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

(11) EP 1 415 005 B1

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
21.11.2012 Bulletin 2012/47

(21) Application number: **01996222.4**

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

(51) Int Cl.:
C07K 14/15 *(2006.01)* ***C12N 7/00*** *(2006.01)*
C12Q 1/68 *(2006.01)* ***G01N 33/574*** *(2006.01)*

(86) International application number:
PCT/US2001/047824

(87) International publication number:
WO 2002/046477 (13.06.2002 Gazette 2002/24)

(54) **ENDOGENOUS RETROVIRUSES UP-REGULATED IN PROSTATE CANCER**

HOCHREGULIERTE ENDOGENE RETROVIREN IN PROSTATAKREBS

RETROVIRUS ENDOGENES REGULES POSITIVEMENT DANS LE CANCER DE LA PROSTATE

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR

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

(43) Date of publication of application:
06.05.2004 Bulletin 2004/19

(60) Divisional application:
10176402.5 / 2 336 368
10176900.8 / 2 339 035

(73) Proprietor: **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 et al**
Carpmaels & Ransford
One Southampton Row
London
WC1B 5HA (GB)

(56) References cited:
US-A- 5 858 723

- **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 cited in the application**
- **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 cited in the application**
- **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**
- **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 cited in the application**
- **HUANG HAOJIE ET AL: "FRA7G extends over a broad region: Coincidence of human endogenous retroviral sequences (HERV-H) and small polydispersed circular DNAs (spcDNA) and fragile sites." ONCOGENE, vol. 16, no. 18, 7 May 1998 (1998-05-07), pages 2311-2319, XP009009633 ISSN: 0950-9232**

Printed by Jouve, 75001 PARIS (FR)

(Cont. next page)

- 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 cited in the application
- MAGIN CHRISTINE ET AL: "Corf, the Rev/Rex homologue of HTDV/HERV-K, encodes an arginine-rich nuclear localization signal that exerts a trans-dominant phenotype when mutated." VIROLOGY, vol. 274, no. 1, 15 August 2000 (2000-08-15), pages 11-16, XP002239089 ISSN: 0042-6822 cited in the application
- 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
- 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

Description

TECHNICAL FIELD

[0001] 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

[0002] 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.

[0003] 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 **pmstate-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.

[0004] 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.

[0005] 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

[0006] 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.

[0007] 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, wherein the patient sample contains prostate cells and for wherein the patient is suspected of having prostate cancer wherein up-regulation of expression of at least 150% relative to a negative control is indicative of prostate cancer.

[0008] 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

[0009] 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.

[0010] 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.

[0011] 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.

[0012] The patient is generally a human, preferably human male, and more preferably an adult human male.

[0013] 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.

[0014] Methods of the invention can be conducted *in vitro* or *in vivo.*

[0015] 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

[0016] 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, S0, 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., contiguous 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, 245, 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 m 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.

[0017] 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.

[0018] 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.

[0019] 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.

[0020] 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.

[0021] 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.

[0022] 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.

[0023] 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.

[0024] Where N_1 is present, it is preferably at the 5' end of the mRNA (*i.e*. 5'- N_1-...).

[0025] Where N_5 is present, it is preferably immediately before a 3' polyA tail (*i.e*.... -N_5-polyA-3').

[0026] 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.

[0027] The RNA will generally have a 5' cap.

B.1 -Enriching RNA in a sample

[0028] 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.

[0029] 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.

[0030] Methods for removing DNA from biological samples without removing mRNA are well known [*e.g*. appendix C of ref. 2] and include DNase digestion.

[0031] 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.

[0032] 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

[0033] 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.

[0034] 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.

[0035] 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.

[0036] 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 of mRNA from prostate cells is reported in references 4, 5, 6 & 7.

B.3 - Indirect detection of RNA

[0037] 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 doublo-stranded, so detection techniques can be based on a sequence, on its complement, or on the double-stranded molecule).

C - POLYPEPTIDE EXPRESSION PRODUCT

[0038] 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.

[0039] 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].

[0040] 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.

[0041] 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 [BFRV-K(II)]; SEQ ID 102 [HERV-K10].

[0042] 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 ['ERVK66'].

[0043] An alignment of gag polypeptide sequences is shown in Figure 7.

[0044] 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.

[0045] Examples of prt nucleotide sequences are: SEQ ID 86 [HERV-K(108)]; SEQ ID 99 [HERV-K(II)]; SEQ ID 105 [HERV-K10].

[0046] Examples of prt polypeptide sequences are: SEQ ID 106 [HERV-K10]; SEQ ID 147 ['ERVK6'].

[0047] 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].

[0048] 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].

[0049] Examples of pol polypeptide sequences are: SEQ ID 94 [HERV-K(C7)]; SEQ ID 108 [HERV-K10]; SEQ ID 148 ['ERVK6'].

[0050] An alignment of pol polypeptide sequences is shown in Figure 8.

[0051] HML-2 env polypeptide is encoded by the fourth long ORF in a complete HML-2 genome. The translated polypeptide is proteolytically cleaved.

[0052] 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].

[0053] Examples of env polypeptide sequences are: SEQ ID 96 [HERV-K(C7)]; SEQ ID 108 [HERV-K10] ; SEQ ID 149 ['ERVK6'].

[0054] Alignments of env polynucleotide and polypeptide sequences are shown in Figures 6 and 9.

[0055] HML-2 cORP 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].

[0056] Examples of cORF nucleotide sequences are: SEQ ID 89 and SEQ ID 90 [HERV-K(108)]

[0057] Examples of cORF polypeptide sequences are SEQ ID 109.

C.1- Direct detection of HML-2 polypeptides

[0058] 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.*

[0059] In general, therefore, the invention provides a method for detecting the presence of and/or measuring a level of a polypeptide 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.

[0060] Polypeptides 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.

[0061] Another way for detecting polypeptides 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.

[0062] 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.

[0063] 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 polypeptide and 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

[0064] 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 which will typically be immobilized.

[0065] Antibodies against HERV-K polypeptides have been detected in humans [143].

[0066] 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.

[0067] 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).

[0068] 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 glyco-

sylation 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

[0069] Antibodies be polyclonal or monoclonal and may be produced by any suitable means (*e.g.* by recombinant expression).

[0070] Antibodies 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.*).

[0071] Antibodies may be attached to a solid support.

[0072] Antibodies 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).

[0073] Antigen-binding fragments of antibodies include Fv, scFv, Fc, Fab, $F(ab')_2$ *etc.*

[0074] 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 administation 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.

[0075] 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].

[0076] 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].

[0077] 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.

[0078] 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]

[0079] 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.

[0080] 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 effect of the corresponding polypeptide.

D - COMPARISON WITH CONTROL SAMPLES

D.1 - The control

[0081] 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.

[0082] 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.

[0083] 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.

[0084] 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.

[0085] 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.

[0086] Other suitable positive and negative controls will be apparent to the skilled person.

[0087] 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).

[0088] 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

[0089] 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

[0090] 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.

[0091] 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. The cancer is prostate cancer.

[0092] 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.

[0093] 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.

[0094] All of these techniques fall within the general meaning of "diagnosis" in the present invention.

G- THE HML-2 FAMILY OF HUMAN ENDOGENOUS RETROVIRUSES

[0095] 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.

[0096] 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 adj acent gene expression in a neutral or beneficial way.

[0097] 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].

[0098] 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.

[0099] 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].

[0100] 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.

[0101] The sequence of HERV-K(II), which locates to chromosome 3, has been disclosed [145].

[0102] 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.

[0103] 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)

[0104] 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)'.

[0105] 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.

[0106] 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-HMV-K(CH) genomic sequences

[0107] The disclosure 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

[0108] The disclosure 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.

[0109] 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 of *x* 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).

[0110] 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).

[0111] 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).

[0112] 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*.

[0113] Probes from more than one polynucleotide sequence of the disclosure 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).

[0114] 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).

[0115] 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).

[0116] 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

[0117] The disclosure 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 nucleotide in length.

[0118] The disclosure also provides an isolated polynucleotide having formula 5'-A-B-C-3', wherein: A is a nucleotide sequence consisting of α 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.

[0119] 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).

[0120] A and/or C may comprise a promoter sequence (or its complement).

H.14 - Homologous sequences

[0121] Also provided is a polynucleotide having at least *s*% identity to: (a) SEQ IDs 7-10; (b) a fragment ofx nucleotides of SEQ IDs 7-10; (c) SEQ IDs 11-13; (b) a fragment of *x* 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, 9S, 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.).*

[0122] These polynucleotides include naturally-occurring variants (*e.g.* degenerate variants, allelic variants, *etc.*), homologs, orthologs, and functional mutants.

[0123] 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.

[0124] The disclosure 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.

[0125] 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(H) hybridizable sequences

[0126] Disclosed is 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.

[0127] 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

[0128] Also disclosed is 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

[0129] Preferred polynucleotides 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 disclosure 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

[0130] General features of the polynucleotides described in this section H.1 are the same as those described in section B.4 above.

[0131] The isolated polynucleotides preferably comprise a polynucleotide having a HERV-K(CH) sequence.

[0132] A polynucleotide 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*.

[0133] Polynucleotides 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.

[0134] 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.

[0135] Polynucleotides 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.

[0136] 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

[0137] Disclosed is an isolated polypeptide: (a) encoded within a HEPV-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.

[0138] Deduced polypeptides encoded by the HERV-K(CH) polynucleotides 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.

[0139] The polypeptides encompassed by the present disclosure include those encoded by polynucleotides of the disclosure *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 disclosure 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.

[0140] 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.

[0141] 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.

[0142] 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.

H.2.2 - HERV-K(CH) fragments

[0143] Disclosed is 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.

[0144] 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.

[0145] The fragment may include an epitope *e.g.* an epitope of the amino acid sequence shown in SEQ IDs 56, 57 or 58.

[0146] 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 disclosure 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

[0147] 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.

[0148] 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.

[0149] 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

[0150] Also disclosed is 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).

[0151] 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

[0152] Also disclosed is 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.

[0153] These polypeptides include naturally-occurring variants (*e.g*. allelic variants, *etc*.), homologs, orthologs, and functional mutants.

[0154] Variants of the naturally-occurring polypeptides, wherein such variants are homologous or substantially similar to the naturally occurring polypeptide, 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, and the genetic code can be used to select appropriate codons to construct the corresponding variants.

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

[0155] Polypeptides, such as those shown in SEQ IDs 46-58, encoded by HERV-K(CH) polynucleotides are differentially expressed in prostate cancer cells. Such polypeptides are referred to herein as "polypeptides associated with prostate cancer" or "HERV-K(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.

H.2.6 - General features of polypeptides

[0156] General features of the polypeptides described in this section H.2 are the same as those described in section C.3 above.

[0157] The isolated polypeptides preferably comprise a polypeptide having a HERV-K(CH) sequence.

[0158] 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).

[0159] 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.

[0160] 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

[0161] The present disclosure also provides isolated antibodies or antigen binding fragments thereof, that bind to a polypeptide disclosed herein. The present disclosure also provides isolated antibodies or antigen binding fragments thereof, that bind to a polypeptide encoded by a polynucleotide disclosure herein. The present disclosure also provides isolated antibodies that bind to a polypeptide, 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, or antigen binding fragment thereof, and ii) collecting cells from said host expressing antibodies against the antigen or antigen binding fragment thereof. Also provided are isolated antibodies that bind to a polypeptide, or antigen binding fragment thereof, encoded by a polynucleotide disclosed herein made by the method comprising the following steps:

providing a cell line producing an antibody, wherein said antibody binds to a polypeptide, or antigen binding fragment thereof, encoded by a polynucleotide disclosed herein 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.

[0162] Antibodies, 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.

[0163] 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.

[0164] 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 know 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.

[0165] 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.

[0166] 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].

[0167] In other embodiments, 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.

[0168] 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].

[0169] 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 supernates assayed for antibody production by ELISA, RIA or bioassay.

[0170] The unique ability of antibodies to recognize and specifically bind to target polypeptides provides an approach for treating an overexpression of the polypeptide.

[0171] 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.6-HERV-K(CH)-based diagnostic methods

[0172] The invention provides methods for diagnosing the presence of prostate 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.

[0173] 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.

[0174] Also disclosed are compositions comprising a test cell sample and an isolated polynucleotide disclosed herein. The present invention further provides methods for detecting prostate 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 disclosed herein, 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 disclosure also provides methods for detecting prostate cancer associated with the presence of an antibody in a test cell sample, comprising the steps of: i) detecting a level of an antibody 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.

[0175] This disclosure also provides methods for detecting prostate cancer associated with elevated levels of HERV-K(CH) polynucleotides, 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 prostate cancer, and monitoring its progression.

[0176] The treatment regimen of a prostate 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.

[0177] The present invention provides methods of using the polynucleotides described herein for detecting prostate cancer cells, facilitating diagnosis of prostate 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 prostate cancer cell and/or detection of a polypeptide encoded by a polynucleotide that is differentially expressed in a prostate 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).

[0178] The detection methods can be provided as part of a kit. Thus, the disclosure 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 disclosure for detecting a polypeptide encoded by a polynucleotide that is differentially expressed in a prostate 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 disclosure 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.

[0179] Accordingly, the present disclosure 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.

[0180] 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.

[0181] 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.

[0182] 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.

[0183] 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.

[0184] 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.

[0185] 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.

[0186] 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. 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 known in the art and can be used in the assay methods of the invention. Specific hybridization can be determined by comparison to appropriate controls.

[0187] 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.

[0188] 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.

[0189] 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.

[0190] 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 (BEX)), 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.

[0191] The present invention further relates to methods of detecting/diagnosing a neoplastic or preneoplastic condition in a mammal (for example, a human).

[0192] 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.

[0193] 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.

[0194] 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.

[0195] 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 HERV-K(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 HERV-K(CH) polypeptide.

[0196] A kit can be used in the detection of a HERV-K(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

[0197] 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.

[0198] The term "about" in relation to a numerical value x means, for example, $x \pm 10\%$.

[0199] 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 characterized 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

[0200]

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

[0201] 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

[0202] 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.

[0203] 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.

[0204] 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 (Molecular 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 101 Geneclean® 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.

[0205] 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.

[0206] 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).

[0207] 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.

[0208] The remaining sequences were compared via BLAST [198] to GenBank and EST databases for gene identifi-

cation 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.

[0209] 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.

[0210] 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

[0211] 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.

[0212] 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.

[0213] 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.

[0214] 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).

[0215] 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% formamide, 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.

[0216] 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.

[0217] 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.

[0218] 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.

[0219] 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

[0220] 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 microarrays and then the full inserts of some of the 16 clones were sequenced (Table 6).

[0221] 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.

[0222] 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.

[0223] 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.

[0224] 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 HERV-K(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

[0225] Homology to HERV-K(II) gag region varied from 87% to 99%. Homology to HERV-K(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.

[0226] 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.

[0227] 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.

[0228] 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

[0229] HERV-K(CH) is a member of the HML-2 subgroup of the HERV-K family. HERV-K(H) and HERV-K(10) are also members of this sub-group.

[0230] 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.

[0231] 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

[0232] 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.

[0233] 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.

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

[0234] 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].

[0235] 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.

[0236] 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.

[0237] 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.

[0238] 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.

[0239] 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.

[0240] 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-K(CH) Gag and/or Pol related sequences.

[0241] To test blood samples for antibodies that bind specifically to recombinantly produced HERV-K(CH) antigens, the following procedure is employed. After the recombinant HERV-K(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 μl). 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.

[0242] 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.

[0243] 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.*

[0244] 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.

[0245] 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.

[0246] 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.

[0247] 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

[0248] 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.

[0249] 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

[0250] The materials listed in Table 7 were deposited with the American Type Culture Collection.

[0251] 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 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'seg) 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
po15'	035JN001.F06	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
po15'	035JN003.E06	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
po15'	035JN013.C11	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
po15'	035JN013.F03	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
pro13'	035JN003.G09	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
po13'	035JN010.A09	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
po13'	035JN015.F06	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
po13'	035JN020.B12	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
po13'	035JN020.D07	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
po13'	035JN022.G09	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
po13'	035JN015.H02	Prostate Cancer Tissue, Patient 101, Gleason Grade 3+3
po13'	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.B12	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.F06	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 Expression Ratio				Tu Expression Ratio mor/Normal mRNA												
HERVK ORF	Chiron Clone ID	>=2x	>=2-5x	>=5x '	<=halfx	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	7.3
qaq	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	9.5	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	2.2	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	035JN020.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
pol3orime	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	035JN018.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 0353N002.E02.
28	Sequence for clone 035JN023.F12.
29	Sequence for clone 035JN013.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	BERV-K(II) gag CDS
98	HERV-K(H) gag amino acid sequence
99	HERV-K(II) prt CDS
100	HERV-K(II) pol CDS
101	BERV-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 Label	Query Length	Target Description	Model Length	Score	P-value	Matches	Similarities	Percent Alignment	Percent Query	Query Start	Query End	Target Start	Target End	Open Gap Penalty	Extension Penalty
N4	7428	NT_022285S1.2 /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=8	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	7328	7328	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=186310 /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=455638 /end=8612270	102399	67086	5.9E-44	7112	7112	95	95	1	7428	164609	172033	-20	-5
N4	7428	NT_024249S1.2 /contig_orient=none /start=1 /end=187403 /chrom=11 Homo	102399	67086	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	66342	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	66610	2.6E-44	7068	7068	95	95	1	7428	144115	151250	-20	-5
N4	7428	NT_004656S5.3 /contig_orient=complement /start=999278 /end=1248561	102399	66624	2.1E-44	7072	7072	95	95	7428	1	23642	30777	-20	-5
N4	7428	NT_005795S3.3 /contig_orient=forward /start=405779 /end=655719 /chrom=3	102399	67968	6.1E-44	7040	7040	94	94	1	7428	122035	129165	-20	-5
N4	7428	NT_025140S3.3 /contig_orient=none /start=449919 /end=649939 /chrom=19	97519	68168	4.4E-44	7049	7049	94	94	7428	1	174979	182103	-20	-5
N4	7428	NT_009394S8.3 /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=965557 /chrom=1	85987	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	59453	5.7E-36	6795	6795	79	91	1	7428	141741	150230	-20	-5
N4	7428	NT_011512S23.3 /contig_orient=forward /start=5505084 /end=5765084	102399	64066	3E-41	6818	6818	92	91	7428	38	93292	100383	-20	-5
N4	7428	NT_019638S1.3 /contig_orient=none /start=1 /end=2500001 /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	91348	65850	2.6E-42	6734	6734	95	90	7024	1	140614	147637	-20	-5
N4	7428	NT_005532S1.3 /contig_orient=complement /start=1 /end=214350 /chrom=3	102399	62739	2.6E-40	6630	6630	91	89	7428	236	48116	55040	-20	-5
N4	7428	NT_022504S1.3 /contig_orient=forward /start=1 /end=271841 /chrom=3	102399	59001	1.3E-35	6420	6420	86	86	7428	1	176529	183705	-20	-5
N4	7428	NT_023367S2.3 /contig_orient=complement /start=250002 /end=465242	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	9.8E-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	21321	-20	-5
N4	7428	NT_019483S2.3 /contig_orient=forward /start=2500001 /end=5000001 /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	5636	-20	-5
N4	7428	NT_023323S1.2 /contig_orient=none /start=1 /end=103081 /chrom=5 Homo	102399	45124	4.5E-28	5600	5600	82	75	6704	1	1	6759	-20	-5

REFERENCES

[0252]

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. Immuno/. 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. 98:503-517 (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

[0253]

SEQ ID 1:

CTTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCACCCCTA
CA

SEQ ID 2:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTATGTACTAA
GAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCCACTCAG
GGTTAAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCC
CTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAACGTTTCTC
CCCATGTGATAGCCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGA
GGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAA
TGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACT

SEQ ID 3:

GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTAT
GCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCACATGTTTGTCTGCTGACCCTCTCCC
CACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCACAGATGATCAGTAAATACTAAGGGAACTCAGAGGCTGG
CGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTCTTTCTTTCTCTATA

SEQ ID 4:

GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTAT
GCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCACATGTTTGTCTGCTGACCCTCTCCC
CACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCACAGATGATCAGTAAATACTAAGGGAACTCAGAGGCTGG
CGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTCTTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAA
ATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCACCCCTACA

SEQ ID 5:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTATGTACTAA
GAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCCACTCAG
GGTTAAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCC
CTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAACGTTTCTC
CCCATGTGATAGCCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGA
GGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAA
TGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGC
AATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAA
AGACCTTTGTTCACATGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCA
CAGATGATCAGTAAATACTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTCTTTCT
TTCTCTATA

SEQ ID 6:

TCTGGTGCCCAACGTGGAGGCTTTTCTCTAGGGTGAAGGTACGCTCGAGCGTGGTCATTGAGGACAAGTCGACGAGAGATCCCGAGT
ACATCTACAGTCAGCCTTACG

SEQ ID 7:

GGGAAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATAAAGGAAG
GAGTTAAACAATATGGRTCCAACTCCCCTTATATAAGAACAKTATTAGATTCCATTGCYCATGGAAATAGACTTACTCCTTATGACT
GGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTGATGGRGTACARGAAACAGGTAC
GAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACAGGTCCAAATTGGAGCACCATTAACCAAC
AATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTT
TCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTA
CAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAA
AAGGAAAAGTTCCAGCAGGAGTTGATGTAATTACMGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATGC
TAATGGCTCAAGCAATGAGGGGGCTCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATCGGTC
ATCKGAAAAGGAGTTGCCCAGGCTTAAAYAARCAGAATATAATAAATCAAGCTATTACAGCAAAAAATAAAAAGCCATCTGGCCTGT
GTCCAAAATGTGGAAAAGCAAAACATTGGGCCAATCAATGTCATTCTAAATTTGATAAAGATGGGCAACCATTGTCTGGAAACAGGA
AGAGGGGCCAGCCTCAGGCCCCCCAACAAACTGGGGCATTCCCAGTTAAACTGTTTGTTCCTCAGGGTTTTCAAGGACAACAACCCC
TACAGAAAATACCACCACTTCAGGGAGTCAGCCAATTACAACAATCCAACAGCTGTCCCGCGCCACAGCAGGCAGCACCGCAGTAGA
TTTATGTTCCACCCAAATGGTCTTTTTACTCCCTGGAAAGCCCCCACAAAAGATTCCTAGAGGGGTATATGGCCCGCTGCCAGAAGG
GAGGGTAGGCCTTTGAGGGAGATCAAGTCTAAATTTGAAGGGAGTCCAAATTCATACTGGGGTAATTTACTCAGATTATAAAGGGGG
AATTCAGTTAGTGATCAGCTCCACTGTTCCCCGGAGTGCCAATCCAGGTGATAGAATTGCTCAATTACTGCTTTTGCCTTATGCA

SEQ ID 8:

ACAACAATGGCATGCAGAGATTACTATCCCAGCCTCCCTATACAGCCCCAGGAATCAAAAAATCATGACTAAAATGGGATAGCTCCC
TAAAAAGGGACTAGGAAAGAAAGAAGTCCCAATTGAGGCTGAAAAAAATYAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGGAGCG
GTCACTGTAGAGCCTCCAAAACCCATTCCATTAACTTGGGAAAAAAAMAACTGTATGGTAAATCAGCAGCCGCTTCCAAAACAAAAR
CTGGAGGCYTTACAYTTATTAGCAAAGAAACMATTAGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTRATTC
AGAAAAAATCCGGCAGATGGCGTATGCTAACTGAGCCATTAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTC
TCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGARAAATTTG
CTTTYACCACACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAAC
TATTTGTCAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGCTGCA
GAAACGAGAGACAAATTAATTGACCGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGGGRCTGACAATAACATCTGATAAGATT

CAARCCTCTACTCCTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCMCAAAAAAATAGAAATAAGAAAAGACAC
ATTAAAAGCATTAAATGAGTTTCAAAAGTTGCTAGGAGATACTAATTGGATTTGGAGATATTAATTGGATTTGGCCAACTCTAGGCA
TTCCTACTTATGCCATGTCAAATTTGTWCTCTTTCTTAAGAGGGGACTCGGAATTAAATAGTGAAAGAACGTTAACTCCAGAGGCAA
CTAAAGAAATTAAATTAATTGAAGAAAAAATTCGGTCAGCACAAGTAAATAGAATAGATCACTTGGCCCCACTCCAAATTTTGATTT
TTACTACTGCACATTCCCTAACAGGCATCATTGTTCAAAACACAGATCTTGTGGAGTGGTCCTTCCTTCCTCACAGTACAATTAAGA
CTTTTACATTGTACTTGGATCAAATGGCTACATTAATTGGTCAGGGAAGATTATGAATAATAACATTGTGTGGAAATGACCCAGATA
AAATCACTGTTCCTTTCAACAAGCAACAGGTTAGACAAGCCTTTATCAATTCTGGTGCATGGCAGATTGGTCTTGCCGATTTTGTGG
GAATTATTGACAATCGTTACCACA

SEQ ID 9:

ACAACAATGGCATGCAGAGATTACTATCCCAGCCTCCCTATACAGCCCCAGGAATCAAAAAATCATGACTAAAATGGGATAGCTCCC
TAAAAAGGGACTAGGAAAGAAAGAAGTCCCAATTGAGGCTGAAAAAAATYAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGGAGCG
GTCACTGTAGAGCCTCCAAAACCCATTCCATTAACTTGGGGGAAAAAAAAAAMAACTGTATGGTAAATCAGCAGCCGCTTCCAAAACA
AAARCTGGAGGCYTTACAYTTATTAGCAAAGAAACMATTAGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTR
ATTCAGAAAAAATCCGGCAGATGGCGTATGCTAACTGAGCCATTAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGC
CCTCTCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGARAAA
TTTGCTTTYACCACACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTT
CAACTATTTGTCAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGC
TGCAGAAACGAGAGACAAATTAATTGACCGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGGGRCTGACAATAACATCTGATAA
GATTCAARCCTCTACTCCTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCMCAAAAAAATAGAAATAAGAAAAG
ACACATTAAAAGCATTAAATGAGTTTCAAAAGTTGCTAGGAGATACTAATTGGATTTGGAGATATTAATTGGATTTGGCCAACTCTA
GGCATTCCTACTTATGCCATGTCAAATTTGTWCTCTTTCTTAAGAGGGGACTCGGAATTAAATAGTGAAAGAACGTTAACTCCAGAG
GCAACTAAAGAAATTAAATTAATTGAAGAAAAAATTCGGTCAGCACAAGTAAATAGAATAGATCACTTGGCCCCACTCCAAATTTTG
ATTTTTACTACTGCACATTCCCTAACAGGCATCATTGTTCAAAACACAGATCTTGTGGAGTGGTCCTTCCTTCCTCACAGTACAATT
AAGACTTTTACATTGTACTTGGATCAAATGGCTACATTAATTGGTCAGGGAAGATTATGAATAATAACATTGTGTGGAAATGACCCA
GATAAAATCACTGTTCCTTTCAACAAGCAACAGGTTAGACAAGCCTTTATCAATTCTGGTGCATGGCAGATTGGTCTTGCCGATTTT
GTGGGAATTATTGACAATCGTTACCACA

SEQ ID 10:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTYGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAGATTTTAA
TCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATCAAATACATTATGGA
TGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAWGAAATTTCCCATTTTATATTACTCATATTCGAGCACA
CACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATKGARGCACAAGAACTTCA
TGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATARATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCRC
CCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGYAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGT
CATGCACGTACCYTCATTTGGAAAATTGTCATTTGTCCAYGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGAC
AGGAGAAAGTACTTCCCATGTYAAGAGACATTTATTATYTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAARACAGACAATGG
GCCAGGTTACTGTAGTAAAGCAGTTCAARAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGG
ACAGGCCATAATTGAAAGAACTAATAGAACACTCAAAGCTCAATTGGTTAAACA

SEQ ID 11:

GGGAAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATAAAGGAAG
GAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACAGTATTAGATTCCATTGCTCATGGAAATAGACTTACTCCTTATGACT
GGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTGATGGAGTACAAGAACAGGTAC
GnAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACAGGTCCAAATTGGAGCACCATTAACCAA
CAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGGACCCAGGAACAGCT
TTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATT
ACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTA
AAAGGAAAAGTTCCAGCAGGAGTTGATGTAATTACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATG
CTAATGGCTCAAGCAATGAGGGGGCTCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATCGGT
CATCTGAAAAGGAGTTGCCCAGGCTTAAATAAACAGAATATAATAAATCAAGCTATTA

SEQ ID 12:

NCTGAAAAAAATNAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGGAGCGGTCACTGTAGAGCCTCCAAAACCCATTCCATTAACTT
GGGNAAAAAAANAACTGTATGGTAAATCAGCAGNCGCTTCCAAAACAAAANCTGGAGGCNTTACANTTATTAGCAAAGAAACCATTA
GAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTAATTCAGAAAAAATCCGGCAGATGGCGTATGCTAACTGAGC
CATTAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTAATTATAATTGATCTGA
AGGATTGCTTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGAAAAATTTGCTTTTACCACACCAGCCTAAATAATAAAGAACCAGC
CACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAACTATTTGTCAGCTCAAGCTCTGCAACCAGTTAGAGAC
AAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGACCGTTACACATTT
CTGCAGACAGAGGTTGCCAACGCGGGACTGACAATAACATCTGATAAGATTCAAACCTCTACTCCTTTCCGTTACTTGGGAATGCAG
GTAGAGGAAAGGAAAATTAAACCACAAAAAAATAGAAATAAGAAAAGACACATTAAAAGCATTAAATGAGTTTCAAAAGTTGCTAGG
AGATACTAATTGGATTTGGAGATATTAATTGGATTTGGCCAACTCTAGGCATTCCTACTTATGCCATGTCAAATTTGTNCTCTTTCT
TAAGAGGGGACTCGGAATTAAATAGTGAAAGAACGTTAACTCCAGAGGCAACTA

SEQ ID 13:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAGATTTTAA
TCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATCAAATACATTATGGA
TGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTATATTACTCATATTCGAGCACA
CACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATGGAAGCACAAGAACTTCA
TGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCAC
CCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGT
CATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGAC

AGGAGAAAGTACTTCCCATGTTAAGAGACATTTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGG
GCCAGGTTACTGTAGTAAAGCAGTTCAAAAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGG
ACAGGCCATAATTGAAAGAACTAATAGAACACTCAAAGCTCAATTGGTTA

SEQ ID 14:

GGGAAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATAAAGGAAG
GAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACAGTATTAGATTCCATTGCTCATGGAAATAGACTTACTCCTTATGACT
GGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTGATGGAGTACAGGAACAGGTAC
GAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACAGGTCCAAATTGGAGCACCATTAACCAAC
AATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTT
TCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTA
CAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAA
AAGGAAAAGTTCCAGCAGGAGTTGATGTAATTACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATGC
TAATGGCTCAAGCAATGAGGGGGCTCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATCGGTC
ATCTGAAAAGGAGTTGCCCAGGCTTAAATAAACAGAATATAATAAATCAAGCTATTACAGAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 15:

GGGAAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATAAAGGAAG
GAGTTAAACAATATGGGTCCAACTCCCCTTATATAAGAACATTATTAGATTCCATTGCTCATGGAAATAGACTTACTCCTTATGACT
GGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTGATGGAGTACAAGAACAGGTAC
GAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACAGGTCCAAATTGGAGCACCATTAACCAAC
AATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTT
TCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTA
CAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAA
AAGGAAAAGTTCCAGCAGGAGTTGATGTAATTACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATGC
TAATGGCTCAAGCAATGAGGGGGCTCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATCGGTC
ATCGGAAAAGGAGTTGCCCAGGCTTAAATAAACAGAATATAATAAATCAAGCTATTACAGCAAAAAATAAAAAGCCATCTGGCCTGT
GTCCAAAATGTGGAAAAGCAAAACATTGGGCCAATCAATGTCATTCTAAATTTGATAAAGATGGGCAACCATTGTCTGGAAACAGGA
AGAGGGGCCAGCCTCAGGCCCCCCAACAAACTGGGGCATTCCCAGTTAAACTGTTTGTTCCTCAGGGTTTTCAAGGACAACAACCCC
TACAGAAAATACCACCACTTCAGGGAGTCAGCCAATTACAACAATCCAACAGCTGTCCCGCGCCACAGCAGGCAGCACCGCAGTAGA
TTTATGTTCCACCCAAATGGTCTTTTTACTCCCTGGAAAGCCCCCACAAAAGATTCCTAGAGGGGTATATGGCCCGCTGCCAGAAGG
GAGGGTAGGCCTTTGAGGGAGATCAAGTCTAAATTTGAAGGGAGTCCAAATTCATACTGGGGTAATTTACTCAGATTATAAAGGGGG
AATTCAGTTAGTGATCAGCTCCACTGTTCCCCGGAGTGCCAATCCAGGTGATAGAATTGCTCAATTACTGCTTTTGCCTTATGCAAA
AAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 16:

AAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATAAAGGAAGGAG
TTAAACAATATGGATCCAACTCCCCTTATATAAGAACAGTATTAGATTCCATTGCCCATGGAAATAGACTTACTCCTTATGACTGGG
AAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTGATGGGGTACAAGAACAGGTACGAA
AAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACAGGTCCAAATTGGAGCACCATTAACCAACAA
TCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGGACCCAGGAACAGCTTTC
CCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCTATTACA
GATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCATTAAAA
GGAAAAGTTCCAGCAGGAGTTGATGTAATTACCGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCAATGCTA
ATGGCTCAAGCAATGAGGGGGCTCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATCGGTCAT
CTGAAAAGGAGTTGCCCAGGCTTAAACAAGCAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 17:

ACAACAATGGCATGCAGAGATTACTATCCCAGCCTCCCTATACAGCCCCAGGAATCAAAAAATCATGACTAAAATGGGATAGCTCCC
TAAAAAGGGACTAGGAAAGAAAGAAGTCCCAATTGAGGCTGAAAAAAATTAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGGAGCG
GTCACTGTAGAGCCTCCAAAACCCATTCCATTAACTTGGGAAAAAAAAAAACTGTATGGTAAATCAGCAGCCGCTTCCAAAACAAAAG
CTGGAGGCCTTACACTTATTAGCAAAGAAACCATTAGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTGATTC
AGAAAAAATCCGGCAGATGGCGTATGCTAACTGAGCCATTAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTC
TCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGAAAAATTTG
CTTTTACCACACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAAC
TATTTGTCAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGCTGCA
GAAACGAGAGACAAATTAATTGACCGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGGGGCTGACAATAACATCTGATAAGATT
CAAACCTCTACTCCTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCCCAAAAAAAAAAAAAAAAAAAAAAAAAA
AAA

SEQ ID 18:

CTGAAAAAAATCAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGGAGCGGTCACTGTAGAGCCTCCAAAACCCATTCCATTAACTTG
GGGGAAAAAAAAACAACTGTATGGTAAATCAGCAGCGCTTCCAAAACAAAAACTGGAGGCTTTACATTTATTAGCAAAGAAACAATT
AGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTAATTCAGAAAAAATCCGGCAGATGGCGTATAATGCCGTAA
TTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTT
ACCATTCCTCTGGCAAAACAGGATTTTGAGAAATTTGCTTTTACCACACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGT
GGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAACTATTTGTCAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTG
TTACATCGTTCACTATGTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGACCGTTACACATTTCTGCAGACAGAGGT
TGCCAACGCGGGACTGACAATAACATCTGATAAGATTCAAACCTCTACTCCTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAA
AATTAAACCACAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 19:

CATTAGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTAATTCAGAAAAAATCCGGCAGATGGCGTATGCTAAC
TGAGCCATTAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTAATTATAATTGA

TCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGAAAAATTTGCTTTTACCACACCAGCCTAAATAATAAAGAA
CCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAACTATTTGTCAGCTCAAGCTCTGCAACCAGTTA
GAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGACCGTTACA
CATTTCTGCAGACAGAGGTTGCCAACGCGGGACTGACAATAACATCTGATAAGATTCAAACCTCTACTCCTTTCCGTTACTTGGGAA
TGCAGGTAGAGGAAAGGAAAATTAAACCACAAAAAATAGAAATAAGAAAAGACACATTAAAAGCATTAAATGAGTTTCAAAAGTTGC
TAGGAGATACTAATTGGATTTGGAGATATTAATTGGATTTGGCCAACTCTAGGCATTCCTACTTATGCCATGTCAAATTTGTACTCT
TTCTTAAGAGGGGACTCGGAATTAAATAGTGAAAGAACGTTAACTCCAGAGGCAACTAAAGAAAAAAAAAAAAAAAAAAAAAAAAA
AA

SEQ ID 20:

ATCTTTACCCTGTATAAACATCTTTCTCTTCCCAGTATTTCTAAGCATGTGACAATGAATATGCAAAGGAAGCGCAGCAGTCCACCA
GGTGTGGGATATGTGTGGCACAATTCAAGACAATGATTAAACCTCCACTTGATGTTGCAAAAGAGATTTTGAAAAATTTGCTTTCAC
CACACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAACTATTTGT
CAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGCTGCAGAAACGA
GAGACAAATTAATTGACCGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGGGACTGACAATAACATCTGATAAGATTCAAGCCT
CTACTCCTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCACAAAAAAATAGAAATAAGAAAAGACACATTAAAA
GCATTAAATGAGTTTCAAAAGTTGCTAGGAGATACTAATTGGATTTGGAGATATTAATTGGATTTGGCCAACTCTAGGCATTCCTAC
TTATGCCATGTCAAATTTGTTCTCTTTCTTAAGAGGGGACTCGGAATTAAATAGTGAAAGAACGTTAACTCCAGAGGCAACTAAAGA
AATTAAATTAATTGAAGAAAAAATTCGGTCAGCACAAGTAAATAGAATAGATCACTTGGCCCCACTCCAAATTTTGATTTTTACTAC
TGCACATTCCCTAACAGGCATCATTGTTCAAAACACAGATCTTGTGGAGTGGTCCTTCCTTCCTCACAGTACAATTAAGACTTTTAC
ATTGTACTTGGATCAAATGGCTACATTAATTGGTCAGGGAAGATTATGAATAATAACATTGTGTGGAAATGACCCAGATAAAATCAC
TGTTCCTTTCAACAAGCAACAGGTTAGACAAGCCTTTATCAATTCTGGTGCATGGCAGATTGGTCTTGCCGATTTTGTGGGAATTAT
TGACAATCGTTACCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 21:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAGATTTTAA
TCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATCAAATACATTATGGA
TGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTATATTACTCATATTCGAGCACA
CACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATGGAAGCACAAGAACTTCA
TGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCAC
CCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGT
CATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGAC
AGGAGAAAGTACTTCCCATGTCAAGAGACATTTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAA

SEQ ID 22:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTCGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAGATTTTAA
TCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATCAAATACATTATGGA
TGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTATATTACTCATATTCGAGCACA
CACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATGGAAGCACAAGAACTTCA
TGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCGC
CCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGTAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGT
CATGCACGTACCCTCATTTGGAAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGAC
AGGAGAAAGTACTTCCCATGTTAAGAGACATTTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGG
GCCAGGTTACTGTAGTAAAGCAGTTCAAAAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGG
ACAGGCCATAATTGAAAGAACTAATAGAACACTCAAAGCTCAATTGGTTAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 23:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAGATTTTAA
TCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATCAAATACATTATGGA
TGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTATATTACTCATATTCGAGCACA
CACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATGGAGGCACAAGAACTTCA
TGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAGATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCAC
CCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGT
CATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGAC
AGGAGAAAGTACTTCCCATGTTAAGAGACATTTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGG
GCCAGGTTACTGTAGTAAAGCAGTTCAAAAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGG
ACAGGCCATAATTGAAAGAACTAATAGAACACTCAAAGCTCAATTGGTTAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 24:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAGATTTTAA
TCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATCAAATACATTATGGA
TGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAATGAAATTTCCCATTTTATATTACTCATATTCGAGCACA
CACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATGGAAGCACAAGAACTTCA
TGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCAC
CCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGT
CATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGAC
AGGAGAAAGTACTTCCCATGTTAAGAGACATTTATTATTTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGG
GCCAGGTTACTGTAGTAAAGCAGTTCAAGAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGG
ACAGGCCATAATTGAAAGAACTAATAGAACACTCAAAGCTCAATTGGTTAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 25:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAGATTTTAA
TCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATCAAATACATTATGGA

TGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTATATTACTCATATTCGAGCACA
CACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATTGAAGCACAAGAACTTCA
TGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCAC
CCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGT
CATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCCACGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGAC
AGGAGAAAGTACTTCCCATGTTAAGAGACATTTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAGACAGACAATGG
GCCAGGTTACTGTAGTAAAGCAGTTCAAAAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGG
ACAGGCCATAATTGAAAGAACTAATAGAACACTCAAAGCTCAATTGGTTAAGCAAAAAAAAAAAAAAAAAAAAAAAAACATGTCG
GCCGCCTCGGCC

SEQ ID 26:

CCAAAAGAATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTGTTAACAAGATTTTAA
TCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAATCAAATACATTATGGA
TGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTATATTACTCATATTCGAGCACA
CACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATGGAAGCACAAGAACTTCA
TGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCAC
CCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGT
CATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGAC
AGGAGAAAGTACTTCCCATGTTAAGAGACATTTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGG
GCCAGGTTACTGTAGTAAAGCAGTTCAAAAATTCTTAAATCAGTGGAAAATTACACATACAATAGGAATTCTCTATAATTCCCAAGG
ACAGGCCATAATTGAAAGAACTAATAGAACACTCAAAGCTCAATTGGTTAAACAAAAAGAAAAAAAAAAAAAAAAA

SEQ ID 27:

ACCGGCCTTACGGCCGGGGAAGAGNTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAA
AAGATATAAAGGAAGGAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACAGTATTAGATTCCATTGCTCATGGAAATAGAC
TTACTCCTTATGACTGGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTGATGGAG
TACAGGAACAGGTACGAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACAGGTCCAAATTGGA
GCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGG
ACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTG
CTCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGG
CCATAAAGCCATTAAAAGGAAAAGTTCCAGCAGGAGTTGATGTAATTACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTA
TGCNTAAGGCAATGCTAATGGCTCAAGCAATGAGGGGGCTCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTT

SEQ ID 28:

TTACGGCCTTACGGCCGGGGAAGNNTNTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTT
AAAAGATATAAAGGAAGGAGTTAAACAATATGGGTCCAACTCCCCTTATATAAGAACATTATTAGATTCCATTGCTCATGGAAATAG
ACTTACTCCTTATGACTGGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTGATGG
AGTACAAGAACAGGTACGAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACAGGTCCAAATTG
GAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGGCCTGGGGAAAAATTCA
GGACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGC
TGCTCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTC
GGCCATAAAGCCATTAAAAGGAAAAGTTCCAGCAGGAGTTGATGTAATTACAGAATATN

SEQ ID 29:

CGGCCTTACGGCCCGGGGAGANNTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAA
GATATAAAGGAAGGAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACAGTATTAGATTCCATTGCCCATGGAAATAGACTT
ACTCCTTATGACTGGGAAATTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTGATGGGGTA
CAAGAACAGGTACGAAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACAGGTCCAAATTGGAG
CACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGGA
CCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTGC
TCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGC
CATAAAGCCATTAAAAGGAAAAGTTCCAGCAGGAGTTGATGTAATTACCG

SEQ ID 30:

NCCGGCCTTACGGCCGGGNCAACAATGGCATGCAGAGNTTACTATCCCAGCCTCCCTATACAGCCCCAGGAATCAAAAAATCATGAC
TAAAATGGGATAGCTCCCTAAAAAGGGACTAGGAAAGAAAGAAGTCCCAATTGAGGCTGAAAAAAATTAAAAAAGAAAAGGAATAGG
GCATCCTTTTTAGGAGCGGTCACTGTAGAGCCTCCAAAACCCATTCCATTAACTTGGGAAAAAAAAAACTGTNTGGTAAATCAGCAG
CCGNTTCCAAAACAAAAGCTGGAGGCCTTACACTTATTANCAAAGAANCCATTANAAAAAGGACATTGAGCCTTCATTTTCGCCTTG
GAATTCTGTTTGTGATTCAAAAAAAATCCGGCANATGGCGTATGCTAACTGANCCATTAATGCCGTAATTCAACCCATGGGGGCTCT
CCCACCCCGGTTGCCCTNTCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTTTACCATTCCTCTGGCAAAAC
AGGATTTTGAAAAATTTGCTTTTACCACACCAGCCTAAATAATAAANAACCANCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGG
AATGCTTAATAGTTCAACTATTNGTCAGCTCAAGCTCTGCAACCAGTTAGAGACN

SEQ ID 31:

NCCTGGCCTTACGGCCGGGGCTGAAAAAAATCAAAAAAGAAAAGGAATAGGGCATCCTTTTTTAGGAGCGGTCACTGTAGAGCCTCCA
AAACCCATTCCATTAACTTGGGGGAAAAAAAAACAACTGTATGGTAAATCAGCAGCGCTTCCAAAACAAAAACTGGAGGCTTTACAT
TTATTAGCAAAGAAACAATTAGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTAATTCAGAAAAAATCCGGCA
GATGGCGTATAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTAATTATAATTG
ATCTGAAGGATTGCTTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGAGAAATTTGCTTTTACCACACCAGCCTAAATAATAAAGA
ACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAGTTCAACTATTTGTCAGCTCAAGCTCTGCAACCAGTT
AGAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGACCGTTAC
ACATTTCTGCAGACAGAGGTTGCCAACGCGGGACTGACAATAACATCTGATAAGATTCAAACCTCTACTCCTTTCCGTTACTTGGGA
ATGCAGGTAGAGGAAAGGAAAATTAAACCACAAAAAAAAAAAAAAN

SEQ ID 32:

NNNNNCNCGGGCATTAGAAAAAGGACATTGAGCCTTCATTTTCGCCTTGGAATTCTGTTTGTAATTCAGAAAAAATCCGGCAGATGG
CGTATGCTAACTGAGCCATTAATGCCGTAATTCAACCCATGGGGGCTCTCCCACCCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTA
ATTATAATTGATCTGAAGGATTGCTTTTTTTACCATTCCTCTGGCAAAACAGGATTTTGAAAAATTTGCTTTTACCACACCAGCCTAA
ATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCANGGAATGCTTAATAGTTCAACTATTTGTCAGCTCAAAGCTC
TGCACCCAGNTAGAGACAAGTTTCAGACTGGTTCATCGTCCTATGTGATATTTTGTGTGCTGCAGAACGAGAGACAAATTATTGGCC
GTTCACATTTTTGCAGACAGAGGTTGCCAACGCGGGACTGACAATAACATCTGATAAGATTAAACCTCTACTCCTTCCGTACTTGGG
AATGCAGGTGGAGGAAAGGAAAATTAACCCCCNNAAAATTGAATTANGAAAAGACCCNTTAAAGCCTTAAATGAGTTCAAAAAGTTG
CTAGGAGAAACTAATTGGATTTGGAGANATTAATTGGATTTGGCAACTNTAGGCATTCCTACTTATGCCN

SEQ ID 33:

TCCGGCCTTACGGCCGGGNTCTTTACCCTGTATAAACATCTTTCTCTTCCCAGTATTTCTAAGCATGTGACAATGAATATGCAAAGG
AAGCGCAGCAGTCCACCAGGTGTGGGATATGTGTGGCACAATTCAAGACAATGATTAAACCTCCACTTGATGTTGCAAAAGAGATTT
TGAAAAATTTGCTTTCACCACACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTT
AATAGTTCAACTATTTGTCAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTACATCGTTCACTATGTTGATATTT
TGTGTGCTGCAGAAACGAGAGACAAATTAATTGACCGTTACACATTTCTGCAGACAGAGGTTGCCAACGCGGGACTGACAATAACAT
CTGATAAGATTCAAGCCTCTACTCCTTTCCGTTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCACAAAAAAATAGAAATA
AGAAAAGACACATTAAAAGCATTAAATGAGTTTCAAAAGTTGCTAGGAGATACTAATTGGATTTGGAGATATTAATTGGATTTGGCC
AACTCTAGGCATTCCTACTTATGCCATGTCAAATTTGTTCTCTTTCT

SEQ ID 34:

TTNCGGCCTTACGGCCGGGCCAAGATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGT
GTTAACAAGATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAA
TCAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTATATTA
CTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATGG
AAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACATGGAAACAGACAAAAAATA
TTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGT
TATGGCAAATGGATGTCATTGCACGTACCTTCATTTGGAAAATTGTCATTTGTCCATGTGACAGNTGATACTTATTCACATTTCATA
TGGGCAACCTGCCAGACAGGAGAAAGTACTTNCCATGTCAAGAGACATTTATTATCTTGGTTTCCTGGNTGGGGAGNTCCCNNNNNN
NANNNNNNNNNAAAAAAAANANNCNNN

SEQ ID 35:

TTACGGCCTTACGGCCGGGCCAAAGATGAGTCATCAAAACTCAGTATCACTCGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAG
TGTTAACAAGATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTA
ATCAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTATATT
ACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATG
GAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACATGGAAACAGACAAAAAAT
ATTGTACAACATTGCGCCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGTAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTG
TTATGGCAAATGGATGTCATGCACGTACCCTCATTTGGAAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATA
TGGGCAACCTGCCAGACAGGAGAAAGTACTTCCCATGTTAAGAGACA

SEQ ID 36:

ATTTGCCTTACGGCCGGGCCAAAAGTATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACA
GTGTTAACAAGATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCT
AATCAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTATAT
TACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCAT
GGAGGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAGATTTGATATCACATGGAAACAGACAAAAAA
TATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGT
GTTATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCAT
ATGGGCAACCTGCCAGACAGGAGAAAGTACTTCCCATGTTAAGAGACATTTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAA
AGTTAAAACAGACAATGGGCCANGTTACTGTAGTAAAGCAGTTCAAAAATTCTTAAATCAGTGGAAAATTACACATN

SEQ ID 37:

CGGCCTTACGGCCGGGCCAAANATGAAGGGNNNAANGNCGGTTCCCAGGGACNNAGGCGCNTTNCATGGTTGCNGTNGTTACACCTG
TTAACAAGATTNTAATCAGTCTATTAACATTGTATCAAATTCTGCATATGTAGNACAGGCTACAAAGGATATTGAGAGAGCCCTAAT
CAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAATGAAATTTCCCATTTTATATTAC
TCATATTCGAGCACACACTAATTTACCAGGGCCTTTNACTAAAGCAAATGAACAAGCTGACTTGCTNGTATCATCTGCATTCATGGA
AGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACATGGAAACAGACAAAAAATAT
TGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTT
ATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCCATGTGACAGNTGATACTTATTCACATTTCATATG
GGCAACCTGCCAGACANGAGAAAGTNCTTCCCATGTTAAGAGACATTTATTATTTTGNTNTCCTGNCATTGGGAGTTCCANAAAAAG
TAAAACAGACANTGGGCCAGGTTAC

SEQ ID 38:

TACGGCCTTACGGCCGGGCCAAGATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCNTTACAGTG
TTAACAAGATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAAT
CAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTATATTAC
TCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATTGA
AGCACAAGAACTTCATGCCTTGACTCATGTAAATGCNATAGGATTAAAAAATAAATTTGATATCACCTGGAAACAGACAAAAAATAT
TGTACAACATTGCACCCNNNGTCAGATTCTACACCTGGCCNCTCNNGAGGCAAGAGTTAATCCCNCANGGCTATGTCCTNATGTGTT
ATGGCAAANGGATGTNATGCNCCNNCCTTCCTTTNGAAAANNNNNNTTTGTNCCCCNNACANNNGATACTTATTCACNNTTNNTATN
GGNNACCCCCCCCACNNGANAAANAACCTNCCCNNTNNANANAAANTNNTTATTTTTNTTTTN

SEQ ID 39:

NCCGGCCTTACGGCCGGGCCAAGATGAGTCATCAAAACTCAGTATCACTTGACTCAAAGAGCAGAGTTGGTTGCCGTCATTACAGTG
TTAACAAGATTTTAATCAGTCTATTAACATTGTATCAGATTCTGCATATGTAGTACAGGCTACAAAGGATATTGAGAGAGCCCTAAT
CAAATACATTATGGATGATCAGTTAAACCCGCTGTTTAATTTGTTACAACAAAATGTAAGAAAAAGAAATTTCCCATTTTATATTAC
TCATATTCGAGCACACACTAATTTACCAGGGCCTTTAACTAAAGCAAATGAACAAGCTGACTTGCTAGTATCATCTGCATTCATGGA
AGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAATTTGATATCACATGGAAACAGACAAAAAATAT
TGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCAGGAGGCAAGAGTTAATCCCAGAGGTCTATGTCCTAATGTGTT
ATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCCATGTGACAGTTGATACTTATTCACATTTCATATG
GGCAACCTGCCAGACAGGAGAAAGTACTTCCCATGTTAAGAGACATTTATTATCTTGTTTTCCTGTCATGGGAGTTCCAGAAAAAGT
TAAAACAGACAATGGGCCAGGTTACTGGAGTAAAGCAGTTCAAAAATTCTTAAATCAGTGG

SEQ ID 40:

AAGGCAGTCAAGCAGGAGTTAAACAATATGGACCTAACTCTCCTTATATTAGAATATTATTAAATTCCATTGCTCATGGAAATAGAC
TTATTTCTTATGATTGGGAAATTCTGGCTATATCTTCCCTTTCACCCTCTCAGTATCTCCAGTTTAAAACCTGGTGGATTGATGGGG
TACAAGAACAGGTACGAAAAAATCAGGCTACTAATCCTGTTGCTTATATAGATGAAGACCAATTGCTAGGAAGAGGTCCAAACTGGG
ACACTATTAACCAACAATCAGTAATGAAAATGAGGCTATTGAACAACTATAAGGGCTATTTGCCTCAGGGCCTGGGAAAACATTCAG
GACCCAGGAACCTCATGCCCTTCTTTTAGTTCAATCAGACAAGGCTCTAAAGAGCCATATCCAGACTTTGTGGCAAGGTTGCAAGAT
GCAGCTCAAAAATCCATTGCAGGTAACGCCCGAAAAGTTATTGTAGAAATAATGGCTTATCAAAACGCAAATTCAGAGTGTCAATCA
GCCATAAAGCCATTAAGAGGAAATGTTTCAGCAGGAGTTGATGTAATTACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCT
ATGCATAAGGCAATGCCATTGGCTCAAGCAATTACAGGGGTTGCTATAGGAGGACAAGTTAAAACATTTGGGGGAAAATGTTATAAT
TGTGGTCAAATCGGTCATCTAAAAAAGAATTGCCCGAGCTTAAATTACCCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAA

SEQ ID 41:

NCCGGCCTTACGGCCGGGAAAGGCAGTCAAGCAGGAGTTAAACAATATGGACCTAACTCTCCTTATATTAGAATATTATTAAATTCC
ATTGCTCATGGAAATAGACTTATTTCTTATGATTGGGAAATTCTGGCTATATCTTCCCTTTCACCCTCTCAGTATCTCCAGTTTAAA
ACCTGGTGGATTGATGGGGTACAAGAACAGGTACCGAAAAAATCAGGCTACTAATCCTGTTGCTTATATAGATGAAGACCAATTGCT
AGGAAGAGGTCCAAACTGGGACACTATTAACCAACAATCAGTAATGAAAATGAGGCTATTGAACAACTATAAGGGCTATTTGCCTCA
GGGCCTGGGAAAACATTCAGGACCCAGGAACCTCATGCCCTTCTTTTAGTTCAATCAGACAAGGCTCTAAAGAGCCATATCCAGACT
TTGTGGCAAGGTTGCAAGATGCAGCTCAAAAATCCATTGCAGGTAACGCCCGAAAAGTTATTGTAGAAATAATGGCTTATCAAAACG
CAAATTCAGAGTGTCAATCAGCCATAAAGCCATTAAGAGGAAATGTTTCAGCAGGAGTTGATGTAATTACAGAATATGTGAAGGCTT
GTGATGGGATTGGAGGAGCTATGCATAAGGCAATGCCATTGGCTCAAGCAATTACAGGGGT

SEQ ID 42:

AAAGGCAGTCAAGCAGGAGTTAAACAATATGGACCTAACTCTCCTTATATGAGAACATTATTAAATTCCATTGCTCATGGAAATAGA
CTTATTTCTTATGATTGGGAAATTCTGGCTAAATCTTCCCTTTCACCCTCTCAGTATCTCCAGTTTAAAACCTGGTGGATTGATGGG
GTACAAGAACAGGTACGAAAAAATCAGGCTACTAATCCTGTTGCTTATATAGATGAAGACCAATTGCTAGGAAGAGGTCCAAACTGG
GACACTATTAACCAACAATCAGTAATGAAAATGAGGCTATTGAACAACTATAAGGGCTATTTGCCTCAGGGGCCTGGGAAAACATTC
AGGACCCAGGGAACCTCATGCCCTTCTTTTAGGTTCAATCAGACAAGGT

SEQ ID 43:

GCTGACTTGCTAGTATCATCTGCATTCATTGAAGCACAAGAACTTCATGCCTTGACTCATGTAAATGCAATAGGATTAAAAAATAAA
TTTGATATCACATGGAAACAGACAAAAAATATTGTACAACATTGCACCCAGTGTCAGATTCTACACCTGGCCACTCAGGAAGCAAGA
GTTAATCCCAGAGGTCTATGTCCTAATGTGTTATGGCAAATGGATGTCATGCACGTACCTTCATTTGGAAAATTGTCATTTGTCCAT
GTGACAGTTGATACTTATTCACATTTCATATGGGCAACCTGCCAGACAGGAGAAAGTCTTCCCATGTTAAAAGACATTTATTATCTT
GTTTTCCTGTCATGGGAGTTCCAGAAAAAGTTAAAACAGACAATGGGCCAGGTTCTGTAGTAAAG

SEQ ID 44:

GCCAAGGTGGGAGGATTGCTTGAGCACAGGAGTTTGAGGCTGAAGTGAGCTATGATCGCACCACTGCAATCAATCAATCAATAAACT
TCAGTCAACCCTGCCAGGAGCTATGGAACAATTATTGTTTGTTGGAGTGTTCTGTGTTGGGCTAAATGTGAAGCCTCTTTATACTTC
TACCTTACTCAGTCACCATATGGGGGCTGCCCCAGAGAGGTCATGACCTCAAGTGAGGAAGTACTCAGCAGCTGAGCCAGGCCCTAC
TGATAGCTGGAGGATGCTGCTGCCCATGCTGCCCACTGTGAGGCAGCAAGCCCTTGCTTGAAGGGGGATCTGGATAGTATGTTTCTG
TGTCTACCACCCCTAGAAATGGTGCCTAGAGTGAGTCATCACAAAAAGAATCAGGATAGCTTGGTGTAGTGGCAGGTGCCTATAATC
CCAGCTACTCAGGAGACTGTGGCAGGAGAATGACTTAAACCAGGGAGTTGGAGGTTGCAGTGAGGTGAGGTCACACAACTGCACTCC
AGACTGGGTGACAGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAGAAAAGAAAAGAAAAAAGAAAAAGAATCAGGAAATACTAA
TATTTAAAGGATAGGTGAATGGAGGAAAATAATCAATTGAAGGAGGCTGAGCAGATGAGGTCAAAGAAGATAGAGATCCATAACAGT
AACCTCATAGAAGCTTATGGAAGCATTTTGACAGTGCTAAAAGCCACATAAAGTTCAAGTAAGACAGTTTCAGAAATGTATAAACAT
GAATGCCTTTGCAGTGACTTAAGTGTGATTCTGGTGTTTCCTTCTAAAAATACTGCCTTCTCAGGTGTGGGAAGGATTCTATCTTTT
TAGGCTTTACCACCATAGTTCTCTGCAGGCTTGCAATCCTGAATCAGGCTTGACTTCAGAAAGTGCTTTAAAAGGGAGGCTGGGCGC
GGTGGCTCATGCCTGTAATCCCAGCACTCTGAGAGGCTGAGGTTGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGT
AGAAAGAAGTAGACATAGGAGACTCCATTTTGTTCTGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTAC
CCCCAACCCCGTGCTCTCTGAAACAGGTGCTGTGTCAAACTCAGGGTTAAATGGATTAAGGGTTGTGCAAGATGTGCTTTGTTAAAC
AAATGCTTGAAGGCAGCATGGTCCTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAACACTGCGGAAGGCCGC
AGAGACCTCTGCCTAGGAAAGCAAGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGA
CCTGACCGTCCCCCAGCCTGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGTGTAAGAGGAAGGCATGCCTCTTGCAGTTGAGA
CAAGAGGAAGGCATCTGTCTCCTGCCCGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGATTGTACGTTCCATCTACTG
AGATAGGAAGAAAACGCCTTAGGGCTGGAGGTGTGGGACAAGCCGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGTGTA
TGCATATCTAAAAGCACAGCACTTGATTCTTTACCTTGTCTGTGATGCAAAGACCTTTGTTCACGTGTTTGTCTGCTGACCCTCTCC
CCACTATTGTCTTGTGACCATGACACATCCCCCTCTCAGAGAAACACCCACGAATGATCAATAAATACTAAGGGAACTCAGAGACGG
CGCGGATCCTCCATATGCTGAACGCTGGTTCCCTGGGTCCCCTTATTTCTTTCTCTATACTTTGTGTCTTTTTCTTTTCCAAGTCTC
TCGTTCCACCTTACGAGAAACACCCACAGGTGTGGAGGGGCAACCCACCCCTTCATCTGGTGCCCAACGTGGAGGCTTTTCTCTAGG
GTGAAGGTACGCTCGAGCGTGGTCATTGAGGACAAGTTGACGAGAGATCCCGAGTACATCTACAGTCAGCCTTGCGGTAAGTTTGTG
CGCTCGGAAGAAGCTAGGGTGATAATGGGGCAAACTAAAAGTAAAACTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATT
CTTTTAAAAAGAGGGGGAGTTAGAGTATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCA
GAACAAGGAACTTTAGATCTAAAAGATTGGAAAAGAATTGGCGAGGAACTAAAACAAGCAGGTAGAAAGGGTAATATCATTCCACTT

ACAGTATGGAATGATTGGGCCATTATTAAAGCAGCTTTAGAACCATTTCAAACAAAAGAAGATAGCGTTTCAGTTTCTGATGCCCCT
GGAAGCTGTGTAATAGATTGTAATGAAAAGACAGGGAGAAAATCCCAGAAAGAAACAGAAAGTTTACATTGCGAATATGTAACAGAG
CCAGTAATGGCTCAGTCAACGCAAAATGTTGACTATAATCAATTACAGGGGGTGATATATCCTGAAACGTTAAAATTAGAAGGAAAA
GGTCCAGAATTAGTGGGGCCATCAGAGTCTAAACCACGAGGGCCAAGTCCTCTTCCAGCAGGTCAGGTGCCCGTAACATTACAACCT
CAAACGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTAGCTTATCAATACTGGCCGCCGGCTGAACTTCAGTATCTGCCACCCCCA
GAAAGTCAGTATGGATATCCAGGAATGCCCCCAGCACTACAGGGCAGGGCGCCATATCCTCAGCCGCCCACTGTGAGACTTAATCCT
ACAGCATCACGTAGTGGACAAGGTGGTACACTGCACGCAGTCATTGATGAAGCCAGAAAACAGGGAGATCTTGAGGCATGGCGGTTC
CTGGTAATTTTACAACTGGTACAGGCCGGGGAAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACC
ATGAAAATGTTAAAAGATATAAAGGAAGGAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACATTATTAGATTCCATTGCT
CATGGAAATAGACTTACTCCTTATGACTGGGAAAGTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGG
TGGATTGATGGAGTACAAGAACAGGTACGAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACA
GGTCCAAATTGGAGCACCATTAACCAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGGCCTGG
GGAAAAATTCAGGACCCAGGAACAGCTTTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGACTTTGTGGCAAGA
TTACAAGATGCTGCTCAAAAGTCTATTACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCA
GAATGTCAGTCGGCCATAAAGCCATTAAAAGGAAAAGTTCCAGCAGGAGTTGATGTAATTACAGAATATGTGAAGGCTTGTGATGGG
ATTGGAGGAGCTATGCATAAGGCAATGCTAATGGCTCAAGCAATGAGGGGGCTCACTCTAGGAGGACAAGTTAGAACATTTGGGAAA
AAATGTTATAATTGTGGTCAAATCGGTCATCTGAAAAGGAGTTGCCCAGTCTTAAATAAACAGAATATAATAAATCAAGCTATTACA
GCAAAAAATAAAAAGCCATCTGGCCTGTGTCCAAAATGTGGAAAAGGAAAACATTGGGCCAATCAATGTCATTCTAAATTTGATAAA
GATGGGCAACCATTGTCGGGAAACAGGAAGAGGGGCCAGCCTCAGGCCCCCCAACAAACTGGGGCATTCCCAGTTCAACTGTTTGTT
CCTCAGGGTTTTCAAGGACAACAACCCCTACAGAAAATACCACCACTTCAGGGAGTCAGCCAATTACAACAATCCAACAGCTGTCCC
GCGCCACAGCAGGCAGCGCCACAGTAGATTTATGTTCCACCCAAATGGTCTCTTTACTCCCTGGAGAGCCCCCACAAAAGATTCCTA
GAGGGGTATATGGCCCGCTGCCAGAAGGGAGGGTAGGCCTTATTTTAGGGAGATCAAGTCTAAATTTGAAGGGAGTCCAAATTCATA
CTGGGGTAATTTATTCAGATTATAAAGGGGGAATTCAGTTAGTGATCAGCTCCACTGTTCCCTGGAGTGCCAATCCAGGTGATAGAA
TTGCTCAATTACTGCTTTTGCCTTATGTTAAAATTGGGGAAAACAAAACGGAAAGAACAGGAGGGTTTGGAAGTACCAACCCTGCAG
GAAAAGCCACTTATTGGGCTAATCAGGTCTCAGAGGATAGACCCGTGTGTACAGTCACTATTCAGGGAAAGAGTTTGAAGGATTAGT
GGATACCCAGGCTGATGTTTCTATCATCGGCATAGGCACCGCCTCAGAAGTGTATCAAAGTGCCATGATTTTACATTGTCTAGGATC
TGATAATCAAGAAAGTACGGTTCAGCCTATGATCACTTCTATTCCAATCAATTTATGGGGCCGAGACTTGTTACAACAATGGCATGC
AGAGATTACTATCCCAGCCTCCCTATACAGCCCCAGGAATCAAAAAATCATGACTAAAATGGGATAGCTCCCTAAAAAGGGACTAGG
AAAGAATGAAGATGGCATTAAAGTCCCAACTGAGGCTGAAAAAAATCAAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGAAGCGGT
CACTGTAGAGCCTCCAAAACCCATTCCATTAATTTGGGGGGAAAAAAAAAACTGTATGGTAAATCAGTAGCCGCTTCCAAAACAAAA
ACTGGAGGCTTTACACTTATTAGCAAAGAAACAGTTAGAAAAAGGACATATTGAGCCTTCATTTTCGCCTTGGAATTCTCCTGTTTG
TAATTCAGAAAAAATCCGGCAGATGGCGTATGCTAACTGACTTAAGAGCCATTAATGCCATAATTCAACCCATGGGGGCTCTCCCAT
CCCGGTTGCCCTCTCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAAGAGGAT
TTTGAAAAATTTGCTTTTACTATACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGC
TTAATAATTCAACTATTTGTCAGACTTTCATAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTATATCGTTCATT
ATGTTGATATTTTGTGTGCTGCAGAAACGAGAGACAAATTAATTGACCGTTACACATTTCTCAGACAGAGGTTGCCAACGCGGGACT
GACAATAGCATCTGATAAGATTCAAACCTCTCCTCCTTTCCATTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCACAAAA
AATAGAAATAAGAAAAGACACATTAAAAACATTAAATGAGTTTCAAAAGTTGGTAGGAGATACTAATTGGATTCGGAGATATTAATT
GGATTTGGCCAACTCTAGGCATTCCTACTTATGCCATGTCAATTTTGTTCTCTTTCTTAAGAGGGGACTTGGAATTAAATAGTGAAA
GAATGTTACCTCCAGAGGCAACTAAAGAAATTAAATTAATTGAAGAAAAAAATTCGGTCAGCACAAGTAAATAGGATCACTTGGCCC
CACTCCAAATTTTGATTTTTGGTACTGCACATTCTCTAACAGCCATCATTGTTCAAAACACAGATCTTGTGGATTGGTCCTTCCTTC
CTCATAGTACAATTAAGACTTTTACATTGTACTTGGATCAAATGGCTACATTAATTGGTCAGGGAAGATTACGAATAATAACATTGT
GTGGAAATGACCCAGATAAAATCACTGTTCCTTTCAACAAGCAACAAGTTAGACAAGCCTTTATCAGTTCTGGTGCATGGCAGATTG
GTCTTGCTAATTTTCTGGGAATTATTGATAATCATTACCCAAAAACAAAAATCTTCCAGTTCTTAAAATTGACTACTTGGATTCTAC
CTAAAATTACCAGACGTGAACCTTTAGAAAATGCTCTAACAGTATTTACTGATGGTTCCAGCAATGGAAAAGCGGCTTACACAGGGC
CGAAAGAACGAGTAATCAAAACTCCGTATCAATCAGCTCAAAGAGCAGAGTTGGTTGCAGTCATTACAGTGTTACAAGATTTTGACC
AACCTATCAATATTATATCAGATTCTGCATATGTAGTACAGGCTACAAGGGATGTTGAGACAGCTCTAATTAAATATAGCACGGACG
ATCATTTAAACCAGCTATTCAATTTATTACAACAAACTGTAAGAAAAAGAAATTTCCCATTTTATATTACTCATATTCGAGCACACA
CTAATTTACCAGGGCCTTTGACTAAAGCAAATGAACAAGCTGACTTACTGGTATCATCTGCATTCATAAAAGCACAAGAACTTCTTG
CTTTGACTCATGTAAATGCAGCAGGATTAAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACATTGCACCC
AGTGTCAAGTCTTACACCTGTCCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCGTTATGGCAAATGGATGGCA
CGCATGTTCCTTCATTTGGAAGATTATCATATGTTCATGTAACAGTTGATACTTATTCACATTTCATATGGGCAACTTGCCAAACAG
GAGAAAGTACTTCCCATGTTAAAAAACATTTATTATCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCAAAACTGACAATGGAC
CAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAACAGGAATTCCTTATAATTCCCAAGGAC
AGGCCATAGTTGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGAAGGGGGAGACAGTAAGGAGTGTACCACTC
CTCAGATGCAACTTAATCTAGCACTCTATACTTTAAATTTTTTAAACATTTATAGAAATCAGACTACTACTTCTGCAAAACAACATC
TTACTGGTAAAAAGCACAGCCCACATGAAGGAAAACTAATTTGGTGGAAAGATAATAAAAATAAGACATGGGAAATAGGGAAGGTGA
TAACGTGGGGGAGAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAGCTTCCTGTTTGGATACCCACTAGACATTTGAAGTTCTACA
ATGAACCCATCGGAGATGCAAAGAAAAGGGCCTCCACAGAGATGGTAACCCCAGTCACATGGATGGATAATCCTATAGAAGTATATG
TTAATGATAGTGTATGGGTACCTGGCCCCACAGATGATCGCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGATAAATATTTCCA
TTGTGTATCGTTATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAATGCCTGCAGTCCAAAATTGGTTGGTAGAAGTACCTA
CTGTCAGTCCTAACAGTAGATTCACTTATCACATGGTAAGCGGGATGTCACTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTT
ATCAAAGATCATTAAAATTTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAGTTTTAGTTT
GGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTCGGAACTATTATAGATTGGGCACCTCGAGGTCAATTCT
ACCACAATTGCTCAGGACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTCTAGACA
AACATAAGCATAAAAAATTACAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATCTCTACCCCAAGACCAGAAATAATAAGTC
CTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTTGGCCTGACACCACATTAGAATTTGGTCTGGAAATCAAACTTTAGAAAC
AAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCCAGTCTAACGGTTCCTTTACAAAGTTGCGTAAAGCCCTCTTATATGCT
AGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAAACTATAACCTGTGAAAATTGTAGATTGTTTACTTGCATTGATTCAACTTT
TAATTGGCGGCACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCTCTGTGTCCGTGGACTGACCGTGGGAGGCCTCGCC

ATCCATCCATATTTTGACTGAAGTATTAAAAGACATTTTAAATAGATCCAAAAGATTCATTTTTACCTTAATTGCAGTGATTATGGG
ATTAATTGCAGTCACAGCTACGGCTGCTGTGGCAGGAGTTGCATTGCACTCTTCTGTTCAGTCGGTAAACTTTGTTAATGATTGGCA
AAAGAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATCAAAAATTGGCAAATCAAATTAATGATCTTAGACAAACTGTCAT
TTGGATGGGAGACAGACTCATGAGCTTAGAACATTGTTTCCAGTTACAGTGTGACTGGAATACGTCAGATTTTTGTATTACACCCCA
AATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAA
ATTAAAATAACAAATTTTCGAAGCATCAAAAGCCCATTTAAATTTGATGCCAGGAACTGAGGCAATTGCAGGAGTTGCTGATGGCCT
CGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATCGGAAGTACTATGATTATAAATCTCATATTAATCCTTGTGTGCCTGTTTTG
TCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCATGATGACGATGGCGGTTTTGTC
GAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTC
CATTTTGTTCTGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACA
GGTGCTGTGTCAAACTCAGGGTTAAATGGATTAAGGGTTGTGCAAGATGTGCTTTGTTAAACAAATGCTTGAAGGCAGCATGCTCCT
TAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAG
GTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCTCATGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACC
CGTAAAGGGTCTGTGCTGAGGAGGATTAGTATAAGAGGAAGGCATTCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGC
CCGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTACGTTCCATCTACTGAGATAGGAAGAAAACGCCTTAGGGCTGGA
GGTGGGACATGCAGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTGATTCTT
TACCTTGTCTATGATGCAAAGACCTTTGTTCACCTGTTTGTCTGCTGACCCTCTCCCCACTATTGTCTTGTGACCATGACACATCCC
CCTCTCAGAGAAACACCCACGAATGATCAATAAATACTAAGGGAACTCAGAGACGGCGCGGATCCTCCATATGCTGAACGCTGGTTC
CCTGGGTCCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAGTCTCTCATTCCACCTTAAGAGAAACACTC
ACAGGTGTGGAGGGGCAACCCATCCCTTCAGAGGTGGGTGGATCACCTGAGGTCAGGAGTTCAAGACAAGCCTGGCCAACATGGTGA
AACCCCATCTCTACTAAAAATACAAAATTAGCCAGGTGTGGTGGCAGGTGTCTGTAGTCCCAGCTACTTGGGAGGCTGACGAGAATC
GCTTGAACCTGGGAGGGGGAGGTTTCAGTGAGCCGAGATTGCACCACTGCACTCCAGCCTGGGGGACAGAGTGAAACTCTGTCTCAA
AAAAACAACAAAAAACCCCACCTATAGACAGGACTAGCTACATAAATAACTTGCAGGGCTCAGTGTAAAATGAAAGTGTGAGGTCCC
TTTTTCAAAGACGTAGAAGGCCGGGTGCGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCTGAGGCAGGCAGGTTATGAGGT
CAGGAGTTCGAGACAGCCTGACCAATATGGTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTAGCGGGCGCCT
GTAGTCCCAGCTACTCAGGAGGCTGAGGCAGAAGAATTACTTGAACCCAGGAGACGGAGGTTGCAGTGAGCTGAGATCGTGCCACTG
CACTCTCCAGCCTCCTCGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAACAGAAAAAGGTGCTATTAAAGATACCAAA
ATATAAGGCACTTTCCTTTATTCTGCAATCTGTCTCTCCACTTTTCATAGTATTTTTTCATTTGTTATTTAACATCATGTTTTGTCA
GGTGAGGACATTTACTCAGCCAGTGCAGCACTCACTGGTATCCAGGGGCCATAGGTGATTTGACGCACCCACATGGCCCACCAGCTG
TTGAGTTCCACCTCCAGCCAGCCACTGGACCAACATGCAGTGCCCTGGCTGGGGGCAGGAAAGTCTAACAAACCATTTCATTCCACT
GTCCTCCTGGCCAAACCCACAGAGGACAGGTAAACCCCCTTGTATGTGTTTTGTACTTGGATCTGGGGTGGGC

SEQ ID 45:

TGTGGGGAAAAGCAAGAGAGATCAAATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTATGTGCTAA
GAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACGTGTGCTGTGTCAACTCAG
GGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCC
CTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTC
CCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCTGTAAAGGGTCTGTGCTGA
GGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAA
TGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGC
AATACTGCTTTGTAAAGCATTGAGATGTTTATGTGTATGCATATCCAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCCA
AGACCTTTGTTCACGTGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACACCCA
CAGATGATCAATAAATACTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTATTTCT
TTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAAC
CCACCCCTACATCTGGTGCCCAACGTGGAGGCTTTTCTCTAGGGTGAAGGTACGCTCGAGCGTAATCATTGAGGACAAGTCGACGAG
AGATCCCGAGTACATCTACAGTCAGCCTTACGGTAAGCTTGCGCGCTCGGAAGAAGCTAGGGTGATAATGGGGCAAACTAAAAGTAA
AATTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGGAGTTAAAGTATCTACAAAAAATCTAAT
CAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAACTTCAGATCTAAAAGATTGGAAAAGAATTGGTAA
GGAACTAAAACAAGCAGGTAGGAAGGGTAATATCATTCCACTTACAGTATGGAATGATTGGGCCATTATTAAAGCAGCTTTAGAACC
ATTTCAAACAGAAGAAGATAGCATTTCAGTTTCTGATGCCCCTGGAAGCTGTTTAATAGATTGTAATGAAAACACAAGGAAAAAATC
CCAGAAAGAAACCGAAAGTTTACATTGCGAATATGTAGCAGAGCCGGTAATGGCTCAGTCAACGCAAAATGTTGACTATAATCAATT
ACAGGAGGTGATATATCCTGAAACGTTAAAATTAGAAGGAAAAGGTCCAGAATTAATGGGGCCATCAGAGTCTAAACCACGAGGCAC
AAGTCCTCTTCCAGCAGGTCAGGTGCTCGTAAGATTACAACCTCAAAAGCAGGTTAAAGAAAATAAGACCCAACCGCAAGTAGCCTA
TCAATACTGCCGCTGGCTGAACTTCAGTATCGGCCACCCCCAGAAAGTCAGTATGGATATCCAGGAATGCCCCCAGCACCACAGGGC
AGGGCGCCATACCATCAGCCGCCCACTAGGAGACTTAATCCTATGGCACCACCTAGTAGACAGGGTAGTGAATTACATGAAATTATT
GATAAATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAACCGATGCCACCTGGAGAAGGAGCCCAAGAG
GGAGAGCCTCCCACAGTTGAGGCCAGATACAAGTCTTTTTCGATAAAAATGCTAAAAGATATGAAAGAGGGAGTAAAACAGTATGGA
CCCAACTCCCCTTATATGAGGACATTATTAGATTCCATTGCTTATGGACATAGACTCATTCCTTATGATTGGGAGATTCTGGCAAAA
TCGTCTCTCTCACCCTCTCAATTTTTACAATTTAAGACTTGGTGGATTGATGGGGTACAAGAACAGGTCCGAAGAAATAGGGCTGCC
AATCCTCCAGTTAACATAGATGCAGATCAACTATTAGGAATAGGTCAAAATTGGAGTACTATTAGTCAACAAGCATTAATGCAAAAT
GAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCTTGGGAAAAAATCCAAGACCCAGGAAGTACCTGCCCCTCATTTAATACA
GTAAGACAAGGTTCAAAAGAGCCCTACCCTGATTTTGTGGCAAGGCTCCAAGATGTTGCTCAAAAGTCAATTGCCGATGAAAAAGCC
GGTAAGGTCATAGTGGAGTTGATGGCATATGAAAACGCCAATCCTGAGTGTCAATCAGCCATTAAGCCATTAAAAGGAAAGGTTCCT
GCAGGATCAGATGTAATCTCAGAATATGTAAAAGCCTGTGATGGAATCGGAGGAGCTATGCATAAAGCTATGCTTATGGCTCAAGCA
ATAACAGGAGTTGTTTTAGGAGGACAAGTTAGAACATTTGGAGGAAAATGTTATAATTGTGGTCAAATTGGTCACTTAAAAAAGAAT
TGCCCAGTCTTAAACAAACAGAATATAACTATTCAAGCAACTACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAAA
GGAAAACATTGGGCTAGTCAATGTCGTTCTAAATTTGATAAAAATGGGCAACCATTGTCGGGAAACGAGCAAAGGGGCCAGCCTCAG
GCCCCACAACAAACTGGGGCATTCCCAATTCAGCCATTTGTTCCTCAGGGTTTTCAGGGACAACAACCCCCACTGTCCCAAGTGTTT
CAGGGAATAAGCCAGTTACCACAATACAACAATTGTCCCTCACCACAAGCGGCAGTGCAGCAGTAGATTTATGTACTATACAAGCAG
TCTCTCTGCTTCCAGGGGAGCCCCCACAAAAAATCCCTACAGGGGTATATGGCCCACTGCCTGAGGGGACTGTAGGACTAATCTTGG

GAAGATCAAGTCTAAATCTAAAAGGAGTTCAAATTCATACTAGTGTGGTTGATTCAGACTATAAAGGCGAAATTCAATTGGTTATTA
GCTCTTCAATTCCTTGGAGTGCCAGTCCAAGAGACAGGATTGCTCAATTATTACTCCTGCCATATATTAAGGGTGGAAATAGTGAAA
TAAAAAGAATAGGAGGGCTTGTAAGCACTGATCCAACAGGAAAGGCTGCATATTGGGCAAGTCAGGTCTCAGAGAACAGACCTGTGT
GTAAGGCCATTATTCAAGGAAAACAGTTTGAAGGGTTGGTAGACACTGGAGCAGATGTCTCTATTATTGCTTTAAATCAGTGGCCAA
AAAACTGGCCTAAACAAAAGGCTGTTACAGGACTTGTCGGCATAGGCACAGCCTCAGAAGTGTATCAAAGTATGGAGATTTTACATT
GCTTAGGGCCAGATAATCAAGAAAGTACTGTTCAGCCAATGATTACTTCAATTCCTCTTAATCTGTGGGGTCGAGATTTATTACAAC
AATGGGGTGCGGAAATCACCATGCCCGCTCCATTATATAGCCCCACGAGTCAAAAAATCATGACCAAGATGGGATATATACCAGGAA
AGGGACTAGGGAAAAATGAAGATGGCATTAAAGTTCCAGTTGAGGCTAAAATAAATCAAGAAAGAGAAGGAATAGGGTATCCTTTTT
AGGGGCGGTCACTGTAGAGCCTCCTAAACCCATACCACTAACTTGGAAAACAGAAAAACCGGTGTGGGTAAATCAGTGGCCGCTACC
AAAACAAAAACTGGAGGCTTTACATTTATTAGCAAATGAACAGTTAGAAAAGGGTCACATTGAGCCTTCGTTCTCACCTTGGAATTC
TCCTGTGTTTGTAATTCAGAAGAAATCAGGCAAATGGCATACGTTAACTGACTTAAGGGCTGTAAACGCCGTAATTCAACCCATGGG
GCCTCTCCAACCCGGGTTGCCCTCTCCGGCCATGATCCCAAAAGATTGGCCTTTAATTATAATTGATCTAAAGGATTGCTTTTTTAC
CATCCCTCTGGCAGAGCAGGATTGTGAAAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGAACCAGCCACCAGGTTTCAGTG
GAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATTTGTCAGACTTTTGTAGGTCGAGCTCTTCAACCAGTGAGAGAAAAGTT
TTCAGACTGTTATATTATTCATTATATTGATGATATTTTATGTGCTGCAGAAACGAAAGATAAATTAATTGACTGTTATACATTTCT
GCAAGCAGAGGTTGCCAATGCTGGACTGGCAATAGCATCCGATAAGATCCAAACCTCTACTCCTTTTCATTATTTAGGGATGCAGAT
AGAAAATAGAAAAATTAAGCCACAAAAAATAGAAATAAGAAAAGACACATTAAAAACACTAAATGATTTTCAAAAATTACTAGGAGA
TATTAATTGGATTCGGCCAACTCTAGGCATTCCTACTTATGCCATGTCAAATTTGTTCTCTATCTTAAGAGGAGACTCAGACTTAAA
TAGTCAAAGAATATTAACCCCAGAGGCAACAAAAGAAATTAAATTAGTGGAAGAAAAAATTCAGTCAGCGCAAATAAATAGAATAGA
TCCCTTAGCCCCACTCCAACTTTTGATTTTTGCCACTGCACATTCTCCAACAGGCATCATTATTCAAAATACTGATCTTGTGGAGTG
GTCATTCCTTCCTCACAGTACAGTTAAGACTTTTACATTGTACTTGGATCAAATAGCTACATTAATCGGTCAGACAAGATTACGAAT
AACAAAATTATGTGGAAATGACCCAGACAAAATAGTTGTCCCTTTAACCAAGGAACAAGTTAGACAAGCCTTTATCAATTCTGGTGC
ATGGCAGATTGGTCTTGCTAATTTTGTGGGACTTATTGATAATCATTACCCAAAAACAAAGATCTTCCAGTTCTTAAAATTGACTAC
TTGGATTCTACCTAAAATTACCAGACGTGAACCTTTAGAAAATGCTCTAACAGTATTTACTGATGGTTCCAGCAATGGAAAAGCAGC
TTACACAGGGCCGAAAGAACGAGTAATCAAAACTCCATATCAATCGGCTCAAAGAGACGAGTTGGTTGCAGTCATTACAGTGTTACA
AGATTTTGACCAACCTATCAATATTATATCAGATTCTGCATATGTAGTACAGGCTACAAGGGATGTTGAGACAGCTCTAATTAAATA
TAGCATGGATGATCAGTTAAACCAGCTATTCAATTTATTACAACAAACTGTAAGAAAAAGAAATTTCCCATTTTATATTACTTATAT
TCGAGCACACACTAATTTACCAGGGCCTTTGACTAAAGCAAATGAACAAGCTGACTTACTGGTATCATCTGCACTCATAAAAGCACA
AGAACTTCATGCTTTGACTCATGTAAATGCAGCAGGATTAAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACA
ACATTGCACCCAGTGTCAAGTCTTACACCTGCCCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCATTATGGCA
AATGGATGTCACGCATGTACCTTCATTTGGAAGATTATCATATGTTCATGTAACAGTTGATACTTATTCACATTTCATATGGGCAAC
TTGCCAAACAGGAGAAAGTACTTCCCATGTTAAAAAACATTTATTGTCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCAAAAC
TGACAATGGACCAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAACAGGAATTCCTTATAA
TTCCCAAGGACAGGCCATAGTTGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGAAGGGGGAGACAGTAAGGA
GTGTACCACTCCTCAGATGCAACTTAATCTAGCACTCTATACTTTAAATTTTTTAAACATTTATAGAAATCAGACTACTACTTCTGC
AGAACAACATCTTACTGGTAAAAAGAACAGCCCACATGAAGGAAAACTAATTTGGTGGAAAGATAATAAAAATAAGACATGGGAAAT
AGGGAAGGTGATAACGTGGGGGAGAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAGCTTCCTGTTTGGTTACCCACTAGACATTT
GAAGTTCTACAATGAACCCATCGGAGATGCAAAGAAAAGGGCCTCCACGGAGATGGTAACACCAGTCACATGGATGGATAATCCTAT
AGAAGTATATGTTAATGATAGTATATGGGTACCTGGCCCCATAGATGATCGCTGCCCTGCCAAACCTGAGGAAGAAGGGATGATGAT
AAATATTTCCATTGGGTATCGTTATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAATGCCTGCAGTCCAAAATTGGTTGGT
AGAAGTACCTACTGTCAGTCCCATCAGTAGATTCACTTATCACATGGTAAGCGGGATGTCACTCAGGCCACGGGTAAATTATTTACA
AGACTTTTCTTATCAAAGATCATTAAAATTTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGA
AGTTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTATAAAACAATGAATTTGGAACTATTATAGATTGGGCACCTCG
AGGTCAATTCTACCACAATTGCTCAGGACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGA
AAGTTTAGACAAACATAAGCATAAAAAATTGCAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATCTCTACCCCAAGACCAAA
AATAGTAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTACTGTGGCCTCACACCACATTAGAATTTGGTCTGGAAA
TCAAACTTTAGAAACAAGAGATTGTAAGCCATTTTATACTGTCGACCTAAATTCCAGTCTAACAGTTCCTTTACAAAGTTGCGTAAA
GCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAGACTATAACCTGTGAAAATTGTAGATTGCTTACTTG
CATTGATTCAACTTTTAATTGGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGGACCGACC
GTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAAAGGTGTTTTAAATAGATCCAAAAGATTCATTTTTACTTTAAT
TGCAGTGATTATGGGATTAATTGCAGTCACAGCTACGGCTGCTGTAGCAGGAGTTGCATTGCACTCTTCTGTTCAGTCAGTAAACTT
TGTTAATGATTGGCAAAAGAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATCAAAAATTGGCAAATCAAATTAATGATCT
TAGACAAACTGTCATTTGGATGGGAGACAGACTCATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTT
TTGTATTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAGGGAAGAGAAGATAATCTCAC
TTTAGACATTTCCAAATTAAAAGAACAAATTTTCGAAGCATCAAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAATTGCAGG
AGTTGCTGATGGCCTCGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATTGGAAGTACATCGATTATAAATCTCATATTAATCCT
TGTGTGCCTGTTTTGTCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCATGATGAC
GATGGCGGTTTTGTCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAAATTGTTACTGTGTCTGTGTAGAAAGAAGT
AGACATAGGAGACTCCATTTTGTTATGTGCTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCC
GTGCTCTCTGAAACATGTGCTGTGTCAACTCAGGGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAA
GGCAGCATGCTCCTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCAGAAGGCCGCAGGGACCTCTG
CCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCC
CCCAGCCCGACACCTGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACAAGAGGAAGG
CATCTGTCTCCTGCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACC
GCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCATTGAGATGTTTATGTGTATGCATATCCAAAAGCAC
AGCACTTAATCCTTTACATTGTCTATGATGCCAAGACCTTTGTTCACGTGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTGA
CCCTGACACATCCCCCTCTTTGAGAAACACCCACAGATGATCAATAAATACTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATG
CTGAACGCTGGTTCCCCGGGTCCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCT
TACGAGAAACACCCACAGGTGTGTAGGGGCAACCCACCCCTACA

SEQ ID 46:

ETQVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTVLDSIAHGNRLTPYDWEILAKSSLSSSQYLQFKTWWIDGVQEQVRK
KSGY*AHC*YRRRPIVRNRSKLEHH*PTISDAE*GY*TSKGYLPQGLGKNSGPRNSFPY*FN*TRL*RAIS*LCGKITRCCSKVYYR
*QCPKSYCRINGL*KCKSRMSVGHKAIKRKSSSRS*CNYRICEGL*WDWRSYA*GNANGSSNEGAHSRRTS*NIWEKML*LWSNRSS
EKELPRLKQAKKKKKKKK

SEQ ID 47:

EETQVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTVLDSIAHGNRLTPYDWEILAKSSLSSSQYLQFKTWWIDGVQEQVR
KNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEAIEQVRAICLRAWGKIQDPGTAFPINSIRQGSKEPYPDFVARLQDAAQKSIT
DDNARKVIVELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGLTLGGQVRTFGKKCYNCGQIGH
LKRSCPGLNKQNIINQAITEKKKKKKK

SEQ ID 48:

EETQVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTLLDSIAHGNRLTPYDWEILAKSSLSSSQYLQFKTWWIDGVQEQVR
KNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEAIEQVRAICLRAWGKIQDPGTAFPINSIRQGSKEPYPDFVARLQDAAQKSIT
DDNARKVIVELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGLTLGGQVRTFGKKCYNCGQIGH
RKRSCPGLNKQNIINQAITAKNKKPSGLCPKCGKAKHWANQCHSKFDKDGQPLSGNRKRGQPQAPQQTGAFPVKLFVPQGFQGQQPL
QKIPPLQGVSQLQQSNSCPAPQQAAPQ*IYVPPKWSFYSLESPHKRFLEGYMARCQKGG*AFEGDQV*I*RESKFILG*FTQIIKGE
FS**SAPLFPGVPIQVIELLNYCFCLMQKKKKKKKK

SEQ ID 49:

GSQAGVKQYGPNSPYIRILLNSIAHGNRLISYDWEILAISSLSPSQYLQFKTWWIDGVQEQVRKNQATNPVAYIDEDQLLGRGPNWD
TINQQSVMKMRLLNNYKGYLPQGLGKHSGPRNLMPFF*FNQTRL*RAISRLCGKVARCSSKIHCR*RPKSYCRNNGLSKRKFRVSIS
HKAIKRKCFSRS*CNYRICEGL*WDWRSYA*GNAIGSSNYRGCYRRTS*NIWGKML*LWSNRSSKKELPELKLPPKKKKKKKKK

SEQ ID 50:

QKNESSKLSIT*LKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFYITHIRAH
TNLPGPLTKANEQADLLVSSAFMEAQELHALTHVNAIGLKNRFDITWKQTKNIVQHCTQCQILHLATQEARVNPRGLCPNVLWQMDV
MHV**PSFGK**LSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLSCFPVMGVPEKVKTDNGPGYCSKAVQKFLNQWKITHTIGILYNSQG
QAIIERTNRTLKAQLVKQKKKKKKKK

SEQ ID 51:

QKNESSKLSIT*LKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFYITHIRAH
TNLPGPLTKANEQADLLVSSAFMEAQELHALTHVNAIGLKNKFDITWKQTKNIVQHCTQCQILHLATQEARVNPRGLCPNVLWQMDV
MHV**PSFGK**LSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLSCFPVMGVPEKVKTDNGPGYCSKAVQKFLNQWKITHTIGILYNSQG
QAIIERTNRTLKAQLVKQKEKKKKK

SEQ ID 52:

QKNESSKLSITRLKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFYITHIRAH
TNLPGPLTKANEQADLLVSSAFMEAQELHALTHVNAIGLKNKFDITWKQTKNIVQHCAQCQILHLATQEVRVNPRGLCPNVLWQMDV
MHV**PSFGK**LSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLSCFPVMGVPEKVKTDNGPGYCSKAVQKFLNQWKITHTIGILYNSQG
QAIIERTNRTLKAQLVKQKKKKKKKK

SEQ ID 53:

QKNESSKLSIT*LKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRK*NFPFYITHIRAH
TNLPGPLTKANEQADLLVSSAFMEAQELHALTHVNAIGLKNKFDITWKQTKNIVQHCTQCQILHLATQEARVNPRGLCPNVLWQMDV
MHV**PSFGK**LSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLFCFPVMGVPEKVKTDNGPGYCSKAVQEFLNQWKITHTIGILYNSQG
QAIIERTNRTLKAQLVKQKKKKKKKK

SEQ ID 54:

QKNESSKLSIT*LKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFYITHIRAH
TNLPGPLTKANEQADLLVSSAFMEAQELHALTHVNAIGLKNKFDITWKQTKNIVQHCTQCQILHLATQEARVNPRGLCPNVLWQMDV
MHV**PSFGK**LSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLSCFPVMGVPEKKKKKKKKK

SEQ ID 55:

QKNESSKLSIT*LKEQSWLPSLQC*QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFYITHIRAH
TNLPGPLTKANEQADLLVSSAFMEAQELHALTHVNAIGLKNKFDITWKQTKNIVQHCTQCQILHLATQEARVNPRGLCPNVLWQMDV
MHV**PSFGK**LSFVHVTVDTYSHFIWATCQTGESTSHVKRHLLSCFPVMGVPEKVKTDNGPGYCSKAVQKFLNQWKITHT
IGILYNSQGQAIIERTNRTLKAQLVKQKKKKKKKTCRPPR

SEQ ID 56:

EETQVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTLLDSIAHGNRLTPYDWEILAKSSLSSSQYLQFKTWWIDGVQEQVR
KNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEAIEQVRAICLRAWGKIQDPGTAFPINSIRQGSKEPYPDFVARLQDAAQKSIT
DDNARKVIVELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGLTLGGQVRTFGKKCYNCGQIGH
RKRSCPGLNKQNIINQAITAKNKKPSGLCPKCGKAKHWANQCHSKFDKDGQPLSGNRKRGQPQAPQQTGAFPVKLFVPQGFQGQQPL
QKIPPLQGVSQLQQSNSCPAPQQAAPQ

SEQ ID 57:

EETQVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTVLDSIAHGNRLTPYDWEILAKSSLSSSQYLQFKTWWIDGVQEQVR
KNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEAIEQVRAICLRAWGKIQDPGTAFPINSIRQGSKEPYPDFVARLQDAAQKSIT
DDNARKVIVELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKAMLMAQAMRGLTLGGQVRTFGKKCYNCGQIGH
LKRSCPGLNKQNIINQAITEKKKKKKK

SEQ ID 58:

QDFNQSINIVSDSAYVVQATKDIERALIKYIMDDQLNPLFNLLQQNVRKRNFPFYITHIRAHTNLPGPLTKANEQADLLVSSAFMEA
QELHALTHVNAIGLKNKFDITWKQTKNIVQHCTQCQILHLATQEARVNPRGLCPNVLWQMDVMHVPSFGKLSFVHVTVDTYSHFIWA
TCQTGESTSHVKRHLLSCFPVMGVPEKVKTDNGPGYCSKAVQKFLNQWKITHTIGILYNSQGQAIIERTNRTLKAQLVKQKKKKKKK
KTCRPPR

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:

ATGGGGCAAACTAAAAGTAAAATTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGGAGTTAAA
GTATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAACTTTAGATCTAAAA
GATTGGAAAAGAATTGGTAAGGAACTAAAACAAGCAGGTAGGAAGGGTAATATCATTCCACTTACAGTATGGAATGATTGGGCCATT
ATTAAAGCAGCTTTAGAACCATTTCAAACAGAAGAAGATAGCGTTTCAGTTTCTGATGCCCCTGGAAGCTGTATAATAGATTGTAAT
GAAAACACAAGGAAAAAATCCCAGAAAGAAACGGAAGGTTTACATTGCGAATATGTAGCAGAGCCGGTAATGGCTCAGTCAACGCAA
AATGTTGACTATAATCAATTACAGGAGGTGATATATCCTGAAACGTTAAAATTAGAAGGAAAAGGTCCAGAATTAGTGGGGCCATCA
GAGTCTAAACCACGAGGCACAAGTCCTCTTCCAGCAGGTCAGGTGCCTGTAACATTACAACCTCAAAAGCAGGTTAAAGAAAATAAG
ACCCAACCGCCAGTAGCCTATCAATACTGGCCTCCGGCTGAACTTCAGTATCGGCCACCCCCAGAAAGTCAGTATGGATATCCAGGA
ATGCCCCCAGCACCACAGGGCAGGGCGCCATACCCTCAGCCGCCCACTAGGAGACTTAATCCTACGGCACCACCTAGTAGACAGGGT

AGTAAATTACATGAAATTATTGATAAATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAACCGATGCCA
CCTGGAGAAGGAGCCCAAGAGGGAGAGCCTCCCACAGTTGAGGCCAGATACAAGTCTTTTTCGATAAAAAAGCTAAAAGATATGAAA
GAGGGAGTAAAACAGTATGGACCCAACTCCCCTTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCCTTAT
GATTGGGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTTACAATTTAAGACTTGGTGGATTGATGGGGTACAAGAACAG
GTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACATAGATGCAGATCAACTATTAGGAATAGGTCAAAATTGGAGTACTATTAGT
CAACAAGCATTAATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAATCCAAGACCCAGGAAGT
ACCTGCCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCCCTATCCTGATTTTGTGGCAAGGCTCCAAGATGTTGCTCAAAAG
TCAATTGCTGATGAAAAAGCCCGTAAGGTCATAGTGGAGTTGATGGCATATGAAAACGCCAATCCTGAGTGTCAATCAGCCATTAAG
CCATTAAAAGGAAAGGTTCCTGCAGGATCAGATGTAATCTCAGAATATGTAAAAGCCTGTGATGGAATCGGAGGAGCTATGCATAAA
GCTATGCTTATGGCTCAAGCAATAACAGGAGTTGTTTTAGGAGGACAAGTTAGAACATTTGGAAGAAAATGTTATAATTGTGGTCAA
ATTGGTCACTTAAAAAAGAATTGCCCAGTCTTAAATAAACAGAATATAACTATTCAAGCAACTACAACAGGTAGAGAGCCACCTGAC
TTATGTCCAAGATGTAAAAAAGGAAAACATTGGGCTAGTCAATGTCGTTCTAAATTTGATAAAAATGGGCAACCATTGTCGGGAAAC
GAGCAAAGGGGCCAGCCTCAGGCCCCACAACAAACTGGGGCATTCCCAATTCAGCCATTTGTTCCTCAGGGTTTTCAGGGACAACAA
CCCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCAGTTACCACAATACAACAATTGTCCCCCGCCACAAGCGGCAGTGCAGCAG

SEQ ID 86:

ATGGGCAACCATTGTCGGGAAACGAGCAAAGGGGCCAGCCTCAGGCCCCACAACAAACTGGGGCATTCCCAATTCAGCCATTTGTTC
CTCAGGGTTTTCAGGGACAACAACCCCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCAGTTACCACAATACAACAATTGTCCCCCGC
CACAAGCGGCAGTGCAGCAGTAGATTTATGTACTATACAAGCAGTCTCTCTGCTTCCAGGGGAGCCCCCACAAAAAACCCCCCACAGG
GGTATATGGACCCCTGCCTAAGGGGACTGTAGGACTAATCTTGGGACGATCAAGTCTAAATCTAAAAGGAGTTCAAATTCATACTAG
TGTGGTTGATTCAGACTATAAAGGCGAAATTCAATTGGTTATTAGCTCTTCAATTCCTTGGAGTGCCAGTCCAAGAGACAGGATTGC
TCAATTATTACTCCTGCCATACATTAAGGGTGGAAATAGTGAAATAAAAAGAATAGGAGGGCTTGGAAGCACTGATCCAACAGGAAA
GGCTGCATATTGGGCAAGTCAGGTCTCAGAGAACAGACCTGTGTGTAAGGCCATTATTCAAGGAAAACAGTTTGAAGGGTTGGTAGA
CACTGGAGCAGATGTCTCTATCATTGCTTTAAATCAGTGGCCAAAAAATTGGCCTAAACAAAAGGCTGTTACAGGACTTGTCGGCAT
AGGCACAGCCTCAGAAGTGTATCAAAGTACGGAGATTTTACATTGCTTAGGGCCAGATAATCAAGAAAGTACTGTTCAGCCAATGAT
TACTTCAATTCCTCTTAATCTGTGGGGTCGAGATTTATTACAACAATGGGGTGCGGAAATCACCATGCCCGCTCCATCATATAGCCC
CACGAGTCAAAAAATCATGACCAAGATGGGATATATACCAGGAAAGGGACTAGGGAAAAATGAAGATGGCATTAAAATTCCAGTTGA
GGCTAAAATAAATCAAGAAAGAGAAGGAATAGGGAATCCTTGC

SEQ ID 87:

ATGGCATTAAAATTCCAGTTGAGGCTAAAATAAATCAAGAAAGAGAAGGAATAGGGAATCCTTGCTAGGGGCGGCCACTGTAGAGCC
TCCTAAACCCATACCATTAACTTGGAAAACAGAAAAACCAGTGTGGGTAAATCAGTGGCCGCTACCAAAACAAAAACTGGAGGCTTT
ACATTTATTAGCAAATGAACAGTTAGAAAAGGGTCATATTGAGCCTTCGTTCTCACCTTGGAATTCTCCTGTGTTTGTAATTCAGAA
GAAATCAGGCAAATGGCGTATGTTAACTGACTTAAGGGCTGTAAACGCCGTAATTCAACCCATGGGGCCTCTCCAACCCGGGTTGCC
CTCTCCGGCCATGATCCCAAAAGATTGGCCTTTAATTATAATTGATCTAAAGGATTGCTTTTTTACCATCCCTCTGGCAGAGCAGGA
TTGCGAAAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTGTTACCTCAGGGAAT
GCTTAATAGTCCAACTATTTGTCAGACTTTTGTAGGTCGAGCTCTTCAACCAGTTAGAGAAAAGTTTTCAGACTGTTATATTATTCA
TTGTATTGATGATATTTTATGTGCTGCAGAAACGAAAGATAAATTAATTGACTGTTATACATTTCTGCAAGCAGAGGTTGCCAATGC
TGGACTGGCAATAGCATCTGATAAGATCCAAACCTCTACTCCTTTTCATTATTTAGGGATGCAGATAGAAAATAGAAAAATTAAGCC
ACAAAAAATAGAAATAAGAAAAGACACATTAAAAACACTAAATGATTTTCAAAAATTACTAGGAGATATTAATTGGATTCGGCCAAC
TCTAGGCATTCCTACTTATGCCATGTCAAATTTGTTCTCTATCTTAAGAGGAGACTCAGACTTAAATAGTAAAAGAATGTTAACCCC
AGAGGCAACAAAAGAAATTAAATTAGTGGAAGAAAAAATTCAGTCAGCGCAAATAAATAGAATAGATCCCTTAGCCCCACTCCAACT
TTTGATTTTTGCCACTGCACATTCTCCAACAGGCATCATTATTCAAAATACTGATCTTGTGGAGTGGTCATTCCTTCCTCACAGTAC
AGTTAAGACTTTTACATTGTACTTGGATCAAATAGCTACATTAATCGGTCAGACAAGATTACGAATAATAAAATTATGTGGGAATGA
CCCAGACAAAATAGTTGTCCCTTTAACCAAGGAACAAGTTAGACAAGCCTTTATCAATTCTGGTGCATGGAAGATTGGTCTTGCTAA
TTTTGTGGGAATTATTGATAATCATTACCCAAAAACAAAGATCTTCCAGTTCTTAAAATTGACTACTTGGATTCTACCTAAAATTAC
CAGACGTGAACCTTTAGAAAATGCTCTAACAGTATTTACTGATGGTTCCAGCAATGGAAAAGCAGCTTACACAGGACCGAAAGAACG
AGTAATCAAAACTCCATATCAATCGGCTCAAAGAGCAGAGTTGGTTGCAGTCATTACAGTGTTACAAGATTTTGACCAACCTATCAA
TATTATATCAGATTCTGCATATGTAGTACAGGCTACAAGGGATGTTGAGACAGCTCTAATTAAATATAGCATGGATGATCAGTTAAA
CCAGCTATTCAATTTATTACAACAAACTGTAAGAAAAAGAAATTTCCCATTTTATATTACACATATTCGAGCACACACTAATTTACC
AGGGCCTTTGACTAAAGCAAATGAACAAGCTGACTTACTGGTATCATCTGCACTCATAAAAGCACAAGAACTTCATGCTTTGACTCA
TGTAAATGCAGCAGGATTAAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACATTGCACCCAGTGTCAAGT
CTTACACCTGCCCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCATTATGGCAAATGGATGTCACGCATGTACC
TTCATTTGGAAGATTATCATATGTTCACGTAACAGTTGATACTTATTCACATTTCATATGGGCAACTTGCCAAACAGGAGAAAGTAC
TTCCCATGTTAAAAAACATTTATTGTCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCAAAACTGACAATGGACCAGGATATTG
TAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAACAGGAATTCCTTATAATTCCCAAGGACAGGCCATAGT
TGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGAAGGGGGAGACAGTAAGGAGTGTACCACTCCTCAGATGCA
ACTTAATCTAGCACTCTATACTTTAAATTTTTTAAACATTTATAGAAATCAGACTACTACTTCTGCAGAACAACATCTTACTGGTAA
AAAGAACAGCCCACATGAAGGAAAACTAATTTGGTGGAAAGATAATAAAAATAAGACATGGGAAATAGGGAAGGTGATAACGTGGGG
GAGAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAGCTTCCTGTTTGGATACCCACTAGACATTTGAAGTTCTACAATGAACCCAT
CAGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACATCGCAATCGAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGT
CAGAAGAACAGATGAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAACTAAAGAAGCTGACGCAGTTAGCTACAA
AATATCTAGAGAACACAAAGGTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCCTTGATGATTGTATCAATGGTGGTAAGTCTCC
CTATGCCTGCAGGAGCAGCTGCAGC

SEQ ID 88:

ATGAACCCATCAGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACATCGCAATCGAGCACCGTTGACTCACAAGATGAACAAA
ATGGTGACGTCAGAAGAACAGATGAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAACTAAAGAAGCTGACGCAG
TTAGCTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCCTTGATGATTGTATCAATGGTG
GTAAGTCTCCCTATGCCTGCAGGAGCAGCTGCAGCTAACTATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTC
ACATGGATGGATAATCCTACAGAAGTATATGTTAATGATAGTGTATGGGTACCTGGCCCCATAGATGATCGCTGCCCTGCCAAACCT
GAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCATTATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAATGCCT

GCAGTCCAAAATTGGTTGGTAGAAGTACCTACTGTCAGTCCCATCTGTAGATTCACTTATCACATGGTAAGCGGGATGTCACTCAGG
CCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTAAAATTTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCC
AAAGAATCAAAAAATACAGAAGTTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTCGGAACT
ATTATAGATTGGGCACCTCGAGGTCAATTCTACCACAATTGCTCAGGACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCAGCT
GTTGATAGCGACTTAACAGAAAGTTTAGACAAACATAAGCATAAAAAATTGCAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGA
ATCTCTACCCCAAGACCAAAAATAGTAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTACTGTGGCCTCACACCAC
ATTAGAATTTGGTCTGGAAATCAAACTTTAGAAACAAGAGATCGTAAGCCATTTTATACTATTGACCTGAATTCCAGTCTAACAGTT
CCTTTACAAAGTTGCGTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAGACTATAACCTGTGAA
AATTGTAGATTGCTTACTTGCATTGATTCAACTTTTAATTGGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATC
CCTGTGTCCATGGACCGACCGTGGGAGGCCTCGCCATCCGTCCATATTTTGACTGAAGTATTAAAAGGTGTTTTAAATAGATCCAAA
AGATTCATTTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAGTCACAGCTACGGCTGCTGTAGCAGGAGTTGCATTGCACTCT
TCTGTTCAGTCAGTAAACTTTGTTAATGATTGGCAAAAAAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATCAAAAATTG
GCAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGACTCATGAGCTTAGAACATCGTTTCCAGTTACAATGT
GACTGGAATACGTCAGATTTTTGTATTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAG
GGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAAGAACAAATTTTCGAAGCATCAAAAGCCCATTTAAATTTGGTGCCA
GGAACTGAGGCAATTGCAGGAGTTGCTGATGGCCTCGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATTGGAAGTACTACGATT
ATAAATCTCATATTAATCCTTGTGTGCCTGTTTTGTCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCA

SEQ ID 89:

AGTTCTACAATGAACCCATCAGAGATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACATCGCAATCGAGCACCGTTGACTCACAAG
ATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAACTAAAGAAG
CTGACGCAGTTAGCTACAAAATATCTAGAGAACACAAAGGTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCCTTGATGATTGTA
TCAATGGTGGTAAGTCTCCCTATGCCTGCAGGA

SEQ ID 90:

TCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCATGA

SEQ ID 91:

ATGGGGCAAACTAAAAGTAAAATTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGGAGTTAAA
GTATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAACTTTAGATCTAAAA
GATTGGAAAAGAATTGGTAAGGAACTAAAACAAGCAGGTAGGAAGGGTAATATCATTCCACTTACAGTATGGAATGATTGGGCCATT
ATTAAAGCAGCTTTAGAACCATTTCAAACAGAAGAAGATAGCGTTTCAGTTTCTGATGCCCCTGGAAGCTGTATAATAGATTGTAAT
GAAAACACAAGGAAAAAATCCCAGAAAGAAACGGAAGGTTTACATTGCGAATATGTAGCAGAGCCGGTAATGGCTCAGTCAACGCAA
AATGTTGACTATAATCAATTACAGGAGGTGATATATCCTGAAACGTTAAAATTAGAAGGAAAAGGTCCAGAATTAGTGGGGCCATCA
GAGTCTAAACCACGAGGCACAAGTCCTCTTCCAGCAGGTCAGGTGCCTGTAACATTACAACCTCAAAAGCAGGTTAAAGAAAATAAG
ACCCAACCGCCAGTAGCCTATCAATACTGGCCTCCGGCTGAACTTCAGTATCGGCCACCCCCAGAAAGTCAGTATGGATATCCAGGA
ATGCCCCCAGCACCACAGGGCAGGGCGCCATACCCTCAGCCGCCCACTAGGAGACTTAATCCTACGGCACCACCTAGTAGACAGGGT
AGTAAATTACATGAAATTATTGATAAATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAACCGATGCCA
CCTGGAGAAGGAGCCCAAGAGGGAGAGCCTCCCACAGTTGAGGCCAGATACAAGTCTTTTTCGATAAAAAAGCTGAAAGATATGAAA
GAGGGAGTAAAACAGTATGGACCCAACTCCCCTTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCCTTAT
GATTGGGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTTACAATTTAAGACTTGGTGGATTGATGGGGTACAAGAACAG
GTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACATAGATGCAGATCAACTATTAGGAATAGGTCAAAATTGGAGTACTATTAGT
CAACAAGCATTAATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAATCCAAGACCCAGGAAGT
ACCTGCCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCCCTATCCTGATTTTGTGGCAAGGCTCCAAGATGTTGCTCAAAAG
TCAATTGCTGATGAAAAAGCCCGTAAGGTCATAGTGGAGTTGATGGCATATGAAAACGCCAATCCTGAGTGTCAATCAGCCATTAAG
CCATTAAAAGGAAAGGTTCCTGCAGGATCAGATGTAATCTCAGAATATGTAAAAGCCTGTGATGGAATCGGAGGAGCTATGTATAAA
GCTATGCTTATGGCTCAAGCAATAACAGGAGTTGTTTTAGGAGGACAAGTTAGAACATTTGGAAGAAAATGTTATAATTGTGGTCAA
ATTGGTCACTTAAAAAAGAATTGCCCAGTCTTAAATAAACAGAATATAACTATTCAAGCAACTACAACAGGTAGAGAGCCACCTGAC
TTATGTCCAAGATGTAAAAAAGGAAAACATTGGGCTAGTCAATGTCGTTCTAAATTTGATAAAAATGGGCAACCATTGTCGGGAAAC
GAGCAAAGGGGCCAGCCTCAGGCCCCACAACAAACTGGGGCATTCCCAATTCAGCCATTTGTTCCTCAGGGTTTTCAGGGACAACAA
CCCCCACTGTCCCAAGTGTTTCAGGGAATAAGCCAGTTACCACAATACAACAATTGTCCCCCGCCACAAGCGGCAGTGCAGCAGTAG

SEQ ID 92:

MGQTKSKIKSKYASYLSFIKILLKRGGVKVSTKNLIKLFQIIEQFCPWFPEQGTLDLKDWKRIGKELKQAGRKGNIIPLTVWNDWAI
IKAALEPFQTEEDSVSVSDAPGSCIIDCNENTRKKSQKETEGLHCEYVAEPVMAQSTQNVDYNQLQEVIYPETLKLEGKGPELVGPS
ESKPRGTSPLPAGQVPVTLQPQKQVKENKTQPPVAYQYWPPAELQYRPPPESQYGYPGMPPAPQGRAPYPQPPTRRLNPTAPPSRQG
SKLHEIIDKSRKEGDTEAWQFPVTLEPMPPGEGAQEGEPPTVEARYKSFSIKKLKDMKEGVKQYGPNSPYMRTLLDSIAHGHRLIPY
DWEILAKSSLSPSQFLQFKTWWIDGVQEQVRRNRAANPPVNIDADQLLGIGQNWSTISQQALMQNEAIEQVRAICLRAWEKIQDPGS
TCPSFNTVRQGSKEPYPDFVARLQDVAQKSIADEKARKVIVELMAYENANPECQSAIKPLKGKVPAGSDVISEYVKACDGIGGAMYK
AMLMAQAITGVVLGGQVRTFGRKCYNCGQIGHLKKNCPVLNKQNITIQATTTGREPPDLCPRCKKGKHWASQCRSKFDKNGQPLSGN
EQRGQPQAPQQTGAFPIQPFVPQGFQGQQPPLSQVFQGISQLPQYNNCPPPQAAVQQ

SEQ ID 93:

ATGTTAACTGACTTAAGGGCTGTAAACGCCGTAATTCAACCCATGGGGCCTCTCCAACCCGGGTTGCCCTCTCCGGCCATGATCCCA
AAAGATTGGCCTTTAATTATAATTGATCTAAAGGATTGCTTTTTTACCATCCCTCTGGCAGAGCAGGATTGCGAAAAATTTGCCTTT
ACTATACCAGCCATAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATT
TGTCAGACTTTTGTAGGTCGAGCTCTTCAACCAGTTAGAGAAAAGTTTTCAGACTGTTATATTATTCATTGTATTGATGATATTTTA
TGTGCTGCAGAAACGAAAGATAAATTAATTGACTGTTATACATTTCTGCAAGCAGAGGTTGCCAATGCTGGACTGGCAATAGCATCT
GATAAGATCCAAACCTCTACTCCTTTTCATTATTTAGGGATGCAGATAGAAAATAGAAAAATTAAGCCACAAAAAATAGAAATAAGA
AAAGACACATTAAAAACACTAAATGATTTTCAAAAATTACTAGGAGATATTAATTGGATTCGGCCAACTCTAGGCATTCCTACTTAT
GCCATGTCAAATTTGTTCTCTATCTTAAGAGGAGACTCAGACTTAAATAGTAAAAGAATGTTAACCCCAGAGGCAACAAAAGAAATT
AAATTAGTGGAAGAAAAAATTCAGTCAGCGCAAATAAATAGAATAGATCCCTTAGCCCCACTCCAACTTTTGATTTTTGCCACTGCA
CATTCTCCAACAGGCATCATTATTCAAAATACTGATCTTGTGGAGTGGTCATTCCTTCCTCACAGTACAGTTAAGACTTTTACATTG

TACTTGGATCAAATAGCTACATTAATCGGTCAGACAAGATTACGAATAATAAAATTATGTGGGAATGACCCAGACAAAATAGTTGTC
CCTTTAACCAAGGAACAAGTTAGACAAGCCTTTATCAATTCTGGTGCATGGAAGATTGGTCTTGCTAATTTTGTGGGAATTATTGAT
AATCATTACCCAAAAACAAAGATCTTCCAGTTCTTAAAATTGACTACTTGGATTCTACCTAAAATTACCAGACGTGAACCTTTAGAA
AATGCTCTAACAGTATTTACTGATGGTTCCAGCAATGGAAAAGCAGCTTACACAGGACCGAAAGAACGAGTAATCAAAACTCCATAT
CAATCGGCTCAAAGAGCAGAGTTGGTTGCAGTCATTACAGTGTTACAAGATTTTGACCAACCTATCAATATTATATCAGATTCTGCA
TATGTAGTACAGGCTACAAGGGATGTTGAGACAGCTCTAATTAAATATAGCATGGATGATCAGTTAAACCAGCTATTCAATTTATTA
CAACAAACTGTAAGAAAAAGAAATTTCCCATTTTATATTACACATATTCGAGCACACACTAATTTACCAGGGCCTTTGACTAAAGCA
AATGAACAAGCTGACTTACTGGTATCATCTGCACTCATAAAAGCACAAGAACTTCATGCTTTGACTCATGTAAATGCAGCAGGATTA
AAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACATTGCACCCAGTGTCAAGTCTTACACCTGCCCACTCAA
GAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCATTATGGCAAATGGATGTCACGCATGTACCTTCATTTGGAAGATTATCA
TATGTTCACGTAACAGTTGATACTTATTCACATTTCATATGGGCAACTTGCCAAACAGGAGAAAGTACTTCCCATGTTAAAAAACAT
TTATTGTCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCAAAACTGACAATGGACCAGGATATTGTAGTAAAGCTTTCCAAAAA
TTCTTAAGTCAGTGGAAAATTTCACATACAACAGGAATTCCTTATAATTCCCAAGGACAGGCCATAGTTGAAAGAACTAATAGAACA
CTCAAAACTCAATTAGTTAAACAAAAAGAAGGGGGAGACAGTAAGGAGTGTACCACTCCTCAGATGCAACTTAATCTAGCACTCTAT
ACTTTAAATTTTTTAAACATTTATAGAAATCAGACTACTACTTCTGCAGAACAACATCTTACTGGTAAAAAGAACAGCCCACATGAA
GGAAAACTAATTTGGTGGAAAGATAGTAAAAATAAGACATGGGAAATAGGGAAGGTGATAACGTGGGGGAGAGGTTTTGCTTGTGTT
TCACCAGGAGAAAATCAGCTTCCTGTTTGGATACCCACTAGACATTGAAGTTCTACAATGAACCCATCAGAGATGCAAAGAAAAGC
ACCTCCGCGGAGACGGAGACATCGCAATCGAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGATGAAGTT
GCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAACTAAAGAAGCTGACGCAGTTAGCTACAAAATATCTAGAGAACACAAA
GGTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATGCCTGCAGGAGCAGC
TGCAGCTAA

SEQ ID 94:

MLTDLRAVNAVIQPMGPLQPGLPSPAMIPKDWPLIIIDLKDCFFTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGMLNSPTI
CQTFVGRALQPVREKFSDCYIIHCIDDILCAAETKDKLIDCYTFLQAEVANAGLAIASDKIQTSTPFHYLGMQIENRKIKPQKIEIR
KDTLKTLNDFQKLLGDINWIRPTLGIPTYAMSNLFSILRGDSDLNSKRMLTPEATKEIKLVEEKIQSAQINRIDPLAPLQLLIFATA
HSPTGIIIQNTDLVEWSFLPHSTVKTFTLYLDQIATLIGQTRLRIIKLCGNDPDKIVVPLTKEQVRQAFINSGAWKIGLANFVGIID
NHYPKTKIFQFLKLTTWILPKITRREPLENALTVFTDGSSNGKAAYTGPKERVIKTPYQSAQRAELVAVITVLQDFDQPINIISDSA
YVVQATRDVETALIKYSMDDQLNQLFNLLQQTVRKRNFPFYITHIRAHTNLPGPLTKANEQADLLVSSALIKAQELHALTHVNAAGL
KNKFDVTWKQAKDIVQHCTQCQVLHLPTQEAGVNPRGLCPNALWQMDVTHVPSFGRLSYVHVTVDTYSHFIWATCQTGESTSHVKKH
LLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQGQAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALY
TLNFLNIYRNQTTTSAEQHLTGKKNSPHEGKLIWWKDSKNKTWEIGKVITWGRGFACVSPGENQLPVWIPTRHLKFYNEPIRDAKKS
TSAETETSQSSTVDSQDEQNGDVRRTDEVAIHQEGRAANLGTTKEADAVSYKISREHKGDTNPREYAACSLDDCINGGKSPYACRSS
CS

SEQ ID 95:

ATGCAAAGAAAAGCACCTCCGCGGAGACGGAGACATCGCAATCGAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAA
GAACAGATGAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCAACTTGGGCACAACTAAAGAAGCTGACGCAGTTAGCTACAAAATAT
CTAGAGAACACAAAGGTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCCTTGATGATTGTATCAATGGTGGTAAGTCTCCCTATG
CCTGCAGGAGCAGCTGCAGCTAACTATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCACATGGATGGATAAT
CCTACAGAAGTATATGTTAATGATAGTGTATGGGTACCTGGCCCCATAGATGATCGCTGCCCTGCCAAACCTGAGGAAGAAGGGATG
ATGATAAATATTTCCATTGGGTATCATTATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAATGCCTGCAGTCCAAAATTGG
TTGGTAGAAGTACCTACTGTCAGTCCCATCTGTAGATTCACTTATCACATGGTAAGCGGGATGTCACTCAGGCCACGGGTAAATTAT
TTACAAGACTTTTCTTATCAAAGATCATTAAAATTTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAAT
ACAGAAGTTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTCGGAACTATTATAGATTGGGCA
CCTCGAGGTCAATTCTACCACAATTGCTCAGGACAAACTCAGTCGTGTCAAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTA
ACAGAAAGTTTAGACAAACATAAGCATAAAAAATTGCAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATCTCTACCCCAAGA
CCAAAAATAGTAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTACTGTGGCCTCACACCACATTAGAATTTGGTCT
GGAAATCAAACTTTAGAAACAAGAGATCGTAAGCCATTTTATACTATTGACCTGAATTCCAGTCTAACAGTTCCTTTACAAAGTTGC
GTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAGACTATAACCTGTGAAAATTGTAGATTGCTT
ACTTGCATTGATTCAACTTTTAATTGGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGGAC
CGACCGTGGGAGGCCTCGCCATCCGTCCATATTTTGACTGAAGTATTAAAAGGTGTTTTAAATAGATCCAAAAGATTCATTTTTACT
TTAATTGCAGTGATTATGGGATTAATTGCAGTCACAGCTACGGCTGCTGTAGCAGGAGTTGCATTGCACTCTTCTGTTCAGTCAGTA
AACTTTGTTAATGATTGGCAAAAAAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATCAAAAATTGGCAAATCAAATTAAT
GATCTTAGACAAACTGTCATTTGGATGGGAGACAGACTCATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCA
GATTTTTGTATTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAGGGAAGAGAAGATAAT
CTCACTTTAGACATTTCCAAATTAAAAGAACAAATTTTCGAAGCATCAAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAATT
GCAGGAGTTGCTGATGGCCTCGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATTGGAAGTACTACGATTATAAATCTCATATTA
ATCCTTGTGTGCCTGTTTTGTCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCATG
ATGACGATGGCGGTTTTGTCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAG

SEQ ID 96:

MQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKLPSTKKAEPPTWAQLKKLTQLATKYLENTKVTQTPESMLLAALMIVSMVVSLPM
PAGAAAANYTYWAYVPFPPLIRAVTWMDNPTEVYVNDSVWVPGPIDDRCPAKPEEEGMMINISIGYHYPPICLGRAPGCLMPAVQNW
LVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQRSLKFRPKGKPCPKEIPKESKNTEVLVWEECVANSAVILQNNEFGTIIDWA
PRGQFYHNCSGQTQSCQSAQVSPAVDSDLTESLDKHKHKKLQSFYPWEWGEKGISTPRPKIVSPVSGPEHPELWRLTVASHHIRIWS
GNQTLETRDRKPFYTIDLNSSLTVPLQSCVKPPYMLVVGNIVIKPDSQTITCENCRLLTCIDSTFNWQHRILLVRAREGVWIPVSMD
RPWEASPSVHILTEVLKGVLNRSKRFIFTLIAVIMGLIAVTATAAVAGVALHSSVQSVNFVNDWQKNSTRLWNSQSSIDQKLANQIN
DLRQTVIWMGDRLMSLEHRFQLQCDWNTSDFCITPQIYNESEHHWDMVRRHLQGREDNLTLDISKLKEQIFEASKAHLNLVPGTEAI
AGVADGLANLNPVTWVKTIGSTTIINLILILVCLFCLLLVCRCTQQLRRDSDHRERAMMTMAVLSKRKGGNVGKSKRDQIVTVSV

SEQ ID 97:

ATGGGGCAAACTAAAAGTAAAACTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGGAGTTAGA
GTATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAACTTTAGATCTAAAA
GATTGGAAAAGAATTGGCGAGGAACTAAAACAAGCAGGTAGAAAGGGTAATATCATTCCACTTACAGTATGGAATGATTGGGCCATT
ATTAAAGCAGCTTTAGAACCATTTCAAACAAAAGAAGATAGCGTTTCAGTTTCTGATGCCCCTGGAAGCTGTGTAATAGATTGTAAT
GAAAAGACAGGGAGAAAATCCCAGAAAGAAACAGAAAGTTTACATTGCGAATATGTAACAGAGCCAGTAATGGCTCAGTCAACGCAA
AATGTTGACTATAATCAATTACAGGGGGTGATATATCCTGAAACGTTAAAATTAGAAGGAAAAGGTCCAGAATTAGTGGGGCCATCA
GAGTCTAAACCACGAGGGCCAAGTCCTCTTCCAGCAGGTCAGGTGCCCGTAACATTACAACCTCAAACGCAGGTTAAAGAAAATAAG
ACCCAACCGCCAGTAGCTTATCAATACTGGCCGCCGGCTGAACTTCAGTATCTGCCACCCCCAGAAAGTCAGTATGGATATCCAGGA
ATGCCCCCAGCACTACAGGGCAGGGCGCCATATCCTCAGCCGCCCACTGTGAGACTTAATCCTACAGCATCACGTAGTGGACAAGGT
GGTACACTGCACGCAGTCATTGATGAAGCCAGAAAACAGGGAGATCTTGAGGCATGGCGGTTCCTGGTAATTTTACAACTGGTACAG
GCCGGGGAAGAGACTCAAGTAGGAGCGCCTGCCCGAGCTGAGACTAGATGTGAACCTTTCACCATGAAAATGTTAAAAGATATAAAG
GAAGGAGTTAAACAATATGGATCCAACTCCCCTTATATAAGAACATTATTAGATTCCATTGCTCATGGAAATAGACTTACTCCTTAT
GACTGGGAAAGTTTGGCCAAATCTTCCCTTTCATCCTCTCAGTATCTACAGTTTAAAACCTGGTGGATTGATGGAGTACAAGAACAG
GTACGAAAAAATCAGGCTACTAAGCCCACTGTTAATATAGACGCAGACCAATTGTTAGGAACAGGTCCAAATTGGAGCACCATTAAC
CAACAATCAGTGATGCAGAATGAGGCTATTGAACAAGTAAGGGCTATTTGCCTCAGGGCCTGGGGAAAAATTCAGGACCCAGGAACA
GCTTTCCCTATTAATTCAATTAGACAAGGCTCTAAAGAGCCATATCCTGACTTTGTGGCAAGATTACAAGATGCTGCTCAAAAGTCT
ATTACAGATGACAATGCCCGAAAAGTTATTGTAGAATTAATGGCCTATGAAAATGCAAATCCAGAATGTCAGTCGGCCATAAAGCCA
TTAAAAGGAAAAGTTCCAGCAGGAGTTGATGTAATTACAGAATATGTGAAGGCTTGTGATGGGATTGGAGGAGCTATGCATAAGGCA
ATGCTAATGGCTCAAGCAATGAGGGGGCTCACTCTAGGAGGACAAGTTAGAACATTTGGGAAAAAATGTTATAATTGTGGTCAAATC
GGTCATCTGAAAAGGAGTTGCCCAGTCTTAAATAAACAGAATATAATAAATCAAGCTATTACAGCAAAAAATAAAAAGCCATCTGGC
CTGTGTCCAAAATGTGGAAAAGGAAAACATTGGGCCAATCAATGTCATTCTAAATTTGATAAAGATGGGCAACCATTGTCGGGAAAC
AGGAAGAGGGGCCAGCCTCAGGCCCCCCAACAAACTGGGGCATTCCCAGTTCAACTGTTTGTTCCTCAGGGTTTTCAAGGACAACAA
CCCCTACAGAAAATACCACCACTTCAGGGAGTCAGCCAATTACAACAATCCAACAGCTGTCCCGCGCCACAGCAGGCAGCGCCACAG
TAG

SEQ ID 98:

MGQTKSKTKSKYASYLSFIKILLKRGGVRVSTKNLIKLFQIIEQFCPWFPEQGTLDLKDWKRIGEELKQAGRKGNIIPLTVWNDWAI
IKAALEPFQTKEDSVSVSDAPGSCVIDCNEKTGRKSQKETESLHCEYVTEPVMAQSTQNVDYNQLQGVIYPETLKLEGKGPELVGPS
ESKPRGPSPLPAGQVPVTLQPQTQVKENKTQPPVAYQYWPPAELQYLPPPESQYGYPGMPPALQGRAPYPQPPTVRLNPTASRSGQG
GTLHAVIDEARKQGDLEAWRFLVILQLVQAGEETQVGAPARAETRCEPFTMKMLKDIKEGVKQYGSNSPYIRTLLDSIAHGNRLTPY
DWESLAKSSLSSSQYLQFKTWWIDGVQEQVRKNQATKPTVNIDADQLLGTGPNWSTINQQSVMQNEAIEQVRAICLRAWGKIQDPGT
AFPINSIRQGSKEPYPDFVARLQDAAQKSITDDNARKVIVELMAYENANPECQSAIKPLKGKVPAGVDVITEYVKACDGIGGAMHKA
MLMAQAMRGLTLGGQVRTFGKKCYNCGQIGHLKRSCPVLNKQNIINQAITAKNKKPSGLCPKCGKGKHWANQCHSKFDKDGQPLSGN
RKRGQPQAPQQTGAFPVQLFVPQGFQGQQPLQKIPPLQGVSQLQQSNSCPAPQQAAPQ

SEQ ID 99:

ATGGGCAACCATTGTCGGGAAACAGGAAGAGGGGCCAGCCTCAGGCCCCCCAACAAACTGGGGCATTCCCAGTTCAACTGTTTGTTC
CTCAGGGTTTTCAAGGACAACAACCCCTACAGAAAATACCACCACTTCAGGGAGTCAGCCAATTACAACAATCCAACAGCTGTCCCG
CGCCACAGCAGGCAGCGCCACAGTAGATTTATGTTCCACCCAAATGGTCTCTTTACTCCCTGGAGAGCCCCCACAAAAGATTCCTAG
AGGGGTATATGGCCCGCTGCCAGAAGGGAGGGTAGGCCTTATTTTAGGGAGATCAAGTCTAAATTTGAAGGGAGTCCAAATTCATAC
TGGGGTAATTTATTCAGATTATAAAGGGGGAATTCAGTTAGTGATCAGCTCCACTGTTCCCTGGAGTGCCAATCCAGGTGATAGAAT
TGCTCAATTACTGCTTTTGCCTTATGTTAAAATTGGGGAAAACAAAACGGAAAGAACAGGAGGGTTTGGAAGTACCAACCCTGCAGG
AAAAGCCACTTATTGGGCTAATCAGGTCTCAGAGGATAGACCCGTGTGTACAGTCACTATTCAGGGAAAGAGTTTGAAGGATTAGTG
GATACCCAGGCTGATGTTTCTATCATCGGCATAGGCACCGCCTCAGAAGTGTATCAAAGTGCCATGATTTTACATTGTCTAGGATCT
GATAATCAAGAAAGTACGGTTCAGCCTATGATCACTTCTATTCCAATCAATTTATGGGGCCGAGACTTGTTACAACAATGGCATGCA
GAGATTACTATCCCAGCCTCCCTATACAGCCCCAGGAATCAAAAAATCATGACTAAAATGGGATAGCTCCCTAAAAAGGGACTAGGA
AAGAATGAAGATGGCATTAAAGTCCCAACTGAGGCTGAAAAAAATCAAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGAAGCGGTC
ACTGTAGAGCCTCCAAAACCCATTCCATTAATTTGGGGGGAAAAAAA

SEQ ID 100:

ATGGCATTAAAGTCCCAACTGAGGCTGAAAAAAATCAAAAAAAGAAAAGGAATAGGGCATCCTTTTTAGAAGCGGTCACTGTAGAGC
CTCCAAAACCCATTCCATTAATTTGGGGGGAAAAAAAAAACTGTATGGTAAATCAGTAGCCGCTTCCAAAACAAAAACTGGAGGCTT
TACACTTATTAGCAAAGAAACAGTTAGAAAAAGGACATATTGAGCCTTCATTTTCGCCTTGGAATTCTCCTGTTTGTAATTCAGAAA
AAATCCGGCAGATGGCGTATGCTAACTGACTTAAGAGCCATTAATGCCATAATTCAACCCATGGGGGCTCTCCCATCCCGGTTGCCC
TCTCCAGCCATGGTCCCCTTTAATTATAATTGATCTGAAGGATTGCTTTTTTACCATTCCTCTGGCAAAAGAGGATTTTGAAAAATT
TGCTTTTACTATACCAGCCTAAATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTATTGCCTCAGGGAATGCTTAATAATTCA
ACTATTTGTCAGACTTTCATAGCTCAAGCTCTGCAACCAGTTAGAGACAAGTTTTCAGACTGTTATATCGTTCATTATGTTGATATT
TTGTGTGCTGCAGAAACGAGAGACAAATTAATTGACCGTTACACATTTCTCAGACAGAGGTTGCCAACGCGGGACTGACAATAGCAT
CTGATAAGATTCAAACCTCTCCTCCTTTCCATTACTTGGGAATGCAGGTAGAGGAAAGGAAAATTAAACCACAAAAAATAGAAATAA
GAAAAGACACATTAAAAACATTAAATGAGTTTCAAAAGTTGGTAGGAGATACTAATTGGATTCGGAGATATTAATTGGATTTGGCCA
ACTCTAGGCATTCCTACTTATGCCATGTCAATTTTGTTCTCTTTCTTAAGAGGGGACTTGGAATTAAATAGTGAAAGAATGTTACCT
CCAGAGGCAACTAAAGAAATTAAATTAATTGAAGAAAAAAATTCGGTCAGCACAAGTAAATAGGATCACTTGGCCCCACTCCAAATT
TTGATTTTTGGTACTGCACATTCTCTAACAGCCATCATTGTTCAAAACACAGATCTTGTGGATTGGTCCTTCCTTCCTCATAGTACA
ATTAAGACTTTTACATTGTACTTGGATCAAATGGCTACATTAATTGGTCAGGGAAGATTACGAATAATAACATTGTGTGGAAATGAC
CCAGATAAAATCACTGTTCCTTTCAACAAGCAACAAGTTAGACAAGCCTTTATCAGTTCTGGTGCATGGCAGATTGGTCTTGCTAAT
TTTCTGGGAATTATTGATAATCATTACCCAAAAACAAAAATCTTCCAGTTCTTAAAATTGACTACTTGGATTCTACCTAAAATTACC
AGACGTGAACCTTTAGAAAATGCTCTAACAGTATTTACTGATGGTTCCAGCAATGGAAAAGCGGCTTACACAGGGCCGAAAGAACGA
GTAATCAAAACTCCGTATCAATCAGCTCAAAGAGCAGAGTTGGTTGCAGTCATTACAGTGTTACAAGATTTTGACCAACCTATCAAT
ATTATATCAGATTCTGCATATGTAGTACAGGCTACAAGGGATGTTGAGACAGCTCTAATTAAATATAGCACGGACGATCATTTAAAC
CAGCTATTCAATTTATTACAACAAACTGTAAGAAAAAGAAATTTCCCATTTTATATTACTCATATTCGAGCACACACTAATTTACCA
GGGCCTTTGACTAAAGCAAATGAACAAGCTGACTTACTGGTATCATCTGCATTCATAAAAGCACAAGAACTTCTTGCTTTGACTCAT
GTAAATGCAGCAGGATTAAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACATTGCACCCAGTGTCAAGTC

TTACACCTGTCCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCGTTATGGCAAATGGATGGCACGCATGTTCCT
TCATTTGGAAGATTATCATATGTTCATGTAACAGTTGATACTTATTCACATTTCATATGGGCAACTTGCCAAACAGGAGAAAGTACT
TCCCATGTTAAAAAACATTTATTATCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCAAAACTGACAATGGACCAGGATATTGT
AGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAACAGGAATTCCTTATAATTCCCAAGGACAGGCCATAGTT
GAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGAAGGGGGAGACAGTAAGGAGTGTACCACTCCTCAGATGCAA
CTTAATCTAGCACTCTATACTTTAAATTTTTTAAACATTTATAGAAATCAGACTACTACTTCTGCAAAACAACATCTTACTGGTAAA
AAGCACAGCCCACATGAAGGAAAACTAATTTGGTGGAAAGATAATAAAAATAAGACATGGGAAATAGGGAAGGTGATAACGTGGGGG
AGAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAGCTTCCTGTTTGGATACCCACTAGACATTTGAAGTTCTACAATGAACCCATC
GGAGATGCAAAGAAAAGGGCCTCCACAGAGATGGTAACCCCAGTCACATGGATGGATAATC

SEQ ID 101:

GTCACATGGATGGATAATCCTATAGAAGTATATGTTAATGATAGTGTATGGGTACCTGGCCCCACAGATGATCGCTGCCCTGCCAAA
CCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGTGTATCGTTATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAATG
CCTGCAGTCCAAAATTGGTTGGTAGAAGTACCTACTGTCAGTCCTAACAGTAGATTCACTTATCACATGGTAAGCGGGATGTCACTC
AGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTAAAATTTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATT
CCCAAAGAATCAAAAAATACAGAAGTTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTACAAAACAATGAATTCGGA
ACTATTATAGATTGGGCACCTCGAGGTCAATTCTACCACAATTGCTCAGGACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCA
GCTGTTGATAGCGACTTAACAGAAAGTCTAGACAAACATAAGCATAAAAAATTACAGTCTTTCTACCCTTGGGAATGGGGAGAAAAA
GGAATCTCTACCCCAAGACCAGAAATAATAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTTGGCCTGACACCACA
TTAGAATTTGGTCTGGAAATCAAACTTTAGAAACAAGAGATCGTAAGCCATTTTATACTATCGACCTAAATTCCAGTCTAACGGTTC
CTTTACAAAGTTGCGTAAAGCCCTCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAAACTATAACCTGTGAAA
ATTGTAGATTGTTTACTTGCATTGATTCAACTTTTAATTGGCGGCACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCT
CTGTGTCCGTGGACTGACCGTGGGAGGCCTCGCCATCCATCCATATTTTGACTGAAGTATTAAAAGACATTTTAAATAGATCCAAAA
GATTCATTTTTACCTTAATTGCAGTGATTATGGGATTAATTGCAGTCACAGCTACGGCTGCTGTGGCAGGAGTTGCATTGCACTCTT
CTGTTCAGTCGGTAAACTTTGTTAATGATTGGCAAAAGAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATCAAAAATTGG
CAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGACTCATGAGCTTAGAACATTGTTTCCAGTTACAGTGTG
ACTGGAATACGTCAGATTTTTGTATTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAGG
GAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAATAACAAATTTTCGAAGCATCAAAAGCCCATTTAAATTTGATGCCAG
GAACTGAGGCAATTGCAGGAGTTGCTGATGGCCTCGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATCGGAAGTACTATGATTA
TAAATCTCATATTAATCCTTGTGTGCCTGTTTTGTCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATC
GAGAACGGGCCA

SEQ ID 102:

ATGGGGCAAACTAAAAGTAAAATTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGGAGTTAAA
GTATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAACTTCAGATCTAAAA
GATTGGAAAAGAATTGGTAAGGAACTAAAACAAGCAGGTAGGAAGGGTAATATCATTCCACTTACAGTATGGAATGATTGGGCCATT
ATTAAAGCAGCTTTAGAACCATTTCAAACAGAAGAAGATAGCATTTCAGTTTCTGATGCCCCTGGAAGCTGTTTAATAGATTGTAAT
GAAAACACAAGGAAAAAATCCCAGAAAGAAACCGAAAGTTTACATTGCGAATATGTAGCAGAGCCGGTAATGGCTCAGTCAACGCAA
AATGTTGACTATAATCAATTACAGGAGGTGATATATCCTGAAACGTTAAAATTAGAAGGAAAAGGTCCAGAATTAATGGGGCCATCA
GAGTCTAAACCACGAGGCACAAGTCCTCTTCCAGCAGGTCAGGTGCTCGTAAGATTACAACCTCAAAAGCAGGTTAAAGAAAATAAG
ACCCAACCGCAAGTAGCCTATCAATACTGCCGCTGGCTGAACTTCAGTATCGGCCACCCCCAGAAAGTCAGTATGGATATCCAGGAA
TGCCCCCAGCACCACAGGGCAGGGCGCCATACCATCAGCCGCCCACTAGGAGACTTAATCCTATGGCACCACCTAGTAGACAGGGTA
GTGAATTACATGAAATTATTGATAAATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAA

SEQ ID 103:

MGQTKSKIKSKYASYLSFIKILLKRGGVKVSTKNLIKLFQIIEQFCPWFPEQGTSDLKDWKRIGKELKQAGRKGNIIPLTVWNDWAI
IKAALEPFQTEEDSISVSDAPGSCLIDCNENTRKKSQKETESLHCEYVAEPVMAQSTQNVDYNQLQEVIYPETLKLEGKGPELMGPS
ESKPRGTSPLPAGQVLVRLQPQKQVKENKTQPQVAYQYCRWLNFSIGHPQKVSMDIQECPQHHRAGRHTISRPLGDLILWHHLVDRV
VNYMKLLINQERKEILRHGNSQ

SEQ ID 104:

MPPAPQGRAPYHQPPTRRLNPMAPPSRQGSELHEIIDKSRKEGDTEAWQFPVTLEPMPPGEGAQEGEPPTVEARYKSFSIKMLKDMK
EGVKQYGPNSPYMRTLLDSIAYGHRLIPYDWEILAKSSLSPSQFLQFKTWWIDGVQEQVRRNRAANPPVNIDADQLLGIGQNWSTIS
QQALMQNEAIEQVRAICLRAWEKIQDPGSTCPSFNTVRQGSKEPYPDFVARLQDVAQKSIADEKAGKVIVELMAYENANPECQSAIK
PLKGKVPAGSDVISEYVKACDGIGGAMHKAMLMAQAITGVVLGGQVRTFGGKCYNCGQIGHLKKNCPVLNKQNITIQATTTGREPPD
LCPRCKKGKHWASQCRSKFDKNGQPLSGNEQRGQPQAPQQTGAFPIQPFVPQGFQGQQPPLSQVFQGISQLPQYNNCPSPQAAVQQ

SEQ ID 105:

ATGGAGATTTTACATTGCTTAGGGCCAGATAATCAAGAAAGTACTGTTCAGCCAATGATTACTTCAATTCCTCTTAATCTGTGGGGT
CGAGATTTATTACAACAATGGGGTGCGGAAATCACCATGCCCGCTCCATTATATAGCCCCACGAGTCAAAAAATCATGACCAAGATG
GGATATATACCAGGAAAGGGACTAGGGAAAAATGAAGATGGCATTAAAGTTCCAGTTGAGGCTAAAATAAATCAAGAAAGAGAAGGA
ATAGGGTATCCTTTTTAG

SEQ ID 106:

MEILHCLGPDNQESTVQPMITSIPLNLWGRDLLQQWGAEITMPAPLYSPTSQKIMTKMGYIPGKGLGKNEDGIKVPVEAKINQEREG
IGYPF

SEQ ID 107:

ATGGGGCCTCTCCAACCCGGGTTGCCCTCTCCGGCCATGATCCCAAAAGATTGGCCTTTAATTATAATTGATCTAAAGGATTGCTTT
TTTACCATCCCTCTGGCAGAGCAGGATTGTGAAAAATTTGCCTTTACTATACCAGCCATAAATAATAAAGAACCAGCCACCAGGTTT
CAGTGGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATTTGTCAGACTTTTGTAGGTCGAGCTCTTCAACCAGTGAGAGAA
AAGTTTTCAGACTGTTATATTATTCATTATATTGATGATATTTTATGTGCTGCAGAAACGAAAGATAAATTAATTGACTGTTATACA
TTTCTGCAAGCAGAGGTTGCCAATGCTGGACTGGCAATAGCATCCGATAAGATCCAAACCTCTACTCCTTTTCATTATTTAGGGATG
CAGATAGAAAATAGAAAAATTAAGCCACAAAAAATAGAAATAAGAAAAGACACATTAAAAACACTAAATGATTTTCAAAAATTACTA

GGAGATATTAATTGGATTCGGCCAACTCTAGGCATTCCTACTTATGCCATGTCAAATTTGTTCTCTATCTTAAGAGGAGACTCAGAC
TTAAATAGTCAAAGAATATTAACCCCAGAGGCAACAAAAGAAATTAAATTAGTGGAAGAAAAAATTCAGTCAGCGCAAATAAATAGA
ATAGATCCCTTAGCCCCACTCCAACTTTTGATTTTTGCCACTGCACATTCTCCAACAGGCATCATTATTCAAAATACTGATCTTGTG
GAGTGGTCATTCCTTCCTCACAGTACAGTTAAGACTTTTACATTGTACTTGGATCAAATAGCTACATTAATCGGTCAGACAAGATTA
CGAATAACAAAATTATGTGGAAATGACCCAGACAAAATAGTTGTCCCTTTAACCAAGGAACAAGTTAGACAAGCCTTTATCAATTCT
GGTGCATGGCAGATTGGTCTTGCTAATTTTGTGGGACTTATTGATAATCATTACCCAAAAACAAAGATCTTCCAGTTCTTAAAATTG
ACTACTTGGATTCTACCTAAAATTACCAGACGTGAACCTTTAGAAAATGCTCTAACAGTATTTACTGATGGTTCCAGCAATGGAAAA
GCAGCTTACACAGGGCCGAAAGAACGAGTAATCAAAACTCCATATCAATCGGCTCAAAGAGACGAGTTGGTTGCAGTCATTACAGTG
TTACAAGATTTTGACCAACCTATCAATATTATATCAGATTCTGCATATGTAGTACAGGCTACAAGGGATGTTGAGACAGCTCTAATT
AAATATAGCATGGATGATCAGTTAAACCAGCTATTCAATTTATTACAACAAACTGTAAGAAAAAGAAATTTCCCATTTTATATTACT
TATATTCGAGCACACACTAATTTACCAGGGCCTTTGACTAAAGCAAATGAACAAGCTGACTTACTGGTATCATCTGCACTCATAAAA
GCACAAGAACTTCATGCTTTGACTCATGTAAATGCAGCAGGATTAAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATT
GTACAACATTGCACCCAGTGTCAAGTCTTACACCTGCCCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCATTA
TGGCAAATGGATGTCACGCATGTACCTTCATTTGGAAGATTATCATATGTTCATGTAACAGTTGATACTTATTCACATTTCATATGG
GCAACTTGCCAAACAGGAGAAAGTACTTCCCATGTTAAAAAACATTTATTGTCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATC
AAAACTGACAATGGACCAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAACAGGAATTCCT
TATAATTCCCAAGGACAGGCCATAGTTGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGAAGGGGGAGACAGT
AAGGAGTGTACCACTCCTCAGATGCAACTTAATCTAGCACTCTATACTTTAAATTTTTTAAACATTTATAGAAATCAGACTACTACT
TCTGCAGAACAACATCTTACTGGTAAAAAGAACAGCCCACATGAAGGAAAACTAATTTGGTGGAAAGATAATAAAAATAAGACATGG
GAAATAGGGAAGGTGATAACGTGGGGGAGAGGTTTTGCTTGTGTTTCACCAGGAGAAAATCAGCTTCCTGTTTGGTTACCCACTAGA
CATTTGAAGTTCTACAATGAACCCATCGGAGATGCAAAGAAAAGGGCCTCCACGGAGATGGTAACACCAGTCACATGGATGGATAAT
CCTATAGAAGTATATGTTAATGATAGTATATGGGTACCTGGCCCCATAGATGATCGCTGCCCTGCCAAACCTGAGGAAGAAGGGATG
ATGATAAATATTTCCATTGGGTATCGTTATCCTCCTATTTGCCTAGGGAGAGCACCAGGATGTTTAATGCCTGCAGTCCAAAATTGG
TTGGTAGAAGTACCTACTGTCAGTCCCATCAGTAGATTCACTTATCACATGGTAAGCGGGATGTCACTCAGGCCACGGGTAAATTAT
TTACAAGACTTTTCTTATCAAAGATCATTAAAATTTAGACCTAAAGGGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAAT
ACAGAAGTTTTAGTTTGGGAAGAATGTGTGGCCAATAGTGCGGTGATATTATAAAACAATGAATTTGGAACTATTATAGATTGGGCA
CCTCGAGGTCAATTCTACCACAATTGCTCAGGACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTA
ACAGAAAGTTTAGACAAACATAAGCATAAAAAATTGCAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATCTCTACCCCAAGA
CCAAAAATAGTAAGTCCTGTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTACTGTGGCCTCACACCACATTAGAATTTGGTCT
GGAAATCAAACTTTAGAAACAAGAGATTGTAAGCCATTTTATACTGTCGACCTAAATTCCAGTCTAACAGTTCCTTTACAAAGTTGC
GTAAAGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAGACTATAACCTGTGAAAATTGTAGATTGCTT
ACTTGCATTGATTCAACTTTTAATTGGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGTGTCCATGGAC
CGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAAAGGTGTTTTAAATAGATCCAAAAGATTCATTTTTACT
TTAATTGCAGTGATTATGGGATTAATTGCAGTCACAGCTACGGCTGCTGTAGCAGGAGTTGCATTGCACTCTTCTGTTCAGTCAGTA
AACTTTGTTAATGATTGGCAAAAGAATTCTACAAGATTGTGGAATTCACAATCTAGTATTGATCAAAAATTGGCAAATCAAATTAAT
GATCTTAGACAAACTGTCATTTGGATGGGAGACAGACTCATGAGCTTAGAACATCGTTTCCAGTTACAATGTGACTGGAATACGTCA
GATTTTTGTATTACACCCCAAATTTATAATGAGTCTGAGCATCACTGGGACATGGTTAGACGCCATCTACAGGGAAGAGAAGATAAT
CTCACTTTAGACATTTCCAAATTAAAAGAACAAATTTTCGAAGCATCAAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAATT
GCAGGAGTTGCTGATGGCCTCGCAAATCTTAACCCTGTCACTTGGGTTAAGACCATTGGAAGTACATCGATTATAAATCTCATATTA
ATCCTTGTGTGCCTGTTTTGTCTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCATG
ATGACGATGGCGGTTTTGTCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAAATTGTTACTGTGTCTGTGTAG

SEQ ID 108:

MGPLQPGLPSPAMIPKDWPLIIIDLKDCFFTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGMLNSPTICQTFVGRALQPVRE
KFSDCYIIHYIDDILCAAETKDKLIDCYTFLQAEVANAGLAIASDKIQTSTPFHYLGMQIENRKIKPQKIEIRKDTLKTLNDFQKLL
GDINWIRPTLGIPTYAMSNLFSILRGDSDLNSQRILTPEATKEIKLVEEKIQSAQINRIDPLAPLQLLIFATAHSPTGIIIQNTDLV
EWSFLPHSTVKTFTLYLDQIATLIGQTRLRITKLCGNDPDKIVVPLTKEQVRQAFINSGAWQIGLANFVGLIDNHYPKTKIFQFLKL
TTWILPKITRREPLENALTVFTDGSSNGKAAYTGPKERVIKTPYQSAQRDELVAVITVLQDFDQPINIISDSAYVVQATRDVETALI
KYSMDDQLNQLFNLLQQTVRKRNFPFYITYIRAHTNLPGPLTKANEQADLLVSSALIKAQELHALTHVNAAGLKNKFDVTWKQAKDI
VQHCTQCQVLHLPTQEAGVNPRGLCPNALWQMDVTHVPSFGRLSYVHVTVDTYSHFIWATCQTGESTSHVKKHLLSCFAVMGVPEKI
KTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQGQAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLNFLNIYRNQTTT
SAEQHLTGKKNSPHEGKLIWWKDNKNKTWEIGKVITWGRGFACVSPGENQLPVWLPTRHLKFYNEPIGDAKKRASTEMVTPVTWMDN
PIEVYVNDSIWVPGPIDDRCPAKPEEEGMMINISIGYRYPPICLGRAPGCLMPAVQNWLVEVPTVSPISRFTYHMVSGMSLRPRVNY
LQDFSYQRSLKFRPKGKPCPKEIPKESKNTEVLVWEECVANSAVILXNNEFGTIIDWAPRGQFYHNCSGQTQSCPSAQVSPAVDSDL
TESLDKHKHKKLQSFYPWEWGEKGISTPRPKIVSPVSGPEHPELWRLTVASHHIRIWSGNQTLETRDCKPFYTVDLNSSLTVPLQSC
VKPPYMLVVGNIVIKPDSQTITCENCRLLTCIDSTFNWQHRILLVRAREGVWIPVSMDRPWEASPSVHILTEVLKGVLNRSKRFIFT
LIAVIMGLIAVTATAAVAGVALHSSVQSVNFVNDWQKNSTRLWNSQSSIDQKLANQINDLRQTVIWMGDRLMSLEHRFQLQCDWNTS
DFCITPQIYNESEHHWDMVRRHLQGREDNLTLDISKLKEQIFEASKAHLNLVPGTEAIAGVADGLANLNPVTWVKTIGSTSIINLIL
ILVCLFCLLLVCRCTQQLRRDSDHRERAMMTMAVLSKRKGGNVGKSKRDQIVTVSV

SEQ ID 109:

MNPSEMQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKLPSTKKAGPPTWAQLKKLTQLATKYLENTKVTQTPESMLLAALMIVSMV
SAGVPNSSEETATIENGP

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:

MGQTKSKIKSKYASYLSFIKILLKRGGVKVSTKNLIKLFQIIEQFCPWFPEQGTLDLKDWKRIGKELKQAGRKGNIIPLTVWNDWAI
IKAALEPFQTEEDSVSVSDAPGSCIIDCNENTGKKSQKETEGLHCEYVAEPVMAQSTQNVDYNQLQEVIYPETLKLEGKGPELVGPS
ESKPRGTSPLPAGQVPVTLQPQKQVKENKTQPPVAYQYWPPAELQYRPPPESQYGYPGMPPAPQGRAPYPQPPTRRLNPTAPPSRQG
SKLHEIIDKSRKEGDTEAWQFPVTLEPMPPGEGAQEGEPPTVEARYKSFSIKKLKDMKEGVKQYGPNSPYMRTLLDSIAHGHRLIPY
DWEIQAKSSLSPSQFLQFKTWWIDGVQEQVRRNRAANPPVNIDADQLLGIGQNWSTISQQALMQNEAIEQVRAICLRAWEKIQDPGS
TCPSFNTVRQGSKEPYPDFVARLQDVAQKSIADEKARKVIVELMAYENANPECQSAIKPLKGKVPAGSDVISEYVKACDGIGGAMHK
AMLMAQAITGVVLGGQVRTFGRKCYNCGQIGHLKKNCPVLNKQNITIQATTTGREPPDLCPRCKKGKHWASQCRSKFDKNGQPLSGN
EQRGQPQAPQQTGAFPIQPFVPQGFQGQQPPLSQVFQGISQLPQYNNCPPPQAAVQQ

SEQ ID 147:

WATIVGKRAKGPASGPTTNWGIPNSAICSSGFSGTTTPTVPSVSGNKPVTTIQQLSPATSGSAAVDLCTIQAVSLLPGEPPQKTPTG
VYGPLPKGTVGLILGRSSLNLKGVQIHTSVVDSDYKGEIQLVISSSIPWSASPRDRIAQLLLLPYIKGGNSEIKRIGGLGSTDPTGK
AAYWASQVSENRPVCKAIIQGKQFEGLVDTGADVSIIALNQWPKNWPKQKAVTGLVGIGTASEVYQSTEILHCLGPDNQESTVQPMI
TSIPLNLWGRDLLQQWGAEITMPAPSYSPTSQKIMTKMGYIPGKGLGKNEDGIKIPVEAKINQEREGIGNPC

SEQ ID 148:

NKSRKRRNRESLLGAATVEPPKPIPLTWKTEKPVWVNQWPLPKQKLEALHLLANEQLEKGHIEPSFSPWNSPVFVIQKKSGKWRMLT
DLRAVNAVIQPMGPLQPGLPSPAMIPKDWPLIIIDLKDCFFTIPLAEQDCEKFAFTIPAINNKEPATRFQWKVLPQGMLNSPTICQT
FVGRALQPVREKFSDCYIIHCIDDILCAAETKDKLIDCYTFLQAEVANAGLAIASDKIQTSTPFHYLGMQIENRKIKPQKIEIRKDT
LKTLNDFQKLLGDINWIRPTLGIPTYAMSNLFSILRGDSDLNSKRMLTPEATKEIKLVEEKIQSAQINRIDPLAPLQLLIFATAHSP
TGIIIQNTDLVEWSFLPHSTVKTFTLYLDQIATLIGQTRLRIIKLCGNDPDKIVVPLTKEQVRQAFINSGAWKIGLANFVGIIDNHY
PKTKIFQFLKLTTWILPKITRREPLENALTVFTDGSSNGKAAYTGPKERVIKTPYQSAQRAELVAVITVLQDFDQPINIISDSAYVV
QATRDVETALIKYSMDDQLNQLFNLLQQTVRKRNFPFYITHIRAHTNLPGPLTKANEQADLLVSSALIKAQELHALTHVNAAGLKNK
FDVTWKQAKDIVQHCTQCQVLHLPTQEAGVNPRGLCPNALWQMDVTHVPSFGRLSYVHVTVDTYSHFIWATCQTGESTSHVKKHLLS
CFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQGQAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNLALYTLN
FLNIYRNQTTTSAEQHLTGKKNSPHEGKLIWWKDNKNKTWEIGKVITWGRGFACVSPGENQLPVWIPTRHLKFYNEPIRDAKKSTSA
ETETSQSSTVDSQDEQNGDVRRTDEVAIHQEGRAANLGTTKEADAVSYKISREHKGDTNPREYAACSLDDCINGGKSPYACRSSCS

SEQ ID 149:

MNPSEMQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKLPSTKKAEPPTWAQLKKLTQLATKYLENTKVTQTPESMLLAALMIVSMV
VSLPMPAGAAAANYTYWAYVPFPPLIRAVTWMDNPTEVYVNDSVWVPGPIDDRCPAKPEEEGMMINISIGYHYPPICLGRAPGCLMP
AVQNWLVEVPTVSPICRFTYHMVSGMSLRPRVNYLQDFSYQRSLKFRPKGKPCPKEIPKESKNTEVLVWEECVANSAVILQNNEFGT
IIDWAPRGQFYHNCSGQTQSCPSAQVSPAVDSDLTESLDKHKHKKLQSFYPWEWGEKGISTPRPKIVSPVSGPEHPELWRLTVASHH
IRIWSGNQTLETRDRKPFYTIDLNSSLTVPLQSCVKPPYMLVVGNIVIKPDSQTITCENCRLLTCIDSTFNWQHRILLVRAREGVWI
PVSMDRPWEASPSVHILTEVLKGVLNRSKRFIFTLIAVIMGLIAVTATAAVAGVALHSSVQSVNFVNDWQKNSTRLWNSQSSIDQKL
ANQINDLRQTVIWMGDRLMSLEHRFQLQCDWNTSDFCITPQIYNESEHHWDMVRRHLQGREDNLTLDISKLKEQIFEASKAHLNLVP
GTEAIAGVADGLANLNPVTWVKTIGSTTIINLILILVCLFCLLLVCRCTQQLRRDSDHRERAMMTMAVLSKRKGGNVGKSKRDQIVT
VSV

SEQ ID 150:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTATGTACTAA
GAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCCACTCAG
GGTTAAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCC
CTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAACGTTTCTC
CCCATGTGATAGCCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGA
GGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAA
TGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGC
AATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAA
AGACCTTTGTTCACATGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCA
CAGATGATCAGTAAATACTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTCTTTCT
TTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAAC
CCACCCCTACA

SEQ ID 151:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTCACTGTATCTGTGTAGAAAGAAGTAGACATGGGAGACTCCATTTTGTTATGTACTAA
GAAAAATTCTTCTGCCTTGAGATTCTGTGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAAACTCAGGGTTAAA
TGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCCCTAATCT
CAAGTACCCAGGGACACAAACACTGCGGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGT
GATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGAT
TAGTAAAAGAGGAAGGCATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCCGTCCCTGGGCAATGGAATGTCTCG
GTATAAAACCGGATTGTACGTTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTG
CTTTTTAAAGCATTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACCTTGTCTATGATGCAAAGATCTT
TGTTCACGTGTTTGTCTGCTGACCCTCTCCCCACTATTGTCTTGTGACCCTGACACATCCCCCTCTCGGAGAAACACCCACGAATGA
CCAATAAATACTAAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGCCCCCTTATTTCTTTCTCT
ACACTTTGTCTCTGTGTCTTTTTCTTTCCTAAGTCTCTCGTTCCACCTTACGAGAAACACCCACAGGTGTGGAGGGGCAACCCACCC
CTACA

SEQ ID 152:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATGGGAGACTCCATTTTGTTATGTGCTAA
GAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTGTGTCAACTCAG
GGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCC
CTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTC
CCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCATCCCCCAGCCCGACACCCATAAAGGGTCTGTGCTGA
GGAGGATTAGTATAAGAGGAAGGCATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAA
TGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGC
AATACTGCCTTGTAAAGCATTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAA
AGACCTTTGTTCACGTGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACACCCA
CAGATGATCAATAAATACTAAGGGAACTCAGAGGCTGGCGGATCCTCCATATGCTGAACGCTGGTTCCCCGGTTCCCCTTATTTCT
TTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAAC
CCACCCCTACA

SEQ ID 153:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACAGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTCTGTACTAA
GAAAAATTCTTCTGCCTTGAAATTCTGTTAATCTATAACCTTACCCCCAACCCCGTGCTCTTTGAAACATGTGCTGTGTCAACTCAG
AGTTAAATGGATTAAGTGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTCCTTGAGAGTCATCACCACTCC
CTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCTCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTC
CCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCATCCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGA
GGAGGATTAGTAAAAGAGGAAGGAACGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAA
TGTCCCGGTATAAAACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGC
AATACTGCTTTGTAAAGCATTGAGCTGTTTATGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAA
AGACCTTTGTTCACGTGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCA
CGAATGATGAATAAATACTAAGGGAACTCAGAGGCTGGCGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTACTTCT
TTCTCTGTACTTTGTCTCTGTGTCTTTTTCTTTCCTAAGTCTCTCGTTCCACCTTACGAGAAATACCCACAGGTGTGGAGGGGCAAC
CCACCCCTACA

SEQ ID 154:

TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTCTGTACTAA
GAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATAACCTTACCCCCAACCCCGTGCTCTCTGAAACATGTGCTATGTCAACTCAG
AGTTGAATGGATTAAGGGCGGTGCAAGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCACGCTCCTTAAGAGTCATCACCACTCC
CTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTC
CCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCATCCCCCAGCCCGACACCTGTAAAGGGTCTGTGCTGA
GGAGGATTAGTATAAGAGGAAGGCATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCCGTCCCTGGGCAATGGAA
TGTCTCGGTATAAAACCCGATTGTATGTTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGC
AATACTGCTTTGTAAAGCATTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACCTTGTCTATGATGCAA
AGACCTTTGTTCACGTGTTTGTCTGCTGACCCTCTCCCCACGATTGTCTTGTGACCCTGACACATCCCCGTCTTCGAGAAACACCCA
CGAATGATCAATAAATACTAAGGGAACTCAGAGGCTGGCGGATCCTCCATATGCTGAACGCTGGTTCCCCAGGTCCCCTTATTTCT
TTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAGTCTCTCGTTCCATCTTACGAGAAACACCCACAGGTGTGGAGGGGCAAC
CCACCCCTACA

SEQ ID 155:

GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTATGTGTAT
GCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCAC

SEQ ID 156:

ATGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTCGAGAAACACCCACAGATGATCAGTAA
ATACTAAGGGAACTCAGAGGCTGGCGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTCTTTCTTTCTCTATACTTTG
TCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCACCCCTACA

SEQ ID 157:

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXACATTTGAAGTTCTACAATGAACC
CATCXGAGATGCAAAGAAAXXXXXXXXXXXXAGCXCCTCCXCGGAGACGGAAACACCGCAATCGAGCAXCXXXXXXXXXXXXXXTXGAC
TCACAAGATGAAXAAAATGGTGAXXTCAGAAGAACAGATGAAGTTGCCATCCACCAAGAAXGCXGAGCCGCCGACTTGGGCACAAXT
AAAGAAGCTGACACAGTTAGCTAXAAAAXXXXXXCTXGAGAACACAAAGGTGACACAAACTCCAGAGAXTATGCTGCTTGCAGCTTT
GATGATTGTATCAATGGTGGTAAGTCTCCCXATGCCTGCAGGAGCAGCTGCAGCTAAXTATACXTACTGGGCCTATGTGCCTTTCCC
GCCCTTAATTCGGGCAGTCACATGGATGGATAATCCTATTGAAGTATATGTTAATAATAGTGTATGGGXTACCTGGCCCCACAGATG
ATCGTTGCCCTGCCAAACCTGAGGAAGAAGGAATGATGATAAATATTTCCATTGGGTATCXTTATCCTCCTATTTGCCTAGGGAGAG
CACCAGGATGTTTAATXGCCTGCAXTCCAAAATTGGTTGGTAGAAGTACCTACTGTCAGTXCCAXCAGTAGATTCACTTATCACATG
GTAAGXGGXATGTCACTCAGGCCACXGGTAAATXATTTACAXGACTTTTCTTATCAAAGATCATTAAAATTTAGXCCTAAAGGGAAA
CCTTGCCCCAAGGAAATTCCCAAAGXATCAAAAXAXXCAGAAGTTTTAGTTTGGGAAGAATGTGTGGCXAATAGTGCXGTGATATTA
CAAAACAATGAATTTGGAACTATTATAGATTGGCACCTCGAGGTCAATTCTAXCACAXXXXXXXXXXXXXXXXXXXXXXXXXATTGC
XCAGGXCAAACTCAXTCXTGTCCXAGXGCACAAGXXXXXXXXXXXXXXAGTCCAGCTGTTGATAGXGACTTAACAGAAAGTXTAGAC
XAAXXTXAXXXTAXAAXXTTAXAXTCXXTCTAXCCXTGGXAATGGGGXGAAAAXGGAATXTCXXCXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCCXXGACCAAAXXTAXTXAGTCCTGTTX
CTGGTCCTGAACATCCAGAATTATGGAXGCTTACTGTGGCCTCAXXACCACATTAGAATTTGGTCTGGAAATCAAXCTXTAGAAACA
AGAGATCXTAAGCCATXTTATACTATCXACCTAAATTCCAGTCTXACAXTTCCTTTXCAAAGTTGXGTAAAGCCCCCTTATATXGCT
AGTTGTAGGAAATAXXTAGTTATTAAACCAGAXTCCCAAACTATAXXACCTGTGAAAATTGTAGATTGTTTACTTGCATTGATTCAA
CTTTTAATTGGCAGCACCGTATTCTGCTXGTGAGAGCAAGAGAXGGXGTGTGGATCCCTGTGTCCATGGACCGACCGTGGGAGGCXT
CXCCATCCXTCCATATTTTXACXGAAGTATTAAAAGGXXTTXTAAXTAGATCCAAAAGATTCATTTTTACTTTAATTGCAGTGATTA

TGGGXXTXATTGCAGTCACAGCTACXGCTGCXGXXGCXGGAXTTGCXTTXCACTCXTCTGTTCAXXCXGXAXAXTXTGTXAATXATT
GGCAAAAXAAXTTCXXCAAXATTGTGGAATTCXCAXAXCXXXXATXGATCAAAAATTGGCAAATCAAATTAATGATCTTAGACAAAC
TGTCATTTGGATGGGAGAXAGXCTCATGAGCTTXGAAXATCXTTTXCAGTTACAXTGTGACTGGAATACGTCAGATTTTTGTATTAC
ACCXCAAXXXTATAATGAGTCTGAGCATCACTGGGACATGGTTAGAXGCCATCTACAXGGAAGAGAAGATAATCTXACTTTAGACAT
TTCXAAATTAAAAGAAXXXXXXXXXXXXXXCAAATTTTXXAAXCATCAAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAATX
GXXXXAGXTGCTGATGGCCTCXCAAATCTTAACCCTGTCACTTGGGTTAAXACCATXXGAAGTXCXACXATTXTAAATXTCATATTA
ATCCTTGTXTGCCTGTTXTGTCTGTTGTTXXAGTCTXCAGGTGTAXCCAXCAGCTCCGAAGAGACAGCGACCAXCXAGAACGGGCCA
TGATGACGATGGXGGTTTTGTCXAAAAGAAAAGGGGGXXAXATGTXGGGAAAAGXXAGAGAGATCAGAXTGTTACTGTXGTCTXTGT
AGAAAXAXGXAGACATAXGAGACTCCATTTTGXXXTGTACXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXX

SEQ ID 158:

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCPWFPEQGXLDLXDWKRIGXELKQAGRKGNXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXDAPGSCIIDCNEXTXKKSQKETEXLHCEYVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXAGQVXVTLQPQXQVKENKTQXPVAYQYWPPXXXXXXXXXXXSQYGYXGMPPAXQXRXPYPQPPTXRXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCXSKFDKXG
QPLSGNXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

SEQ ID 159:

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXDXLAPLQLLIF
ATAHSXTGIIIQNTDLVEWSFLPHSTVKTFTLYLDQMATLIGQXRLRIIXLCGNDPDKIXVPXXKXQVRQAFIXSGAWXIGLANFLG
IIDNHYPKTKIFQFLKLTTWILPKITRREPLENALTVFTDGSSNGKAAYTGPKERVIKTPYQSAQRAELVAVITVLQDFDQPINIIS
DSAYVVQATRDVETALIKYSXDDXLNQLFNLLQQTVRKRNFPFYITHIRAHTNLPGPLTKANEQADLLVSSAXIKAQELXALTHVNA
AGLKNKFDVTWKQAKDIVQHCTQCQVLHLXTQEAGVNPRGLCPNALWQMDXTHVXSFGRLSYVHVTVDTYSHFIWATCQTGESTSHV
KKHLLSCFAVMGVPEKIKTDNGPGYCSKAFQKFLSQWKISHTTGIPYNSQGQAIVERTNRTLKTQLVKQKEGGDSKECTTPQMQLNL
ALYTLNFLNIYRNQTTTSAXQHLTGKKXSPHEGKLIWWKDXKNKTWEIGKVITWGRGFACVSPGENQLPVWIPTRHLKFYNEPIXDA
KKXXSXEXXTXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXTIXXXXXXXXXXXXXXXX
XXXXXXXXXXXAXXXDXXXXXXKXXXXXXXXXXXPXEWGXXXIXXXXXXXXXSPXSXXXXXXXXXXXXXXXXXXXXXX

SEQ ID 160:

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXLIXXVTWMDNPXEVYVNDSVWVPGPXDDX
CPAKPEEEGMMINISIXYXYPPICLGRAPGCLMPAVQNWLVEVPTVSPXXRFTYHMVSGMSLRPRVNXLQDFSYQRSLKFRPKGKPC
PKEIPKESKNTEVLVWEECVANSXVILQNNEFGTIIDWAPRGQFYHNCSGQTQSCXSAQVSPAVDSDLTESLDKHKHKKLQSFYPWE
WGEKGISTPRPXIISPVSGPEHPELWXXXXXXXXXXRIXXXXXXXXXXXXXXXXXXXXLNSXLTVPLQSCVKPXXXXXXXXXXXXXX
XXXXXXXXXXXXXDSTXXWXXXIXLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXRCTQQLR
RDSDXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

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 method for diagnosing 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, wherein the patient sample contains prostate cells and/or wherein the patient is suspected of having prostate cancer wherein up-regulation of expression of at least 150% relative to a negative control is indicative of prostate cancer.

2. The method of claim 1, wherein the expression product is a RNA or a polypeptide.

3. The method of any preceding claim, wherein the patient sample is a prostate sample or a blood sample.

4. The method of any preceding claim, wherein the expression product is a RNA having the following 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; N_2 has at least 75% sequence identity to SEQ ID 156; N_3 has at least 75% sequence identity to SEQ ID 6; N_4 comprises any RNA sequence; N_5 has at least 75% sequence identity to SEQ ID 5; and at least one of N_1 or N_5 is present, but N_2, N_3, N_4 and polyA are optional.

5. The method of claim 4, wherein the RNA comprises N_1.

6. The method of claim 5, wherein N_1 is at the 5' end of the RNA.

7. The method of claim 4, wherein N_4 comprises a polypeptide-coding sequence.

8. The method of claim 2, wherein the polypeptide is encoded by a mRNA having the following formula: N_1-N_2-N_3-N_4-N_5-polyA, as defined in claim 4.

9. The method of claim 8, wherein the mRNA encodes one or more of the following HML-2 polypeptides: gag, prt, pol, env, cORF, tat.

10. The method of claim 8 or claim 9, wherein the polypeptide is detected using an antibody.

11. The method of any one of claims 1 to 6, wherein said step is preceded by a step of enriching RNA in the patient sample.

12. The method of any one of claims 1 to 6 or 11, wherein the expression product is detected using PCR, SDA, SSSR, LCR, TMA or NASBA.

13. The method of claim 12, wherein the PCR is RT-PCR.

Patentansprüche

1. Verfahren zum Diagnostizieren von Prostatakrebs, wobei das Verfahren den Schritt des Nachweisens des Vorhandenseins oder Fehlens eines Expressionsprodukts eines endogenen HML-2-Retrovirus in einer Patientenprobe umfasst, wobei die Patientenprobe Prostatazellen enthält und/oder wobei der Verdacht besteht, dass der Patient Prostatakrebs aufweist, wobei eine Hochregulation der Expression von wenigstens 150 % bezogen auf eine Negativkontrolle indikativ für Prostatakrebs ist.

2. Verfahren gemäß Anspruch 1, wobei das Expressionsprodukt eine RNA oder ein Polypeptid ist.

3. Verfahren gemäß einem der vorstehenden Ansprüche, wobei die Patientenprobe eine Prostataprobe oder eine Blutprobe ist.

4. Verfahren gemäß einem der vorstehenden Ansprüche, wobei das Expressionsprodukt eine RNA mit der folgenden Formel ist: N_1-N_2-N_3-N_4-N_5-polyA, wobei: N_1 wenigstens 75 % Sequenzidentität mit SEQ ID 155 aufweist; N_2 wenigstens 75 % Sequenzidentität mit SEQ ID 156 aufweist; N_3 wenigstens 75 % Sequenzidentität mit SEQ ID 6 aufweist; N_4 eine beliebige RNA-Sequenz umfasst; N_5 wenigstens 75 % Sequenzidentität mit SEQ ID 5 aufweist; und N_1 und/oder N_5 vorhanden sind, während N_2, N_3, N_4 und polyA optional sind.

5. Verfahren gemäß Anspruch 4, wobei die RNA N_1 umfasst.

6. Verfahren gemäß Anspruch 5, wobei N_1 an dem 5'-Ende der RNA angeordnet ist.

7. Verfahren gemäß Anspruch 4, wobei N_4 eine Polypeptid-codierende Sequenz umfasst.

8. Verfahren gemäß Anspruch 2, wobei das Polypeptid von einer mRNA, die folgende Formel gemäß Anspruch 4 aufweist : N_1-N_2-N_3-N_4-N_5-polyA, codiert wird.

9. Verfahren gemäß Anspruch 8, wobei die mRNA ein oder mehrere der folgenden HML-2-Polypeptide codiert: gag, prt, pol, env, cORF, tat.

10. Verfahren gemäß Anspruch 8 oder Anspruch 9, wobei das Polypeptid unter Verwendung eines Antikörpers nachgewiesen wird.

11. Verfahren gemäß einem der Ansprüche 1 bis 6, wobei dem Schritt ein Schritt des Anreicherns von RNA in der Patientenprobe vorausgeht.

12. Verfahren gemäß einem der Ansprüche 1 bis 6 oder 11, wobei das Expressionsprodukt unter Verwendung von PCR, SDA, SSSR, LCR, TMA oder NASBA nachgewiesen wird.

13. Verfahren gemäß Anspruch 12, wobei die PCR RT-PCR ist.

Revendications

1. Méthode de diagnostic du cancer de la prostate, la méthode comprenant l'étape consistant à détecter la présence ou l'absence d'un produit d'expression d'un rétrovirus endogène HML-2 dans un prélèvement d'un patient, le prélèvement d'un patient contenant des cellules de la prostate et/ou le patient étant suspecté d'avoir le cancer de la prostate, la régulation positive de l'expression d'au moins 150 % par rapport à un témoin négatif étant indicatrice du cancer de la prostate.

2. Méthode selon la revendication 1, **caractérisée en ce que** le produit d'expression est un ARN ou un polypeptide.

3. Méthode selon l'une ou l'autre des revendications 1 et 2, **caractérisée en ce que** le prélèvement d'un patient est un prélèvement de la prostate ou un prélèvement de sang.

4. Méthode selon l'une quelconque des revendications précédentes, **caractérisée en ce que** le produit d'expression est un ARN ayant la formule suivante : N_1-N_2-N_3-N_4-N_5-polyA, dans laquelle : N_1 a au moins 75 % d'identité de séquence avec SEQ ID 155 ; N_2 a au moins 75 % d'identité de séquence avec SEQ ID 156 ; N_3 a au moins 75 % d'identité de séquence avec SEQ ID 6 ; N_4 comprend une séquence ARN quelconque ; N_5 a au moins 75 % d'identité de séquence avec SEQ ID 5 ; et au moins l'un de N_1 ou N_5 est présent, mais N_2, N_3, N_4 et polyA sont facultatifs.

5. Méthode selon la revendication 4, **caractérisée en ce que** l'ARN comprend N_1.

6. Méthode selon la revendication 5, **caractérisée en ce que** N_1 est à l'extrémité 5' de l'ARN.

7. Méthode selon la revendication 4, **caractérisée en ce que** N_4 comprend une séquence codant pour un polypeptide.

8. Méthode selon la revendication 2, **caractérisée en ce que** le polypeptide est codé par un ARNm ayant la formule suivante : N_1-N_2-N_3-N_4-N_5-polyA, telle que définie dans la revendication 4.

9. Méthode selon la revendication 8, **caractérisée en ce que** l'ARNm code pour un ou plusieurs des polypeptides de HML-2 suivants : gag, prt, pol, env, cORF, tat.

10. Méthode selon la revendication 8 ou la revendication 9, **caractérisée en ce que** le polypeptide est détecté à l'aide d'un anticorps.

11. Méthode selon l'une quelconque des revendications 1 à 6, **caractérisée en ce que** ladite étape est précédée d'une étape d'enrichissement de l'ARN dans le prélèvement d'un patient.

12. Méthode selon l'une quelconque des revendications 1 à 6 ou 11, **caractérisée en ce que** le produit d'expression est détecté par PCR, SDA, SSSR, LCR, TMA ou NASBA.

13. Méthode selon la revendication 12, **caractérisée en ce que** la PCR est la RT-PCR.

FIGURE 1

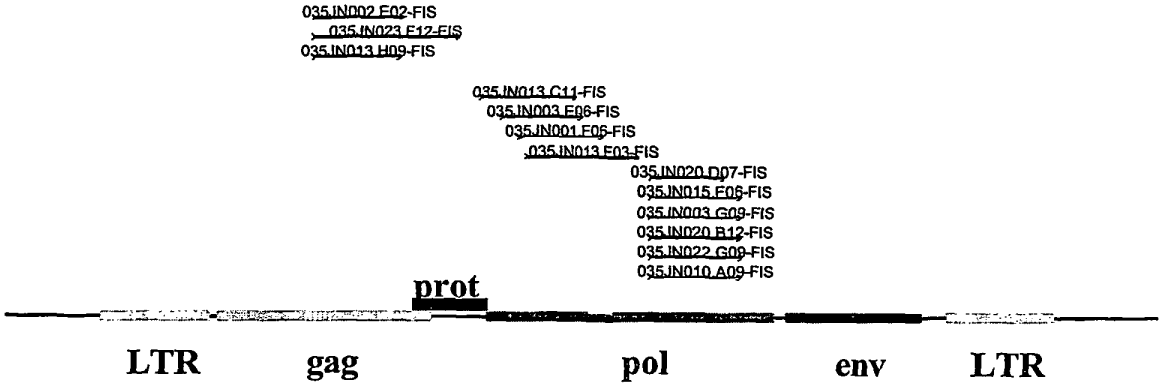

FIGURE 2

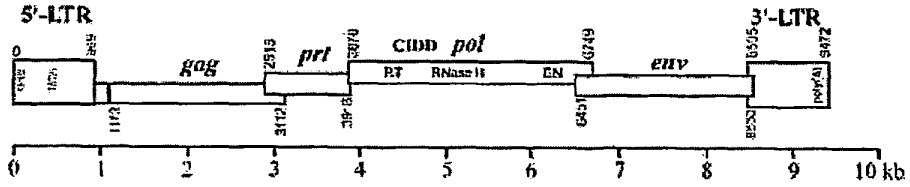

FIGURE 3

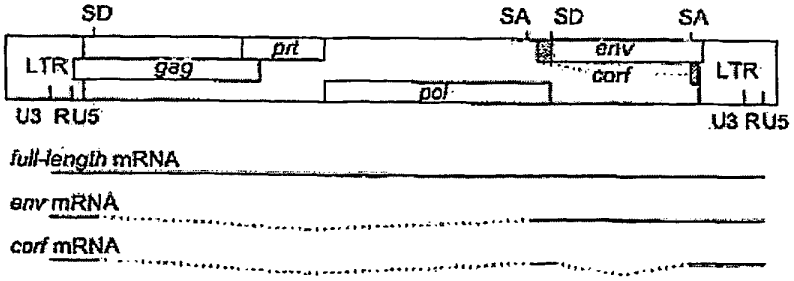

FIGURE 4

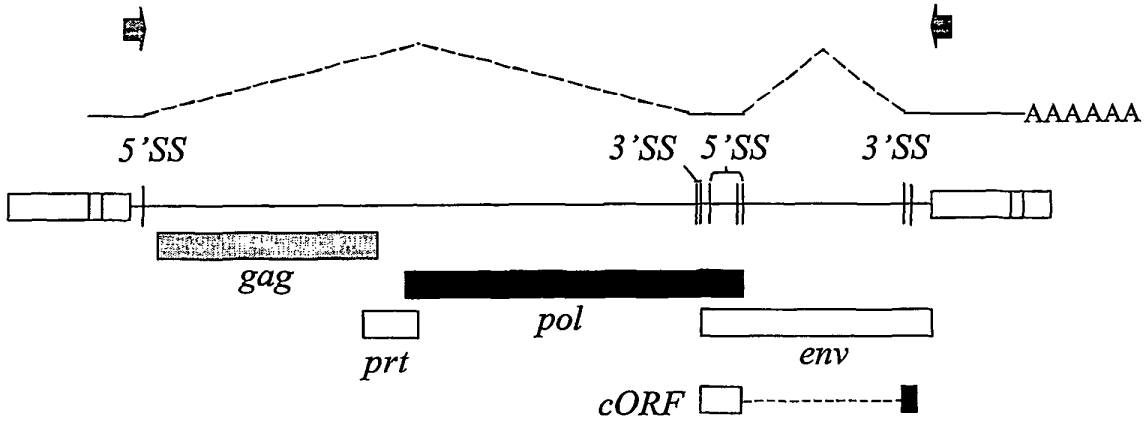

FIGURE 5

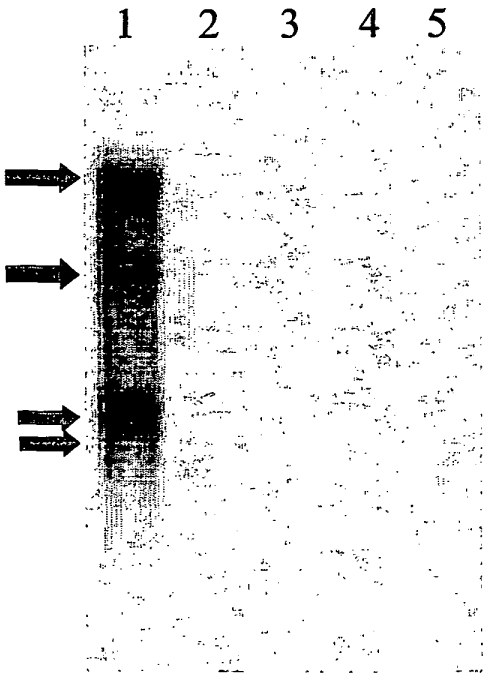

FIGURE 6

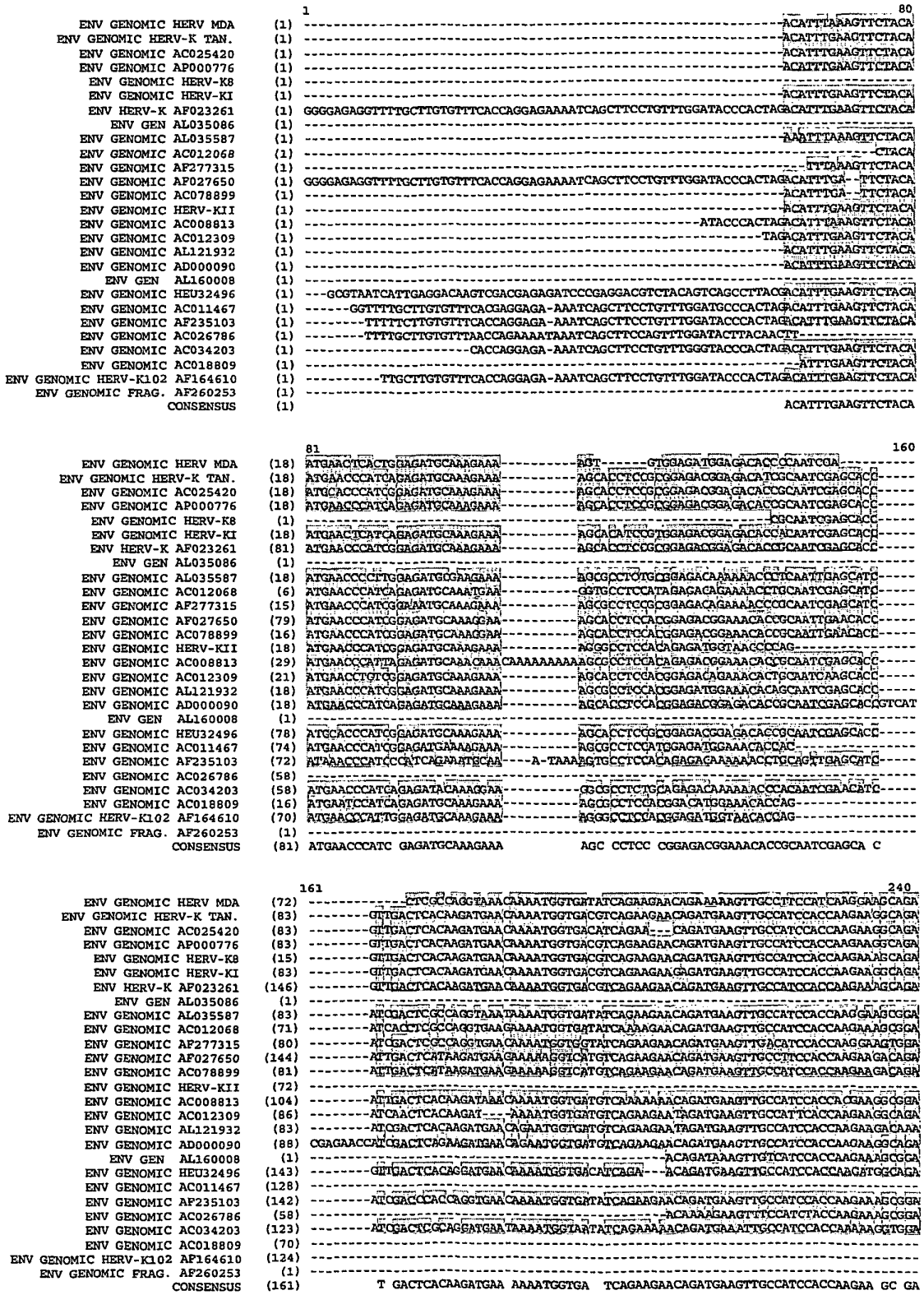
ENV GENOMIC HERV MDA
ENV GENOMIC HERV-K TAN.
ENV GENOMIC AC025420
ENV GENOMIC AP000776
ENV GENOMIC HERV-K8
ENV GENOMIC HERV-KI
ENV HERV-K AF023261
ENV GEN AL035086
ENV GENOMIC AL035587
ENV GENOMIC AC012068
ENV GENOMIC AF277315
ENV GENOMIC AF027650
ENV GENOMIC AC078899
ENV GENOMIC HERV-KII
ENV GENOMIC 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

FIGURE 6 CONTD...

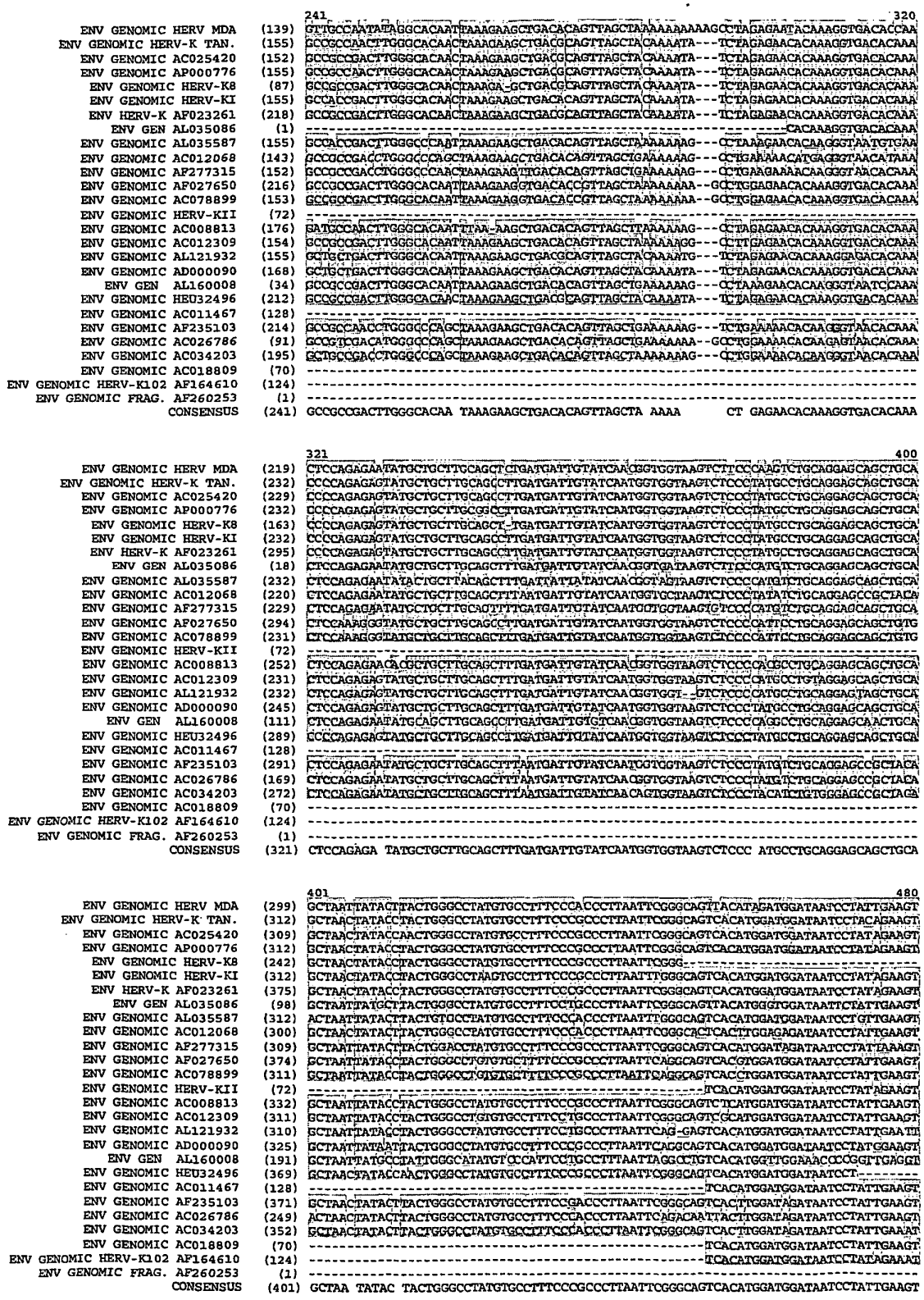

FIGURE 6 CONTD...

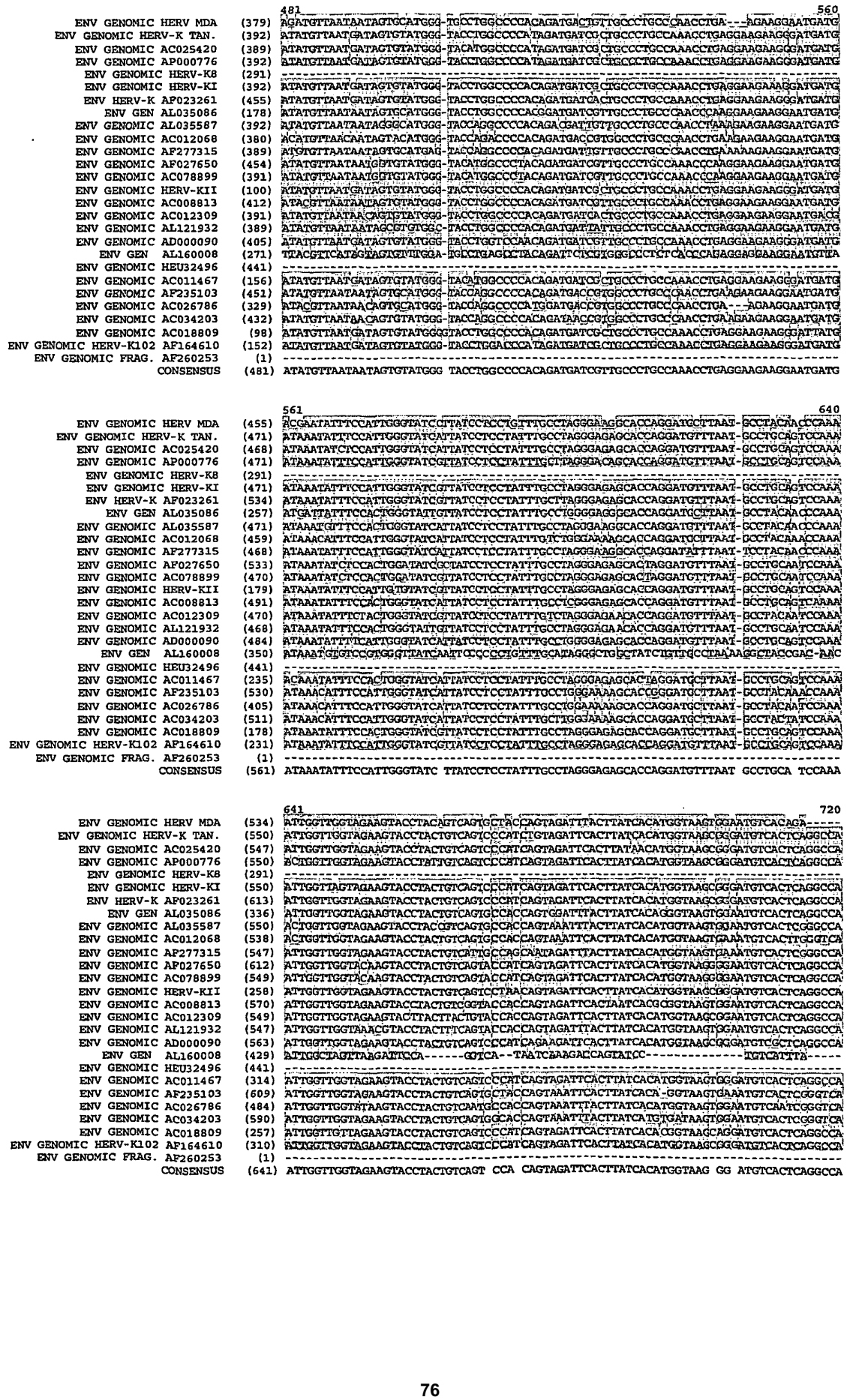

FIGURE 6 CONTD...

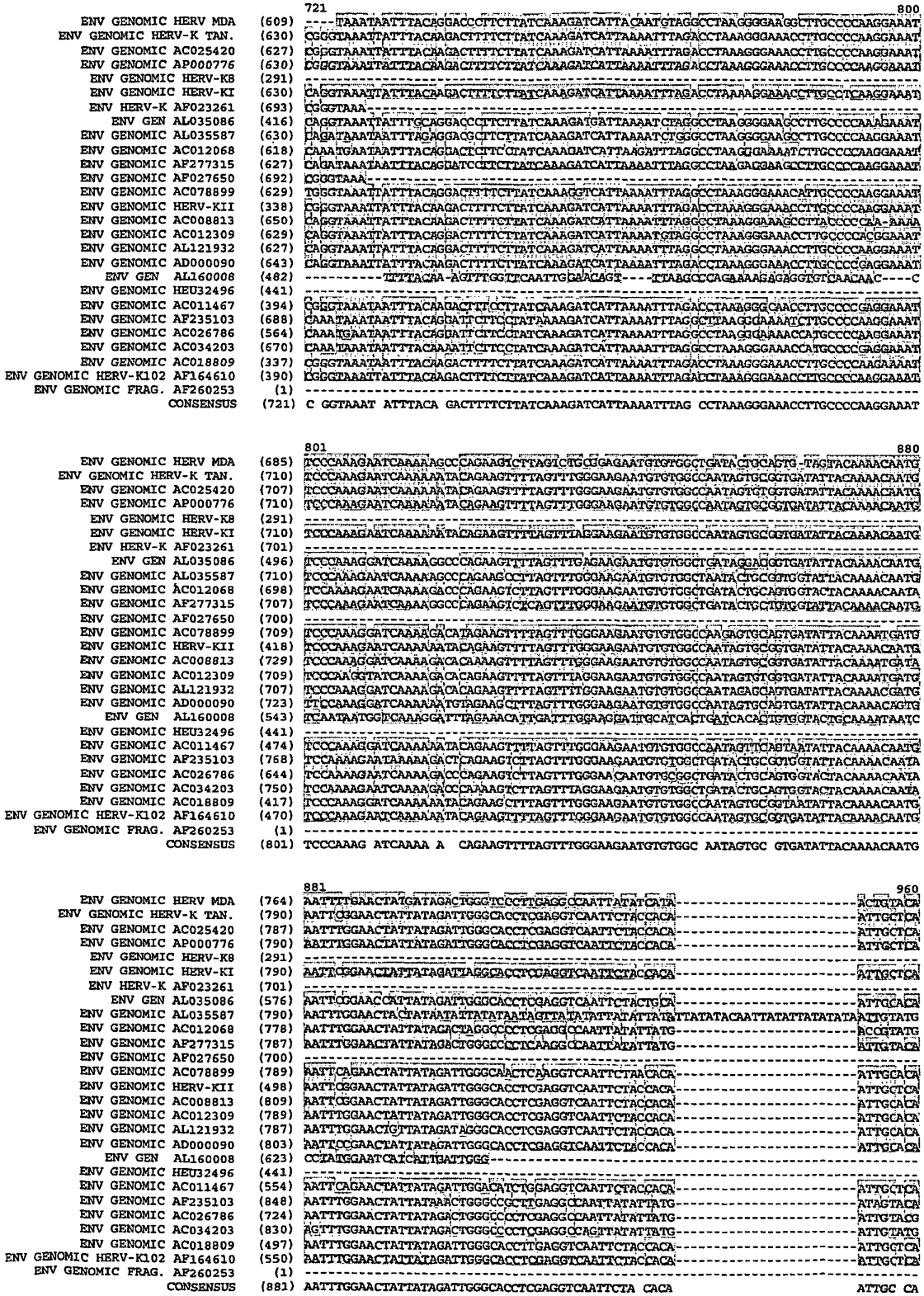

FIGURE 6 CONTD...

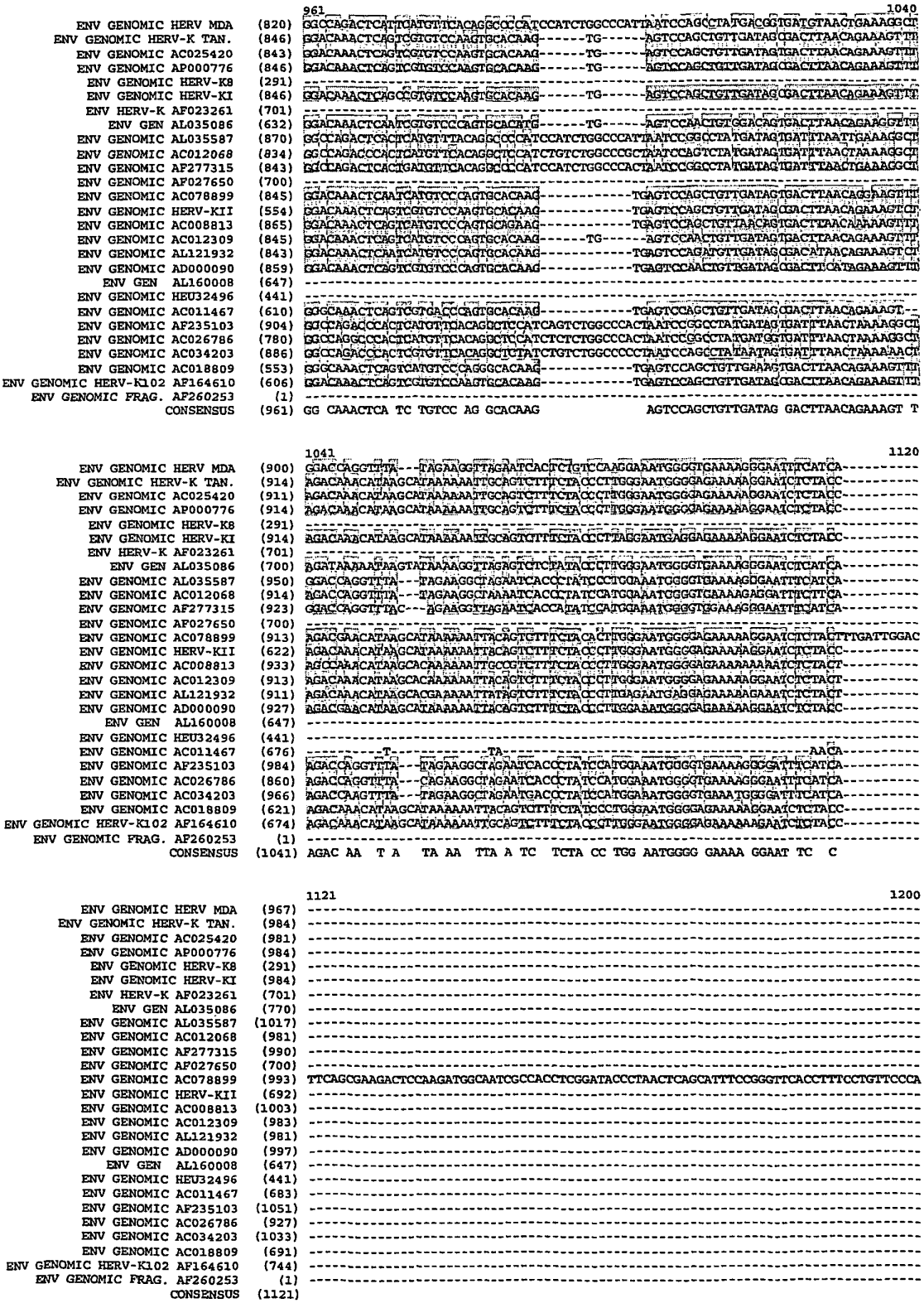

FIGURE 6 CONTD...

```
                                             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...

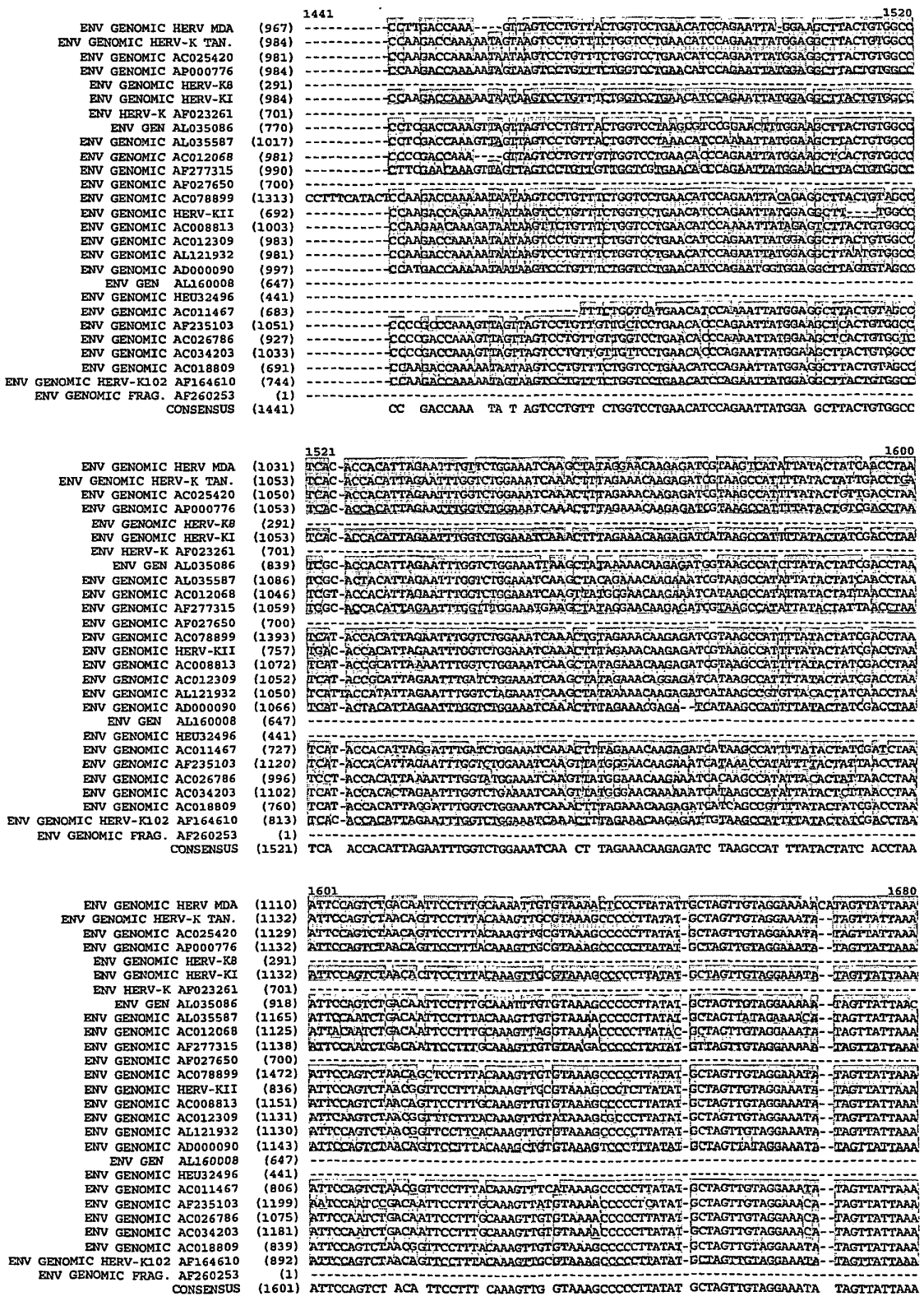

FIGURE 6 CONTD...

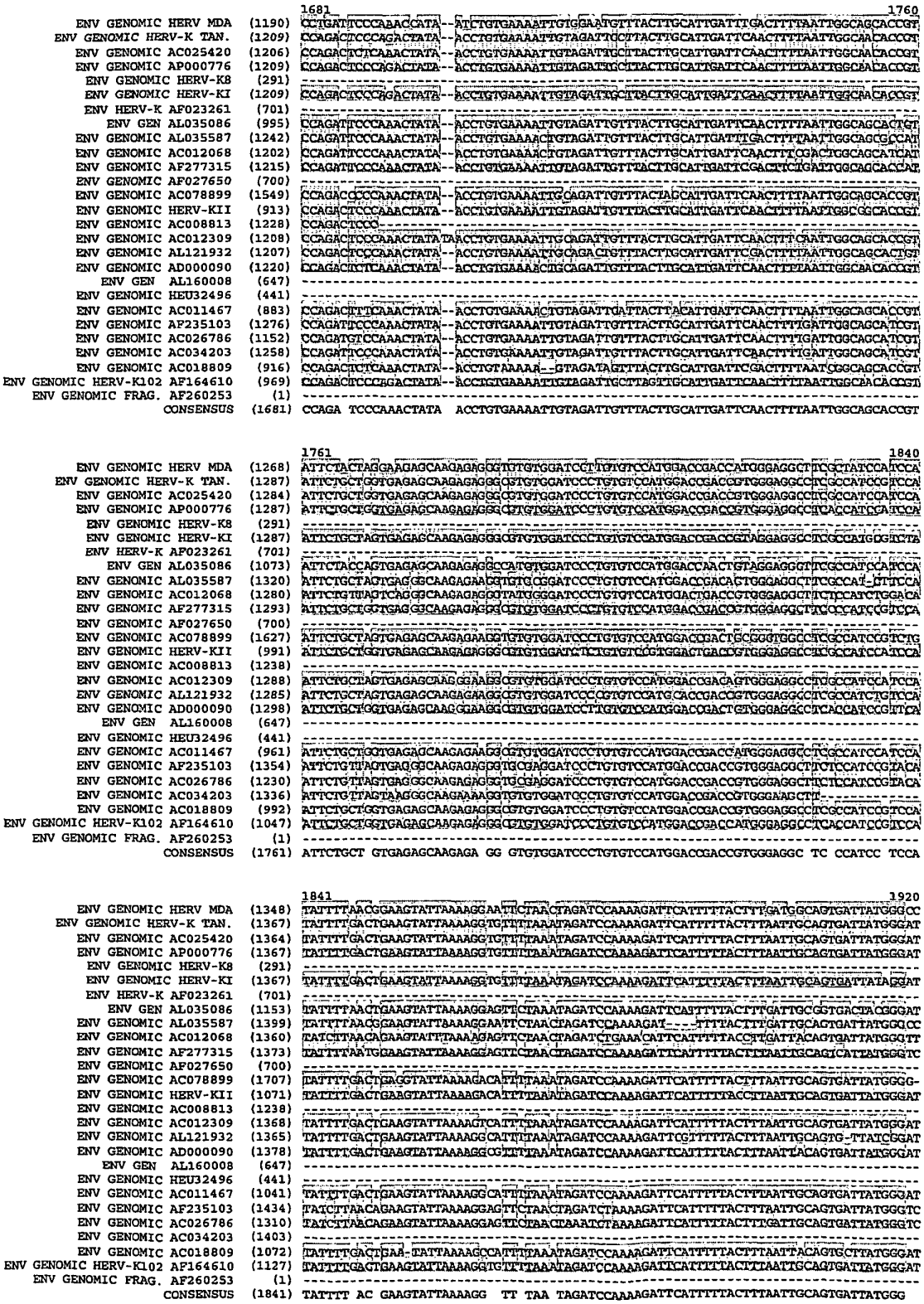

FIGURE 6 CONTD...

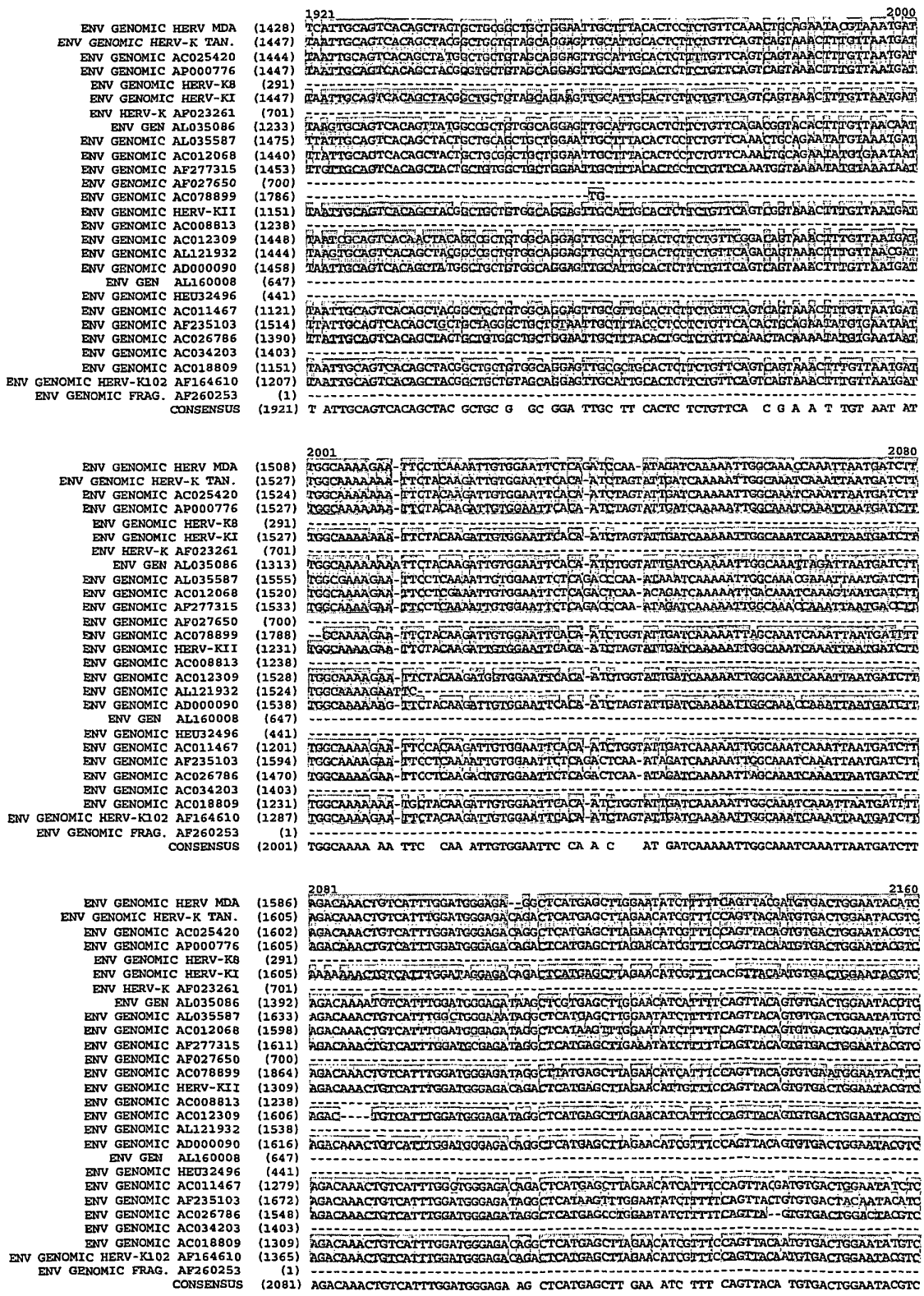

FIGURE 6 CONTD...

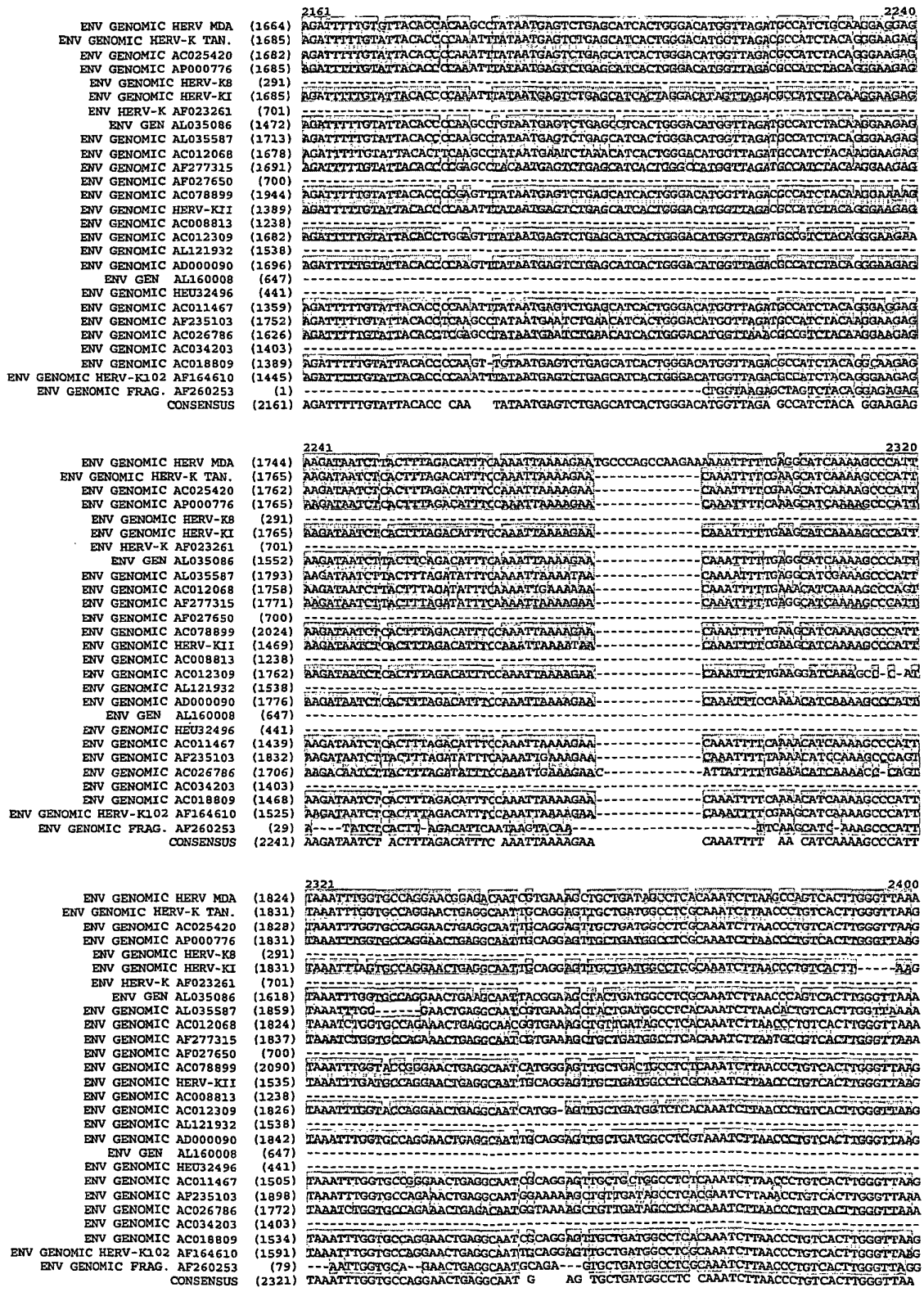

FIGURE 6 CONTD...

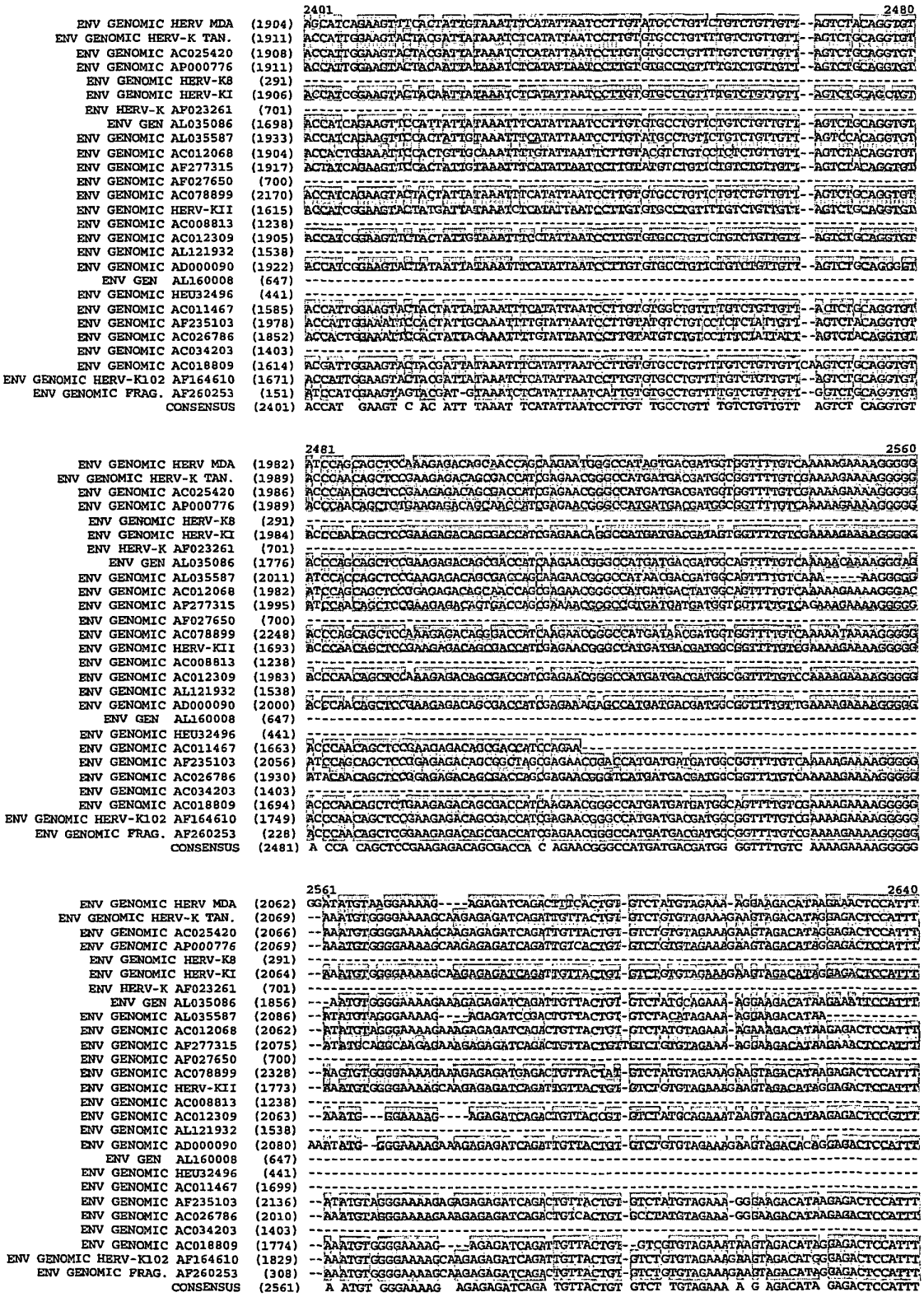

FIGURE 6 CONTD...

```
                                          2641                                                            2707
          ENV GENOMIC HERV MDA   (2136)  TGATCTGTACTAA------------------------------------------------------
      ENV GENOMIC HERV-K TAN.    (2146)  TGTTATGTACTAA------------------------------------------------------
          ENV GENOMIC AC025420   (2143)  TGTTATGTACTAA------------------------------------------------------
          ENV GENOMIC AP000776   (2146)  TGTTATGTACTAA------------------------------------------------------
           ENV GENOMIC HERV-K8    (291)  -------------------------------------------------------------------
           ENV GENOMIC HERV-KI   (2141)  TGTTATGTACTAA------------------------------------------------------
           ENV HERV-K AF023261    (701)  -------------------------------------------------------------------
              ENV GEN AL035086   (1931)  TGATCTGTACTAA------------------------------------------------------
          ENV GENOMIC AL035587   (2146)  -------------------------------------------------------------------
          ENV GENOMIC AC012068   (2138)  TGAAAAAGACCTGTACTTTGAACAATT----------------------------------------
          ENV GENOMIC AF277315   (2152)  TGACCTGTA----------------------------------------------------------
          ENV GENOMIC AF027650    (700)  -------------------------------------------------------------------
          ENV GENOMIC AC078899   (2405)  TGTTCTGTACTAAGAGAAATTCTTCTGCCTTGAGATGCTGTTAA-----------------------
          ENV GENOMIC HERV-KII   (1850)  TGTTCTGTACTAA------------------------------------------------------
          ENV GENOMIC AC008813   (1238)  -------------------------------------------------------------------
          ENV GENOMIC AC012309   (2133)  TGTTCTGTACCAAG-----------------------------------------------------
          ENV GENOMIC AL121932   (1538)  -------------------------------------------------------------------
          ENV GENOMIC AD000090   (2157)  TGTTCTGTACTAA------------------------------------------------------
             ENV GEN  AL160008    (647)  -------------------------------------------------------------------
          ENV GENOMIC HEU32496    (441)  -------------------------------------------------------------------
          ENV GENOMIC AC011467   (1699)  -------------------------------------------------------------------
          ENV GENOMIC AF235103   (2212)  TGAAAAAGACCTGTACTTTGAACAATTGCTTTGCTCAGATGTTGTTAATTTGTAGTTTT--------
          ENV GENOMIC AC026786   (2086)  TGAAAAAGACCTGTACTTTAAACAATTGCTTTGCTGAGATGTTGTTAATTTGTAGCTTTCCCCAGCC
          ENV GENOMIC AC034203   (1403)  -------------------------------------------------------------------
          ENV GENOMIC AC018809   (1846)  TGCTCTGTACTAAG-----------------------------------------------------
ENV GENOMIC HERV-K102 AF164610   (1906)  TGTTATGTGTTAAGAAAAATTCTT-------------------------------------------
    ENV GENOMIC FRAG. AF260253    (385)  TGTTCTGTACTAA------------------------------------------------------
                     CONSENSUS   (2641)  TG   TGTAC
```

FIGURE 7

```
                                     1                                                          60
  GI_4185938_EMB_CAA76878.1_   (1)  -----MGQTKSKIKSKYASYLSFIKILLKRGGVKVSTKNLIKLFQIIEQFCPWFPEQGTL
  GI_4185942_EMB_CAA76881.1_   (1)  -----MGQTKSKIKSKYASYLSFIKILLKRGGVKVSTKNLIKLFQIIEQFCPWFPEQGTL
  GI_4185946_EMB_CAA76884.1_   (1)  -----MGQTKSKIKSKYASYLSFIKILLKRGGVKVSTKNLIKLFQIIEQFCPWFPEQGTL
  GI_5931704_EMB_CAB56602.1_   (1)  -------MGQTKTKSKYASYLSFIKILLKRGGVRVSTKNLIKLFQTTEQFCPWFPEQGNL
             GAG OF AB047240   (1)  -----MGQTKSKTKSKYASYLSFIKILLKRGGVRVSTKNLIKLFQIIEQFCPWFPEQGTL
        TRANSLATION OF ORF99   (1)  YKKAGLGQTKSKTKSKYASYLSFIKILLKRGGVRVSTKNLIKLFQIIEQFCPWFPEQGTL
TRANSLATION OF G226TOP-LINK    (1)  ------------------------------------------------------------
TRANSLATION OF G591TOP-LINK    (1)  ------------------------------------------------------------
   TRANSLATION OF LNCAP-GAG    (1)  -----MGQTKSKTKSKYASYLSFIKILLKRGGVRVSTKNLIKLFQIIEQFCPWFPEQGTL
                  GAG106-135   (1)  ------------------------------------------------------------
                  GAG186-215   (1)  ------------------------------------------------------------
                    GAG46-75   (1)  --------------------------------------------------CPWFPEQGTL
                      PDG-G1   (1)  ------------------------------------------------------------
                      PGD-G2   (1)  ------------------------------------------------------------
                      PGD-G3   (1)  ------------------------------------------------------------
                   CONSENSUS   (1)                                                    CPWFPEQG L

                                     61                                                        120
  GI_4185938_EMB_CAA76878.1_  (56)  DLKDWKRIGKELKQAGRKGNIIPLTVWNDWAIIKAALEPFQTEEDSVSVSDAPGSCTIDC
  GI_4185942_EMB_CAA76881.1_  (56)  DLKDWKRIGKELKQAGRKGNIIPLTVWNDWAIIKAALEPFQTEEDSVSVSDAPGSCIIDC
  GI_4185946_EMB_CAA76884.1_  (56)  DLKDWKRIGKELKQAGRKGNIIPLTVWNDWAIIKAALEPFQTEEDSVSVSDAPGSCIIDC
  GI_5931704_EMB_CAB56602.1_  (54)  DLEDWKRIGKELKQAGRKGNIIPLTVWNDWPIIKAALEPFQTEDS-VSVSDAPGSCIIDC
             GAG OF AB047240  (56)  DLKDWKRIGEELKQAGRKGNIIPLTVWNDWAIIKAALEPFQTKEDSVSVSDAPGSCVIDC
        TRANSLATION OF ORF99  (61)  DLKDWKRIGEELKQAGRKGNIIPLTVWNDWAIIKAALEPFQTKEDSVSVSDAPGSCVIDC
TRANSLATION OF G226TOP-LINK    (1)  ------------------------------------------------------------
TRANSLATION OF G591TOP-LINK    (1)  ------------------------------------------------------------
   TRANSLATION OF LNCAP-GAG   (56)  DLKDWKRIGEELKQAGRKGNIIPLTVWNDWAIIKAALEPFQTKEDSVSVSDAPGSCVIDC
                  GAG106-135   (1)  --------------------------------------------------DAPGSCIIDC
                  GAG186-215   (1)  ------------------------------------------------------------
                    GAG46-75  (11)  DLKDWKRIGKELKQAGRKGN----------------------------------------
                      PDG-G1   (1)  ---DWKRIGKELKQAGRKG-----------------------------------------
                      PGD-G2   (1)  ------------------------------------------------------------
                      PGD-G3   (1)  ------------------------------------------------------------
                   CONSENSUS  (61)  DL DWKRIG ELKQAGRKGN                              DAPGSCIIDC
```

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) NEKTGRKSQKETESLHCEYVTEPVMAQSTQNVDYNQLQGVIYPETLKLEGKGPELVGPSE
     TRANSLATION OF ORF99  (121) NEKTGRKSQKETESLHCEYVTEPVMAQSTQNVDYNQLQGVIYPETLKLEGKGPELVGPSE
TRANSLATION OF G226TOP-LINK    (1) ------------------------------------------------------------
TRANSLATION OF G591TOP-LINK    (1) ------------------------------------------------------------
 TRANSLATION OF LNCAP-GAG  (116) NEKTGRKSQKETESLHCEYVTEPVMAQSTQNVDYNQLQGVIYPETLKLEGKGPELVGPSE
              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) SKPRGPSPLSAGQVTVTLQPQAQVRENKTQLPVAYQYWPPAELQYRPPPESQYGYLGMPP
         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) CRSKFDKNGQPLSGNEQRGQPQAPQQTGAFPIQPFVPHGFQGQQP-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) CHSKFDKDGQPLSGNRKRGQPQAPQQTGAFPVQLFVPQGFQGQQPLQKIPPLQGVSQLQQ
                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...

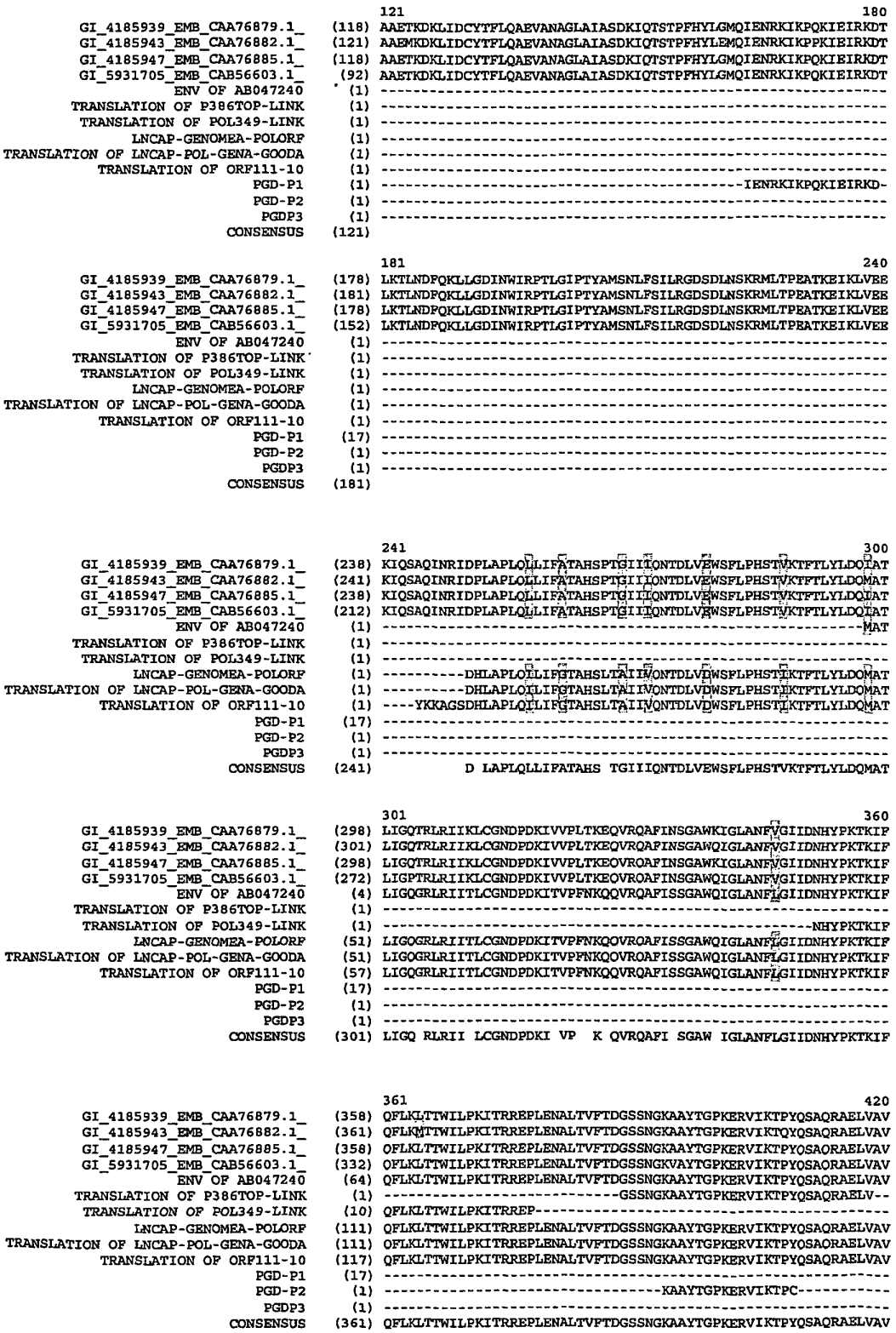

FIGURE 8 CONTD...

```
                                               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) THIRAHTNLPGPLTKANEQADLLVSSAFIKAQELHALTHVNAAGLKNKFDVTWKQAKDIV
                     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...

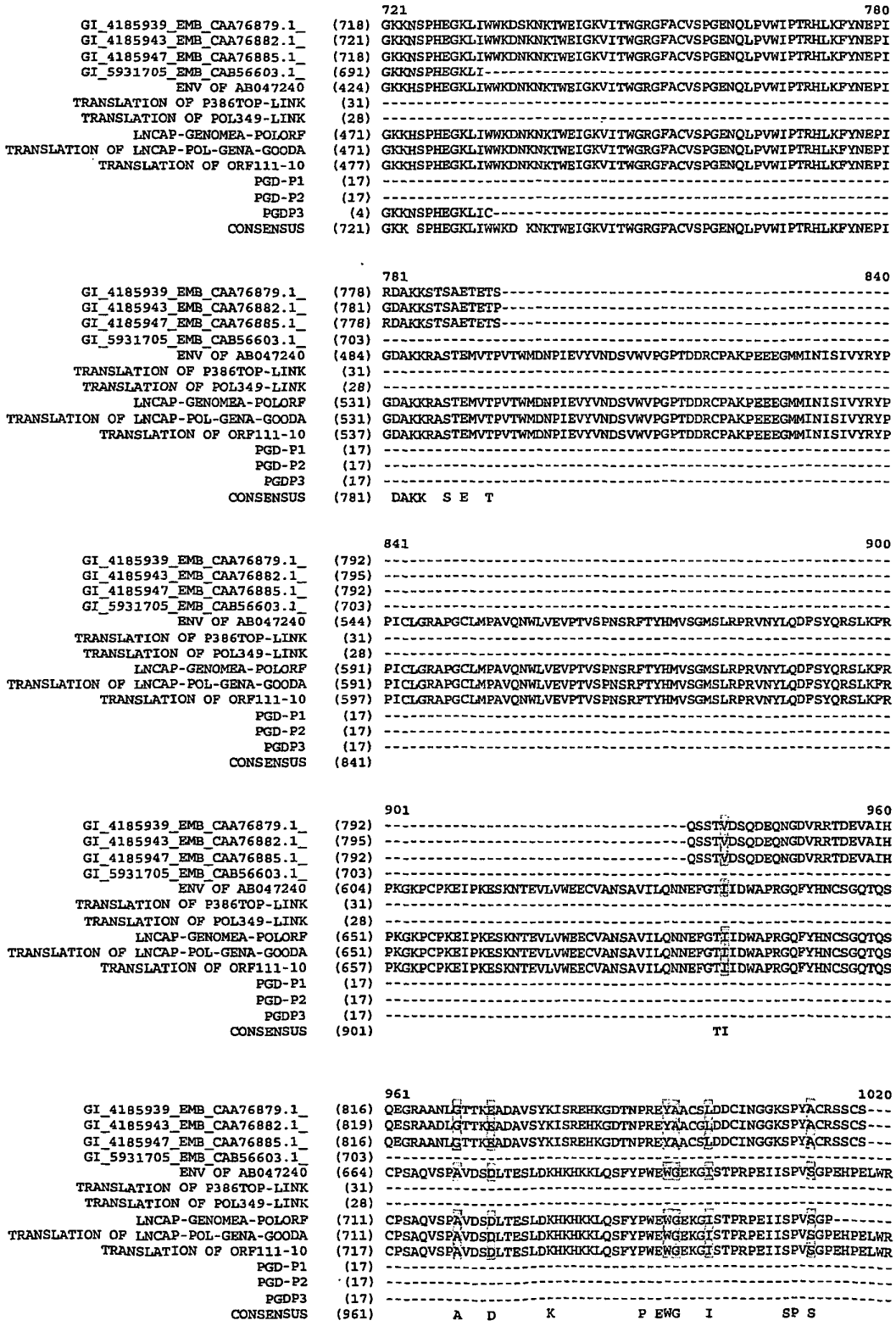

FIGURE 8 CONTD...

```
                                          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...

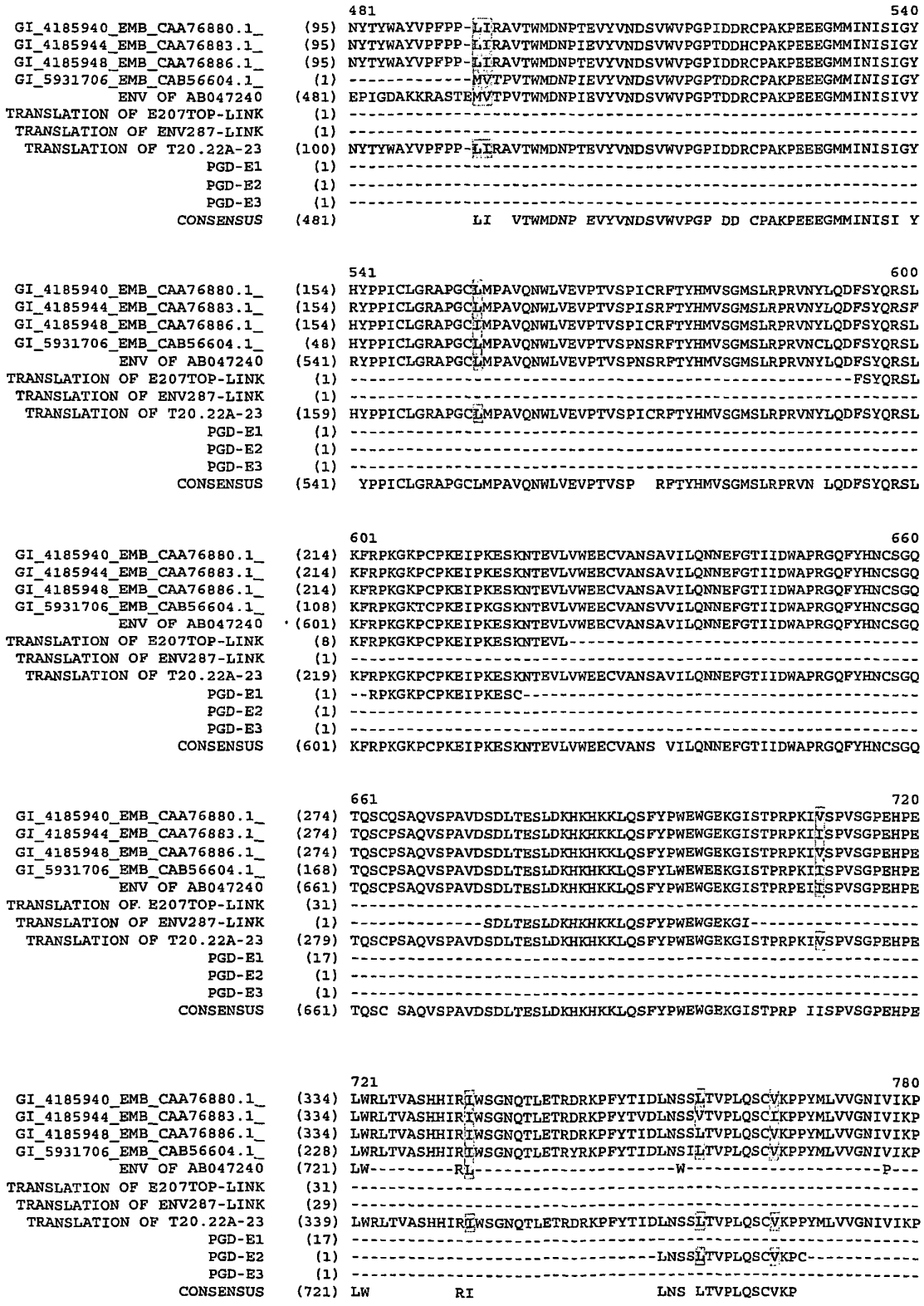

FIGURE 9 CONTD...

```
                                     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) ----------------DTTLEFGLEIKL--------------------------------
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) -----------------RCTQQLRRDSDHERA-----------------------------
                   CONSENSUS  (1021)                  RCTQQLRRDSD
```

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 **[0251]**
- US 5707829 A **[0252]**
- EP 0509612 B **[0252]**
- EP 0505012 B **[0252]**
- US 4959314 A **[0252]**
- US 5530101 A **[0252]**
- US 5585089 A **[0252]**
- WO 9824893 A **[0252]**
- WO 9110741 A **[0252]**
- WO 9630498 A **[0252]**
- WO 9402602 A **[0252]**
- US 5939598 A **[0252]**
- WO 9633735 A **[0252]**
- WO 9314778 A **[0252]**
- WO 9007936 A **[0252]**
- WO 9403622 A **[0252]**
- WO 9325698 A **[0252]**
- WO 9325234 A **[0252]**
- US 5219740 A **[0252]**
- WO 9311230 A **[0252]**
- WO 9310218 A **[0252]**
- US 4777127 A **[0252]**
- GB 2200651 A **[0252]**
- EP 0345242 A **[0252]**
- WO 9102805 A **[0252]**
- WO 9412649 A **[0252]**
- WO 9303769 A **[0252]**
- WO 9319191 A **[0252]**
- WO 9428938 A **[0252]**
- WO 9511984 A **[0252]**
- WO 9500655 A **[0252]**
- US 5814482 A **[0252]**
- WO 9507994 A **[0252]**
- WO 9617072 A **[0252]**
- WO 9530763 A **[0252]**
- WO 9742338 A **[0252]**
- WO 9011092 A **[0252]**
- US 5580859 A **[0252]**
- US 5422120 A **[0252]**
- WO 9513796 A **[0252]**
- WO 9423697 A **[0252]**
- WO 9114445 A **[0252]**
- EP 0524968 A **[0252]**
- US 5206152 A **[0252]**
- WO 9211033 A **[0252]**
- US 5149655 A **[0252]**
- WO 9014837 A **[0252]**
- WO 0007621 A **[0252]**
- GB 2220221 A **[0252]**
- EP 0689454 A **[0252]**
- EP 0835318 A **[0252]**
- EP 0735898 A **[0252]**
- EP 0761231 A **[0252]**
- WO 9952549 A **[0252]**
- WO 0121207 A **[0252]**
- WO 0121152 A **[0252]**
- WO 0062800 A **[0252]**
- WO 0023105 A **[0252]**
- WO 9911241 A **[0252]**
- WO 9857659 A **[0252]**
- WO 9313202 A **[0252]**
- US 5858723 A **[0252]**
- US 5010175 A **[0252]**
- WO 9117823 A **[0252]**
- US 4816567 A **[0252]**
- US 154 A **[0252]**
- US 5807522 A **[0252]**
- EP 0799897 A **[0252]**
- WO 9729212 A **[0252]**
- WO 9727317 A **[0252]**
- EP 0785280 A **[0252]**
- WO 9702357 A **[0252]**
- US 5593839 A **[0252]**
- US 5578832 A **[0252]**
- EP 0728520 A **[0252]**
- US 5599695 A **[0252]**
- EP 0721016 A **[0252]**
- US 5556752 A **[0252]**
- WO 9522058 A **[0252]**
- US 5631734 A **[0252]**
- US 5134854 A **[0252]**
- US 5445934 A **[0252]**
- WO 9535505 A **[0252]**
- US 5800992 A **[0252]**
- WO 9202526 A **[0252]**
- US 5124246 A **[0252]**
- US 187 A **[0252]**
- US 4683195 A **[0252]**
- US 4683202 A **[0252]**

Non-patent literature cited in the description

- **MAYER et al.** *Nat. Genet.,* 1999, vol. 21 (3), 257-258 **[0252]**
- **FARRELL.** RNA Methodologies. Academic Press, 1998 **[0252]**
- **YANG et al.** *Proc Natl Acad Sci USA,* 1999, vol. 96 (23), 13404-8 **[0252]**
- **ROBBINS et al.** *Clin Lab Sci,* 1997, vol. 10 (5), 265-71 **[0252]**
- **YLIKOSKI et al.** *Clin Chem,* 1999, vol. 45 (9), 1397-407 **[0252]**
- **YLIKOSKI et al.** *Biotechniques,* 2001, vol. 30, 832-840 **[0252]**
- **SHIRAHATA ; PEGG.** *J. Biol. Chem.,* 1986, vol. 261 (29), 13833-7 **[0252]**
- **SAMBROOK et al.** Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, 1989 **[0252]**
- Short protocols in molecular biology. 1999 **[0252]**
- Current Protocols in Molecular Biology. 1987, 30 **[0252]**
- **BERKHOUT et al.** *J. Virol.,* 1999, vol. 73, 2365-2375 **[0252]**
- **LÖWER et al.** *J. Virol.,* 1995, vol. 69, 141-149 **[0252]**
- **MAGIN et al.** *J. Virol.,* 1999, vol. 73, 9496-9507 **[0252]**
- **MAGIN-LACHMANN.** *J. Virol.,* 2001, vol. 75 (21), 10359-71 **[0252]**
- **HASHIDO et al.** *Biochem. Biophys. Res. Comm.,* 1992, vol. 187, 1241-1248 **[0252]**
- **GEYSEN et al.** *PNAS USA,* 1984, vol. 81, 3998-4002 **[0252]**
- **CARTER.** *Methods Mol Biol,* 1994, vol. 36, 207-23 **[0252]**
- **JAMESON, BA et al.** *CABIOS,* 1988, vol. 4 (1), 181-186 **[0252]**
- **RADDRIZZANI ; HAMMER.** *Brief Bioinform,* 2000, vol. 1 (2), 179-89 **[0252]**
- **DE LALLA et al.** *J. Immunol.,* 1999, vol. 163, 1725-29 **[0252]**
- **BRUSIC et al.** *Bioinformatics,* 1998, vol. 14 (2), 121-30 **[0252]**
- **MEISTER et al.** *Vaccine,* 1995, vol. 13 (6), 581-91 **[0252]**
- **ROBERTS et al.** *AIDS Res Hum Retroviruses,* 1996, vol. 12 (7), 593-610 **[0252]**
- **MAKSYUTOV ; ZAGREBELNAYA.** *Comput Appl Biosci,* 1993, vol. 9 (3), 291-7 **[0252]**
- **FELLER ; DE LA CRUZ.** *Nature,* 1991, vol. 349 (6311), 720-1 **[0252]**
- **HOPP.** *Peptide Research,* 1993, vol. 6, 183-190 **[0252]**
- **WELLING et al.** *FEBS Lett.,* 1985, vol. 188, 215-218 **[0252]**
- **DAVENPORT et al.** *Immunogenetics,* 1995, vol. 42, 392-297 **[0252]**
- **SMITH ; WATERMAN.** *Adv. Appl. Math.,* 1981, vol. 2, 482-489 **[0252]**
- **GO et al.** *Int. J. Peptide Protein Res.,* 1980, vol. 15, 211 **[0252]**
- **QUEROL et al.** *Prot. Eng.,* 1996, vol. 9, 265 **[0252]**
- **OLSEN ; THOMSEN.** *J. Gen. Microbiol.,* 1991, vol. 137, 579 **[0252]**
- **CLARKE et al.** *Biochemistry,* vol. 32, 4322 **[0252]**
- **WAKARCHUK et al.** *Protein Eng.,* 1994, vol. 7, 1379 **[0252]**
- **TOMA et al.** *Biochemistry,* 1991, vol. 30, 97 **[0252]**
- **HAEZERBROUCK et al.** *Protein Eng.,* 1993, vol. 6, 643 **[0252]**
- **MASUL et al.** *Appl. Env. Microbiol.,* 1994, vol. 60, 3579 **[0252]**
- **BREEDVELD.** *Lancet,* 2000, vol. 355 (9205), 735-740 **[0252]**
- **GORMAN ; CLARK.** *Semin. Immunol.,* 1990, vol. 2, 457-466 **[0252]**
- **JONES et al.** *Nature,* 1986, vol. 321, 522-525 **[0252]**
- **MORRISON et al.** *Proc. Natl. Acad. Sci, US.A.,* 1984, vol. 81, 6851-6855 **[0252]**
- **MORRISON ; OI.** *Adv. Immunol.,* 1988, vol. 44, 65-92 **[0252]**
- **VERHOEYER et al.** *Science,* 1988, vol. 239, 1534-1536 **[0252]**
- **PADLAN.** *Molec. Immun.,* 1991, vol. 28, 489-498 **[0252]**
- **PADLAN.** *Molec. Immunol.,* 1994, vol. 31 (3), 169-217 **[0252]**
- **KETTLEBOROUGH, C.A. et al.** *Protein Eng.,* 1991, vol. 4 (7), 773-83 **[0252]**
- **CHOTHIA et al.** *J. Mol. Biol.,* 1987, vol. 196, 901-917 **[0252]**
- **FINDEIS et al.** *Trends Biotechnol.,* 1993, vol. 11, 202 **[0252]**
- **CHIOU et al.** *Gene Therapeutics: Methods And Applications Of Direct Gene Transfer,* 1994 **[0252]**
- **WU et al.** *J. Biol. Chem.,* 1988, vol. 263, 621 **[0252]**
- **WU et al.** *J. Biol. Chem.,* 1994, vol. 269, 542 **[0252]**
- **ZENKE et al.** *Proc. Natl. Acad. Sci. (USA),* 1990, vol. 87, 3655 **[0252]**
- **WU et al.** *J. Biol. Chem.,* 1991, vol. 266, 338 **[0252]**
- **JOLLY.** *Cancer Gene Therapy,* 1994, vol. 1, 51 **[0252]**
- **KIMURA.** *Human Gene Therapy,* 1994, vol. 5, 845 **[0252]**
- **CONNELLY.** *Human Gene Therapy,* 1995, vol. 1, 185 **[0252]**
- **KAPLITT.** *Nature Genetics,* 1994, vol. 6, 148 **[0252]**
- **CURIEL.** *Hum. Gene Ther.,* 1992, vol. 3, 147 **[0252]**
- **WU.** *J. Biol. Chem.,* 1989, vol. 264, 16985 **[0252]**
- **PHILIP.** *Mol. Cell Biol.,* 1994, vol. 14, 2411 **[0252]**
- **WOFFENDIN.** *Proc. Natl. Acad. Sci.,* 1994, vol. 91, 11581 **[0252]**
- Vaccine Design - the subunit and adjuvant approach. 1995 **[0252]**

- **MCSHARRY.** *Antiviral Res,* 1999, vol. 43 (1), 1-21 **[0252]**
- **KUHELJ et al.** *J Biol Chem,* 2001, vol. 276 (20), 16674-82 **[0252]**
- **SCHOMMER et al.** *J Gen Virol,* 1996, vol. 77, 375-379 **[0252]**
- **MAGIN et al.** *Virology,* 2000, vol. 274, 11-16 **[0252]**
- **BOESE et al.** *FEBS Lett,* 2001, vol. 493 (2-3), 117-21 **[0252]**
- **LARSSON, E. et al.** *Current Topics in Microbiology and Immunology,* 1989, vol. 148, 115 **[0252]**
- **MARIANI-COSTANTINI et al.** *J. Virol.,* 1989, vol. 63, 4982 **[0252]**
- **TÖNJES et al.** *J. AIDS Hum. Retrovir.,* 1996, vol. 13 (1), S261-S267 **[0252]**
- **BARBULESCU et al.** *Curr. Biol.,* 1999, vol. 9, 861 **[0252]**
- **ONO et al.** *J. Virol.,* 1986, vol. 58, 937 **[0252]**
- **LÖWER et al.** *Proc. Natl. Acad. Sci USA,* 1993, vol. 90, 4480 **[0252]**
- **ONO et al.** *J. Virol.,* 1986, vol. 60, 589 **[0252]**
- **BOLLER et al.** *Virol.,* 1993, vol. 196, 349 **[0252]**
- **YANG et al.** *Proc. Natl. Acad. Sci USA,* 1999, vol. 96, 13404 **[0252]**
- **MUELLER- LANTZSCH et al.** *AIDS Research and Human Retroviruses,* 1993, vol. 9, 343-350 **[0252]**
- **HERBST et al.** *Amer. J. Pathol.,* 1996, vol. 149, 1727 **[0252]**
- **LÖWER et al.** *Proc. Natl. Acad. Sci USA,* 1996, vol. 93, 5177 **[0252]**
- **LÖWER et al.** *Virology,* 1993, vol. 192, 501 **[0252]**
- **ANDERSSON et al.** *J. Gen. Virol.,* 1999, vol. 80, 255-260 **[0252]**
- **ZSÍROS et al.** *J. Gen. Virol.,* 1998, vol. 79, 61-70 **[0252]**
- **TÖNJES et al.** *J. Virol.,* 1999, vol. 73, 9187-9195 **[0252]**
- **JOHNSTON et al.** *Ann Neurol,* vol. 50 (4), 434-42 **[0252]**
- **MEDSTRAND et al.** *J Virol,* vol. 72 (12), 9782-7 **[0252]**
- **MERRIFELD.** *J. Am. Chem. Soc.,* 1963, vol. 85, 2149 **[0252]**
- **CAPRINO ; HAN.** *J. Org. Chem.,* 1972, vol. 37, 3404 **[0252]**
- **MILSTEIN ; KOHLER.** *Nature,* 1975, vol. 256, 495-497 **[0252]**
- **GULFRE ; MILSTEIN.** *Methods in Enzymology: Immunochemical Techniques,* vol. 73, 1-46 **[0252]**
- **ALTSCHUL et al.** *Nucleic Acids Res.,* 1997, vol. 25, 3389-3402 **[0252]**
- **BRUTLAG et al.** *Comp. Chem.,* 1993, vol. 17, 203 **[0252]**
- **SCHENA et al.** *Proc Natl Acad Sci U S A.,* 1996, vol. 93 (20), 10614-9 **[0252]**
- **SCHENA et al.** *Science,* 1995, vol. 270 (5235), 467-70 **[0252]**
- **SHALON et al.** *Genome Res.,* 1996, vol. 6 (7), 639-45 **[0252]**
- **PAPPALARADO et al.** *Sem. Radiation Oncol.,* vol. 8, 217 **[0252]**
- **RAMSAY.** *Nature Biotechnol.,* 1998, vol. 16, 40 **[0252]**
- **MULLIS et al.** *Meth. Enzymol.,* 1987, vol. 155, 335 **[0252]**
- **SAIKI et al.** *Science,* vol. 239, 487 **[0252]**
- **HANAHAN et al.** *Cell,* 2000, vol. 100, 57-70 **[0252]**
- **WEISSMAN SM.** *Mol Biol. Med.,* 1987, vol. 4 (3), 133-143 **[0252]**
- **PATANJALI et al.** *Proc. Natl. Acad. Sci. USA,* 1991, vol. 88 **[0252]**
- **SIMONE et al.** *Am J Pathol.,* 2000, vol. 156 (2), 445-52 **[0252]**
- **CLAVERIE.** *Meth. Enzymol.,* 1996, vol. 266, 212-227 **[0252]**
- Automated DNA Sequencing and Analysis Techniques. Academic, 1994, 267 **[0252]**
- **CLAVERIE et al.** *Comput. Chem.,* 1993, vol. 17, 191 **[0252]**
- **ALTSCHUL.** *J. Mol. Biol.,* 1990, vol. 215, 403-410 **[0252]**
- **PEARSON ; LIPMAN.** *PNAS,* 1988, vol. 85, 2444 **[0252]**
- **LUO et al.** *Nature Med,* 1999, vol. 5, 117-122 **[0252]**
- **HIGGINS ; SHARP.** *CABIOS,* 1989, vol. 5, 151-153 **[0252]**
- **DELLI BOVI et al.** *Cancer Res.,* 1986, vol. 46, 6333-6338 **[0252]**
- **CESARONE, C. et al.** *Anal Biochem,* 1979, vol. 100, 188-197 **[0252]**
- **SOUTHERN, E. M.** *J. Mol. Biol.,* 1975, vol. 98, 503-517 **[0252]**
- **FEINBERG, A. P. et al.** *Anal. Biochem.,* 1983, vol. 132, 6-13 **[0252]**
- **WRIGHT ; MANOS et al.** *PCR Protocols,* 1990, 153-158 **[0252]**
- **KEOWN et al.** *Methods in Enzymology,* 1990, vol. 185, 527-537 **[0252]**
- **MARKS et al.** *Brit. J. Urol.,* 1995, vol. 75, 225 **[0252]**
- **SKEA et al.** *J. Immunol.,* 1993, vol. 151, 3557 **[0252]**
- **MATHER et al.** *J. Nucl. Med.,* 1990, vol. 31, 692 **[0252]**
- **ZHANG et al.** *Nucl. Med. Biol.,* 1992, vol. 19, 607 **[0252]**

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	EP1415005B1	公开(公告)日	2012-11-21
申请号	EP2001996222	申请日	2001-12-07
[标]申请(专利权)人(译)	希龙公司		
申请(专利权)人(译)	Chiron公司		
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	C07K14/15 C12N7/00 C12Q1/68 G01N33/574 G01N33/53 A61K38/00 A61K39/00 A61K39/395 A61K45/00 A61K48/00 A61P3/10 A61P15/00 A61P25/00 A61P35/00 C07K16/10 C12N15/09 C12P21/08 C12Q1/02 C12Q1/70 G01N33/569 G01N33/577		
CPC分类号	A61K39/00 A61K2039/505 A61K2039/53 A61P13/08 A61P15/00 A61P25/00 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		
其他公开文献	EP1415005A2		
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。

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