGTCT
AAACCACGAGGGCCAAGTCCTCTTCCAGCAGGTCAGGTGCCCGTAACATTACAACCTCAAACGCAGGTTAAAGAAAATAA
GACCCAACCGCCAGTAGCTTATCAATACTGGCCGCCGGCTGAACTTCAGTATCTGCCACCCCCAGAAAGTCAGTATGGAT
ATCCAGGAATGCCCCCAGCACTACAGGGCAGGGCGCCATATCCTCAGCCGCCCACTGTGAGACTTAATCCTACAGCATCA
CGTAGTGGACAAGGTGGTACACTGCACGCAGTCATTGATGAAGCCAGAAAACAGGGAGATCTTGAGGCATGGCGGTTCCT
GGTAATTTTACAACTGGTACAGGCCGGGGAAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTT
TCACCATGAAAATGTTAAAAGATATAAAGGAAGGAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACATTATTA
GATTCCATTGCTCATGGAAATAGACTTACTCCTTATGACTGGGAAAGTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTA
TCTACAGTTTAAAACCTGGTGGATTGATGGAGTACAAGAACAGGTACGAAAAAATCAGGCTACTAAGCCCACTGTTAATA
TAGACGCAGACCAATTGTTAGGAACAGGTCCAAATTGGAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATT
GAACAAGTAAGGGCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTTTCCCTATTAATTCAATTAG
ACAAGGCTCTAAAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTACAGATGACAATG
CCCGAAAAGTTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAAAAGGA
AAAGTTCCAGCAGGAGTTGATGTAATTACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAAT
GCTAATGGCTCAAGCAATGAGGGGGCTCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTC
AAATCGGTCATCTGAAAAGGAGTTGCCCAGTCTTAAATAAACAGAATATAATAAATCAAGCTATTACAGCAAAAAATAAA
AAGCCATCTGGCCTGTGTCCAAAATGTGGAAAAGGAAAACATTGGGCCAATCAATGTCATTCTAAATTTGATAAAGATGG
GCAACCATTGTCGGGAAACAGGAAGAGGGGCCAGCCTCAGGCCCCCCAACAAACTGGGGCATTCCCAGTTCAACTGTTTG
TTCCTCAGGGTTTTCAAGGACAACAACCCCTACAGAAAATACCACCACTTCAGGGAGTCAGCCAATTACAACAATCCAAC
AGCTGTCCCGCGCCACAGCAGGCAGCGCCACAGTAGATTTATGTTCCACCCAAATGGTCTCTTTACTCCCTGGAGAGCCC
CCACAAAAGATTCCTAGAGGGGTATATGGCCCGCTGCCAGAAGGGAGGGTAGGCCTTATTTTAGGGAGATCAAGTCTAAA
TTTGAAGGGAGTCCAAATTCATACTGGGGTAATTTATTCAGATTATAAAGGGGGAATTCAGTTAGTGATCAGCTCCACTG
TTCCCTGGAGTGCCAATCCAGGTGATAGAATTGCTCAATTACTGCTTTTGCCTTATGTTAAAATTGGGGAAAACAAAACG
GAAAGAACAGGAGGGTTTGGAAGTACCAACCCTGCAGGAAAAGCCACTTATTGGGCTAATCAGGTCTCAGAGGATAGACC
CGTGTGTACAGTCACTATTCAGGGAAAGAGTTTGAAGGATTAGTGGATACCCAGGCTGATGTTTCTATCATCGGCATAGG
CACCGCCTCAGAAGTGTATCAAAGTGCCATGATTTTACATTGTCTAGGATCTGATAATCAAGAAAGTACGGTTCAGCCTA
TGATCACTTCTATTCCAATCAATTTATGGGGCCGAGACTTGTTACAACAATGGCATGCAGAGATTACTATCCCAGCCTCC
CTATACAGCCCCAGGAATCAAAAAATCATGACTAAAATGGGATAGCTCCCTAAAAAGGGACTAGGAAAGAATGAAGATGG
CATTAAAGTCCCAACTGAGGCTGAAAAAAATCAAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGAAGCGGTCACTGTAG
AGCCTCCAAAACCCATTCCATTAATTTGGGGGGAAAAAAAAAACTGTATGGTAAATCAGTAGCCGCTTCCAAAACAAAAA
CTGGAGGCTTTACACTTATTAGCAAAGAAACAGTTAGAAAAAGGACATATTGAGCCTTCATTTTCGCCTTGGAATTCTCC
TGTTTGTAATTCAGAAAAAATCCGGCAGATGGCGTATGCTAACTGACTTAAGAGCCATTAATGCCATAATTCAACCCATG
GGGGCTCTCCCATCCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTTACCA
TTCCTCTGGCAAAAGAGGATTTTGAAAAATTTGCTTTTACTATACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTC
AGTGGAAAGTATTGCCTCAGGGAATGCTTAATAATTCAACTATTTGTCAGACTTTCATAGCTCAAGCTCTGCAACCAGTT
AGAGACAAGTTTTCAGACTGTTATATCGTTCATTATGTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGA
CCGTTACACATTTCTCAGACAGAGGTTGCCAACGCGGGACTGACAATAGCATCTGATAAGATTCAAACCTCTCCTCCTTT
CCATTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCACAAAAAATAGAAATAAGAAAAGACACATTAAAAACAT
TAAATGAGTTTCAAAAGTTGGTAGGAGATACTAATTGGATTCGGAGATATTAATTGGATTTGGCCAACTCTAGGCATTCC
TACTTATGCCATGTCAATTTTGTTCTCTTTCTTAAGAGGGGACTTGGAATTAAATAGTGAAAGAATGTTACCTCCAGAGG

CAACTAAAGAAATTAAATTAATTGAAGAAAAAAATTCGGTCAGCACAAGTAAATAGGATCACTTGGCCCCACTCCAAATT
TTGATTTTTGGTACTGCACATTCTCTAACAGCCATCATTGTTCAAAACACAGATCTTGTGGATTGGTCCTTCCTTCCTCA
TAGTACAATTAAGACTTTTACATTGTACTTGGATCAAATGGCTACATTAATTGGTCAGGGAAGATTACGAATAATAACAT
TGTGTGGAAATGACCCAGATAAAATCACTGTTCCTTTCAACAAGCAACAAGTTAGACAAGCCTTTATCAGTTCTGGTGCA
TGGCAGATTGGTCTTGCTAATTTTCTGGGAATTATTGATAATCATTACCCAAAAACAAAAATCTTCCAGTTCTTAAAATT
GACTACTTGGATTCTACCTAAAATTACCAGACGTGAACCTTTAGAAAATGCTCTAACAGTATTTACTGATGGTTCCAGCA
ATGGAAAAGCGGCTTACACAGGGCCGAAAGAACGAGTAATCAAAACTCCGTATCAATCAGCTCAAAGAGCAGAGTTGGTT
GCAGTCATTACAGTGTTACAAGATTTTGACCAACCTATCAATATTATATCAGATTCTGCATATGTAGTACAGGCTACAAG
GGATGTTGAGACAGCTCTAATTAAATATAGCACGGACGATCATTTAAACCAGCTATTCAATTTATTACAACAAACTGTAA
GAAAAAGAAATTTCCCATTTTATATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTGACTAAAGCAAATGAA
CAAGCTGACTTACTGGTATCATCTGCATTCATAAAAGCACAAGAACTTCTTGCTTTGACTCATGTAAATGCAGCAGGATT
AAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACATTGCACCCAGTGTCAAGTCTTACACCTGT
CCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCGTTATGGCAAATGGATGGCACGCATGTTCCTTCA
TTTGGAAGATTATCATATGTTCATGTAACAGTTGATACTTATTCACATTTCATATGGGCAACTTGCCAAACAGGAGAAAG
TACTTCCCATGTTAAAAAACATTTATTATCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCAAAACTGACAATGGAC
CAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAACAGGAATTCCTTATAATTCC
CAAGGACAGGCCATAGTTGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGAAGGGGGAGACAGTAA
GGAGTGTACCACTCCTCAGATGCAACTTAATCTAGCACTCTATACTTTAAATTTTTTAAACATTTATAGAAATCAGACTA
CTACTTCTGCAAAACAACATCTTACTGGTAAAAAGCACAGCCCACATGAAGGAAAACTAATTTGGTGGAAAGATAATAAA
AATAAGACATGGGAAATAGGGAAGGTGATAACGTGGGGGAGAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAGCTTCC
TGTTTGGATACCCACTAGACATTTGAAGTTCTACAATGAACCCATCGGAGATGCAAAGAAAAGGGCCTCCACAGAGATGG
TAACCCCAGTCACATGGATGGATAATCCTATAGAAGTATATGTTAATGATAGTGTATGGGTACCTGGCCCCACAGATGAT
CGCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGTGTATCGTTATCCTCCTATTTGCCTAGG
GAGAGCACCAGGATGTTTAATGCCTGCAGTCCAAAATTGGTTGGTAGAAGTACCTACTGTCAGTCCTAACAGTAGATTCA
CTTATCACATGGTAAGCGGGATGTCACTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTAAAA
TTTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAGTTTTAGTTTGGGAAGAATG
TGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTCGGAACTATTATAGATTGGGCACCTCGAGGTCAATTCTACC
ACAATTGCTCAGGACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTCTA
GACAAACATAAGCATAAAAAATTACAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATCTCTACCCCAAGACCAGA
AATAATAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTTGGCCTGACACCACATTAGAATTTGGTCTGG
AAATCAAACTTTAGAAACAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCCAGTCTAACGGTTCCTTTACAAA
GTTGCGTAAAGCCCTCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAAACTATAACCTGTGAAAAT
TGTAGATTGTTTACTTGCATTGATTCAACTTTTAATTGGCGGCACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTG
GATCTCTGTGTCCGTGGACTGACCGTGGGAGGCCTCGCCATCCATCCATATTTTGACTGAAGTATTAAAAGACATTTTAA
ATAGATCCAAAAGATTCATTTTTACCTTAATTGCAGTGATTATGGGATTAATTGCAGTCACAGCTACGGCTGCTGTGGCA
GGAGTTGCATTGCACTCTTCTGTTCAGTCGGTAAACTTTGTTAATGATTGGCAAAAGAATTCTACAAGATTGTGGAATTC
ACAATCTAGTATTGATCAAAAATTGGCAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGACTCA
TGAGCTTAGAACATTGTTTCCAGTTACAGTGTGACTGGAATACGTCAGATTTTTGTATTACACCCCAAATTTATAATGAG
TCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAA
ATAACAAATTTTCGAAGCATCAAAAGCCCATTTAAATTTGATGCCAGGAACTGAGGCAATTGCAGGAGTTGCTGATGGCC
TCGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATCGGAAGTACTATGATTATAAATCTCATATTAATCCTTGTGTGC
CTGTTTTGTCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCATGATGAC
GATGGCGGTTTTGTCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGA
AAGAAGTAGACATAGGAGACTCCATTTTGTTCTGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACC
TTACCCCCAACCCCGTGCTCTCTGAAACAGGTGCTGTGTCAAACTCAGGGTTAAATGGATTAAGGGTTGTGCAAGATGTG
CTTTGTTAAACAAATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACA
AAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAA
ATATGGCCTCATGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGT
ATAAGAGGAAGGCATTCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCCGTCCCTGGGCAATGGAATGTC
TCGGTATAAAACCCGATTGTACGTTCCATCTACTGAGATAGGAAGAAAACGCCTTAGGGCTGGAGGTGGGACATGCAGGC
AGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTGATTCTTTACCTTGTC
TATGATGCAAAGACCTTTGTTCACCTGTTTGTCTGCTGACCCTCTCCCCACTATTGTCTTGTGACCATGACACATCCCCC
TCTCAGAGAAACACCCACGAATGATCAATAAATACTAAGGGAACTCAGAGACGGCGCGGATCCTCCATATGCTGAACGCT
GGTTCCCTGGGTCCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAGTCTCTCATTCCACCTTA
AGAGAAACACTCACAGGTGTGGAGGGGCAACCCATCCCTTCAGAGGTGGGTGGATCACCTGAGGTCAGGAGTTCAAGACA
AGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTAGCCAGGTGTGGTGGCAGGTGTCTGTAGTCCC
AGCTACTTGGGAGGCTGACGAGAATCGCTTGAACCTGGGAGGGGGAGGTTTCAGTGAGCCGAGATTGCACCACTGCACTC
CAGCCTGGGGGACAGAGTGAAACTCTGTCTCAAAAAAACAACAAAAAACCCCCACCTATAGACAGGACTAGCTACATAAAT
AACTTGCAGGGCTCAGTGTAAAATGAAAGTGTGAGGTCCCTTTTTCAAAGACGTAGAAGGCCGGGTGCGGTGGCTCATGC
CTGTAATCCCAGCACTTTGGGAGGCTGAGGCAGGCAGGTTATGAGGTCAGGAGTTCGAGACAGCCTGACCAATATGGTGA
AACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTAGCGGGCGCCTGTAGTCCCAGCTACTCAGGAGGCTGA
GGCAGAAGAATTACTTGAACCCAGGAGACGGAGGTTGCAGTGAGCTGAGATCGTGCCACTGCACTCTCCAGCCTCCTCGG
TGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAACAGAAAAAGGTGCTATTAAAGATACCAAAATATAAGGCACT
TTCCTTTATTCTGCAATCTGTCTCTCCACTTTTCATAGTATTTTTTCATTTGTTATTTAACATCATGTTTTGTCAGGTGA
GGACATTTACTCAGCCAGTGCAGCACTCACTGGTATCCAGGGGCCATAGGTGATTTGACGCACCCACATGGCCCACCAGC
TGTTGAGTTCCACCTCCAGCCAGCCACTGGACCAACATGCAGTGCCCTGGCTGGGGGCAGGAAAGTCTAACAAACCATTT
CATTCCACTGTCCTCCTGGCCAAACCCACAGAGGACAGGTAAACCCCCTTGTATGTGTTTTGTACTTGGATCTGGGGTGG
GC

SEQ ID 45:

TGTGGGGAAAAGCAAGAGAGATCAAATTGTTACTGTGTCTGTGTAGAAGAAGTAGACATAGGAGACTCCATTTTGTTAT
GTGCTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACGTGTG
CTGTGTCAACTCAGGGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTC
CTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAG
GAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGT
CCCCCAGCCCGACACCTGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA
AGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACT
GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTA
TGTGTATGCATATCCAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTTTGTTCACGTGTTTGTCT
GCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACACCCACAGATGATCAATAAATA
CTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTATTTCTTTCTCTATAC
TTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA
CCCCTACATCTGGTGCCCAACGTGGAGGCTTTTCTCTAGGGTGAAGGTACGCTCGAGCGTAATCATTGAGGACAAGTCGA
CGAGAGATCCCGAGTACATCTACAGTCAGCCTTACGGTAAGCTTGCGCGCTCGGAAGAAGCTAGGGTGATAATGGGGCAA
ACTAAAAGTAAAATTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGGAGTTAAAGT
ATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAACTTCAGATC
TAAAAGATTGGAAAAGAATTGGTAAGGAACTAAAACAAGCAGGTAGGAAGGGTAATATCATTCCACTTACAGTATGGAAT
GATTGGGCCATTATTAAAGCAGCTTTAGAACCATTTCAAACAGAAGAAGATAGCATTTCAGTTTCTGATGCCCCTGGAAG
CTGTTTAATAGATTGTAATGAAAACACAAGGAAAAAATCCCAGAAAGAAACCGAAAGTTTACATTGCGAATATGTAGCAG
AGCCGGTAATGGCTCAGTCAACGCAAAATGTTGACTATAATCAATTACAGGAGGTGATATATCCTGAAACGTTAAAATTA
GAAGGAAAAGGTCCAGAATTAATGGGGCCATCAGAGTCTAAACCACGAGGCACAAGTCCTCTTCCAGCAGGTCAGGTGCT
CGTAAGATTACAACCTCAAAAGCAGGTTAAAGAAAATAAGACCCAACCGCAAGTAGCCTATCAATACTGCCGCTGGCTGA
ACTTCAGTATCGGCCACCCCCAGAAAGTCAGTATGGATATCCAGGAATGCCCCCAGCACCACAGGGCAGGGCGCCATACC
ATCAGCCGCCCACTAGGAGACTTAATCCTATGGCACCACCTAGTAGACAGGGTAGTGAATTACATGAAATTATTGATAAA
TCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAACCGATGCCACCTGGAGAAGGAGCCCAAGA
GGGAGAGCCTCCCACAGTTGAGGCCAGATACAAGTCTTTTTCGATAAAAATGCTAAAAGATATGAAAGAGGGAGTAAAAC
AGTATGGACCCAACTCCCCTTATATGAGGACATTATTAGATTCCATTGCTTATGGACATAGACTCATTCCTTATGATTGG
GAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTTACAATTTAAGACTTGGTGGATTGATGGGGTACAAGAACA
GGTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACATAGATGCAGATCAACTATTAGGAATAGGTCAAAATTGGAGTA
CTATTAGTCAACAAGCATTAATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCTTGGGAAAAAATC
CAAGACCCAGGAAGTACCTGCCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCCCTACCCTGATTTTGTGGCAAG
GCTCCAAGATGTTGCTCAAAAGTCAATTGCCGATGAAAAAGCCGGTAAGGTCATAGTGGAGTTGATGGCATATGAAAACG
CCAATCCTGAGTGTCAATCAGCCATTAAGCCATTAAAAGGAAAGGTTCCTGCAGGATCAGATGTAATCTCAGAATATGTA
AAAGCCTGTGATGGAATCGGAGGAGCTATGCATAAAGCTATGCTTATGGCTCAAGCAATAACAGGAGTTGTTTTAGGAGG
ACAAGTTAGAACATTTGGAGGAAAATGTTATAATTGTGGTCAAATTGGTCACTTAAAAAAGAATTGCCCAGTCTTAAACA
AACAGAATATAACTATTCAAGCAACTACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAAAGGAAAACAT
TGGGCTAGTCAATGTCGTTCTAAATTTGATAAAAATGGGCAACCATTGTCGGGAAACGAGCAAAGGGGCCAGCCTCAGGC
CCCACAACAAACTGGGGCATTCCCAATTCAGCCATTTGTTCCTCAGGGTTTTCAGGGACAACAACCCCCACTGTCCCAAG
TGTTTCAGGGAATAAGCCAGTTACCACAATACAACAATTGTCCCTCACCACAAGCGGCAGTGCAGCAGTAGATTTATGTA
CTATACAAGCAGTCTCTCTGCTTCCAGGGGAGCCCCCACAAAAAATCCCTACAGGGGTATATGGCCCACTGCCTGAGGGG
ACTGTAGGACTAATCTTGGGAAGATCAAGTCTAAATCTAAAAGGAGTTCAAATTCATACTAGTGTGGTTGATTCAGACTA
TAAAGGCGAAATTCAATTGGTTATTAGCTCTTCAATTCCTTGGAGTGCCAGTCCAAGAGACAGGATTGCTCAATTATTAC
TCCTGCCATATATTAAGGGTGGAAATAGTGAAATAAAAAGAATAGGAGGGCTTGTAAGCACTGATCCAACAGGAAAGGCT
GCATATTGGGCAAGTCAGGTCTCAGAGAACAGACCTGTGTGTAAGGCCATTATTCAAGGAAAACAGTTTGAAGGGTTGGT
AGACACTGGAGCAGATGTCTCTATTATTGCTTTAAATCAGTGGCCAAAAAACTGGCCTAAACAAAAGGCTGTTACAGGAC
TTGTCGGCATAGGCACAGCCTCAGAAGTGTATCAAAGTATGGAGATTTTACATTGCTTAGGGCCAGATAATCAAGAAAGT
ACTGTTCAGCCAATGATTACTTCAATTCCTCTTAATCTGTGGGGTCGAGATTTATTACAACAATGGGGTGCGGAAATCAC
CATGCCCGCTCCATTATATAGCCCCACGAGTCAAAAAATCATGACCAAGATGGGATATATACCAGGAAAGGGACTAGGGA
AAAATGAAGATGGCATTAAAGTTCCAGTTGAGGCTAAAATAAATCAAGAAAGAGAAGGAATAGGGTATCCTTTTTAGGGG
CGGTCACTGTAGAGCCTCCTAAACCCATACCACTAACTTGGAAAACAGAAAAACCGGTGTGGGTAAATCAGTGGCCGCTA
CCAAAACAAAAACTGGAGGCTTTACATTTATTAGCAAATGAACAGTTAGAAAAGGGTCACATTGAGCCTTCGTTCTCACC
TTGGAATTCTCCTGTGTTTGTAATTCAGAAGAAATCAGGCAAATGGCATACGTTAACTGACTTAAGGGCTGTAAACGCCG
TAATTCAACCCATGGGGCCTCTCCAACCCGGGTTGCCCTCTCCGGCCATGATCCCAAAAGATTGGCCTTTAATTATAATT
GATCTAAAGGATTGCTTTTTTACCATCCCTCTGGCAGAGCAGGATTGTGAAAAATTTGCCTTTACTATACCAGCCATAAA
TAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATTTGTCAGACTT
TTGTAGGTCGAGCTCTTCAACCAGTGAGAGAAAAGTTTTCAGACTGTTATATTATTCATTATATTGATGATATTTTATGT
GCTGCAGAAACGAAAGATAAATTAATTGACTGTTATACATTTCTGCAAGCAGAGGTTGCCAATGCTGGACTGGCAATAGC
ATCCGATAAGATCCAAACCTCTACTCCTTTTCATTATTTAGGGATGCAGATAGAAAATAGAAAAATTAAGCCACAAAAAA
TAGAAATAAGAAAAGACACATTAAAAACACTAAATGATTTTCAAAAATTACTAGGAGATATTAATTGGATTCGGCCAACT
CTAGGCATTCCTACTTATGCCATGTCAAATTTGTTCTCTATCTTAAGAGGAGACTCAGACTTAAATAGTCAAAGAATATT
AACCCCAGAGGCAACAAAAGAAATTAAATTAGTGGAAGAAAAAATTCAGTCAGCGCAAATAAATAGAATAGATCCCTTAG
CCCCACTCCAACTTTTGATTTTTGCCACTGCACATTCTCCAACAGGCATCATTATTCAAAATACTGATCTTGTGGAGTGG
TCATTCCTTCCTCACAGTACAGTTAAGACTTTTACATTGTACTTGGATCAAATAGCTACATTAATCGGTCAGACAAGATT
ACGAATAACAAAATTATGTGGAAATGACCCAGACAAAATAGTTGTCCCTTTAACCAAGGAACAAGTTAGACAAGCCTTTA
TCAATTCTGGTGCATGGCAGATTGGTCTTGCTAATTTTGTGGGACTTATTGATAATCATTACCCAAAAACAAAGATCTTC
CAGTTCTTAAAATTGACTACTTGGATTCTACCTAAAATTACCAGACGTGAACCTTTAGAAAATGCTCTAACAGTATTTAC
TGATGGTTCCAGCAATGGAAAAGCAGCTTACACAGGGCCGAAAGAACGAGTAATCAAAACTCCATATCAATCGGCTCAAA
GAGACGAGTTGGTTGCAGTCATTACAGTGTTACAAGATTTTGACCAACCTATCAATATTATATCAGATTCTGCATATGTA
GTACAGGCTACAAGGGATGTTGAGACAGCTCTAATTAAATATAGCATGGATGATCAGTTAAACCAGCTATTCAATTTATT

ACAACAAACTGTAAGAAAAAGAAATTTCCCATTTTATATTACTTATATTCGAGCACACACTAATTTACCAGGGCCTTTGA
CTAAAGCAAATGAACAAGCTGACTTACTGGTATCATCTGCACTCATAAAAGCACAAGAACTTCATGCTTTGACTCATGTA
AATGCAGCAGGATTAAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACATTGCACCCAGTGTCA
AGTCTTACACCTGCCCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCATTATGGCAAATGGATGTCA
CGCATGTACCTTCATTTGGAAGATTATCATATGTTCATGTAACAGTTGATACTTATTCACATTTCATATGGGCAACTTGC
CAAACAGGAGAAAGTACTTCCCATGTTAAAAAACATTTATTGTCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCAA
AACTGACAATGGACCAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAACAGGAA
TTCCTTATAATTCCCAAGGACAGGCCATAGTTGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGAA
GGGGGAGACAGTAAGGAGTGTACCACTCCTCAGATGCAACTTAATCTAGCACTCTATACTTTAAATTTTTTAAACATTTA
TAGAAATCAGACTACTACTTCTGCAGAACAACATCTTACTGGTAAAAAGAACAGCCCACATGAAGGAAAACTAATTTGGT
GGAAAGATAATAAAAATAAGACATGGGAAATAGGGAAGGTGATAACGTGGGGGAGAGGTTTTGCTTGTGTTTCACCAGGA
GAAAATCAGCTTCCTGTTTGGTTACCCACTAGACATTTGAAGTTCTACAATGAACCCATCGGAGATGCAAAGAAAAGGGC
CTCCACGGAGATGGTAACACCAGTCACATGGATGGATAATCCTATAGAAGTATATGTTAATGATAGTATATGGGTACCTG
GCCCCATAGATGATCGCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCGTTATCCT
CCTATTTGCCTAGGGAGAGCACCAGGATGTTTAATGCCTGCAGTCCAAAATTGGTTGGTAGAAGTACCTACTGTCAGTCC
CATCAGTAGATTCACTTATCACATGGTAAGCGGGATGTCACTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATC
AAAGATCATTAAAATTTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAGTTTTA
GTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTATAAAACAATGAATTTGGAACTATTATAGATTGGGCACCTCG
AGGTCAATTCTACCACAATTGCTCAGGACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACT
TAACAGAAAGTTTAGACAAACATAAGCATAAAAAATTGCAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATCTCT
ACCCCAAGACCAAAAATAGTAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTACTGTGGCCTCACACCA
CATTAGAATTTGGTCTGGAAATCAAACTTTAGAAACAAGAGATTGTAAGCCATTTTATACTGTCGACCTAAATTCCAGTC
TAACAGTTCCTTTACAAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAG
ACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATTGGCAACACCGTATTCTGCTGGTGAG
AGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGGACCGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAG
TATTAAAAGGTGTTTTAAATAGATCCAAAAGATTCATTTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAGTCACA
GCTACGGCTGCTGTAGCAGGAGTTGCATTGCACTCTTCTGTTCAGTCAGTAAACTTTGTTAATGATTGGCAAAAGAATTC
TACAAGATTGTGGAATTCACAATCTAGTATTGATCAAAAATTGGCAAATCAAATTAATGATCTTAGACAAACTGTCATTT
GGATGGGAGACAGACTCATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTTGTATTACA
CCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAGGGAAGAGAAGATAATCTCACTTT
AGACATTTCCAAATTAAAAGAACAAATTTTCGAAGCATCAAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAATTG
CAGGAGTTGCTGATGGCCTCGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATTGGAAGTACATCGATTATAAATCTC
ATATTAATCCTTGTGTGCCTGTTTTGTCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCG
AGAACGGGCCATGATGACGATGGCGGTTTTGTCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAAATTG
TTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTATGTGCTAAGAAAAATTCTTCTGCCTTGAGA
TTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAACTCAGGGTTGAATGGATTAA
GGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCCCTAATCTC
AAGTACCCAGGGACACAAAAACTGCAGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCC
CCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCTGTAAAGGGTCTG
TGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCTGTCC
CTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGG
AGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGTGTATGCATATCCAAAAGCACAGCACT
TAATCCTTTACATTGTCTATGATGCCAAGACCTTTGTTCACGTGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTG
ACCCTGACACATCCCCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACTCAGAGGCTGGCGGGATCC
TCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAA
TCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCACCCCTACA

SEQ ID 46:

ETQVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTVLDSIAHGNRLTPYDWEILAKSSLSSSQYLQFKTWWIDG
VQEQVRKKSGY*AHC*YRRRPIVRNRSKLEHH*PTISDAE*GY*TSKGYLPQGLGKNSGPRNSFPY*FN*TRL*RAIS*L
CGKITRCCSKVYYR*QCPKSYCRINGL*KCKSRMSVGHKAIKRKSSSRS*CNYRICEGL*WDWRSYA*GNANGSSNEGAH
SRRTS*NIWEKML*LWSNRSSEKELPRLKQAKKKKKKKK

SEQ ID 47:

EETQVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTVLDSIAHGNRLTPYDWEILAKSSLSSSQYLQFKTWWID
GVQEQVRKNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEAIEQVRAICLRAWGKIQDPGTAFPINSIRQGSKEPYPD
FVARLQDAAQKSITDDNARKVIVELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGL
TLGGQVRTFGKKCYNCGQIGHLKRSCPGLNKQNIINQAITEKKKKKKK

SEQ ID 48:

EETQVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTLLDSIAHGNRLTPYDWEILAKSSLSSSQYLQFKTWWID
GVQEQVRKNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEAIEQVRAICLRAWGKIQDPGTAFPINSIRQGSKEPYPD
FVARLQDAAQKSITDDNARKVIVELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGL
TLGGQVRTFGKKCYNCGQIGHRKRSCPGLNKQNIINQAITAKNKKPSGLCPKCGKAKHWANQCHSKFDKDGQPLSGNRKR
GQPQAPQQTGAFPVKLFVPQGFQGQQPLQKIPPLQGVSQLQQSNSCPAPQQAAPQ*IYVPPKWSFYSLESPHKRFLEGYM
ARCQKGG*AFEGDQV*I*RESKFILG*FTQIIKGEFS**SAPLFPGVPIQVIELLNYCFCLMQKKKKKKKK

SEQ ID 49:

GSQAGVKQYGPNSPYIRILLNSIAHGNRLISYDWEILAISSLSPSQYLQFKTWWIDGVQEQVRKNQATNPVAYIDEDQLL
GRGPNWDTINQQSVMKMRLLNNYKGYLPQGLGKHSGPRNLMPFF*FNQTRL*RAISRLCGKVARCSSKIHCR*RPKSYCR
NNGLSKRKFRVSISHKAIKRKCFSRS*CNYRICEGL*WDWRSYA*GNAIGSSNYRGCYRRTS*NIWGKML*LWSNRSSKK
ELPELKLPPKKKKKKKKK

SEQ ID 50:

QKNESSKLSIT*LKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFY
ITHIRAHTNLPGPLTKANEQADLLVSSAFMEAQELHALTHVNAIGLKNRFDITWKQTKNIVQHCTQCQILHLATQEARVN
PRGLCPNVLWQMDVMHVPSFGKLSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLSCFPVMGVPEKVKTDNGPGYCSKAV
QKFLNQWKITHTIGILYNSQGQAIIERTNRTLKAQLVKQKKKKKKKK

SEQ ID 51:

QKNESSKLSIT*LKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFY
ITHIRAHTNLPGPLTKANEQADLLVSSAFMEAQELHALTHVNAIGLKNKFDITWKQTKNIVQHCTQCQILHLATQEARVN
PRGLCPNVLWQMDVMHVPSFGKLSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLSCFPVMGVPEKVKTDNGPGYCSKAV
QKFLNQWKITHTIGILYNSQGQAIIERTNRTLKAQLVKQKEKKKKK

SEQ ID 52:

QKNESSKLSITRLKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFY
ITHIRAHTNLPGPLTKANEQADLLVSSAFMEAQELHALTHVNAIGLKNKFDITWKQTKNIVQHCAQCQILHLATQEVRVN
PRGLCPNVLWQMDVMHVPSFGKLSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLSCFPVMGVPEKVKTDNGPGYCSKAV
QKFLNQWKITHTIGILYNSQGQAIIERTNRTLKAQLVKQKKKKKKKK

SEQ ID 53:

QKNESSKLSIT*LKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRK*NFPFY
ITHIRAHTNLPGPLTKANEQADLLVSSAFMEAQELHALTHVNAIGLKNKFDITWKQTKNIVQHCTQCQILHLATQEARVN
PRGLCPNVLWQMDVMHVPSFGKLSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLFCFPVMGVPEKVKTDNGPGYCSKAV
QEFLNQWKITHTIGILYNSQGQAIIERTNRTLKAQLVKQKKKKKKKK

SEQ ID 54:

QKNESSKLSIT*LKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFY
ITHIRAHTNLPGPLTKANEQADLLVSSAFMEAQELHALTHVNAIGLKNKFDITWKQTKNIVQHCTQCQILHLATQEARVN
PRGLCPNVLWQMDVMHVPSFGKLSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLSCFPVMGVPEKKKKKKKKK

SEQ ID 55:

QKNESSKLSIT*LKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFY
ITHIRAHTNLPGPLTKANEQADLLVSSAFIEAQELHALTHVNAIGLKNKFDITWKQTKNIVQHCTQCQILHLATQEARVN
PRGLCPNVLWQMDVMHVPSFGKLSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLSCFPVMGVPEKVKTDNGPGYCSKAV
QKFLNQWKITHT
IGILYNSQGQAIIERTNRTLKAQLVKQKKKKKKKTCRPPR

SEQ ID 56:

EETQVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTLLDSIAHGNRLTPYDWEILAKSSLSSSQYLQFKTWWID
GVQEQVRKNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEAIEQVRAICLRAWGKIQDPGTAFPINSIRQGSKEPYPD
FVARLQDAAQKSITDDNARKVIVELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGL
TLGGQVRTFGKKCYNCGQIGHRKRSCPGLNKQNIINQAITAKNKKPSGLCPKCGKAKHWANQCHSKFDKDGQPLSGNRKR
GQPQAPQQTGAFPVKLFVPQGFQGQQPLQKIPPLQGVSQLQQSNSCPAPQQAAPQ

SEQ ID 57:

EETQVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTVLDSIAHGNRLTPYDWEILAKSSLSSSQYLQFKTWWID
GVQEQVRKNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEAIEQVRAICLRAWGKIQDPGTAFPINSIRQGSKEPYPD
FVARLQDAAQKSITDDNARKVIVELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGL
TLGGQVRTFGKKCYNCGQIGHLKRSCPGLNKQNIINQAITEKKKKKKK

SEQ ID 58:

QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFYITHIRAHTNLPGPLTKANEQADLLV
SSAFMEAQELHALTHVNAIGLKNKFDITWKQTKNIVQHCTQCQILHLATQEARVNPRGLCPNVLWQMDVMHVPSFGKLSF
VHVTVDTYSHFIWATCQTGESTSHVKRHLLSCFPVMGVPEKVKTDNGPGYCSKAVQKFLNQWKITHTIGILYNSQGQAII
ERTNRTLKAQLVKQKKKKKKKTCRPPR

SEQ ID 59:

TAGGCCTTTGAGGGA

SEQ ID 60:

CATTAGAAAAAGGACATTG

SEQ ID 61:

TTGGAATTCTGTTTGTA

SEQ ID 62:

TAACTGAGCCATTAAT

SEQ ID 63:

AGCCATGGTCCCCTTTAATTA

SEQ ID 64:

TTTTACCACACCAGCCT

SEQ ID 65:

TTGTCAGCTCAAGCT

SEQ ID 66:

TACATCGTTCACTAT

SEQ ID 67:

TTAAAAGCATTAAAT

SEQ ID 68:

AGAAGTCCCAATTGAGG

SEQ ID 69:

GGTCTTGCCGATTTT

SEQ ID 70:

ACAATCGTTACCACA

SEQ ID 71:
AAAAGAATGAGTCAT
SEQ ID 72:
CAGTATCACTTGACT
SEQ ID 73:
TTTTAATCAGTCTATTAACATTG
SEQ ID 74:
AAAGGATATTGAGAGA
SEQ ID 75:
CCTAATCAAATACATT
SEQ ID 76:
CGCTGTTTAATTTGT
SEQ ID 77:
TGCATTCATGGAAGCA
SEQ ID 78:
ACTCAGGAGGCAAGA
SEQ ID 79:
TTAAGAGACATTTATT
SEQ ID 80:
TAAAGCAGTTCAAAAA
SEQ ID 81:
AATAGGAATTCTCTA
SEQ ID 82:
AAAGCTCAATTGGTTA
SEQ ID 83:
TAGGAGGACAAGTTAGAACATTTGG
SEQ ID 84:
AAAATGTTATAATTGTGGTCAAAT
SEQ ID 85:
ATGGGGCAAACTAAAAGTAAAATTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGG
AGTTAAAGTATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAA
CTTTAGATCTAAAAGATTGGAAAAGAATTGGTAAGGAACTAAAACAAGCAGGTAGGAAGGGTAATATCATTCCACTTACA
GTATGGAATGATTGGGCCATTATTAAAGCAGCTTTAGAACCATTTCAAACAGAAGAAGATAGCGTTTCAGTTTCTGATGC
CCCTGGAAGCTGTATAATAGATTGTAATGAAAACACAAGGAAAAAATCCCAGAAAGAAACGGAAGGTTTACATTGCGAAT
ATGTAGCAGAGCCGGTAATGGCTCAGTCAACGCAAAATGTTGACTATAATCAATTACAGGAGGTGATATATCCTGAAACG
TTAAAATTAGAAGGAAAAGGTCCAGAATTAGTGGGGCCATCAGAGTCTAAACCACGAGGCACAAGTCCTCTTCCAGCAGG
TCAGGTGCCTGTAACATTACAACCTCAAAAGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTAGCCTATCAATACTGGC
CTCCGGCTGAACTTCAGTATCGGCCACCCCCAGAAAGTCAGTATGGATATCCAGGAATGCCCCCAGCACCACAGGGCAGG
GCGCCATACCCTCAGCCGCCCACTAGGAGACTTAATCCTACGGCACCACCTAGTAGACAGGGTAGTAAATTACATGAAAT
TATTGATAAATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAACCGATGCCACCTGGAGAAG
GAGCCCAAGAGGGAGAGCCTCCCACAGTTGAGGCCAGATACAAGTCTTTTTCGATAAAAAAGCTAAAAGATATGAAAGAG
GGAGTAAAACAGTATGGACCCAACTCCCCTTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCC
TTATGATTGGGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTTACAATTTAAGACTTGGTGGATTGATGGGG
TACAAGAACAGGTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACATAGATGCAGATCAACTATTAGGAATAGGTCAA
AATTGGAGTACTATTAGTCAACAAGCATTAATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTG
GGAAAAAATCCAAGACCCAGGAAGTACCTGCCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCCCTATCCTGATT
TTGTGGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCTGATGAAAAAGCCCGTAAGGTCATAGTGGAGTTGATGGCA
TATGAAAACGCCAATCCTGAGTGTCAATCAGCCATTAAGCCATTAAAAGGAAAGGTTCCTGCAGGATCAGATGTAATCTC
AGAATATGTAAAAGCCTGTGATGGAATCGGAGGAGCTATGCATAAAGCTATGCTTATGGCTCAAGCAATAACAGGAGTTG
TTTTAGGAGGACAAGTTAGAACATTTGGAAGAAAATGTTATAATTGTGGTCAAATTGGTCACTTAAAAAAGAATTGCCCA
GTCTTAAATAAACAGAATATAACTATTCAAGCAACTACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAA
AGGAAAACATTGGGCTAGTCAATGTCGTTCTAAATTTGATAAAAATGGGCAACCATTGTCGGGAAACGAGCAAAGGGGCC
AGCCTCAGGCCCCACAACAAACTGGGGCATTCCCAATTCAGCCATTTGTTCCTCAGGGTTTTCAGGGACAACAACCCCCA
CTGTCCCAAGTGTTTCAGGGAATAAGCCAGTTACCACAATACAACAATTGTCCCCCGCCACAAGCGGCAGTGCAGCAG
SEQ ID 86:
ATGGGCAACCATTGTCGGGAAACGAGCAAAGGGGCCAGCCTCAGGCCCCACAACAAACTGGGGCATTCCCAATTCAGCCA
TTTGTTCCTCAGGGTTTTCAGGGACAACAACCCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCAGTTACCACAATACAA
CAATTGTCCCCCGCCACAAGCGGCAGTGCAGCAGTAGATTTATGTACTATACAAGCAGTCTCTCTGCTTCCAGGGGAGCC
CCCACAAAAAACCCCCACAGGGGTATATGGACCCCTGCCTAAGGGGACTGTAGGACTAATCTTGGGACGATCAAGTCTAA
ATCTAAAAGGAGTTCAAATTCATACTAGTGTGGTTGATTCAGACTATAAAGGCGAAATTCAATTGGTTATTAGCTCTTCA
ATTCCTTGGAGTGCCAGTCCAAGAGACAGGATTGCTCAATTATTACTCCTGCCATACATTAAGGGTGGAAATAGTGAAAT
AAAAAGAATAGGAGGGCTTGGAAGCACTGATCCAACAGGAAAGGCTGCATATTGGGCAAGTCAGGTCTCAGAGAACAGAC
CTGTGTGTAAGGCCATTATTCAAGGAAAACAGTTTGAAGGGTTGGTAGACACTGGAGCAGATGTCTCTATCATTGCTTTA
AATCAGTGGCCAAAAAATTGGCCTAAACAAAAGGCTGTTACAGGACTTGTCGGCATAGGCACAGCCTCAGAAGTGTATCA
AAGTACGGAGATTTTACATTGCTTAGGGCCAGATAATCAAGAAAGTACTGTTCAGCCAATGATTACTTCAATTCCTCTTA
ATCTGTGGGGTCGAGATTTATTACAACAATGGGGTGCGGAAATCACCATGCCCGCTCCATCATATAGCCCCACGAGTCAA
AAAATCATGACCAAGATGGGATATATACCAGGAAAGGGACTAGGGAAAAATGAAGATGGCATTAAAATTCCAGTTGAGGC
TAAAATAAATCAAGAAAGAGAAGGAATAGGGAATCCTTGC
SEQ ID 87:

ATGGCATTAAAATTCCAGTTGAGGCTAAAATAAATCAAGAAAGAGAAGGAATAGGGAATCCTTGCTAGGGGCGGCCACTG
TAGAGCCTCCTAAACCCATACCATTAACTTGGAAAACAGAAAAACCAGTGTGGGTAAATCAGTGGCCGCTACCAAAACAA
AAACTGGAGGCTTTACATTTATTAGCAAATGAACAGTTAGAAAAGGGTCATATTGAGCCTTCGTTCTCACCTTGGAATTC
TCCTGTGTTTGTAATTCAGAAGAAATCAGGCAAATGGCGTATGTTAACTGACTTAAGGGCTGTAAACGCCGTAATTCAAC
CCATGGGGCCTCTCCAACCCGGGTTGCCCTCTCCGGCCATGATCCCAAAAGATTGGCCTTTAATTATAATTGATCTAAAG
GATTGCTTTTTTACCATCCCTCTGGCAGAGCAGGATTGCGAAAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGA
ACCAGCCACCAGGTTTCAGTGGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATTTGTCAGACTTTTGTAGGTC
GAGCTCTTCAACCAGTTAGAGAAAAGTTTTCAGACTGTTATATTATTCATTGTATTGATGATATTTTATGTGCTGCAGAA
ACGAAAGATAAATTAATTGACTGTTATACATTTCTGCAAGCAGAGGTTGCCAATGCTGGACTGGCAATAGCATCTGATAA
GATCCAAACCTCTACTCCTTTTCATTATTTAGGGATGCAGATAGAAAATAGAAAAATTAAGCCACAAAAAATAGAAATAA
GAAAAGACACATTAAAAACACTAAATGATTTTCAAAAATTACTAGGAGATATTAATTGGATTCGGCCAACTCTAGGCATT
CCTACTTATGCCATGTCAAATTTGTTCTCTATCTTAAGAGGAGACTCAGACTTAAATAGTAAAAGAATGTTAACCCCAGA
GGCAACAAAAGAAATTAAATTAGTGGAAGAAAAAATTCAGTCAGCGCAAATAAATAGAATAGATCCCTTAGCCCCACTCC
AACTTTTGATTTTTGCCACTGCACATTCTCCAACAGGCATCATTATTCAAAATACTGATCTTGTGGAGTGGTCATTCCTT
CCTCACAGTACAGTTAAGACTTTTACATTGTACTTGGATCAAATAGCTACATTAATCGGTCAGACAAGATTACGAATAAT
AAAATTATGTGGGAATGACCCAGACAAAATAGTTGTCCCTTTAACCAAGGAACAAGTTAGACAAGCCTTTATCAATTCTG
GTGCATGGAAGATTGGTCTTGCTAATTTTGTGGGAATTATTGATAATCATTACCCAAAAACAAAGATCTTCCAGTTCTTA
AAATTGACTACTTGGATTCTACCTAAAATTACCAGACGTGAACCTTTAGAAAATGCTCTAACAGTATTTACTGATGGTTC
CAGCAATGGAAAAGCAGCTTACACAGGACCGAAAGAACGAGTAATCAAAACTCCATATCAATCGGCTCAAAGAGCAGAGT
TGGTTGCAGTCATTACAGTGTTACAAGATTTTGACCAACCTATCAATATTATATCAGATTCTGCATATGTAGTACAGGCT
ACAAGGGATGTTGAGACAGCTCTAATTAAATATAGCATGGATGATCAGTTAAACCAGCTATTCAATTTATTACAACAAAC
TGTAAGAAAAAGAAATTTCCCATTTTATATTACACATATTCGAGCACACACTAATTTACCAGGGCCTTTGACTAAAGCAA
ATGAACAAGCTGACTTACTGGTATCATCTGCACTCATAAAAGCACAAGAACTTCATGCTTTGACTCATGTAAATGCAGCA
GGATTAAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACATTGCACCCAGTGTCAAGTCTTACA
CCTGCCCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCATTATGGCAAATGGATGTCACGCATGTAC
CTTCATTTGGAAGATTATCATATGTTCACGTAACAGTTGATACTTATTCACATTTCATATGGGCAACTTGCCAAACAGGA
GAAAGTACTTCCCATGTTAAAAAACATTTATTGTCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCAAAACTGACAA
TGGACCAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAACAGGAATTCCTTATA
ATTCCCAAGGACAGGCCATAGTTGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGAAGGGGGAGAC
AGTAAGGAGTGTACCACTCCTCAGATGCAACTTAATCTAGCACTCTATACTTTAAATTTTTTAAACATTTATAGAAATCA
GACTACTACTTCTGCAGAACAACATCTTACTGGTAAAAAGAACAGCCCACATGAAGGAAAACTAATTTGGTGGAAAGATA
ATAAAAATAAGACATGGGAAATAGGGAAGGTGATAACGTGGGGGAGAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAG
CTTCCTGTTTGGATACCCACTAGACATTTGAAGTTCTACAATGAACCCATCAGAGATGCAAAGAAAAGCACCTCCGCGGA
GACGGAGACATCGCAATCGAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGCCA
TCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAACTAAAGAAGCTGACGCAGTTAGCTACAAAATATCTAGAGAACAC
AAAGGTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTG
CAGGAGCAGCTGCAGC

SEQ ID 88:

ATGAACCCATCAGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACATCGCAATCGAGCACCGTTGACTCACAAGAT
GAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAACTAA
AGAAGCTGACGCAGTTAGCTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCC
TTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTAACTATACCTACTGGGCCTATGT
GCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGGATAATCCTACAGAAGTATATGTTAATGATAGTGTATGGGTAC
CTGGCCCCATAGATGATCGCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCATTAT
CCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAATGCCTGCAGTCCAAAATTGGTTGGTAGAAGTACCTACTGTCAG
TCCCATCTGTAGATTCACTTATCACATGGTAAGCGGGATGTCACTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTT
ATCAAAGATCATTAAAATTTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAGTT
TTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTCGGAACTATTATAGATTGGGCACC
TCGAGGTCAATTCTACCACAATTGCTCAGGACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCG
ACTTAACAGAAAGTTTAGACAAACATAAGCATAAAAAATTGCAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATC
TCTACCCCAAGACCAAAAATAGTAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTACTGTGGCCTCACA
CCACATTAGAATTTGGTCTGGAAATCAAACTTTAGAAACAAGAGATCGTAAGCCATTTTATACTATTGACCTGAATTCCA
GTCTAACAGTTCCTTTACAAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCC
CAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATTGGCAACACCGTATTCTGCTGGT
GAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGGACCGACCGTGGGAGGCCTCGCCATCCGTCCATATTTTGACTG
AAGTATTAAAAGGTGTTTTAAATAGATCCAAAAGATTCATTTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAGTC
ACAGCTACGGCTGCTGTAGCAGGAGTTGCATTGCACTCTTCTGTTCAGTCAGTAAACTTTGTTAATGATTGGCAAAAAAA
TTCTACAAGATTGTGGAATTCACAATCTAGTATTGATCAAAAATTGGCAAATCAAATTAATGATCTTAGACAAACTGTCA
TTTGGATGGGAGACAGACTCATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTTGTATT
ACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAGGGAAGAGAAGATAATCTCAC
TTTAGACATTTCCAAATTAAAAGAACAAATTTTCGAAGCATCAAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAA
TTGCAGGAGTTGCTGATGGCCTCGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATTGGAAGTACTACGATTATAAAT
CTCATATTAATCCTTGTGTGCCTGTTTTGTCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCA

SEQ ID 89:

AGTTCTACAATGAACCCATCAGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACATCGCAATCGAGCACCGTTGAC
TCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGG
CACAACTAAAGAAGCTGACGCAGTTAGCTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCCAGAGAGTATGCTG
CTTGCAGCCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGA

SEQ ID 90:

TCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCATGA

SEQ ID 91:

ATGGGGCAAACTAAAAGTAAAATTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGG
AGTTAAAGTATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAA
CTTTAGATCTAAAAGATTGGAAAAGAATTGGTAAGGAACTAAAACAAGCAGGTAGGAAGGGTAATATCATTCCACTTACA
GTATGGAATGATTGGGCCATTATTAAAGCAGCTTTAGAACCATTTCAAACAGAAGAAGATAGCGTTTCAGTTTCTGATGC
CCCTGGAAGCTGTATAATAGATTGTAATGAAAACACAAGGAAAAAATCCCAGAAAGAAACGGAAGGTTTACATTGCGAAT
ATGTAGCAGAGCCGGTAATGGCTCAGTCAACGCAAAATGTTGACTATAATCAATTACAGGAGGTGATATATCCTGAAACG
TTAAAATTAGAAGGAAAAGGTCCAGAATTAGTGGGGCCATCAGAGTCTAAACCACGAGGCACAAGTCCTCTTCCAGCAGG
TCAGGTGCCTGTAACATTACAACCTCAAAAGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTAGCCTATCAATACTGGC
CTCCGGCTGAACTTCAGTATCGGCCACCCCCAGAAAGTCAGTATGGATATCCAGGAATGCCCCCAGCACCACAGGGCAGG
GCGCCATACCCTCAGCCGCCCACTAGGAGACTTAATCCTACGGCACCACCTAGTAGACAGGGTAGTAAATTACATGAAAT
TATTGATAAATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAACCGATGCCACCTGGAGAAG
GAGCCCAAGAGGGAGAGCCTCCCACAGTTGAGGCCAGATACAAGTCTTTTTCGATAAAAAAGCTGAAAGATATGAAAGAG
GGAGTAAAACAGTATGGACCCAACTCCCCTTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCC
TTATGATTGGGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTTACAATTTAAGACTTGGTGGATTGATGGGG
TACAAGAACAGGTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACATAGATGCAGATCAACTATTAGGAATAGGTCAA
AATTGGAGTACTATTAGTCAACAAGCATTAATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTG
GGAAAAAATCCAAGACCCAGGAAGTACCTGCCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCCCTATCCTGATT
TTGTGGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCTGATGAAAAAGCCCGTAAGGTCATAGTGGAGTTGATGGCA
TATGAAAACGCCAATCCTGAGTGTCAATCAGCCATTAAGCCATTAAAAGGAAAGGTTCCTGCAGGATCAGATGTAATCTC
AGAATATGTAAAAGCCTGTGATGGAATCGGAGGAGCTATGTATAAAGCTATGCTTATGGCTCAAGCAATAACAGGAGTTG
TTTTAGGAGGACAAGTTAGAACATTTGGAAGAAAATGTTATAATTGTGGTCAAATTGGTCACTTAAAAAAGAATTGCCCA
GTCTTAAATAAACAGAATATAACTATTCAAGCAACTACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAA
AGGAAAACATTGGGCTAGTCAATGTCGTTCTAAATTTGATAAAAATGGGCAACCATTGTCGGGAAACGAGCAAAGGGGCC
AGCCTCAGGCCCCACAACAAACTGGGGCATTCCCAATTCAGCCATTTGTTCCTCAGGGTTTTCAGGGACAACAACCCCCA
CTGTCCCAAGTGTTTCAGGGAATAAGCCAGTTACCACAATACAACAATTGTCCCCCGCCACAAGCGGCAGTGCAGCAGTA
G

SEQ ID 92:

MGQTKSKIKSKYASYLSFIKILLKRGGVKVSTKNLIKLFQIIEQFCPWFPEQGTLDLKDWKRIGKELKQAGRKGNIIPLT
VWNDWAIIKAALEPFQTEEDSVSVSDAPGSCIIDCNENTRKKSQKETEGLHCEYVAEPVMAQSTQNVDYNQLQEVIYPET
LKLEGKGPELVGPSESKPRGTSPLPAGQVPVTLQPQKQVKENKTQPPVAYQYWPPAELQYRPPPESQYGYPGMPPAPQGR
APYPQPPTRRLNPTAPPSRQGSKLHEIIDKSRKEGDTEAWQFPVTLEPMPPGEGAQEGEPPTVEARYKSFSIKKLKDMKE
GVKQYGPNSPYMRTLLDSIAHGHRLIPYDWEILAKSSLSPSQFLQFKTWWIDGVQEQVRRNRAANPPVNIDADQLLGIGQ
NWSTISQQALMQNEAIEQVRAICLRAWEKIQDPGSTCPSFNTVRQGSKEPYPDFVARLQDVAQKSIADEKARKVIVELMA
YENANPECQSAIKPLKGKVPAGSDVISEYVKACDGIGGAMYKAMLMAQAITGVVLGGQVRTFGRKCYNCGQIGHLKKNCP
VLNKQNITIQATTTGREPPDLCPRCKKGKHWASQCRSKFDKNGQPLSGNEQRGQPQAPQQTGAFPIQPFVPQGFQGQQPP
LSQVFQGISQLPQYNNCPPPQAAVQQ

SEQ ID 93:

ATGTTAACTGACTTAAGGGCTGTAAACGCCGTAATTCAACCCATGGGGCCTCTCCAACCCGGGTTGCCCTCTCCGGCCAT
GATCCCAAAAGATTGGCCTTTAATTATAATTGATCTAAAGGATTGCTTTTTTACCATCCCTCTGGCAGAGCAGGATTGCG
AAAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTGTTACCTCAGGGA
ATGCTTAATAGTCCAACTATTTGTCAGACTTTTGTAGGTCGAGCTCTTCAACCAGTTAGAGAAAAGTTTTCAGACTGTTA
TATTATTCATTGTATTGATGATATTTTATGTGCTGCAGAAACGAAAGATAAATTAATTGACTGTTATACATTTCTGCAAG
CAGAGGTTGCCAATGCTGGACTGGCAATAGCATCTGATAAGATCCAAACCTCTACTCCTTTTCATTATTTAGGGATGCAG
ATAGAAAATAGAAAAATTAAGCCACAAAAAATAGAAATAAGAAAAGACACATTAAAAACACTAAATGATTTTCAAAAATT
ACTAGGAGATATTAATTGGATTCGGCCAACTCTAGGCATTCCTACTTATGCCATGTCAAATTTGTTCTCTATCTTAAGAG
GAGACTCAGACTTAAATAGTAAAAGAATGTTAACCCCAGAGGCAACAAAAGAAATTAAATTAGTGGAAGAAAAAATTCAG
TCAGCGCAAATAAATAGAATAGATCCCTTAGCCCCACTCCAACTTTTGATTTTTGCCACTGCACATTCTCCAACAGGCAT
CATTATTCAAAATACTGATCTTGTGGAGTGGTCATTCCTTCCTCACAGTACAGTTAAGACTTTTACATTGTACTTGGATC
AAATAGCTACATTAATCGGTCAGACAAGATTACGAATAATAAAATTATGTGGGAATGACCCAGACAAAATAGTTGTCCCT
TTAACCAAGGAACAAGTTAGACAAGCCTTTATCAATTCTGGTGCATGGAAGATTGGTCTTGCTAATTTTGTGGGAATTAT
TGATAATCATTACCCAAAAACAAAGATCTTCCAGTTCTTAAAATTGACTACTTGGATTCTACCTAAAATTACCAGACGTG
AACCTTTAGAAAATGCTCTAACAGTATTTACTGATGGTTCCAGCAATGGAAAAGCAGCTTACACAGGACCGAAAGAACGA
GTAATCAAAACTCCATATCAATCGGCTCAAAGAGCAGAGTTGGTTGCAGTCATTACAGTGTTACAAGATTTTGACCAACC
TATCAATATTATATCAGATTCTGCATATGTAGTACAGGCTACAAGGGATGTTGAGACAGCTCTAATTAAATATAGCATGG
ATGATCAGTTAAACCAGCTATTCAATTTATTACAACAAACTGTAAGAAAAAGAAATTTCCCATTTTATATTACACATATT
CGAGCACACACTAATTTACCAGGGCCTTTGACTAAAGCAAATGAACAAGCTGACTTACTGGTATCATCTGCACTCATAAA
AGCACAAGAACTTCATGCTTTGACTCATGTAAATGCAGCAGGATTAAAAAACAAATTTGATGTCACATGGAAACAGGCAA
AAGATATTGTACAACATTGCACCCAGTGTCAAGTCTTACACCTGCCCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTG
TGTCCTAATGCATTATGGCAAATGGATGTCACGCATGTACCTTCATTTGGAAGATTATCATATGTTCACGTAACAGTTGA
TACTTATTCACATTTCATATGGGCAACTTGCCAAACAGGAGAAAGTACTTCCCATGTTAAAAAACATTTATTGTCTTGTT
TTGCTGTAATGGGAGTTCCAGAAAAAATCAAAACTGACAATGGACCAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTA
AGTCAGTGGAAAATTTCACATACAACAGGAATTCCTTATAATTCCCAAGGACAGGCCATAGTTGAAAGAACTAATAGAAC
ACTCAAAACTCAATTAGTTAAACAAAAAGAAGGGGGAGACAGTAAGGAGTGTACCACTCCTCAGATGCAACTTAATCTAG
CACTCTATACTTTAAATTTTTTAAACATTTATAGAAATCAGACTACTACTTCTGCAGAACAACATCTTACTGGTAAAAAG
AACAGCCCACATGAAGGAAAACTAATTTGGTGGAAAGATAGTAAAAATAAGACATGGGAAATAGGGAAGGTGATAACGTG
GGGGAGAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAGCTTCCTGTTTGGATACCCACTAGACATTTGAAGTTCTACA
ATGAACCCATCAGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACATCGCAATCGAGCACCGTTGACTCACAAGAT

GAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAACTAA
AGAAGCTGACGCAGTTAGCTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCC
TTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTAA

SEQ ID 94:

MLTDLRAVNAVIQPMGPLQPGLPSPAMIPKDWPLIIIDLKDCFFTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQG
MLNSPTICQTFVGRALQPVREKFSDCYIIHCIDDILCAAETKDKLIDCYTFLQAEVANAGLAIASDKIQTSTPFHYLGMQ
IENRKIKPQKIEIRKDTLKTLNDFQKLLGDINWIRPTLGIPTYAMSNLFSILRGDSDLNSKRMLTPEATKEIKLVEEKIQ
SAQINRIDPLAPLQLLIFATAHSPTGIIIQNTDLVEWSFLPHSTVKTFTLYLDQIATLIGQTRLRIIKLCGNDPDKIVVP
LTKEQVRQAFINSGAWKIGLANFVGIIDNHYPKTKIFQFLKLTTWILPKITRREPLENALTVFTDGSSNGKAAYTGPKER
VIKTPYQSAQRAELVAVITVLQDFDQPINIISDSAYVVQATRDVETALIKYSMDDQLNQLFNLLQQTVRKRNFPFYITHI
RAHTNLPGPLTKANEQADLLVSSALIKAQELHALTHVNAAGLKNKFDVTWKQAKDIVQHCTQCQVLHLPTQEAGVNPRGL
CPNALWQMDVTHVPSFGRLSYVHVTVDTYSHFIWATCQTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFL
SQWKISHTTGIPYNSQGQAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAEQHLTGKK
NSPHEGKLIWWKDSKNKTWEIGKVITWGRGFACVSPGENQLPVWIPTRHLKFYNEPIRDAKKSTSAETETSQSSTVDSQD
EQNGDVRRTDEVAIHQEGRAANLGTTKEADAVSYKISREHKGDTNPREYAACSLDDCINGGKSPYACRSSCS

SEQ ID 95:

ATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACATCGCAATCGAGCACCGTTGACTCACAAGATGAACAAAATGGTGAC
GTCAGAAGAACAGATGAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAACTAAAGAAGCTGACGCAGT
TAGCTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCCTTGATGATTGTATCA
ATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTAACTATACCTACTGGGCCTATGTGCCTTTCCCGCCCTT
AATTCGGGCAGTCACATGGATGGATAATCCTACAGAAGTATATGTTAATGATAGTGTATGGGTACCTGGCCCCATAGATG
ATCGCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCATTATCCTCCTATTTGCCTA
GGGAGAGCACCAGGATGTTTAATGCCTGCAGTCCAAAATTGGTTGGTAGAAGTACCTACTGTCAGTCCCATCTGTAGATT
CACTTATCACATGGTAAGCGGGATGTCACTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTAA
AATTTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAGTTTTAGTTTGGGAAGAA
TGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTCGGAACTATTATAGATTGGGCACCTCGAGGTCAATTCTA
CCACAATTGCTCAGGACAAACTCAGTCGTGTCAAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTT
TAGACAAACATAAGCATAAAAAATTGCAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATCTCTACCCCAAGACCA
AAAATAGTAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTACTGTGGCCTCACACCACATTAGAATTTG
GTCTGGAAATCAAACTTTAGAAACAAGAGATCGTAAGCCATTTTATACTATTGACCTGAATTCCAGTCTAACAGTTCCTT
TACAAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAGACTATAACCTGT
GAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATTGGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGG
CGTGTGGATCCCTGTGTCCATGGACCGACCGTGGGAGGCCTCGCCATCCGTCCATATTTTGACTGAAGTATTAAAAGGTG
TTTTAAATAGATCCAAAAGATTCATTTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAGTCACAGCTACGGCTGCT
GTAGCAGGAGTTGCATTGCACTCTTCTGTTCAGTCAGTAAACTTTGTTAATGATTGGCAAAAAAATTCTACAAGATTGTG
GAATTCACAATCTAGTATTGATCAAAAATTGGCAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACA
GACTCATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTTGTATTACACCCCAAATTTAT
AATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAA
ATTAAAAGAACAAATTTTCGAAGCATCAAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAATTGCAGGAGTTGCTG
ATGGCCTCGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATTGGAAGTACTACGATTATAAATCTCATATTAATCCTT
GTGTGCCTGTTTTGTCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCAT
GATGACGATGGCGGTTTTGTCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTG
TGTAG

SEQ ID 96:

MQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKLPSTKKAEPPTWAQLKKLTQLATKYLENTKVTQTPESMLLAALMIVS
MVVSLPMPAGAAAANYTYWAYVPFPPLIRAVTWMDNPTEVYVNDSVWVPGPIDDRCPAKPEEEGMMINISIGYHYPPICL
GRAPGCLMPAVQNWLVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQRSLKFRPKGKPCPKEIPKESKNTEVLVWEE
CVANSAVILQNNEFGTIIDWAPRGQFYHNCSGQTQSCQSAQVSPAVDSDLTESLDKHKHKKLQSFYPWEWGEKGISTPRP
KIVSPVSGPEHPELWRLTVASHHIRIWSGNQTLETRDRKPFYTIDLNSSLTVPLQSCVKPPYMLVVGNIVIKPDSQTITC
ENCRLLTCIDSTFNWQHRILLVRAREGVWIPVSMDRPWEASPSVHILTEVLKGVLNRSKRFIFTLIAVIMGLIAVTATAA
VAGVALHSSVQSVNFVNDWQKNSTRLWNSQSSIDQKLANQINDLRQTVIWMGDRLMSLEHRFQLQCDWNTSDFCITPQIY
NESEHHWDMVRRHLQGREDNLTLDISKLKEQIFEASKAHLNLVPGTEAIAGVADGLANLNPVTWVKTIGSTTIINLILIL
VCLFCLLLVCRCTQQLRRDSDHRERAMMTMAVLSKRKGGNVGKSKRDQIVTVSV

SEQ ID 97:

ATGGGGCAAACTAAAAGTAAAACTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGG
AGTTAGAGTATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAA
CTTTAGATCTAAAAGATTGGAAAAGAATTGGCGAGGAACTAAAACAAGCAGGTAGAAAGGGTAATATCATTCCACTTACA
GTATGGAATGATTGGGCCATTATTAAAGCAGCTTTAGAACCATTTCAAACAAAAGAAGATAGCGTTTCAGTTTCTGATGC
CCCTGGAAGCTGTGTAATAGATTGTAATGAAAAGACAGGGAGAAAATCCCAGAAAGAAACAGAAAGTTTACATTGCGAAT
ATGTAACAGAGCCAGTAATGGCTCAGTCAACGCAAAATGTTGACTATAATCAATTACAGGGGGTGATATATCCTGAAACG
TTAAAATTAGAAGGAAAAGGTCCAGAATTAGTGGGGCCATCAGAGTCTAAACCACGAGGGCCAAGTCCTCTTCCAGCAGG
TCAGGTGCCCGTAACATTACAACCTCAAACGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTAGCTTATCAATACTGGC
CGCCGGCTGAACTTCAGTATCTGCCACCCCCAGAAAGTCAGTATGGATATCCAGGAATGCCCCCAGCACTACAGGGCAGG
GCGCCATATCCTCAGCCGCCCACTGTGAGACTTAATCCTACAGCATCACGTAGTGGACAAGGTGGTACACTGCACGCAGT
CATTGATGAAGCCAGAAAACAGGGAGATCTTGAGGCATGGCGGTTCCTGGTAATTTTACAACTGGTACAGGCCGGGGAAG
AGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATAAAGGAA
GGAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACATTATTAGATTCCATTGCTCATGGAAATAGACTTACTCC
TTATGACTGGGAAAGTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTGATGGAG
TACAAGAACAGGTACGAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACAGGTCCA

AATTGGAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGGCCTG
GGGAAAAATTCAGGACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGACTTTG
TGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGCCTAT
GAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAAAAGGAAAAGTTCCAGCAGGAGTTGATGTAATTACAGA
ATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATGCTAATGGCTCAAGCAATGAGGGGGCTCACTC
TAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATCGGTCATCTGAAAAGGAGTTGCCCAGTC
TTAAATAAACAGAATATAATAAATCAAGCTATTACAGCAAAAAATAAAAAGCCATCTGGCCTGTGTCCAAAATGTGGAAA
AGGAAAACATTGGGCCAATCAATGTCATTCTAAATTTGATAAAGATGGGCAACCATTGTCGGGAAACAGGAAGAGGGGCC
AGCCTCAGGCCCCCCAACAAACTGGGGCATTCCCAGTTCAACTGTTTGTTCCTCAGGGTTTTCAAGGACAACAACCCCTA
CAGAAAATACCACCACTTCAGGGAGTCAGCCAATTACAACAATCCAACAGCTGTCCCGCGCCACAGCAGGCAGCGCCACA
GTAG

SEQ ID 98:

MGQTKSKTKSKYASYLSFIKILLKRGGVRVSTKNLIKLFQIIEQFCPWFPEQGTLDLKDWKRIGEELKQAGRKGNIIPLT
VWNDWAIIKAALEPFQTKEDSVSVSDAPGSCVIDCNEKTGRKSQKETESLHCEYVTEPVMAQSTQNVDYNQLQGVIYPET
LKLEGKGPELVGPSESKPRGPSPLPAGQVPVTLQPQTQVKENKTQPPVAYQYWPPAELQYLPPPESQYGYPGMPPALQGR
APYPQPPTVRLNPTASRSGQGGTLHAVIDEARKQGDLEAWRFLVILQLVQAGEETQVGAPARAETRCEPFTMKMLKDIKE
GVKQYGSNSPYIRTLLDSIAHGNRLTPYDWESLAKSSLSSSQYLQFKTWWIDGVQEQVRKNQATKPTVNIDADQLLGTGP
NWSTINQQSVMQNEAIEQVRAICLRAWGKIQDPGTAFPINSIRQGSKEPYPDFVARLQDAAQKSITDDNARKVIVELMAY
ENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGLTLGGQVRTFGKKCYNCGQIGHLKRSCPV
LNKQNIINQAITAKNKKPSGLCPKCGKGKHWANQCHSKFDKDGQPLSGNRKRGQPQAPQQTGAFPVQLFVPQGFQGQQPL
QKIPPLQGVSQLQQSNSCPAPQQAAPQ

SEQ ID 99:

ATGGGCAACCATTGTCGGGAAACAGGAAGAGGGGCCAGCCTCAGGCCCCCCAACAAACTGGGGCATTCCCAGTTCAACTG
TTTGTTCCTCAGGGTTTTCAAGGACAACAACCCCTACAGAAAATACCACCACTTCAGGGAGTCAGCCAATTACAACAATC
CAACAGCTGTCCCGCGCCACAGCAGGCAGCGCCACAGTAGATTTATGTTCCACCCAAATGGTCTCTTTACTCCCTGGAGA
GCCCCCACAAAAGATTCCTAGAGGGGTATATGGCCCGCTGCCAGAAGGGAGGGTAGGCCTTATTTTAGGGAGATCAAGTC
TAAATTTGAAGGGAGTCCAAATTCATACTGGGGTAATTTATTCAGATTATAAAGGGGGAATTCAGTTAGTGATCAGCTCC
ACTGTTCCCTGGAGTGCCAATCCAGGTGATAGAATTGCTCAATTACTGCTTTTGCCTTATGTTAAAATTGGGGAAAACAA
AACGGAAAGAACAGGAGGGTTTGGAAGTACCAACCCTGCAGGAAAAGCCACTTATTGGGCTAATCAGGTCTCAGAGGATA
GACCCGTGTGTACAGTCACTATTCAGGGAAAGAGTTTGAAGGATTAGTGGATACCCAGGCTGATGTTTCTATCATCGGCA
TAGGCACCGCCTCAGAAGTGTATCAAAGTGCCATGATTTTACATTGTCTAGGATCTGATAATCAAGAAAGTACGGTTCAG
CCTATGATCACTTCTATTCCAATCAATTTATGGGGCCGAGACTTGTTACAACAATGGCATGCAGAGATTACTATCCCAGC
CTCCCTATACAGCCCCAGGAATCAAAAAATCATGACTAAAATGGGATAGCTCCCTAAAAAGGGACTAGGAAAGAATGAAG
ATGGCATTAAAGTCCCAACTGAGGCTGAAAAAAATCAAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGAAGCGGTCACT
GTAGAGCCTCCAAAACCCATTCCATTAATTTGGGGGGAAAAAAA

SEQ ID 100:

ATGGCATTAAAGTCCCAACTGAGGCTGAAAAAAATCAAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGAAGCGGTCACT
GTAGAGCCTCCAAAACCCATTCCATTAATTTGGGGGGAAAAAAAAAACTGTATGGTAAATCAGTAGCCGCTTCCAAAACA
AAAACTGGAGGCTTTACACTTATTAGCAAAGAAACAGTTAGAAAAAGGACATATTGAGCCTTCATTTTCGCCTTGGAATT
CTCCTGTTTGTAATTCAGAAAAAATCCGGCAGATGGCGTATGCTAACTGACTTAAGAGCCATTAATGCCATAATTCAACC
CATGGGGGCTCTCCCATCCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTT
ACCATTCCTCTGGCAAAAGAGGATTTTGAAAAATTTGCTTTTACTATACCAGCCTAAATAATAAAGAACCAGCCACCAGG
TTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAATTCAACTATTTGTCAGACTTTCATAGCTCAAGCTCTGCAACC
AGTTAGAGACAAGTTTTCAGACTGTTATATCGTTCATTATGTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAA
TTGACCGTTACACATTTCTCAGACAGAGGTTGCCAACGCGGGACTGACAATAGCATCTGATAAGATTCAAACCTCTCCTC
CTTTCCATTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCACAAAAAATAGAAATAAGAAAAGACACATTAAAA
ACATTAAATGAGTTTCAAAAGTTGGTAGGAGATACTAATTGGATTCGGAGATATTAATTGGATTTGGCCAACTCTAGGCA
TTCCTACTTATGCCATGTCAATTTTGTTCTCTTTCTTAAGAGGGGACTTGGAATTAAATAGTGAAAGAATGTTACCTCCA
GAGGCAACTAAAGAAATTAAATTAATTGAAGAAAAAAATTCGGTCAGCACAAGTAAATAGGATCACTTGGCCCCACTCCA
AATTTTGATTTTTGGTACTGCACATTCTCTAACAGCCATCATTGTTCAAAACACAGATCTTGTGGATTGGTCCTTCCTTC
CTCATAGTACAATTAAGACTTTTACATTGTACTTGGATCAAATGGCTACATTAATTGGTCAGGGAAGATTACGAATAATA
ACATTGTGTGGAAATGACCCAGATAAAATCACTGTTCCTTTCAACAAGCAACAAGTTAGACAAGCCTTTATCAGTTCTGG
TGCATGGCAGATTGGTCTTGCTAATTTTCTGGGAATTATTGATAATCATTACCCAAAAACAAAAATCTTCCAGTTCTTAA
AATTGACTACTTGGATTCTACCTAAAATTACCAGACGTGAACCTTTAGAAAATGCTCTAACAGTATTTACTGATGGTTCC
AGCAATGGAAAAGCGGCTTACACAGGGCCGAAAGAACGAGTAATCAAAACTCCGTATCAATCAGCTCAAAGAGCAGAGTT
GGTTGCAGTCATTACAGTGTTACAAGATTTTGACCAACCTATCAATATTATATCAGATTCTGCATATGTAGTACAGGCTA
CAAGGGATGTTGAGACAGCTCTAATTAAATATAGCACGGACGATCATTTAAACCAGCTATTCAATTTATTACAACAAACT
GTAAGAAAAAAGAAATTTCCCATTTTATATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTGACTAAAGCAAA
TGAACAAGCTGACTTACTGGTATCATCTGCATTCATAAAAGCACAAGAACTTCTTGCTTTGACTCATGTAAATGCAGCAG
GATTAAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACATTGCACCCAGTGTCAAGTCTTACAC
CTGTCCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCGTTATGGCAAATGGATGGCACGCATGTTCC
TTCATTTGGAAGATTATCATATGTTCATGTAACAGTTGATACTTATTCACATTTCATATGGGCAACTTGCCAAACAGGAG
AAAGTACTTCCCATGTTAAAAAACATTTATTATCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCAAAACTGACAAT
GGACCAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAACAGGAATTCCTTATAA
TTCCCAAGGACAGGCCATAGTTGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGAAGGGGGAGACA
GTAAGGAGTGTACCACTCCTCAGATGCAACTTAATCTAGCACTCTATACTTTAAATTTTTTAAACATTTATAGAAATCAG
ACTACTACTTCTGCAAAACAACATCTTACTGGTAAAAAGCACAGCCCACATGAAGGAAAACTAATTTGGTGGAAAGATAA
TAAAAATAAGACATGGGAAATAGGGAAGGTGATAACGTGGGGGAGAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAGC
TTCCTGTTTGGATACCCACTAGACATTTGAAGTTCTACAATGAACCCATCGGAGATGCAAAGAAAAGGGCCTCCACAGAG

ATGGTAACCCCAGTCACATGGATGGATAATC

SEQ ID 101:

GTCACATGGATGGATAATCCTATAGAAGTATATGTTAATGATAGTGTATGGGTACCTGGCCCCACAGATGATCGCTGCCC
TGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGTGTATCGTTATCCTCCTATTTGCCTAGGGAGAGCAC
CAGGATGTTTAATGCCTGCAGTCCAAAATTGGTTGGTAGAAGTACCTACTGTCAGTCCTAACAGTAGATTCACTTATCAC
ATGGTAAGCGGGATGTCACTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTAAAATTTAGACC
TAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAGTTTTAGTTTGGGAAGAATGTGTGGCCA
ATAGTGCGGTGATATTACAAAACAATGAATTCGGAACTATTATAGATTGGGCACCTCGAGGTCAATTCTACCACAATTGC
TCAGGACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTCTAGACAAACA
TAAGCATAAAAAATTACAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATCTCTACCCCAAGACCAGAAATAATAA
GTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTTGGCCTGACACCACATTAGAATTTGGTCTGGAAATCAAA
CTTTAGAAACAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCCAGTCTAACGGTTCCTTTACAAAGTTGCGTA
AAGCCCTCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAAACTATAACCTGTGAAAATTGTAGATT
GTTTACTTGCATTGATTCAACTTTTAATTGGCGGCACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCTCTG
TGTCCGTGGACTGACCGTGGGAGGCCTCGCCATCCATCCATATTTTGACTGAAGTATTAAAAGACATTTTAAATAGATCC
AAAAGATTCATTTTTACCTTAATTGCAGTGATTATGGGATTAATTGCAGTCACAGCTACGGCTGCTGTGGCAGGAGTTGC
ATTGCACTCTTCTGTTCAGTCGGTAAACTTTGTTAATGATTGGCAAAAGAATTCTACAAGATTGTGGAATTCACAATCTA
GTATTGATCAAAAATTGGCAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGACTCATGAGCTTA
GAACATTGTTTCCAGTTACAGTGTGACTGGAATACGTCAGATTTTTGTATTACACCCCAAATTTATAATGAGTCTGAGCA
TCACTGGGACATGGTTAGACGCCATCTACAGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAATAACAAA
TTTTCGAAGCATCAAAAGCCCATTTAAATTTGATGCCAGGAACTGAGGCAATTGCAGGAGTTGCTGATGGCCTCGCAAAT
CTTAACCCTGTCACTTGGGTTAAGACCATCGGAAGTACTATGATTATAAATCTCATATTAATCCTTGTGTGCCTGTTTTG
TCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCA

SEQ ID 102:

ATGGGGCAAACTAAAAGTAAAATTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGG
AGTTAAAGTATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAA
CTTCAGATCTAAAAGATTGGAAAAGAATTGGTAAGGAACTAAAACAAGCAGGTAGGAAGGGTAATATCATTCCACTTACA
GTATGGAATGATTGGGCCATTATTAAAGCAGCTTTAGAACCATTTCAAACAGAAGAAGATAGCATTTCAGTTTCTGATGC
CCCTGGAAGCTGTTTAATAGATTGTAATGAAAACACAAGGAAAAAATCCCAGAAAGAAACCGAAAGTTTACATTGCGAAT
ATGTAGCAGAGCCGGTAATGGCTCAGTCAACGCAAAATGTTGACTATAATCAATTACAGGAGGTGATATATCCTGAAACG
TTAAAATTAGAAGGAAAAGGTCCAGAATTAATGGGGCCATCAGAGTCTAAACCACGAGGCACAAGTCCTCTTCCAGCAGG
TCAGGTGCTCGTAAGATTACAACCTCAAAAGCAGGTTAAAGAAAATAAGACCCAACCGCAAGTAGCCTATCAATACTGCC
GCTGGCTGAACTTCAGTATCGGCCACCCCCAGAAAGTCAGTATGGATATCCAGGAATGCCCCCAGCACCACAGGGCAGGG
CGCCATACCATCAGCCGCCCACTAGGAGACTTAATCCTATGGCACCACCTAGTAGACAGGGTAGTGAATTACATGAAATT
ATTGATAAATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAA

SEQ ID 103:

MGQTKSKIKSKYASYLSFIKILLKRGGVKVSTKNLIKLFQIIEQFCPWFPEQGTSDLKDWKRIGKELKQAGRKGNIIPLT
VWNDWAIIKAALEPFQTEEDSISVSDAPGSCLIDCNENTRKKSQKETESLHCEYVAEPVMAQSTQNVDYNQLQEVIYPET
LKLEGKGPELMGPSESKPRGTSPLPAGQVLVRLQPQKQVKENKTQPQVAYQYCRWLNFSIGHPQKVSMDIQECPQHHRAG
RHTISRPLGDLILWHHLVDRVVNYMKLLINQERKEILRHGNSQ

SEQ ID 104:

MPPAPQGRAPYHQPPTRRLNPMAPPSRQGSELHEIIDKSRKEGDTEAWQFPVTLEPMPPGEGAQEGEPPTVEARYKSFSI
KMLKDMKEGVKQYGPNSPYMRTLLDSIAYGHRLIPYDWEILAKSSLSPSQFLQFKTWWIDGVQEQVRRNRAANPPVNIDA
DQLLGIGQNWSTISQQALMQNEAIEQVRAICLRAWEKIQDPGSTCPSFNTVRQGSKEPYPDFVARLQDVAQKSIADEKAG
KVIVELMAYENANPECQSAIKPLKGKVPAGSDVISEYVKACDGIGGAMHKAMLMAQAITGVVLGGQVRTFGGKCYNCGQI
GHLKKNCPVLNKQNITIQATTTGREPPDLCPRCKKGKHWASQCRSKFDKNGQPLSGNEQRGQPQAPQQTGAFPIQPFVPQ
GFQGQQPPLSQVFQGISQLPQYNNCPSPQAAVQQ

SEQ ID 105:

ATGGAGATTTTACATTGCTTAGGGCCAGATAATCAAGAAAGTACTGTTCAGCCAATGATTACTTCAATTCCTCTTAATCT
GTGGGGTCGAGATTTATTACAACAATGGGGTGCGGAAATCACCATGCCCGCTCCATTATATAGCCCCACGAGTCAAAAAA
TCATGACCAAGATGGGATATATACCAGGAAAGGGACTAGGGAAAAATGAAGATGGCATTAAAGTTCCAGTTGAGGCTAAA
ATAAATCAAGAAAGAGAAGGAATAGGGTATCCTTTTTAG

SEQ ID 106:

MEILHCLGPDNQESTVQPMITSIPLNLWGRDLLQQWGAEITMPAPLYSPTSQKIMTKMGYIPGKGLGKNEDGIKVPVEAK
INQEREGIGYPF

SEQ ID 107:

ATGGGGCCTCTCCAACCCGGGTTGCCCTCTCCGGCCATGATCCCAAAAGATTGGCCTTTAATTATAATTGATCTAAAGGA
TTGCTTTTTTACCATCCCTCTGGCAGAGCAGGATTGTGAAAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGAAC
CAGCCACCAGGTTTCAGTGGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATTTGTCAGACTTTTGTAGGTCGA
GCTCTTCAACCAGTGAGAGAAAAGTTTTCAGACTGTTATATTATTCATTATATTGATGATATTTTATGTGCTGCAGAAAC
GAAAGATAAATTAATTGACTGTTATACATTTCTGCAAGCAGAGGTTGCCAATGCTGGACTGGCAATAGCATCCGATAAGA
TCCAAACCTCTACTCCTTTTCATTATTTAGGGATGCAGATAGAAAATAGAAAAATTAAGCCACAAAAAATAGAAATAAGA
AAAGACACATTAAAAACACTAAATGATTTTCAAAAATTACTAGGAGATATTAATTGGATTCGGCCAACTCTAGGCATTCC
TACTTATGCCATGTCAAATTTGTTCTCTATCTTAAGAGGAGACTCAGACTTAAATAGTCAAAGAATATTAACCCCAGAGG
CAACAAAAGAAATTAAATTAGTGGAAGAAAAAATTCAGTCAGCGCAAATAAATAGAATAGATCCCTTAGCCCCACTCCAA
CTTTTGATTTTTGCCACTGCACATTCTCCAACAGGCATCATTATTCAAAATACTGATCTTGTGGAGTGGTCATTCCTTCC
TCACAGTACAGTTAAGACTTTTACATTGTACTTGGATCAAATAGCTACATTAATCGGTCAGACAAGATTACGAATAACAA
AATTATGTGGAAATGACCCAGACAAAATAGTTGTCCCTTTAACCAAGGAACAAGTTAGACAAGCCTTTATCAATTCTGGT
GCATGGCAGATTGGTCTTGCTAATTTTGTGGGACTTATTGATAATCATTACCCAAAAACAAAGATCTTCCAGTTCTTAAA

ATTGACTACTTGGATTCTACCTAAAATTACCAGACGTGAACCTTTAGAAAATGCTCTAACAGTATTTACTGATGGTTCCA
GCAATGGAAAAGCAGCTTACACAGGGCCGAAAGAACGAGTAATCAAAACTCCATATCAATCGGCTCAAAGAGACGAGTTG
GTTGCAGTCATTACAGTGTTACAAGATTTTGACCAACCTATCAATATTATATCAGATTCTGCATATGTAGTACAGGCTAC
AAGGGATGTTGAGACAGCTCTAATTAAATATAGCATGGATGATCAGTTAAACCAGCTATTCAATTTATTACAACAAACTG
TAAGAAAAAGAAATTTCCCATTTTATATTACTTATATTCGAGCACACACTAATTTACCAGGGCCTTTGACTAAAGCAAAT
GAACAAGCTGACTTACTGGTATCATCTGCACTCATAAAAGCACAAGAACTTCATGCTTTGACTCATGTAAATGCAGCAGG
ATTAAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACATTGCACCCAGTGTCAAGTCTTACACC
TGCCCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCATTATGGCAAATGGATGTCACGCATGTACCT
TCATTTGGAAGATTATCATATGTTCATGTAACAGTTGATACTTATTCACATTTCATATGGGCAACTTGCCAAACAGGAGA
AAGTACTTCCCATGTTAAAAAACATTTATTGTCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCAAAACTGACAATG
GACCAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAACAGGAATTCCTTATAAT
TCCCAAGGACAGGCCATAGTTGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGAAGGGGGAGACAG
TAAGGAGTGTACCACTCCTCAGATGCAACTTAATCTAGCACTCTATACTTTAAATTTTTTAAACATTTATAGAAATCAGA
CTACTACTTCTGCAGAACAACATCTTACTGGTAAAAAGAACAGCCCACATGAAGGAAAACTAATTTGGTGGAAAGATAAT
AAAAATAAGACATGGGAAATAGGGAAGGTGATAACGTGGGGGAGAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAGCT
TCCTGTTTGGTTACCCACTAGACATTTGAAGTTCTACAATGAACCCATCGGAGATGCAAAGAAAAGGGCCTCCACGGAGA
TGGTAACACCAGTCACATGGATGGATAATCCTATAGAAGTATATGTTAATGATAGTATATGGGTACCTGGCCCCATAGAT
GATCGCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCGTTATCCTCCTATTTGCCT
AGGGAGAGCACCAGGATGTTTAATGCCTGCAGTCCAAAATTGGTTGGTAGAAGTACCTACTGTCAGTCCCATCAGTAGAT
TCACTTATCACATGGTAAGCGGGATGTCACTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTA
AAATTTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAGTTTTAGTTTGGGAAGA
ATGTGTGGCCAATAGTGCGGTGATATTATAAAACAATGAATTTGGAACTATTATAGATTGGGCACCTCGAGGTCAATTCT
ACCACAATTGCTCAGGACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGT
TTAGACAAACATAAGCATAAAAAATTGCAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATCTCTACCCCAAGACC
AAAAATAGTAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTACTGTGGCCTCACACCACATTAGAATTT
GGTCTGGAAATCAAACTTTAGAAACAAGAGATTGTAAGCCATTTTATACTGTCGACCTAAATTCCAGTCTAACAGTTCCT
TTACAAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAGACTATAACCTG
TGAAAATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATTGGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGG
GCGTGTGGATCCCTGTGTCCATGGACCGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAAAGGT
GTTTTAAATAGATCCAAAAGATTCATTTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAGTCACAGCTACGGCTGC
TGTAGCAGGAGTTGCATTGCACTCTTCTGTTCAGTCAGTAAACTTTGTTAATGATTGGCAAAAGAATTCTACAAGATTGT
GGAATTCACAATCTAGTATTGATCAAAAATTGGCAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGAC
AGACTCATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTTGTATTACACCCCAAATTTA
TAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAGGGAAGAGAAGATAATCTCACTTTAGACATTTCCA
AATTAAAAGAACAAATTTTCGAAGCATCAAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAATTGCAGGAGTTGCT
GATGGCCTCGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATTGGAAGTACATCGATTATAAATCTCATATTAATCCT
TGTGTGCCTGTTTTGTCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCA
TGATGACGATGGCGGTTTTGTCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAAATTGTTACTGTGTCT
GTGTAG

SEQ ID 108:

MGPLQPGLPSPAMIPKDWPLIIIDLKDCFFTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGMLNSPTICQTFVGR
ALQPVREKFSDCYIIHYIDDILCAAETKDKLIDCYTFLQAEVANAGLAIASDKIQTSTPFHYLGMQIENRKIKPQKIEIR
KDTLKTLNDFQKLLGDINWIRPTLGIPTYAMSNLFSILRGDSDLNSQRILTPEATKEIKLVEEKIQSAQINRIDPLAPLQ
LLIFATAHSPTGIIIQNTDLVEWSFLPHSTVKTFTLYLDQIATLIGQTRLRITKLCGNDPDKIVVPLTKEQVRQAFINSG
AWQIGLANFVGLIDNHYPKTKIFQFLKLTTWILPKITRREPLENALTVFTDGSSNGKAAYTGPKERVIKTPYQSAQRDEL
VAVITVLQDFDQPINIISDSAYVVQATRDVETALIKYSMDDQLNQLFNLLQQTVRKRNFPFYITYIRAHTNLPGPLTKAN
EQADLLVSSALIKAQELHALTHVNAAGLKNKFDVTWKQAKDIVQHCTQCQVLHLPTQEAGVNPRGLCPNALWQMDVTHVP
SFGRLSYVHVTVDTYSHFIWATCQTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYN
SQGQAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAEQHLTGKKNSPHEGKLIWWKDN
KNKTWEIGKVITWGRGFACVSPGENQLPVWLPTRHLKFYNEPIGDAKKRASTEMVTPVTWMDNPIEVYVNDSIWVPGPID
DRCPAKPEEEGMMINISIGYRYPPICLGRAPGCLMPAVQNWLVEVPTVSPISRFTYHMVSGMSLRPRVNYLQDFSYQRSL
KFRPKGKPCPKEIPKESKNTEVLVWEECVANSAVILXNNEFGTIIDWAPRGQFYHNCSGQTQSCPSAQVSPAVDSDLTES
LDKHKHKKLQSFYPWEWGEKGISTPRPKIVSPVSGPEHPELWRLTVASHHIRIWSGNQTLETRDCKPFYTVDLNSSLTVP
LQSCVKPPYMLVVGNIVIKPDSQTITCENCRLLTCIDSTFNWQHRILLVRAREGVWIPVSMDRPWEASPSVHILTEVLKG
VLNRSKRFIFTLIAVIMGLIAVTATAAVAGVALHSSVQSVNFVNDWQKNSTRLWNSQSSIDQKLANQINDLRQTVIWMGD
RLMSLEHRFQLQCDWNTSDFCITPQIYNESEHHWDMVRRHLQGREDNLTLDISKLKEQIFEASKAHLNLVPGTEAIAGVA
DGLANLNPVTWVKTIGSTSIINLILILVCLFCLLLVCRCTQQLRRDSDHRERAMMTMAVLSKRKGGNVGKSKRDQIVTVS
V

SEQ ID 109:

MNPSEMQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKLPSTKKAGPPTWAQLKKLTQLATKYLENTKVTQTPESMLLAA
LMIVSMVSAGVPNSSEETATIENGP

SEQ ID 110:

GAAAAAAATCAAAAAAAGAA

SEQ ID 111:

AGCCATTAATGCCATAA

SEQ ID 112:

TAAATAGGATCACTT

SEQ ID 113:

GGTGCGGAAATCACCATGCCCGCTCCAT

SEQ ID 114:
ATTATATAGCCCCACGAG
SEQ ID 115:
CAAGATGGGATATATACCAGG
SEQ ID 116:
AAAACAGAAAAACCGGTG
SEQ ID 117:
AAATCAGTGGCCGCTA
SEQ ID 118:
AGTTAGAAAAGGGTCAC
SEQ ID 119:
TGAGCCTTCGTTCTCA
SEQ ID 120:
AGGCAAATGGCATACGT
SEQ ID 121:
GGCCTCTCCAACCCG
SEQ ID 122:
GAGCAGGATTGTGAAAA
SEQ ID 123:
TCTTCAACCAGTGAGAGAAAA
SEQ ID 124:
ATTATATTGATGATATTTTA
SEQ ID 125:
AACGAAAGATAAATT
SEQ ID 126:
TGACTGTTATACATT
SEQ ID 127:
TTCATTATTTAGGGAT
SEQ ID 128:
AGATAGAAAATAGAAAAAT
SEQ ID 129:
ATTATTCAAAATACT
SEQ ID 130:
AATAACAAAATTATGT
SEQ ID 131:
AGACAAAATAGTTGT
SEQ ID 132:
TCCCTTTAACCAAGGAA
SEQ ID 133:
AAAAGAATGAGTCAT
SEQ ID 134:
CAGTATCACTTGACT
SEQ ID 135:
TTTTAATCAGTCTATTAACATTG
SEQ ID 136:
AAAGGATATTGAGAGA
SEQ ID 137:
CCTAATCAAATACATT
SEQ ID 138:
CGCTGTTTAATTTGT
SEQ ID 139:
TGCATTCATGGAAGCA
SEQ ID 140:
ACTCAGGAGGCAAGA
SEQ ID 141:
TTAAGAGACATTTATT
SEQ ID 142:
TAAAGCAGTTCAAAAA
SEQ ID 143:
AATAGGAATTCTCTA
SEQ ID 144:
AAAGCTCAATTGGTTA
SEQ ID 145:
ACGGACGATCATTTAA
SEQ ID 146:
MGQTKSKIKSKYASYLSFIKILLKRGGVKVSTKNLIKLFQIIEQFCPWFPEQGTLDLKDWKRIGKELKQAGRKGNIIPLT
VWNDWAIIKAALEPFQTEEDSVSVSDAPGSCIIDCNENTGKKSQKETEGLHCEYVAEPVMAQSTQNVDYNQLQEVIYPET
LKLEGKGPELVGPSESKPRGTSPLPAGQVPVTLQPQKQVKENKTQPPVAYQYWPPAELQYRPPPESQYGYPGMPPAPQGR
APYPQPPTRRLNPTAPPSRQGSKLHEIIDKSRKEGDTEAWQFPVTLEPMPPGEGAQEGEPPTVEARYKSFSIKKLKDMKE

GVKQYGPNSPYMRTLLDSIAHGHRLIPYDWEIQAKSSLSPSQFLQFKTWWIDGVQEQVRRNRAANPPVNIDADQLLGIGQ
NWSTISQQALMQNEAIEQVRAICLRAWEKIQDPGSTCPSFNTVRQGSKEPYPDFVARLQDVAQKSIADEKARKVIVELMA
YENANPECQSAIKPLKGKVPAGSDVISEYVKACDGIGGAMHKAMLMAQAITGVVLGGQVRTFGRKCYNCGQIGHLKKNCP
VLNKQNITIQATTTGREPPDLCPRCKKGKHWASQCRSKFDKNGQPLSGNEQRGQPQAPQQTGAFPIQPFVPQGFQGQQPP
LSQVFQGISQLPQYNNCPPPQAAVQQ

SEQ ID 147:

WATIVGKRAKGPASGPTTNWGIPNSAICSSGFSGTTTPTVPSVSGNKPVTTIQQLSPATSGSAAVDLCTIQAVSLLPGEP
PQKTPTGVYGPLPKGTVGLILGRSSLNLKGVQIHTSVVDSDYKGEIQLVISSSIPWSASPRDRIAQLLLLPYIKGGNSEI
KRIGGLGSTDPTGKAAYWASQVSENRPVCKAIIQGKQFEGLVDTGADVSIIALNQWPKNWPKQKAVTGLVGIGTASEVYQ
STEILHCLGPDNQESTVQPMITSIPLNLWGRDLLQQWGAEITMPAPSYSPTSQKIMTKMGYIPGKGLGKNEDGIKIPVEA
KINQEREGIGNPC

SEQ ID 148:

NKSRKRRNRESLLGAATVEPPKPIPLTWKTEKPVWVNQWPLPKQKLEALHLLANEQLEKGHIEPSFSPWNSPVFVIQKKS
GKWRMLTDLRAVNAVIQPMGPLQPGLPSPAMIPKDWPLIIIDLKDCFFTIPLAEQDCEKFAFTIPAINNKEPATRFQWKV
LPQGMLNSPTICQTFVGRALQPVREKFSDCYIIHCIDDILCAAETKDKLIDCYTFLQAEVANAGLAIASDKIQTSTPFHY
LGMQIENRKIKPQKIEIRKDTLKTLNDFQKLLGDINWIRPTLGIPTYAMSNLFSILRGDSDLNSKRMLTPEATKEIKLVE
EKIQSAQINRIDPLAPLQLLIFATAHSPTGIIIQNTDLVEWSFLPHSTVKTFTLYLDQIATLIGQTRLRIIKLCGNDPDK
IVVPLTKEQVRQAFINSGAWKIGLANFVGIIDNHYPKTKIFQFLKLTTWILPKITRREPLENALTVFTDGSSNGKAAYTG
PKERVIKTPYQSAQRAELVAVITVLQDFDQPINIISDSAYVVQATRDVETALIKYSMDDQLNQLFNLLQQTVRKRNFPFY
ITHIRAHTNLPGPLTKANEQADLLVSSALIKAQELHALTHVNAAGLKNKFDVTWKQAKDIVQHCTQCQVLHLPTQEAGVN
PRGLCPNALWQMDVTHVPSFGRLSYVHVTVDTYSHFIWATCQTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAF
QKFLSQWKISHTTGIPYNSQGQAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAEQHL
TGKKNSPHEGKLIWWKDNKNKTWEIGKVITWGRGFACVSPGENQLPVWIPTRHLKFYNEPIRDAKKSTSAETETSQSSTV
DSQDEQNGDVRRTDEVAIHQEGRAANLGTTKEADAVSYKISREHKGDTNPREYAACSLDDCINGGKSPYACRSSCS

SEQ ID 149:

MNPSEMQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKLPSTKKAEPPTWAQLKKLTQLATKYLENTKVTQTPESMLLAA
LMIVSMVVSLPMPAGAAAANYTYWAYVPFPPLIRAVTWMDNPTEVYVNDSVWVPGPIDDRCPAKPEEEGMMINISIGYHY
PPICLGRAPGCLMPAVQNWLVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQRSLKFRPKGKPCPKEIPKESKNTEV
LVWEECVANSAVILQNNEFGTIIDWAPRGQFYHNCSGQTQSCPSAQVSPAVDSDLTESLDKHKHKKLQSFYPWEWGEKGI
STPRPKIVSPVSGPEHPELWRLTVASHHIRIWSGNQTLETRDRKPFYTIDLNSSLTVPLQSCVKPPYMLVVGNIVIKPDS
QTITCENCRLLTCIDSTFNWQHRILLVRAREGVWIPVSMDRPWEASPSVHILTEVLKGVLNRSKRFIFTLIAVIMGLIAV
TATAAVAGVALHSSVQSVNFVNDWQKNSTRLWNSQSSIDQKLANQINDLRQTVIWMGDRLMSLEHRFQLQCDWNTSDFCI
TPQIYNESEHHWDMVRRHLQGREDNLTLDISKLKEQIFEASKAHLNLVPGTEAIAGVADGLANLNPVTWVKTIGSTTIIN
LILILVCLFCLLLVCRCTQQLRRDSDHRERAMMTMAVLSKRKGGNVGKSKRDQIVTVSV

SEQ ID 150:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTAT
GTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTG
CTGTGTCCACTCAGGGTTAAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTC
CTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAG
GAAAGCCAGGTATTGTCCAACGTTTCTCCCCATGTGATAGCCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGT
CCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA
AGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACT
GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTA
TGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCACATGTTTGTCT
GCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCACAGATGATCAGTAAATA
CTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTCTTTCTTTCTCTATAC
TTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA
CCCCTACA

SEQ ID 151:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTCACTGTATCTGTGTAGAAAGAAGTAGACATGGGAGACTCCATTTTGTTAT
GTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCA
AACTCAGGGTTAAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGA
GTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAACACTGCGGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCC
AGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCCAG
CCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGCATGCCTCTTGCAGTTGAGACAAGAGGAA
GGCATCTGTCTCCTGCCCGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCGGATTGTACGTTCCATCTACTGAGATAG
GGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTTTAAAGCATTGAGATGTTTATGTGTAT
GCATATCTAAAAGCACAGCACTTAATCCTTTACCTTGTCTATGATGCAAAGATCTTTGTTCACGTGTTTGTCTGCTGACC
CTCTCCCCACTATTGTCTTGTGACCCTGACACATCCCCCTCTCGGAGAAACACCCACGAATGACCAATAAATACTAAAGG
GAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGCCCCCTTATTTCTTTCTCTACACTTTGTC
TCTGTGTCTTTTTCTTTCCTAAGTCTCTCGTTCCACCTTACGAGAAACACCCACAGGTGTGGAGGGGCAACCCACCCCTA
CA

SEQ ID 152:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATGGGAGACTCCATTTTGTTAT
GTGCTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTG
CTGTGTCAACTCAGGGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTC
CTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAG
GAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCAT
CCCCCAGCCCGACACCCATAAAGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGCATGCCTCTTGCAGTTGAGACA

AGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACT
GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCCTTGTAAAGCATTGAGATGTTTA
TGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCACGTGTTTGTCT
GCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACACCCACAGATGATCAATAAATA
CTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCTTTCTCTATAC
TTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA
CCCCTACA

SEQ ID 153:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACAGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTCT
GTACTAAGAAAAATTCTTCTGCCTTGAAATTCTGTTAATCTATAACCTTACCCCCAACCCCGTGCTCTTTGAAACATGTG
CTGTGTCAACTCAGAGTTAAATGGATTAAGTGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTC
CTTGAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCTCAGGGACCTCTGCCTAG
GAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCAT
CCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAACGCCTCTTGCAGTTGAGACA
AGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAATGTCCCGGTATAAAACCCGATTGTATGCTCCATCTACT
GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGCTGTTTA
TGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCACGTGTTTGTCT
GCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCACGAATGATGAATAAATA
CTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTACTTCTTTCTCTGTAC
TTTGTCTCTGTGTCTTTTTCTTTCCTAAGTCTCTCGTTCCACCTTACGAGAAATACCCACAGGTGTGGAGGGGCAACCCA
CCCCTACA

SEQ ID 154:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTCT
GTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATAACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTG
CTATGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCACGCTC
CTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAG
GAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCAT
CCCCCAGCCCGACACCTGTAAAGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGCATGCCTCTTGCAGTTGAGACA
AGAGGAAGGCATCTGTCTCCTGCCCGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTATGTTCCATCTACT
GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTA
TGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACCTTGTCTATGATGCAAAGACCTTTGTTCACGTGTTTGTCT
GCTGACCCTCTCCCCACGATTGTCTTGTGACCCTGACACATCCCCGTCTTCGAGAAACACCCACGAATGATCAATAAATA
CTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCAGGTCCCCTTATTTCTTTCTCTATAC
TTTGTCTCTGTGTCTTTTTCTTTTCCAAGTCTCTCGTTCCATCTTACGAGAAACACCCACAGGTGTGGAGGGGCAACCCA
CCCCTACA

SEQ ID 155:

GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTA
TGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCAC

SEQ ID 156:

ATGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCACAGATGA
TCAGTAAATACTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTCTTTCT
TTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAG
GGGCAACCCACCCCTACA

SEQ ID 157:

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXACATTTGAAGTTCTACA
ATGAACCCATCXGAGATGCAAAGAAAXXXXXXXXXXAGCXCCTCCXCGGAGACGGAAACACCGCAATCGAGCAXCXXXXX
XXXXXXXXXTXGACTCACAAGATGAAXAAAATGGTGAXXTCAGAAGAACAGATGAAGTTGCCATCCACCAAGAAXGCXGA
GCCGCCGACTTGGGCACAAXTAAAGAAGCTGACACAGTTAGCTAXAAAAXXXXXXCTXGAGAACACAAAGGTGACACAAA
CTCCAGAGAXTATGCTGCTTGCAGCTTTGATGATTGTATCAATGGTGGTAAGTCTCCCXATGCCTGCAGGAGCAGCTGCA
GCTAAXTATACXTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGGATAATCCTATTGAAGT
ATATGTTAATAATAGTGTATGGGXTACCTGGCCCCACAGATGATCGTTGCCCTGCCAAACCTGAGGAAGAAGGAATGATG
ATAAATATTTCCATTGGGTATCXTTATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAATXGCCTGCAXTCCAAA
ATTGGTTGGTAGAAGTACCTACTGTCAGTXCCAXCAGTAGATTCACTTATCACATGGTAAGXGGXATGTCACTCAGGCCA
CXGGTAAATXATTTACAXGACTTTTCTTATCAAAGATCATTAAAATTTAGXCCTAAAGGGAAACCTTGCCCCAAGGAAAT
TCCCAAAGXATCAAAAXAXXCAGAAGTTTTAGTTTGGGAAGAATGTGTGGCXAATAGTGCXGTGATATTACAAAACAATG
AATTTGGAACTATTATAGATTGGGCACCTCGAGGTCAATTCTAXCACAXXXXXXXXXXXXXXXXXXXXXXXXATTGCXCA
GGXCAAACTCAXTCXTGTCCXAGXGCACAAGXXXXXXXXXXXXXXAGTCCAGCTGTTGATAGXGACTTAACAGAAAGTXT
AGACXAAXXTXAXXXTAXAAXXTTAXAXTCXXTCTAXCCXTGGXAATGGGGXGAAAAXGGAATXTCXXCXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXCCXXGACCAAAXXTAXTXAGTCCTGTTXCTGGTCCTGAACATCCAGAATTATGGAXGCTTACTGTGGCC
TCAXXACCACATTAGAATTTGGTCTGGAAATCAAXCTXTAGAAACAAGAGATCXTAAGCCATXTTATACTATCXACCTAA
ATTCCAGTCTXACAXTTCCTTTXCAAAGTTGXGTAAAGCCCCCTTATATXGCTAGTTGTAGGAAATAXXTAGTTATTAAA
CCAGAXTCCCAAACTATAXXACCTGTGAAAATTGTAGATTGTTTACTTGCATTGATTCAACTTTTAATTGGCAGCACCGT
ATTCTGCTXGTGAGAGCAAGAGAXGGXGTGTGGATCCCTGTGTCCATGGACCGACCGTGGGAGGCXTCXCCATCCXTCCA
TATTTTXACXGAAGTATTAAAAGGXXTTXTAAXTAGATCCAAAAGATTCATTTTTACTTTAATTGCAGTGATTATGGGXX
TXATTGCAGTCACAGCTACXGCTGCXGXXGCXGGAXTTGCXTTXCACTCXTCTGTTCAXXCXGAXAXTXTGTXAATXAT

TGGCAAAAXAAXTTCXXCAAXATTGTGGAATTCXCAXAXCXXXXATXGATCAAAAATTGGCAAATCAAATTAATGATCTT
AGACAAACTGTCATTTGGATGGGAGAXAGXCTCATGAGCTTXGAAXATCXTTTXCAGTTACAXTGTGACTGGAATACGTC
AGATTTTTGTATTACACCXCAAXXXTATAATGAGTCTGAGCATCACTGGGACATGGTTAGAXGCCATCTACAXGGAAGAG
AAGATAATCTXACTTTAGACATTTCXAAATTAAAAGAAXXXXXXXXXXXXXXXCAAATTTTXXAAXCATCAAAAGCCCATT
TAAATTTGGTGCCAGGAACTGAGGCAATXGXXXXAGXTGCTGATGGCCTCXCAAATCTTAACCCTGTCACTTGGGTTAAX
ACCATXXGAAGTXCXACXATTXTAAATXTCATATTAATCCTTGTXTGCCTGTTXTGTCTGTTGTTXXAGTCTXCAGGTGT
AXCCAXCAGCTCCGAAGAGACAGCGACCAXCXAGAACGGGCCATGATGACGATGGXGGTTTTGTCXAAAAGAAAAGGGGG
XXAXATGTXGGGAAAAGXXAGAGAGATCAGAXTGTTACTGTXGTCTXTGTAGAAAXAXGXAGACATAXGAGACTCCATTT
TGXXXTGTACXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

SEQ ID 158:

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCPWFPEQGXLDLXDWKRIGXELKQAGRKGN
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXDAPGSCIIDCNEXTXKKSQKETEXLHCEYVXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAGQVXVTLQPQXQVKENKTQXPVAYQYWPPXXXXXXXXXXSQYGYXGMPP
AXQXRXPYPQPPTXRXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCXSKFDKXGQPLSGNXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

SEQ ID 159:

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXDXLAPLQLLIFATAHSXTGIIIQNTDLVEWSFLPHSTVKTFTLYLDQMATLIGQXRLRIIXLCGNDPDKI
XVPXXKXQVRQAFIXSGAWXIGLANFLGIIDNHYPKTKIFQFLKLTTWILPKITRREPLENALTVFTDGSSNGKAAYTGP
KERVIKTPYQSAQRAELVAVITVLQDFDQPINIISDSAYVVQATRDVETALIKYSXDDXLNQLFNLLQQTVRKRNFPFYI
THIRAHTNLPGPLTKANEQADLLVSSAXIKAQELXALTHVNAAGLKNKFDVTWKQAKDIVQHCTQCQVLHLXTQEAGVNP
RGLCPNALWQMDXTHVXSFGRLSYVHVTVDTYSHFIWATCQTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQ
KFLSQWKISHTTGIPYNSQGQAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAXQHLT
GKKXSPHEGKLIWWKDXKNKTWEIGKVITWGRGFACVSPGENQLPVWIPTRHLKFYNEPIXDAKKXXSXEXXTXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXTIXXXXXXXXXXXXXXXXXXX
XXXXXXXXAXXXDXXXXXXKXXXXXXXXXXPXEWGXXXIXXXXXXXXSPXSXXXXXXXXXXXXXXXXXXXXXXXX

SEQ ID 160:

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXLIXXVTWMDNPXEVYVNDSVWVPGPXDDXCPAKPEEEGMMINISIXYXYPPICLGRAPGCLMPAVQN
WLVEVPTVSPXXRFTYHMVSGMSLRPRVNXLQDFSYQRSLKFRPKGKPCPKEIPKESKNTEVLVWEECVANSXVILQNNE
FGTIIDWAPRGQFYHNCSGQTQSCXSAQVSPAVDSDLTESLDKHKHKKLQSFYPWEWGEKGISTPRPXIISPVSGPEHPE
LWXXXXXXXXRIXXXXXXXXXXXXXXXXXXXXLNSXLTVPLQSCVKPXXXXXXXXXXXXXXXXXXXXXXXXXXXXXDSTX
XWXXXIXLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXRCT
QQLRRDSDXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

SEQ ID 161:

TAGGCCTTTGAGGGA

SEQ ID 162:

TAGGCCTTATTTTAGGG

SEQ ID 163:

GAGAAGGAGCCCAAGAG

SEQ ID 164:

GAGCCTCCCACAGTT

SEQ ID 165:

AGGCCAGATACAAGTCT

SEQ ID 166:

TTTTCGATAAAAATGCTA

SEQ ID 167:

TTATATGAGGACATTA

SEQ ID 168:

TTATGGACATAGACTCAT

SEQ ID 169:

TTGGGAGATTCTGGCAAA

SEQ ID 170:

AATCGTCTCTCTCACC

SEQ ID 171:

AATTTTTACAATTTAAGACT
SEQ ID 172:
GTCCGAAGAAATAGG
SEQ ID 173:
TGCCAATCCTCCAGTT
SEQ ID 174:
AACATAGATGCAGATCAACTAT
SEQ ID 175:
AGTACTATTAGTCAACAA
SEQ ID 176:
GTCAACAAGCATTAATGCAA
SEQ ID 177:
CCATTGAGCAAGTTAGAG
SEQ ID 178:
GAGCTATCTGCCTTAGAG
SEQ ID 179:
CTTGGGAAAAAATCCAAGAC
SEQ ID 180:
GAAGTACCTGCCCCTCATTTAA-TACAGTAA
SEQ ID 181:
CCCTACCCTGATTTT
SEQ ID 182:
AAGGCTCCAAGATGTT
SEQ ID 183:
TCAATTGCCGATGAAAAAG
SEQ ID 184:
CGGTAAGGTCATAGTGG
SEQ ID 185:
TGGAGTTGATGGCATAT
SEQ ID 186:
AAACGCCAATCCTGAGT
SEQ ID 187:
TCAATCAGCCATTAA
SEQ ID 188:
AAAGGTTCCTGCAGGATCAGA
SEQ ID 189:
AGGATCAGATGTAATCTCA
SEQ ID 190:
AATATGTAAAAGCCTGT
SEQ ID 191:
ATAAAGCTATGCTTAT
SEQ ID 192:
AATAACAGGAGTTGTTTTAG
SEQ ID 193:
ACATTTGGAGGAAAAT
SEQ ID 194:
ATTGGTCACTTAAAAAA
SEQ ID 195:
ATTGGTCACTTAAAAAA
SEQ ID 196:
GGTAGAGAGCCACCTGACTTAT
SEQ ID 197:
AAGATGTAAAAAAGG
SEQ ID 198:
GCTAGTCAATGTCGTT
SEQ ID 199:
GGGAAACGAGCAAAG
SEQ ID 200:
CCAATTCAGCCATTTG
SEQ ID 201:
CCACTGTCCCAAGTGTTTC
SEQ ID 202:
AATAAGCCAGTTACCA
SEQ ID 203:
ACAATACAACAATTG
SEQ ID 204:
CTCACCACAAGCGGCAGTGCAGC
SEQ ID 205:
TACTATACAAGCAGTCTCTCTGCTTCCAGGGGAGC

SEQ ID 206:
AAAAAATCCCTACAGG
SEQ ID 207:
CACTGCCTGAGGGGACTG
SEQ ID 208:
GACTAATCTTGGGAAGA
SEQ ID 209:
AAATCTAAAAGGAGTTCA
SEQ ID 210:
CTAGTGTGGTTGATTCAGACT
SEQ ID 211:
CGAAATTCAATTGGTTATTA
SEQ ID 212:
TCTTCAATTCCTTGG
SEQ ID 213:
AGTCCAAGAGACAGGAT
SEQ ID 214:
TTATTACTCCTGCCATATA
SEQ ID 215:
CATTAGAAAAAGGACATTG
SEQ ID 216:
TTGGAATTCTGTTTGTA
SEQ ID 217:
TAACTGAGCCATTAAT
SEQ ID 218:
AGCCATGGTCCCCTTTAATTA
SEQ ID 219:
TTTTACCACACCAGCCT
SEQ ID 220:
TTGTCAGCTCAAGCT
SEQ ID 221:
TACATCGTTCACTAT
SEQ ID 222:
TTAAAAGCATTAAAT
SEQ ID 223:
AGAAGTCCCAATTGAGG
SEQ ID 224:
GGTCTTGCCGATTTT
SEQ ID 225:
ACAATCGTTACCACA

Claims

1. A composition comprising an isolated HERV-K (CH) gag and/ or pol expression product.

2. The composition of claim 1, wherein the expression product is a polypeptide or a polynucleotide.

3. The composition of claim 2, wherein the polynucleotide:

i. is an isolated polynucleotide comprising: (a) the nucleotide sequence of any of SEQ IDs 7-10; (b) the nucleotide sequence of any of SEQ IDs 27-39; (c) the complement of a nucleotide sequence of any of SEQ IDs 7-10; or (d) the complement of the nucleotide sequence of any of SEQ IDs 27-39;
ii. is a fragment of at least 7 nucleotides of: (a) a nucleotide sequence shown in SEQ IDs 7-10; (b) the nucleotide sequence shown in any of SEQ IDs 27-39; (c) the complement of a nucleotide sequence shown in SEQ IDs 7-10; or (d) the complement of the nucleotide sequence shown in any of SEQ IDs 27-39;
iii. has the formula 5'-A-B-C-3', wherein: A is a nucleotide sequence consisting of at least 1 nucleotide; B is a nucleotide sequence consisting of a fragment of at least 7 nucleotides from (i) the nucleotide sequence shown in SEQ IDs 7-10, (ii) the nucleotide sequence shown in any of SEQ IDs 27-39, (iii) the complement of the nucleotide sequence shown in SEQ IDs 7-10, or (iv) the complement of the nucleotide sequence shown in any of SEQ IDs 27-39; C is a nucleotide sequence consisting of at least 1 nucleotide; and wherein said polynucleotide is not a fragment of (i) the nucleotide sequence shown in SEQ IDs 7-10, (ii) the nucleotide sequence shown in any of SEQ IDs 27-39, (iii) the complement of the nucleotide sequence shown in SEQ IDs 7-10, or (iv) the complement of the nucleotide sequence shown in any of SEQ IDs 27-39;

iv. has at least 50% identity to : (a) SEQ IDs 7-10; (b) a fragment of 20 nucleotides of SEQ IDs 7-10; (c) SEQ IDs 11-13; (d) a fragment of 20 nucleotides of SEQ IDs 11-13;
v. comprises a polynucleotide that selectively hybridizes, relative to a known polynucleotide, to: (a) the nucleotide sequence shown in SEQ IDs 7-10; (b) the nucleotide sequence shown in any of SEQ IDs 27-39; (c) the complement of the nucleotide sequence shown in SEQ IDs 7-10; (d) the complement of the nucleotide sequence shown in any of SEQ IDs 27-39; (e) a fragment of the nucleotide sequence shown in SEQ IDs 7-10; (f) a fragment of the nucleotide sequence shown in any of SEQ IDs 27-39; (g) the complement of a fragment of the nucleotide sequence shown in SEQ IDs 7-10; (h) the complement of a fragment of the nucleotide sequence shown in any of SEQ IDs 27-39; (j) a nucleotide sequence shown in SEQ IDs 14-39; or (k) polynucleotides found in ATCC deposits having ATCC accession numbers given in Table 7; or
vi. comprises any of SEQ IDs 14-26.

4. The composition of claim 2, wherein the polypeptide:

i. is encoded by a polynucleotide sequence according to claim 3, or a variant thereof;
ii. is (a) encoded within a HERV-K(CH) open reading frame; (b) encoded by a polynucleotide shown in SEQ ID 11, 12 or 13; or (c) comprising an amino acid sequence as shown in any one of SEQ IDs 46-49, 50-55, 56-57 or 58;
iii. comprises a fragment of at least 5 amino acids in length of: (a) a polypeptide sequence encoded within a HERV-K(CH) open reading frame; (b) a polypeptide sequence encoded by a polynucleotide shown in SEQ ID 11, 12 or 13; or (c) an amino acid sequence as shown in any one of SEQ IDs 46-49, 50-55, 56-57 or 58;
iv. is an isolated polypeptide having formula 5'-A-B-C-3', wherein: A is an amino acid sequence consisting of at least 1 amino acid; B is an amino acid sequence consisting of a fragment of at least 7 amino acids from (i) the amino acid sequence encoded by a polynucleotide shown in SEQ ID 11, 12 or 13; (ii) any one of SEQ IDs 46-49, 50-55, 56-57 or 58; C is an amino acid sequence consisting of at least 1 amino acid; and wherein said polypeptide is not a fragment of the amino acid sequence defined in (i) or (ii); or
v. has at least 50% identity to: (a) the polypeptide sequences encoded by SEQ IDs 7-39; (b) a fragment of at least 20 amino acids of the polypeptide sequences encoded by SEQ IDs 7-39; (c) the polypeptide sequences SEQ IDs 46-58; or (d) a fragment of at least 20 amino acids of the polypeptide sequences SEQ IDs 46-58.

5. A composition of any of claim 2 or claim 4 where the polypeptide comprises an antigenic region corresponding to any of:

i. amino acids 1-40; 45-60; 80-105; 130-145; 147-183; 186-220; 245-253; 255-288, or 1-8; 2-10; 1-15; 5-15; 7-15; 10-20; 12-20; 15-23; 20-28; 28-35; 30-37; 33-40; 1-20; 20-40; 1-15; 15-30; 15-40; 45-52; 50-57; 55-62; 50-60; 1-60; 80-87; 85-92; 80-90; 90-97; 95-102; 98-105; 85-100; 90-105; 80-100; 85-105; 130-137; 135-142; 140-147; 145-152; 150-157; 155-162; 160-167; 165-172; 170-177; 175-183; 180-187; 185-192; 190-197; 195-202; 200-207; 205-212; 210-217; 213-220; 185-220; 190-220; 195-220; 200-220; 205-220; 255-262; 260-267; 265-272; 270-277; 275-282; 280-288; 245-288; 250-288; 260-288; 265-288; 270-288 of SEQ ID 56;
ii. amino acids 1-40; 80-105; 145-180; 185-225; 240-335, or 1-8; 2-10; 1-15; 5-15; 7-15; 10-20; 12-20; 15-23; 20-28; 28-35; 30-37; 33-40; 1-20; 20-40; 1-15; 15-30; 15-40; 80-87; 85-92; 80-90; 90-97; 95-102; 98-105; 85-100; 90-105; 80-100; 85-105; 145-152; 150-157; 155-162; 160-167; 165-172; 170-177; 175-182; 180-187; 185-192; 190-197; 195-202; 200-207; 205-212; 210-217; 215-212; 218-225; 145-160; 150-165; 155-170; 160-175; 170-185; 180-225; 185-225; 190-225; 195-225; 200-225; 205-225; 210-225; 215-225; 240-247; 245-252; 250-257; 255-262; 260-267; 265-272; 270-277; 275-282; 280-287; 285-292; 290-297; 295-302; 300-307; 305-312; 310-317; 315-322; 320-327; 325-332; 328-335; 245-285; 250-285; 260-285; 265-285; 270-295; 275-300; 280-305; 285-310; 295-315; 300-320; 305-325; 325-335; 245-335; 250-335; 255-335; 260-335; 270-335; 275-335; 280-335; 285-335; 290-335; 295-335; 305-335; 310-335; 315-335; 320-335 of SEQ ID 57; or
iii. amino acids: 1-10; 15-35; 45-55; 60-85; 100-115; 125-140; 170-190; 195-215; 230-268; or 1-8; 2-10; 1-15; 5-15; 7-15; 10-20; 12-20; 15-23; 20-28; 28-35; 15-30; 15-40; 20-30; 45-52; 48-55; 60-68; 60-70; 65-73; 70-78; 75-83; 70-80; 65-75; 68-75; 75-85; 78-85; 65-85; 60-75; 100-108; 103-110; 105-113; 108-115; 125-133; 128-135; 132-140; 170-178; 175-182; 180-187; 182-190; 195-202; 200-208; 205-212; 208-215; 230-237; 235-242; 240-247; 245-252; 250-257; 255-262; 260-268; 230-250; 235-255; 240-260; 245-268; 230-245; 235-245; 235-250; 240-255; 245-260; 250-268; 15-55; 170-215; 45-85 of SEQ ID 58.

6. The composition of any preceding claim further comprising a pharmaceutically acceptable carrier.

7. The composition of any preceding claim, wherein the composition is immunogenic.

8. The composition of any preceding claim, wherein the composition is a vaccine composition and optionally further comprises a vaccine adjuvant.

9. The composition of any of claims 6-8 for use in a method of raising an immune response in a patient, wherein the method comprises administering an immunogenic dose of the composition to an animal, preferably a human.

10. The composition of any of claims 6-9 for use in treating prostate cancer.

11. A method of testing the efficacy of a composition of any of claims 6-10, the method comprising monitoring the expression of a polynucleotide and/or polypeptide as defined in any of claims 3-5, in a patient sample, after administering an immunogenic dose of said composition.

12. A method of diagnosing prostate cancer, the method comprising the step of detecting the presence or absence of a HERV-K (CH) gag and/ or pol expression product in a patient sample.

13. The method of claim 11 or 12, wherein said sample is a prostate sample or a blood sample.

14. The method of any of claims 11-13, wherein the expression product is a polynucleotide or a polypeptide as defined in according to any of claims 3-5.

15. The method of any of claims 11-14, wherein said method comprises the use of at least one probe, wherein said probe comprises a nucleotide sequence selected from the SEQ IDs 161, 215-225 or 133-144.

FIGURE 1

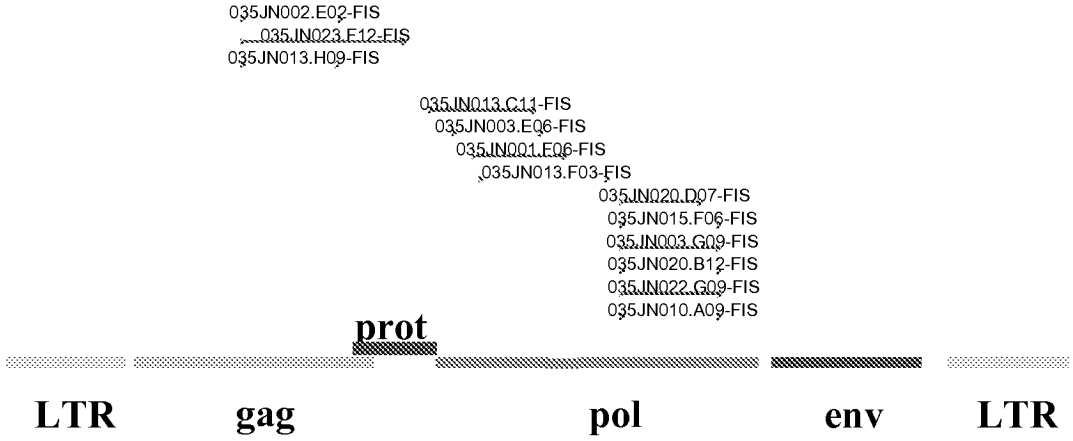


FIGURE 2

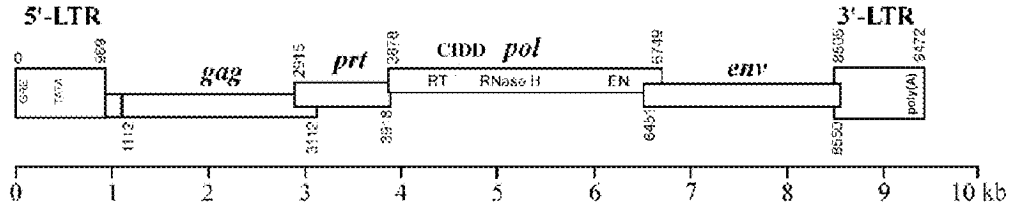


FIGURE 3

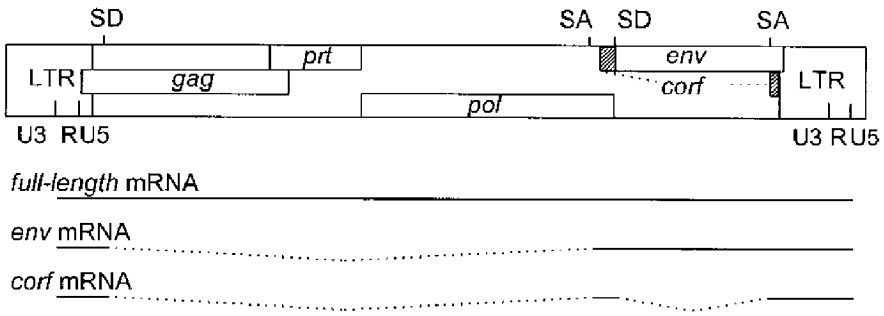


FIGURE 4

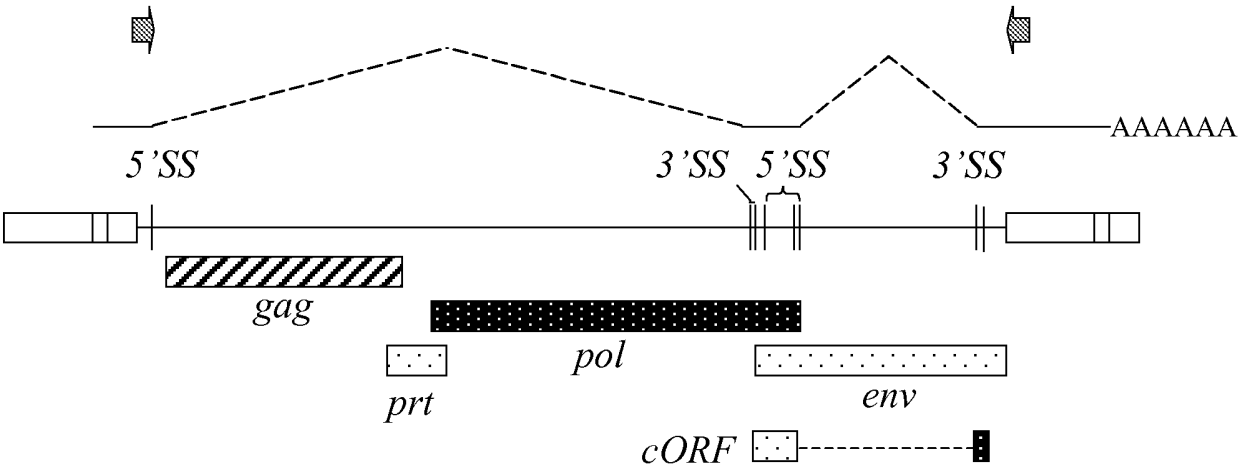


FIGURE 5

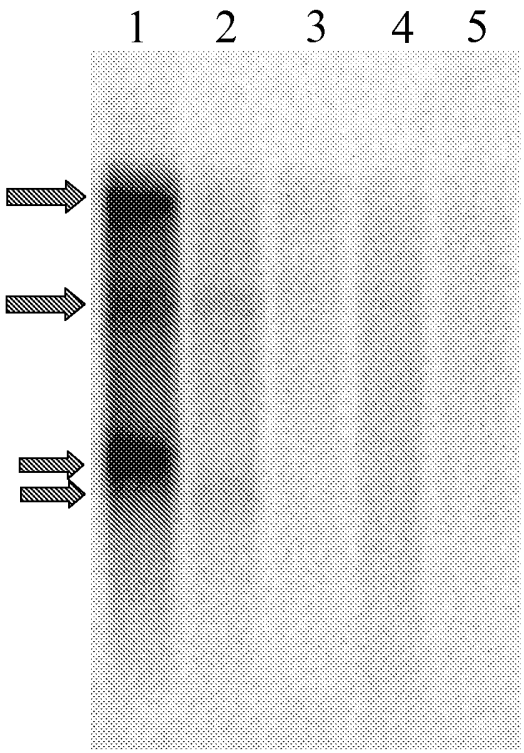


FIGURE 6

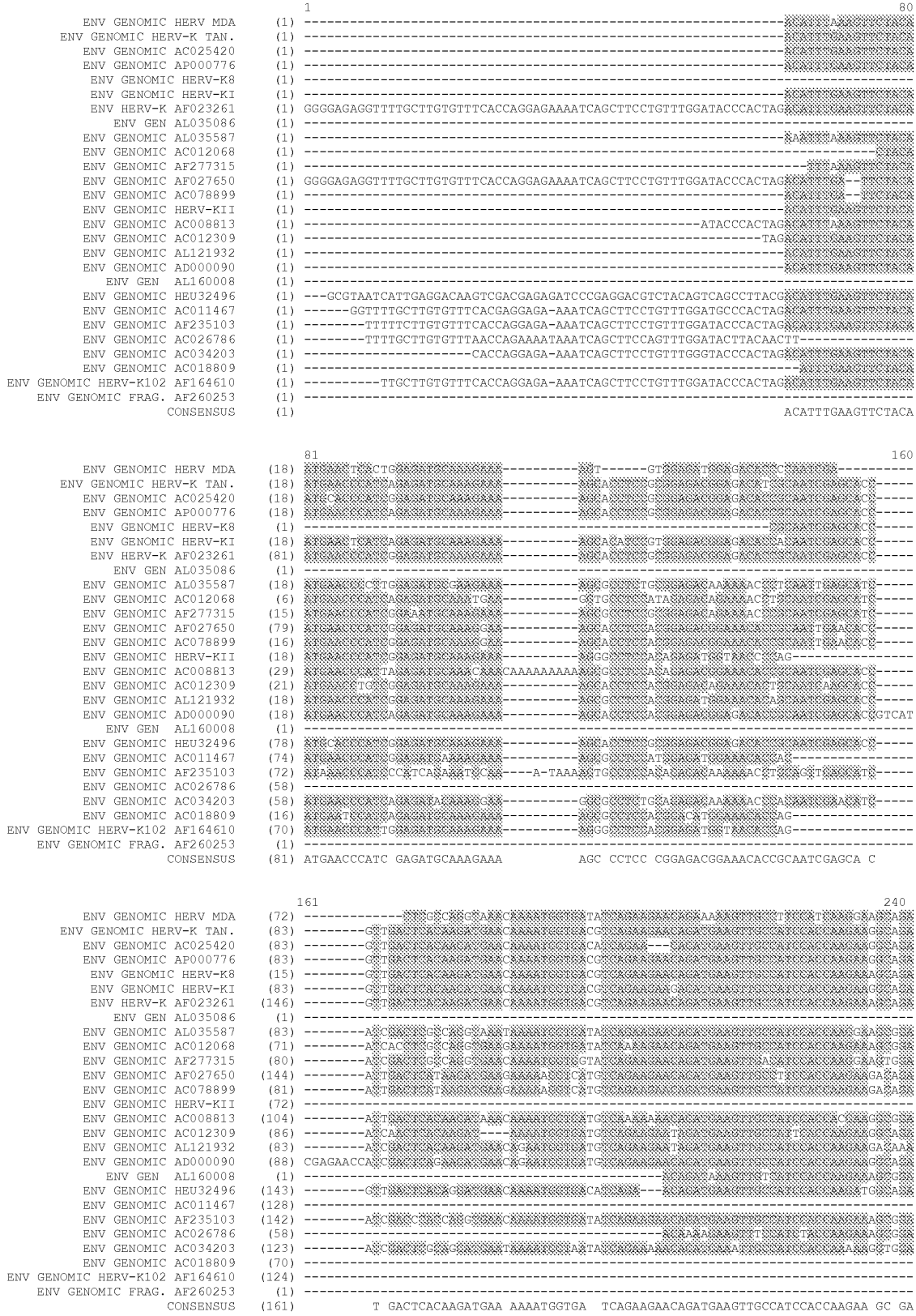

ENV GENOMIC HERV MDA
ENV GENOMIC HERV-K TAN.
ENV GENOMIC AC025420
ENV GENOMIC AP000776
ENV GENOMIC HERV-K8
ENV GENOMIC HERV-KI
ENV HERV-K AF023261
ENV GEN AL035086
ENV GENOMIC AL035587
ENV GENOMIC AC012068
ENV GENOMIC AF277315
ENV GENOMIC AF027650
ENV GENOMIC AC078899
ENV GENOMIC HERV-KII
ENV GENOMIC AC008813
ENV GENOMIC AC012309
ENV GENOMIC AL121932
ENV GENOMIC AD000090
ENV GEN AL160008
ENV GENOMIC HEU32496
ENV GENOMIC AC011467
ENV GENOMIC AF235103
ENV GENOMIC AC026786
ENV GENOMIC AC034203
ENV GENOMIC AC018809
ENV GENOMIC HERV-K102 AF164610
ENV GENOMIC FRAG. AF260253
CONSENSUS
ACATTTGAAGTTCTACA

FIGURE 6 CONTD...

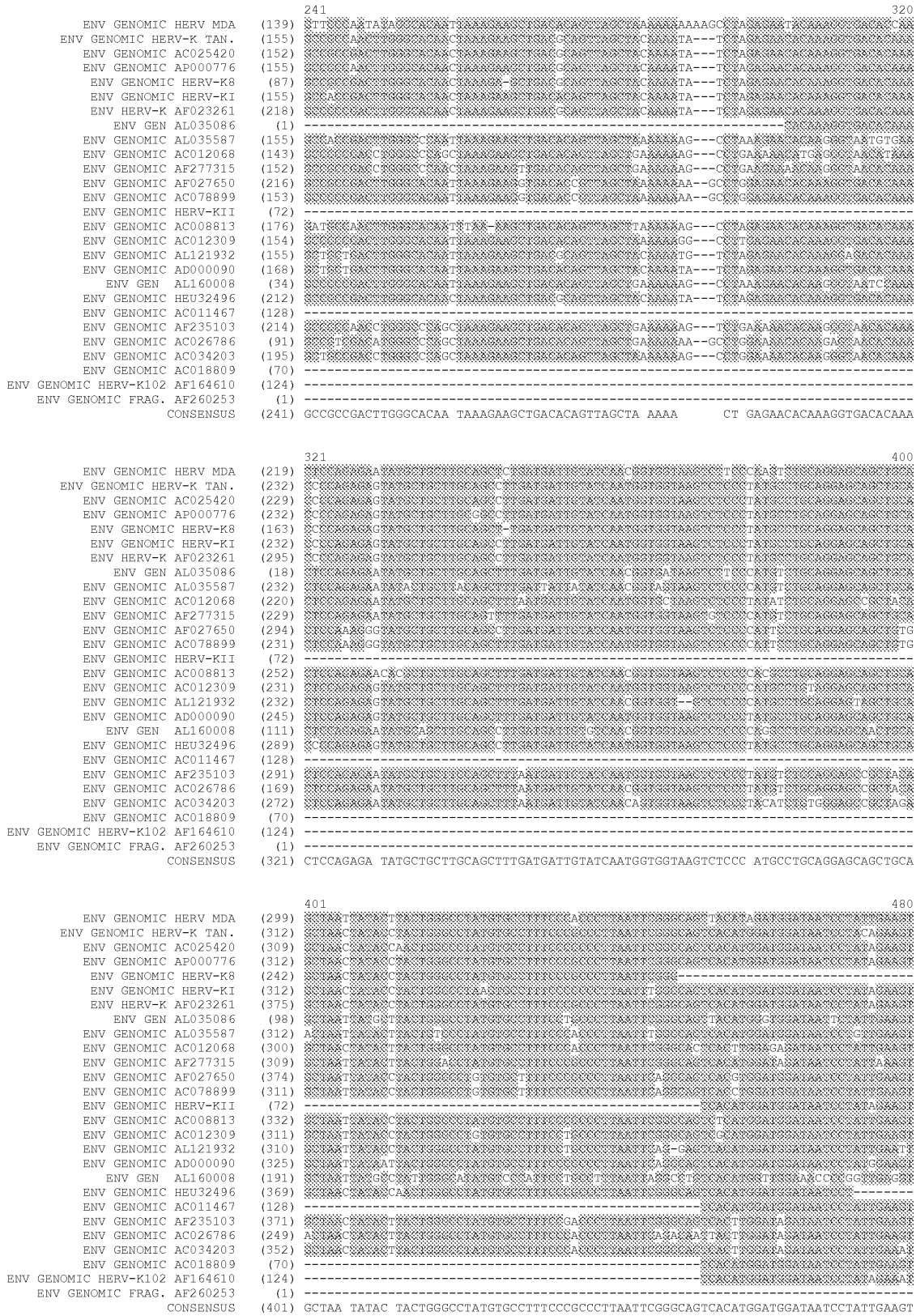


FIGURE 6 CONTD...

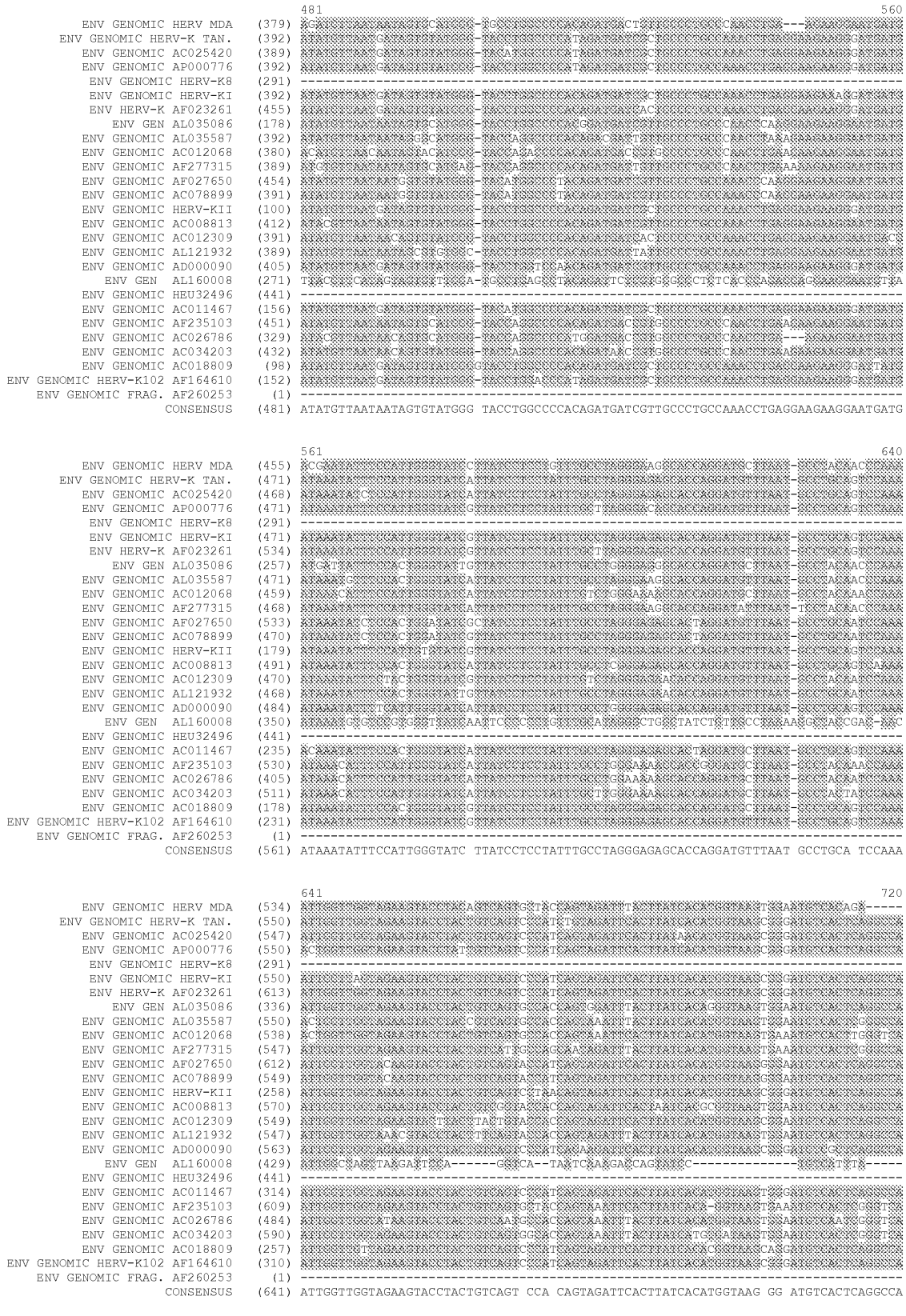


FIGURE 6 CONTD...

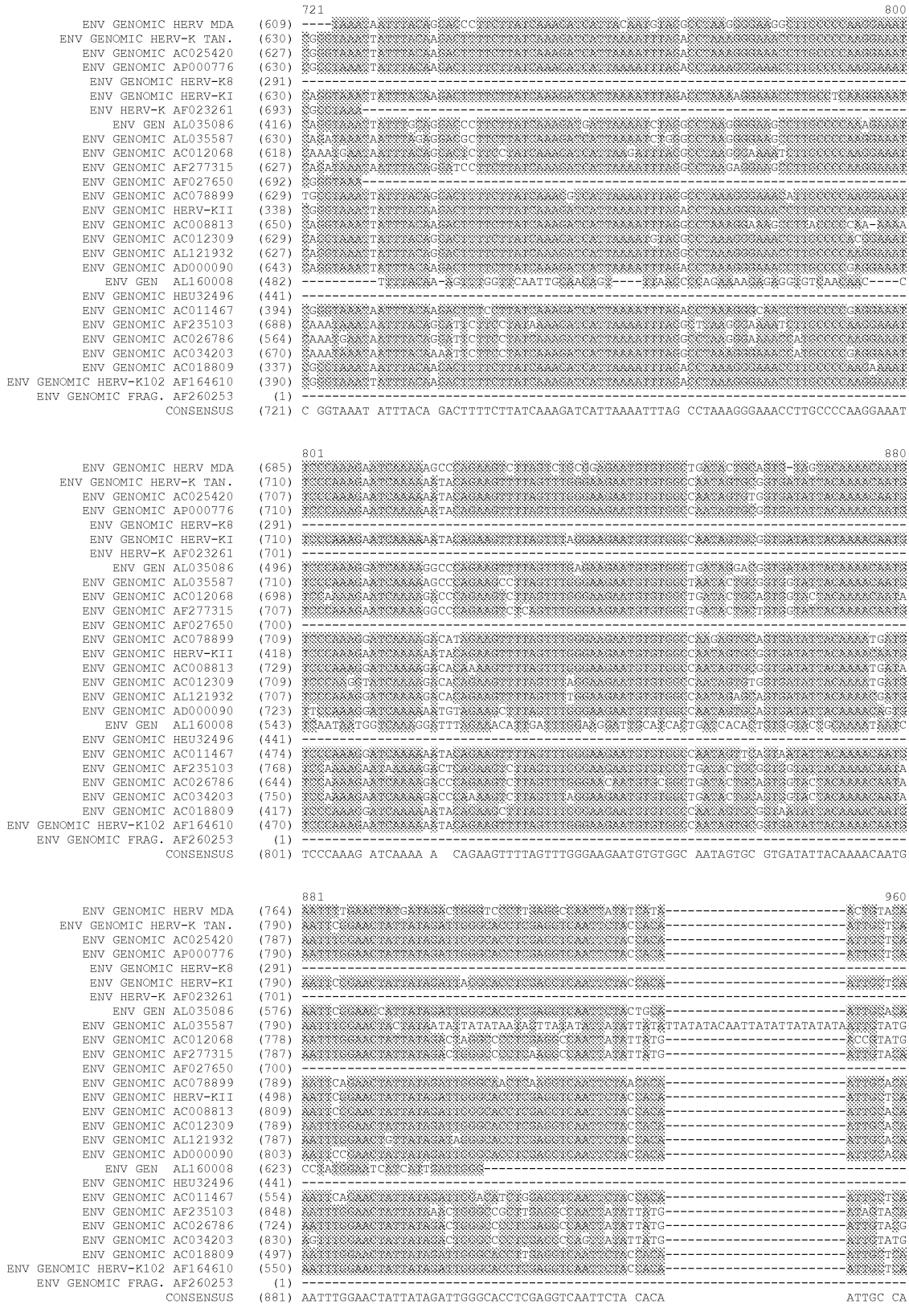


FIGURE 6 CONTD...

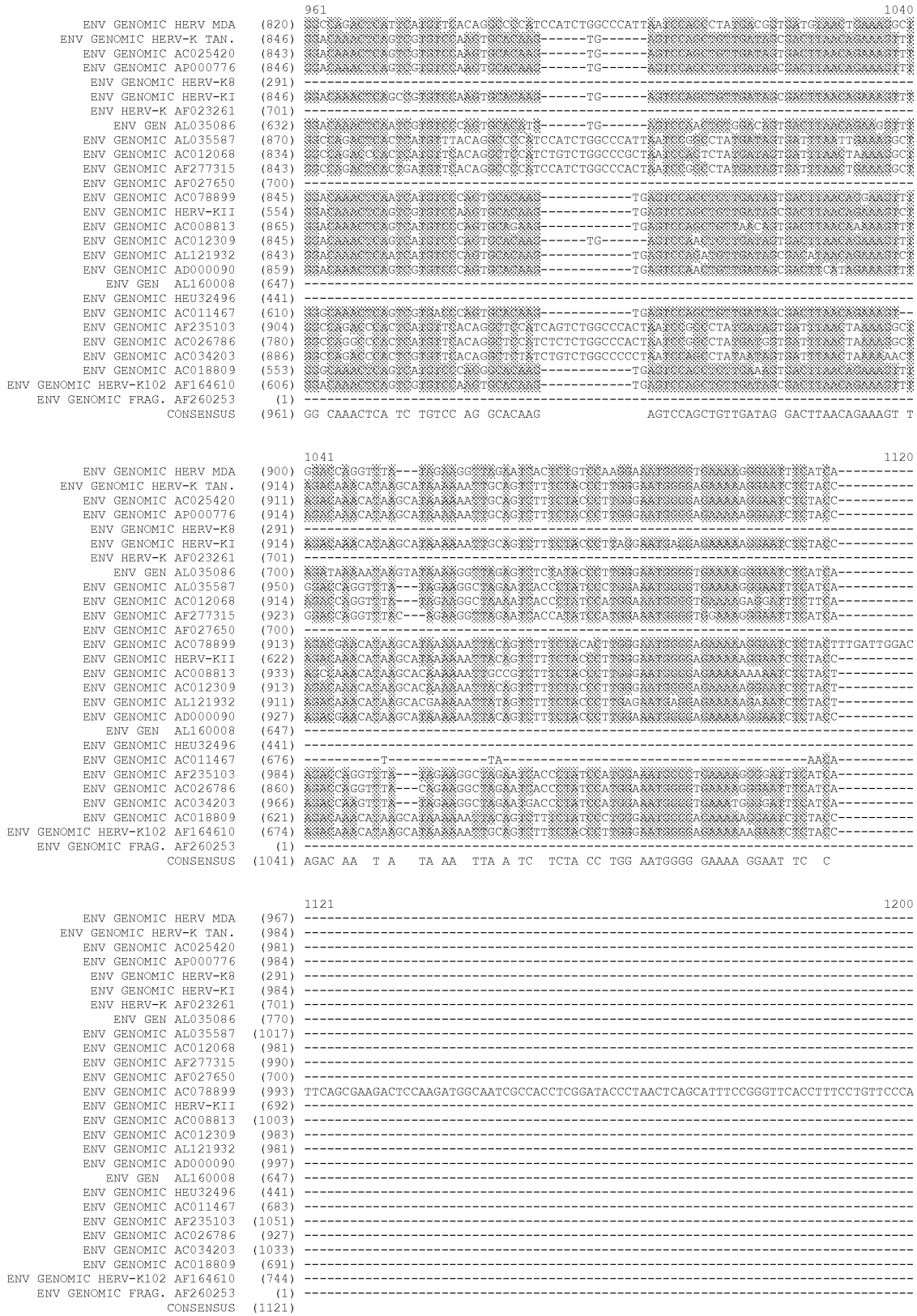


FIGURE 6 CONTD...

```
                                        1201                                                                           1280
         ENV GENOMIC HERV MDA   (967) --------------------------------------------------------------------------------
      ENV GENOMIC HERV-K TAN.   (984) --------------------------------------------------------------------------------
         ENV GENOMIC AC025420   (981) --------------------------------------------------------------------------------
         ENV GENOMIC AP000776   (984) --------------------------------------------------------------------------------
          ENV GENOMIC HERV-K8   (291) --------------------------------------------------------------------------------
          ENV GENOMIC HERV-KI   (984) --------------------------------------------------------------------------------
          ENV HERV-K AF023261   (701) --------------------------------------------------------------------------------
             ENV GEN AL035086   (770) --------------------------------------------------------------------------------
         ENV GENOMIC AL035587  (1017) --------------------------------------------------------------------------------
         ENV GENOMIC AC012068   (981) --------------------------------------------------------------------------------
         ENV GENOMIC AF277315   (990) --------------------------------------------------------------------------------
         ENV GENOMIC AF027650   (700) --------------------------------------------------------------------------------
         ENV GENOMIC AC078899  (1073) CCACCCCGACTAACGCACATGCCCACTAGGGCGTGTCACACTCAGAAGTGTGAAACTCAACCGATCCCGCCCCTACCCCG
         ENV GENOMIC HERV-KII   (692) --------------------------------------------------------------------------------
         ENV GENOMIC AC008813  (1003) --------------------------------------------------------------------------------
         ENV GENOMIC AC012309   (983) --------------------------------------------------------------------------------
         ENV GENOMIC AL121932   (981) --------------------------------------------------------------------------------
         ENV GENOMIC AD000090   (997) --------------------------------------------------------------------------------
            ENV GEN  AL160008   (647) --------------------------------------------------------------------------------
         ENV GENOMIC HEU32496   (441) --------------------------------------------------------------------------------
         ENV GENOMIC AC011467   (683) --------------------------------------------------------------------------------
         ENV GENOMIC AF235103  (1051) --------------------------------------------------------------------------------
         ENV GENOMIC AC026786   (927) --------------------------------------------------------------------------------
         ENV GENOMIC AC034203  (1033) --------------------------------------------------------------------------------
         ENV GENOMIC AC018809   (691) --------------------------------------------------------------------------------
ENV GENOMIC HERV-K102 AF164610  (744) --------------------------------------------------------------------------------
    ENV GENOMIC FRAG. AF260253    (1) --------------------------------------------------------------------------------
                    CONSENSUS  (1201)

                                        1281                                                                           1360
         ENV GENOMIC HERV MDA   (967) --------------------------------------------------------------------------------
      ENV GENOMIC HERV-K TAN.   (984) --------------------------------------------------------------------------------
         ENV GENOMIC AC025420   (981) --------------------------------------------------------------------------------
         ENV GENOMIC AP000776   (984) --------------------------------------------------------------------------------
          ENV GENOMIC HERV-K8   (291) --------------------------------------------------------------------------------
          ENV GENOMIC HERV-KI   (984) --------------------------------------------------------------------------------
          ENV HERV-K AF023261   (701) --------------------------------------------------------------------------------
             ENV GEN AL035086   (770) --------------------------------------------------------------------------------
         ENV GENOMIC AL035587  (1017) --------------------------------------------------------------------------------
         ENV GENOMIC AC012068   (981) --------------------------------------------------------------------------------
         ENV GENOMIC AF277315   (990) --------------------------------------------------------------------------------
         ENV GENOMIC AF027650   (700) --------------------------------------------------------------------------------
         ENV GENOMIC AC078899  (1153) ACCACTCCTCACCCAGCATCCATAAAAGCGCGCTGCACCTTTCGCACAGCGTGACTTCCCCTGGCGGACCAGTGAACCTC
         ENV GENOMIC HERV-KII   (692) --------------------------------------------------------------------------------
         ENV GENOMIC AC008813  (1003) --------------------------------------------------------------------------------
         ENV GENOMIC AC012309   (983) --------------------------------------------------------------------------------
         ENV GENOMIC AL121932   (981) --------------------------------------------------------------------------------
         ENV GENOMIC AD000090   (997) --------------------------------------------------------------------------------
            ENV GEN  AL160008   (647) --------------------------------------------------------------------------------
         ENV GENOMIC HEU32496   (441) --------------------------------------------------------------------------------
         ENV GENOMIC AC011467   (683) --------------------------------------------------------------------------------
         ENV GENOMIC AF235103  (1051) --------------------------------------------------------------------------------
         ENV GENOMIC AC026786   (927) --------------------------------------------------------------------------------
         ENV GENOMIC AC034203  (1033) --------------------------------------------------------------------------------
         ENV GENOMIC AC018809   (691) --------------------------------------------------------------------------------
ENV GENOMIC HERV-K102 AF164610  (744) --------------------------------------------------------------------------------
    ENV GENOMIC FRAG. AF260253    (1) --------------------------------------------------------------------------------
                    CONSENSUS  (1281)

                                        1361                                                                           1440
         ENV GENOMIC HERV MDA   (967) --------------------------------------------------------------------------------
      ENV GENOMIC HERV-K TAN.   (984) --------------------------------------------------------------------------------
         ENV GENOMIC AC025420   (981) --------------------------------------------------------------------------------
         ENV GENOMIC AP000776   (984) --------------------------------------------------------------------------------
          ENV GENOMIC HERV-K8   (291) --------------------------------------------------------------------------------
          ENV GENOMIC HERV-KI   (984) --------------------------------------------------------------------------------
          ENV HERV-K AF023261   (701) --------------------------------------------------------------------------------
             ENV GEN AL035086   (770) --------------------------------------------------------------------------------
         ENV GENOMIC AL035587  (1017) --------------------------------------------------------------------------------
         ENV GENOMIC AC012068   (981) --------------------------------------------------------------------------------
         ENV GENOMIC AF277315   (990) --------------------------------------------------------------------------------
         ENV GENOMIC AF027650   (700) --------------------------------------------------------------------------------
         ENV GENOMIC AC078899  (1233) ACCGGAGAGCTCAATAAAGAAGATTTTTGCCCTCTTTGTCTTGCCTCTTGGCCTTATTGATCCACGGTGCCTTTCCATTG
         ENV GENOMIC HERV-KII   (692) --------------------------------------------------------------------------------
         ENV GENOMIC AC008813  (1003) --------------------------------------------------------------------------------
         ENV GENOMIC AC012309   (983) --------------------------------------------------------------------------------
         ENV GENOMIC AL121932   (981) --------------------------------------------------------------------------------
         ENV GENOMIC AD000090   (997) --------------------------------------------------------------------------------
            ENV GEN  AL160008   (647) --------------------------------------------------------------------------------
         ENV GENOMIC HEU32496   (441) --------------------------------------------------------------------------------
         ENV GENOMIC AC011467   (683) --------------------------------------------------------------------------------
         ENV GENOMIC AF235103  (1051) --------------------------------------------------------------------------------
         ENV GENOMIC AC026786   (927) --------------------------------------------------------------------------------
         ENV GENOMIC AC034203  (1033) --------------------------------------------------------------------------------
         ENV GENOMIC AC018809   (691) --------------------------------------------------------------------------------
ENV GENOMIC HERV-K102 AF164610  (744) --------------------------------------------------------------------------------
    ENV GENOMIC FRAG. AF260253    (1) --------------------------------------------------------------------------------
                    CONSENSUS  (1361)
```

FIGURE 6 CONTD...

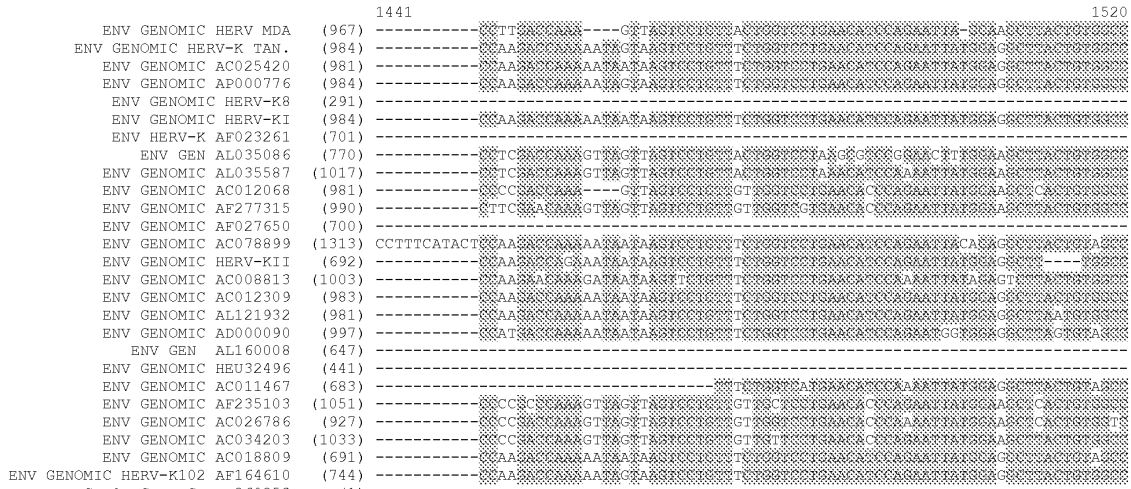


FIGURE 6 CONTD...

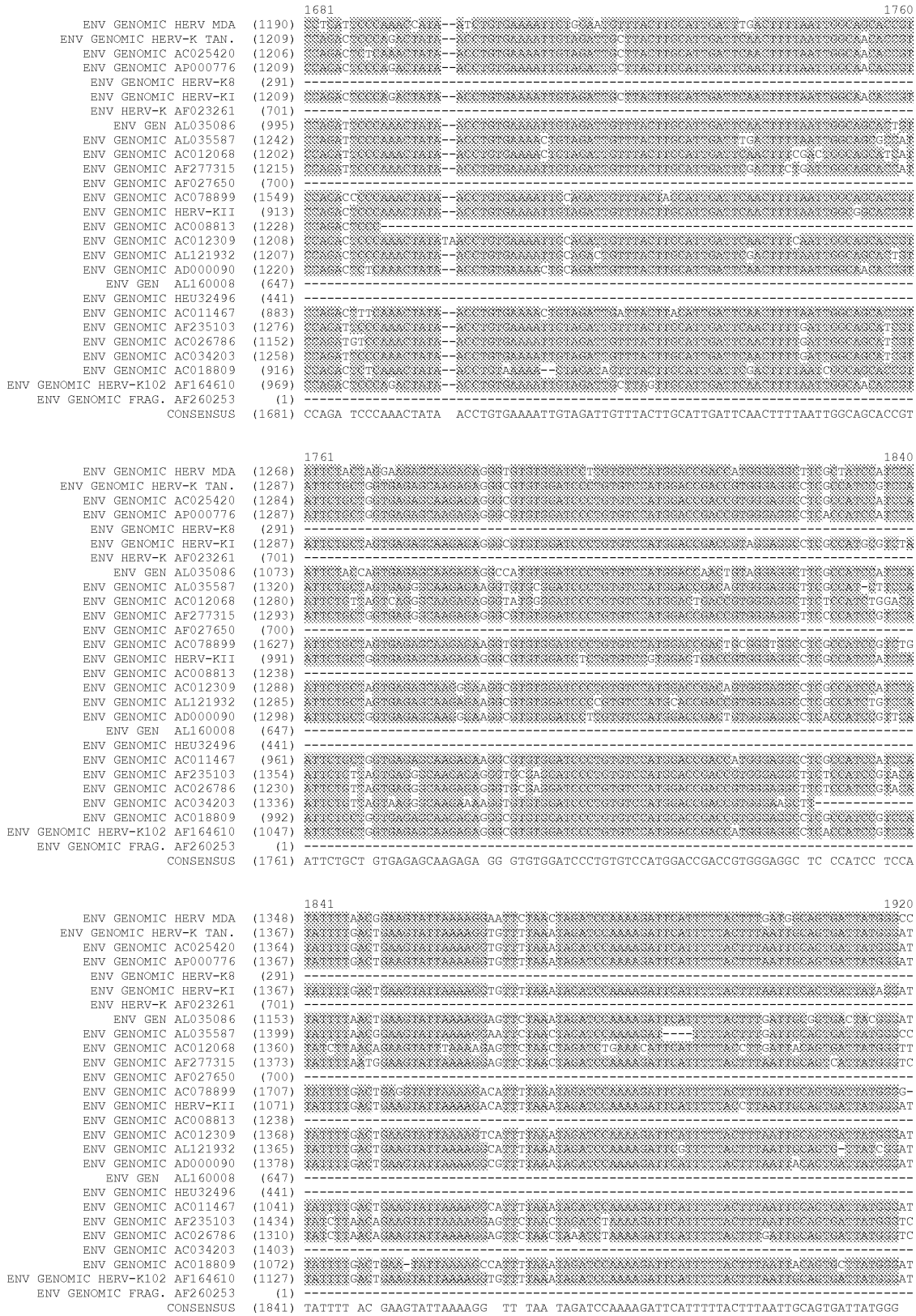


FIGURE 6 CONTD...

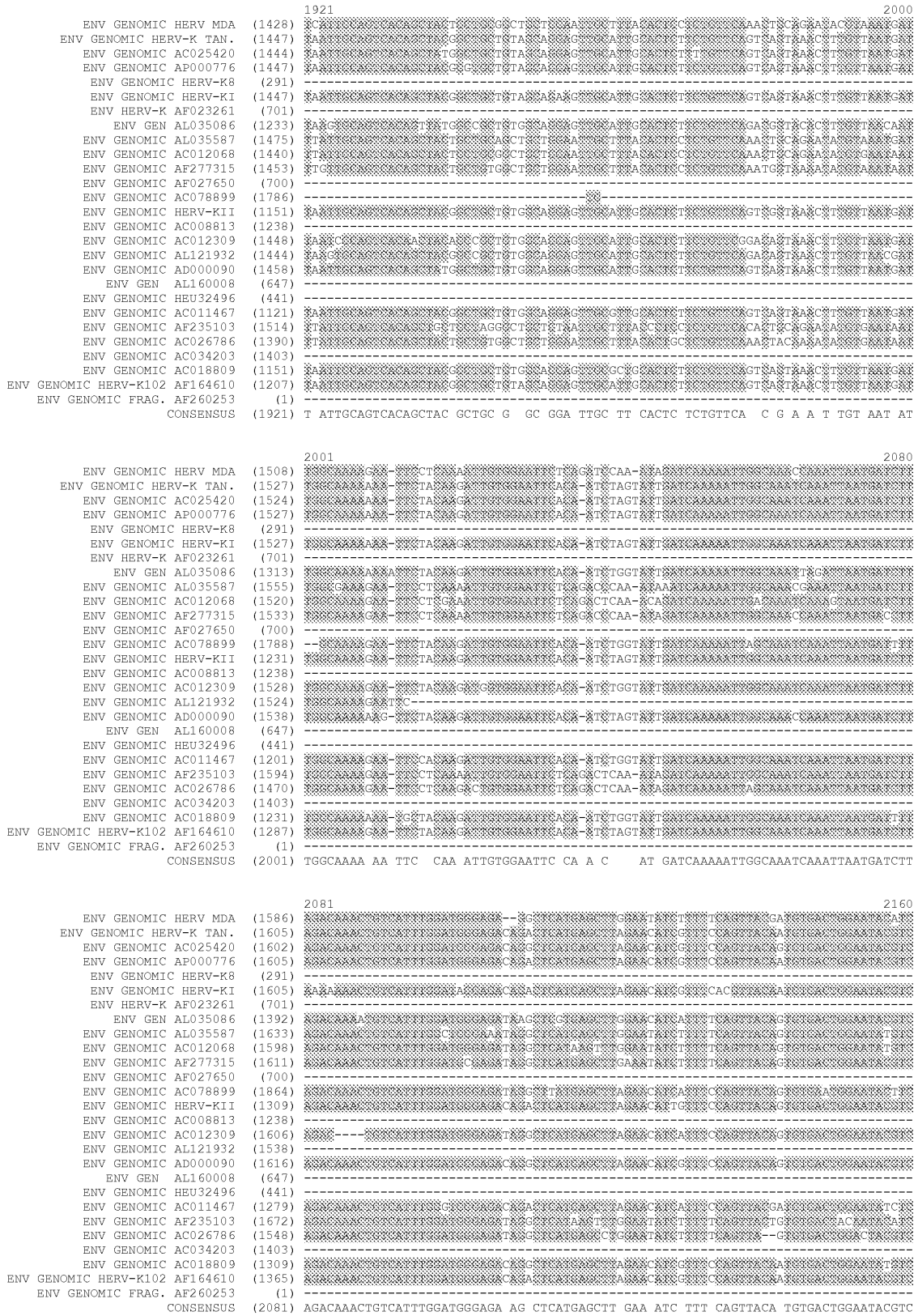
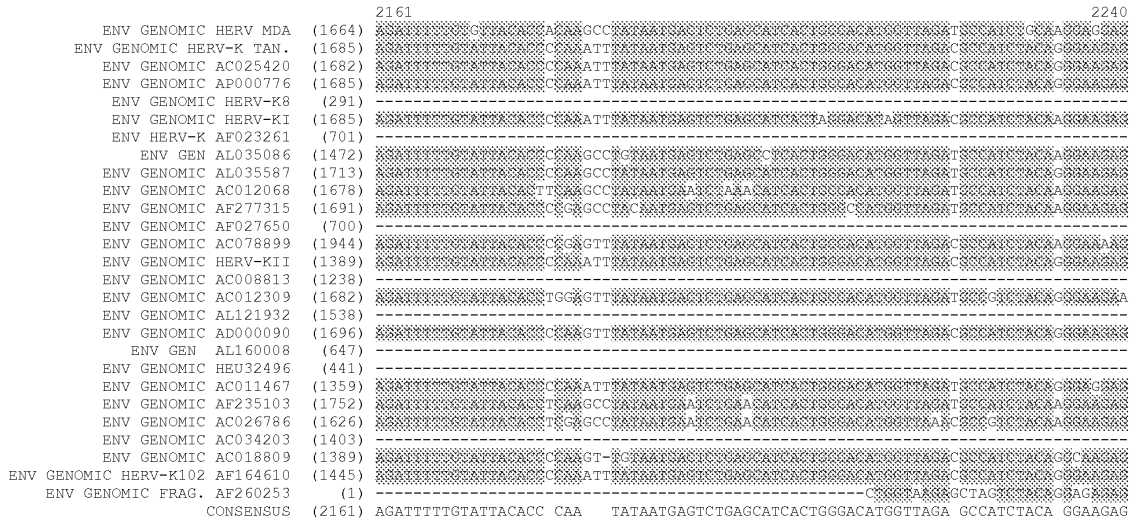
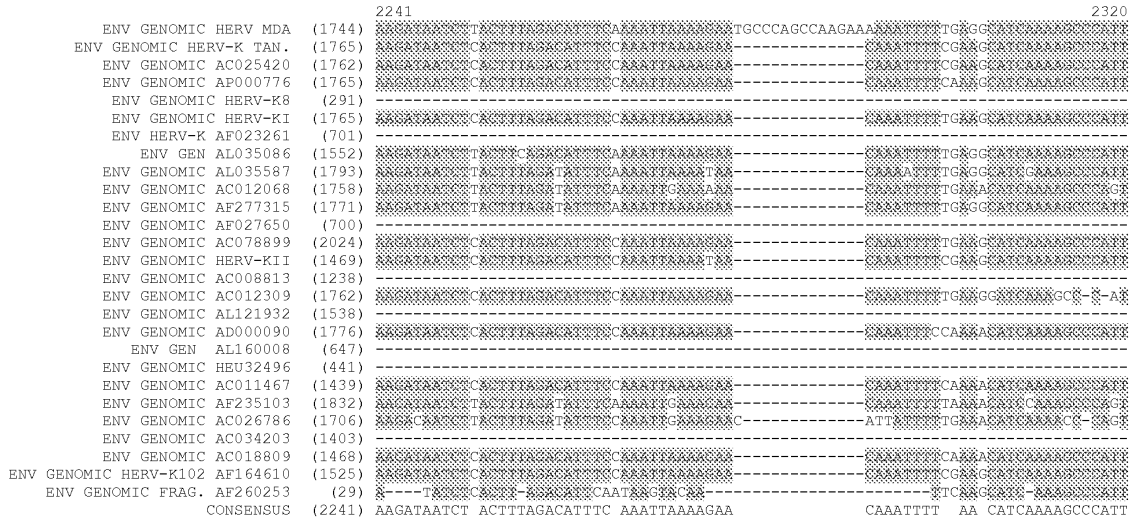


FIGURE 6 CONTD...





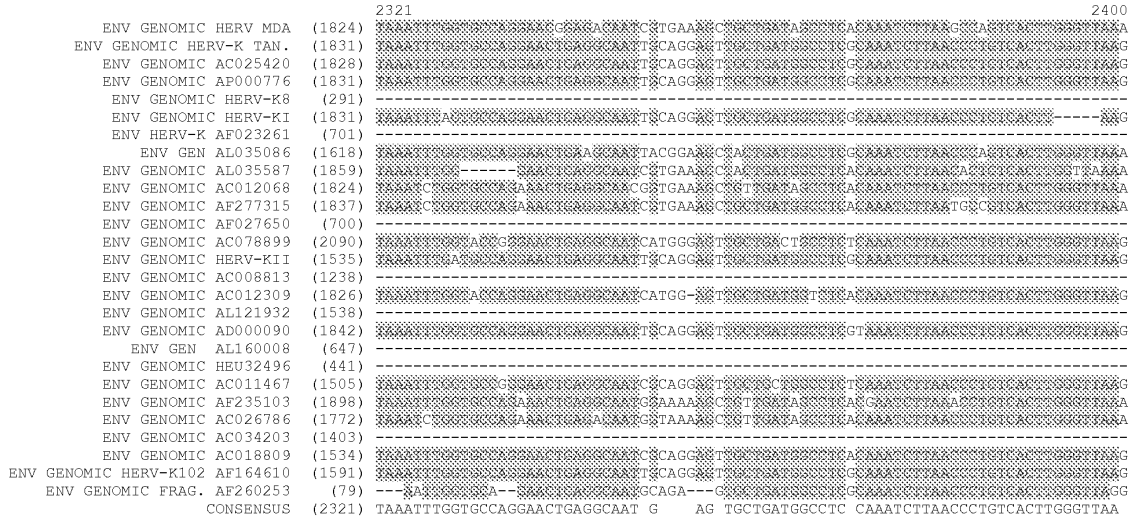


FIGURE 6 CONTD...

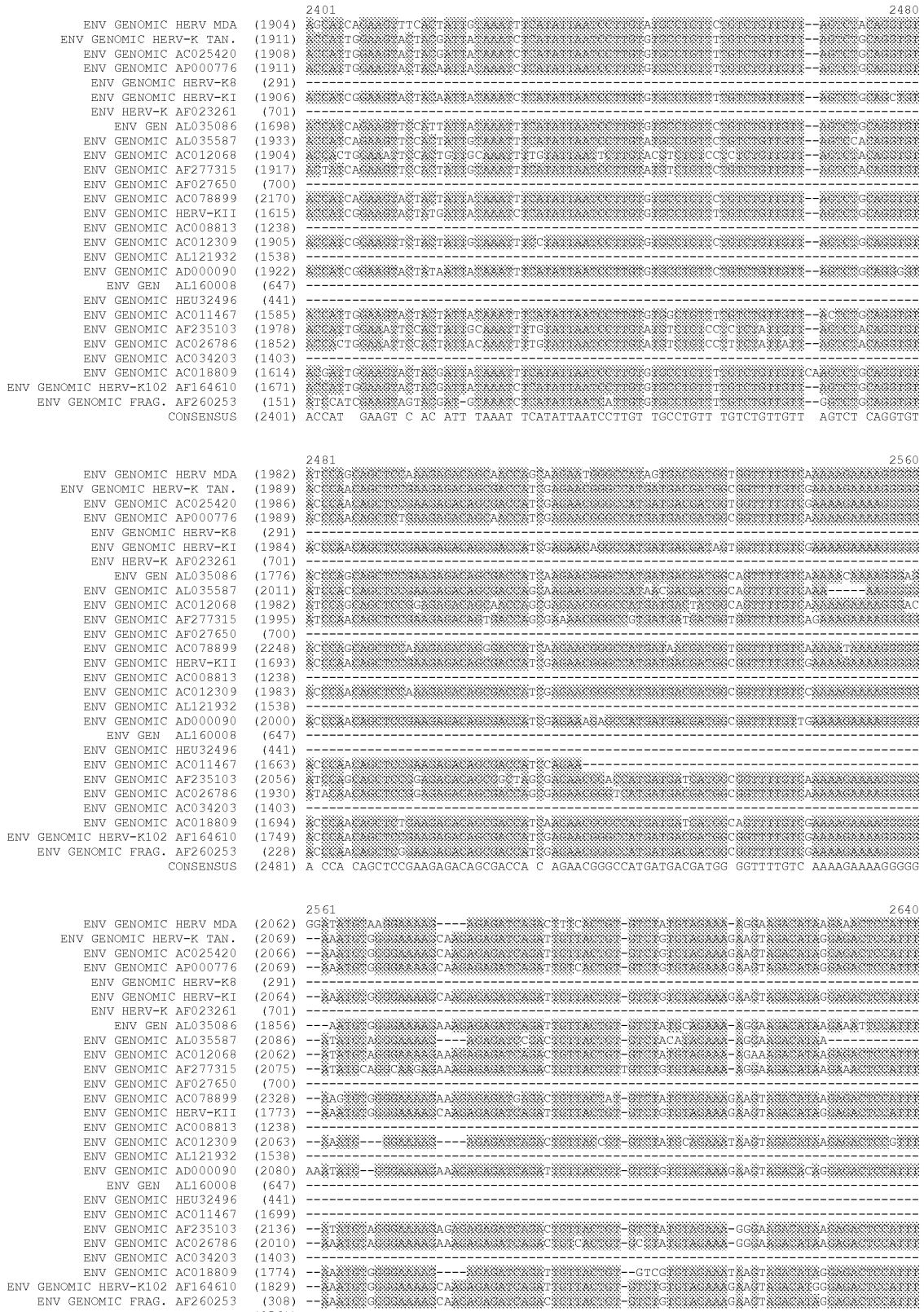


FIGURE 6 CONTD...

FIGURE 7

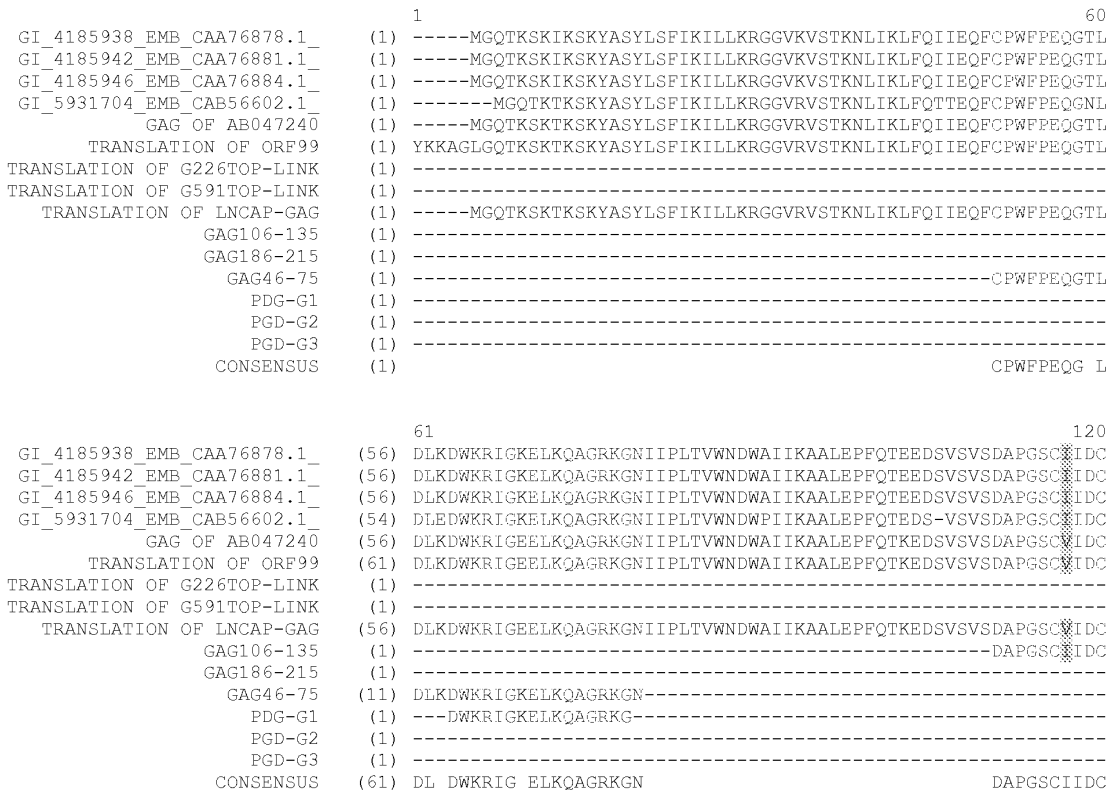


FIGURE 7 CONTD...

```
                                        121                                                                 180
 GI_4185938_EMB_CAA76878.1_  (116) NENTRKKSQKETEGLHCEYVAEPVMAQSTQNVDYNQLQEVIYPETLKLEGKGPELVGPSE
 GI_4185942_EMB_CAA76881.1_  (116) NENTRKKSQKETESLHCEYVAEPVMAQSTQNVDYNQLQEVIYPETLKLEGKGPELVGPSE
 GI_4185946_EMB_CAA76884.1_  (116) NENTRKKSQKETEGLHCEYVAEPVMAQSTQNVDYNQLQEVIYPETLKLEGKGPELVGPSE
 GI_5931704_EMB_CAB56602.1_  (113) NEKTRKKSQKETETLHCEYVAEPLMAQSTQNVDYNQLQEVIYPETLKLEGKGPELVGPLE
           GAG OF AB047240   (116) NEKTGKKSQKETESLHCEYVTEPVMAQSTQNVDYNQLQGVIYPETLKLEGKGPELVGPSE
      TRANSLATION OF ORF99   (121) NEKTGKKSQKETESLHCEYVTEPVMAQSTQNVDYNQLQGVIYPETLKLEGKGPELVGPSE
TRANSLATION OF G226TOP-LINK    (1) ------------------------------------------------------------
TRANSLATION OF G591TOP-LINK    (1) ------------------------------------------------------------
   TRANSLATION OF LNCAP-GAG  (116) NEKTGKKSQKETESLHCEYVTEPVMAQSTQNVDYNQLQGVIYPETLKLEGKGPELVGPSE
                GAG106-135   (11) NENTRKKSQKETEGLHCEYV----------------------------------------
                GAG186-215    (1) ------------------------------------------------------------
                  GAG46-75   (31) ------------------------------------------------------------
                     PDG-G1  (17) ------------------------------------------------------------
                     PGD-G2   (1) ------------------------------------------------------------
                     PGD-G3   (1) ------------------------------------------------------------
                  CONSENSUS (121) NE T KKSQKETE LHCEYV

                                        181                                                                 240
 GI_4185938_EMB_CAA76878.1_  (176) SKPRGTSPLPAGQVPVTLQPQKQVKENKTQPPVAYQYWPPAELQYRPPPESQYGYPGMPP
 GI_4185942_EMB_CAA76881.1_  (176) SKPRGTSRLPAGQVPVTLQPQTQVKENKTQPPVAYQYWPPAELQYRPPVESQYGYPGMPP
 GI_4185946_EMB_CAA76884.1_  (176) SKPRGTSPLPAGQVPVTLQPQKQVKENKTQPPVAYQYWPPAELQYRPPPESQYGYPGMPP
 GI_5931704_EMB_CAB56602.1_  (173) SKPRGPSPLSAGQVTVTLQPQAQVKENKTQLPVAYQYWPPAELQYRPPPESQYGYLGMPP
           GAG OF AB047240   (176) SKPRGPSPLPAGQVPVTLQPQTQVKENKTQPPVAYQYWPPAELQYLPPPESQYGYPGMPP
      TRANSLATION OF ORF99   (181) SKPRGPSPLPAGQVPVTLQPQTQVKENKTQPPVAYQYWPPAELQYLPPPESQYGYPGMPP
TRANSLATION OF G226TOP-LINK    (1) ----------------------------------------------------SQYGYPGMPP
TRANSLATION OF G591TOP-LINK    (1) ------------------------------------------------------------
   TRANSLATION OF LNCAP-GAG  (176) SKPRGPSPLPAGQVPVTLQPQTQVKENKTQPPVAYQYWPPAELQYLPPPESQYGYPGMPP
                GAG106-135   (31) ------------------------------------------------------------
                GAG186-215    (1) ----------AGQVPVTLQPQKQVKENKTQPPVAYQYWPP--------------------
                  GAG46-75   (31) ------------------------------------------------------------
                     PDG-G1  (17) ------------------------------------------------------------
                     PGD-G2   (1) ------------------------------------------------------------
                     PGD-G3   (1) ------------------------------------------------------------
                  CONSENSUS (181)           AGQV VTLQPQ QVKENKTQ PVAYQYWPP          SQYGY GMPP

                                        241                                                                 300
 GI_4185938_EMB_CAA76878.1_  (236) APQGRAPYPQPPTRRLNPTAPPSRQGSKLHEIIDKSRKEGDTEAWQFPVTLEPMPPGEGA
 GI_4185942_EMB_CAA76881.1_  (236) APQGRAPYPQPPTRRLNPTAPPSRRGSELHEIIDKSRKEGDTEAWQFPVMLEPMPPGEGA
 GI_4185946_EMB_CAA76884.1_  (236) APQGRAPYPQPPTRRLNPTAPPSRQGSKLHEIIDKSRKEGDTEAWQFPVTLEPMPPGEGA
 GI_5931704_EMB_CAB56602.1_  (233) APQDREPYPQPPTRRQCYGTT-------------------------------------
           GAG OF AB047240   (236) ALQGRAPYPQPPTVRLNPTASRSGQGGTLHAVIDEARKQGDLEAWRFLVILQLVQAGEET
      TRANSLATION OF ORF99   (241) ALQGRAPYPQPPTVRLNPTASRSGQGGTLHAVIDEARKQGDLEAWRFLVILQLVQAGEET
TRANSLATION OF G226TOP-LINK   (11) APQGRAPYPQPPTRRLNPTA-------------------------------------
TRANSLATION OF G591TOP-LINK    (1) ------------------------------------------------------------
   TRANSLATION OF LNCAP-GAG  (236) ALQGRAPYPQPPTVRLNPTASRSGQGGTLHAVIDEARKQGDLEAWRFLVILQLVQAGEET
                GAG106-135   (31) ------------------------------------------------------------
                GAG186-215   (31) ------------------------------------------------------------
                  GAG46-75   (31) ------------------------------------------------------------
                     PDG-G1  (17) ------------------------------------------------------------
                     PGD-G2   (1) --------------------------SKLHEIIDKSRKEGDT------------------
                     PGD-G3   (1) ------------------------------------------------------------
                  CONSENSUS (241) A Q R PYPQPPT R

                                        301                                                                 360
 GI_4185938_EMB_CAA76878.1_  (296) QEGEPPTVEARYKSFSIKKLKDMKEGVKQYGPNSPYMRTLLDSIAHGHRLIPYDWEILAK
 GI_4185942_EMB_CAA76881.1_  (296) QEGEPPTVEARYKSFSIKMLKDMKEGVKQYGPNSPYMRTLLDSIAHGHRLIPYDWEILAK
 GI_4185946_EMB_CAA76884.1_  (296) QEGEPPTVEARYKSFSIKKLKDMKEGVKQYGPNSPYMRTLLDSIAHGHRLIPYDWEILAK
 GI_5931704_EMB_CAB56602.1_  (254) ------------------------------------------------------------
           GAG OF AB047240   (296) QVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTLLDSIAHGNRLTPYDWESLAK
      TRANSLATION OF ORF99   (301) QVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTLLDSIAHGNRLTPYDWESLAK
TRANSLATION OF G226TOP-LINK   (31) ------------------------------------------------------------
TRANSLATION OF G591TOP-LINK    (1) ------------------------------------------------------------
   TRANSLATION OF LNCAP-GAG  (296) QVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTLLDSIAHGNRLTPYDWESLAK
                GAG106-135   (31) ------------------------------------------------------------
                GAG186-215   (31) ------------------------------------------------------------
                  GAG46-75   (31) ------------------------------------------------------------
                     PDG-G1  (17) ------------------------------------------------------------
                     PGD-G2  (17) ------------------------------------------------------------
                     PGD-G3   (1) ------------------------------------------------------------
                  CONSENSUS (301)
```

FIGURE 7 CONTD...

```
                                    361                                                        420
   GI_4185938_EMB_CAA76878.1_  (356) SSLSPSQFLQFKTWWIDGVQEQVRRNRAANPPVNIDADQLLGIGQNWSTISQQALMQNEA
   GI_4185942_EMB_CAA76881.1_  (356) SSLSPSQFLQFKTWWIDGVQEQVRRNRAANPPVNIDADQLLGIGQNWSTISQQALMQNEA
   GI_4185946_EMB_CAA76884.1_  (356) SSLSPSQFLQFKTWWIDGVQEQVRRNRAANPPVNIDADQLLGIGQNWSTISQQALMQNEA
   GI_5931704_EMB_CAB56602.1_  (254) ------------------------------------------------------------
              GAG OF AB047240  (356) SSLSSSQYLQFKTWWIDGVQEQVRKNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEA
        TRANSLATION OF ORF99  (361) SSLSSSQYLQFKTWWIDGVQEQVRKNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEA
TRANSLATION OF G226TOP-LINK   (31) ------------------------------------------------------------
TRANSLATION OF G591TOP-LINK    (1) ------------------------------------------------------------
   TRANSLATION OF LNCAP-GAG  (356) SSLSSSQYLQFKTWWIDGVQEQVRKNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEA
                  GAG106-135   (31) ------------------------------------------------------------
                  GAG186-215   (31) ------------------------------------------------------------
                    GAG46-75   (31) ------------------------------------------------------------
                      PDG-G1   (17) ------------------------------------------------------------
                      PGD-G2   (17) ------------------------------------------------------------
                      PGD-G3    (1) ------------------------------------------------------------
                   CONSENSUS  (361)

                                    421                                                        480
   GI_4185938_EMB_CAA76878.1_  (416) IEQVRAICLRAWEKIQDPGSTCPSFNTVRQGSKEPYPDFVARLQDVAQKSIADEKARKVI
   GI_4185942_EMB_CAA76881.1_  (416) IEQVRAICLRAWEKIQDPGSTCPSFNTVRQGSKEPYPDFVARLQDVAQKSIANEKARKVI
   GI_4185946_EMB_CAA76884.1_  (416) IEQVRAICLRAWEKIQDPGSTCPSFNTVRQGSKEPYPDFVARLQDVAQKSIADEKARKVI
   GI_5931704_EMB_CAB56602.1_  (254) ------------------------------------------------------------
              GAG OF AB047240  (416) IEQVRAICLRAWGKIQDPGTAFP-INSIRQGSKEPYPDFVARLQDAAQKSITDDNARKVI
        TRANSLATION OF ORF99  (421) IEQVRAICLRAWGKIQDPGTAF-PINSIRQGSKEPYPDFVARLQDAAQKSITDDNARKVI
TRANSLATION OF G226TOP-LINK   (31) ------------------------------------------------------------
TRANSLATION OF G591TOP-LINK    (1) ------------------------------------------------------------
   TRANSLATION OF LNCAP-GAG  (416) IEQVRAICLRAWGKIQDPGTAFP-INSIRQGSKEPYPDFVARLQDAAQKSITDDNARKVI
                  GAG106-135   (31) ------------------------------------------------------------
                  GAG186-215   (31) ------------------------------------------------------------
                    GAG46-75   (31) ------------------------------------------------------------
                      PDG-G1   (17) ------------------------------------------------------------
                      PGD-G2   (17) ------------------------------------------------------------
                      PGD-G3    (1) ------------------------------------------------------------
                   CONSENSUS  (421)

                                    481                                                        540
   GI_4185938_EMB_CAA76878.1_  (476) VELMAYENANPECQSAIKPLKGKVPAGSDVISEYVKACDGIGGAMYKAMLMAQAITGVVL
   GI_4185942_EMB_CAA76881.1_  (476) VELMAYENPNPECQSAIKPLKGKVPAGSDVISEYVKACDGMGGAMHKAMLMAQAITGVVL
   GI_4185946_EMB_CAA76884.1_  (476) VELMAYENANPECQSAIKPLKGKVPAGSDVISEYVKACDGIGGAMHKAMLMAQAITGVVL
   GI_5931704_EMB_CAB56602.1_  (254) ------------------------------------------------------------
              GAG OF AB047240  (475) VELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGLTL
        TRANSLATION OF ORF99  (480) VELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGLTL
TRANSLATION OF G226TOP-LINK   (31) ------------------------------------------------------------
TRANSLATION OF G591TOP-LINK    (1) ------------------------------------------------------------
   TRANSLATION OF LNCAP-GAG  (475) VELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGLTL
                  GAG106-135   (31) ------------------------------------------------------------
                  GAG186-215   (31) ------------------------------------------------------------
                    GAG46-75   (31) ------------------------------------------------------------
                      PDG-G1   (17) ------------------------------------------------------------
                      PGD-G2   (17) ------------------------------------------------------------
                      PGD-G3    (1) ------------------------------------------------------------
                   CONSENSUS  (481)

                                    541                                                        600
   GI_4185938_EMB_CAA76878.1_  (536) GGQVRTFGRKCYNCGQIGHLKKNCPVLNKQNITIQATTTG-REPPDLCPRCKKGKHWASQ
   GI_4185942_EMB_CAA76881.1_  (536) GGQVRTFGGKCYNCGQIGHLKKNCPVLNKQNITIQATTTG-REPPDLCPRCKKGKHWASQ
   GI_4185946_EMB_CAA76884.1_  (536) GGQVRTFGRKCYNCGQIGHLKKNCPVLNKQNITIQATTTG-REPPDLCPRCKKGKHWASQ
   GI_5931704_EMB_CAB56602.1_  (254) ------------------------------------------------------------
              GAG OF AB047240  (535) GGQVRTFGKKCYNCGQIGHLKRSCPVLNKQNIINQAITAKNKKPSGLCPKCGKGKHWANQ
        TRANSLATION OF ORF99  (540) GGQVRTFGKKCYNCGQIGHLKRSCPVLNKQNIINQAITAKNKKPSGLCPKCGKGKHWANQ
TRANSLATION OF G226TOP-LINK   (31) ------------------------------------------------------------
TRANSLATION OF G591TOP-LINK    (1) --------------------------------------------------------WASQ
   TRANSLATION OF LNCAP-GAG  (535) GGQVRTFGKKCYNCGQIGHLKRSCPVLNKQNIINQAITAKNKKPSGLCPKCGKGKHWANQ
                  GAG106-135   (31) ------------------------------------------------------------
                  GAG186-215   (31) ------------------------------------------------------------
                    GAG46-75   (31) ------------------------------------------------------------
                      PDG-G1   (17) ------------------------------------------------------------
                      PGD-G2   (17) ------------------------------------------------------------
                      PGD-G3    (1) ------------------------------------------------------------
                   CONSENSUS  (541)
```

FIGURE 7 CONTD...

```
                                        601                                                                 660
 GI_4185938_EMB_CAA76878.1_   (595) CRSKFDKNGQPLSGNEQRGQPQAPQQTGAFPIQPFVPQGFQGQQP-PLSQVFQGISQLPQ
 GI_4185942_EMB_CAA76881.1_   (595) CRSKFDKNGQFLSGNEQRGQPQAPQQTGAFPIQPFVPHGFQGQQP-PLSQVFQGISQLPQ
 GI_4185946_EMB_CAA76884.1_   (595) CRSKFDKNGQPLSGNEQRGQPQAPQQTGAFPIQPFVPQGFQGQQP-PLSQVFQGISQLPQ
 GI_5931704_EMB_CAB56602.1_   (254) ------------------------------------------------------------
            GAG OF AB047240   (595) CHSKFDKDGQPLSGNRKRGQPQAPQQTGAFPVQLFVPQGFQGQQPLQKIPPLQGVSQLQQ
       TRANSLATION OF ORF99   (600) CHSKFDKDGQPLSGNRKRGQPQAPQQTGAFPVQLFVPQGFQGQQPLQKIPPLQGVSQLQQ
TRANSLATION OF G226TOP-LINK   (31) ------------------------------------------------------------
TRANSLATION OF G591TOP-LINK    (5) CRSKFDKNGQPLSGNEQRGQPQAPQQ----------------------------------
   TRANSLATION OF LNCAP-GAG   (595) CHSKFDKDGQFLSGNRKRGQPQAPQQTGAFPVQLFVPQGFQGQQPLQKIPPLQGVSQLQQ
                GAG106-135    (31) ------------------------------------------------------------
                GAG186-215    (31) ------------------------------------------------------------
                  GAG46-75    (31) ------------------------------------------------------------
                     PDG-G1   (17) ------------------------------------------------------------
                     PGD-G2   (17) ------------------------------------------------------------
                     PGD-G3    (1) CRSKFDKNGQPLSGNE--------------------------------------------
                  CONSENSUS  (601) C SKFDK GQPLSGN

                                        661       673
 GI_4185938_EMB_CAA76878.1_   (654) YNNCPPPQAAVQQ
 GI_4185942_EMB_CAA76881.1_   (654) YNNCPPPQAAVQQ
 GI_4185946_EMB_CAA76884.1_   (654) YNNCPPPQAAVQQ
 GI_5931704_EMB_CAB56602.1_   (254) -------------
            GAG OF AB047240   (655) SNSCPAPQQAAPQ
       TRANSLATION OF ORF99   (660) SNSCPAPQQAAPQ
TRANSLATION OF G226TOP-LINK   (31) -------------
TRANSLATION OF G591TOP-LINK   (31) -------------
   TRANSLATION OF LNCAP-GAG   (655) SNSCPAPQQAAPQ
                GAG106-135    (31) -------------
                GAG186-215    (31) -------------
                  GAG46-75    (31) -------------
                     PDG-G1   (17) -------------
                     PGD-G2   (17) -------------
                     PGD-G3   (17) -------------
                  CONSENSUS  (661)
```

FIGURE 8

```
                                        1                                                                   60
        GI_4185939_EMB_CAA76879.1_     (1) MLTDLRAVN---AVIQPMGPLQPGLPSPAMIPKDWPLIIIDLKDCFFTIPLAEQDCEKFA
        GI_4185943_EMB_CAA76882.1_     (1) MLTDLRAVNAVNAVIQPMGPLQPGLPSLAMIPKDWPLIIIDLKDCFFTIPLAEQDCEKFA
        GI_4185947_EMB_CAA76885.1_     (1) MLTDLRAVN---AVIQPMGPLQPGLPSPAMIPKDWPLIIIDLKDCFFTIPLAEQDCEKFA
        GI_5931705_EMB_CAB56603.1_     (1) ----------------------------MIPKDWPLIIIDLKDCFFTIPLAEQDCEKFA
                    ENV OF AB047240    (1) ------------------------------------------------------------
          TRANSLATION OF P386TOP-LINK  (1) ------------------------------------------------------------
           TRANSLATION OF POL349-LINK  (1) ------------------------------------------------------------
                 LNCAP-GENOMEA-POLORF  (1) ------------------------------------------------------------
TRANSLATION OF LNCAP-POL-GENA-GOODA    (1) ------------------------------------------------------------
             TRANSLATION OF ORF111-10  (1) ------------------------------------------------------------
                              PGD-P1   (1) ------------------------------------------------------------
                              PGD-P2   (1) ------------------------------------------------------------
                               PGDP3   (1) ------------------------------------------------------------
                           CONSENSUS   (1)

                                        61                                                                  120
        GI_4185939_EMB_CAA76879.1_    (58) FTIPAINNKEPATRFQWKVLPQGMLNSPTICQTFVGRALQPVREKFSDCYIIHCIDDILC
        GI_4185943_EMB_CAA76882.1_    (61) FTIPAINNKEPATRFQWKVLPQGMLNSPTICQTFVGRALQPVREKFSDCYIIHYIDDILC
        GI_4185947_EMB_CAA76885.1_    (58) FTIPAINNKEPATRFQWKVLPQGMLNSPTICQTFVGRALQPVREKFSDCYIIHCIDDILC
        GI_5931705_EMB_CAB56603.1_    (32) FTIPAINNKEPATRFQWKVLPQGMLNSPTLCQTFVGRALQPVRDKFSDCYIIHYFDDILC
                    ENV OF AB047240    (1) ------------------------------------------------------------
          TRANSLATION OF P386TOP-LINK  (1) ------------------------------------------------------------
           TRANSLATION OF POL349-LINK  (1) ------------------------------------------------------------
                 LNCAP-GENOMEA-POLORF  (1) ------------------------------------------------------------
TRANSLATION OF LNCAP-POL-GENA-GOODA    (1) ------------------------------------------------------------
             TRANSLATION OF ORF111-10  (1) ------------------------------------------------------------
                              PGD-P1   (1) ------------------------------------------------------------
                              PGD-P2   (1) ------------------------------------------------------------
                               PGDP3   (1) ------------------------------------------------------------
                           CONSENSUS  (61)
```

FIGURE 8 CONTD...

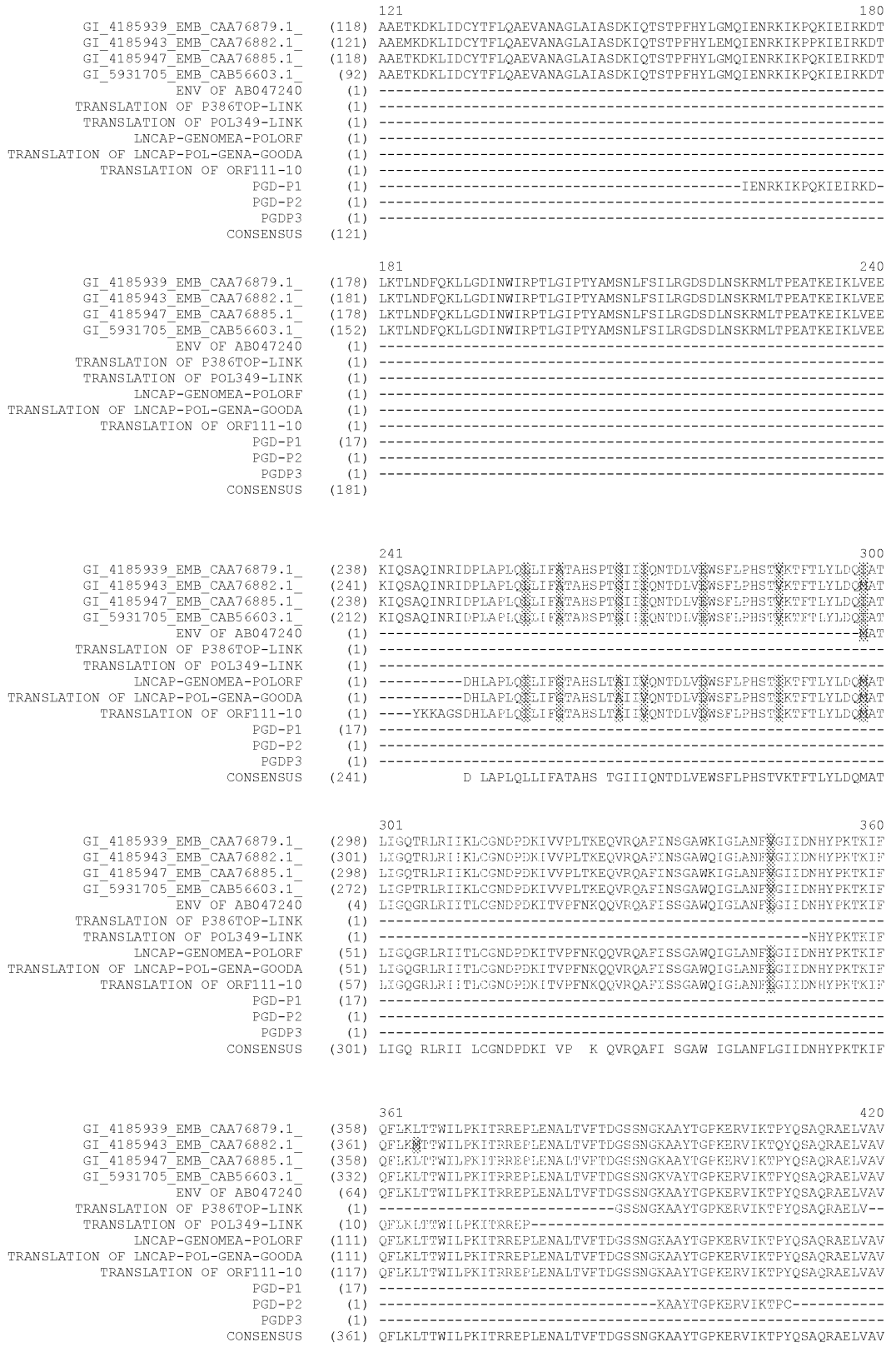


FIGURE 8 CONTD...

		421 … 480
GI_4185939_EMB_CAA76879.1_	(418)	ITVLQDFDQPINIISDSAYVVQATRDVETALIKYSMDDQLNQLFNLLQQTVRKRNFPFYI
GI_4185943_EMB_CAA76882.1_	(421)	ITVLQDFDQPINIISDSAYVVQATRDVETALIKYSMDDQLNQLFNLLQQTVRKRNFPFYI
GI_4185947_EMB_CAA76885.1_	(418)	ITVLQDFDQPINIISDSAYVVQATRDVETALIKYSMDDQLNQLFNLLQQTVRKRNFPFYI
GI_5931705_EMB_CAB56603.1_	(392)	ITVLQDFDQPINIISDSAYVVQATRDVETALIKYSMDDQLNQLFNLLQQTVRKRNFPFYI
ENV OF AB047240	(124)	ITVLQDFDQPINIISDSAYVVQATRDVETALIKYSTDDHLNQLFNLLQQTVRKRNFPFYI
TRANSLATION OF P386TOP-LINK	(31)	--
TRANSLATION OF POL349-LINK	(28)	--
LNCAP-GENOMEA-POLORF	(171)	ITVLQDFDQPINIISDSAYVVQATRDVETALIKYSTDDHLNQLFNLLQQTVRKRNFPFYI
TRANSLATION OF LNCAP-POL-GENA-GOODA	(171)	ITVLQDFDQPINIISDSAYVVQATRDVETALIKYSTDDHLNQLFNLLQQTVRKRNFPFYI
TRANSLATION OF ORF111-10	(177)	ITVLQDFDQPINIISDSAYVVQATRDVETALIKYSTDDHLNQLFNLLQQTVRKRNFPFYI
PGD-P1	(17)	--
PGD-P2	(17)	--
PGDP3	(1)	--
CONSENSUS	(421)	ITVLQDFDQPINIISDSAYVVQATRDVETALIKYS DD LNQLFNLLQQTVRKRNFPFYI

		481 … 540
GI_4185939_EMB_CAA76879.1_	(478)	THIRAHTNLPGPLTKANEQADLLVSSALIKAQELHALTHVNAAGLKNKFDVTWKQAKDIV
GI_4185943_EMB_CAA76882.1_	(481)	THIRAHTNLPGPLTKANEQADLLVSSALIKAQELHALTHVNVAGLKNKFDVTWKQAKDIV
GI_4185947_EMB_CAA76885.1_	(478)	THIRAHTNLPGPLTKANEQADLLVSSALIKAQELHALAHVNAAGLKNKFDVTWKQAKDIV
GI_5931705_EMB_CAB56603.1_	(452)	THIRAHTNLPGPLTKANEQADLLVSSALIKAQELHALTHVNAAGLKNKFDVTWKQAKDIV
ENV OF AB047240	(184)	THIRAHTNLPGPLTKANEQADLLVSSAFIKAQELLALTHVNAAGLKNKFDVTWKQAKDIV
TRANSLATION OF P386TOP-LINK	(31)	--
TRANSLATION OF POL349-LINK	(28)	--
LNCAP-GENOMEA-POLORF	(231)	THIRAHTNLPGPLTKANEQADLLVSSAFIKAQELLALTHVNAAGLKNKFDVTWKQAKDIV
TRANSLATION OF LNCAP-POL-GENA-GOODA	(231)	THIRAHTNLPGPLTKANEQADLLVSSAFIKAQELLALTHVNAAGLKNKFDVTWKQAKDIV
TRANSLATION OF ORF111-10	(237)	THIRAHTNLPGPLTKANEQADLLVSSAFIKAQELLALTHVNAAGLKNKFDVTWKQAKDIV
PGD-P1	(17)	--
PGD-P2	(17)	--
PGDP3	(1)	--
CONSENSUS	(481)	THIRAHTNLPGPLTKANEQADLLVSSA IKAQEL ALTHVNAAGLKNKFDVTWKQAKDIV

		541 … 600
GI_4185939_EMB_CAA76879.1_	(538)	QHCTQCQVLHLPTQEAGVNPRGLCPNALWQMDVTHVPSFGRLSYVHVTVDTYSHFIWATC
GI_4185943_EMB_CAA76882.1_	(541)	QHCTQCQVLHLPTQEAGVNPRGLCPNALWQMDVTHVSSFGRLSYIHVTVDTYSHFIWATC
GI_4185947_EMB_CAA76885.1_	(538)	QHCTQCQVLHLPTQEAGVNPRGLCPNALWQMDVTHVPSFGRLSYVHVTVDTYSHFIWATC
GI_5931705_EMB_CAB56603.1_	(512)	QHCTQCQVLDLPTQEAGVNPEVCVLMHYGKWMSHMYLHLGRLSYVHVTVDTYSHFMCATC
ENV OF AB047240	(244)	QHCTQCQVLHLSTQEAGVNPRGLCPNALWQMDGTHVPSFGRLSYVHVTVDTYSHFIWATC
TRANSLATION OF P386TOP-LINK	(31)	--
TRANSLATION OF POL349-LINK	(28)	--
LNCAP-GENOMEA-POLORF	(291)	QHCTQCQVLHLSTQEAGVNPRGLCPNALWQMDGTHVPSFGRLSYVHVTVDTYSHFIWATC
TRANSLATION OF LNCAP-POL-GENA-GOODA	(291)	QHCTQCQVLHLSTQEAGVNPRGLCPNALWQMDGTHVPSFGRLSYVHVTVDTYSHFIWATC
TRANSLATION OF ORF111-10	(297)	QHCTQCQVLHLSTQEAGVNPRGLCPNALWQMDGTHVPSFGRLSYVHVTVDTYSHFIWATC
PGD-P1	(17)	--
PGD-P2	(17)	--
PGDP3	(1)	--
CONSENSUS	(541)	QHCTQCQVLHL TQEAGVNPRGLCPNALWQMD THV SFGRLSYVHVTVDTYSHFIWATC

		601 … 660
GI_4185939_EMB_CAA76879.1_	(598)	QTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQG
GI_4185943_EMB_CAA76882.1_	(601)	QTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQG
GI_4185947_EMB_CAA76885.1_	(598)	QTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQG
GI_5931705_EMB_CAB56603.1_	(572)	QTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQG
ENV OF AB047240	(304)	QTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQG
TRANSLATION OF P386TOP-LINK	(31)	--
TRANSLATION OF POL349-LINK	(28)	--
LNCAP-GENOMEA-POLORF	(351)	QTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQG
TRANSLATION OF LNCAP-POL-GENA-GOODA	(351)	QTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQG
TRANSLATION OF ORF111-10	(357)	QTGESTSHAKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQG
PGD-P1	(17)	--
PGD-P2	(17)	--
PGDP3	(1)	--
CONSENSUS	(601)	QTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQG

		661 … 720
GI_4185939_EMB_CAA76879.1_	(658)	QAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAEQHLT
GI_4185943_EMB_CAA76882.1_	(661)	QAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAEQHLT
GI_4185947_EMB_CAA76885.1_	(658)	QAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAEQHLT
GI_5931705_EMB_CAB56603.1_	(632)	QAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAE-HLT
ENV OF AB047240	(364)	QAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAKQHLT
TRANSLATION OF P386TOP-LINK	(31)	--
TRANSLATION OF POL349-LINK	(28)	--
LNCAP-GENOMEA-POLORF	(411)	QAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAKQHLT
TRANSLATION OF LNCAP-POL-GENA-GOODA	(411)	QAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAKQHLT
TRANSLATION OF ORF111-10	(417)	QAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAKQHLT
PGD-P1	(17)	--
PGD-P2	(17)	--
PGDP3	(1)	---HLT
CONSENSUS	(661)	QAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSA QHLT

FIGURE 8 CONTD...

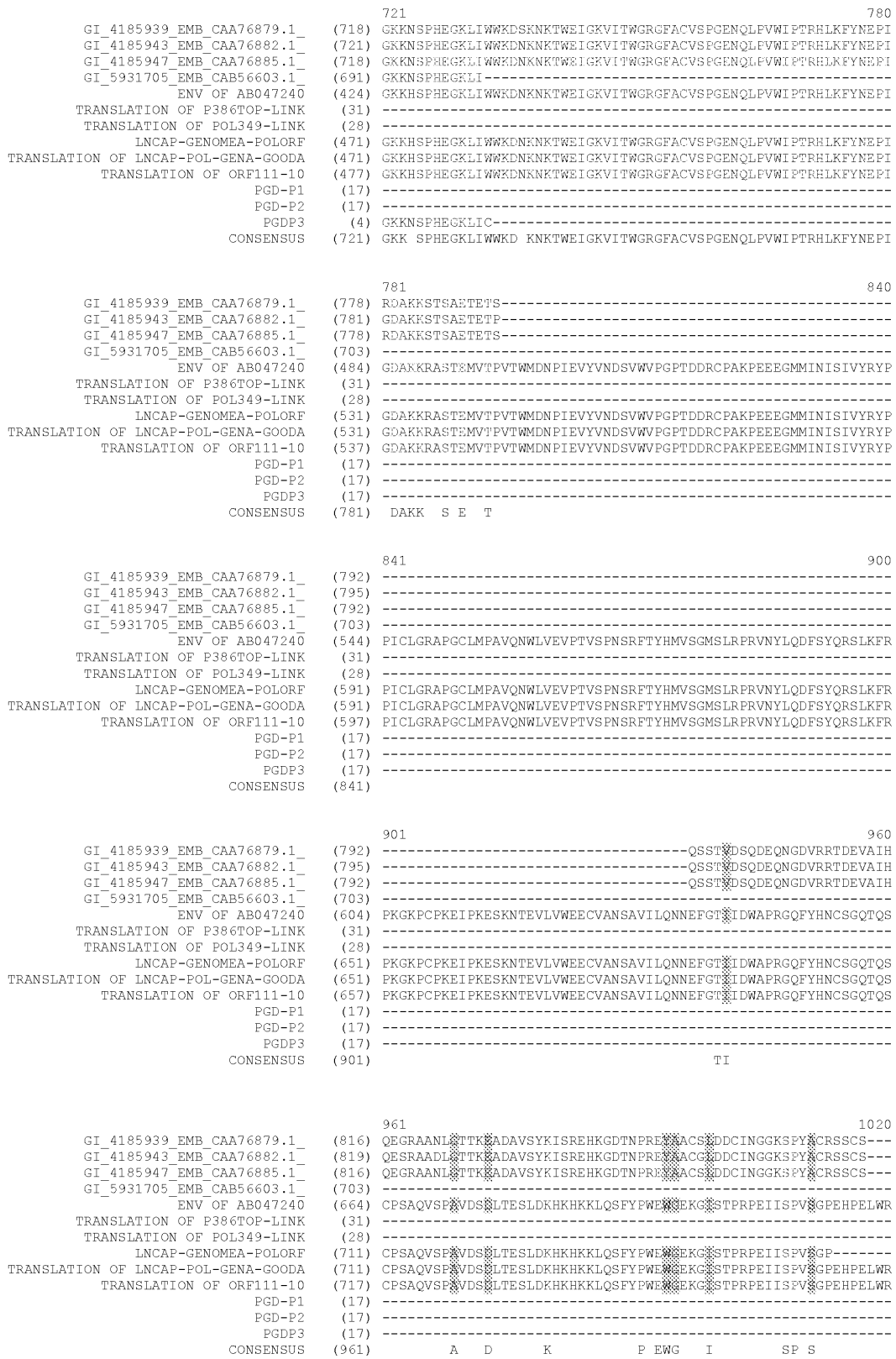


FIGURE 8 CONTD...

```
                                          1021      1035
        GI_4185939_EMB_CAA76879.1_   (873) ----------
        GI_4185943_EMB_CAA76882.1_   (876) ----------
        GI_4185947_EMB_CAA76885.1_   (873) ----------
        GI_5931705_EMB_CAB56603.1_   (703) ----------
                    ENV OF AB047240  (724) LWPDTTLEFGLEIKL
       TRANSLATION OF P386TOP-LINK    (31) ----------
        TRANSLATION OF POL349-LINK    (28) ----------
              LNCAP-GENOMEA-POLORF   (764) ----------
TRANSLATION OF LNCAP-POL-GENA-GOODA  (771) LWPDTTLEFGLEIKL
          TRANSLATION OF ORF111-10   (777) LWPDTTLEFGLEIKL
                           PGD-P1    (17) ----------
                           PGD-P2    (17) ----------
                            PGDP3    (17) ----------
                        CONSENSUS  (1021)
```

FIGURE 9

```
                                          1                                                          60
 GI_4185940_EMB_CAA76880.1_      (1) ------------------------------------------------------------
 GI_4185944_EMB_CAA76883.1_      (1) ------------------------------------------------------------
 GI_4185948_EMB_CAA76886.1_      (1) ------------------------------------------------------------
 GI_5931706_EMB_CAB56604.1_      (1) ------------------------------------------------------------
             ENV OF AB047240     (1) MATLIGQGRLRIITLCGNDPDKITVPFNKQQVRQAFISSGAWQIGLANFLGIIDNHYPKT
TRANSLATION OF E207TOP-LINK      (1) ------------------------------------------------------------
 TRANSLATION OF ENV287-LINK      (1) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23      (1) ------------------------------------------------------------
                      PGD-E1      (1) ------------------------------------------------------------
                      PGD-E2      (1) ------------------------------------------------------------
                      PGD-E3      (1) ------------------------------------------------------------
                  CONSENSUS       (1)

                                          61                                                         120
 GI_4185940_EMB_CAA76880.1_      (1) ------------------------------------------------------------
 GI_4185944_EMB_CAA76883.1_      (1) ------------------------------------------------------------
 GI_4185948_EMB_CAA76886.1_      (1) ------------------------------------------------------------
 GI_5931706_EMB_CAB56604.1_      (1) ------------------------------------------------------------
             ENV OF AB047240    (61) KIFQFLKLTTWILPKITRREPLENALTVFTDGSSNGKAAYTGPKERVIKTPYQSAQRAEL
TRANSLATION OF E207TOP-LINK      (1) ------------------------------------------------------------
 TRANSLATION OF ENV287-LINK      (1) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23      (1) ------------------------------------------------------------
                      PGD-E1      (1) ------------------------------------------------------------
                      PGD-E2      (1) ------------------------------------------------------------
                      PGD-E3      (1) ------------------------------------------------------------
                  CONSENSUS      (61)


                                          121                                                        180
 GI_4185940_EMB_CAA76880.1_      (1) ------------------------------------------------------------
 GI_4185944_EMB_CAA76883.1_      (1) ------------------------------------------------------------
 GI_4185948_EMB_CAA76886.1_      (1) ------------------------------------------------------------
 GI_5931706_EMB_CAB56604.1_      (1) ------------------------------------------------------------
             ENV OF AB047240    (121) VAVITVLQDFDQPINIISDSAYVVQATRDVETALIKYSTDDHLNQLFNLLQQTVRKRNFP
TRANSLATION OF E207TOP-LINK      (1) ------------------------------------------------------------
 TRANSLATION OF ENV287-LINK      (1) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23      (1) ------------------------------------------------------------
                      PGD-E1      (1) ------------------------------------------------------------
                      PGD-E2      (1) ------------------------------------------------------------
                      PGD-E3      (1) ------------------------------------------------------------
                  CONSENSUS      (121)
```

FIGURE 9 CONTD...

```
                                      181                                                                 240
GI_4185940_EMB_CAA76880.1_      (1) ------------------------------------------------------------
GI_4185944_EMB_CAA76883.1_      (1) ------------------------------------------------------------
GI_4185948_EMB_CAA76886.1_      (1) ------------------------------------------------------------
GI_5931706_EMB_CAB56604.1_      (1) ------------------------------------------------------------
            ENV OF AB047240   (181) FYITHIRAHTNLPGPLTKANEQADLLVSSAFIKAQELLALTHVNAAGLKNKFDVTWKQAK
TRANSLATION OF E207TOP-LINK     (1) ------------------------------------------------------------
 TRANSLATION OF ENV287-LINK     (1) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23     (1) ------------------------------------------------------------
                      PGD-E1    (1) ------------------------------------------------------------
                      PGD-E2    (1) ------------------------------------------------------------
                      PGD-E3    (1) ------------------------------------------------------------
                   CONSENSUS  (181)

                                      241                                                                 300
GI_4185940_EMB_CAA76880.1_      (1) ------------------------------------------------------------
GI_4185944_EMB_CAA76883.1_      (1) ------------------------------------------------------------
GI_4185948_EMB_CAA76886.1_      (1) ------------------------------------------------------------
GI_5931706_EMB_CAB56604.1_      (1) ------------------------------------------------------------
            ENV OF AB047240   (241) DIVQHCTQCQVLHLSTQEAGVNPRGLCPNALWQMDGTHVPSFGRLSYVHVTVDTYSHFIW
TRANSLATION OF E207TOP-LINK     (1) ------------------------------------------------------------
 TRANSLATION OF ENV287-LINK     (1) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23     (1) ------------------------------------------------------------
                      PGD-E1    (1) ------------------------------------------------------------
                      PGD-E2    (1) ------------------------------------------------------------
                      PGD-E3    (1) ------------------------------------------------------------
                   CONSENSUS  (241)

                                      301                                                                 360
GI_4185940_EMB_CAA76880.1_      (1) ------------------------------------------------------------
GI_4185944_EMB_CAA76883.1_      (1) ------------------------------------------------------------
GI_4185948_EMB_CAA76886.1_      (1) ------------------------------------------------------------
GI_5931706_EMB_CAB56604.1_      (1) ------------------------------------------------------------
            ENV OF AB047240   (301) ATCQTGESTSHVKKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYN
TRANSLATION OF E207TOP-LINK     (1) ------------------------------------------------------------
 TRANSLATION OF ENV287-LINK     (1) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23     (1) ------------------------------------------------------------
                      PGD-E1    (1) ------------------------------------------------------------
                      PGD-E2    (1) ------------------------------------------------------------
                      PGD-E3    (1) ------------------------------------------------------------
                   CONSENSUS  (301)

                                      361                                                                 420
GI_4185940_EMB_CAA76880.1_      (1) -------------------------MQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKL
GI_4185944_EMB_CAA76883.1_      (1) -------------------------MQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKL
GI_4185948_EMB_CAA76886.1_      (1) -------------------------MQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKL
GI_5931706_EMB_CAB56604.1_      (1) ------------------------------------------------------------
            ENV OF AB047240   (361) SQGQAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTTSAKQ
TRANSLATION OF E207TOP-LINK     (1) ------------------------------------------------------------
 TRANSLATION OF ENV287-LINK     (1) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23     (1) --------------------MNPSEMQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKL
                      PGD-E1    (1) ------------------------------------------------------------
                      PGD-E2    (1) ------------------------------------------------------------
                      PGD-E3    (1) ------------------------------------------------------------
                   CONSENSUS  (361)

                                      421                                                                 480
GI_4185940_EMB_CAA76880.1_     (35) PSTKKAEPPTWAQLKKLTQLATKYLENTKVTQTPESMLLAALMIVSMVVSLPMPAGAAAA
GI_4185944_EMB_CAA76883.1_     (35) PSTKKAEPPTWAQLKKLTQLATKYLENTKVTQTPESMLLAALMIVSMVVSLPMPAGAAAA
GI_4185948_EMB_CAA76886.1_     (35) PSTKKAEPPTWAQLKKLTQLATKYLENTKVTQTPESMLLAALMIVSMVVSLPMPAGAAAA
GI_5931706_EMB_CAB56604.1_      (1) ------------------------------------------------------------
            ENV OF AB047240   (421) HLTGKKHSPHEGKLIWWKDNKNKTWEIGKVITWGRGFACVSPGENQLPVWIPTRHLKFYN
TRANSLATION OF E207TOP-LINK     (1) ------------------------------------------------------------
 TRANSLATION OF ENV287-LINK     (1) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23     (40) PSTKKAEPPTWAQLKKLTQLATKYLENTKVTQTPESMLLAALMIVSMVVSLPMPAGAAAA
                      PGD-E1    (1) ------------------------------------------------------------
                      PGD-E2    (1) ------------------------------------------------------------
                      PGD-E3    (1) ------------------------------------------------------------
                   CONSENSUS  (421)
```

FIGURE 9 CONTD...

```
                                 481                                                        540
GI_4185940_EMB_CAA76880.1_    (95) NYTYWAYVPFPP-LIRAVTWMDNPTEVYVNDSVWVPGPIDDRCPAKPEEEGMMINISIGY
GI_4185944_EMB_CAA76883.1_    (95) NYTYWAYVPFPP-LIRAVTWMDNPIEVYVNDSVWVPGPTDDHCPAKPEEEGMMINISIGY
GI_4185948_EMB_CAA76886.1_    (95) NYTYWAYVPFPP-LIRAVTWMDNPTEVYVNDSVWVPGPIDDRCPAKPEEEGMMINISIGY
GI_5931706_EMB_CAB56604.1_     (1) -------------MVTPVTWMDNPIEVYVNDSVWVPGPTDDRCPAKPEEEGMMINISIGY
            ENV OF AB047240   (481) EPIGDAKKRASTEMVTPVTWMDNPIEVYVNDSVWVPGPTDDRCPAKPEEEGMMINISIVY
TRANSLATION OF E207TOP-LINK    (1) ------------------------------------------------------------
 TRANSLATION OF ENV287-LINK    (1) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23  (100) NYTYWAYVPFPP-LIRAVTWMDNPTEVYVNDSVWVPGPIDDRCPAKPEEEGMMINISIGY
                     PGD-E1    (1) ------------------------------------------------------------
                     PGD-E2    (1) ------------------------------------------------------------
                     PGD-E3    (1) ------------------------------------------------------------
                  CONSENSUS  (481)              LI  VTWMDNP EVYVNDSVWVPGP DD CPAKPEEEGMMINISI Y

                                 541                                                        600
GI_4185940_EMB_CAA76880.1_   (154) HYPPICLGRAPGCLMPAVQNWLVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQRSL
GI_4185944_EMB_CAA76883.1_   (154) RYPPICLGRAPGCLMPAVQNWLVEVPTVSPISRFTYHMVSGMSLRPRVNYLQDFSYQRSF
GI_4185948_EMB_CAA76886.1_   (154) HYPPICLGRAPGCIMPAVQNWLVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQRSL
GI_5931706_EMB_CAB56604.1_    (48) HYPPICLGRAPGCLMPAVQNWLVEVPTVSPNSRFTYHMVSGMSLRPRVNCLQDFSYQRSL
            ENV OF AB047240   (541) RYPPICLGRAPGCLMPAVQNWLVEVPTVSPNSRFTYHMVSGMSLRPRVNYLQDFSYQRSL
TRANSLATION OF E207TOP-LINK    (1) -----------------------------------------------------FSYQRSL
 TRANSLATION OF ENV287-LINK    (1) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23  (159) HYPPICLGRAPGCLMPAVQNWLVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQRSL
                     PGD-E1    (1) ------------------------------------------------------------
                     PGD-E2    (1) ------------------------------------------------------------
                     PGD-E3    (1) ------------------------------------------------------------
                  CONSENSUS  (541)  YPPICLGRAPGCLMPAVQNWLVEVPTVSP  RFTYHMVSGMSLRPRVN LQDFSYQRSL

                                 601                                                        660
GI_4185940_EMB_CAA76880.1_   (214) KFRPKGKPCPKEIPKESKNTEVLVWEECVANSAVILQNNEFGTIIDWAPRGQFYHNCSGQ
GI_4185944_EMB_CAA76883.1_   (214) KFRPKGKPCPKEIPKESKNTEVLVWEECVANSAVILQNNEFGTIIDWAPRGQFYHNCSGQ
GI_4185948_EMB_CAA76886.1_   (214) KFRPKGKPCPKEIPKESKNTEVLVWEECVANSAVILQNNEFGTIIDWAPRGQFYHNCSGQ
GI_5931706_EMB_CAB56604.1_   (108) KFRPKGKTCPKEIPKGSKNTEVLVWEECVANSVVILQNNEFGTIIDWAPRGQFYHNCSGQ
            ENV OF AB047240   (601) KFRPKGKPCPKEIPKESKNTEVLVWEECVANSAVILQNNEFGTIIDWAPRGQFYHNCSGQ
TRANSLATION OF E207TOP-LINK    (8) KFRPKGKPCPKEIPKESKNTEVL-------------------------------------
 TRANSLATION OF ENV287-LINK    (1) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23  (219) KFRPKGKPCPKEIPKESKNTEVLVWEECVANSAVILQNNEFGTIIDWAPRGQFYHNCSGQ
                     PGD-E1    (1) --RPKGKPCPKEIPKESC------------------------------------------
                     PGD-E2    (1) ------------------------------------------------------------
                     PGD-E3    (1) ------------------------------------------------------------
                  CONSENSUS  (601) KFRPKGKPCPKEIPKESKNTEVLVWEECVANS VILQNNEFGTIIDWAPRGQFYHNCSGQ

                                 661                                                        720
GI_4185940_EMB_CAA76880.1_   (274) TQSCQSAQVSPAVDSDLTESLDKHKHKKLQSFYPWEWGEKGISTPRPKIVSPVSGPEHPE
GI_4185944_EMB_CAA76883.1_   (274) TQSCPSAQVSPAVDSDLTESLDKHKHKKLQSFYPWEWGEKGISTPRPKIISPVSGPEHPE
GI_4185948_EMB_CAA76886.1_   (274) TQSCPSAQVSPAVDSDLTESLDKHKHKKLQSFYPWEWGEKGISTPRPKIVSPVSGPEHPE
GI_5931706_EMB_CAB56604.1_   (168) TQSCPSAQVSPAVDSDLTESLDKHKHKKLQSFYLWEWEEKGISTPRPKIISPVSGPEHPE
            ENV OF AB047240   (661) TQSCPSAQVSPAVDSDLTESLDKHKHKKLQSFYPWEWGEKGISTPRPEIISPVSGPEHPE
TRANSLATION OF E207TOP-LINK   (31) ------------------------------------------------------------
 TRANSLATION OF ENV287-LINK    (1) --------------SDLTESLDKHKHKKLQSFYPWEWGEKGI------------------
  TRANSLATION OF T20.22A-23  (279) TQSCPSAQVSPAVDSDLTESLDKHKHKKLQSFYPWEWGEKGISTPRPKIVSPVSGPEHPE
                     PGD-E1   (17) ------------------------------------------------------------
                     PGD-E2    (1) ------------------------------------------------------------
                     PGD-E3    (1) ------------------------------------------------------------
                  CONSENSUS  (661) TQSC SAQVSPAVDSDLTESLDKHKHKKLQSFYPWEWGEKGISTPRP IISPVSGPEHPE

                                 721                                                        780
GI_4185940_EMB_CAA76880.1_   (334) LWRLTVASHHIRIWSGNQTLETRDRKPFYTIDLNSSLTVPLQSCVKPPYMLVVGNIVIKP
GI_4185944_EMB_CAA76883.1_   (334) LWRLTVASHHIRIWSGNQTLETRDRKPFYTVDLNSSVTVPLQSCIKPPYMLVVGNIVIKP
GI_4185948_EMB_CAA76886.1_   (334) LWRLTVASHHIRIWSGNQTLETRDRKPFYTIDLNSSLTVPLQSCVKPPYMLVVGNIVIKP
GI_5931706_EMB_CAB56604.1_   (228) LWRLTVASHHIRIWSGNQTLETRYRKPFYTIDLNSILTVPLQSCVKPPYMLVVGNIVIKP
            ENV OF AB047240   (721) LW---------RL---------------------W---------------------P---
TRANSLATION OF E207TOP-LINK   (31) ------------------------------------------------------------
 TRANSLATION OF ENV287-LINK   (29) ------------------------------------------------------------
  TRANSLATION OF T20.22A-23  (339) LWRLTVASHHIRIWSGNQTLETRDRKPFYTIDLNSSLTVPLQSCVKPPYMLVVGNIVIKP
                     PGD-E1   (17) ------------------------------------------------------------
                     PGD-E2    (1) --------------------------------LNSSLTVPLQSCVKPC------------
                     PGD-E3    (1) ------------------------------------------------------------
                  CONSENSUS  (721) LW         RI                   LNS LTVPLQSCVKP
```

FIGURE 9 CONTD...

```
                                      781                                                      840
  GI_4185940_EMB_CAA76880.1_   (394) DSQTITCENCRLLTCIDSTFNWQHRILLVRAREGVWIPVSMDRPWEASPSVHILTEVLKG
  GI_4185944_EMB_CAA76883.1_   (394) DSQTITCENCRLLTCIDSTFNWQHRILLVRAREGVWIPVSMDRPWETSPSIHTLTEVLKG
  GI_4185948_EMB_CAA76886.1_   (394) DSQTITCENCRLLTCIDSTFNWQHRILLVRAREGVWIPVSMDRPWEASPSVHILTEVLKG
  GI_5931706_EMB_CAB56604.1_   (288) ASQTITCENCRLFTCIDSTFNWQHRILLVRAREGMWIPVSTDRPWEASPSIHILTEILKG
            ENV OF AB047240    (727) ----------------DSTLEWGLEIKL--------------------------------
 TRANSLATION OF E207TOP-LINK    (31) ------------------------------------------------------------
  TRANSLATION OF ENV287-LINK    (29) ------------------------------------------------------------
   TRANSLATION OF T20.22A-23   (399) DSQTITCENCRLLTCIDSTFNWQHRILLVRAREGVWIPVSMDRPWEASPSVHILTEVLKG
                      PGD-E1    (17) ------------------------------------------------------------
                      PGD-E2    (17) ------------------------------------------------------------
                      PGD-E3     (1) ------------------------------------------------------------
                   CONSENSUS   (781)                 DST  W   I L

                                      841                                                      900
  GI_4185940_EMB_CAA76880.1_   (454) VLNRSKRFIFTLIAVIMGLIAVTATAAVAGVALHSSVQSVNFVNDWQKNSTRLWNSQSSI
  GI_4185944_EMB_CAA76883.1_   (454) VLNRSKRFIFTLIAVIMGLIAVTATAAVAGVALHSSVQSVNFVNDWQKNSTRLWNSQSSI
  GI_4185948_EMB_CAA76886.1_   (454) VLNRSKRFIFTLIAVIMGLIAVTATAAVAGVALHSSVQSVNFVNDWQKNSTRLWNSQSSI
  GI_5931706_EMB_CAB56604.1_   (348) VLNRSKRFIFTLIAVIMGLIAVTATAAVAGVALHSSVQSVNFVNYWQKNSTRLWNSQSSI
            ENV OF AB047240    (739) ------------------------------------------------------------
 TRANSLATION OF E207TOP-LINK    (31) ------------------------------------------------------------
  TRANSLATION OF ENV287-LINK    (29) ------------------------------------------------------------
   TRANSLATION OF T20.22A-23   (459) VLNRSKRFIFTLIAVIMGLIAVTATAAVAGVALHSSVQSVNFVNDWQKNSTRLWNSQSSI
                      PGD-E1    (17) ------------------------------------------------------------
                      PGD-E2    (17) ------------------------------------------------------------
                      PGD-E3     (1) ------------------------------------------------------------
                   CONSENSUS   (841)

                                      901                                                      960
  GI_4185940_EMB_CAA76880.1_   (514) DQKLANQINDLRQTVIWMGDRLMSLEHRFQLQCDWNTSDFCITPQIYNESEHHWDMVRRH
  GI_4185944_EMB_CAA76883.1_   (514) DQKLANQINDLRQTVIWMGDRLMSLEHRFQLQCDWNTSDFSITPQIYNESEHHWDMVRRH
  GI_4185948_EMB_CAA76886.1_   (514) DQKLANQINDLRQTVIWMGDRLMSLEHRFQLQCDWNTSDFCITPQIYNESELHWDMVRRH
  GI_5931706_EMB_CAB56604.1_   (408) DQKLASQINDLRQTVIWMGDRLMTLEHHFQLQCDWNTSDFCITPQIYNESEHHWDMVRRH
            ENV OF AB047240    (739) ------------------------------------------------------------
 TRANSLATION OF E207TOP-LINK    (31) ------------------------------------------------------------
  TRANSLATION OF ENV287-LINK    (29) ------------------------------------------------------------
   TRANSLATION OF T20.22A-23   (519) DQKLANQINDLRQTVIWMGDRLMSLEHRFQLQCDWNTSDFCITPQIYNESEHHWDMVRRH
                      PGD-E1    (17) ------------------------------------------------------------
                      PGD-E2    (17) ------------------------------------------------------------
                      PGD-E3     (1) ------------------------------------------------------------
                   CONSENSUS   (901)

                                      961                                                     1020
  GI_4185940_EMB_CAA76880.1_   (574) LQGREDNLTLDISKLKEQIFEASKAHLNLVPGTEAIAGVADGLANLNPVTWVKTIGSTTI
  GI_4185944_EMB_CAA76883.1_   (574) LQGREDNLTLDISKLKEQIFEASKAHLNLVPGTEAIAGVADGLANLNPVTWVKTIGSTTI
  GI_4185948_EMB_CAA76886.1_   (574) LQGREDNLTLDISKLKEQIFEASKAHLNLVPGTEAIAGVADGLANLNPVTWVKTIGSTTI
  GI_5931706_EMB_CAB56604.1_   (468) LQGREDNLTLDISKLKEQIFEASKAHLNLVPGTEAIAGVADGLANLNPVTWIKTIRSTMI
            ENV OF AB047240    (739) ------------------------------------------------------------
 TRANSLATION OF E207TOP-LINK    (31) ------------------------------------------------------------
  TRANSLATION OF ENV287-LINK    (29) ------------------------------------------------------------
   TRANSLATION OF T20.22A-23   (579) LQGREDNLTLDISKLKEQIFEASKAHLNLVPGTEAIAGVADGLANLNPVTWVKTIGSTTI
                      PGD-E1    (17) ------------------------------------------------------------
                      PGD-E2    (17) ------------------------------------------------------------
                      PGD-E3     (1) ------------------------------------------------------------
                   CONSENSUS   (961)

                                      1021                                                     1081
  GI_4185940_EMB_CAA76880.1_   (634) INLILILVCLFCLLLVCRCTQQLRRDSDHRERAMMTMAVLSKRKGGNVGKSKRDQIVTVSV
  GI_4185944_EMB_CAA76883.1_   (634) INLILILVCLFCLLLVCRCTQQLRRDSDHRERAMMTMAVLSKRKGGNVGKSKRDQIVTVSV
  GI_4185948_EMB_CAA76886.1_   (634) INLILILVCLFCLLLVCRCTQQLRRDSDHRERAMMTMAVLSKRKGGNVGKSKRDQIVTVSV
  GI_5931706_EMB_CAB56604.1_   (528) INLILIVVCLFCLLLVCRCTQQLRRDSDIENGP----------------------------
            ENV OF AB047240    (739) -------------------------------------------------------------
 TRANSLATION OF E207TOP-LINK    (31) -------------------------------------------------------------
  TRANSLATION OF ENV287-LINK    (29) -------------------------------------------------------------
   TRANSLATION OF T20.22A-23   (639) INLILILVCLFCLLLVCRCTQQLRRDSDHRERAMMTMAVLSKRKGGNVGKSKRDQIVTVSV
                      PGD-E1    (17) -------------------------------------------------------------
                      PGD-E2    (17) -------------------------------------------------------------
                      PGD-E3     (1) -----------------RCTQQLRRDSDHRERA----------------------------
                   CONSENSUS  (1021)                  RCTQQLRRDSD
```

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 10 17 6402

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X,D	TÖNJES R R ET AL: "Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K.", JOURNAL OF VIROLOGY. UNITED STATES NOV 1999, vol. 73, no. 11, November 1999 (1999-11), pages 9187-9195, XP002238923, ISSN: 0022-538X	1-10	INV. C12Q1/70 G01N33/574 C07K14/15 A61K48/00 A61P35/00 C12Q1/68 A61K39/21
A	* abstract * * page 9187, column 2, paragraph 3 - page 9189, column 1, paragraph 6 * * page 9189, column 2, paragraph 4 - page 9190, column 1, paragraph 1 * * page 9190, column 2, paragraph 2 - page 9193, column 1, paragraph 1 * * page 9193, column 2, paragraph 2 - page 9194, column 1, paragraph 3; figures 1-3,5,6 * * EMBL Acc No Y17832 *	11-15	
	-----		TECHNICAL FIELDS SEARCHED (IPC)
X,D	BARBULESCU M ET AL: "MANY HUMAN ENDOGENOUS RETROVIRUS K (HERV-K) PROVIRUSES ARE UNIQUE TO HUMANS", CURRENT BIOLOGY, CURRENT SCIENCE,, GB, vol. 9, 1999, pages 861-868, XP000953273, ISSN: 0960-9822	1-11	C12Q G01N A61K C07K
A	* GenBank Acc No AF164614 * * abstract * * page 861, column 2, paragraph 2 - page 862, column 1, paragraph 1 * * page 864, column 2, paragraph 2 - page 866, column 1, paragraph 1 * * page 867, column 1, paragraph 2 - column 2, paragraph 3; figures 1-4; tables 1,2 *	12-15	
	----- -/--		

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	13 April 2011	Tilkorn, A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 10 17 6402

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X	MAYER J ET AL: "Chromosomal assignment of human endogenous retrovirus K (HERV-K) env open reading frames.", CYTOGENETICS AND CELL GENETICS, vol. 79, no. 1-2, 1997, pages 157-161, XP009009741, ISSN: 0301-0171	1-11	
A	Embl Acc NO AF023261; * abstract * * page 158, column 2, paragraph 3 - page 159, column 1, paragraph 2 * * page 160, column 1, paragraph 2 *	12-15	
X,D	MAYER JENS ET AL: "An almost-intact human endogenous retrovirus K on human chromosome 7.", NATURE GENETICS, vol. 21, no. 3, March 1999 (1999-03), pages 257-258, XP002239083, ISSN: 1061-4036	1-11	
A	GenBank Acc No:AF074086; * abstract; figure 1 *	12-15	TECHNICAL FIELDS SEARCHED (IPC)
X,D	US 5 858 723 A (MUELLER-LANTZSCH NIKOLAUS ET AL) 12 January 1999 (1999-01-12)	1-11	
A	* abstract * * column 4, line 48 - column 5, line 28 * * column 6, line 56 - column 7, line 2 * * column 7, line 42 - column 9, line 22 * * column 10, line 6 - line 48 * * column 13, line 29 - column 14, line 18; examples 4-6; tables I-IV *	12-15	

-/--

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	13 April 2011	Tilkorn, A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 10 17 6402

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X,D	LOEWER R ET AL: "The viruses in all of us: Characteristics and biological significance of human endogenous retrovirus sequences", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 93, no. 11, May 1996 (1996-05), pages 5177-5184, XP002233322, ISSN: 0027-8424	1-11	
A	* the whole document *	12-15	

X	DATABASE EMBL [Online] 16 August 2000 (2000-08-16), "Human endogenous retrovirus HERV-K(II) DNA, complete sequence and flanking region.", XP002626827, retrieved from EBI accession no. EMBL:AB047240 Database accession no. AB047240	1,2,4,5	
A	* compound *	3,6-15	TECHNICAL FIELDS SEARCHED (IPC)

X	BOLLER K ET AL: "Characterization of the antibody response specific for the human endogenous retrovirus HTDV/HERV-K", JOURNAL OF VIROLOGY, THE AMERICAN SOCIETY FOR MICROBIOLOGY, US, vol. 71, no. 6, 1 June 1997 (1997-06-01), pages 4581-4588, XP002967165, ISSN: 0022-538X	1,2,4-11	
A	* Y08032 *	3,12-15	
	-/--		

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	13 April 2011	Tilkorn, A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 10 17 6402

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
	-& DATABASE UniProt [Online] 1 February 1997 (1997-02-01), "RecName: Full=Gag polyprotein;", XP2626899, retrieved from EBI accession no. UNIPROT:Q96897 Database accession no. Q96897 * compound * -----		
X	DATABASE UniProt [Online] 1 July 1989 (1989-07-01), "RecName: Full=HERV-K_5q33.3 provirus ancestral Pol protein; AltName: Full=HERV-K10 Pol protein; AltName: Full=HERV-K107 Pol protein; Includes: RecName: Full=Reverse transcriptase; Short=RT; EC=2.7.7.49;", XP002629551, retrieved from EBI accession no. UNIPROT:P10266 Database accession no. P10266	1,2,4,5	
A	* compound * -----	3,6-15	TECHNICAL FIELDS SEARCHED (IPC)
X,P	SMITH R D ET AL: "The human endogenous retrovirus HERV-K in prostate cancer.", AMERICAN JOURNAL OF HUMAN GENETICS, vol. 69, no. 4 Supplement, October 2001 (2001-10), page 275, XP009009588, 51st Annual Meeting of the American Society of Human Genetics;San Diego, California, USA; October 12-16, 2001 ISSN: 0002-9297 * the whole document * -----	1-15	
	-/--		

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	13 April 2011	Tilkorn, A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

Europäisches Patentamt
European Patent Office
Office européen des brevets

EUROPEAN SEARCH REPORT

Application Number
EP 10 17 6402

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (IPC)
X,P	SMITH RICHARD DARON ET AL: "Human endogenous retrovirus HERV-K expression in prostate cancer.", JOURNAL OF UROLOGY, vol. 165, no. 5 Supplement, May 2001 (2001-05), pages 136-137, XP009009589, Annual Meeting of the American Urological Association, Inc.;Anaheim, California, USA; June 02-07, 2001 ISSN: 0022-5347 * the whole document * -----	1-15	

TECHNICAL FIELDS SEARCHED (IPC)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
Munich	13 April 2011	Tilkorn, A

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

ANNEX TO THE EUROPEAN SEARCH REPORT ON EUROPEAN PATENT APPLICATION NO. EP 10 17 6402

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

13-04-2011

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5858723 A	12-01-1999	NONE	

EPO FORM P0459

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- WO 0004149 A **[0370]**
- US 5707829 A **[0371]**
- EP 0509612 B **[0371]**
- EP 0505012 B **[0371]**
- US 4959314 A **[0371]**
- US 5530101 A **[0371]**
- US 5585089 A **[0371]**
- WO 9824893 A **[0371]**
- WO 9110741 A **[0371]**
- WO 9630498 A **[0371]**
- WO 9402602 A **[0371]**
- US 5939598 A **[0371]**
- WO 9633735 A **[0371]**
- WO 9314778 A **[0371]**
- WO 9007936 A **[0371]**
- WO 9403622 A **[0371]**
- WO 9325698 A **[0371]**
- WO 9325234 A **[0371]**
- US 5219740 A **[0371]**
- WO 9311230 A **[0371]**
- WO 9310218 A **[0371]**
- US 4777127 A **[0371]**
- GB 2200651 A **[0371]**
- EP 0345242 A **[0371]**
- WO 9102805 A **[0371]**
- WO 9412649 A **[0371]**
- WO 9303769 A **[0371]**
- WO 9319191 A **[0371]**
- WO 9428938 A **[0371]**
- WO 9511984 A **[0371]**
- WO 9500655 A **[0371]**
- US 5814482 A **[0371]**
- WO 9507994 A **[0371]**
- WO 9617072 A **[0371]**
- WO 9530763 A **[0371]**
- WO 9742338 A **[0371]**
- WO 9011092 A **[0371]**
- US 5580859 A **[0371]**
- US 5422120 A **[0371]**
- WO 9513796 A **[0371]**
- WO 9423697 A **[0371]**
- WO 9114445 A **[0371]**
- EP 0524968 A **[0371]**
- US 5206152 A **[0371]**
- WO 9211033 A **[0371]**
- US 5149655 A **[0371]**
- WO 9014837 A **[0371]**
- WO 0007621 A **[0371]**
- GB 2220221 A **[0371]**
- EP 0689454 A **[0371]**
- EP 0835318 A **[0371]**
- EP 0735898 A **[0371]**
- EP 0761231 A **[0371]**
- WO 9952549 A **[0371]**
- WO 0121207 A **[0371]**
- WO 0121152 A **[0371]**
- WO 0062800 A **[0371]**
- WO 0023105 A **[0371]**
- WO 9911241 A **[0371]**
- WO 9857659 A **[0371]**
- WO 9313202 A **[0371]**
- US 5858723 A **[0371]**
- US 5010175 A **[0371]**
- WO 9117823 A **[0371]**
- US 4816567 A **[0371]**
- US 5807522 A **[0371]**
- EP 0799897 A **[0371]**
- WO 9729212 A **[0371]**
- WO 9727317 A **[0371]**
- EP 0785280 A **[0371]**
- WO 9702357 A **[0371]**
- US 5593839 A **[0371]**
- US 5578832 A **[0371]**
- EP 0728520 A **[0371]**
- US 5599695 A **[0371]**
- EP 0721016 A **[0371]**
- US 5556752 A **[0371]**
- WO 9522058 A **[0371]**
- US 5631734 A **[0371]**
- US 5134854 A **[0371]**
- US 5445934 A **[0371]**
- WO 9535505 A **[0371]**
- US 5800992 A **[0371]**
- WO 9202526 A **[0371]**
- US 5124246 A **[0371]**
- US 4683195 A **[0371]**
- US 4683202 A **[0371]**

Non-patent literature cited in the description

- **ALFONSO GENNARO.** Remington: The Science and Practice of Pharmacy. Lippincott, Williams, & Wilkins, 1995 **[0145]**
- **MAYER et al.** *Nat. Genet.,* 1999, vol. 21 (3), 257-258 **[0371]**
- **FARRELL.** RNA Methodologies. Academic Press, 1998 **[0371]**
- **YANG et al.** *Proc Natl Acad Sci USA,* 1999, vol. 96 (23), 13404-8 **[0371]**
- **ROBBINS et al.** *Clin Lab Sci,* 1997, vol. 10 (5), 265-71 **[0371]**
- **YLIKOSKI et al.** *Clin Chem,* 1999, vol. 45 (9), 1397-407 **[0371]**
- **YLIKOSKI et al.** *Biotechniques,* 2001, vol. 30, 832-840 **[0371]**
- **SHIRAHATA ; PEGG.** *J Biol. Chem.,* 1986, vol. 261 (29), 13833-7 **[0371]**
- **SAMBROOK et al.** Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, 1989 **[0371]**
- Short protocols in molecular biology. 1999 **[0371]**
- Current Protocols in Molecular Biology. 1987, vol. 30 **[0371]**
- **BERKHOUT et al.** *J. Virol.,* 1999, vol. 73, 2365-2375 **[0371]**
- **LÖWER et al.** *J. Virol.,* 1995, vol. 69, 141-149 **[0371]**
- **MAGIN et al.** *J. Virol.,* 1999, vol. 73, 9496-9507 **[0371]**
- **MAGIN-LACHMANN.** *J Virol.,* 2001, vol. 75 (21), 10359-71 **[0371]**
- **HASHIDO et al.** *Biochem. Biophys. Res. Comm.,* 1992, vol. 187, 1241-1248 **[0371]**
- **GEYSEN et al.** *PNAS USA,* 1984, vol. 81, 3998-4002 **[0371]**
- **CARTER.** *Methods Mol Biol,* 1994, vol. 36, 207-23 **[0371]**
- **JAMESON, BA et al.** *CABIOS,* 1988, vol. 4 (1), 181-186 **[0371]**
- **RADDRIZZANI ; HAMMER.** *Brief Bioinform,* 2000, vol. 1 (2), 179-89 **[0371]**
- **DE LALLA et al.** *J. Immunol.,* 1999, vol. 163, 1725-29 **[0371]**
- **BRUSIC et al.** *Bioinformatics,* 1998, vol. 14 (2), 121-30 **[0371]**
- **MEISTER et al.** *Vaccine,* 1995, vol. 13 (6), 581-91 **[0371]**
- **ROBERTS et al.** *AIDS Res Hum Retroviruses,* 1996, vol. 12 (7), 593-610 **[0371]**
- **MAKSYUTOV ; ZAGREBELNAYA.** *Comput Appl Biosci,* 1993, vol. 9 (3), 291-7 **[0371]**
- **FELLER ; DE LA CRUZ.** *Nature,* 1991, vol. 349 (6311), 720-1 **[0371]**
- **HOPP.** *Peptide Research,* 1993, vol. 6, 183-190 **[0371]**
- **WELLING et al.** *FEBS Lett.,* 1985, vol. 188, 215-218 **[0371]**
- **DAVENPORT et al.** *Immunogenetics,* 1995, vol. 42, 392-297 **[0371]**
- **SMITH ; WATERMAN.** *Adv. Appl. Math.,* 1981, vol. 2, 482-489 **[0371]**
- **GO et al.** *Int. J. Peptide Protein Res.,* 1980, vol. 15, 211 **[0371]**
- **QUEROL et al.** *Prot. Eng.,* 1996, vol. 9, 265 **[0371]**
- **OLSEN ; THOMSEN.** *J. Gen. Microbiol.,* 1991, vol. 137, 579 **[0371]**
- **CLARKE et al.** *Biochemistry,* 1993, vol. 32, 4322 **[0371]**
- **WAKARCHUK et al.** *Protein Eng.,* 1994, vol. 7, 1379 **[0371]**
- **TOMA et al.** *Biochemistry,* 1991, vol. 30, 97 **[0371]**
- **HAEZERBROUCK et al.** *Protein Eng.,* 1993, vol. 6, 643 **[0371]**
- **MASUL et al.** *Appl. Env. Microbiol.,* 1994, vol. 60, 3579 **[0371]**
- **BREEDVELD.** *Lancet,* 2000, vol. 355 (9205), 735-740 **[0371]**
- **GORMAN ; CLARK.** *Semin. Immunol.,* 1990, vol. 2, 457-466 **[0371]**
- **JONES et al.** *Nature,* 1986, vol. 321, 522-525 **[0371]**
- **MORRISON et al.** *Proc. Natl. Acad. Sci, US.A.,* 1984, vol. 81, 6851-6855 **[0371]**
- **MORRISON ; OI.** *Adv. Immunol.,* 1988, vol. 44, 65-92 **[0371]**
- **VERHOEYER et al.** *Science,* 1988, vol. 239, 1534-1536 **[0371]**
- **PADLAN.** *Molec. Immun.,* 1991, vol. 28, 489-498 **[0371]**
- **PADLAN.** *Molec. Immunol.,* 1994, vol. 31 (3), 169-217 **[0371]**
- **KETTLEBOROUGH, C.A. et al.** *Protein Eng.,* 1991, vol. 4 (7), 773-83 **[0371]**
- **CHOTHIA et al.** *J. Mol. Biol.,* 1987, vol. 196, 901-917 **[0371]**
- **KABAT et al.** *U.S. Dept. of Health and Human Services NIH Publication No. 91-3242,* 1991 **[0371]**
- **FINDEIS et al.** *Trends Biotechnol.,* 1993, vol. 11, 202 **[0371]**
- **CHIOU et al.** Gene Therapeutics: Methods And Applications Of Direct Gene Transfer. 1994 **[0371]**
- **WU et al.** *J. Biol. Chem.,* 1988, vol. 263, 621 **[0371]**
- **WU et al.** *J. Biol. Chem.,* 1994, vol. 269, 542 **[0371]**
- **ZENKE et al.** *Proc. Natl. Acad. Sci. (USA),* 1990, vol. 87, 3655 **[0371]**
- **WU et al.** *J. Biol. Chem.,* 1991, vol. 266, 338 **[0371]**
- **JOLLY.** *Cancer Gene Therapy,* 1994, vol. 1, 51 **[0371]**
- **KIMURA.** *Human Gene Therapy,* 1994, vol. 5, 845 **[0371]**
- **CONNELLY.** *Human Gene Therapy,* 1995, vol. 1, 185 **[0371]**
- **KAPLITT.** *Nature Genetics,* 1994, vol. 6, 148 **[0371]**
- **CURIEL.** *Hum. Gene Ther.,* 1992, vol. 3, 147 **[0371]**

- **WU.** *J. Biol. Chem.,* 1989, vol. 264, 16985 **[0371]**
- **PHILIP.** *Mol. Cell Biol.,* 1994, vol. 14, 2411 **[0371]**
- **WOFFENDIN.** *Proc. Natl. Acad. Sci.,* 1994, vol. 91, 11581 **[0371]**
- Vaccine Design - the subunit and adjuvant approach. 1995 **[0371]**
- **MCSHARRY.** *Antiviral Res,* 1999, vol. 43 (1), 1-21 **[0371]**
- **KUHELJ et al.** *J Biol Chem,* 2001, vol. 276 (20), 16674-82 **[0371]**
- **SCHOMMER et al.** *J Gen Virol,* 1996, vol. 77, 375-379 **[0371]**
- **MAGIN et al.** *Virology,* 2000, vol. 274, 11-16 **[0371]**
- **BOESE et al.** *FEBS Lett,* 2001, vol. 493 (2-3), 117-21 **[0371]**
- **LARSSON, E. et al.** *Current Topics in Microbiology and Immunology,* 1989, vol. 148, 115 **[0371]**
- **MARIANI-COSTANTINI et al.** *J. Virol.,* 1989, vol. 63, 4982 **[0371]**
- **SHIH et al.** *Virology,* 1991, vol. 182, 495 **[0371]**
- **TÖNJES et al.** *J. AIDS Hum. Retrovir.,* 1996, vol. 13 (1), S261-S267 **[0371]**
- **BARBULESCU et al.** *Curr. Biol.,* 1999, vol. 9, 861 **[0371]**
- **ONO et al.** *J. Virol.,* 1986, vol. 58, 937 **[0371]**
- **LÖWER et al.** *Proc. Natl. Acad. Sci USA,* 1993, vol. 90, 4480 **[0371]**
- **ONO et al.** *J. Virol.,* 1986, vol. 60, 589 **[0371]**
- **BOLLER et al.** *Virol.,* 1993, vol. 196, 349 **[0371]**
- **YANG et al.** *Proc. Natl. Acad. Sci USA,* 1999, vol. 96, 13404 **[0371]**
- **MUELLER- LANTZSCH et al.** *AIDS Research and Human Retroviruses,* 1993, vol. 9, 343-350 **[0371]**
- **HERBST et al.** *Amer. J. Pathol.,* 1996, vol. 149, 1727 **[0371]**
- **LÖWER et al.** *Proc. Natl. Acad. Sci USA,* 1996, vol. 93, 5177 **[0371]**
- **LÖWER et al.** *Virology,* 1993, vol. 192, 501 **[0371]**
- Genbank. AB047240 **[0371]**
- **ANDERSSON et al.** *J. Gen. Virol.,* 1999, vol. 80, 255-260 **[0371]**
- **ZSÍROS et al.** *J. Gen. Virol.,* 1998, vol. 79, 61-70 **[0371]**
- **TÖNJES et al.** *J. Virol.,* 1999, vol. 73, 9187-9195 **[0371]**
- **JOHNSTON et al.** *Ann Neurol,* 2001, vol. 50 (4), 434-42 **[0371]**
- **MEDSTRAND et al.** *J Virol,* 1998, vol. 72 (12), 9782-7 **[0371]**
- **MERRIFELD.** *J. Am. Chem. Soc.,* 1963, vol. 85, 2149 **[0371]**
- **CAPRINO ; HAN.** *J. Org. Chem.,* 1972, vol. 37, 3404 **[0371]**
- **MILSTEIN ; KOHLER.** *Nature,* 1975, vol. 256, 495-497 **[0371]**
- **ALTSCHUL et al.** *Nucleic Acids Res.,* 1997, vol. 25, 3389-3402 **[0371]**
- **BRUTLAG et al.** *Comp. Chem.,* 1993, vol. 17, 203 **[0371]**
- **SCHENA et al.** *Proc Natl Acad Sci U S A.,* 1996, vol. 93 (20), 10614-9 **[0371]**
- **SCHENA et al.** *Science,* 1995, vol. 270 (5235), 467-70 **[0371]**
- **SHALON et al.** *Genome Res.,* 1996, vol. 6 (7), 639-45 **[0371]**
- **PAPPALARADO et al.** *Sem. Radiation Oncol.,* 1998, vol. 8, 217 **[0371]**
- **RAMSAY.** *Nature Biotechnol.,* 1998, vol. 16, 40 **[0371]**
- **MULLIS et al.** *Meth. Enzymol.,* 1987, vol. 155, 335 **[0371]**
- **SAIKI et al.** *Science,* 1985, vol. 239, 487 **[0371]**
- **HANAHAN et al.** *Cell,* 2000, vol. 100, 57-70 **[0371]**
- **WEISSMAN SM.** *Mol Biol. Med.,* 1987, vol. 4 (3), 133-143 **[0371]**
- **PATANJALI et al.** *Proc. Natl. Acad. Sci. USA,* 1991, 88 **[0371]**
- **SIMONE et al.** *Am J Pathol.,* 2000, vol. 156 (2), 445-52 **[0371]**
- **CLAVERIE.** *Meth. Enzymol.,* 1996, vol. 266, 212-227 **[0371]**
- Automated DNA Sequencing and Analysis Techniques. Academic Press, 1994, 267 **[0371]**
- **CLAVERIE et al.** *Comput. Chem.,* 1993, vol. 17, 191 **[0371]**
- **ALTSCHUL.** *J. Mol. Biol.,* 1990, vol. 215, 403-410 **[0371]**
- **PEARSON ; LIPMAN.** *PNAS,* 1988, vol. 85, 2444 **[0371]**
- **LUO et al.** *Nature Med,* 1999, vol. 5, 117-122 **[0371]**
- **HIGGINS ; SHARP.** *CABIOS,* 1989, vol. 5, 151-153 **[0371]**
- **DELLI BOVI et al.** *Cancer Res.,* 1986, vol. 46, 6333-6338 **[0371]**
- **CESARONE, C. et al.** *Anal Biochem,* 1979, vol. 100, 188-197 **[0371]**
- **SOUTHERN, E. M.** *J. Mol. Biol.,* 1975, vol. 95, 503-517 **[0371]**
- **FEINBERG, A. P. et al.** *Anal. Biochem.,* 1983, vol. 132, 6-13 **[0371]**
- **WRIGHT ; MANOS et al.** PCR Protocols. Academic Press, 1990, 153-158 **[0371]**
- **KEOWN et al.** *Methods in Enzymology,* 1990, vol. 185, 527-537 **[0371]**
- **MARKS et al.** *Brit. J. Urol.,* 1995, vol. 75, 225 **[0371]**
- **SKEA et al.** *J. Immunol.,* 1993, vol. 151, 3557 **[0371]**
- **MATHER et al.** *J. Nucl. Med.,* 1990, vol. 31, 692 **[0371]**
- **ZHANG et al.** *Nucl. Med. Biol.,* 1992, vol. 19, 607 **[0371]**

专利名称(译)	内源性逆转录病毒在前列腺癌中上调		
公开(公告)号	EP2336368A1	公开(公告)日	2011-06-22
申请号	EP2010176402	申请日	2001-12-07
[标]申请(专利权)人(译)	诺华疫苗和诊断		
申请(专利权)人(译)	诺华疫苗与诊断，INC.		
当前申请(专利权)人(译)	诺华疫苗与诊断，INC.		
[标]发明人	GARCIA PABLO HARDY STEPHEN F WILLIAMS LEWIS T ESCOBEDO JAIME		
发明人	GARCIA, PABLO HARDY, STEPHEN F. WILLIAMS, LEWIS T. ESCOBEDO, JAIME		
IPC分类号	C12Q1/70 G01N33/574 C07K14/15 A61K48/00 A61P35/00 C12Q1/68 A61K39/21 G01N33/53 A61K38/00 A61K39/00 A61K39/395 A61K45/00 A61P3/10 A61P15/00 A61P25/00 C07K16/10 C12N7/00 C12N15/09 C12P21/08 C12Q1/02 G01N33/569 G01N33/577		
CPC分类号	A61K39/00 A61K2039/505 A61K2039/53 A61P3/10 A61P13/08 A61P15/00 A61P25/00 A61P31/14 A61P35/00 A61P37/04 C07K14/005 C12N7/00 C12N2740/10021 C12N2740/10022 C12Q1/6886 C12Q1/702 C12Q2600/158 G01N33/57434 G01N2333/15		
代理机构(译)	MARSHALL，CAMERON JOHN		
优先权	60/251830 2000-12-07 US 10/016604 2001-12-07 US		
外部链接	Espacenet		

摘要(译)

HML-2家族的人内源性逆转录病毒在前列腺肿瘤中表现出上调的表达。该发现可用于前列腺癌筛查，诊断和治疗。