

EUKARYOTIC CELLS COMPRISING ADENOVIRUS-ASSOCIATED VIRUS POLYNUCLEOTIDES

This Application claims priority to U.S. Application Serial No. 63/256,730, filed October 18, 2021, which is hereby incorporated by reference in its entirety.

FIELD OF THE INVENTIONS

[0001] The present inventions provide eukaryotic cells that comprise adeno-associated virus (AAV) polynucleotides, including AAV capsid proteins (Cap). The cells are capable of expressing the polypeptides encoded by the AAV polynucleotides, and thereby are capable of producing AAV, including recombinant AAV. The eukaryotic cells also may comprise adenovirus (Ad) polynucleotides. The present inventions also provide methods of expressing AAV polynucleotides, as well as Ad polynucleotides, in eukaryotic cells. The present inventions also provide methods for producing recombinant adeno-associated virus utilizing eukaryotic cells that express AAV and Ad polypeptides encoded by polynucleotides, as well as recombinant AAV produced by these inventive methods. The present inventions further provide other products and methods described herein.

REFERENCE TO ELECTRONIC SEQUENCE LISTING

[0002] The application contains a Sequence Listing, which has been submitted electronically in .XML format and is hereby incorporated by reference in its entirety. Said .XML copy, created on October 5, 2022, is named "135975-61702.xml" and is 229,879 bytes in size. The sequence listing contained in this .XML file is part of the specification and is hereby incorporated by reference herein in its entirety.

BACKGROUND OF THE INVENTIONS

[0003] Adeno-associated virus (AAV) is a non-enveloped, single-stranded DNA virus and is used as a gene delivery vector for both research and therapeutics. Weitzman and Linden, Adeno-Associated Virus Biology (chapter 1), *Meth. Molec. Biol.* 807: 1-23 (2011). Gene transfer vectors based on AAV have demonstrated promise for human gene therapy based on their safety profile and potential to achieve long-term efficacy in animal models. Wang *et al.*, *Nature*, 18: 358-78 (2019). A major challenge for advancing AAV-based therapies into clinical development is the difficulty and cost of producing sufficient quantities of AAV through transient methodologies.

[0004] AAV has been produced in HEK 293, BHK, human amniotic (for example, epithelial cells such as HAEpiC) and SF9 lines. However, expression is transient due to the use of plasmid vectors containing the necessary AAV and helper virus genes. For example, recombinant AAV production in HEK 293 cells using adenovirus helper gene products utilizes adenovirus E2A, E4, VA RNA and AAV Rep and Cap, along with AAV inverted terminal repeats (ITR) flanking the polynucleotide of interest. The reliance on non-integrated plasmid vectors means that the requisite gene products will be lost over time and need to be continually re-established.

[0005] The AAV genome includes a capsid gene referred to as “Cap” or “CAP”. Cap in nature is translated to produce, via alternative start codons and transcript splicing, three size-variant structural proteins referred to as VP1 (about 90kDa), VP2 (about 72kDa) and VP3 (about 60kDa). An AAV capsid contains 60 subunits total of the VP proteins. A ratio of 1:1:10 is considered the most typical

ratio for VP1:VP2:VP3, which is a stoichiometry of 5 VP1 subunits:5 VP2 subunits:50 VP3 subunits. However, there can be variation. Wörner *et al.*, *Nature Communications* 12:1642 (2021). AAV polynucleotides and proteins, including CAP, can be selected from any serotype.

[0006] Thus, there exists the need to develop improved cells and production methods that avoid the transient nature of non-integrated plasmid vectors.

SUMMARY OF THE INVENTIONS

[0007] The present inventions provide stable eukaryotic cells, such as mammalian cells (for example, primate, rodent and canine cells), comprising integrated AAV polynucleotides and Ad polynucleotides. The protein and VA RNA products of the Ad polynucleotides act in a helper capacity. All AAV and Ad types are amenable for use according to the present inventions. The present inventions advantageously can employ site-specific integration into the cell genome, which refers to pre-selected genomic sites for exogenous DNA to be inserted into a cellular genome. Random insertion can be employed as well.

[0008] Herein described are polynucleotides, where each can comprise (i) a promoter, (ii) an intron, (iii) an internal ribosome entry site, (iv) a polynucleotide encoding Adeno-associated virus (AAV) Cap protein, and (v) a polyadenylation site. For example, the polynucleotide can be in a CHO cell and have the (i) promoter, (ii) intron, (iii) internal ribosome entry site and (iv) polynucleotide encoding AAV Cap protein can be operably linked. The polynucleotide can be integrated into a CHO cell genome, such as a CHO chromosome. Alternatively, the polynucleotide can be

in a HEK 293 cell and have the (i) promoter, (ii) intron, (iii) internal ribosome entry site and (iv) polynucleotide encoding AAV Cap protein operably linked. The polynucleotide can be integrated into a HEK 293 cell genome, such as a HEK 293 cell chromosome. Another alternative is the polynucleotide can be in a BHK cell and have the (i) promoter, (ii) intron, (iii) internal ribosome entry site and (iv) polynucleotide encoding AAV Cap protein operably linked. The polynucleotide can be integrated into a BHK cell genome, such as a BHK cell chromosome. In yet another alternative is the polynucleotide can be in a human amniotic cell and have the (i) promoter, (ii) intron, (iii) internal ribosome entry site and (iv) polynucleotide encoding AAV Cap protein operably linked. The polynucleotide can be integrated into a human amniotic cell genome, such as a human amniotic cell chromosome. Additionally, the polynucleotide can be integrated into non-chromosomal locations as known by the person skilled in the art, such as episomes.

[0009] The polynucleotide can further comprise an operator. The promoter can be a CMV promoter and the operator can be a Tet operator. AAV proteins and polynucleotides, including CAP, can be selected from any serotype. When Cap is from serotype 5 ("Cap5") and expressed in a CHO cell, the polynucleotide allows for production of AAV Cap5 VP2 and VP3 protein, wherein the amount of VP3 produced is greater than the amount of VP2 produced. Production of VP1 can be less than 1% the level of VP2 production. Production ratios can vary based upon experimental conditions and analytical techniques.

[0010] Additionally, polynucleotides are described, wherein each can comprise (i) a promoter, (ii) an intron, (iii) a first internal ribosome entry site, (iv) a first polynucleotide encoding AAV Cap protein, (v) a second internal ribosome entry

site, (vi) a second polynucleotide encoding AAV Cap protein, and (vii) a polyadenylation site. For example, the polynucleotide can be in a CHO cell and the (i) promoter, (ii) intron, (iii) first internal ribosome entry site, (iv) first polynucleotide encoding AAV Cap protein, (v) a second internal ribosome entry site, and (vi) second polynucleotide encoding AAV Cap protein can be operably linked. The polynucleotide can be integrated into a CHO cell genome. Alternatively, the polynucleotide can be in a HEK 293 cell and the (i) promoter, (ii) intron, (iii) first internal ribosome entry site, (iv) first polynucleotide encoding AAV Cap protein, (v) a second internal ribosome entry site, and (vi) second polynucleotide encoding AAV Cap protein are operably linked. The polynucleotide can be integrated into a HEK 293 cell genome. In another alternative, the polynucleotide can be in a BHK cell and the (i) promoter, (ii) intron, (iii) first internal ribosome entry site, (iv) first polynucleotide encoding AAV Cap protein, (v) a second internal ribosome entry site, and (vi) second polynucleotide encoding AAV Cap protein are operably linked. The polynucleotide can be integrated into a BHK cell genome. In another alternative, the polynucleotide can be in a human amniotic cell and the (i) promoter, (ii) intron, (iii) first internal ribosome entry site, (iv) first polynucleotide encoding AAV Cap protein, (v) a second internal ribosome entry site, and (vi) second polynucleotide encoding AAV Cap protein are operably linked. The polynucleotide can be integrated into a human amniotic cell genome. The polynucleotide can further comprise an operator. The promoter can be a CMV promoter and the operator can be a Tet operator. AAV proteins and polynucleotides, including CAP, can be selected from any serotype. When Cap is from serotype 5 ("Cap5") and expressed in a CHO cell, the polynucleotide allows for production of AAV Cap5 proteins VP1, VP2 and VP3. The

amount of VP3 production can be greater than the amount of VP1 production and the amount of VP2 production. The amount of VP2 production can be greater than the amount of VP1 production. Production ratios can vary based upon experimental conditions and analytical techniques.

[0011] Moreover, there are described polynucleotides, wherein each polynucleotide can comprise (i) a promoter, (ii) an intron, (iii) a polynucleotide encoding AAV Cap protein, and (iv) a polyadenylation site, wherein the polynucleotide allows for production of AAV Cap VP1 protein when expressed. For example, the polynucleotide can be in a CHO cell and the (i) promoter, (ii) intron, and (iii) polynucleotide encoding AAV Cap protein can be operably linked. The polynucleotide can be integrated into a CHO cell genome. Alternatively, the polynucleotide can be in a HEK 293 cell and the (i) promoter, (ii) intron, and (iii) polynucleotide encoding AAV Cap protein can be operably linked. The polynucleotide can be integrated into a HEK 293 cell genome. In another alternative, the polynucleotide can be in a BHK cell and the (i) promoter, (ii) intron, and (iii) polynucleotide encoding AAV Cap protein can be operably linked. The polynucleotide can be integrated into a BHK cell genome. In still another alternative, the polynucleotide can be in a human amniotic cell and the (i) promoter, (ii) intron, and (iii) polynucleotide encoding AAV Cap protein can be operably linked. The polynucleotide can be integrated into a human amniotic cell genome. The polynucleotide can further comprise an operator. The promoter can be a CMV promoter and the operator can be a Tet operator.

[0012] Furthermore, there are described polynucleotides, wherein each can comprise (i) a promoter, (ii) an internal ribosome binding site, (iii) a

polynucleotide encoding AAV Cap protein, and (iv) a polyadenylation site, wherein the polynucleotide allows for production of AAV VP1 Cap protein when expressed. The polynucleotide can be in a CHO cell and the (i) promoter, (ii) internal ribosome binding site, and (iii) polynucleotide encoding AAV Cap protein can be operably linked. The polynucleotide can be integrated into a CHO genome. Alternatively, the polynucleotide can be in a HEK 293 cell and the (i) promoter, (ii) internal ribosome binding site, and (iii) polynucleotide encoding AAV Cap protein can be operably linked. The polynucleotide can be integrated into a HEK 293 cell genome. In another alternative, the polynucleotide can be in a BHK cell and the (i) promoter, (ii) internal ribosome binding site, and (iii) polynucleotide encoding AAV Cap protein can be operably linked. The polynucleotide can be integrated into a BHK cell genome. In yet another alternative, the polynucleotide can be in a human amniotic cell and the (i) promoter, (ii) internal ribosome binding site, and (iii) polynucleotide encoding AAV Cap protein can be operably linked. The polynucleotide can be integrated into a human amniotic cell genome. The polynucleotide can comprise an operator. The promoter can be a CMV promoter and the operator can be a Tet operator.

[0013] Additionally, there are described eukaryotic cells, where each cell can comprise a polynucleotide comprising (i) a promoter, (ii) an intron, (iii) a first internal ribosome entry site, (iv) a first polynucleotide encoding AAV Cap protein, (v) a second internal ribosome entry site, (vi) a second polynucleotide encoding AAV Cap protein, and (vii) a polyadenylation site. The (i) promoter, (ii) intron, (iii) first internal ribosome entry site, (iv) first polynucleotide encoding AAV Cap protein, (v) a second internal ribosome entry site, and (vi) second polynucleotide encoding AAV

Cap protein can be operably linked. The polynucleotide can be integrated into a cell genome. The cell can be a CHO cell, a HEK 293 cell, a BHK cell, a Human Amniotic Cell or other eukaryotic cell. The cell can further comprise an operator. The promoter can be a CMV promoter and the operator can be a Tet operator. The cell can further comprise: a polynucleotide encoding AAV Rep, a polynucleotide encoding Ad E1A, a polynucleotide encoding Ad E1B, a polynucleotide encoding Ad E2A or E2A orf, a polynucleotide encoding Ad E4 or E4 orf 6, a polynucleotide encoding VA RNA, and a polynucleotide encoding AAV ITRs and a protein of interest.

[0014] There are also described eukaryotic cells, where each cell can comprise (A) a first polynucleotide comprising (i) a promoter, (ii) an intron, (iii) an internal ribosome entry site, (iv) a polynucleotide encoding AAV Cap protein, and (v) a polyadenylation site; and (B) a second polynucleotide comprising (i) a promoter, (ii) an intron, (iii) a polynucleotide encoding AAV Cap protein, and (iv) a polyadenylation site. The (i) promoter, (ii) intron, (iii) internal ribosome entry site and (iv) polynucleotide encoding AAV Cap protein of (A) first polynucleotide can be operably linked, and wherein the (i) promoter, (ii) intron, and (iii) polynucleotide encoding AAV Cap protein of (B) second polynucleotide can be operably linked. The cell can have at least one polynucleotide integrated into a cell genome. The cell can be a CHO cell, a HEK 293 cell, a BHK cell, a human amniotic cell or other eukaryotic cell. The cell can further comprise an operator. The promoter can be a CMV promoter and the operator can be a Tet operator. The cell can further comprise a polynucleotide encoding AAV Rep, a polynucleotide encoding Ad E1A, a polynucleotide encoding Ad E1B, a polynucleotide encoding Ad E2A or E2A orf, a

polynucleotide encoding E4 or E4 orf 6, a polynucleotide encoding VA RNA, and a polynucleotide encoding AAV ITRs and a protein of interest.

[0015] Additionally, there are described eukaryotic cells, where each cell can comprise: (A) a first polynucleotide comprising (i) a promoter, (ii) an intron, (iii) an internal ribosome entry site, (iv) a polynucleotide encoding AAV Cap protein, and (v) a polyadenylation site; and (B) a second polynucleotide comprising (i) a promoter, (ii) an internal ribosome entry site, (iii) a polynucleotide encoding AAV Cap protein, and (iv) a polyadenylation site. The cells can have the (i) promoter, (ii) intron, (iii) internal ribosome entry site and (iv) polynucleotide encoding AAV Cap protein of (A) first polynucleotide operably linked, and the (i) promoter, (ii) internal ribosome entry site, and (iii) polynucleotide encoding AAV Cap protein of (B) second polynucleotide operably linked. At least one polynucleotide can be integrated into a cell genome. The cell can be a CHO cell, a HEK 293 cell, a BHK cell, a human amniotic cell or other eukaryotic cell. The cell can further comprise an operator. The promoter can be a CMV promoter and the operator can be a Tet operator. The cell can further comprise: a polynucleotide encoding AAV Rep, a polynucleotide encoding Ad E1A, a polynucleotide encoding Ad E1B, a polynucleotide encoding Ad E2A or E2A orf, a polynucleotide encoding Ad E4 or E4 orf 6, a polynucleotide encoding VA RNA, and a polynucleotide encoding AAV ITRs and a protein of interest.

[0016] There also are described cell cultures comprising any of the above cells in any type of media, including growth media and maintenance media. Additionally, there are described methods of producing AAV proteins, including Cap proteins, and methods that can result in the production of recombinant AAV.

[0017] There are described methods of producing adeno-associated virus (AAV) Cap protein in cell culture, wherein a method comprises the steps of: providing eukaryotic cells, wherein a cell comprises a polynucleotide comprising (i) a promoter, (ii) an intron, (iii) a first internal ribosome entry site, (iv) a first polynucleotide encoding AAV Cap protein, (v) a second internal ribosome entry site, (vi) a second polynucleotide encoding AAV Cap protein, and (vii) a polyadenylation site; and culturing the cells in a culture medium to allow the cells to produce AAV Cap protein, wherein the polynucleotide allows for production of AAV Cap proteins VP1, VP2 and VP3. The (i) promoter, (ii) intron, (iii) first internal ribosome entry site, (iv) first polynucleotide encoding AAV Cap protein, (v) a second internal ribosome entry site, and (vi) second polynucleotide encoding AAV Cap protein can be operably linked. The polynucleotide can be integrated into a cell genome. The cell can be a CHO cell, a HEK 293 cell, a BHK cell, a human amniotic cell or other eukaryotic cell. The cell can further comprise an operator. The cell can further comprise: a polynucleotide encoding AAV Rep, a polynucleotide encoding Ad E1A, a polynucleotide encoding Ad E1B, a polynucleotide encoding Ad E2A or E2A orf, a polynucleotide encoding Ad E4 or E4 orf 6, a polynucleotide encoding VA RNA, and a polynucleotide encoding AAV ITRs and a protein of interest, wherein the cell can produce recombinant AAV.

[0018] Also described are methods of producing adeno-associated virus (AAV) Cap protein in cell culture, wherein a method comprises the steps of providing eukaryotic cells, where a cell comprises (a) a first polynucleotide comprising (i) a promoter, (ii) an intron, (iii) an internal ribosome entry site, (iv) a polynucleotide encoding AAV Cap protein, and (v) a polyadenylation site; and (b)

a second polynucleotide comprising (i) a promoter, (ii) an intron, (iii) a polynucleotide encoding AAV Cap protein, and (iv) a polyadenylation site; and culturing the cells in a culture medium to allow the cells to produce AAV Cap protein, wherein the polynucleotide allows for production of AAV Cap proteins VP1, VP2 and VP3. The (i) promoter, (ii) intron, (iii) internal ribosome entry site and (iv) polynucleotide encoding AAV Cap protein of (a) first polynucleotide can be operably linked, and the (i) promoter, (ii) intron, and (iii) polynucleotide encoding AAV Cap protein of (b) second polynucleotide can be operably linked. The polynucleotide can be integrated into a cell genome. The cell can be a CHO cell, a HEK 293 cell, a BHK cell, a human amniotic cell or other eukaryotic cell. The cell can further comprise an operator. The promoter can be a CMV promoter and the operator can be a Tet operator. The cell can further comprise: a polynucleotide encoding AAV Rep, a polynucleotide encoding Ad E1A, a polynucleotide encoding Ad E1B, a polynucleotide encoding Ad E2A or E2A orf, a polynucleotide encoding Ad E4 or E4 orf 6, a polynucleotide encoding VA RNA, and a polynucleotide encoding AAV ITRs and a protein of interest, wherein the cell can produce recombinant AAV.

[0019] There are also described methods of producing adeno-associated virus (AAV) Cap protein in cell culture, wherein a method comprises the steps of: providing eukaryotic cells, where a cell comprises (a) a first polynucleotide comprising (i) a promoter, (ii) an intron, (iii) an internal ribosome entry site, (iv) a polynucleotide encoding AAV Cap protein, and (v) a polyadenylation site; and (b) a second polynucleotide comprising (i) a promoter, (ii) an internal ribosome entry site, (iii) a polynucleotide encoding AAV Cap protein, and (iv) a polyadenylation site; and culturing the cells in a culture medium to allow the cells to produce AAV Cap

protein, wherein the polynucleotide allows for production of AAV Cap proteins VP1, VP2 and VP3. The (i) promoter, (ii) intron, (iii) internal ribosome entry site and (iv) polynucleotide encoding AAV Cap protein of (a) first polynucleotide can be operably linked, and the (i) promoter, (ii) internal ribosome entry site, and (iii) polynucleotide encoding AAV Cap protein of (b) second polynucleotide can be operably linked. The polynucleotide can be integrated into a cell genome. The cell can be a CHO cell, a HEK 293 cell, a BHK cell, a human amniotic cell or other eukaryotic cell. The cell can further comprise an operator. The promoter can be a CMV promoter and the operator can be a Tet operator. The cell can further comprise: a polynucleotide encoding AAV Rep, a polynucleotide encoding Ad E1A, a polynucleotide encoding Ad E1B, a polynucleotide encoding Ad E2A or E2A orf, a polynucleotide encoding Ad E4 or E4 orf 6, a polynucleotide encoding VA RNA, and a polynucleotide encoding AAV ITRs and a protein of interest, wherein the cell can produce recombinant AAV.

BRIEF DESCRIPTION OF THE FIGURES

[0020] **FIGURE 1** is a schematic diagram of a polynucleotide comprising a promoter, an intron, an internal ribosome entry site (IRES), a polynucleotide encoding AAV Cap protein and a polyadenylation site (poly A).

[0021] **FIGURE 2** is a schematic diagram of a polynucleotide comprising a promoter, an intron, two internal ribosome entry sites, two polynucleotides encoding AAV Cap protein and a polyadenylation site.

[0022] **FIGURE 3** is a schematic diagram of a polynucleotide comprising a promoter, an intron, a polynucleotide encoding AAV Cap protein and a polyadenylation site.

[0023] **FIGURE 4** is a schematic diagram of a polynucleotide comprising a promoter, an internal ribosome entry site, a polynucleotide encoding AAV Cap protein and a polyadenylation site.

[0024] **FIGURE 5** is a schematic diagram of a polynucleotide comprising a promoter with operator (Op), an intron, an internal ribosome entry site, a polynucleotide encoding AAV Cap protein and a polyadenylation site.

[0025] **FIGURE 6** is a schematic diagram of a polynucleotide comprising a promoter with operator (Op), an intron, two internal ribosome entry sites, two polynucleotides encoding AAV Cap protein and a polyadenylation site.

[0026] **FIGURE 7** is a schematic diagram of a polynucleotide comprising a promoter with operator (Op), an intron, a polynucleotide encoding AAV Cap protein and a polyadenylation site.

[0027] **FIGURE 8** is a schematic diagram of a polynucleotide comprising a promoter with operator (Op), an internal ribosome entry site, a polynucleotide encoding AAV Cap protein and a polyadenylation site.

DETAILED DESCRIPTION OF THE INVENTIONS

[0028] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which these inventions belong.

Definitions

[0029] The term “about” in the context of numerical values and ranges refers to values or ranges that approximate or are close to the recited values or ranges such that the inventions can perform, such as having a sought rate, amount,

degree, increase, decrease, or extent of expression, concentration, or time, as is apparent from the teachings contained herein. Thus, this term encompasses values beyond those simply resulting from systematic error. For example, “about” can signify values either above or below the stated value in a range of approx. +/- 10% or more or less depending on the ability to perform.

[0030] “Intron” is a section of DNA located between exons. An intron is removed to form a mature messenger RNA. Preferred introns are those that can affect the starting point of translation, and exemplars are the hCMV-IE intron (Human cytomegalovirus immediate early protein) and FMDV intron (Foot and Mouth Disease Virus). The globin gene intron also has been reportedly used for expression.

[0031] A “nucleic acid moiety” includes any arrangement of single stranded or double stranded nucleotide sequences. Nucleic acid moieties can include, but are not limited to, polynucleotides, promoters, enhancers, operators, repressors, transcription termination signals, ribosomal entry sites and polyadenylation signals.

[0032] A “DNA cassette” or “cassette” is a type of nucleic acid moiety that comprises at least a promoter, at least one open reading frame and optionally a polyadenylation signal, for example an SV40 polyadenylation signal. Other nucleic acid moieties, such as operators, also are optional. A DNA cassette thus is a polynucleotide that comprises two or more shorter polynucleotides.

[0033] “Operably linked” refers to one or more nucleotide sequences in functional relationships with one or more other nucleotide sequences. Such functional relationships can directly or indirectly control, cause, regulate, enhance,

facilitate, permit, attenuate, repress or block an action or activity in accordance with the selected design. Exemplars include single-stranded or double-stranded nucleic acid moieties, and can comprise two or more nucleotide sequences arranged within a given moiety in such a way that sequence(s) can exert at least one functional effect on other(s). For example, a promoter operably linked to the coding region of a DNA polynucleotide sequence can facilitate transcription of the coding region. Other elements, such as enhancers, operators, repressors, transcription termination signals, ribosomal entry sites and polyadenylation signals also can be operably linked with a polynucleotide of interest to control its expression. Arrangements and spacing to achieve operable linkages can be ascertained by approaches available to the person skilled in the art, such as screening using western blots and RT-PCR.

[0034] “Operator” indicates a DNA sequence that is introduced in or near a polynucleotide sequence in such a way that the polynucleotide sequence may be regulated by the interaction of a molecule capable of binding to the operator and, as a result, prevent or allow transcription of the polynucleotide sequence, as the case may be. One skilled in the art will recognize that the operator must be located sufficiently in proximity to the promoter such that it is capable of controlling or influencing transcription by the promoter, which can be considered a type of operable linkage. The operator may be placed either downstream or upstream of the promoter. These include, but are not limited to, the operator region of the Lex A gene of *E. coli*, which binds the Lex A peptide and the lactose and 45 tryptophan operators, which bind the repressor proteins encoded by the *Lac* and *trpR* genes of *E. coli*. The bacteriophage operators from the lambda P_i and the phage P22 Mnt and Arc. Preferred operators are the Tet (tetracycline) operator and the Arc

operator. Operators can have a native sequence or a mutant sequence. For example, mutant sequences of the Tet operator are disclosed in Wissmann *et al.*, *Nucleic Acids Res.* 14: 4253-66 (1986).

[0035] The phrases "percent identity" or "% identical," in their various grammatical forms, when describing a sequence is meant to include homologous sequences that display the recited identity along regions of contiguous homology, but the presence of gaps, deletions, or insertions that have no homolog in the compared sequence are not taken into account in calculating percent identity. As used herein, a "percent identity" or "% identical" determination between homologs would not include a comparison of sequences where the homolog has no homologous sequence to compare in an alignment. Thus, "percent identity" and "% identical" do not include penalties for gaps, deletions, and insertions.

[0036] A "homologous sequence" in the context of nucleic acid sequences refers to a sequence that is substantially homologous to a reference nucleic acid sequence. In some embodiments, two sequences are considered to be substantially homologous if at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more of their corresponding nucleotides are identical over a relevant stretch of residues. In some embodiments, the relevant stretch is a complete (i.e., full) sequence.

[0037] "Polynucleotide" includes a sequence of nucleotides covalently joined, and includes RNA and DNA. Oligonucleotides are considered shorter polynucleotides. Genes are DNA polynucleotides (polydeoxyribonucleic acid) that ultimately encode polypeptides, which are translated from RNA (polyribonucleic acid) that was typically transcribed from DNA. DNA polynucleotides also can

encode RNA polynucleotides that is not translated, but rather function as RNA “products”. The type of polynucleotide (that is, DNA or RNA) is apparent from the context of the usage of the term. A polynucleotide referred to or identified by the polypeptide it encodes sets forth and covers all suitable sequences in accordance with codon degeneracy. Polynucleotides, including those disclosed herein, include percent identity sequences and homologous sequences when indicated.

[0038] “Polypeptide” or “peptide” refers to sequence(s) of amino acids covalently joined. Polypeptides include natural, semi-synthetic and synthetic proteins and protein fragments. “Polypeptide” and “protein” can be used interchangeably. Oligopeptides are considered shorter polypeptides.

[0039] “Protein of interest” or “polypeptide of interest” can have any amino acid sequence, and includes any protein, polypeptide, or peptide, and derivatives, components, domains, chains and fragments thereof. Included are, but not limited to, viral proteins, bacterial proteins, fungal proteins, plant proteins and animal (including human) proteins. Protein types can include, but are not limited to, antibodies, bi-specific antibodies, multi-specific antibodies, antibody chains (including heavy and light), antibody fragments, Fv fragments, Fc fragments, Fc-containing proteins, Fc-fusion proteins, receptor Fc-fusion proteins, receptors, receptor domains, trap and mini-trap proteins, enzymes, factors, repressors, activators, ligands, reporter proteins, selection proteins, protein hormones, protein toxins, structural proteins, storage proteins, transport proteins, neurotransmitters and contractile proteins. Derivatives, components, chains and fragments of the above also are included. The sequences can be natural, semi-synthetic or synthetic. Proteins of interest and polypeptides of interest are encoded by “genes

of interest,” which also can be referred to as “polynucleotides of interest.” Where multiple genes (same or different) are integrated, they can be referred to as “first,” “second,” “third,” “fourth,” “fifth,” “sixth,” “seventh,” “eighth,” “ninth,” “tenth,” etc. as is apparent from the context of use.

[0040] “Promoter” indicates a DNA sequence that cause transcription of a DNA sequence to which it is operably linked, i.e., linked in such a way as to permit transcription of the nucleotide sequence of interest when the appropriate signals are present and repressors are absent. The expression of a polynucleotide of interest may be placed under control of any promoter or enhancer element known in the art. A eukaryotic promoter can be operably linked to a TATA Box. The TATA Box is typically located upstream of the transcription start site.

[0041] Useful promoters that may be used include, but are not limited to, the SV40 early promoter region, SV40 E/L (early late) promoter, the promoter contained in the 3' long terminal repeat of Rous sarcoma virus, the regulatory sequences of the metallothionein gene, mouse or human cytomegalovirus major immediate early (CMV-MIE) promoter and other CMV promoters, including CMVmin promoters. Plant expression vectors comprising the nopaline synthetase promoter region, the cauliflower mosaic virus 35S RNA promoter, and the promoter of the photosynthetic enzyme ribulose biphosphate carboxylase; promoter elements from yeast or other fungi such as the Gal 4 promoter, the ADC (alcohol dehydrogenase) promoter, PGK (phosphoglycerol kinase) promoter, alkaline phosphatase promoter, and the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: elastase I; insulin; immunoglobulin; mouse mammary tumor virus; albumin; C.-feto protein; C.1-

antitrypsin; 3-globin, and myosin light chain-2. Various forms of the CMV promoter can be used according to the inventions.

[0042] Minimal promoters, such as CMVmin promoters, can be truncated promoters or core promoters and are preferred for use in controlled expression systems. Minimal promoters and development approaches are widely known and disclosed in, for example, Saxena *et al.*, *Methods Molec. Biol.* 1651:263-73 (2017); Ede *et al.*, *ACS Synth Biol.* 5:395-404 (2016); Brown *et al.*, *Biotech Bioeng.* 111:1638-47 (2014); Morita *et al.*, *Biotechniques* 0:1-5 (2012); Lagrange *et al.*, *Genes Dev.* 12:34-44 (1998). There are many CMVmin promoters described in the field.

[0043] “Reporter proteins” as used herein, refers to any protein capable of generating a detectable signal. Reporter proteins typically fluoresce, or catalyze a colorimetric or fluorescent reaction, and often are referred to as “fluorescent proteins” or “color proteins.” However, a reporter protein also can be non-enzymatic and non-fluorescent as long as it can be detected by another protein or moiety, such as a cell surface protein detected with a fluorescent ligand. A reporter protein also can be an inactive protein that is made functional through interaction with another protein that is fluorescent or catalyzes a reaction. Accordingly, any suitable reporter protein, as understood by one of skill in the art, could be used. In some aspects, the reporter protein may be selected from fluorescent protein, luciferase, alkaline phosphatase, β -galactosidase, β -lactamase, dihydrofolate reductase, ubiquitin, and variants thereof. Fluorescent proteins are useful for the recognition of gene cassettes that have or have not been successfully inserted and/or replaced, as the case may be. Fluid cytometry and fluorescence-activated cell sorting are

suitable for detection. Examples of fluorescent proteins are well-known in the art, including, but not limited to Discosoma coral (DsRed), green fluorescent protein (GFP), enhanced green fluorescent protein (eGFP), cyano fluorescent protein (CFP), enhanced cyano fluorescent protein (eCFP), yellow fluorescent protein (YFP), enhanced yellow fluorescent protein (eYFP) and far-red fluorescent protein (e.g. mKate, mKate2, mPlum, mRaspberry or E2-crimson. See, for example, U.S. Patent Nos. 9,816,110. Reporter proteins are encoded by polynucleotides, and are referred to herein as “reporter genes” or “reporter protein genes.” Reporters can be considered a type of marker. “Color” or “fluorescent,” in their various grammatical forms, also can be used the more specifically refer to a reporter protein or gene..

[0044] A “ repressor protein”, also referred to as a “repressor,” is a protein that can bind to DNA in order to repressor transcription. Repressors are of eukaryotic and prokaryotic origin. Prokaryotic repressors are preferred. Examples of repressor families include: TetR, LysR, LacI, ArsR, IclR, MerR, AsnC, MarR, DeoR, GntR and Crp families. Repressor proteins in the TetR family include: ArcR, ActII, AmeR, AmrR, ArpR, BpeR, EnvR, EthR, HemR, HydR, IfeR, LanK, LfrR, LmrA, MtrR, Pip, PqrA, QacR, RifQ, RmrR, SimReg2, SmeT, SrpR, TcmR, TetR, TtgR, TrgW, UrdK, VarR YdeS, ArpA., BarA, Aur1B, CalR1, CprB, FarA, JadR*, JadR2, MphB, NonG, PhIF, TyIQ, VanT, TarA, TyIP, BM1P1, Bm3R1, ButR, CampR, CamR, DhaR, KstR, LexA-like, AcnR, PaaRR, PsbI, Th1R, UidR, YDH1, BetI, McbR, MphR, PhaD, Q9ZF45, TtK, Yhgd, YixD, CasR, IcaR, LitR, LuxR, LuxT, OpaR, Orf2, SmcR, HapR, Ef0113, HlyIIR, BarB, ScbR, MmfR, AmtR, PsrA and YjdC proteins See Ramos *et al.*, *Microbiol. Mol. Biol. Rev.*, 69: 326-56 (2005). Still other repressors

include PurR, LacR, MetJ and PadR, Repressor proteins are encoded by genes referred to as “repressor genes” or “repressor protein genes.”

[0045] “Selectable” or “selection” marker proteins include proteins conferring certain traits, including but not limited to drug resistance or other selective advantages. Selection markers can give the cell receiving the selectable marker gene resistance towards a certain toxin, drug, antibiotic or other compound and permit the cell to produce protein and propagate in the presence of the toxin, drug, antibiotic or other compound, and are often referred to as “positive selectable markers.” Suitable examples of antibiotic resistance markers include, but are not limited to, proteins that impart resistance to various antibiotics, such as kanamycin, spectinomycin, neomycin, gentamycin (G418), ampicillin, tetracycline, chloramphenicol, puromycin, hygromycin, zeocin, and/or blasticidin. There are other selectable markers, often referred to as “negative selectable markers,” which cause a cell to stop propagating, stop protein production and/or are lethal to the cell in the presence of the negative selectable marker proteins. Thymidine kinase and certain fusion proteins can serve as negative selectable markers, including but not limited to GyrB-PKR. See White *et al.*, *Biotechniques*, 50: 303-309 (May 2011).

Selectable marker proteins and corresponding genes can be referred to generically as first (1), second (2), third (3), fourth (4), fifth (5), sixth (6), seventh (7), eighth (8), ninth (9), tenth (10), etc., as is apparent from the context of usage..

[0046] All numerical limits and ranges set forth herein include all numbers or values thereabout or there between of the numbers of the range or limit. The ranges and limits described herein expressly denominate and set forth all integers, decimals and fractional values defined and encompassed by the range or limit.

Description

[0047] The inventions provide cells comprising AAV and optionally Ad polynucleotide sequences to allow production of recombinant AAV comprising a polynucleotide of interest, such as a gene or other sequence encoding a polypeptide of interest. The AAV and Ad polynucleotides provide the requisite structural and helper products required for AAV production.

[0048] AAV polynucleotides, and optionally Ad polynucleotides, can be integrated using a recombinase-mediated cassette exchange (RMCE), for example. “Stable” in the context of cell integration refers to a polynucleotide of interest, such as a gene, introduced into the genome of a cell and can be passed to subsequent generations of cells, and thereby can provide cell lines that are genetically homogeneous for a period of time.

[0049] Cells that are suitable for use with the inventions can be readily selected by those of skill in the art. In some embodiments the cell line is a eukaryotic cell line such as a yeast cell line, insect cell line (for example, Sf9 and Sf21 cells) or a mammalian cell line. Preferred mammalian cells include primate cells (including human), canine cells and rodent cells. Cells can be primary cells or immortalized cells. Suitable cells can be selected from Vero cells, COS cells, HEK 293 cells, HeLa cells, CHO cells, BHK cells, MDCK cells, amniotic cells (human),

embryonic cells, cell lines transfected with viral genes, for example, AD5 E1, including but not limited to an immortalized human retinal cell transfected with an adenovirus gene, for example, a PER.C6 cell, or an NSO cell. In some embodiments, the cell is a Chinese hamster ovary (CHO) cell line. Some examples of CHO cells include, but are not limited to, CHO-ori, CHO-K1, CHO-s, CHO-DHB11, CHO-DXB11, CHO-K1SV, and mutants and variants thereof. In other embodiments, the cell is a HEK293 cell. Some examples of HEK293 cells include, but are not limited, to HEK293, HEK293A, HEK293E, HEK293F, HEK293FT, HEK293FTM, HEK293H, HEK293MSR, HEK293S, HEK293SG, HEK293SGGD, HEK293T and mutants and variants thereof.

[0050] For hamster cells such as CHO and BHK, integration can be accomplished by inventions disclosed in U.S. Patent Nos. 7,771,997 (“Stable Site 1”) and 9,816,110 (“Stable Site 2”), which are hereby incorporated by reference, including sequence information. Regeneron provides a suite of goods and services referred to as EESYR®. CHO cells with integrated sequences in Stable Site 1 and Stable Site 2 are disclosed in US 2019/0233544 A1, which is hereby incorporated by reference, including sequence information. Sequences set forth in these patents and Examples 14 and 15 can be used according to the inventions described and depicted herein. Additionally, an AAVS1-like region and the COSMC locus in hamster cells can be used according to the inventions.

[0051] Where human cells are employed, integration into adeno-associated virus integration site 1 (AAVS1) can be undertaken. See Lou *et al.*, *Human Gene Therapy Methods*, 28: 124-38 (2017); Liu *et al.*, *BMC Research Note*,

7:626 (2014). AAVS1 is reported to be located on chromosome 19. Other integration sites in human cells can be used as well, such as CCR5 and hROSA26.

[0052] Modification of cellular genomes can be undertaken with known approaches, such as Cre/Lox, Flp/Frt, transcription activator-like effector nuclease (TALEN), a TAL effector domain fusion protein, zinc finger nuclease (ZFN), a ZFN dimer, or a RNA-guided DNA endonuclease system, such as CRISPR/Cas9. See U.S. Patent No. 9,816,110 at cols. 17-18. Integration using Bxb1 integrase in human, mouse and rat cells also can be undertaken. Russell *et al.*, *Biotechniques* 40: 460-64 (2006).

[0053] To maximize stability and efficiency and facilitate integration and control of the inventions, Stable Integration Sites (SIS) can be created using Genomic Safe Harbors and the like in a wide variety of cell types and lines according to the teachings of U.S. Serial No. 63/256,675. The descriptions (including examples) and figures providing methods and cells resulting from the methods of U.S. Serial No. 63/256,675 are hereby incorporated by reference.

[0054] For production of recombinant AAV, the inventions provide for integration of AAV Cap to produce size variants VP1 (about 90kDa), VP2 (about 72kDa) and VP3 (about 60kDa). The variants differ at their N-terminus.

[0055] Usually, recombinant AAV will contain a gene-of interest (GOI) flanked by AAV ITRs (inverted terminal repeats), For production of recombinant AAV, seven additional polynucleotides, namely adenovirus E1A, E1B, E4, E2A, VA RNA and AAV Rep and Cap, are typically employed for production.

[0056] Figures 1-4 depict constructs that permit constitutive expression of Cap protein. Figures 5-8 depict constructs that permit controlled expression of

Cap protein by inclusion of an operator downstream of the promoter. A preferred operator is the tetracycline operator (TetO), which binds the tetracycline repressor (TetR). Tetracycline, doxycycline and derivatives thereof can bind TetR so that TetR no longer binds TetO, and thus is permissive for transcription. An example of a CMV promoter and TetO is set forth in Example 13.

[0057] The invention is further described by the following examples, which are illustrative of the many aspects of the invention, but do not limit the invention in any manner.

EXAMPLE 1 – CHO Cells

[0058] One or more Cap-containing polynucleotides according to Figures 1-8 are stably inserted into the CHO genome. A preferred promoter is the hCMV-IE promoter, and optionally a tet operator can be operably linked to the promoter for expression control. Optionally, an intron can be located 3' of the promoter. A preferred intron is an hCMV-IE intron. AAV Cap, Rep and ITRs can be obtained from any AAV serotype. Preferred AAV serotypes are AAV2 and AAV5. AAV polynucleotide sequences are set forth in Example 11. Promoter, operator, IRES and intron sequences are set forth in Example 13.

[0059] AAV ITRs flanking a gene of interest, AAV Rep and Ad E1A, E1B, E2A (or E2A partial sequence (E2A orf)), E4 (or E4 partial sequence (E4 orf 6)) and VA RNA can be randomly integrated, site-specifically integrated or remain on a plasmid. Adenovirus polynucleotide sequences are available and are exemplified in Example 12. Adenovirus (Ad) proteins and polynucleotides can be selected from any serotype.

EXAMPLE 2 – HEK 293 Cells

[0060] One or more Cap-containing polynucleotides according to Figures 1-8 are stably inserted into the HEK 293 genome. A preferred promoter is the hCMV-IE promoter, and optionally a tet operator can be operably linked to the promoter for expression control. Optionally, an intron can be located 3' of the promoter. A preferred intron is an hCMV-IE intron. AAV Cap, Rep and ITRs can be obtained from any AAV serotype. Preferred AAV serotypes of AAV2 and AAV5. AAV polynucleotide sequences are set forth in Example 11.

[0061] AAV ITRs and Rep and Ad E1A, E1B, E2A (or E2A partial sequence (E2A orf)), E4 (or E4 partial sequence (E4 orf 6)) and VA RNA can be randomly integrated, site-specifically integrated or remain on a plasmid. Adenovirus polynucleotide sequences are available and are exemplified in Example 12.

EXAMPLE 3 – BHK Cells

[0062] BHK cells are fibroblast cells from baby hamster kidneys. There are adherent BHK lines and BHK lines that can propagate in suspension. Wentz and Schügerl, *Enzyme Microbial Tech.* 14: 68-75 (1992).

[0063] One or more Cap-containing polynucleotides according to Figures 1-8 are stably inserted into the BHK genome. A preferred promoter is the hCMV-IE promoter, and optionally a tet operator can be operably linked to the promoter for expression control. Optionally, an intron can be located 3' of the promoter. A preferred intron is an hCMV-IE intron. AAV Cap, Rep and ITRs can be obtained from any AAV serotype. Preferred AAV serotypes of AAV2 and AAV5. AAV polynucleotide sequences are set forth in Example 11.

[0064] AAV ITRs and Rep and Ad E1A, E1B, E2A (or E2A partial sequence (E2A orf)), E4 (or E4 partial sequence (E4 orf 6)) and VA RNA can be randomly integrated, site-specifically integrated or remain on a plasmid. Adenovirus polynucleotide sequences are available and are exemplified in Example 12.

EXAMPLE 4 – Intron IRES CAP

[0065] Embodiments of this construct are depicted in Figures 1 (constitutive) and 5 (controllable). In a cell, both embodiments can primarily produce VP2 and VP3, with more VP3 being produced than VP2. In an experiment in CHO cells using Cap5, an average ratio of about 1:5.7 of VP2 to VP3 was observed by conducting densitometry analyses on western blots. Some VP1 also can be produced, but the level observed was typically less than 1% the level of VP2 that is produced. Accordingly, this construct can be used with a construct that will produce primarily VP1. See Examples 6 and 7. Production ratios can vary based upon experimental conditions and analytical techniques. A preferred IRES is the encephalomyocarditis virus (referred to as “EMCV” or “ECMV”) IRES.

EXAMPLE 5 – Intron IRES CAP IRES CAP

[0066] Embodiments of this construct are depicted in Figures 2 (constitutive) and 6 (controllable). This construct contains two IRES polynucleotides and two Cap polynucleotides. In a cell, both embodiments can produce a VP1, VP2 and VP3. The amount of VP3 production can be greater than the amount of VP1 production and the amount of VP2 production. The amount of VP2 production can be greater than the amount of VP1 production. In an experiment using CHO cells containing Cap5, an average ratio of about 1:2:9.3 of VP1 to VP2 to VP3 was observed by conducting densitometry analyses on western blots. Production ratios

can vary based upon experimental conditions and analytical techniques. A preferred IRES is the encephalomyocarditis virus (referred to as “EMCV” or “ECMV”) IRES.

EXAMPLE 6 – Intron CAP

[0067] Embodiments of this construct are depicted in Figures 3 (constitutive) and 7 (controllable). This construct contains an Intron and Cap polynucleotide, and produces predominantly VP1, and can be used with the constructs of Example 4 to produce VP1, VP2 and VP3.

Example 7 – IRES CAP

[0068] Embodiments of this construct are depicted in Figures 4 (constitutive) and 8 (controllable). This construct contains an Intron and Cap polynucleotide, and produces predominantly VP1, and can be used with the constructs of Example 4 to produce VP1, VP2 and VP3. A preferred IRES is the encephalomyocarditis virus (referred to as “EMCV” or “ECMV”) IRES.

Example 8 – CHO Cells Comprising Cap constructs

[0069] CHO cells of Example 1 can comprise an Intron IRES CAP IRES CAP polynucleotide of Example 5 (Figures 2 or 6). In an experiment, CHO cells containing Cap5 expressed VP1:VP2:VP3 in an observed ratio of about 1:2:9.3 by conducting densitometry analyses on western blots.

[0070] As an alternative, CHO cells of Example 1 can comprise CAP polynucleotides of Example 4 (Figs. 1 or 5) and Examples 6 (Figs. 3 or 7) or 6 (Figs. 4 and 8). Such cells would express VP1, VP2 and VP3, although possibly at different ratios than according to Example 5.

Example 9 – HEK 293 Cells Comprising Cap constructs

[0071] HEK 293 cells of Example 2 can comprise an Intron IRES CAP IRES CAP polynucleotide of Example 5 (Figures 2 or 6).

[0072] As an alternative, HEK 293 cells of Example 2 can comprise CAP polynucleotides of Example 4 (Figs. 1 or 5) and Examples 6 (Figs. 3 or 7) or 7 (Figs. 4 and 8). Such cells would express VP1, VP2 and VP3, although possibly at different ratios than according to Example 5.

EXAMPLE 10 – BHK Cells Comprising Cap constructs

[0073] BHK cells of Example 3 can comprise an Intron IRES CAP IRES CAP polynucleotide of Example 5 (Figures 2 or 6).

[0074] As an alternative, BHK cells of Example 3 can comprise CAP polynucleotides of Example 4 (Figs. 1 or 5) and Examples 6 (Figs. 3 or 7) or 7 (Figs. 4 and 8). Such cells would express VP1, VP2 and VP3, although possibly at different ratios than according to Example 5.

EXAMPLE 11 – AAV Polynucleotide Sequences

[0075] AAV Rep, Cap and ITR sequences are known in the art. The present inventions are amenable to all AAV serotypes. AAV sequences from various AAV serotypes are set forth below. Many of these sequences are available from the National Center for Biotechnology Information (NCBI).

AAV-1

Full Genome: NC_002077

CapVP1: (SEQ ID NO: 1)

ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCATTTCGCGAGTGGTGGGACTTG
AAACCTGGAGCCCCGAAGCCCAAGCCAACCAGCAAAAGCAGGACGACGGCCGGGGTCTGGTGCTTCCTGGCTAC
AAGTACCTCGGACCTTCAACGGACTCGACAAGGGGGAGCCCGTCAACGCGGCGGACGCAGCGGCCCTCGAGCAC
GACAAGGCCTACGACCAGCAGCTCAAAGCGGGTGACAATCCGTACCTGCGGTATAACCACGCCGACGCCGAGTTT
CAGGAGCGTCTGCAAGAAGATACGTCTTTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAGAAGCGGGTT

CTCGAACCTCTCGGTCTGGTTGAGGAAGGCGCTAAGACGGCTCCTGGAAAAGAAACGTCCGGTAGAGCAGTCGCCA
CAAGAGCCAGACTCCTCCTCGGGCATCGGCAAGACAGGCCAGCAGCCCCGCTAAAAAGAGACTCAATTTTGGTCAG
ACTGGCGACTCAGAGTCAGTCCCCGATCCACAACCTCTCGGAGAACCTCCAGCAACCCCCGCTGCTGTGGGACCT
ACTACAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAACGAAGGCGCCGACGGAGTGGGTAATGCCTCA
GGAAATTGGCATTGCGATTCCACATGGCTGGGCGACAGAGTCATCACCACCAGCACCCGCACCTGGGCTTGCCC
ACCTACAATAACCACCTCTACAAGCAAATCTCCAGTGCTTCAACGGGGGCCAGCAACGACAACCACTACTTCGGC
TACAGCACCCCTGGGGGTATTTTGATTTCAACAGATTCCACTGCCACTTTTCACCACGTGACTGGCAGCGACTC
ATCAACAACAATTGGGGATTCCGGCCCAAGAGACTCAACTTCAAACCTCTTCAACATCCAAGTCAAGGAGGTCACG
ACGAATGATGGCGTCACAACCATCGCTAATAACCTTACCAGCACGGTTCAAGTCTTCTCGGACTCGGAGTACCAG
CTTCCGTACGTCTCGGCTCTGCGCACCAAGGGCTGCCTCCCTCCGTTCGCGGCGACGTGTTTCATGATTCCGCAA
TACGGCTACCTGACGCTCAACAATGGCAGCCAAGCCGTGGGACGTTTCATCCTTTTACTGCCTGGAATATTTCCCT
TCTCAGATGCTGAGAACGGGCAACAACCTTTACCTTCAGCTACACCTTTGAGGAAGTGCCTTTCCACAGCAGCTAC
GCGCACAGCCAGAGCCTGGACCGGCTGATGAATCCTCTCATCGACCAATACCTGTATTACCTGAACAGACTCAA
AATCAGTCCGGAAGTGCCCAAAACAAGGACTTGCTGTTTACCGGTGGGTCTCCAGCTGGCATGTCTGTTTCAGCCC
AAAAACTGGCTACCTGGACCCTGTTATCGGCAGCAGCGCTTTCTAAAAACAAAAACAGACAACAACAACAGCAAT
TTTACCTGGACTGGTGCTTCAAAATATAACCTCAATGGGCGTGAATCCATCATCAACCTGGCACTGCTATGGCC
TCACACAAAGACGACGAAGACAAGTTCTTTCCCATGAGCGGTGTCATGATTTTTGGAAAAGAGAGCGCCGGAGCT
TCAAACACTGCATTGGACAATGTCATGATTACAGACGAAGAGGAAATTAAGCCACTAACCCGTGTGGCCACCGAA
AGATTTGGGACCGTGGCAGTCAATTTCCAGAGCAGCAGCACAGACCCTGCGACCGGAGATGTGCATGCTATGGGA
GCATTACCTGGCATGGTGTGGCAAGATAGAGACGTGTACCTGCAGGGTCCCATTTGGGCCAAAATTCCTCACACA
GATGGACACTTTTACCCGTCTCCTCTTATGGGCGCTTTGGACTCAAGAACCCGCCTCCTCAGATCCTCATCAAA
AACACGCCTGTTCTGCGAATCCTCCGGCGGAGTTTTAGCTACAAAGTTTGCTTCATTTCATCACCAATACTCC
ACAGGACAAGTGAGTGTGGAAATTGAATGGGAGCTGCAGAAAAGAAAACAGCAAGCGCTGGAATCCCGAAGTGCAG
TACACATCCAATTATGCAAAATCTGCCAACGTTGATTTTACTGTGGACAACAATGGACTTTTATACTGAGCCTCGC
CCCATTTGGCACCCGTTACCTTACCCGTCCCCTGTAA

Rep78: (SEQ ID NO: 2)

ATGCCGGGCTTCTACGAGATCGTGATCAAGGTGCCGAGCGACCTGGACGAGCACCTGCCGGGCATTTCTGACTCG
TTTGTGAGCTGGGTGGCCGAGAAGGAATGGGAGCTGCCCCCGGATTCTGACATGGATCTGAATCTGATTGAGCAG
GCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTCTGGTCCAATGGCGCCGCGTGAGTAAGGCCCCGGAG
GCCCTCTTCTTTGTTTCAGTTCGAGAAGGGCGAGTCTACTTCCACCTCCATATTCTGGTGGAGACCACGGGGGTC
AAATCCATGGTGCTGGGCCGCTTCTGAGTCAGATTAGGGACAAGCTGGTGCAGACCATCTACCGCGGGATCGAG
CCGACCCTGCCCACTGGTTTCGCGGTGACCAAGACGCGTAATGGCGCCGAGGGGGGAACAAGGTGGTGGACGAG
TGCTACATCCCCAACTACCTCCTGCCAAGACTCAGCCGAGCTGCAGTGGGCGTGGACTAACATGGAGGAGTAT
ATAAGCGCCTGTTTGAACCTGGCCGAGCGCAAACGGCTCGTGGCGCAGCACCTGACCCACGTCAGCCAGACCCAG
GAGCAGAACAAGGAGAATCTGAACCCCAATTCTGACGCGCCTGTATCCGGTCAAAAACCTCCGCGCGCTACATG
GAGCTGGTTCGGGTGGCTGGTGGACCGGGGCATCACCTCCGAGAAGCAGTGGATCCAGGAGGACCAGGCCTCGTAC
ATCTCCTTCAACGCCGCTTCCAACCTCGCGGTCCCAGATCAAGGCCGCTCTGGACAATGCCGGCAAGATCATGGCG
CTGACCAAATCCGCGCCCCGACTACCTGGTAGGCCCGCTCCGCCCGCGGACATTAAAACCAACCGCATCTACCGC
ATCCTGGAGCTGAACGGCTACGAACCTGCCTACGCCGGCTCCGTCTTTCTCGGCTGGGCCCAGAAAAGGTTCCGG
AAGCGCAACACCATCTGGCTGTTTGGGCCGGCCACCACGGGCAAGACCAACATCGCGGAAGCCATCGCCACGCC
GTGCCCTTCTACGGCTGCGTCAACTGGACCAATGAGAACTTTCCCTTCAATGATTGCGTCGACAAGATGGTGATC
TGGTGGGAGGAGGGCAAGATGACGGCCAAGGTCTGAGGTCCGCCAAGGCCATTCTCGGCGGCAGCAAGGTGCGC
GTGGACCAAAAAGTGCAAGTCTGTCGCGCCAGATCGACCCACCCCGTGATCGTCACCTCCAACACCAACATGTGC
CCCGTGATTGACGGGAACAGCACCCACCTTCGAGCACCAGCAGCCGTTGCAGGACCGGATGTTCAAATTTGAATC
ACCCGCCGTCTGGAGCATGACTTTGGCAAGGTGACAAAGCAGGAAAGTCAAAGAGTTCTTCCGTGGGCGCAGGAT
CACGTGACCGAGGTGGCGCATGAGTTCTACGTCAGAAAAGGTGGAGCCAACAAAAGACCCGCCCGCATGACGCG
GATAAAAGCGAGCCCAAGCGGGCCTGCCCTCAGTCGCGGATCCATCGACGTCAGACGCGGAAGGAGCTCCGGTG
GACTTTTGGCAGAGGTACCAAAACAAATGTTCTCGTCACGCGGGCATGCTTCAGATGCTGTTTCCCTGCAAGACA
TGCGAGAGAATGAATCAGAATTTCAACATTTGCTTCACGCACGGGACGAGAGACTGTTTCAGAGTGCTTCCCCGGC
GTGTCAGAATCTCAACCGGTCTGTCAGAAAGAGGACGTATCGGAAACTCTGTGCCATTTCATCATCTGCTGGGGCGG
GCTCCCGAGATTGCTTGCTCGGCCTGCGATCTGGTCAACGTGGACCTGGATGACTGTGTTTCTGAGCAATAA

AAV-2

Full Genome: NC_001401

Rep78: (SEQ ID NO: 3)

ATGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGCCCGGCATTTCTGACAGC
TTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCAGATTCTGACATGGATCTGAATCTGATTGAGCAG
GCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTTCTGACGGAATGGCGCCGTGTGAGTAAGGCCCGGAG
GCCCTTTTTCTTTGTGCAATTTGAGAAGGGGAGAGAGCTACTTCCACATGCACGTGCTCGTGGAAAACCACCGGGGTG
AAATCCATGGTTTTTGGGACGTTTCTGAGTCAGATTCGCGAAAAACTGATTTCAGAGAATTTACCGCGGGATCGAG
CCGACTTTGCCAAACTGGTTCGCGGTACAAAAGACCAGAAATGGCGCCGAGGCGGGAACAAGGTGGTGGATGAG
TGCTACATCCCCAATTACTTGTCCCCAAAACCCAGCCTGAGCTCCAGTGGGCGTGGACTAATATGGAACAGTAT
TTAAGCGCCTGTTTGAATCTCACGGAGCGTAAACGGTTGGTGGCGCAGCATCTGACGCACGTGTGCGAGACGCAG
GAGCAGAACAAAGAGAATCAGAATCCCAATTCTGATGCGCCGGTGATCAGATCAAAAACTTCAGCCAGGTACATG
GAGCTGGTCGGGTGGCTCGTGGACAAGGGGATTACCTCGGAGAAGCAGTGGATCCAGGAGGACCAGGCCTCATAC
ATCTCCTTCAATGCGGCCTCCAACTCGCGGTCCCAATCAAGGCTGCCTTGGACAATGCGGGAAAGATTATGAGC
CTGACTAAAACCGCCCCCGACTACCTGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTATAAA
ATTTTGGAACTAAACGGGTACGATCCCCAATATGCGGCTTCCGTCTTTCTGGGATGGGCCACGAAAAAGTTCGGC
AAGAGGAACACCATCTGGCTGTTTGGGCCTGCAACTACCGGGAAGACCAACATCGCGGAGGCCATAGCCACACT
GTGCCCTTCTACGGGTGCGTAAACTGGACCAATGAGAACTTTCCCTTCAACGACTGTGTGACAAAGATGGTGATC
TGGTGGGAGGAGGGGAAGATGACCGCCAAGGTCTGGAGTCGGCCAAAGCCATTCTCGGAGGAAGCAAGGTGCGC
GTGGACCAGAAATGCAAGTCCTCGGCCAGATAGACCCGACTCCCGTGATCGTCACCTCCAACACCAACATGTGC
GCCGTGATTGACGGGAACCTCAACGACCTTCGAACACCAGCAGCCGTGCAAGACCGGATGTTCAAATTTGAACTC
ACCCGCCGTCTGGATCATGACTTTGGGAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGAT
CACGTGGTTGAGGTGGAGCATGAATTCTACGTCAAAAAGGTTGGAGCCAAGAAAAGACCCGCCCCAGTGACGCA
GATATAAGTGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTGACGCGGAAGCTTCGATCAAC
TACGACAGACAGGTACCAAAAACAAATGTTCTCGTCACGTGGGCATGAATCTGATGCTGTTTCCCTGCAGACAATGC
GAGAGAATGAATCAGAATTCAAATATCTGCTTCACTCACGGACAGAAAAGACTGTTTAGAGTGCTTTCCCGTGTC
GAATCTCAACCCGTTTCTGTGCTCAAAAAGGCGTATCAGAACTGTGCTACATTATCATATCATGGGAAAGGTG
CCAGACGCTTGCCTGCGATCTGGTCAATGTGGATTGGATGACTGCATCTTTGAACAATAA

Rep52: (SEQ ID NO: 4)

ATGGAGCTGGTTCGGGTGGCTCGTGGACAAGGGGATTACCTCGGAGAAGCAGTGGATCCAGGAGGACCAGGCCTCA
TACATCTCCTTCAATGCGGCCTCCAACTCGCGGTCCCAATCAAGGCTGCCTTGGACAATGCGGGAAAGATTATG
AGCCTGACTAAAACCGCCCCCGACTACCTGGTGGGCCAGCAGCCGTGGAGGACATTTCCAGCAATCGGATTTAT
AAAATTTTGGAACTAAACGGGTACGATCCCCAATATGCGGCTTCCGTCTTTCTGGGATGGGCCACGAAAAAGTTC
GGCAAGAGGAACACCATCTGGCTGTTTGGGCCTGCAACTACCGGGAAGACCAACATCGCGGAGGCCATAGCCAC
ACTGTGCCCTTCTACGGGTGCGTAAACTGGACCAATGAGAACTTTCCCTTCAACGACTGTGTGACAAAGATGGTG
ATCTGGTGGGAGGAGGGGAAGATGACCGCCAAGGTCTGGAGTCGGCCAAAGCCATTCTCGGAGGAAGCAAGGTG
CGCGTGGACCAGAAATGCAAGTCCTCGGCCAGATAGACCCGACTCCCGTGATCGTCACCTCCAACACCAACATG
TGCGCCGTGATTGACGGGAACCTCAACGACCTTCGAACACCAGCAGCCGTTCGAAGACCGGATGTTCAAATTTGAA
CTCACCCGCCGTCTGGATCATGACTTTGGGAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAG
GATCACGTGGTTGAGGTGGAGCATGAATTCTACGTCAAAAAGGTTGGAGCCAAGAAAAGACCCGCCCCAGTGAC
GCAGATATAAGTGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTGACGCGGAAGCTTCGATC
AACTACGCAGACAGGTACCAAAAACAAATGTTCTCGTCACGTGGGCATGAATCTGATGCTGTTTCCCTGCAGACAA
TGCGAGAGAATGAATCAGAATTCAAATATCTGCTTCACTCACGGACAGAAAAGACTGTTTAGAGTGCTTTCCCGTG
TCAGAATCTCAACCCGTTTCTGTGCTCAAAAAGGCGTATCAGAACTGTGCTACATTATCATATCATGGGAAAG
GTGCCAGACGCTTGCCTGCGATCTGGTCAATGTGGATTGGATGACTGCATCTTTGAACAATAA

CapVP1: (SEQ ID NO: 5)

ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACACTCTCTCTGAAGGAATAAGACAGTGGTGGAAAGCTC
AAACCTGGCCACCACCACCAAGCCCGCAGAGCGGCATAAGGACGACAGCAGGGGTCTTGTGCTTCTGGGTAC
AAGTACCTCGGACCCTTCAACGGAATCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGCAC
GACAAAGCCTACGACCGGCAGCTCGACAGCGGAGACAACCCGTACCTCAAGTACAACCACGCCGACGCGGAGTTT
CAGGAGCGCCTTAAAGAAGATACGTCTTTTGGGGCAACCTCGGACGAGCAGTCTTCCAGGCAGAAAAGAGGGTT
CTTGAACCTCTGGGCCTGGTTGAGGAACCTGTTAAGACGGCTCCGGGAAAAAAGAGGCCGGTAGAGCACTCTCCT
GTGGAGCCAGACTCCTCCTCGGGAACCGGAAAGGCGGGCCAGCAGCTGCAAGAAAAGATTGAATTTTGGTTCAG
ACTGGAGACGCAGACTCAGTACCTGACCCCCAGCCTCTCGGACAGCCACCAGCAGCCCCCTCTGGTCTGGGAACT
AATACGATGGCTACAGGCAGTGGCGCACCAATGGCAGACAATAACGAGGGCGCCGACGGAGTGGGTAATTCCTCG

GGAAATTGGCATTGCGATTCCACATGGATGGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGCCC
ACCTACAACAACCACCTCTACAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACAATCACTACTTTGGCTAC
AGCACCCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCACCACGTGACTGGCAAAGACTCATC
AACAACAACCTGGGGATTCCGACCCAAGAGACTCAACTTCAAGCTCTTTAACATTCAAGTCAAAGAGGTACGCGAG
AATGACGGTACGACGACGATTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTC
CCGTACGTCTCTGGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCCAGCAGACGTCTTCATGGTGCCACAGTAT
GGATACCTCACCTGAACAACGGGAGTCAGGCAGTAGGACGCTCTTCATTTTACTGCCTGGAGTACTTTCTTTCT
CAGATGCTGCGTACCGGAAACAACCTTTACCTTCAGCTACACTTTTGAGGACGTTCCCTTTCCACAGCAGCTACGCT
CACAGCCAGAGTCTGGACCGTCTCATGAATCCTCTCATCGACCAGTACCTGTATTACTTGAGCAGAACAAACACT
CCAAGTGAACACCACGAGTCAAGGCTTCAGTTTTCTCAGGCCGGAGCGAGTGACATTCGGGACCAGTCTAGG
AACTGGCTTCTTGACCCCTGTTACCGCCAGCAGCGAGTATCAAAGACATCTGCGGATAACAACAACAGTGAATAC
TCGTGGACTGGAGCTACCAAGTACCCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCCGGCCATGGCAAGC
CACAAGGACGATGAAGAAAAGTTTTTCTCAGAGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACA
AATGTGGACATTGAAAAGGTTCATGATTACAGACGAAGAGGAAATCAGGACAACCAATCCCGTGGCTACGGAGCAG
TATGGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAAGCAGCTACCGCAGATGTCAACACACAAGGCGTT
CTTCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCCCATCTGGGCAAAGATTCCACACACGGAC
GGACATTTTACCCCTCTCCCTCATGGGTGGATTTCGGACTTAAACACCCCTCTCCACAGATTCTCATCAAGAAC
ACCCCGGTACCTGCGAATCCTTCGACCACCTTCAGTGCGGCAAAGTTTGCTTCCTTCATCACACAGTACTCCACG
GGACAGGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAAAACAGCAAACGCTGGAATCCCGAAATTCAGTAC
ACTTCCAACCTACAACAAGTCTGTTAATGTGGACTTTACTGTGGACACTAATGGCGTGTATTACAGAGCCTCGCCCC
ATTGGCACCAAGATACCTGACTCGTAATCTGTAA

CapVP2: (SEQ ID NO: 6)

ACGGCTCCGGGAAAAAAGAGGCCGGTAGAGCACTCTCCTGTGGAGCCAGACTCCTCCTCGGGAACCGGAAAGGCG
GGCCAGCAGCCTGCAAGAAAAAGATTGAATTTTGGTCAGACTGGAGACGCAGACTCAGTACCTGACCCCCAGCCT
CTCGGACAGCCACCAGCAGCCCCCTCTGGTCTGGGAACTAATACGATGGCTACAGGCAGTGGCGCACCAATGGCA
GACAATAACGAGGGCGCCGACGGAGTGGGTAATTCCTCGGGAAATTTGGCATTGCGATTCCACATGGATGGGCGAC
AGATCATCTACCACAGCAGCCGAACCTGGGCCCTGCCACCTACAACAACCACTCTACAAACAAATTTCCAGC
CAATCAGGAGCCTCGAACGACAATCACTACTTTGGCTACAGCACCCCTTGGGGGTATTTTGACTTCAACAGATTTC
CACTGCCACTTTTACCACGTGACTGGCAAAGACTCATCAACAACAACCTGGGGATTCCGACCCAAGAGACTCAAC
TTCAAGCTCTTTAACATTCAAGTCAAAGAGGTACGCGAGAATGACGGTACGACGACGATTGCCAATAACCTTACC
AGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTCCCGTACGTCTCCTCGGCTCGGCGCATCAAGGATGCCTC
CCGCCGTTCCAGCAGACGTCTTCATGGTGCCACAGTATGGATACCTCACCTGAACAACGGGAGTCAGGCAGTA
GGACGCTCTTCATTTTACTGCCTGGAGTACTTTCTTCTCAGATGTGCGTACCGGAAACAACCTTTACCTTCAGC
TACACTTTTGAGGACGTTCTTTCCACAGCAGCTACGCTCACAGCCAGAGTCTGGACCGTCTCATGAATCCTCTC
ATCGACCAGTACCTGTATTACTTGAGCAGAACAACACTCCAAGTGAACACCACGAGTCAAGGCTTCAGTTT
TCTCAGGCCGGAGCGAGTGACATTCGGGACCAGTCTAGGAACTGGCTTCTTGACCCCTGTTACCGCCAGCAGCGA
GTATCAAAGACATCTGCGGATAACAACAACAGTGAATACCTCGTGGACTGGAGCTACCAAGTACCACCTCAATGGC
AGAGACTCTCTGGTGAATCCGGGCCCGGCCATGGCAAGCCACAAGGACGATGAAGAAAAGTTTTTCTCAGAGC
GGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAAACAAATGTGGACATTGAAAAGGTTCATGATTACAGACGAA
GAGGAAATCAGGACAACCAATCCCGTGGCTACGGAGCAGTATGGTTCTGTATCTACCAACCTCCAGAGAGGCAAC
AGACAAGCAGTACCGCAGATGTCAACACACAAGGCGTTCTTCAGGCATGGTCTGGCAGGACAGAGATGTGTAC
CTTCAGGGGCCCATCTGGGCAAAGATTCCACACACGGACGGACATTTTACCCCTCTCCCTCATGGGTGGATTTC
GGACTTAAACACCCCTCTCCACAGATTCTCATCAAGAACACCCCGGTACCTGCGAATCCTTCGACCACCTTCAGT
GCGGCAAAGTTTGCTTCTTCATCACACAGTATCCACGGGACAGGTCAGCGTGGAGATCGATGGGAGCTGCAG
AAGGAAAACAGCAAACGCTGGAATCCCGAAATTCAGTACACTTCCAACCTACAACAAGTCTGTTAATGTGGACTTT
ACTGTGGACACTAATGGCGTGTATTACAGAGCCTCGCCCCATTGGCACCAAGATACCTGACTCGTAATCTGTAA

CapVP3: (SEQ ID NO:7)

ATGGCTACAGGCAGTGGCGCACCAATGGCAGACAATAACGAGGGCGCCGACGGAGTGGGTAATTCCTCGGGAAAT
TGGCATTGCGATTCCACATGGATGGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGCCCACCTAC
AACAACCACCTCTACAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACAATCACTACTTTGGCTACAGCACC
CCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCACCACGTGACTGGCAAAGACTCATCAACAAC
AACTGGGGATTCCGACCCAAGAGACTCAACTTCAAGCTCTTTAACATTCAAGTCAAAGAGGTACGCGAGAATGAC
GGTACGACGACGATTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTCCCGTAC
GTCCTCGGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCCAGCAGACGTCTTCATGGTGCCACAGTATGGATAC

CTCACCTGAACAACGGGAGTCAGGCAGTAGGACGCTCTTCATTTTACTGCCTGGAGTACTTTCTTCTCAGATG
CTGCGTACCGGAAACAACCTTTACCTTCAGCTACACTTTTGAGGACGTTCTTTCCACAGCAGCTACGCTCACAGC
CAGAGTCTGGACCGTCTCATGAATCCTCTCATCGACCAGTACCTGTATTACTTGAGCAGAACAAACACTCCAAGT
GGAACCACCACGCAGTCAAGGCTTCAGTTTTCTCAGGCCGGAGCGAGTGACATTCGGGACCAGTCTAGGAACCTGG
CTTCCTGGACCTGTTACCGCCAGCAGCGAGTATCAAAGACATCTGCGGATAACAACAACAGTGAATACTCGTGG
ACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCCCGGCCATGGCAAGCCACAAG
GACGATGAAGAAAAGTTTTTCTCAGAGCGGGGTTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACAAATGTG
GACATTGAAAAGGTCATGATTACAGACGAAGAGGAAATCAGGACAACCAATCCCGTGGCTACGGAGCAGTATGGT
TCTGTATCTACCAACCTCCAGAGAGGCAACAGACAAGCAGCTACCGCAGATGTCAACACACAAGGCGTTCTTCCA
GGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCCCATCTGGGCAAAGATTCCACACACGGACGGACAT
TTTACCCCTCTCCCTCATGGGTGGATTTCGGAATTAAACACCTCTCCACAGATTCTCATCAAGAACACCCCG
GTACCTGCGAATCCTTCGACCACCTTCAGTGCGGCAAAGTTTGCTTCTTCATCACACAGTACTCCACGGGACAG
GTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAAAACAGCAAACGCTGGAATCCCGAAATTCAGTACACTTCC
AACTACAACAAGTCTGTTAATGTGGACTTTACTGTGGACACTAATGGCGTGTATTAGAGCCTCGCCCCATTGGC
ACCAGATACCTGACTCGTAATCTGTAA

CapAAP: (SEQ ID NO:8)

CTGGAGACGCAGACTCAGTACCTGACCCCCAGCCTCTCGGACAGCCACCAGCAGCCCCCTCTGGTCTGGGAACTA
ATACGATGGCTACAGGCAGTGGCGCACCAATGGCAGACAATAACGAGGGCGCCGACGGAGTGGGTAATTCCTCGG
GAAATTGGCATTGCGATTCCACATGGATGGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGCCCA
CCTACAACAACCACCTCTACAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACAATCACTACTTTGGCTACA
GCACCCCTTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTACCACGTGACTGGCAAAGACTCATCA
ACAACAACCTGGGGATTCCGACCCAAGAGACTCAACTTCAAGCTCTTTAACATTCAAGTCAAAGAGGTACGCAGA
ATGACGGTACGACGACGATTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTCC
CGTACGTCTCTCGGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCACAGCAGACGTCTTCATGGTGCCACAGTATG
GATACCTCACCTGA

AAV-3

Full Genome: NC_001729

Rep78: (SEQ ID NO:9)

ATGCCGGGGTTCTACGAGATTGTCCTGAAGGTCCCGAGTGACCTGGACGAGCGCCTGCCGGGCATTTCTAACTCG
TTTGTTAACTGGGTGGCCGAGAAGGAATGGGACGTGCCGCCGATTCTGACATGGATCCGAATCTGATTGAGCAG
GCACCCCTGACCGTGGCCGAAAAGCTTCAGCGCGAGTTCCTGGTGGAGTGGCGCCGCGTGAGTAAGGCCCGGAG
GCCCTCTTTTTTGTCCAGTTCGAAAAGGGGGAGACCTACTTCCACCTGCACGTGCTGATTGAGACCATCGGGGTC
AAATCCATGGTGGTTCGGCCGCTACGTGAGCCAGATTAAAGAGAAGCTGGTGACCCGCATCTACCGCGGGTTCGAG
CCGCAGCTTCCGAACCTGGTTCGCGGTGACCAAAACGCGAAATGGCGCCGGGGGCGGGAACAAGGTGGTGGACGAC
TGCTACATCCCCAACTACCTGCTCCCCAAGACCCAGCCGAGCTCCAGTGGGCGTGGACTAACATGGACCAGTAT
TTAAGCGCCTGTTTGAATCTCGCGGAGCGTAAACGGCTGGTGGCGCAGCATCTGACGCACGTGTGCGACAGCGAG
GAGCAGAACAAAGAGAATCAGAACCCCAATTCTGACGCGCCGGTTCATCAGGTCAAAAACCTCAGCCAGGTACATG
GAGCTGGTTCGGGTGGCTGGTGGACCGCGGGATCACGTGAGAAAAGCAATGGATTACAGGAGGACCAGGCCTCGTAC
ATCTCCTTCAACGCCGCTCCAACCTCGCGGTCCCAGATCAAGGCCGCGCTGGACAATGCCCTCAAGATCATGAGC
CTGACAAAGACGGCTCCGGACTACCTGGTGGGCGAGCAACCCGCCGAGGACATTACCAAAAATCGGATCTACCAA
ATCCTGGAGCTGAACGGGTACGATCCGCAGTACGCGGCTCCGTCTTCTGGGCTGGGCGAAAAGAAGTTCGGG
AAGAGGAACACCATCTGGCTCTTTGGGCCGGCCACGACGGGTAAACCAACATCGCGGAAGCCATCGCCACGCC
GTGCCCTTCTACGGCTGCGTAAACTGGACCAATGAGAACTTTCCCTTCAACGATTGCGTGCACAAGATGGTGATC
TGGTGGGAGGAGGGCAAGATGACGGCCAAGGTCGTGGAGAGCGCCAAGGCCATTCTGGGCGGAAGCAAGGTGCGC
GTGGACCAAAAGTGCAAGTCATCGGCCAGATCGAACCCACTCCCGTGATCGTCACCTCCAACACCAACATGTGC
GCCGTGATTGACGGGAACAGCACCACTTCGAGCATCAGCAGCCGCTGCAGGACCGGATGTTTGAATTTGAACCTT
ACCCGCCGTTTGGACCATGACTTTGGGAAGGTCACCAAACAGGAAGTAAAGGACTTTTTCCGGTGGGCTTCCGAT
CACGTGACTGACGTGGCTCATGAGTCTACGTGAGAAAGGTTGGAGCTAAGAAACGCCCCGCTCCAATGACGCG
GATGTAAGCGAGCCAAAACGGGAGTGCACGTCACTTGCGCAGCCGACAACGTCAGACGCGGAAGCACCGGCGGAC
TACGCGGACAGGTACCAAAAACAAATGTTCTCGTCACGTGGGCATGAATCTGATGCTTTTTCCCTGTAAAACATGC
GAGAGAATGAATCAAATTTCCAATGTCTGTTTTACGCATGGTCAAAGAGACTGTGGGGAATGCTTCCCTGGAATG
TCAGAATCTCAACCCGTTTCTGTCTGTCAAAAAGAACTTATCAGAACTGTGTCCAATTCATCATATCCTGGGA
AGGGCACCCGAGATTGCCTGTTTCGGCCTGCGATTTGGCCAATGTGGACTTGGATGACTGTGTTTCTGAGCAATAA

CapVP1: (SEQ ID NO:10)

ATGGCTGCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTTTCTGAAGGCATTTCGTGAGTGGTGGGCTCTG
AAACCTGGAGTCCCTCAACCCAAAGCGAACCAACAACACCAGGACAACCGTCGGGGTCTTGTGCTTCCGGGTAC
AAATACCTCGGACCCGGTAACGGACTCGACAAAGGAGAGCCGGTCAACGAGGCGGACGCGGCAGCCCTCGAACAC
GACAAAGCTTACGACCAGCAGCTCAAGGCCGGTGACAACCCGTACCTCAAGTACAACCACGCCGACGCCGAGTTT
CAGGAGCGTCTTCAAGAAGATACGTCTTTTGGGGGCAACCTTGGCAGAGCAGTCTTCCAGGCCAAAAAGAGGATC
CTTGAGCCTCTTGGTCTGGTTGAGGAAGCAGCTAAAAACGGCTCCTGGAAAAGAGGGGGCTGTAGATCAGTCTCCT
CAGGAACCGGACTCATCATCTGGTGTGGCAAATCGGGCAAACAGCCTGCCAGAAAAAGACTAAATTTCCGTCAG
ACTGGAGACTCAGAGTCAGTCCCAGACCCCTCAACCTCTCGGAGAACCACCAGCAGCCCCACAAGTTTGGGATCT
AATACAATTGGCTTCAGGCGGTGGCGCACCAATGGCAGAGATCATCAACCACCAGCACCAGAACCTGGGCCCTGCC
GGAAATTGGCATTTCGATTCCCAATGGCTGGGCGACAGAGTCATCAACCACCAGCACCAGAACCTGGGCCCTGCC
ACTTACAACAACCATCTCTACAAGCAAATCTCCAGCCAATCAGGAGCTTCAAACGACAACCCTACTTTGGCTAC
AGCACCCCTTGGGGGTATTTTGAATTTAACAGATTCCACTGCCACTTCTCACCACGTGACTGGCAGCGACTCATT
AACAACAACCTGGGGATTCCGGCCCCAAGAACTCAGCTTCAAGCTCTTCAACATCCAAGTTAGAGGGGTACGCGAG
AACGATGGCAGCAGACTATTGCCAATAACCTTACCAGCACGGTTCAAGTGTTTACGGACTCGGAGTATCAGCTC
CCGTACGTGCTCGGGTCGGCGCACCAAGGCTGTCTCCCGCCGTTTCCAGCGGACGTCTTCATGGTCCCTCAGTAT
GGATACCTCACCTGAACAACGGAAGTCAAGCGGTGGGACGCTCATCTTTTACTGCCTGGAGTACTTCCCTTCG
CAGATGCTAAGGACTGGAAATAACTTCCAATTCAGCTATACCTTCGAGGATGTACCTTTTACAGCAGCTACGCT
CACAGCCAGAGTTTGGATCGCTTGATGAATCCTCTTATTGATCAGTATCTGTACTACCTGAACAGAACGCAAGGA
ACAACCTCTGGAACAACCAACCAATCACGGCTGCTTTTTAGCCAGGCTGGGCCTCAGTCTATGTCTTTCAGGCC
AGAAATTGGCTACCTGGGCCCTGTACCCGGCAACAGAGACTTTCAAAGACTGCTAACGACAACAACAACAGTAAC
TTTCTTGGACAGCGGCCAGCAAATATCATCTCAATGGCCGCGACTCGCTGGTGAATCCAGGACCAGCTATGGCC
AGTCACAAGGACGATGAAGAAAAATTTTTCCCTATGCACGGCAATCTAATATTTGGCAAAGAAGGGACAACGGCA
AGTAACGCAGAATTAGATAATGTAATGATTACGGATGAAGAAGAGATTTCGTACCACCAATCCTGTGGCAACAGAG
CAGTATGGAACGTGTGGCAAATAACTTGCAGAGCTCAAATACAGCTCCACGACTGGAACGTCAATCATCAGGGG
GCCTTACCTGGCATGGTGTGGCAAGATCGTGACGTGTACCTTCAAGGACCTATCTGGGCAAAGATTCTTCACACG
GATGGACACTTTTACCTTCTCCTCTGATGGGAGGCTTTGGACTGAAACATCCGCCTCCTCAAATCATGTGACAAA
AATATCCGGTACCGGCAAAATCCTCCGACGACTTTCAGCCCCGCGCAAGTTTGCTTCATTTTACTACTCAGTATCC
ACTGGACAGGTTCAGCGTGGAAATTGAGTGGGAGCTACAGAAAAGAAACAGCAAACGTTGGAATCCAGAGATTAG
TACACTTCCAACATAACAAGTCTGTTAATGTGGACTTTACTGTAGACACTAATGGTGTTTATAGTGAACCTCGC
CCTATTGGAACCCGGTATCTCACACGAAACTTGTGA

AAV-4

Full Genome: NC_001829

Rep78: (SEQ ID NO:11)

ATGCCGGGGTTCTACGAGATCGTGCTGAAGGTGCCCAGCGACCTGGACGAGCACCTGCCCGGCATTTCTGACTCT
TTTGTGAGCTGGGTGGCCGAGAAGGAATGGGAGCTGCCGCCGGATTCTGACATGGACTTGAATCTGATTGAGCAG
GCACCCCTGACCGTGGCCGAAAAGCTGCAACGCGAGTTCCTGGTCGAGTGGCGCCGCGTGAGTAAGGCCCGGAG
GCCCTCTTCTTTGTCCAGTTCGAGAAGGGGGACAGCTACTTCCACCTGCACATCCTGGTGGAGACCGTGGGCGTC
AAATCCATGGTGGTGGGCCGCTACGTGAGCCAGATTAAAGAGAAGCTGGTGACCCGCATCTACCGCGGGGTCGAG
CCGACACTTCCGAACCTGGTTTCGCGGTGACCAAGACGCGTAATGGCGCCGGAGGCGGGAACAAGGTGGTGACGAC
TGCTACATCCCCAACTACCTGCTCCCCAAGACCTTACGCCCGAGCTCCAGTGGGCGTGGACTAACATGGACAGTAT
ATAAGCGCCTGTTTGAATCTCGCGGAGCGTAAACGGCTGGTGGCGCAGCATCTGACGCACGTGTGCGAGACGCAG
GAGCAGAACAAGGAAAACAGAACCCCAATTCTGACGCGCCGGTTCATCAGGTCAAAAACCTCCGCCAGGTACATG
GAGCTGGTGGGTGGCTGGTGGACCGCGGGATCACGTGAGAAAAGCAATGGATCCAGGAGGACCAGGCGTCTTAC
ATCTCCTTCAACGCCGCTCCAACCTCGCGGTACAAAATCAAGGCCGCGCTGGACAATGCCCTCAAAATCATGAGC
CTGACAAAGACGGCTCCGGAATACCTGGTGGGCCAGAACCCGCCGAGGACATTTCCAGCAACCGCATCTACCGA
ATCCTCGAGATGAACGGGTACGATCCGAGTACGCGGCTCCGTCTTCTGGGCTGGGCGCAAAAGAAGTTCCGGG
AAGAGGAACACCATCTGGCTCTTTGGGCCGGCCACGACGGGTAAACCAACATCGCGGAAGCCATCGCCACGCC
GTGCCCTTCTACGGCTGCGTGAACCTGGACCAATGAGAACTTCCGTTCAACGATTGCGTCGACAAGATGGTGATC
TGGTGGGAGGAGGGCAAGATGACGGCCAAGGTCGTAGAGAGCGCCAAGGCCATCCTGGGCGGAAGCAAGGTGCGC
GTGGACCAAAAGTGCAAGTCATCGGCCAGATCGACCCAACTCCCGTGATCGTCACCTCCAACACCAACATGTGC
GCGGTATCGACGGAACCTCGACACCTTCGAGCACCAACAACCACTCCAGGACCGGATGTTCAAGTTCGAGCTC

ACCAAGCGCCTGGAGCACGACTTTGGCAAGGTCACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCGTCAGAT
CACGTGACCGAGGTGACTCACGAGTTTTACGTCAGAAAGGGTGGAGCTAGAAAGAGGCCCCCCCCAATGACGCA
GATATAAGTGAGCCCAAGCGGGCCTGTCCGTGAGTTGCGCAGCCATCGACGTCAGACGCGGAAGCTCCGGTGGAC
TACGCGGACAGGTACCAAAACAAATGTTCTCGTCACGTGGGTATGAATCTGATGCTTTTTCCCTGCCGGCAATGC
GAGAGAATGAATCAGAATGTGGACATTTGCTTCACGCACGGGGTCATGGACTGTGCCGAGTGCTTCCCCGTGTCA
GAATCTCAACCCGTGTCTGTCTCGTCAGAAAGCGGACGTATCAGAACTGTGTCCGATTTCATCACATCATGGGGAGG
GCGCCCCGAGGTGGCCTGTCTGGCCTGCGAACTGGCCAATGTGGACTTGGATGACTGTGACATGGAACAATAA

CapVP1: (SEQ ID NO:12)

ATGACTGACGGTTACCTTCCAGATTGGCTAGAGGACAACCTCTCTGAAGGCGTTCGAGAGTGGTGGGCGCTGCAA
CCTGGAGCCCCCTAAACCCAAGGCAAATCAACAACATCAGGACAACGCTCGGGGTCTTGTGCTTCCGGGTTACAAA
TACCTCGGACCCGGCAACGGACTCGACAAGGGGGAACCCGTCAACGCAGCGGACGCGGCAGCCCTCGAGCACGAC
AAGGCCTACGACCAGCAGCTCAAGGCCGGTGACAACCCCTACCTCAAGTACAACCACGCCGACGCGGAGTTCCAG
CAGCGGCTTCAGGGCGACACATCGTTTGGGGGCAACCTCGGCAGAGCAGTCTTCCAGGCCAAAAAGAGGGTTCTT
GAACCTCTTGGTCTGGTTGAGCAAGCGGGTGGAGCGGCTCCTGGAAAAGAGAGACCGTTGATTGAATCCCCCAG
CAGCCCCGACTCCTCCACGGGTATCGGCAAAAAAGGCAAGCAGCCGGCTAAAAAGAGCTCGTTTTTCGAAGACGAA
ACTGGAGCAGGCGACGACACCCCTGAGGGATCAACTTCCGGAGCCATGTCTGATGACAGTGAGATGCGTGCAGCA
GCTGGCGGAGCTGCAGTCGAGGGCGGACAAGGTGCCGATGGAGTGGGTAATGCCCTCGGGTGATTGGCATTGCGAT
TCCACCTGGTCTGAGGGCCACGTACGACCACCAGCACCAGAACCTGGGTCTTGCCACCTACAACAACCACCTC
TACAAGCGACTCGGAGAGAGCCTGCAGTCCAACACCTACAACGATTCTCCACCCCTGGGGATACTTTGACTTC
AACCGCTTCCACTGCCACTTCTCACCACGTGACTGGCAGCGACTCATCAACAACAACTGGGGCATGCGACCCAAA
GCCATGCGGGTCAAAATCTTCAACATCCAGGTCAAGGAGGTACAGACGTCGAACGCGCAGACAACGGTGGCTAAT
AACCTTACCAGCACGGTTCAGATCTTTGCGGACTCGTCGTACGAACTGCCGTACGTGATGGATGCGGGTCAAGAG
GGCAGCCTGCCTCCTTTTCCCAACGACGTCTTTATGGTGCCCCAGTACGGCTACTGTGGACTGGTGACCGGCAAC
ACTTCGCAGCAACAGACTGACAGAAATGCCTTCTACTGCCTGGAGTACTTTCCTTCGCAGATGCTGCGGACTGGC
AACAACCTTTGAAATTACGTACAGTTTTGAGAAGGTGCCTTTCCTACTCGATGTACGCGCACAGCCAGAGCCTGGAC
CGGCTGATGAACCTCTCATCGACCAGTACCTGTGGGGACTGCAATCGACCACCACCGGAACCACCTGAATGCC
GGGACTGCCACCACCACTTTACCAAGCTGCGGCCTACCAACTTTTCCAACCTTTAAAAAGAACTGGCTGCCCGGG
CCTTCAATCAAGCAGCAGGGCTTCTCAAAGACTGCCAATCAAACTACAAGATCCCTGCCACCGGGTCAGACAGT
CTCATCAAATACGAGACGCACAGCACTCTGGACGGAAGATGGAGTGCCCTGACCCCGGACCTCCAATGGCCACG
GCTGGACCTGCGGACAGCAAGTTTACGCAACAGCCAGCTCATCTTTGCGGGGCCCTAAACAGAACGGCAACACGGCC
ACCGTACCCGGGACTCTGATCTTACCTCTGAGGAGGAGCTGGCAGCCACCAACGCCACCGATACGGACATGTGG
GGCAACCTACCTGGCGGTGACCAGAGCAACAGCAACCTGCCGACCGTGGACAGACTGACAGCCTTGGGAGCCGTG
CCTGGAATGGTCTGGCAAAACAGAGACATTTACTACCAGGGTCCCATTGCGGCAAGATTTCCTCATACCGATGGA
CACTTTTACCCCTCACCGCTGATTGGTGGGTTTGGGCTGAAACACCCGCCCTCCTCAAATTTTTATCAAGAACACC
CCGGTACCTGCGAATCCTGCAACGACCTTCAGCTCTACTCCGGTAACTCCTTCATTACTCAGTACAGCACTGGC
CAGGTGTCGGTGAGATTGACTGGGAGATCCAGAAGGAGCGGTCCAACGCTGGAACCCCGAGGTCCAGTTTACC
TCCAACCTACGACAGCAAACTCTCTGTGTGGGCTCCCGATGCGGCTGGGAAATACACTGAGCCTAGGGCTATC
GGTACCCGCTACCTCACCCACCACCTGTAA

AAV-5

Full Genome: NC_006152

Rep78: (SEQ ID NO:13)

ATGGCTACCTTCTATGAAGTCATTGTTTCGCGTCCCATTGACGTGGAGGAACATCTGCCTGGAATTTCTGACAGC
TTTGTGGACTGGGTAACCTGGTCAAATTTGGGAGCTGCCTCCAGAGTCAGATTTAAATTTGACTCTGGTTGAACAG
CCTCAGTTGACGGTGGCTGATAGAATTCGCCGCGTGTTCCTGTACGAGTGGAACAAATTTTCCAAGCAGGAGTCC
AAATTCTTTGTGCAGTTTGAAAAGGGATCTGAATATTTTCATCTGCACACGCTTGTGGAGACCTCCGGCATCTCT
TCCATGGTCTCTCGCCGCTACGTGAGTCAGATTTCGCGCCAGCTGGTGAAAGTGGTCTTCCAGGGAATTGAACCC
CAGATCAACGACTGGGTGCGCATACCAAGGTAAAGAAGGGCGGAGCCAATAAGGTGGTGGATTCTGGGTATATT
CCCGCCTACCTGCTGCCGAAGGTCCAACCGGAGCTTCAGTGGGCGTGGACAAACCTGGACGAGTATAAATTGGCC
GCCCTGAATCTGGAGGAGCGCAAACGGCTCGTCGCGCAGTTTCTGGCAGAATCCTCGCAGCGCTCGCAGGAGGCG
GCTTCGCAGCGTGAGTTCTCGGCTGACCCGGTCATCAAAAGCAAGACTTCCCAGAAATACATGGCGCTCGTCAAC
TGGCTCGTGGAGCACGGCATCACTTCGAGAAGCAGTGATCCAGGAAAAACAGGAGAGCTACCTCTCCTTCAAC
TCCACCGGCAACTCTCGGAGCCAGATCAAGGCCGCGCTCGACAACGCGACCAAAATTATGAGTCTGACAAAAAGC

GCGGTGGACTACCTCGTGGGGAGCTCCGTTCCCGAGGACATTTCAAAAAACAGAATCTGGCAAATTTTTGAGATG
 AATGGCTACGACCCGGCTACGCGGGATCCATCCTCTACGGCTGGTGTACGCGCTCCTTCAACAAGAGGAACACC
 GTCTGGCTCTACGGACCCGCCACGACCGGCAAGACCAACATCGCGGAGGCCATCGCCACACTGTGCCCTTTTAC
 GGCTGCGTGAACGGACCAATGAAAACCTTTCCCTTTAATGACTGTGTGGACAAAATGCTCATTGTGGTGGGAGGAG
 GGAAAGATGACCAACAAGGTGGTTGAATCCGCCAAGGCCATCCTGGGGGGCTCAAAGGTGCGGGTCGATCAGAAA
 TGTAATCCTCTGTTCAAATTGATTCTACCCCTGTCAATTGTAACCTCCAATACAAACATGTGTGTGGTGGTGGAT
 GGAATTCACGACCTTTGAACACCAGCAGCCGCTGGAGGACCGCATGTTCAAATTTGAACTGACTAAGCGGCTC
 CCGCCAGATTTTGGCAAGATTACTAAGCAGGAAGTCAAGGACTTTTTTGTCTGGGCAAAGGTCAATCAGGTGCCG
 GTGACTCACGAGTTTAAAGTTCCAGGGAATTGGCGGGAACATAAGGGGCGGAGAAATCTCTAAAACGCCACTG
 GGTGACGTACCAATACTAGCTATAAAAGTCTGGAGAAGCGGGCCAGGCTCTCATTGTTCCCGAGACGCCCTCGC
 AGTTCAGACGTGACTGTTGATCCCGCTCCTCTGCGACCGCTCAATTGGAATTCAAGGTATGATTGCAAATGTGAC
 TATCATGCTCAATTTGACAACATTTCTAACAATGTGATGAATGTGAATATTTGAATCGGGGCAAAAATGGATGT
 ATCTGTCACAATGTAACCTACTGTCAAATTTGTCAATGGGATTCCCCCTGGGAAAAGGAAAACCTTGTGAGATTT
 GGGGATTTTGACGATGCCAATAAAGAACAGTAA

CapVP1: (SEQ ID NO:14)

ATGTCTTTTGTGATCACCTCCAGATTGGTTGGAAGAAGTTGGTGAAGGTCTTCGCGAGTTTTTGGGCCTTGAA
 GCGGGCCCCACCGAAACCAAAACCAATCAGCAGCATCAAGATCAAGCCCGTGGTCTTGTGCTGCCTGGTTATAAC
 TATCTCGGACCCGAAACGGTCTCGATCGAGGAGAGCCTGTCAACAGGGCAGACGAGGTGCGCGAGAGCACGAC
 ATCTCGTACAACGAGCAGCTTGAGGCGGGAGACAACCCCTACCTCAAGTACAACCACGCGGACGCCGAGTTTCAG
 GAGAAGCTCGCCGACGACACATCCTTCGGGGGAAACCTCGGAAAGGCAGTCTTTCAGGCCAAGAAAAGGGTTCTC
 GAACCTTTTGGCCTGGTTGAAGAGGGTGCTAAGACGGCCCCCTACCGGAAAGCGGATAGACGACCACCTTCCAAAA
 AGAAAGAAGGCTCGGACCGAAGAGGACTCCAAGCCTTCCACCTCGTCAGACGCCGAAGCTGGACCCAGCGGATCC
 CAGCAGCTGCAAATCCAGCCCAACCAGCCTCAAGTTTGGGAGCTGATACAATGTCTGCGGGAGGTGGCGGGCCA
 TTGGGCGACAATAACCAAGGTGCCGATGGAGTGGGCAATGCCCTCGGGAGATTGGCATTGCGATTCCACGTGGATG
 GGGGACAGAGTCGTACCAAGTCCACCCGAACCTGGGTGCTGCCAGCTACAACAACCACAGTACCGAGAGATC
 AAAAGCGGCTCCGTCGACGGAAGCAACGCCAACGCCTACTTTGGATACAGCACCCCTTGGGGGTACTTTGACTTT
 AACCGCTTCCACAGCCACTGGAGCCCCCGAGACTGGCAAAGACTCATCAACAATACTGGGGCTTCAGACCCCGG
 TCCCTCAGAGTCAAAATCTTCAACATTCAAGTCAAAGAGGTACGGTGCAGGACTCCACCACCACCATCGCCAAC
 AACCTCACCTCCACCGTCCAAGTGTTTACGGACGACGACTACCAGCTGCCCTACGTGCTCGGCAACGGGACCGAG
 GGATGCCTGCCGGCCTTCCCTCCGAGGTCTTTACGCTGCCGAGTACGGTTACGCGACGCTGAACCGCGACAAC
 ACAGAAAATCCCACCGAGAGGAGCAGCTTCTTCTGCCTAGAGTACTTTCCCAGCAAGATGCTGAGAACGGGCAAC
 AACTTTGAGTTTACCTACAACCTTTGAGGAGGTGCCCTTCCACTCCAGCTTCGCTCCCAGTCAGAACCTGTTCAAG
 CTGGCCAACCCGCTGGTGGACAGTACTTGTACCGCTTCGTGAGCACAAATAACACTGGCGGAGTCCAGTTCAAC
 AAGAACCTGGCCGGGAGATACGCCAACACCTACAAAACTGGTTCCCGGGGCCCATGGGCCGAACCCAGGGCTGG
 AACCTGGGCTCCGGGGTCAACCGCGCCAGTGTGAGCGCTTCGCCACGACCAATAGGATGGAGCTCGAGGGCGCG
 AGTTACAGGTGCCCCCGCAGCCGAACGGCATGACCAACAACCTCCAGGGCAGCAACACCTATGCCCTGGAGAAC
 ACTATGATCTTCAACAGCCAGCCGGCGAACCCGGGCACCAACCGCCACGTACCTCGAGGGCAACATGCTCATCACC
 AGCGAGAGCGAGACGACGCGCGGTGAACCGCGTGGCGTACAACGTGGCGGGCAGATGGCCACCAACAACAGAGC
 TCCACCACTGCCCCCGCAGCCGGCACGTACAACCTCCAGGAAATCGTGCCCGGCAGCGTGTGGATGGAGAGGGAC
 GTGTACCTCCAAGGACCCATCTGGGCCAAGATCCAGAGACGGGGGCGCACTTTACCCCTCTCCGGCCATGGGC
 GGATTTCGACTCAAACACCCACCGCCCATGATGCTCATCAAGAACACGCTGTGCCCGGAAATATCACCAGCTTC
 TCGGACGTGCCCGTCAGCAGCTTCATCACCCAGTACAGCACCGGGCAGGTCACCGTGGAGATGGAGTGGGAGCTC
 AAGAAGGAAAACCTCAAGAGGTGGAACCCAGAGATCCAGTACACAAACAATAACGACCCCCAGTTTGTGGAC
 TTTGCCCGGACAGACCCGGGAATACGAACCAAGAGCTATCGGAACCCGATACCTTACCCGACCCCTTTAA

AAV-6

Full Genome: AF028704

Rep78: (SEQ ID NO:15)

ATGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGCCCCGCATTTCTGACAGC
 TTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCAGATTCTGACATGGATCTGAATCTGATTGAGCAG
 GCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTCTGGTCCAGTGGCGCCGCGTGAGTAAGGCCCGGAG
 GCCCTCTTCTTTGTTTCAAGTTCGAGAAGGGCGAGTCTTACTTCCACCTCCATATTTCTGGTGGAGACCACGGGGGTC
 AAATCCATGGTGTGGGCCGCTTCTGAGTCAGATTAGGGACAAGCTGGTGCAGACCATCTACCGCGGGATCGAG
 CCGACCCCTGCCCAACTGGTTTCGCGGTGACCAAGACGCGTAATGGCGCCGGAGGGGGGAACAAGGTGGTGGACGAG

TGCTACATCCCCAACTACCTCCTGCCCAAGACTCAGCCCGAGCTGCAGTGGGCGTGGACTAACATGGAGGAGTAT
ATAAGCGCGTGTTTTAAACCTGGCCGAGCGCAAACGGCTCGTGGCGCACGACCTGACCCACGTCAGCCAGACCCAG
GAGCAGAACAAGGAGAATCTGAACCCCAATTCTGACGCGCCTGTATCCGGTCAAAAACCTCCGCACGCTACATG
GAGCTGGTCGGGTGGCTGGTGGACCGGGGCATCACCTCCGAGAAGCAGTGGATCCAGGAGGACCAGGCCCTCGTAC
ATCTCCTTCAACGCCGCTCCAACCTCGCGGTCCCAGATCAAGGCCGCTCTGGACAATGCCGGCAAGATCATGGCG
CTGACCAAAATCCGCGCCCCGACTACCTGGTAGGCCCGCTCCGCCCGCCGACATTAAAACCAACCGCATTTACCGC
ATCCTGGAGCTGAACGGCTACGACCTTGCTACGCCGGCTCCGCTCTTTCTCGGCTGGGCCAGAAAAGGTTTCGGA
AAACGCAACACCATCTGGCTGTTTGGGCCGGCCACCACGGGCAAGACCAACATCGCGGAAGCCATCGCCACGCC
GTGCCCTTCTACGGCTGCGTCAACTGGACCAATGAGAACTTTCCCTTCAACGATTGCGTCGACAAGATGGTGATC
TGGTGGGAGGAGGGCAAGATGACGGCCAAGGTCGTGGAGTCCGCCAAGGCCATTCTCGGCGGCAGCAAGGTGCGC
GTGGACCAAAAGTGCAAGTCGTCCGCCCAGATCGATCCACCCCCGTGATCGTCACCTCCAACACCAACATGTGC
GCCGTGATTGACGGGAACAGCACCACTTCGAGCACCAGCAGCCGTTGCAGGACCGGATGTTCAAATTTGAATC
ACCCGCCGTCTGGAGCATGACTTTGGCAAGGTGACAAAGCAGGAAGTCAAAGAGTTCTTCCGCTGGGCGCAGGAT
CACGTGACCGAGGTGGCGCATGAGTTCTACGTGAGAAAGGTTGGAGCCAACAAGAGACCCGCCCCGATGACGCG
GATAAAAGCGAGCCCAAGCGGGCTGCCCTCAGTCGCGGATCCATCGACGTGACGCGGAAGGAGCTCCGGTG
GACTTTGCCGACAGGTACCAAAACAAATGTTCTCGTCACGCGGGCATGCTTCAGATGCTGTTTCCCTGCAAAACA
TGCGAGAGAATGAATCAGAATTTCAACATTTGCTTCACGCACGGGACCAGAGACTGTTTCAAGATGTTTCCCGGC
GTGTGAGAATCTCAACCGGTGTCGAGAAAGAGGACGTATCGGAAACTCTGTGCCATTATCATCTGCTGGGGCGG
GCTCCCGAGATTGCTTGCTCGGCCTGCGATCTGGTCAACGTGGATCTGGATGACTGTGTTTCTGAGCAATAA

CapVP1: (SEQ ID NO:16)

ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCATTTCGCGAGTGGTGGGACTTG
AAACCTGGAGCCCCGAAACCCAAAGCCAACCAGCAAAAAGCAGGACGACGGCCGGGGTCTGGTGCTTCTGGCTAC
AAGTACCTCGGACCTTCAACGGACTCGACAAGGGGGAGCCCGTCAACGCGGCGGATGCAGCGGCCCTCGAGCAC
GACAAGGCCTACGACCAGCAGCTCAAAGCGGGTGACAATCCGTACCTGCGGTATAACCACGCCGACGCCGAGTTT
CAGGAGCGTCTGCAAGAAGATACGTCTTTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAGAAGAGGGTT
CTCGAACCTTTTGGTCTGGTTGAGGAAGGTGCTAAGACGGCTCCTGGAAAGAAACGTCCGGTAGAGCAGTCGCCA
CAAGAGCCAGACTCCTCCTCGGGCATTGGCAAGACAGGCCAGCAGCCCGCTAAAAAGAGACTCAATTTTGGTCAG
ACTGGCGACTCAGAGTCAGTCCCCGACCCACAACCTCTCGGAGAACCTCCAGCAACCCCCGCTGCTGTGGGACCT
ACTACAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAACGAAGGCGCCGACGGAGTGGGTAATGCCTCA
GGAAATTGGCATTGCGATTCCACATGGCTGGGCGACAGAGTCATCACCACCAGCACCCGAACATGGGCCCTTGCCC
ACCTATAACAACCACCTCTACAAGCAAATCTCCAGTGCTTCAACGGGGGCCAGCAACGACAACCACTACTTCGGC
TACAGCACCCCTGGGGGTATTTTGATTTCACAGATTCCACTGCCATTTCTCACCACGTGACTGGCAGCGACTC
ATCAACAACAATTGGGGATTCCGGCCCAAGAGACTCAACTTCAAGCTCTTCAACATCCAAGTCAAGGAGGTCACG
ACGAATGATGGCGTCACGACCATCGCTAATAACCTTACCAGCACGGTTCAAGTCTTCTCGGACTCGGAGTACCAG
TTGCCGTACGTCTCGGCTCTGCGCACCAAGGGCTGCCTCCCTCCGTTCGCGGCGGACGTGTTTCATGATTCCGCAG
TACGGCTACCTAACGCTCAACAATGGCAGCCAGGCAGTGGGACGGTCATCCTTTTACTGCCTGGAATATTTCCCA
TCGAGATGCTGAGAACGGGCAATAACTTTACCTTCAGCTACACCTTCGAGGACGTGCCTTTCCACAGCAGCTAC
GCGCACAGCCAGAGCCTGGACCGGCTGATGAATCCTCTCATCGACCAGTACCTGTATTACCTGAACAGAACTCAG
AATCAGTCCGGAAGTGCCCAAAACAAGGACTTGCTGTTTAGCCGGGGGTCTCCAGCTGGCATGTCTGTTTCAGCCC
AAAAACTGGCTACCTGGACCTGTTACCGGCAGCAGCGGTTTCTAAAAACAAAAACAGACAACAACAGCAAC
TTTACCTGGACTGGTGCTTCAAAATATAACCTTAATGGGCGTGAATCTATAATCAACCTGGCACTGCTATGGCC
TCACACAAAGACGACAAAGACAAGTTCTTTCCCATGAGCGGTGTCATGATTTTTGGAAAGGAGAGCGCCGGAGCT
TCAAACACTGCATTGGACAATGTATGATCACAGACGAAGAGGAAATCAAAGCCACTAACCCCGTGCCACCCGAA
AGATTTGGGACTGTGGCAGTCAATCTCCAGAGCAGCAGCAGACCCCTGCGACCGGAGATGTGCATGTTATGGGA
GCCTTACCTGGAATGGTGTGGCAAGACAGAGACGTATACCTGACGGGTCTTATTTGGGCCAAAATTCCTACACG
GATGGACACTTTTACCCGTCTCCTCTCATGGGCGGCTTTGGACTTAAGCACCCGCTCCTCAGATCCTCATCAAA
AACACGCTGTTCTGCGAATCCTCCGGCAGAGTTTTCGGCTACAAAGTTTGCTTCATTCATCACCAGTATTCC
ACAGGACAAGTGAGCGTGAGATTGAATGGGAGCTGCAGAAAGAAAAACAGCAAACGCTGGAATCCCGAAGTGCAG
TATACATCTAACTATGCAAAATCTGCCAACGTTGATTTTACTGTGGACAACAATGGACTTTTATACTGAGCCTCGC
CCCATTTGGCACCCGTTACCTCACCCGTCCCCTGTAA

AAV-7

Full Genome: NC_006260

Rep78: (SEQ ID NO:17)

ATGCCGGGTTTCTACGAGATCGTGATCAAGGTGCCGAGCGACCTGGACGAGCACCTGCCGGGCATTTCTGACTCG
TTTGTGAAGTGGGTGGCCGAGAAGGAATGGGAGCTGCCCCGGATTCTGACATGGATCTGAATCTGATCGAGCAG
GCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTCTGGTCCAATGGCGCCGCGTGAGTAAGGCCCGGAG
GCCCTGTTCTTTGTTTCAGTTCGAGAAGGGCGAGAGCTACTTCCACCTTACGTTCTGGTGGAGACCACGGGGGTC
AAGTCCATGGTGCTAGGCCGCTTCTGAGTCAGATTGCGGAGAAGCTGGTCCAGACCATCTACCGCGGGGTCGAG
CCCACGCTGCCCAACTGGTTCGCGGTGACCAAGACGCGTAATGGCGCCGGCGGGGGGAACAAGGTGGTGGACGAG
TGCTACATCCCCAACTACCTCTGCCAAGACCCAGCCGAGCTGCAGTGGGCGTGGACTAACATGGAGGAGTAT
ATAAGCGCGTGTTTGAACCTGGCCGAACGCAAACGGCTCGTGGCGCAGCACCTGACCCACGTCAGCCAGACGCAG
GAGCAGAACAAGGAGAATCTGAACCCCAATTCTGACGCGCCCGTGATCAGGTCAAAAACCTCCGCGCGCTACATG
GAGCTGGTCGGGTGGCTGGTGGACCGGGGCATCACCTCCGAGAAGCAGTGGATCCAGGAGGACCAGGCCTCGTAC
ATCTCCTTCAACGCCGCTTCAACTCGCGGTCCCAGATCAAGGCCGCGCTGGACAATGCCGGCAAGATCATGGCG
CTGACCAATCCGCGCCCGACTACCTGGTGGGGCCCTCGCTGCCCGCGGACATTAAACCAACCGCATCTACCGC
ATCCTGGAGCTGAACGGGTACGATCCTGCCTACGCCGGCTCCGTCTTTCTCGGCTGGGCCAGAAAAAGTTCGGG
AAGCGCAACACCATCTGGCTGTTTGGGCCCCGCCACCACCGGCAAGACCAACATTGCGGAAGCCATCGCCACGCC
GTGCCCTTCTACGGCTGCGTCAACTGGACCAATGAGAACTTTCCCTTCAACGATTGCGTCGACAAGATGGTGATC
TGGTGGGAGGAGGGCAAGATGACGGCCAAGGTCTGGAGTCCGCCAAGGCCATTCTCGGCGGCAGCAAGGTGCGC
GTGGACCAAAAGTGCAAGTCGTCCGCCAGATCGACCCACCCCCGTGATCGTCACCTCCAACACCAACATGTGC
GCCGTGATTGACGGGAACAGCACCACTTTCGAGCACCAGCAGCCGTTGCAGGACCGGATGTTCAAATTTGAACCTC
ACCCGCCGTCTGGAGCAGCACTTTGGCAAGGTGACGAAGCAGGAAGTCAAAGAGTTCTTCCGCTGGGCCAGTGAT
CACGTGACCGAGGTGGCGCATGAGTTCTACGTCAGAAAGGGCGGAGCCAGCAAAAGACCCGCCCCCGATGACGCG
GATATAAGCGAGCCCAAGCGGGCCTGCCCTCAGTCGCGGATCCATCGACGTCAGACGCGGAAGGAGCTCCGGTG
GACTTTGCCGACAGGTACCAAAACAAATGTTCTCGTCACGCGGGCATGATTCAGATGCTGTTTCCCTGCAAAACG
TGCGAGAGAATGAATCAGAAATTTCAACATTTGCTTCACACACGGGGTCAGAGACTGTTTAGAGTGTTTCCCCGGC
GTGTGAGAATCTCAACCGGTCTGAGAAAAAGACGTATCGGAAACTCTGCGCGATTATCATCTGCTGGGGCGG
GCGCCCAGATTGCTTGCTCGGCCTGCGACCTGGTCAACGTGGACCTGGACGACTGCGTTTCTGAGCAATAA

CapVP1: (SEQ ID NO:18)

ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCATTCGCGAGTGGTGGGACCTG
AAACCTGGAGCCCCGAAACCCAAAGCCAACCAGCAAAAGCAGGACAACGGCCGGGGTCTGGTGCTTCCCTGGCTAC
AAGTACCTCGGACCCCTTCAACGGACTCGACAAGGGGGGAGCCCGTCAACGCGGCGGACGCAGCGGCCCTCGAGCAC
GACAAGGCCTACGACCAGCAGCTCAAAGCGGGTGACAATCCGTACCTGCGGTATAACCACGCCGACGCCGAGTTT
CAGGAGCGTCTGCAAGAAGATACGTCAATTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAGAAGCGGGTT
CTCGAACCTCTCGGTCTGGTTGAGGAAGGCGTAAGACGGCTCTTGCAAAAGAGAGACCGGTAGAGCCGTCACCT
CAGCGTTCCCCCGACTCCTCCACGGGCATCGGCAAGAAAGGCCAGCAGCCCGCCAGAAAGAGACTCAATTTCCGT
CAGACTGGCGACTCAGAGTCAGTCCCCGACCCTCAACCTCTCGGAGAACCTCCAGCAGCGCCCTCTAGTGTGGGA
TCTGGTACAGTGGCTGCAGGCGGTGGCGCACCAATGGCAGACAATAACGAAGGTGCCGACGGAGTGGGTAATGCC
TCAGGAAATTGGCATTGCGATTCCACATGGCTGGGCGACAGAGTCATTACCACCAGCACCCGAACCTGGGCCCTG
CCCACCTACAACAACCACTCTACAAGCAAAATCTCCAGTGAAACTGCAGGTAGTACCAACGACAACACCTACTTC
GGCTACAGCACCCCCCTGGGGGTATTTTGAATTTAACAGATTCCACTGCCACTTCTCACCACGTGACTGGCAGCGA
CTCATCAACAACAACCTGGGGATTCCGGCCCCAAGAAGCTGCGGTTCAAGCTCTTCAACATCCAGGTCAAGGAGGT
ACGACGAATGACGGCGTTACGACCATCGCTAATAACCTTACCAGCACGATTACAGGTATTCTCGGACTCGGAATAC
CAGCTGCCGTACGTCTCGGCTCTGCGCACCAAGGCTGCTGCTCCGTTCCCGGCGGACGTCTTCATGATTCTCT
CAGTACGGCTACCTGACTCTCAACAATGGCAGTCAGTCTGTGGGACGTTTCTCTCTTACTGCCTGGAGTACTTC
CCCTCTCAGATGCTGAGAACGGGCAACAACCTTTGAGTTCAGCTACAGCTTCGAGGACGTGCCTTTCCACAGCAGC
TACGCACACAGCCAGAGCCTGGACCGGCTGATGAATCCCTCATCGACCAGTACTTGTACTACTTGGCCAGAAC
CAGAGTAACCCAGGAGGCACAGCTGGCAATCGGGAACGAGTCTTACCAGGGCGGGCCTTCAACTATGGCCGAA
CAAGCCAAGAATTGGTTACCTGGACCTTGCTTCCGGCAACAAAGAGTCTCCAAAACGCTGGATCAAAACAACAAC
AGCAACTTTGCTTGGACTGGTGCCACCAATATCACCTGAACGGCAGAAAACCTCGTTGGTTAATCCCGGCGTCGCC
ATGGCAACTCACAAGGACGACGAGGACCGCTTTTTCCCATCCAGCGGAGTCTTGATTTTTTGGAAAAACTGGAGCA
ACTAACAAAACCTACATTGGAAAATGTGTTAATGACAAATGAAGAAGAAATTCGTCTACTAATCTGTAGCCACG
GAAGAATACGGGATAGTCAGCAGCAACTTACAAGCGGCTAATACTGCAGCCCAGACACAAGTTGTCAACAACAG
GGAGCCTTACCTGGCATGGTCTGGCAGAACCGGGACGTGTACCTGCAGGGTCCCATCTGGGCCAAGATTCTCTAC
ACGGATGGCAACTTTACCCGCTCTCCTTTGATGGGCGGCTTTGGACTTAAACATCCGCCTCCTCAGATCCTGATC
AAGAACACTCCCGTTCCCGCTAATCTCCGGAGGTGTTTACTCTGCAAGTTTGCTTCGTTTATCACACAGTAC
AGCACCGGACAAGTCAGCGTGGAAATCGAGTGGGAGCTGCAGAAGGAAAACAGCAAGCGCTGGAACCCGGAGATT
CAGTACACCTCCAACCTTTGAAAAGCAGACTGGTGTGGACTTTGCCGTTGACAGCCAGGGTGTCTTACTCTGAGCCT
CGCCCTATTGGCACTCGTTACCTCACCCGTAATCTGTAA

AAV-8

Full Genome: NC_006261

Rep78: (SEQ ID NO:19)

ATGCCGGGCTTCTACGAGATCGTGATCAAGGTGCCGAGCGACCTGGACGAGCACCTGCCGGGCATTTCTGACTCG
TTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGCTGCCCCGGATTCTGACATGGATCGGAATCTGATCGAGCAG
GCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTCTTGGTCCAATGGCGCCGCGTGAGTAAGGCCCCGGAG
GCCCTCTTCTTTGTTTCAGTTCGAGAAGGGCGAGAGCTACTTTCACCTGCACGTTCTGGTCGAGACCACGGGGGTC
AAGTCCATGGTGTAGGCCGCTTCTGAGTCAGATTTCGGGAAAAGCTTGGTCCAGACCATCTACCCGCGGGGTCG
AGCCCCACCTTGCCCAACTGGTTTCGCGGTGACCAAAGACGCGTAATGGCGCCGGCGGGGGGGAACAAGGTGGTG
GACGAGTGCTACATCCCCAACTACCTCCTGCCCCAAGACTCAGCCCCGAGCTGCAGTGGGCGTGGACTAACATGGAG
GAGTATATAAGCGCGTGCTTGAACCTGGCCGAGCGCAAACGGCTCGTGGCGCAGCACCTGACCCACGTCAGCCAG
ACGCAGGAGCAGAACAAGGAGAATCTGAACCCCAATTCTGACGCGCCCGTGATCAGGTCAAAAACCTCCGCGCGC
TATATGGAGCTGGTCGGGTGGCTGGTGGACCGGGGCATCACCTCCGAGAAAGCAGTGGATCCAGGAGGACCAGGCC
TCGTACATCTCCTTCAACGCGCCTCCAACCTCGCGGTCCCAGATCAAGGCCGCGCTGGACAATGCCGGCAAGATC
ATGGCGCTGACCAAATCCGCGCCCCGACTACCTGGTGGGGCCCTCGTGTCCCGCGGACATTACCCAGAACC GCATC
TACCGCATCTCTCGTCTCAACGGCTACGACCCTGCCTACGCGGCTCCGTCTTTCTCGGCTGGGCTCAGAAAAAG
TTCGGGAAACGCAACACCATCTGGCTGTTTGGACCCGCCACCACCGGCAAGACCAACATTGCGGAAGCCATCGCC
CACGCCGTGCCCTTCTACGGCTGCGTCAACTGGACCAATGAGAACCTTCCCTTCAATGATTGCGTCGACAAGATG
GTGATCTGGTGGGAGGAGGGCAAGATGACGGCCAAGGTCGTGGAGTCCGCCAAGGCCATTCTCGGCGGCAGCAAG
GTGCGCGTGGACCAAAAGTGCAAGTCGTCCGCCCAGATCGACCCCAACCCCGTGATCGTCACCTCCAACACCAAC
ATGTGCGCCGTGATTGACGGGAACAGCACCACTTCGAGCACCAGCAGCCTCTCCAGGACCGGATGTTTAAGTTC
GAACTACCCGCGCTCTGGAGCACGACTTTGGCAAGGTGACAAAAGCAGGAAGTCAAAGAGTTCTTCCGCTGGGCC
AGTGATCACGTGACCGAGGTGGCGCATGAGTTTACGTGAGAAAGGCGGAGCCAGCAAAAGACCCGCCCCCGAT
GACGCGGATAAAAGCGAGCCCAAGCGGGCCTGCCCTCAGTCGCGGATCCATCGACGTGACGCGGAAGGAGCT
CCGGTGGACTTTGCCGACAGGTACCAAAACAAATGTTCTCGTCACGCGGGCATGCTTCAGATGCTGTTTCCCTGC
AAAACGTGCGAGAGAAATGAATCAGAATTTCAACATTTGCTTCACACACGGGGTCAGAGACTGCTCAGAGTGTTC
CCCGCGTGTGAGAAATCTCAACCGGTGCTCAGAAAGAGGACGTATCGGAAACTCTGTGCGATTATCATCTGCTG
GGGCGGGCTCCCGAGATTGCTTGCTCGGCCTGCGATCTGGTCAACGTGGACCTGGATGACTGTGTTTCTGAGCAA
TAA

CapVP1: (SEQ ID NO:20)

ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCATTTCGCGAGTGGTGGGCGCTG
AAACCTGGAGCCCCGAAGCCCAAGCCAACCAGCAAAAGCAGGACGACGGCCGGGGTCTGGTGCTTCTTGGCTAC
AAGTACCTCGGACCTTCAACGGACTCGACAAGGGGGAGCCCGTCAACGCGGCGGACGCAGCGGCCCTCGAGCAC
GACAAGGCCTACGACCAGCAGCTGCAGGCGGGTGACAATCCGTACCTGCGGTATAACCACGCCGACGCCGAGTTT
CAGGAGCGTCTGCAAGAAGATACGTCTTTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAGAAGCGGGT
CTCGAACCTCTCGGTCTGGTTGAGGAAGGCGCTAAGACGGCTCCTGGAAAAGAGACCGGTAGAGCCATCACCC
CAGCGTTCTCCAGACTCCTCTACGGGCATCGGCAAGAAAGGCCAACAGCCCGCCAGAAAAAGACTCAATTTTGGT
CAGACTGGCGACTCAGAGTCAGTTCCAGACCCTCAACCTCTCGGAGAACCTCCAGCAGCGCCCTCTGGTGTGGGA
CCTAATACAATGGCTGCAGGCGGTGGCGCACCAATGGCAGACAATAACGAAGGCGCCGACGGAGTGGGTAGTTCC
TCGGGAAATTGGCATTGCGATTCCACATGGCTGGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTG
CCCACCTACAACAACCACTCTACAAGCAAATCTCCAACGGGACATCGGGAGGAGCCACCAACGACAACACCTAC
TTCGGCTACAGCACCCCTGGGGGTATTTTGAATTTAAGAGATTCCACTGCCACTTTTACCACGTGACTGGCAG
CGAGTCATCAACAACAACCTGGGGATTCCGGCCCAAGAGACTCAGCTTCAAGCTCTTCAACATCCAGGTCAAGGAG
GTCACGCAGAATGAAGGCACCAAGACCATCGCCAATAACCTCACCAGCACCATCCAGGTGTTTACGGACTCGGAG
TACCAGCTGCCGTACGTTCTCGGCTCTGCCACCAGGGCTGCCTGCCTCCGTTCCCGGCGGACGTGTTTATGATT
CCCCAGTACGGCTACCTAACACTCAACAACGGTAGTCAGGCGGTGGGACGCTCCTCCTTCTACTGCCTGGAATAC
TTTCTTTCGAGATGCTGAGAACCGGCAACAACCTCCAGTTTACTTACACCTTCGAGGACGTGCCTTTTCCACAGC
AGCTACGCCCACAGCCAGAGCTTGGACCGGCTGATGAATCCTCTGATTGACCAGTACCTGTACTACTTGTCTCGG
ACTCAAACAACAGGAGGCACGGCAAATACGCAGACTCTGGGCTTCAGCCAAGGTGGGCCTAATAACAATGGCCAAT
CAGGCAAAGAACTGGCTGCCAGGACCTGTTACCGCCAACAACGCGTCTCAACGACAACCGGGCAAAACAACAT
AGCAACTTTGCTGACTGCTGGGACCAATAACCATCTGAATGGAAGAAATTCATTGGCTAATCCTGGCATCGCT
ATGGCAACACACAAAGACGACGAGGAGCGTTTTTTTCCAGTAACGGGATCCTGATTTTTTGGCAAACAAAATGCT
GCCAGAGACAATGCGGATTACAGCGATGTCATGCTCACCAGCGAGGAAGAAATCAAACCACTAACCTGTGGCT
ACAGAGGAATACGGTATCGTGGCAGATAACTTGCAGCAGCAAAACACGGCTCCTCAAATTGGAAGTGTCAACAGC

CAGGGGGCCTTACCCGGTATGGTCTGGCAGAACCGGGACGTGTACCTGCAGGGTCCCATCTGGGCCAAGATTTCCT
CACACGGACGGCAACTTCCACCCGTCTCCGCTGATGGGCGGCTTTGGCCTGAAACATCTCCGCCTCAGATCCTG
ATCAAGAACACGCCTGTACCTGCGGATCCTCCGACCACCTTCAACCAGTCAAAGCTGAACTCTTTCATCACGCAA
TACAGCACCGGACAGGTTCAGCGTGGAAATTGAATGGGAGCTGCAGAAGGAAAACAGCAAGCGCTGGAACCCCGAG
ATCCAGTACACCTCCAACCTACTACAAATCTACAAGTGTGGACTTTGCTGTTAATACAGAAGGCGTGTACTCTGAA
CCCCGCCCCATTGGCACCCGTTACCTCACCCGTAATCTGTAA

AAV-9

Cap only: AY530579

CapVP1: (SEQ ID NO:21)

ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACCTTAGTGAAGGAATTCGCGAGTGGTGGGCTTTG
AAACCTGGAGCCCTCAACCCAAAGGCAAATCAACAACATCAAGACAACGCTCGAGGTCTTGTGCTTCCGGGTTAC
AAATACCTTGGACCCGGCAACGGACTCGACAAGGGGGAGCCGGTCAACGCAGCAGACGCGGCGGCCCTCGAGCAC
GACAAGGCCTACGACCAGCAGCTCAAGGCCGGAGACAACCCGTACCTCAAGTACAACCACGCCGACGCCGAGTTC
CAGGAGCGGCTCAAAGAAGATACGTCTTTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAAAAGAGGCTT
CTTGAACCTCTTGGTCTGGTTGAGGAAGCGGCTAAGACGGCTCCTGGAAGAAGAGGCCTGTAGAGCAGTCTCCT
CAGGAACCGGACTCCTCCGCGGGTATTGGCAAATCGGGTGACAGCCCGCTAAAAAGAGACTCAATTTCCGTCAG
ACTGGCGACACAGAGTCAGTCCAGACCCCTCAACCAATCGGAGAACCTCCCGCAGCCCCCTCAGGTGTGGGATCT
CTTACAATGGCTTCAGGTGGTGGCGCACCAGTGGCAGACAATAACGAAGTGCCGATGGAGTGGGTAGTTCCTCG
GGAAATTGGCATTGCGATTCCCAATGGCTGGGGGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGCCC
ACCTACAACAATCACCTCTACAAGCAAATCTCCAACAGCACATCTGGAGGATCTTCAAATGACAACGCCTACTTC
GGCTACAGCACCCCTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTCTCACCACGTGACTGGCAGCGA
CTCATCAACAACAATGGGGATTCCGGCCTAAGCGACTCAACTTCAAGCTCTTCAACATTCAAGTCAAAGAGGTT
ACGGACAACAATGGAGTCAAGACCATCGCCAATAACCTTACCAGCACGGTCCAGGTCTTCACGGACTCAGACTAT
CAGCTCCCGTACGTGCTCGGGTCCGGCTCACGAGGGGTGCTCCCGCCGTTCCAGCGGACGTTTTTCATGATTCTT
CAGTACGGGTATCTGACGCTTAATGATGGAAGCCAGGCCGTGGGTGCTTCGTCTTTTACTGCCTGGAATATTTT
CCGTGCGAAATGCTAAGAACGGGTAACTTCCAGTTCAGCTACGAGTTTGAGAACGTACCTTTCCATAGCAGC
TACGCTCACAGCCAAAGCCTGGACCGACTAATGAATCCACTCATCGACCAATACTTGTACTATCTCTCAAAGACT
ATTAACGGTTCTGGACAGAATCAACAACGCTAAAAATTCAGTGTGGCCGGACCCAGCAACATGGCTGTCCAGGGA
AGAAACTACATACCTGGACCCAGCTACCGACAACAACGTGTCTCAACCACTGTGACTCAAAACAACAACAGCGAA
TTTGTCTGGCCTGGAGCTTCTTCTTGGGCTCTCAATGGACGTAATAGCTTGATGAATCCTGGACCTGCTATGGCC
AGCCACAAGAAGGAGAGGACCGTTTTCTTTCTTTGTCTGGATCTTTAATTTTTTGGCAAACAAGGAACTGGAAGA
GACAACGTGGATGCGGACAAAGTCATGATAACCAACGAAGAAGAAATTAATACTACTAACCCGGTAGCAACGGAG
TCCTATGGACAAGTGGCCACAACACCAGAGTGCCCAAGCACAGGCGCAGACCGGCTGGGTTCAAAACCAAGGA
ATACTTCCGGGTATGGTTTGGCAGGACAGAGATGTGTACCTGCAAGGACCCATTTGGGCCAAAATTCCTCACACG
GACGGCAACTTTACCCCTTCTCCGCTGATGGGAGGGTTTGAATGAAGCACCCGCCTCCTCAGATCCTCATCAA
AACACACCTGTACCTGCGGATCCTCCAACGGCTTCAACAAGGACAAGCTGAACCTCTTTCATCACCCAGTATTC
ACTGGCCAAGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAAAACAGCAAGCGCTGGAACCCGGAGATCCAG
TACACTTCCAACCTATTACAAGTCTAATAATGTTGAATTTGCTGTTAATACTGAAGGTGTATATAGTGAACCCCGC
CCCATTTGGCACCCAGATACCTGACTCGTAATCTGTAA

AAV-10

Partial Genome: AY631965

Rep78: (SEQ ID NO:22)

ATGCCGGGCTTCTACGAGATCGTGATCAAGGTGCCGAGCGACCTGGACGAGCACCTGCCGGGCATTTCTGACTCG
TTTGTGAACCTGGGTGGCCGAGAAGGAATGGGAGCTGCCCCGGATTCTGACATGGATCGGAATCTGATCGAGCAG
GCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTCTGGTCCACTGGCGCCGCGTGAGTAAGGCCCCGGAG
GCCCTCTTCTTTGTTTCAGTTCGAGAAGGGCGAGTCTACTTTACCTGCACGTTCTGGTCGAGACCACGGGGGTC
AAGTCCATGGTCTTGGGCCGCTTCTGAGTCAGATCAGAGACAGGCTGGTGCAGACCATCTACCGCGGGGTAGAG
CCCACGCTGCCCAACTGGTTTCGCGGTGACCAAGACGCGAAATGGCGCCGGCGGGGGGAACAAGGTGGTGGACGAG
TGCTACATCCCCAACTACCTCCTGCCAAGACGCGAGCCGAGCTGCAGTGGGCGTGGACTAACATGGAGGAGTAT
ATAAGCGCGTGTCTGAACCTCGCGGAGCGTAAACGGCTCGTGGCGCAGCACCTGACCCACGTCAGCCAGACGCGAG
GAGCAGAACAAGGAGAATCTGAACCCGAATTCTGACGCGCCCGTGATCAGGTCAAAAACCTCCGCGCGCTACATG
GAGCTGGTCCGGTGGCTGGTGGACCGGGGCATCACCTCCGAGAAGCAGTGGATCCAGGAGGACCAGGCCTCGTAC

ATCTCCTTCAACGCCGCTCCAACCTCGCGGTCCCAGATCAAGGCCGCGCTGGACAATGCCGGAAGATCATGGCG
 CTGACCAAATCCGCGCCCCGACTACCTGGTAGGCCCTCCTTACCCGCGGACATTAAGGCCAACCGCATCTACCGC
 ATCCTGGAGCTCAACGGCTACGACCCCGCTACGCCGGCTCCGTCTTCCTGGGCTGGGCGCAGAAAAAGTTCGGT
 AAAAGGAATACAATTTGGCTGTTCGGGCCCCGCCACCACCGGCAAGACCAACATCGCGGAAGCCATCGCCACGCC
 GTGCCCTTCTACGGCTGCGTCAACTGGACCAATGAGAACTTTCCCTTCAACGATTGCGTCGACAAGATGGTGATC
 TGGTGGGAGGAGGGCAAGATGACCGCCAAGGTCTGGAGTCCGCCAAGGCCATTCTGGGCGGAAGCAAGGTGCGC
 GTCGACCAAAAGTGCAAGTCTCGGCCCAGATCGACCCACGCCCCGTGATCGTCACCTCCAACACCAACATGTGC
 GCCGTGATCGACGGGAACAGCACCACTTCGAGCACCAGCAGCCCCGTCAGGACCGCATGTTCAAGTTCGAGCTC
 ACCCGCCGTCTGGAGCACGACTTTGGCAAGGTGACCAAGCAGGAAGTCAAAGAGTTCTTCCGCTGGGCTCAGGAT
 CACGTGACTGAGGTGACGCATGAGTTCTACGTGAGAAAGGGCGGAGCCACCAAAAGACCCGCCCCCAGTGACGCG
 GATATAAGCGAGCCCAAGCGGGCTGCCCCCTCAGTTGCGGAGCCATCGACGTGACGCGGAAGCACCAGGTGGAC
 TTTGCGGACAGGTACCAAAACAAATGTTCTCGTCACGCGGGCATGCTTCAGATGCTGTTTCCCTGCAAGACATGC
 GAGAGAATGAATCAGAATTTCAACGCTCGCTTCACGCACGGGGTCAGAGACTGCTCAGAGTGTCTTCCCGGCGCG
 TCAGAATCTCAACCTGTGCTGAGAAAAAGACGTATCAGAACTGTGCGCGATTTCATCATCTGCTGGGGCGGGCA
 CCCGAGATTGCGTGTTCGGCCTGCGATCTCGTCAACGTGGACTTGGATGACTGTGTTTCTGAGCAATAA

CapVP1: (SEQ ID NO:23)

ATGGCTGCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCATTTCGCGAGTGGTGGGACCTG
 AAACCTGGAGCCCCCAAGCCCAAGGCCAACCAGCAGAAGCAGGACGACGGCCGGGGTCTGGTGCTTCTGGCTAC
 AAGTACCTCGGACCTTCAACGGACTCGACAAGGGGGAGCCCGTCAACGCGCGGACGCAGCGGCCCTCGAGCAC
 GACAAGGCCTACGACCAGCAGCTCAAAGCGGGTGACAATCCGTACCTGCGGTATAACCACGCCGACGCCAGTTT
 CAGGAGCGTCTGCAAGAAGATACGTCTTTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAGAAGCGGGT
 CTCGAACCTCTCGGTCTGGTTGAGGAAGCTGCTAAGACGGCTCCTGGAAAAGAGACCGGTAGAACCCTCACCT
 CAGCGTTCCCCCGACTCCTCCACGGGCATCGGCAAGAAAGGCCAGCAGCCCGCTAAAAAGAGACTGAACTTTGGG
 CAGACTGGCGAGTCAGAGTCAGTCCCCGACCCTCAACCAATCGGAGAACCACCAGCAGGCCCTCTGGTCTGGGA
 TCTGGTACAATGGCTGCAGGCGGTGGCGCTCCAATGGCAGACAATAACGAAGGCGCCGACGGAGTGGGTAGTTCC
 TCAGGAATTTGGCATTGCGATTCCACATGGCTGGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTG
 CCCACCTACAACAACCACTCTACAAGCAAATCTCCAACGGGACATCGGGAGGAAGCACAACGACAACACCTAC
 TTCGGCTACAGCACCCCTGGGGGTATTTGACTTCAACAGATTCCACTGCCACTTCTACCACGTGACTGGCAG
 CGACTCATCAACAACAACCTGGGGATTCCGGCCAAAAAGACTCAGCTTCAAGCTCTTCAACATCCAGGTCAAGGAG
 GTCACGCAGAATGAAGGCACCAAGACCATCGCCAATAACCTTACCAGCACGATTGAGTATTTACGGACTCGGAA
 TACCAGCTGCCGTACGTCTCGGCTCCGCGCACACGGGCTGCCTGCCTCCGTTCGCGGCGGATGTCTTCATGATT
 CCCCAGTACGGCTACCTGACACTGAACAATGGAAGTCAAGCCGTAGGCCGTTCCTCCTTCTACTGCCTGGAATAT
 TTTCCATCTCAAATGCTGCGAACTGGAAACAATTTTGAATTCAGCTACACCTTCGAGGACGTGCCTTTCCACAGC
 AGCTACGCACACAGCCAGAGCTTGGACCGACTGATGAATCCTCTCATTGACCAGTACCTGTACTACTTATCCAGA
 ACTCAGTCCACAGGAGGAACCTCAAGGTACCCAGCAATTGTTATTTTCTCAAGCTGGGCCTGCAAACATGTCGGCT
 CAGGCCAAGAAGTGGCTGCCTGGACCTTGCTACCGGCAGCAGCGAGTCTCCACGACACTGTCGAAAACAACAAC
 AGCAACTTTGCTTGGACTGGTGCCACCAAAATATCACCCTGAACGGAAGAGACTCTCTGGTGAATCCCGGTGTCGCC
 ATGGCAACCCACAAGGACGACGAGGAACGCTTCTTCCCGTCGAGCGGAGTCCGTGATGTTTGGAAAACAGGGTGCT
 GGAAGAGACAATGTGGACTACAGCAGCGTTATGCTAACAAGCGAAGAAAGAAATTAAACCCTAACCCTGTAGCC
 ACAGAACAATACGGCGTGGTGGCTGACAACCTGACAGCAAGCCAATACAGGGCCCTATTGTGGGAAATGTCAACAGC
 CAAGGAGCCTTACCTGGCATGGTCTGGCAGAACCAGACGTGTACCTGCAGGGTCCCATCTGGGCCAAGATTCTCT
 CACACGGACGGCAACTTTACCCGTCTCCTCTGATGGGCGGCTTTGGACTTAAACACCCGCCCTCCACAGATCCTG
 ATCAAGAACACGCCGTACCTGCGGATCCTCCAACAACGTTTACGCCAGGCGAAATTGGCTTCTTTCATCACGCAG
 TACAGCACCGGACAGGTACGCTGGAATCGAGTGGGAGCTGCAGAAAGGAGAACAGCAAACGCTGGAACCCAGAG
 ATTACAGTACACTTCAAACCTACTACAAATCTACAAATGTGAGCTTTGCTGTCAATACAGAGGGAACCTTATTCTGAG
 CCTCGCCCCATTGGTACTCGTTATCTGACACGTAATCTGTAA

AAV-11

Partial Genome: AY631966

Rep78: (SEQ ID NO:24)

ATGCCGGGCTTCTACGAGATCGTGATCAAGGTGCCGAGCGACCTGGACGAGCACCTGCCGGGCATTTCTGACTCG
 TTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGCTGCCCCCGGATTCTGACATGGATCGGAATCTGATCGAGCAG
 GCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTCTGGTCCACTGGCGCCGCGTGAGTAAGGCCCGGGAG
 GCCCTCTTCTTTGTTTTCAGTTTCGAGAAGGGCGAGTCTTACTTCCACCTCCACGTTCTCGTCGAGACCACGGGGGT

AAGTCCATGGTCCTGGGCCGCTTCTGAGTCAGATCAGAGACAGGCTGGTGCAGACCATCTACCGCGGGGTCGAG
CCCACGCTGCCCCAACTGGTTTCGCGGTGACCAAGACGCGAAATGGCGCCGGCGGGGGGAACAAGGTGGTGGACGAG
TGCTACATCCCCAACTACCTCCTGCCAAGACCCAGCCGAGCTGCAGTGGGCGTGGACTAACATGGAGGAGTAT
ATAAGCGCGTGTCTAAACCTCGCGGAGCGTAAACGGCTCGTGGCGCAGCACCTGACCCACGTCAGCCAGACGCAG
GAGCAGAACAAGGAGAATCTGAACCCGAATTCTGACGCGCCCGTGATCAGGTCAAAAACCTCCGCGCGCTACATG
GAGCTGGTCGGGTGGCTGGTGGACCGGGGCATCACCTCCGAGAAGCAGTGGATCCAGGAGGACCAGGCCTCGTAC
ATCTCCTTCAACGCCGCTCCAACCTCGCGGTCCCAGATCAAGGCCGCGCTGGACAAATGCCGGAAGATCATGGCG
CTGACCAAATCCGCGCCCGACTACCTGGTAGGCCCGTCTTACCCGCGGACATTAAGGCCAACC GCATCTACCGC
ATCCTGGAGCTCAACGGCTACGACCCCGCTACGCCGGCTCCGTCTTCTGGGCTGGGCGCAGAAAAAGTTCCGT
AAACGCAACACCATCTGGCTGTTTGGGCCCGCCACCACCGGCAAGACCAACATCGCGGAAGCCATAGCCACGCC
GTGCCCTTCTACGGCTGCGTGAACCTGGACCAATGAGAACTTTCCTTCAACGATTGCGTCGACAAGATGGTGATC
TGGTGGGAGGAGGGCAAGATGACCGCCAAGGTCGTGGAGTCCGCCAAGGCCATTCTGGGCGGAAGCAAGGTGCGC
GTGGACCAAAAGTGCAAGTCTCGGCCCAAGATCGACCCACGCCCGTGATCGTCACCTCCAACACCAACATGTGC
GCCGTGATCGACGGGAACAGCACCACTTCGAGCACCAGCAGCCGCTGCAGGACCGCATGTTCAAGTTCGAGCTC
ACCCGCCGTCTGGAGCACGACTTTGGCAAGGTGACCAAGCAGGAAGTCAAAGAGTTCTTCCGCTGGGCTCAGGAT
CACGTGACTGAGGTGGCGCATGAGTTCTACGTGAGAAAGGGCGGAGCCACCAAAAGACCCGCCCCCAGTGACGCG
GATATAAGCGAGCCCAAGCGGGCCTGCCCCCAGTTCCGGAGCCATCGACGTGACGCGGAAGCACCAGGTGGAC
TTTGGCGACAGGTACCAAAACAAATGTTCTCGTCACGCGGGCATGCTTCAGATGCTGTTTCCCTGCAAGACATGC
GAGAGAATGAATCAGAATTTCAACGTCTGCTTCACGCACGGGGTCAGAGACTGCTCAGAGTGCTTCCCCGGCGCG
TCAGAATCTCAACCCGTGCTCAGAAAAAAGACGTATCAGAACTGTGCGCGATTATCATCTGCTGGGGCGGGCA
CCCGAGATTGCGTGTTCCGCCTGCGATCTCGTCAACGTGGACTTGGATGACTGTGTTTCTGAGCAATAA

CapVP1: (SEQ ID NO:25)

ATGGCTGCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCATTTCGCGAGTGGTGGGACCTG
AAACCTGGAGCCCCGAAGCCCAAGGCCAACCAGCAGAAGCAGGACGACGGCCGGGGTCTGGTGCTTCTGGCTAC
AAGTACCTCGGACCCCTTCAACGGACTCGACAAGGGGGAGCCCGTCAACGCGGCGGACGACGCGGCCCTCGAGCAC
GACAAGGCCCTACGACCAGCAGCTCAAAGCGGGTGACAATCCGTACCTGCGGTATAACCACGCCGACGCCGAGTTT
CAGGACCGCTGCAAGAAGATACGTCTTTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAGAAGAGGGTA
CTCGAACCTCTGGGCTGGTTGAAGAAGGTGCTAAAACGGCTCCTGGAAAGAAGAGACCCTTAGAGTCAACCAAA
GAGCCCGACTCCTCCTCGGGCATCGGCAAAAAAGGCAAAACAACCAGCCAGAAAGAGGCTCAACTTTGAAGAGGAC
ACTGGAGCCGGAGACGGACCCCCCTGAAGGATCAGATACCAGCGCCATGTCTTCAGACATTGAAATGCGTGCAGCA
CCGGGCGGAAATGCTGTGATGCGGGACAAGGTTCCGATGGAGTGGGTAATGCCCTGGGTGATTGGCATTGCGAT
TCCACCTGGTCTGAGGGCAAGGTCAACAACCTCGACCAGAACCTGGGTCTTGCCACCTACAACAACCACTTG
TACCTGCGTCTCGGAACAACATCAAGCAGCAACACCTACAACGGATTCTCCACCCCTGGGGATATTTTGACTTC
AACAGATTCCACTGTCACTTCTCACCACGTGACTGGCAAAGACTCATCAACAACAACCTGGGGACTACGACCAAAA
GCCATGCGCGTTAAATCTTCAATATCCAAGTTAAGGAGGTCAACACGTCGAACGGCGAGACTACGGTCGCTAAT
AACCTTACCAGCACGGTTCAGATATTTGCGGACTCGTCGTATGAGTCCCCTACGTGATGGACGCTGGACAAGAG
GGGAGCCTGCCTCCTTTCCCAATGACGTGTTCAATGGTGCCTCAATATGGCTACTGTGGCATCGTGACTGGCGAG
AATCAGAACCAACCGGACAGAAACGCTTTTCTACTGCCTGGAGTATTTTCTTCGCAAATGTTGAGAACTGGCAAC
AACTTTGAAATGGCTTACAACCTTTGAGAAGGTGCCGTTCCACTCAATGTATGCTCACAGCCAGAGCCTGGACAGA
CTGATGAATCCCCTCCTGGACCAGTACCTGTGGCACTTACAGTCGACTACCTCTGGAGAGACTCTGAATCAAGGC
AATGCAGCAACCACATTTGGA AAAATCAGGAGTGGAGACTTTGCCTTTTACAGAAAGAACTGGCTGCCTGGGCCT
TGTGTTAAACAGCAGAGATTCTCAAAAACCTGCCAGTCAAAATTACAAGATTCTGCCAGCGGGGGCAACGCTCTG
TTAAAGTATGACACCCACTATACCTTAAACAACCGCTGGAGCAACATCGCGCCCGGACCTCCAATGGCCACAGCC
GGACCTTCGATGGGAGACTTCAGTAACGCCAGCTATATTCCTGGACCATCTGTTACCGGAAATACAACAAC
TCAGCCAACAATCTGTTGTTTACATCAGAAAGAAATGCTGCCACCAACCAAGAGACACGGACATGTTTGGC
CAGATTGCTGACAATAATCAGAATGCTACAACCTGCTCCCATAAACGGCAACGTGACTGCTATGGGAGTGCTGCCT
GGCATGGTGTGGCAAAACAGAGACATTTACTACCAAGGGCCAATTTGGGCCAAGATCCCACACGCGGACGGACAT
TTTCATCCTTACCAGCTGATTGGTGGGTTTGGACTGAAACACCCGCTCCCCAGATATTCATCAAGAACTCCC
GTACCTGCCAATCCTGCGACAACCTTCACTGCAGCCAGAGTGGACTCTTTCATCACACAATACAGACCGGCCAG
GTCGCTGTTTCAAGATTGAATGGGAAATTGAAAAGGAACGCTCCAAACGCTGGAATCCTGAAGTGCAGTTTACTTCA
AACTATGGGAACCAAGTCTTCTATGTTGTGGGCTCCTGATACAACCTGGGAAGTATACAGAGCCGCGGGTTATTGGC
TCTCGTTATTTGACTAATCATTTGTAA

AAV-12

Partial Genome: DQ813647

Rep78: (SEQ ID NO:26)

ATGCCGGGGTTCTACGAGGTGGTGATCAAGGTGCCCAGCGACCTGGACGAGCACCTGCCCGGCATTTCTGACTCC
TTTGTGAACTGGGTGGCCGAGAAGGAATGGGAGTTGCCCCGGATTCTGACATGGATCAGAATCTGATTGAGCAG
GCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGAGTTCTGGTGGAAATGGCGCCGAGTGAGTAAATTTCTGGAG
GCCAAGTTTTTTGTGCAGTTTGAAAAGGGGGACTCGTACTTTTCATTTGCATATTCTGATTGAAATTACCGGCGTG
AAATCCATGGTGGTGGGCCGCTACGTGAGTCAGATTAGGGATAAACTGATCCAGCGCATCTACCGCGGGGTCGAG
CCCCAGCTGCCCAACTGGTTCGCGGTACAAAAGACCCGAAATGGCGCCGAGGCGGGAACAAGGTGGTGGACGAG
TGCTACATCCCCAACTACCTGCTCCCCAAGGTCCAGCCGAGCTTCAGTGGGCGTGGACTAACATGGAGGAGTAT
ATAAGCGCCTGTTTGAACCTCGCGGAGCGTAAACGGCTCGTGGCGCAGCACCTGACGCACGTCTCCCAGACCCAG
GAGGGCGACAAGGAGAATCTGAACCCGAATTCTGACGCGCCGGTGATCCGGTCAAAAACCTCCGCCAGGTACATG
GAGCTGGTCGGGTGGCTGGTGGACAAGGGCATCACGTCCGAGAAGCAGTGGATCCAGGAGGACCAGGCCTCGTAC
ATCTCCTTCAACGCGGCCTCCAACCTCCCGGTTCGCAGATCAAGGCGGCCCTGGACAATGCCTCCAAAATCATGAGC
CTCACAAAACGGCTCCGGACTATCTCATCGGGCAGCAGCCCGTGGGGGACATTACCACCAACCGGATCTACAAA
ATCCTGAACTGAACGGGTACGACCCCCAGTACGCCGCTCCGTCTTTCTCGGCTGGGCCAGAAAAAGTTTGGGA
AAGCGCAACACCATCTGGCTGTTTGGGCCCCGCCACCACCGGCAAGACCAACATCGCGGAAGCCATCGCCACGCG
GTCCCTTCTACGGCTGCGTCAACTGGACCAATGAGAACTTTCCCTTCAACGACTGCGTCGACAAAATGGTGATT
TGGTGGGAGGAGGGCAAGATGACCGCCAAGGTCTGAGAGTCCGCCAAGGCCATTCTGGGCGGCAGCAAGGTGCGC
GTGGACCAAAAATGCAAGGCCTCTGCGCAGATCGACCCACCCCGTGATCGTCACCTCCAACACCAACATGTGC
GCCGTGATTGACGGGAACAGCACACCTTCGAGCACCAGCAGCCCTGCAGGACCGGATGTTCAAGTTTGAACCTC
ACCCGCGCCTCGACCACGACTTTGGCAAGGTCAACAAGCAGGAAGTCAAGGACTTTTTCCGGTGGGCGGCTGAT
CACGTGACTGACGTGGCTCATGAGTTTTACGTACAAAAGGTGGAGCTAAGAAAAGGCCCGCCCCCTCTGACGAG
GATATAAGCGAGCCCAAGCGGCCGCGGTGTCATTTGCGCAGCCGGAGACGTGACGCGGGAAGCTCCCGGAGAC
TTCGCCGACAGGTACCAAAAACAAATGTTCTCGTCACGCGGGTATGCTGCAGATGCTCTTTCCCTGCAAGACGTGC
GAGAGAATGAATCAGAATTCCAACGTCTGCTTCACGCACGGTCAGAAAAGATTGCGGGGAGTGCTTTCCCGGGTCA
GAATCTCAACCGGTTTTCTGTCGTACAGAAAACGTATCAGAACTGTGCATCCTTCATCAGCTCCGGGGGGCACCC
GAGATCGCCTGCTCTGCTTGCGACCAACTCAACCCCGATTGGACGATTGCCAATTTGAGCAATAA

CapVP1: (SEQ ID NO:27)

ATGGCTGCTGACGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAAGGCATTTCGCGAGTGGTGGGCGCTG
AAACCTGGAGCTCCACAACCCAAGGCCAACCAACAGCATCAGGACAACGGCAGGGGTCTTGCTTCTGGGTAC
AAGTACCTCGGACCCTTCAACGGACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGCAC
GACAAGGCCTACGACAAGCAGCTCGAGCAGGGGGACAACCCGTATCTCAAGTACAACCACGCCGACGCCGAGTTC
CAGCAGCGCTTGGCGACCGACACCTCTTTTGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAAAGAGGATT
CTCGAGCCTCTGGGTCTGGTTGAAGAGGGCGTTAAAACGGCTCTTGAAAAGAAACGCCATTAGAAAAGACTCCA
AATCGGCCGACCAACCCGGACTCTGGGAAGGCCCGGCCAAGAAAAAGCAAAAAGACGGCGAACACAGCCGACTCT
GCTAGAAGGACACTCGACTTTGAAGACTCTGGAGCAGGAGACGGACCCCTGAGGGATCATCTTCCGAGAAATG
TCTCATGATGCTGAGATGCGTGCGGCGCCAGGCGGAAATGCTGTGAGGCGGGACAAGGTGCCGATGGAGTGGGT
AATGCCTCCGGTGATTGGCATTGCGATTCCACCTGGTCAGAGGGCCGAGTCACCACCACCAGCACCCGAACCTGG
GTCCTACCCACGTACAACAACCACTGTACCTGCGAATCGGAACAACGGCCAACAGCAACACCTACAACGGATTTC
TCCACCCCTGGGGATACTTTGACTTTAACCGCTTCCACTGCCACTTTTCCCCACGCGACTGGCAGCGACTCATC
AACAACAACCTGGGACTCAGGCCGAAATCGATGCGTGTTAAAATCTTCAACATACAGGTCAAGGAGGTCACGACG
TCAAACGGCGAGACTACGGTCGCTAATAACCTTACCAGCACGGTTCAGATCTTTGCGGATTTCAGCTATGAACCTC
CCATACGTGATGGACGCCGGTCAGGAGGGGAGCTTCTCTCCGTTTCCCAACGACGTCTTTATGGTTCCCCAATAC
GGATACTTCGGAGTTGTCACTGGAAAAAACAGAACAGACAGAGAAATGCCTTTTACTGCCTGGAATACATTT
CCATCCCAAATGCTAAGAACTGGCAACAATTTGAAGTCAGTTACCAATTTGAAAAAGTTCTTTCCATTCAATG
TACGCGCACAGCCAGAGCCTGGACAGAATGATGAATCCTTTACTGGATCAGTACCTGTGGCATCTGCAATCGACC
ACTACCGGAAATTCCTTAATCAAGGAACAGCTACCACCACGTACGGGAAAAATTACCCTGGAGACTTTGCCTAC
TACAGGAAAAACTGGTTGCCTGGAGCCTGCATTAAACAACAAAAATTTCAAAGAAATGCCAATCAAACTACAAG
ATTCCCGCCAGCGGGGAGACGCCCTTTTAAAGTATGACACGCATACCACTCTAAATGGGCGATGGAGTAACATG
GCTCCTGGACCTCCAATGGCAACCGCAGGTGCCGGGGACTCGGATTTTAGCAACAGCCAGCTGATCTTTGCCGGA
CCCAATCCGAGCGGTAACACGACCACATCTTCAAACAATTTGTTGTTTACCTCAGAAGAGGAGATTGCCACAACA
AACCCACGAGACACGGACATGTTTGGACAGATTGCAGATAATAATCAAAATGCCACCACCGCCCTCACATCGCT
AACCTGGACGCTATGGGAATTTGCCGGAATGGTCTGGCAAAACAGAGACATCTACTACCAGGGCCCTATTTGG
GCCAAGGTCCCTCACACGGACGGACACTTTCACCTTCGCCGCTGATGGGAGGATTTGGACTGAAACACCCGCCT
CCACAGATTTTCATCAAAAACACCCCGTACCCGCCAATCCCAATACTACCTTTAGCGCTGCAAGGATTAATTCCT
TTTCTGACGCAGTACAGCACCGGACAAGTTGCCGTTTCAGATCGACTGGGAAATTCAGAAGGAGCATTCCAAACGC

TGGAATCCCGAAGTTCAATTTACTTCAAACCTACGGCACTCAAAATTCTATGCTGTGGGCTCCCGACAATGCTGGC
AACTACCACGAACCTCCGGGCTATTGGGTCCCGTTTCCTCACCCACCACTTGTAA

AAV-13

Partial Genome: EU285562

Rep78: (SEQ ID NO:28)

ATGCCGGGATTCTACGAGATTGTCTGAAGGTGCCAGCGACCTGGACGAGCACCTGCCTGGCATTCTGACTCT
TTTGTAACCTGGGTGGCGGAGAAGGAATGGGAGCTGCCGCCGGATTCTGACATGGATCTGAATCTGATTGAGCAG
GCACCCCTAACCGTGGCCGAAAAGCTGCAACGCGAATTCCTGGTTCGAGTGGCGCCGCGTGAGTAAGGCCCCGGAG
GCCCTCTTCTTTGTTTCAGTTTCGAGAAGGGGGACAGCTACTTCCACCTACACATTCTGGTGGAGACCGTGGGCGTG
AAATCCATGGTGGTGGGCGCTACGTGAGCCAGATTAAAGAGAAGCTGGTGACCCGCATCTACCGCGGGGTCGAG
CCGCAGCTTCCGAACCTGGTTCGCGGTGACCAAGACGCGTAATGGCGCCGGAGGCGGGAACAAGGTGGTGGACGAC
TGCTACATCCCCAACTACCTGCTCCCCAAGACCCAGCCGAGCTCCAGTGGGCGTGGACTAATATGGACCAGTAT
TTAAGCGCCTGTTTGAATCTCGCGGAGCGTAAACGGCTGGTGGCGCAGCATCTGACGCACGTGTGCGAGACGCGAG
GAGCAGAACAAGAGAACCAGAATCCCAATTCTGACGCGCCGGTGATCAGATCAAAAACCTCCGCGAGGTACATG
GAGCTGGTTCGGGTGGCTGGTGGACCGCGGGATCACGTGAGAAAAGCAATGGATCCAGGAGGACCAGGCCTCTTAC
ATCTCCTTCAACGCCGCTCCAACTCGCGGTACAAAATCAAGGCCGCACTGGACAATGCCTCCAAATTTATGAGC
CTGACAAAAACGGCTCCGGACTACCTGGTGGGAAACAACCCGCCGAGGACATTACCAGCAACCGGATCTACAAA
ATCCTCGAGATGAACGGGTACGATCCGCAGTACGCGGCTCCGTCTTCCTGGGCTGGGCGCAAAAGAAGTTCGGG
AAGAGGAACACCATCTGGCTCTTTGGGCCGGCCACGACGGGTAAACCACATCGCTGAAGCTATCGCCACGCC
GTGCCCTTTTACGGCTGCGTGAACCTGGACCAATGAGAACTTTCCGTTCAACGATTGCGTCGACAAGATGGTGATC
TGGTGGGAGGAGGGCAAGATGACGGCCAAGGTCTGGAGTCCGCCAAGGCCATTCTGGGCGGAAGCAAGGTGCGC
GTGGACCAAAAGTGCAAGTCATCGGCCCAGATCGACCCAACCTCCCGTCATCGTCACCTCCAACACCAACATGTGC
GCGGTTCATCGACGGAATTCACACCTTCGAGCACCAACAACCACTCCAAGACCGGATGTTCAAGTTCGAGCTC
ACCAAGCGCCTGGAGCACGACTTTGGCAAGGTACCAAGCAGGAAGTCAAGGACTTTTTCCGGTGGGCGTCAGAT
CACGTGACTGAGGTGTCTCACGAGTTTACGTGAGAAAGGGTGGAGCTAGAAAGAGGCCCGCCCCCAATGACGCA
GATATAAGTGAGCCCAAGCGGGCCTGTCCGTGAGTTCGCGAGCCATCGACGTGAGACGCGGAAGCTCCGGTGGAC
TACGCGGACAGGTACCAAAACAATGTTCTCGTCACGTGGGCATGAATCTGATGCTTTTCCCTGCGGCAATGC
GAGAGAATGAATCAGAATGTGGACATTTGCTTCACGCACGGGGTCATGGACTGTGCCGAGTGCTTCCCCGTGTCA
GAATCTCAACCCGTGTCTGTGTCGAGAAAGCGGACATATCAGAACTGTGTCCGATTTCATCACATCATGGGGAGG
GCGCCCCGAGGTGGCTTGTTCGGCCTGCGATCTGGCCAATGTGGACTTGGATGACTGTGACATGGAGCAATAA

CapVP1: (SEQ ID NO:29)

ATGACTGACGGTTACCTTCCAGATTGGCTAGAGGACAACCTCTCTGAAGGCGTTCGAGAGTGGTGGGCGCTGCAA
CCTGGAGCCCCCTAAACCAAGGCAATCAACAACATCAGGACAACGCTCGGGGTCTTGTGCTTCCGGGTACAAA
TACCTCGGACCCGGCAACGGACTTGACAAGGGGGAACCCGTCAACGCAGCGGACGCGGCAGCCCTCGAACACGAC
AAGGCCTACGACCAGCAGCTCAAGGCCGGTGACAACCCCTACCTCAAGTACAACCACGCCGACGCCGAGTTTCAG
GAGCGTCTTCAAGAAGATACGTCTTTTGGGGGCAACCTCGGACGAGCAGTCTTCCAGGCCAAAAAGAGGATCCTT
GAGCCTCTGGGTCTGGTTGAGGAAGCGGCTAAGACGGCTCCTGGAAAAAGAGACCTGTAGAGCAATCTCCAGCA
GAACCGGACTCCTCTTCGGGCATCGGCAATCAGGCCAGCAGCCGCTAGAAAAAGACTGAATTTTGGTCAGACT
GGCGACACAGAGTCAGTCCCAGACCTCAACCACTCGGACAACCTCCCGCAGCCCCCTCTGGTGTGGGATCTACT
ACAATGGCTTCAGGCGGTGGCGCACCAATGGCAGACAATAACGAGGGTGCCGATGGAGTGGGTAATTCCTCAGGA
AATTGGCATTGCGATTCCCAATGGCTGGGCGCAGACAGTCATCACCACGACCCGACCTGGGCCCTGCCACAC
TACAACAATCACCTCTACAAGCAAATCTCCAGCCAATCAGGAGCCACCAACGACAACCACCTACTTTGGCTACAGC
ACCCCTGGGGGTATTTTGACTTCAACAGATTCCACTGCCACTTTTCAACCACGTGACTGGCAAAGACTCATCAAC
AACAACCTGGGGATTCCGACCCAAGAGACTCAACTTCAAGCTCTTTAACATTCAAGTCAAAGAGGTACGCAGAAT
GACGGTACGACGACGATTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCCGAGTACCAGTCCCCG
TACGTCTCGGCTCGGCGCATCAGGGATGCCTCCCGCGTTCCAGCAGACGTCTTCATGGTCCCACAGTATGGA
TACCTCACCTGAACAACGGGAGTCAGGCGGTAGGACGCTCTTCCTTTTACTGCCTGGAGTACTTTCTTCTCAG
ATGCTGCGTACTGGAAACAACCTTTCAGTTTAGCTACACTTTTGAAGACGTGCCTTTCCACAGCAGCTACGCTCAC
AGCCAAAGTCTGGACCGTCTCATGAATCCTCTGATCGACCACTGTACTATCTGAACAGGACACAAACAGCC
AGTGGAACCTCAGCAGTCTCGGCTACTGTTTAGCCAAGCTGGACCCACCACTATGTCTCTTCAAGCTAAAACTGG
CTGCCTGGACCTTGCTACAGACAGCAGCGTCTGTCAAAGCAGGCAACGACAACAACAACAGCAACTTTCCCTGG
ACTGGTGCCACCAAAATATCATCTGAATGGCCGGGACTCATTGGTGAACCCGGGCCCCGTGTATGGCCAGTCACAAG
GATGACAAAGAAAAGTTTTTCCCCATGCATGGAACCTGATATTTGGTAAAGAAGGAACAAATGCCAACAACGCG

GATTTGGAAAATGTCATGATTACAGATGAAGAAGAAATCCGCACCACCAATCCCGTGGCTACGGAGCAGTACGGG
ACTGTGTCAAATAATTTGCAAACTCAAACGCTGGTCCAATACTGGAAGTGTCAATCACCAGGAGCGTTACCT
GGTATGGTGTGGCAGGATCGAGACGTGTACCTGCAGGGACCCATTTGGGCCAAGATTCTCAGATCATGATCAAAAACACTCCC
TTTCATCCTTCTCCACTGATGGGAGGTTTTGGGCTCAAACACCCGCTCCTCAGATCATGATCAAAAACACTCCC
GTTCCAGCCAATCTCTCCACAACTTTAGTGCGGCAAAAGTTTGCTTCTTCATCACACAGTACTCCACGGGGCAG
GTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAAGGAGAACAGCAAAACGCTGGAATCCCGAAATTTCAGTACACTTCC
AACTACAACAAATCTGTTAATGTGGACTTTACTGTGGACACTAATGGTGTGTATTTCAGAGCCTCGCCCCATTGGC
ACCAGATACCTGACTCGTAATCTGTAA

ITR Sequence (SEQ ID NO:30)

CCTGCAGGCAGCTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTGGGGCGACCTTTGGTC
GCCCCGCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATCACTAGGGGTTCT

Rep2 Sequence - Contains Rep78 and Rep52 (start codon underlined) (SEQ ID NO:31)

ATGCCGGGGTTTTACGAGATTGTGATTAAGGTCCCCAGCGACCTTGACGAGCATCTGCCCCGCATTTCTGACAGC
TTTGTGAAGTGGGTGGCCGAGAAGGAATGGGAGTTGCCGCCAGATTCTGACATGGATCTGAATCTGATTGAGCAG
GCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTTCTGACGGAATGGCGCCGTGTGAGTAAGGCCCGGAG
GCCCTTTTCTTTGTGCAATTTGAGAAGGGAGAGAGCTACTTCACATGCACGTGCTCGTGGAAACCACCGGGGTG
AAATCCATGGTTTTTGGGACGTTTCTGAGTCAGATTTCGCGAAAACTGATTTCAGAGAATTTACCGCGGGATCGAG
CCGACTTTGCCAAACTGGTTCGCGGTCAAAAAGACCAGAAATGGCGCCGGAGGCGGGAACAAGTGGTGGATGAG
TGCTACATCCCCAATTACTTGCTCCCCAAAACCCAGCCTGAGCTCCAGTGGGCGTGGACTAATATGGAACAGTAT
TTAAGCGCCTGTTTGAATCTCACGGAGCGTAAACGGTTGGTGGCGCAGCATCTGACGCACGTGTGCGAGACGCAG
GAGCAGAACAAGAGAATCAGAATCCCAATTCTGATGCGCCGGTGATCAGATCAAAAACCTTCAGCCAGGTACATG
GAGCTGGTTCGGGTGGCTCGTGGACAAGGGGATTACCTCGGAGAAGCAGTGGATCCAGGAGGACCAGGCCTCATAC
ATCTCCTTCAATGCGGCCCTCCAACCTCGCGGTCCCAAATCAAGGCTGCCTTGGACAATGCGGGAAAGATTATGAGC
CTGACTAAAACCGCCCCGACTACCTGGTGGGCCAGCAGCCCGTGGAGGACATTTCCAGCAATCGGATTTATAAAA
ATTTTGGAACTAAACGGGTACGATCCCCAATATGCGGCTTCCGTCTTTCTGGGATGGGCCACGAAAAAGTTCGGC
AAGAGGAACACCATCTGGCTGTTTGGGCCTGCAACTACCGGGAAGACCAACATCGCGGAGGCCATAGCCACACT
GTGCCCTTCTACGGGTGCGTAAACTGGACCAATGAGAACTTTCCCTTCAACGACTGTGTGCGACAAGATGGTGATC
TGGTGGGAGGAGGGGAAGATGACCGCAAGGTCTGGAGTCGGCCAAAAGCCATTCTCGGAGGAAGCAAGGTGCGC
GTGGACCAGAAATGCAAGTCCTCGGCCAGATAGACCCGACTCCCGTGATCGTCACCTCCAACACCAACATGTGC
GCCGTGATTGACGGGAACCTCAACGACCTTCGAACACCAGCAGCCGTGCAAGACCGGATGTTCAAATTTGAATC
ACCCGCCGTCTGGATCATGACTTTGGGAAGGTACCAAGCAGGAAGTCAAAGACTTTTTCCGGTGGGCAAAGGAT
CACGTGGTTGAGGTGGAGCATGAATTCTACGTCAAAAAGGTTGGAGCCAAGAAAAGACCCGCCCCCAGTGACGCA
GATATAAGTGAGCCCAAACGGGTGCGCGAGTCAGTTGCGCAGCCATCGACGTCAGACGCGGAAGCTTCGATCAAC
TACGCAGACAGGTACCAAAAACAAATGTTCTCGTCACGTGGGCATGAATCTGATGCTGTTTCCCTGCAGACAATGC
GAGAGAATGAATCAGAATTCAAATATCTGCTTCACTCACGGACAGAAAAGACTGTTTAGAGTGCTTTCCCGTGTCA
GAATCTCAACCCGTTTCTGTCTGTCAAAAAGGCGTATCAGAACTGTGCTACATTTCATCATATCATGGGAAAGGTG
CCAGACGCTTGCACTGCCTGCGATCTGGTCAATGTGGATTTGGATGACTGCATCTTTGAACAATAA

Cap2 Sequence - contains sequentially VP1, VP2, AAP, VP3 (start codons underlined) (SEQ ID NO:32)

ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACACTCTCTCTGAAGGAATAAGACAGTGGTGGAAAGCTC
AAACCTGGCCACCACCACCAAAAGCCCGCAGAGCGGCATAAGGACGACAGCAGGGGTCTTGCTTCTGGGTAC
AAGTACCTCGGACCCTTCAACGGACTCGACAAGGGAGAGCCGGTCAACGAGGCAGACGCCGCGGCCCTCGAGCAC
GACAAAGCCTACGACCGGCAGCTCGACAGCGGAGACAACCCGTACCTCAAGTACAACCACGCCGACGCGGAGTTT
CAGGAGCGCCTTAAAGAAGATACGTCTTTTGGGGGCAACCTCGGACGAGCAGTCTTCCAGGCGAAAAAGAGGGTT
CTTGAACCTCTGGGCCTGGTTGAGGAACCTGTTAAGACGGCTCCGGGAAAAAAGAGGCCGGTAGAGCACTCTCCT
GTGGAGCCAGACTCCTCCTCGGGAACCGGAAAGGCGGGCCAGCAGCCTGCAAGAAAAAGATTGAATTTTGGTCAG
ACTGGAGACGCAGACTCAGTACCTGACCCCCAGCCTCTCGGACAGCCACCAGCAGCCCCCTCTGGTCTGGGAAC
AATACGATGGCTACAGGCAGTGGCGCACCAATGGCAGACAATAACGAGGGCGCCGACGGAGTGGGTAATTCCTCG
GGAAATTGGCATTGCGATTCCACATGGATGGGCGACAGAGTCATCACCACCAGCACCCGAACCTGGGCCCTGCC
ACCTACAACAACCACCTCTACAAAACAAATTTCCAGCCAATCAGGAGCCTCGAACGACAATCACTACTTTGGCTAC

AGCACCCCTTGGGGGTATTTTGA^{CTTCAACAGATTCCACTGCCACTTTTACCACGTGACTGGCAAAGACTCATC}
AACAACA^{ACTGGGGATTCCGACCCAAGAGACTCAACTTCAAGCTCTTTAACATTCAAGTCAAAGAGGTCACGCAG}
AATGACGGTACGACGACGATTGCCAATAACCTTACCAGCACGGTTCAGGTGTTTACTGACTCGGAGTACCAGCTC
CCGTACGTCTCTGGCTCGGCGCATCAAGGATGCCTCCCGCCGTTCACAGCAGACGTCTTCATGGTGCCACAGTAT
GGATACCTCACCTGAACAACGGGAGTCAGGCAGTAGGACGCTCTTCATTTTACTGCCTGGAGTACTTTCTCTCT
CAGATGCTGCGTACCGGAAACAACCTTTACCTTCAGCTACACTTTTGAGGACGTTCTTTCCACAGCAGCTACGCT
CACAGCCAGAGTCTGGACCGTCTCATGAATCCTCTCATCGACCAGTACCTGTATTACTTGAGCAGAACAACACT
CCAAGTGAACACCACGACGAGTCAAGGCTTCAGTTTTCTCAGGCCGGAGCGAGTGACATTCCGGGACCAGTCTAGG
AACTGGCTTCCTGGACCGTGTACCGCCAGCAGCGAGTATCAAAGACATCTGCGGATAACAACAACAGTGAATAC
TCGTGGACTGGAGCTACCAAGTACCACCTCAATGGCAGAGACTCTCTGGTGAATCCGGGCCCCGCCATGGCAAGC
CACAAGGACGATGAAGAAAAGTTTTTCTCAGAGCGGGGTCTCTCATCTTTGGGAAGCAAGGCTCAGAGAAAACA
AATGTGGACATTGAAAAGGTCATGATTACAGACGAAGAGGAAATCAGGACAACCAATCCCGTGGCTACGGAGCAG
TATGGTTCTGTATCTACCAACCTCCAGAGAGGCAACAGACAAGCAGCTACCGCAGATGTCAACACACAAGGCGTT
CTTCCAGGCATGGTCTGGCAGGACAGAGATGTGTACCTTCAGGGGCCATCTGGGCAAAGATTCCACACACGGAC
GGACATTTTACCCCTCTCCCTCATGGGTGGATTTCGGACTTAAACACCCCTCCTCCACAGATTCTCATCAAGAAC
ACCCCGGTACCTGCGAATCCTTCGACCACCTTCAGTGCGGCAAAGTTTGCTTCCTTCATCACACAGTACTCCACG
GGACAGGTACGCGTGGAGATCGAGTGGGAGCTGCAGAAGGAAAAACAGCAAACGCTGGAATCCCGAAATTCAGTAC
ACTTCCA^{ACTACAACAAGTCTGTAAATGTGGACTTTACTGTGGACACTAATGGCGTGTATT}CAGAGCCTCGCCCC
ATTGGCAC^{CAGATACTGACTCGTAATCTGTAA}

Cap5 Sequence – contains sequentially VP1, VP2, AAP, VP3 (start codons underlined) (SEQ ID NO:33)

ATGGCTTTTGTGATCACCTCCAGATTGGTTGGAAGAAGTTGGTGAAGGTCTTCGCGAGTTTTTGGGCCTTGAA
GCGGGCCCA^{CCGAAACCAAAACCCAATCAGCAGCATCAAGATCAAGCCCGTGGTCTTGTGCTGCCTGGTTATAAC}
TATCTCGGACCCGGAACGGTCTCGATCGAGGAGAGCCTGTCAACAGGGCAGACGAGGTTCGCGGAGAGCACGAC
ATCTCGTACAACGAGCAGCTTGAGGCGGGAGACAACCCCTACCTCAAGTACAACCACGCGGACGCCGAGTTTCAG
GAGAAGCTCGCCGACGACACATCCTTCGGGGGAAACCTCGGAAAGGCAGTCTTTCAGGCCAAGAAAAGGGTTCTC
GAACCTTTTGGCCTGGTTGAAGAGGGTGCTAAGACGCGCCCTACCGGAAAGCGGATAGACGACCATTTC^{CAAAA}
AGAAAGAAGGCTCGGACCGAAGAGGACTCCAAGCCTTCCACCTCGTCAGACGCCGAAGCTGGACCCAGCGGATCC
CAGCAGCTGCAAATCCAGCCCAACCAGCCTCAAGTTTGGGAGCTGATACAATGTCTGCGGGAGGTGGCGGCCCA
TTGGGCGACAATAACCAAGGTGCCGATGGAGTGGGCAATGCCTCGGGAGATTGGCATTGCGATTCCACGTGGATG
GGGGACAGAGTCGTACCAAGTCCACCCGAACCTGGGTGCTGCCAGCTACAACAACCACCAGTACCGAGAGATC
AAAAGCGGCTCCGTCGACGGAAGCAACGCCAACGCCTACTTTGGATACAGCACCCCTCGGGGTACTTTGACTTT
AACCGCTTCCACAGCCACTGGAGCCCCCGAGACTGGCAAAGACTCATCAACA^{ACTACTGGGGCTTCAGACCCCGG}
TCCCTCAGAGTCAAAATCTTCAACATTCAAGTCAAAGAGGTACCGGTGCAGGACTCCACCACCACCATCGCCAAC
AACCTCACCTCCACCGTCCAAGTGTTTACGGACGACGACTACCAGTGCCTTACGTCGTCGGAACGGGACCGAG
GGATGCCTGCCGGCCTTCCCTCCGCAGGTCTTTACGCTGCCGCAGTACGGTTACGCGACGCTGAACCGCGACAAC
ACAGAAAATCCCACCGAGAGGAGCAGCTTCTTCTGCCTAGAGTACTTTCCAGCAAGATGCTGAGAACGGGCAAC
AACTTTGAGTTTACCTACAACCTTTGAGGAGGTGCCCTTCCACTCCAGCTTCGCTCCAGTCAGAACCTCTTCAAG
CTGGCCAACCCGCTGGTGGACCACTTGTACCGCTTCGTGAGCACAAATAACACTGGCGGAGTCCAGTTCAAC
AAGAACCTGGCCGGGAGATACGCCAACACCTACAAAAACTGGTTCCCGGGGCCCATGGGCCGAACCCAGGGCTGG
AACCTGGGCTCCGGGTCAACCGCGCCAGTGTACGCGCTTCGCCACGACCAATAGGATGGAGCTCGAGGGCGCG
AGTTACCAGGTGCCCCCGCAGCCGAACGGCATGACCAACAACCTCCAGGGCAGCAACACCTATGCCCTGGAGAAC
ACTATGATCTTCAACAGCCAGCCGGCGAACCCGGGCACCAACCGCCACGTACCTCGAGGGCAACATGCTCATCACC
AGCGAGAGCGAGACGACCGGTGAACCGCGTGGCGTACAACGTCCGGCGGGCAGATGGCCAAACAACAGAGC
TCCACCACTGCCCGCGACCGGCACGTACAACCTCCAGAAATCGTGCCCGGCAGCGTGTGGATGGAGAGGGAC
GTGTACCTCCAAGGACCCATCTGGGCCAAGATCCAGAGACGGGGGCGCACTTTCACCCCTCTCCGGCCATGGGC
GGATTTCGGACTCAAACACCCACCGCCCATGATGCTCATCAAGAACACGCTGTGCCCGGAAATATCACCAGCTTC
TCGGACGTGCCCGTCAGCAGCTTCATCACCCAGTACAGCACCGGGCAGGTACCGTGGAGATGGAGTGGGAGCTC
AAGAAGGAAA^{ACTCCAAGAGGTGGAACCCAGAGATCCAGTACACAAACA}ACTACAACGACCCCCAGTTTGTGGAC
TTTGCCCCGACAGCACCGGGGAATACAGAAGCACCAAGACCTATCGGAACCCGATACCTTACCCGACCCCTTTAA

EXAMPLE 12 – Adenovirus Polynucleotide Sequences

Adenovirus (Ad) polynucleotides can be selected from any serotype, and representative polynucleotides are exemplified below.

E2A Full Sequence (SEQ ID NO:34)

CGACCGCACCTGTGACGAAAGCCGCCCCGCAAGCTGCGCCCCCTGAGTTAGTCATCTGAACTTCGGCCTGGGCGT
CTCTGGGAAGTACCACAGTGGTGGGAGCGGGACTTTCTGGTACACCAGGGCAGCGGGCCAACCTACGGGGATTAA
GGTTATTACGAGGTGTGGTGGTAATAGCCGCTGTTCGAGGAGAATTCGGTTTCGGTGGGCGCGGATTCCGTTGA
CCCCGGATATCATGTGGGGTCCCGCGCTCATGTAGTTTATTTCGGGTTGAGTAGTCTTGGGCAGCTCCAGCCGCAA
GTCCCATTTGTGGCTGGTAACTCCACATGTAGGGCGTGGGAATTTCTTGCTCATAATGGCGCTGACGACAGGTG
CTGGCGCCGGGTGTGGCCGCTGGAGATGACGTAGTTTTCGCGCTTAAATTTGAGAAAGGGCGCGAAACTAGTCCT
TAAGAGTCAGCGCGCAGTATTTGCTGAAGAGAGCCTCCGCGTCTTCCAGCGTGCGCCGAAGCTGATCTTCGCTTT
TGTGATACAGGCAGCTGCGGGTGAAGGAGCGCAGAGACCTGTTTTTATTTTCAGCTCTTGTCTTGGCCCCGTC
TTTGTGTGAATATAGCATACAGAGTGGGAAAAATCCTATTTCTAAGCTCGCGGGTCGATACGGGTTCTTGGGCG
CCAGACGCAGCGCTCCTCCTCCTGCTGCTGCCGCCGCTGTGGATTCTTGGGCTTTGTCTAGAGTCTTGTCTATCCG
GTCGCCCTTTGCTTCTGTGTGACCGCTGCTGTTGCTGCCGCTGCCGCTGCCGCCGCTGCAGTAGGGGCTGTAGAGA
TGACGGTAGTAATGCAGGATGTTACGGGGGAAGGCCACGCCGTGATGGTAGAGAAGAAAGCGGCGGGCGAAGGAG
ATGTTGCCCCACAGTCTTGCAAGCAAGCAACTATGGCGTTCTTGTGCCCGCGCCACGAGCGGTAGCCTTGGCGC
TGTTGTTGCTCTTGGGCTAACGGCGGCGGCTGCTTAGACTTACCGGCCCTGGTTCCAGTGGTGTCCCATCTACGG
TTGGGTGCGCGAACAGGCAGTGCCGGCGGCGCCTGAGGAGCGGAGGTTGTAGCGATGCTGGGAACGGTTGCCAAT
TTCTGGGGCGCCGGCGAGGGGAATGCGACCGAGGGTGACGGTGTTCGCTGACACCTCTTCGGCCTCGGAAGCT
TCGTCTAGGCTGTCCAGTCTTCCATCATCTCCTCCTCCTCGTCCAAAACCTCCTCTGCCTGACTGTCCAGTAT
TCCTCCTCGTCCGTGGGTGGCGGCGGCGGCAGCTGCAGCTTCTTTTGGGTGCCATCTGGGAAGCAAGGGCCCG
CGGCTGCTGATAGGGCTGCGGCGGCGGGGGGATTGGGTTGAGCTCCTCGCCGACTGGGGGTCCAGGTAAACCCC
CCGTCCCTTTCTAGCAGAACTCTTGGCGGGCTTTGTTGATGGCTTGCAATTGGCCAAGGATGTGGCCCTGGGT
AATGACGCAGGCGGTAAGCTCCGCATTTGGCGGGCGGGATTGGTCTTCGTAGAACCTAATCTCGTGGGCGTGGTA
GTCCTCAGGTACAAATTTGCGAAGGTAAGCCGACGTCCACAGCCCCGGAGTGAGTTTCAACCCCGAGCCGCGGA
CTTTTCGTAGGCGAGGGACCTGCGAGTCAAAGGTACCGATAATTTGACTTTCGCTAAGCAGTTGCGAATTGCA
GACCAGGGAGCGGTGCGGGGTGCATAGGTTGCAGCGACAGTGACACTCCAGTAGGCCGTACCGCTCACGTCTTC
CATGATGTCGGAGTGGTAGGCAAGGTAGTTGGCTAGCTGCAGAAAGTAGCAGTGACCCCAAAGCGGCGGAGGGCA
TTCACGGTACTTAATGGGCACAAAGTCGCTAGGAAGCGCACAGCAGGTGGCGGGCAGAATTCCTGAACGCTCTAG
GATAAGTTCTTAAGTTTTGCAACATGCTTTGACTGGTGAAGTCTGGCAGACCCTGTTGCAGGGTTTTAAGCAG
GCGTTTCGGGGAAGATAATGTCCGCAAGGTGCGCGGCCACGGAGCGCTCGTTGAAGGCCGTCCATAGGTCTTCAA
GTTTTGCTTTAGCAGCTTCTGCAGCTCCTTTAGGTTGCGCTCCTCCAGGCATTGCTGCCACACGCCCATGGCCGT
TTGCCAGGTGTAGCACAGAAATAAGTAAACGCAGTCGCGGACGTAGTCGCGGCGCGCCTCGCCCTTGAGCGTGGA
ATGAAGCACGTTTTTGGCCGAGGCGGTTTTCTGTGCAAAATTCGAAGGTAGGAGACCAGGTTGCAGAGCTCCACGTT
GGAAATTTTGAGGCCTGGCGCACGTAGCCCTGGCGAAAGGTGTAGTGCAACGTTTCTCTAGCTTGCCTGCAT
CTCCGGGTGACGAAAGAACCGCTGCATGCACTCAAGCTCCACGGTAACAAGCACTGCGGCCATCATTAGCTTGCG
TCGCTCCTCCAAGTCGGCAGGCTCGCGCGTCTCAAGCCAGCGCGCCAGCTGCTCATCGCCAACTGCGGGTAGGCC
CTCCTCGGTTTGTCTTGAAGTTGCATCCCTCTCCAGGGGTCTGTGCACGGCGCACGATCAGCTCGCTCATGAC
TGTGCTCATAACCTTGGGGGGTAGGTTAAGTGCCGGGTAGGCAAAGTGGGTGACCTCGATGCTGCGTTTCAGCAC
GGCTAGGCGCGCGTTGTACCCCTCAAGTTCCACCAGCACCTCCACAGTGACTTTCATTTTCGCTGTTTTCTTGTG
CAGAGCGTTTGCCGCGCGTTTCTCGTCGCGTCCAAGACCTCAAAGATTTTTTGGCACTTCGTCGAGCGAGGCGAT
ATCAGGTATGACAGCGCCCTGCCGCAAGGCCAGCTGCTTGTCCGCTCGGCTGCGGTTGGCACGGCAGGATAGGGG
TATCTTGACGTTTTTGGAAAAAGATGTGATAGGTGGCAAGCACCTCTGGCACGGCAAATACGGGGTAGAAGTTGAG
GCGCGGGTTGGGCTCGCATGTGCCGTTTTCTTGGCGTTTGGGGGGTACGCGCGGTGAGAACAGGTGGCGTTTCGTA
GGCAAGGCTGACATCCGCTATGGCGAGGGGCACATCGCTGCGCTCTTGCAACGCGTCGCAGATAATGGCGCACTG
GCGCTGCAGATGCTTCAACAGCACGTCTGCTCCACATCTAGGTAGTCGCCATGCCCTTGGTCCCCCGCCCCGAC
TTGTTCTCTGTTTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
TTGCTTACAAAACCTGGGTCCGTGCTGATAATCACTTCCCTCCTCAAGCGGGAGTGCCTGACGCGGAAGGT
GGTAGGCGCGTTGGCGGCATCGGTGGAGGCGGTGGTGGCGAACTCAAAGGGGGCGGTTAGGCTGTCTCCTTCTC
GACTGACTCCATGATCTTTTTCTGCTATAGGAGAAGGAAATGGCCAGTCGGGAAGAGGAGCAGCGCGAAACCAC
CCCCGAGCGCGGACGCGGTGCGGCGGACGTCCACCAACCATGGAGGACGTGTCGTCCCCGTGCGCGTGGCGCC
GCCTCCCCGCGCGCCCCCAAAAAGCGGCTGAGGCGGCGTCTCGAGTCCGAGGACGAAGAAGACTCGTCACAAGA
TGCGCTGGTGCCGCGCACACCCAGCCGCGGCCATCGACCTCGACGGCGGATTTGGCCATTGCGTCCAAAAGAA
AAAGAAGCGCCCCCTCTCCAAGCCGAGCGCCGCCATCCCCAGAGGTGATCGTGGACAGCGAGGAAGAAAGAGA
AGATGTGGCGCTACAAATGGTGGGTTTCAAGCAACCCACCGGTGCTAATCAAGCACGGCAAGGGAGGTAAGCGCAC
GGTGCGGCGGCTGAATGAAGACGACCCAGTGGCGCGGGGTATGCGGACGCAAGAGGAAAAGGAAGAGTCCAGTGA
AGCGGAAAGTGAAAGCACGGTGATAAACCCGCTGAGCCTGCCGATCGTGTCTGCGTGGGAGAAGGGCATGGAGGC
TGCGCGCGCGTTGATGGACAAGTACCAGTGGATAACGATCTAAAGGCAAACCTTCAAGCTACTGCCTGACCAAGT
GGAAGCTCTGGCGGCGGTATGCAAGACCTGGCTAAACGAGGAGCACCGCGGGTTGCAGCTGACCTTACCAGCAA

CAAGACCTTTGTGACGATGATGGGGCGATTCTGCAGGCGTACCTGCAGTCGTTTGCAGAGGTAACCTACAAGCA
 CCACGAGCCCACGGGCTGCGCGTTGTGGCTGCACCGCTGCGCTGAGATCGAAGGCGAGCTTAAGTGTCTACACGG
 GAGCATTATGATAAATAAGGAGCACGTGATTGAAATGGATGTGACGAGCGAAAACGGGCGAGCGCGCTGAAGGA
 GCAGTCTAGCAAGGCCAAGATCGTGAAGAACCAGTGGGGCCGAAATGTGGTGCAGATCTCCAACACCGACGCAAG
 GTGCTGCGTGCATGACGCGGCCTGTCCGGCCAATCAGTTTTCCGGCAAGTCTTGCGGCATGTTCTTCTCTGAAGG
 CGCAAAGGCTCAGGTGGCTTTTTAAGCAGATCAAGGCTTTCATGCAGGCGCTGTATCCTAACGCCAGACCGGGCA
 CGGTACCTTCTGATGCCACTACGGTGCAGGTGCAACTCAAAGCCTGGGCATGCACCCTTTTTGGGAAGGCAGCT
 ACCAAAGTTGACTCCGTTTCGCCCTGAGCAACGCGGAGGACCTGGACGCGGATCTGATCTCCGACAAGAGCGTGCT
 GGCCAGCGTGCACCACCGGCGCTGATAGTGTTCAGTGCTGCAACCTGTGTATCGCAACTCGCGCGCGCAGGG
 CGGAGGCCCAACTGCGACTTCAAGATATCGGCGCCGACCTGCTAAACGCGTTGGTGATGGTGCGCAGCCTGTG
 GAGTGAAAACCTTACCCGAGCTGCCGCGGATGGTTGTGCTGAGTTTAAAGTGAGCACTAAACACAGTATCGCAA
 CGTGTCCTGCCAGTGGCGCATAGCGATGCGCGGCAGAACCCCTTTGATTTTTTAAACGGCGCAGACGGCAAGGTT
 GGGGGGTAAATAACTACCCGAGAGTGTACAAATAAAACATTTGCCCTTTATTTGAAAGTGTCTCCTAGTACATTAT
 TTTTACATGTTTTTCAAGTGACAAAAAGAAAGTGGCGCTCCTAATCTGCGCACTGTGGCTGCGGAAGTAGGGCGAG
 TGGCGCTCCAGGAAGCTGTAGAGCTGTTCTGGTTGCGACGCAGGGTGGGCTGTACCTGGGGACTGTTAAGCATG
 GAGTTGGGTACC

E2A ORF Sequence (SEQ ID NO:35)

ATGGCCAGTCGGGAAGAGGAGCAGCGCGAAACCACCCCGAGCGCGGACGCGGTGCGGCGCGACGTCCACCAACC
 ATGGAGGACGTGTCTGTCCTCCCGTGCCTGCGCGCCGCTCCCGCGCGCCCCCAAAAAAGCGGCTGAGGCGGCGT
 CTCGAGTCCGAGGACGAAGAAGACTCGTCAAAAGATGCGCTGGTGCCGCGCACACCCAGCCCGCGGCCATCGACC
 TCGACGGCGGATTTGGCCATTGCGTCCAAAAAGAAAAAGAACGCCCTCTCCCAAGCCGAGCGCCCGCCATCC
 CCAGAGGTGATCGTGGACAGCGAGGAAGAAAGAGAAGATGTGGCGCTACAAATGGTGGGTTTTAGCAACCCACCG
 GTGCTAATCAAGCACGGCAAGGGAGGTAAGCGCACGGTGCGGCGGCTGAATGAAGACGACCCAGTGGCGCGGGGT
 ATGCGGACGCAAGAGGAAAAGGAAGAGTCCAGTGAAGCGGAAAGTGAAAGCACGGTGATAAACCCGCTGAGCCTG
 CCGATCGTGTCTGCGTGGGAGAAGGGCATGGAGGCTGCGCGCGCGTTGATGGACAAGTACCACGTGGATAACGAT
 CTAAAGGCAAACTTCAAGCTACTGCCTGACCAAGTGGAAGCTCTGGCGGCCGTATGCAAGACCTGGCTAAACGAG
 GAGCACCGCGGGTTGACGTGACCTTACCAGCAACAAGACCTTTGTGACGATGATGGGGCGATTCTCGCAGGCG
 TACCTGCAGTCGTTTGCAGAGGTAACCTACAAGCACACGAGCCACGGGCTGCGCGTTGTGGCTGCACCGCTGC
 GCTGAGATCGAAGGCGAGCTTAAGTGTCTACACGGGAGCATTATGATAAATAAGGAGCACGTGATTGAAATGGAT
 GTGACGAGCGAAAACGGGCAGCGCGCGCTGAAGGAGCAGTCTAGCAAGGCCAAGATCGTGAAGAACCAGTGGGGC
 CGAAATGTGGTGAGATCTCCAACACCGACGCAAGGTGCTGCGTGCATGACGCGGCCGTGTCGGCCAATCAGTTT
 TCCGGCAAGTCTTGCGGCATGTTCTTCTCTGAAGGCGCAAAGGCTCAGGTGGCTTTTTAAGCAGATCAAGGCTTTC
 ATGCAGGCGCTGTATCCTAACGCCAGACCGGGCACGGTCACCTTCTGATGCCACTACGGTGCAGTGCAACTCA
 AAGCCTGGGCATGCACCCTTTTTGGGAAGGCAGCTACCAAAGTTGACTCCGTTTCGCCCTGAGCAACGCGGAGGAC
 CTGGACGCGGATCTGATCTCCGACAAGAGCGTGCTGGCCAGCGTGACCAACCGGCGCTGATAGTGTTCAGTGC
 TGCAACCTGTGTATCGCAACTCGCGCGCGCAGGGCGGAGGCCCAACTGCGACTTCAAGATATCGGCGCCCGAC
 CTGCTAAACGCGTTGGTGATGGTGCGCAGCCGTGTTGAGTGAAAACCTTACCGAGCTGCCGCGGATGGTTGTGCC
 GAGTTTTAAGTGAGCACTAAACACAGTATCGCAACGTGTCCCTGCCAGTGGCGCATAGCGATGCGCGGCAGAAC
 CCCTTTGATTTTTTAA

E4 Full Sequence (SEQ ID NO:36)

CCCGGGCGTTTTAGGGCGGAGTAACTTGCATGTATTGGGAATTGTAGTTTTTTTTTAAATGGGAAGTGACGTATCG
 TGGGAAAACGGAAGTGAAGATTTGAGGAAGTTGTGGTTTTTTTGGCTTTCTGTTTCTGGGCGTAGGTTTCGCGTGC
 GTTTTCTGGGTGTTTTTGTGGACTTTAACCGTTACGTCATTTTTTAGTCCTATATATACTCGCTCTGTACTTGG
 CCCTTTTTTACACTGTGACTGATTGAGCTGGTGCCGTGTCGAGTGGTGTTTTTTAATAGGTTTTTTTTACTGGTAAG
 GCTGACTGTTATGGCTGCCGCTGTGGAAGCGCTGTATGTTGTTCTGGAGCGGGAGGGTGCTATTTTGCCTAGGCA
 GGAGGGTTTTTACAGGTGTTTATGTGTTTTTCTCTCTATTAATTTTGTATACCTCCTATGGGGGCTGTAATGTT
 GTCTCTACGCTGCGGGTATGTATTCCCCCGGGCTATTTTCGGTTCGCTTTTTAGCACTGACCGATGTTAACCAACC
 TGATGTGTTTTACCGAGTCTTACATTATGACTCCGGACATGACCGAGGAACTGTCGGTGGTGCTTTTTAATCACGG
 TGACCAGTTTTTTTTACGGTCACGCCGGCATGGCCGTAGTCCGTCTTATGCTTATAAGGGTTGTTTTTCTGTGTT
 AAGACAGGCTTCTAATGTTTAAATGTTTTTTTTTTTGTATTTTATTTTGTGTTTAAATGCAGGAACCCGAGACA
 TGTTTGAGAGAAAAATGGTGCTTTTTTCTGTGGTGGTTCCGGAACCTACCTGCCTTTATCTGCATGAGCATGACT
 ACGATGTGCTTGTCTTTTTTGCGCGAGGCTTTGCTGATTTTTTGTAGCAGCACCTTGCATTTTATATCGCCGCCCA
 TGCAACAAGCTTACATAGGGGCTACGCTGGTTAGCATAGCTCCGAGTATGCGTGTCTAATCAGTGTGGGTTCTT
 TTGTCATGGTTTCTGGCGGGGAAGTGGCCGCGCTGGTCCGTGCAGACCTGCACGATTATGTTACAGCTGGCCCTGC

GAAGGGACCTACGGGATCGCGGTATTTTTGTTAATGTTCCGCTTTTGAATCTTATACAGGTCTGTGAGGAACCTG
AATTTTTGCAATCATGATTTCGCTGCTTGAGGCTGAAGGTGGAGGGCGCTCTGGAGCAGATTTTTACAATGGCCGG
ACTTAATATTCGGGATTTGCTTAGAGACATATTGATAAGGTGGCGAGATGAAAATTATTTGGGCATGGTTGAAGG
TGCTGGAATGTTTATAGAGGAGATTACCCCTGAAGGGTTTAGCCTTTACGTCCACTTGGACGTGAGGGCAGTTTG
CCTTTTGGAAAGCCATTGTGCAACATCTTACAAATGCCATTATCTGTTCTTTGGCTGTAGAGTTTGACCACGCCAC
CGGAGGGGAGCGCGTTCACTTAATAGATCTTCATTTTGAAGTTTTGGATAATCTTTTGGAAATAAAAAAAAAAAAA
CATGGTTCTTCCAGCTCTTCCCGCTCCTCCCGTGTGTGACTCGCAGAACGAATGTGTAGGTTGGCTGGGTGTGGC
TTATTCTGCGGTGGTGGATGTTATCAGGGCAGCGGCGCATGAAGGAGTTTACATAGAACCCGAAGCCAGGGGGCG
CCTGGATGCTTTGAGAGAGTGGATATACTACAATACTACACAGAGCGAGCTAAGCGACGAGACCGGAGACGCAG
ATCTGTTTGTACGCCCCGACCTGGTTTTGCTTCAGGAAATATGACTACGTCCGGCGTTCCATTTGGCATGACAC
TACGACCAACACGATCTCGGTTGTCTCGGCGCACTCCGTACAGTAGGGATCGCCTACCTCCTTTTGAGACAGAGA
CCCGCGCTACCATCTGGAGGATCATCCGCTGCTGCCGAATGTAACTTTGACAATGCACAACGTGAGTTACG
TGCGAGGTCTTCCCTGCACTGTGGGATTTACGCTGATTTCAGGAATGGGTGTGTTCCCTGGGATATGGTTCTGACGC
GGGAGGAGCTTGTAATCCTGAGGAAGTGTATGCACGTGTGCCTGTGTTGTGCCAACATTGATATCATGACGAGCA
TGATGATCCATGGTTACGAGTCTTGGGCTCTCCACTGTCATTGTTCCAGTCCCGGTTCCCTGCAGTGCATAGCCG
GCGGGCAGGTTTTGGCCAGCTGGTTTAGGATGGTGGTGGATGGCGCCATGTTTAATCAGAGGTTTTATATGGTACC
GGGAGGTGGTGAATTACAACATGCCAAAAGAGGTAATGTTTATGTCCAGCGTGTGTTATGAGGGGTCGCCACTTAA
TCTACCTGCGCTTGTGGTATGATGGCCACGTGGGTCTGTGGTCCCCGCCATGAGCTTTGGATACAGCGCCTTGC
ACTGTGGGATTTTGAACAATATTGTGGTGTGTGCTGCAGTTACTGTGCTGATTTAAGTGAGATCAGGGTGCCT
GCTGTGCCCCGAGGACAAGGCGTCTCATGCTGCGGGCGGTGCGAATCATCGCTGAGGAGACCACTGCCATGTTGT
ATTCTGACAGGACGGAGCGGCGGCGGCGAGCAGTTTATTCGCGCGCTGTGTCAGCACCACCGCCCTATCCTGATGC
ACGATTATGACTCTACCCCCATGTAGGCGTGGACTTCCCTTCGCCGCCCGTTGAGCAACCGCAAGTTGGACAGC
AGCCTGTGGCTCAGCAGCTGGACAGCGACATGAACCTTAAGCGAGCTGCCCGGGGAGTTTATTAATATCACTGATG
AGCGTTTGGCTCGACAGGAAACCGTGTGGAATATAACACCTAAGAATATGTCTGTTACCCATGATATGATGCTTT
TTAAGGCCAGCCGGGGAGAAAGGACTGTGTACTCTGTGTGTTGGGAGGGAGGTGGCAGGTTGAATACTAGGGTTC
TGTGAGTTTGATTAAAGGTACGGTGATCAATATAAGCTATGTGGTGGTGGGGCTATACTACTGAATGAAAAATGAC
TTGAAATTTTCTGCAATTGAAAAATAAACACGTTGAAACATAACATGCAACAGGTTACGATTCTTTATTCCTGG
GCAATGTAGGAGAAGGTGTAAGAGTTGGTAGCAAAAGTTTCAGTGGTGTATTTTCCACTTTCCAGGACCATGTA
AAAGACATAGAGTAAGTGCTTACCTCGCTAGTTTCTGTGGATTCACTAGAA

E4 Orf6 Sequence (SEQ ID NO:37)

ATGACTACGTCCGGCGTTCCATTTGGCATGACACTACGACCAACACGATCTCGGTTGTCTCGGCGCACTCCGTAC
AGTAGGGATCGCCTACCTCCTTTTGGACAGAGACCCGCGCTACCATACTGGAGGATCATCCGCTGCTGCCCGAA
TGTAACACTTTGACAATGCACAACGTGAGTTACGTGCGAGGTCTTCCCTGCAGTGTGGGATTTACGCTGATTACG
GAATGGGTTGTTCCCTGGGATATGGTTCTGACGCGGGAGGAGCTTGTAACTCCTGAGGAAGTGTATGCACGTGTGC
CTGTGTTGTGCCAACATTGATATCATGACGAGCATGATGATCCATGGTTACGAGTCTTGGGCTCTCCACTGTCAT
TGTTCCAGTCCCGGTTCCCTGCAGTGCATAGCCGGCGGGCAGGTTTTGGCCAGCTGGTTTAGGATGGTGGTGGAT
GGCGCCATGTTTAAATCAGAGGTTTATATGGTACCGGGAGGTGGTGAATTACAACATGCCAAAAGAGGTAATGTTT
ATGTCCAGCGTGTGTTATGAGGGGTCGCCACTTAATCTACCTGCGCTTGTGGTATGATGGCCACGTGGGTTCTGTG
GTCCCCGCCATGAGCTTTGGATACAGCGCCTTGCACTGTGGGATTTTGAACAATATTGTGGTGTGTGCTGCAGT
TACTGTGCTGATTTAAGTGAGATCAGGGTGCCTGCTGTGCCCGGAGGACAAGGCGTCTCATGCTGCGGGCGGTG
CGAATCATCGCTGAGGAGACCACTGCCATGTTGTATTCTGTCAGGACGGAGCGGCGGCGGAGCAGTATTATTCGC
GCGCTGTGTCAGCACCACCGCCCTATCCTGATGCACGATTATGACTCTACCCCCATGTAG

VA Sequence (VA transcripts I and II are underlined) (SEQ ID NO:38)

CGTAATCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGAAAGTCGCGGAC
GCGGTTCCAGATGTTGCGCAGCGGCAAAAAGTGCTCCATGGTTCGGGACGCTCTGGCCGGTGAGGCGTGCGCAGTC
GTTGACGCTCTAGACCGTGCAAAAGGAGAGCCTGTAAAGCGGGCACTCTCCGTGGTCTGGTGGATAAATTCGAA
GGGTATCATGGCGGACGACCGGGGTTTGAACCCCGGATCCGGCCGTCCGCCGTGATCCATGCGGTTACCGCCCGC
GTGTCGAACCCAGGTGTGCGACGTGACACAACGGGGGAGCGCTCCTTTTGGCTTCCTTCCAGGCGCGGCGGCTGC
TGCGCTAGCTTTTTTGGCCACTGGCCGCGCGCGGCGTAAGCGTTAGGCTGGAAAGCGAAAGCATTAAGTGGCTC
GCTCCCTGTAGCCGGAGGGTTATTTTCCAAGGGTTGAGTCGAGGACCCCGGTTTCGAGTCTCGGGCCGGCCGA
CTGCGGCGAACGGGGGTTTGCTCCCCGTATGCAAGACCCCGCTTGCAAATTCCTCCGGAAACAGGGACGAGCC
CCTTTTTTGTCTTTCCAGATGCATCCGGTGTGCGGCAGATGCGCCCCCTCCTCAGCAGCGGCAAGAGCAAGA
GCAGCGGCAGACATGCAGGGCACCTCCCCCTTCTCCTACCGCGTCAGGAGGGGCAACATCCTACATCGA

Sequences for E1A and E1B are both contained within Accession AY339865.1

Ad5 E1A

Two proteins can be transcribed, a 32 kDa protein (first accession number) and a 27 kDa protein (second accession number). These are both splice variants from the transcript:

Accession 1: AAQ19284.1

Accession 2: AAQ19285.1

(SEQ ID NO:39)

ATGAGACATATTATCTGCCACGGAGGTGTTATTACCGAAGAAATGGCCGCCAGTCTTTTGGACCAGCTGATCGAA
GAGGTACTGGCTGATAATCTTCCACCTCCTAGCCATTTTGAACCACCTACCCTTCACGAACTGTATGATTTAGAC
GTGACGGCCCCCGAAGATCCCAACGAGGAGGCGGTTTCGCAGATTTTCCCGACTCTGTAATGTTGGCGGTGCAG
GAAGGGATTGACTTACTCACTTTTCCGCCGGCGCCCGGTTCTCCGAGCCGCCTCACCTTTCCCGGCAGCCCGAG
CAGCCGGAGCAGAGAGCCTTGGGTCCGGTTTCTATGCCAAACCTTGTACCGGAGGTGATCGATCTTACCTGCCAC
GAGGCTGGCTTTCCACCCAGTGACGACGAGGATGAAGAGGGTGAGGAGTTTGTGTTAGATTATGTGGAGACCCCC
GGGCACGGTTGCAGGTCTTGTCTATTATCACCGGAGGAATACGGGGGACCCAGATATTATGTGTTGCTTTGCTAT
ATGAGGACCTGTGGCATGTTTGTCTACAGTCTGTGTCTGAACCTGAGCCTGAGCCCAGCCAGAACCGGAGCCT
GCAAGACCTACCCGCCGTCTAAAATGGCGCCTGCTATCCTGAGACGCCCAGACATCACCTGTGTCTAGAGAATGC
AATAGTAGTACGGATAGCTGTGACTCCGGTCTTCTAACACACCTCCTGAGATACACCCGGTGGTCCCGCTGTGC
CCCATTAACAGTTGCCGTGAGAGTTGGTGGGCGTCGCCAGGCTGTGGAATGTATCGAGGACTTGCTTAACGAG
CCTGGGCAACCTTTGGACTTGAGCTGTAAACGCCCCAGGCCATAA

(SEQ ID NO:40)

ATGAGACATATTATCTGCCACGGAGGTGTTATTACCGAAGAAATGGCCGCCAGTCTTTTGGACCAGCTGATCGAA
GAGGTACTGGCTGATAATCTTCCACCTCCTAGCCATTTTGAACCACCTACCCTTCACGAACTGTATGATTTAGAC
GTGACGGCCCCCGAAGATCCCAACGAGGAGGCGGTTTCGCAGATTTTCCCGACTCTGTAATGTTGGCGGTGCAG
GAAGGGATTGACTTACTCACTTTTCCGCCGGCGCCCGGTTCTCCGAGCCGCCTCACCTTTCCCGGCAGCCCGAG
CAGCCGGAGCAGAGAGCCTTGGGTCCGGTTTCTATGCCAAACCTTGTACCGGAGGTGATCGATCTTACCTGCCAC
GAGGCTGGCTTTCCACCCAGTGACGACGAGGATGAAGAGGGTCTGTGCTGAACCTGAGCCTGAGCCCGAGCCCA
GAACCGGAGCCTGCAAGACCTACCCGCCGTCTAAAATGGCGCCTGCTATCCTGAGACGCCCAGACATCACCTGTG
TCTAGAGAATGCAATAGTAGTACGGATAGCTGTGACTCCGGTCTTCTAACACACCTCCTGAGATACACCCGGTG
GTCCCGCTGTGCCCCATTAAACAGTTGCCGTGAGAGTTGGTGGGCGTCGCCAGGCTGTGGAATGTATCGAGGAC
TTGCTTAACGAGCCTGGGCAACCTTTGGACTTGAGCTGTAAACGCCCCAGGCCATAA

Ad5 E1B_19K

Accession: AAQ19286.1

(SEQ ID NO:41)

ATGGAGGCTTGGGAGTGTGTTTGAAGATTTTTCTGCTGTGCGTAACTTGCTGGAACAGAGCTCTAACAGTACCTCT
TGGTTTTGGAGGTTTCTGTGGGGCTCATCCAGGCAAGTTAGTCTGCAGAAATTAAGGAGGATTACAAGTGGGAA
TTTGAAGAGCTTTTGAAATCCTGTGGTGAGCTGTTTGATTCTTTGAATCTGGGTACCCAGGCGCTTTTCCAAGAG
AAGGTCATCAAGACTTTGGATTTTTCCACACCGGGGCGCGCTGCGGCTGCTGTTTGCTTTTTTTGAGTTTTATAAAG
GATAAATGGAGCGAAGAAACCCATCTGAGCGGGGGGTACCTGCTGGATTTTCTGGCCATGCATCTGTGGAGAGCG
GTTGTGAGACACAAGAATCGCCTGCTACTGTTGTCTTCCGTCCGCCCGGCGATAATACCGACGAGGAGCAGCAG
CAGCAGCAGGAGGAGGACCGGCGGCGGAGGAGCAGAGGCCATGGAACCCGAGAGCCGGCCTGGACCCGATCCG
GAATGA

Ad5 E1B_55K

Accession: AAQ19287.1

(SEQ ID NO:42)

ATGGAGCGAAGAAACCCATCTGAGCGGGGGGTACCTGCTGGATTTTCTGGCCATGCATCTGTGGAGAGCGGTTGT
GAGACACAAGAATCGCCTGCTACTGTTGTCTTCCGTCCGCCCGGCGATAATACCGACGAGGAGCAGCAGCAGCA
GCAGGAGGAAGCCAGGCGGCGGCGGCGGAGGAGCAGAGCCATGGAACCCGAGAGCCGGCCTGGACCCCTCGGGAATG
AATGTTGTACAGGTGGCTGAACTGTATCCAGAACTGAGACGCATTTTGACAAATTACAGAGGATGGGCAGGGGCTA
AAGGGGGTAAAGAGGGAGCGGGGGGCTTGTGAGGCTACAGAGGAGGCTAGGAATCTAGCTTTTAGCTTAATGACC
AGACACCGTCTGAGTGTATTACTTTTCAACAGATCAAGGATAATTGCGCTAATGAGCTTGATCTGCTGGCGCAG

AAGTATTCCATAGAGCAGCTGACCACTTACTGGCTGCAGCCAGGGGATGATTTTGAGGAGGCTATTAGGGTATAT
GCAAAGGTGGCACTTAGGCCAGATTGCAAGTACAAGATCAGCAAACCTTGTAATATCAGGAATTGTTGCTACATT
TCTGGGAACGGGGCCGAGGTGGAGATAGATACGGAGGATAGGGTGGCCTTTAGATGTAGCATGATAAATATGTGG
CCGGGGGTGCTTGGCATGGACGGGGTGGTTATTATGAATGTAAGGTTTACTGGCCCCAATTTTAGCGGTACGGTT
TTCCTGGCCAATACCAACCTTATCCTACACGGTGTAAGCTTCTATGGGTTTAAACAATACCTGTGTGGAAGCCTGG
ACCGATGTAAGGGTTTCGGGGCTGTGCCTTTTACTGCTGCTGGAAGGGGGTGGTGTGTCGCCCCAAAAGCAGGGCT
TCAATTAAGAAATGCCTCTTTGAAAGGTGTACCTTGGGTATCCTGTCTGAGGGTAACTCCAGGGTGCGCCACAAT
GTGGCCTCCGACTGTGGTTGCTTCATGCTAGTGAAAAGCGTGGCTGTGATTAAGCATAACATGGTATGTGGCAAC
TGCGAGGACAGGGCCTCTCAGATGCTGACCTGCTCGGACGGCAACTGTCACCTGCTGAAGACCATTACGTTAGCC
AGCCACTCTCGCAAGGCCTGGCCAGTGTGTTGAGCATAACATACTGACCCGCTGTTTCTTGCATTTGGGTAAACAGG
AGGGGGGTGTTTCTACCTTACCAATGCAATTTGAGTCACACTAAGATATTGCTTGAGCCCCGAGAGCATGTCCAAG
GTGAACCTGAACGGGGTGTGTTGACATGACCATGAAGATCTGGAAGGTGCTGAGGTACGATGAGACCCGCACCAGG
TGCAGACCTGCGAGTGTGGCGGTAAACATATTAGGAACACGCTGTGATGCTGGATGTGACCGAGGAGCTGAGG
CCCGATCACTTGGTGTGCTGGCCTGCACCCGCGCTGAGTTTGGCTCTAGCGATGAAGATACAGATTGA

Sequences for E2A and E4A are both contained within Accession MN088492

Ad5 E2A orf:

Accession: QHX41645.1

(SEQ ID NO:43)

ATGGCCAGTCGGGAAGAGGAGCAGCGCGAAACCACCCCCGAGCGCGGACGCGGTGCGGCGCGACGTCCACCAACC
ATGGAGGACGTGTGCTCCCGTCGCCGTCGCCGCCGCTCCCGCGCGCCCCCAAAAAAGCGGCTGAGGCGGCGT
CTCGAGTCCGAGGACGAAGAAGACTCGTGCACAAGATGCGCTGGTGCCGCGCACACCCAGCCCGCGGCCATCGACC
TCGACGGCGGATTTGGCCATTGCGTCCAAAAAGAAAAAGAAAGCGCCCCCTCTCCCAAGCCCGAGCGCCCGCCATCC
CCAGAGGTGATCGTGGACAGCGAGGAAGAAAGAGAAAGATGTGGCGCTACAAATGGTGGGTTTCAGCAACCCACCG
GTGCTAATCAAGCACGGCAAGGGAGGTAAGCGCACGGTGCGGCGGCTGAATGAAGACGACCCAGTGGCGCGGGGT
ATGCGGACGCAAGAGGAAAAGGAAGAGTCCAGTGAAGCGGAAAGTGAAAGCACGGTGATAAACCCGCTGAGCCTG
CCGATCGTGTCTGCGTGGGAGAAGGGCATGGAGGCTGCGCGCGCGTTGATGGACAAGTACCACGTGGATAACGAT
CTAAAGGCAAACCTCAAGCTACTGCGCTGACCAAGTGGAAGCTCTGGCGGCCGTATGCAAGACCTGGCTAAACGAG
GAGCACCGCGGGTTGCGAGCTGACCTTCACAGCAACAAGACCTTTGTGACGATGATGGGGCGATTCTGCGAGGCG
TACCTGCAGTCGTTTGCAGAGGTAACCTACAAGCACCACGAGCCACGGGCTGCGCGTTGTGGCTGCACCGCTGC
GCTGAGATCGAAGGCGAGCTTAAGTGTCTACACGGGAGCATTATGATAAATAAGGAGCACGTGATTGAAATGGAT
GTGACGAGCGAAAACGGGCAGCGCGCGCTGAAGGAGCAGTCTAGCAAGGCCAAGATCGTGAAGAACCAGTGGGGC
CGAAATGTGGTGCAGATCTCCAACACCGACGCAAGGTGCTGCGTGCATGACGCGGCCGTGCCGGCCAATCAGTTT
TCCGGCAAGTCTTGCGGCATGTTCTTCTCTGAAGGCGCAAAGGCTCAGGTGGCTTTTAAAGCAGATCAAGGCTTTC
ATGCAGGCGCTGTATCCTAACGCCCAGACCGGGCACGGTCACCTTCTGATGCCACTACGGTGCAGTGCAACTCA
AAGCCTGGGCATGCACCTTTTTTGGGAAGGCAGCTACCAAAGTTGACTCCGTTTCGCCCTGAGCAACGCGGAGGAC
CTGGACGCGGATCTGATCTCCGACAAGAGCGTGTGTCAGCGTGCACCACCCGGCGCTGATAGTGTTCAGTGC
TGCAACCTGTGTATCGCAACTCGCGCGCGCAGGGCGGAGGCCCAACTGCGACTTCAAGATATCGGCGCCCGAC
CTGCTAAACGCGTTGGTGATGGTGCAGCCTGTGGAGTGAAAACCTTACCGAGCTGCCGCGGATGGTTGTGCCT
GAGTTTAAAGTGGAGCACTAAACACCAGTATCGCAACGTGTCCCTGCCAGTGGCGCATAGCGATGCGCGGCAGAAC
CCCTTTGATTTTTAA

Ad5 E4A:

Two proteins are present in this ORF. The first is a splice variant contained within the ORF. The second is a non-spliced transcript present in the ORF. Accession 1: QHX41659.1

Accession 2: QHX41660.1

(SEQ ID NO:44)

ATGACTACGTCCGGCGTTCCATTTGGCATGACACTACGACCAACACGATCTCGGTTGTCTCGGCGCACTCCGTAC
AGTAGGGATCGCCTACCTCCTTTTGGACAGAGACCCGCGCTACCATACTGGAGGATCATCCGCTGCTGCCCCGAA
TGTAACACTTTGACAATGCACAACGCGTGGACTTCCCTTCGCCGCCGTTGAGCAACCGCAAGTTGGACAGCAG
CCTGTGGCTCAGCAGCTGGACAGCGACATGAACTTAAGCGAGCTGCCCGGGGAGTTTATTAATATCACTGATGAG
CGTTTGGCTCGACAGGAAACCGTGTGGAATATAACACCTAAGAATATGTCTGTTACCCATGATATGATGCTTTTT
AAGGCCAGCCGGGGAGAAAGGACTGTGTACTCTGTGTGTTGGGAGGGAGGTGGCAGGTTGAATACTAGGGTTCTG
TGA

(SEQ ID NO:45)

ATGACTACGTCCGGCGTTCCATTTGGCATGACACTACGACCAACACGATCTCGGTTGTCTCGGCGCACTCCGTAC
AGTAGGGATCGCCTACCTCCTTTTGAGACAGAGACCCGCGCTACCATACTGGAGGATCATCCGCTGCTGCCCGAA
TGTAACACTTTGACAATGCACAACGTGAGTTACGTGCGAGGTCTTCCCTGCAGTGTGGGATTTACGCTGATTACG
GAATGGGTTGTTCCCTGGGATATGGTTCTGACGCGGGAGGAGCTTGTAATCCTGAGGAAGTGTATGCACGTGTGC
CTGTGTTGTGCCAACATTGATATCATGACGAGCATGATGATCCATGGTTACGAGTCCCTGGGCTCTCCACTGTCTAT
TGTTCCAGTCCCGGTTCCCTGCAGTGCATAGCCGGCGGGCAGGTTTTGGCCAGCTGGTTTAGGATGGTGGTGGAT
GGCGCCATGTTTAATCAGAGGTTTATATGGTACCGGGAGGTGGTGAATTACAACATGCCAAAAGAGGTAATGTTT
ATGTCCAGCGTGTGTTATGAGGGGTCGCCACTTAATCTACCTGCGCTTGTGGTATGATGGCCACGTGGGTTCTGTG
GTCCCGCCATGAGCTTTGGATACAGCGCCTTGCACTGTGGGATTTGAACAATATTGTGGTGTCTGTCTGCAGT
TACTGTGCTGATTTAAGTGAGATCAGGGTGCGCTGTGTGCCCGGAGGACAAGGCGTCTCATGTCTGCGGGCGGTG
CGAATCATCGCTGAGGAGACCACTGCCATGTTGTATTCTTGCAGGACGGAGCGGCGGCGGCGAGCAGTTTATTTCG
GCGCTGCTGCAGCACCAACCGCCCTATCCTGATGCACGATTATGACTCTACCCCCATGTAG

Ad5 VA:

Accession: AF369965.1

(SEQ ID NO:46)

TCGATGTAGGATGTTGCCCTCCTGACGCGGTAGGAGAAGGGGAGGGTGCCCTGCATGTCTGCCGCTGCTCTTG
TCTTGCCGCTGCTGAGGAGGGGGGCGCATCTGCCGAGCACCGGATGCATCTGGGAAAAGCAAAAAGGGGCTCG
TCCCTGTTTCCGGAGGAATTTGCAAGCGGGTCTTGATGACGGGGAGGCAAACCCCGTTCGCCGAGTCCGGC
CGGCCCCGAGACTCGAACCGGGGGTCTTGCAGTCAACCCCTTGAAAAATAACCCCTCCGGCTACAGGGAGCGAGCCA
CTTAATGCTTTTCGCTTTCCAGCCTAACCGCTTACGCCGCGCGCGGCCAGTGGCCAAAAAGCTAGCGCAGCAGCC
GCCGCGCCTGGAAGGAAGCCAAAAGGAGCGCTCCCCGTTGTCTGACGTCGCACACCTGGGTTTCGACACGCGGGC
GGTAACCGCATGGATCACGGCGGACGGCCGGATCCGGGGTTTGAACCCCGGTCGTCCGCCATGATACCTTTCGA
ATTTATCCACCAGACCACGGAAGAGTGCCCGCTTACAGGCTCTCCTTTTGCACGGTCTAGAGCGTCAACGACTGC
GCACGCCTCACCGGCCAGAGCGTCCCGACCATGGAGCACTTTTGCCTGCTGCGCAACATCTGGAACCGCGTCCGC
GACTTTCCGCGCGCCTCCACCACCGCCCGCGGCATCACCTGGATGTCCAGGTACATCTACGGATTACG

EXAMPLE 13 – Promoter, Operator, IRES and Intron Sequences

CMV Promoter (SEQ ID NO:47)

TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTAC
GGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCGCCCATGACGTCAATAATGACGTATGTTCCCATAGT
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCA
AGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA
CATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGTATGCGGTT
TTGGCAGTACATCAATGGGCGTGGATAGCGGTTTACTCACGGGGATTTCGAAGTCTCCACCCCATGACGTCAA
TGGGAGTTTGTGTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAAT
GGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTCGACGTTTAGTGAACCG

2xTet Operator Sequence (SEQ ID NO: 48)

TCCCTATCAGTGATAGAGATCTCCCTATCAGTGATAGAGA

hCMV Intron Sequence (SEQ ID NO: 49)

GTAAGTACCGCCTATAGAGTCTATAGGCCCCACCCCTTGGCTTCTTATGCATGCTATACTGTTTTTGGCTTGGGG
TCTATACACCCCGCTTCCCTCATGTTATAGGTGATGGTATAGCTTAGCCTATAGGTGTGGGTTATTGACCATTAT

TGACCACTCCCCTATTGGTGACGATACTTTCCATTACTAATCCATAACATGGCTCTTTGCCACAACCTCTCTTTAT
TGGCTATATGCCAATACACTGTCTTCAGAGACTGACACGGACTCTGTATTTTTACAGGATGGGGTCTCATTAT
TATTTACAAATTCACATATACAACACCACCGTCCCCAGTGCCCGCAGTTTTTTATTAAACATAACGTGGGATCTCC
ACGCGAATCTCGGGTACGTGTTCCGGACATGGTCTCTTCTCCGGTAGCGGCGGAGCTTCTACATCCGAGCCCTGC
TCCCATGCCTCCAGCGACTCATGGTCGCTCGGCAGCTCCTTGCTCCTAACAGTGGAGGCCAGACTTAGGCACAGC
ACGATGCCCCACCACCACAGTGTGCCGCACAAGGCCGTGGCGGTAGGGTATGTGTCTGAAAATGAGCTCGGGGAG
CGGGCTTGACCGCTGACGCATTTGGAAGACTTAAGGCAGCGGCAGAAGAAGATGCAGGCAGCTGAGTTGTTGTG
TTCTGATAAGAGTCAGAGGTAACCTCCCGTTGCGGTGCTGTTAACGGTGGAGGGCAGTGTAGTCTGAGCAGTACTC
GTTGCTGCCGCGCGGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCCCTTTCCATGGGTCTTTTCTGC
AG

ECMV IRES Sequence (SEQ ID NO: 50)

CCCCCTCTCCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGAATAAGGCCGGTGTGCGTTTGTCTAT
ATGTTATTTTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGC
ATTCTAGGGGTCTTTCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCTCTG
GAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTCAGGCAGCGGAACCCCCACCTGGCGACAGGTGC
CTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGCGGCACAAACCCAGTGCCACGTTGTGAGTTGG
ATAGTTGTGGAAGAGTCAAATGGCTCTCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCAGAAGGTACCC
CATTGTATGGGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAACGTCTAG
GCCCCCGAACCACGGGGACGTGGTTTTCTTTGAAAAACACGATTGCTCGAATCACC

FMDV IRES (SEQ ID NO: 51)

AGCAGGTTTCCCCAACTGACACAAAACGTGCAACTTGAACTCCGCCTGGTCTTTCCAGGTCTAGAGGGGTAACA
CTTTGTACTGTGTTTGGCTCCACGCTCGATCCACTGGCGAGTGTTAGTAACAGCACTGTTGCTTCGTAGCGGAGC
ATGACGGCCGTGGGAACCTCTCTTGGTAACAAGGACCCACGGGGCCAAAAGCCACGCCACACGGGCCCCGTCAT
GTGTGCAACCCACGACGCGCACTTTACTGCGAAACCCACTTTAAAGTGACATTGAACTGGTACCCACACACTG
GTGACAGGCTAAGGATGCCCTTCAGGTACCCCCAGGTAACACGCGACACTCGGGATCTGAGAAGGGGACTGGGGC
TTCTATAAAAGCGCTCGGTTTTAAAAAGCTTCTATGCCTGAATAGGTGACCGGAGGTGGGCACCTTTCTTTACAA
TTAATGACCT

EXAMPLE 14 – CHO and Mouse Stable Site 1 Sequences – U.S. Patent No.

7,771,997

211> 6473

<212> DNA

<213> Cricetulus griseus

<400> 1

(SEQ ID NO:52)

tctagaaaca	aaaccaaaaa	tattaagtca	ggcttggctt	caggtgctgg	ggtggagtgc	60
tgacaaaaat	acacaaattc	ctggctttct	aaggcttttt	cggggattca	ggtattgggt	120
gatggtagaa	taaaaatctg	aaacataggt	gatgtatctg	ccatactgca	tgggtgtgta	180
tgtgtgtgta	tgtgtgtctg	tgtgtgtgcc	cagacagaaa	taccatgaag	gaaaaaaca	240
cttcaaagac	aggagagaag	agtgacctgg	gaaggactcc	ccaatgagat	gagaactgag	300
cacatgccag	aggaggtgag	gactgaacca	ttcaacacaa	gtggtgaata	gtcctgcaga	360
cacagagagg	gccagaagca	ctcagaactc	caggggttca	ggagtgggtc	tctggaggct	420
tctgcccttg	gaggttcctg	aggaggaggc	ttccatattg	aaaatgtagt	tagtggccgt	480
ttccattagt	acagtgacta	gagagagctg	agggaccact	ggactgaggc	ctagatgctc	540
agtcagatgg	ccatgaaagc	ctagacaagc	acttccgggt	ggaaaggaaa	cagcaggtgt	600
gaggggtcag	gggcaagtta	gtgggagagg	tcttccagat	gaagtagcag	gaacggagac	660
gcaactggatg	gccccacttg	tcaaccagca	aaagcttggg	tcttgttcta	agaggccagg	720
gacatgacaa	gggtgatctc	ggttttttaa	aggcttttgt	ttacctaatc	acttctatta	780

gtcagatact	ttgtaacaca	aatgagtact	tggcctgtat	tttagaaact	tctgggatcc	840
tgaaaaaaca	caatgacatt	ctggctgcaa	cacctggaga	ctcccagcca	ggccttggac	900
ccgggtccat	tcatgcaaat	actcagggac	agattcttca	ctaggtactg	atgagctgtc	960
ttggatgcaa	atgtggcctc	ttcattttac	tacaagtcac	catgagtcag	gaggtgctgt	1020
ttgcacagtg	tgactaagtg	atggagtgtt	gactgcagcc	attcccggcc	ccagcttgtg	1080
agagagatcc	ttttaaattg	aaagtaagct	caaagttacc	acgaagccac	acatgtataa	1140
actgtgtgaa	taatctgtgc	acatacacaa	accatgtgaa	taatctgtgt	acatgtataa	1200
actgtgtgaa	taatctgtgt	gcagcctttc	cttacctact	accttccagt	gatcaggttt	1260
ggactgcctg	tgtgctactg	gaccctgaat	gtccccaccg	ctgtcccctg	tcttttacga	1320
ttctgacatt	tttaataaat	tcagcggctt	cccctctgct	ctgtgcctag	ctataccttg	1380
gtactctgca	ttttggtttc	tgtgacattt	ctctgtgact	ctgctacatt	ctcagatgac	1440
atgtgacaca	gaaggtgttc	cctctggaga	catgtgatgt	ccctgtcatt	agtggaatca	1500
gatgccccca	aactgttgtc	cagtgtttgg	gaaagtgaca	cgtgaaggag	gatcaggaaa	1560
agaggggtgg	aaatcaagat	gtgtctgagt	atctcatgtc	cctgagtggg	ccaggtctgt	1620
gacttcaactc	ccccaaagtga	gggaggccat	ggtgagtaca	cacacctcac	acatactata	1680
tccaacacac	acacacacac	acacacacac	acgcacgcac	gcacgcacgc	acgcacacat	1740
gcacacacac	gaactacatt	tcacaaacca	catacgcata	ttacacccca	aacgtatcac	1800
ctatacatac	cacacataca	caccctcca	cacatcacac	acataccaca	cccacacaca	1860
gcacacacat	acataggcac	acattcacac	accacacata	tacattttgtg	tatgcataca	1920
tgcatacaca	cacaggcaca	cagacaccac	acacatgcat	tgtgtacgca	cacatgcata	1980
cacacacata	ggcacacatt	gagcacacac	atacatttgt	gtacgcacac	tacatagaca	2040
tatatgcatt	tgtatatgca	cacatgcatg	cacacataca	taggcacaca	tagagcacac	2100
acatacattt	gtgtatgcac	acatgcacac	accaatcaca	tgggaagact	caggttcttc	2160
actaaggttc	acatgaactt	agcagttcct	ggttatctcg	tgaaacttgg	aagattgctg	2220
tggagaagag	gaagcgttgg	cttgagccct	ggcagcaatt	aaccccgccc	agaagaagta	2280
ggtttaaaaa	tgagagggtc	tcaatgtgga	acccgcaggg	cgccagttca	gagaagagac	2340
ctacccaagc	caactgagag	caaaggcaga	gggatgaacc	tgggatgtag	tttgaacctc	2400
tgtaccagct	gggcttcatg	ctatttttgt	atatctttat	taaatattct	tttagtttta	2460
tgtgcgtgaa	taccttgctt	gcataaaagt	atgggcactg	tatgtgttct	tggtgccggt	2520
ggaggccagg	agagggcatg	gacccctcgg	agctggcggt	tgagacagtt	gtgaccaca	2580
gtgtgggggtc	tgggaactgg	gtcttagtgt	tcgcgaagtg	cagctggggc	tcttaacctc	2640
tgagccatcc	ctccagcttc	aagaaactta	ttttcttagg	acatggggga	agggatccag	2700
ggcttttaggc	ttgtttgttc	agcaaatact	cttttctgtg	attttgaatt	ttattttatt	2760
ttactttttt	gggatagaat	cacattctgc	agctcaggct	gggcctgaac	tcatacaaat	2820
cctcctgtct	cagtctacca	ggtgataaga	ttactgatgt	gagcctggct	ttgacaagca	2880
cttttagagtc	cccagccctt	ctggacactt	gttccaagta	taatatatat	atatatatat	2940
atatatatat	atatatatat	atatatttgt	tgtgtgtgtt	tgtgtgtgta	tgagacactt	3000
gctctaaggg	tatcatatat	atccttgatt	tgcttttaat	ttatttttta	attaaaaatg	3060
attagctaca	tgtcacctgt	atgcgtctgt	atcatctata	tatccttcc	tccttctctc	3120
tctttctctc	ttcttcttct	cacccccaa	catctatatt	caaatccttg	tgccgaggag	3180
atgccaagag	tctcgttggg	ggagatggtg	agggggcgat	acaggggaag	agcaggagga	3240
aagggggaca	gactggtgtg	ggtcttttga	gagctcagga	gaatagcagc	gatcttccct	3300
gtccctggtg	tcacctctta	cagccaacac	catttttgtg	cctggcagaa	gagttgtcaa	3360
gctggtcgca	ggtctgccac	acaaccccaa	tctggcccca	agaaaaggca	cctgtgtgtg	3420
actctggggt	taaaggcgct	gcctggtcgt	ctccagctgg	acttgaaact	cccgtttaat	3480
aaagagttct	gcaaaataat	acccgcagag	tcacagtgcc	aggttcccg	gctttcctga	3540
agcgccaggc	acgggttccc	taggaaatgg	ggccttgctt	gccaaagctcc	cacggcttgc	3600
cctgcaaacg	gcctgaatga	tctggcactc	tgcgttgcca	ctgggatgaa	atggaaaaaa	3660
gaaaaagaag	aagtgtctct	ggaagcgggc	gcgctcacac	aaacccgcaa	cgatttgtgt	3720
aacactctcc	attgagaatc	tggagtgcgg	ttgccctcta	ctggggagct	gaagacagct	3780
agtgggggcg	gggggaggac	cgtgctagca	tccttccacg	gtgctcgctg	gctgtggtgc	3840
atgcccgggaa	ccgaaacgcg	gaactaaagt	caagtcttgc	tttgggtggaa	ctgacaatca	3900
acgaaatcac	ttcgattgtt	ttcctctttt	tactggaatt	cttggatttg	atagatgggg	3960
gaggatcaga	gggggagggg	aggggcgggg	agacggaggg	aggaggggag	gaggggagga	4020
ggggaggagg	ggaggagggg	aagggatgga	ggaaaatact	aacttttcta	attcaacatg	4080
acaaagattc	ggagaaagtg	caccgctagt	gaccgggagg	aggaatgcc	tattgggcat	4140
tatatccct	gtcgtcta	ggaatcaa	tcttggttcc	agcaccaagg	attctgagcc	4200
tatcctat	aagacagtaa	ctacagccca	cacggaagag	gctatacaac	tgaagaaata	4260

aaatttttcac	tttattttcat	ttctgtgact	gcatgtttcac	atgttagagag	ccacctgtgt	4320
ctaggggctg	atgtgctggg	cagtagagtt	ctgagcccgt	taactggaac	aaccagaac	4380
tcccaccaca	gttagagctt	gctgagagag	ggaggccctt	ggtgagattt	ctttgtgtat	4440
ttatttagag	acagggctct	atactgtagt	ccaagctagc	ctccagctca	cagaaattct	4500
cctgttccgg	tttccaaagt	actggagtta	tgagtgtgtg	ttaattgaac	gctaagaatt	4560
tgctgattga	agaaaacctc	aagtgggttt	ggctaataccc	cacgacccca	gaggctgagg	4620
caggaggaat	gagagaattc	aaggtttgcc	agagccacag	ggtgagctca	atgtggagac	4680
tgtgaggggtg	agctcaatgt	ggagactgtg	aggggtgagct	caatgtggag	actgtgaggg	4740
tgagctcaat	gtggagactg	tgaggggtgag	ctcaatgtgg	agactgtgag	ggtgagctca	4800
atgtggagac	ctgtatcaag	ataataatag	tagtagtaac	aatgcaggcg	aggggtgtgg	4860
tgagtggtag	agcagttagt	tgatttgaca	tgcttgaggt	ctcccgggtcc	atctgtggcc	4920
ctgcaacagg	aagggagggg	ggaagggggg	gaacgagaga	gaggaaagag	agacagaagc	4980
taagataggg	aatgagagag	gaagggaagaa	acgggaagaa	attcagactc	cttcctgagt	5040
tccgccaacg	cctagtgaca	tcctgtgcac	accctaaggt	ggcctttgtg	tggcactggc	5100
ttgggtggtc	gggaaaggca	ttttcagctt	gttgcagaac	tgccacagta	gcatgctggg	5160
tccgtgaaag	ttttctgccg	ttaacaagaa	gtctctacta	cttgtgacct	caccagtga	5220
aattttcttta	attgtctcct	ggtgttctgg	gttttgcatt	tttgtttcta	aggatacatt	5280
cctgggtgat	gtcatgaagt	ccccaaagac	acagtggggc	tgtgttggat	tgggaaagat	5340
gattttatctg	gggtgtcaaa	aggaaaagaa	gggaaacagg	cacttgggaa	aatgtcctcc	5400
cgcccacccg	aattttggct	tggcaaccgt	ggtggaggag	caagaaacac	gtggacgttt	5460
gaggaggcat	ggggctcctag	gaggacagga	agcagaagga	gagagctggg	ctgacagcct	5520
gcaggcattg	cacagtttca	gaaggagatt	acagcatgac	tgagttttta	gggatccaac	5580
agggacctgg	gtagagattc	tgtgggctct	gaggcaactt	gacctcagcc	agatggtatt	5640
tgaataacct	gctcttagag	ggaaaacaga	catagcaaac	agagccacgt	ttagtgatga	5700
aactctcact	ttgcctgagt	catgtgcggc	catgccaggg	ggtcaggctg	acactcaact	5760
caaaaacaag	tgagaaattg	aagacaatcc	gtgggtggcag	ctactggaag	ggccaccaca	5820
tccccagaaa	gagtggagct	gctaaaaagc	cattttgtgat	aggcacagtt	atcttgaatg	5880
catggagcag	agattacgga	aaaatcgaga	atgttaatga	ggcaacattc	gagttgagtc	5940
attcagtggtg	ggaaacccag	acgcttccat	cccctaaaag	gaacatcttg	ctctcagtc	6000
aaatggaaat	aaaaattggg	gcttgaattt	ggcaaatgat	tcagaactct	gtgtagggtat	6060
tttcacacgc	acagtggata	attttcatgt	tggagtttat	ttgtgctaaa	aggcagaaaa	6120
gggtaaaaag	cacatcttaa	gagttatgag	gttctacgaa	taaaaataat	gttacttaca	6180
gctattcctt	aattagtagc	cccttccacc	tgtggtaatt	tcctgagata	gtcagtgagg	6240
aaaagatctc	tccttctctt	ctttctcccc	ctcccctcct	ctcccctcct	ccctcccctc	6300
ctcccctcct	tccttccctc	cccctttcct	tcttttcttg	ctccttctcc	tctgctcctc	6360
tctccccttc	ttcttcatth	attctaagta	gcttttaaca	gcacaccaat	tacctgtgta	6420
taacgggaaa	acacaggctc	aagcagctta	gagaagattg	atctgtgttc	act	6473

<211> 7045

<212> DNA

<213> *Cricetulus griseus*

<400> 2

(SEQ ID NO: 53)

actagcgtgc	aattcagagg	tgggtgaaga	taaaaggcaa	acattttgagg	ccatttccctt	60
atlttgccacg	gcacttagga	agtggaaacat	gcctaattcta	ctggttttgta	ccacctttcc	120
ctataatgga	ctgttttgga	agctcctggg	caaccgatcc	tggcatctca	ttggctcagag	180
gcctgttaaa	tggtagctctt	atlttgcaag	aaggctgtaa	cttgtagctt	taaaagcctc	240
tcctcaagaa	agaagggaga	aaggatatgg	ctagacatat	ctaatagact	taaccactgt	300
gaaaagcctt	agtatgaatc	agatagaacc	tatttttaac	tcagttttga	aaaaaataat	360
ctttatatatt	atlttgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	420
gaaccacatg	tagcaggtgc	tggaggaggc	cagaagaggg	caccagatct	cctggaactg	480
acaccacaca	tggttatgag	ctgcctgatg	tgggtgctgg	gaactgaact	ctcgtgttct	540
gcaagagcag	caactgttct	cttaactgat	gagccatctc	tccagccccc	cccataattt	600
taattgttca	tttttagtaaa	ttttattcat	aatcaattat	cacagtataa	aacaatgatt	660

ttatatatat	catatacata	tcaaggatga	cagtgagggg	gatatgtgtg	tgtgtgtgtg	720
tgtgtgtgtg	tgtgtgtgtg	tgtgttattt	gtgtgtgtgc	tttttaagaa	ggtgccatag	780
tactgcatt	tctctgaagg	atttcaaagg	aatgagacat	gtctgtctgc	caggaacct	840
atcttctct	ttgggaatct	gacccaaatg	aggtattctg	aggaactgaa	tgaagagctc	900
aagtagcagt	gtcttaaacc	caaatgtgct	gtctagagaa	agtcaacgtc	atcagtgage	960
tgaggagaga	tttactgagc	ggaagacaag	cgctctttga	tttaagtggc	tcgaacagtc	1020
acggctgtgg	agtggagcct	gtgctcaggt	ctgaggcagt	ctttgctagc	cagctgtgat	1080
gagcagtga	gaaaggggtg	agatggaggc	agggtgggag	cagggtatg	gttcagacta	1140
ggtatcgtga	gcacaccagc	tggttgactt	gtggtctgtg	ggtcaggcgt	tgtaaacgcc	1200
ctcagggtca	ggcagtcaca	ttgcttgaag	ctgaatgggt	gaggcaacac	agagagtga	1260
aagaaggcaa	agtaccacct	cttccccgac	ccaggctcact	tctgggttat	agctgagact	1320
ccggacagca	tgcaaccagc	tggttagagc	ttcagggaaa	acttgatgtc	tgcatgttgc	1380
tatgaaatgt	gattcggtac	atctggagaa	aattttataat	gctgggtcag	tcaagcactg	1440
aacaaaggta	ccttggcttt	gggagctaca	tgacattgac	ttgtaggcag	actttttttt	1500
ttctgcccgc	caattcccag	ataaccaata	tggaggctca	atattaatta	taaatgctcg	1560
gctgatagct	caggcttgtt	actagctaac	tcttccaact	taaatgaacc	cattttctatt	1620
atctacattc	tgccacgtga	ctttaccttg	tacttctctg	ttctctctct	tgtctgactc	1680
tgcccttctg	cttcccagag	tccttagtct	ggttctctct	cctaacctta	tcctgcccag	1740
ctgctgacca	agcattttata	attaatatata	agtctcccag	tgagactctc	atccagggtg	1800
gacttgggtg	ctccccctc	ctcattgcca	tccgtgtctt	cctcttccct	cgcttcccc	1860
tcctcttctc	gctcttctc	ctccaccct	cctttcatag	tattgatggc	aagggtgttc	1920
tagaatggag	gagtgcccat	aggcatgcaa	agaaaccagt	taggatgctc	tgtgaggggt	1980
tgtaatcata	agcgatggac	acaattcaag	ccacagagtg	aagacggaag	gatgcactgt	2040
gctctagagc	aacttctggg	gcagaatcac	agggtgagtt	tctgacttga	gggcgaagag	2100
gccacgagga	agggtgag	tttgtctgag	ctagaagcta	cggcccacct	cttggttagca	2160
gacctgccca	caagcatgct	ttgttaataca	tgtgggatct	gattttctct	taaatctatg	2220
ttcaactctt	aagaaaatgt	gaattctcac	attaaaattt	agatatacgt	cttttggtgg	2280
ggggggtgta	aaaaatcctc	aagaatatgg	atttctgggg	gccggagaga	tggtctcagag	2340
gttaagagaa	ctggttgctc	ttctagacat	tctgagttca	attcccagca	accacatggt	2400
ggctcacaac	catctgtaac	gcgacctgt	gccatcttct	gacatgcacg	tacatagca	2460
ggcagaaagc	tgtatacata	gtaaattgat	aaatcttttt	ttaaaaagag	tatggattct	2520
gccgggtgtt	ggtggcgcac	gcctttaatc	ccagcactct	ggaggcagag	gcagggtggat	2580
ctctgtgagt	tcgagaccag	cctgggtctat	aagagctagt	tccaggacag	cctccaaagc	2640
cacagagaaa	ccctgtctcg	aaaaacccaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaga	2700
gtatggattc	taagaaagcc	gtaacagctg	gagctgtgta	cggagttcag	cgtgggtacta	2760
gaagaacaga	cattcatgat	gaaacacccc	aggattttta	cttagtatct	agtttccatt	2820
gttggttttga	gaccggctct	tatgctctcc	aggctggcct	caaactgctg	atcttccgc	2880
ctctacctct	caagtctctg	gactacttgg	ctcataaaac	agtttttgtc	gggtctccctg	2940
aagttatggt	tgtacaaacc	gtgggggtca	atatactcac	ttgggcagag	agagaagggtc	3000
tgaatcccag	acaatgactg	catctcagga	cagttgggaa	gaggacaatg	gcagaaggac	3060
ttagaaaaga	tagactggag	ggtggaaaag	cagcaggaac	agagaaacaa	aacaggaagc	3120
ttgctatcca	gggccactct	ggagtctctg	ggcaagatgg	aagcgggcta	ggggaataca	3180
tttgtgtctac	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgat	caatgcctat	3240
caatggtgaa	ggggaaatat	gtataaccaca	ttgattctgg	gagcaattct	cagtatctgg	3300
cctagagaaa	ggaatggccc	ctgcagaata	gacagagtga	atggtgccct	ttatcatttg	3360
ctaaagtga	ggagaaataa	acatccttcc	atagagtttc	aggtaaatga	acccacagt	3420
tcactgtgtc	cgtgggtggag	gcctggccaa	cagttaaaaa	gattagacac	ggacaaagtc	3480
tgaaggaaac	acctcgaata	ggaagaggag	agccacctca	ttctgttaact	ttcctcaagg	3540
ggaagatggt	ccaagagtgg	gaataaatgg	tcaaaggggg	gatttttaat	taggaaaacg	3600
atttctctgta	tcacttgtga	aactggaggt	tgatttgggg	cataggacaa	tagattttgat	3660
gctttgcaaa	aagctgtttc	aaagcagaga	aatggaatag	agacaattat	gtagcgagga	3720
gggagggtgg	ggcgaaagatg	gagacagaga	agtggaaagct	gacttttaggg	aagaggaaca	3780
tagaccacag	gggcggggcg	gggggcaggg	gcgggggggcg	gggctcaaag	gaggcagtg	3840
gaacgttgct	agtgttcgca	gcgtaaagct	gaatgtgcaa	gcgtctttgt	ggtgtgtgac	3900
caggagtagc	gtggctggct	tgtgtgctgc	ttgtaatccc	agtctttgag	gtttccacac	3960
tgttccacag	tgggtgtgat	tttccctcgg	agagcatgag	ggctctgctt	tccccacatc	4020
ctccccagcg	ttcggttggt	tttggtttcca	agatgttagt	gggtgagaca	aagcctctct	4080
gttgatttgc	ctttaacagg	tgacaaaaaa	agctcaacca	ggagacattt	ttgccttctt	4140

ggaaggtaat	gctcccatgt	agagcaatgg	gacccatctc	taagggtgagg	ctactcttgc	4200
agtttgcacc	cagctcttct	gatgcaggaa	ggaagttggg	gggcaagcaa	gactgtttgc	4260
ttcttgcgat	ggacacattc	tgcacacaaa	ggctcaggag	gggagaaggc	tgtttgatgt	4320
ttagcactca	ggaaggcccc	tgatgcatct	gtgattagct	gtctccatct	gtggagcaga	4380
cacggactaa	ctaaaaacca	gtgtttttta	attgtcaagc	ctttaagggtg	aggaaattga	4440
cttattgtgc	tggggccatac	gtagagcaag	tgctctgcat	tggggccaacc	cccggtctctg	4500
gtttctaggc	accagaatgg	cctagaacta	actcacaatc	ctcccatgcc	aggtctcagg	4560
tgctagaatg	aaccactata	ccagcctgcc	tgctctgcta	cctgccttcc	taaattttta	4620
atcatgggga	gtaggggaga	atacacttat	cttagttagg	gtttctattg	ctgtgaagag	4680
acaccatgag	catggcaact	cttataaagg	aaaacattta	gttgggtggc	agtttcagag	4740
gttttagtac	attgtcatca	tggctgggaa	catgatggca	tgcagacaga	catggtgctg	4800
gagaaaggga	tgagagtcc	acatcttgca	ggcaacagga	cctcagctga	gacactggct	4860
ggtaccctga	gcataggaaa	cctcacagcc	cacctcaca	gtgacatatt	tccttcaaca	4920
aagccatacc	tcctaatagt	gccactccct	atgagatgac	agggccaatt	acattcaaac	4980
tgctataaca	ctttaaagta	ttttattttt	attattgtaa	attatgtatg	tagctgggtg	5040
gtggcagccg	aggtgcacgc	ctttaatccc	agcacttggg	aggcagaggc	agatggatct	5100
ctgtgagttc	aagaccagcc	tgggtctataa	gagctagttg	caagggaagga	tatacaaga	5160
acagttctag	gatagccttc	aaagccacag	agaagtgtctg	tcttgaaaac	caaaaattgt	5220
gctgggacct	gtctctgctt	tgggttgcctc	ccactcccc	agagctggac	tcttgggtcaa	5280
cactgaatca	gctgcaaaat	aaactcctgg	attcctctct	tgtaacagga	gcccgaagtc	5340
aggcgccac	ttgtcttctc	gcaggattgc	catagacttt	ttctgtgtgc	ccaccattcc	5400
agactgaagt	agagatggca	gtggcagaga	ctgggaaggc	tgcaacgaaa	acaggaagtt	5460
attgcaccct	gggaatagtc	tggaaatgaa	gcttcaaaaac	ttgcttcatg	ttcagttgta	5520
cacagactca	ctcccagggt	gactcacacg	tgtaaataatt	cctgactatg	tctgcaactgc	5580
ttttatctga	tgttcccttc	ccaaaatgcc	aagtgtacaa	ggtgagggaa	tcacccttgg	5640
attcagagcc	cagggctgct	ctccttaacc	tggacttgtc	tttctccggc	agcctctgac	5700
acccctcccc	ccattttctc	tatcagaagg	tctgagcaga	gttggggcac	gctcatgtcc	5760
tgatacactc	cttgtcttcc	tgaagatcta	acttctgacc	cagaaagatg	gctaagggtg	5820
tgaagtgttt	gacatgaaga	cttgggtctta	agaactggag	caggggaaaa	aagtcggatg	5880
tggcagcatg	taccgaaat	cccagaactg	gggaggtaga	gacggatgag	tgcccggggc	5940
tagctggctg	ctcagccagc	ctagctgaat	tgccaaattc	caactcctat	tgaaaaacct	6000
ttaccaaaca	aacaaacaaa	caaataataa	caacaacaac	aacaacaaac	tacccatac	6060
aagggtgggcg	gctcttggct	cttgaggaat	gactcaccca	aacccaaagc	ttgccacagc	6120
tgttctctgg	cctaaatggg	gtgggggtgg	ggcagagaca	gagacagaga	gagacatgac	6180
ttcctgggct	gggctgtgtg	ctctaggcca	ccaggaaactt	tcctgtcttg	ctctctgtct	6240
ggcagagcca	gagcaccagc	acccagcagg	tgcacacacc	tcctccgtg	cttcttgagc	6300
aaacacaggt	gccttgggtct	gtctattgaa	ccggagtaag	ttcttgcaga	tgtatgcatg	6360
gaaacaacat	tgtcctgggt	ttatttctac	tgttgtgata	aaaaccgggg	aactccagga	6420
agcagctgag	gcagaggcaa	atgcaaggaa	tgctgcctcc	tagcttgctc	cccatggctt	6480
gocgggcctg	ctttctgcaa	gcccttctct	ccccattggc	atgcctgaca	tgaacagcgt	6540
ttgaaatgct	ctcaaatgtc	actttcaaag	aaggcttctc	tgatcttgct	aactaaatca	6600
gaccatgttt	caccggtgat	tatctttctg	ctgtctgtct	gtctgtctgt	ctgtctatct	6660
gtctatcatc	tatcaatcat	ctatctatct	atcttctatt	tatctacct	tcattcaatc	6720
atctatcttc	taactagtta	tcattttattt	atttgtttac	ttactttttt	tatttgagac	6780
agtatttctc	tgagtgcag	ccttggctgt	cctggaaccc	attctgtaac	caggctgtcc	6840
tcaaactcac	agagatccaa	ctgcctctgc	ctctctggtg	ctgggggttaa	agacgtgcac	6900
caccaacgcc	ccgctctatc	atctattttat	gtacttatta	ttcagtcatt	atctatcctc	6960
taactatcca	tcactctgtct	atccatcatc	tatctatcta	tctatctatc	tatctatcta	7020
tctatcatcc	atctataatc	aattg				7045

<211> 6473
 <212> DNA
 <213> *Cricetulus griseus*

<400> 3
 (SEQ ID NO: 54)

```

agtgaacaca gatcaatctt ctctaagctg cttgagcctg tgtttttccc ttatacacag 60
gtaattggtg tgctgttaaa agctacttag aataaatgaa gaagaaaggg agaaggaggc 120
agaggagaag gagcaaagaa agaaggaaag ggggagggag ggagaggagg gagggaggga 180
gggagggagg gagaggaggg gagggggaga aagaagagaa ggagagatct tttccccact 240
gactatctca ggaaattacc acaggtggaa gggggtacta attaaggaat agctgttaagt 300
aacattatth ttattcgtag aacctcataa ctcttaagat gtgcttttta cctttttctg 360
ccttttagca caaataaact ccaacatgaa aattatccac tgtgcgtgtg aaaataccta 420
cacagagttc tgaatcattt gccaaattca agccccaatt tttatttcca ttttgactga 480
gagcaagatg ttctcttttag gggatggaag cgtctgggtt tcccacactg aatgactcaa 540
ctcgaatgtt gcctcattaa cattctcgat ttttccgtaa tctctgctcc atgcattcaa 600
gataactgtg cctatcacaa atggcttttt agcagctcca ctctttctgg ggatgtgggtg 660
gcccttccag tagctgccac cacggattgt cttcaatttc tcacttgttt ttgagttgag 720
tgtcagcctg acccctgggc atggccgcac atgactcagg caaagtgaga gtttcatcac 780
taaacgtggc tctgtttgct atgtctgttt tccctctaag agcaggttat tcaaatacca 840
tctggctgag gtcaagttgc ctcaagagccc acagaatctc taccaggtc cctgttggat 900
ccctaaaaac tcagtcatgc tgtaatctcc ttctgaaact gtgcaatgcc tgcaggctgt 960
cagcccagct ctctccttct gcttctgtc ctctaggac cccatgcctc ctcaaacgtc 1020
cacgtgtttc ttgtccttcc accacggttg ccaagccaaa attcgggtgg gcgggaggac 1080
atthttccaa gtgcctgttt ccttctttt ccttttgaca cccagataa atcatctttc 1140
ccaatccaac acagccccac tgtgtctttg gggacttcat gacatcacc aggaatgtat 1200
ccttagaaac aaaaatgcaa aaccagaac accaggagac aattaaagaa atthttactg 1260
gtgaggtcac aagtagtaga gacttcttgt taacgggcag aaactttcac ggaccagca 1320
tgtactgtg gcagttctgc aacaagctga aaatgccttt cccgaccacc caagccagt 1380
ccacacaaag gccaccttag ggtgtgcaca ggatgtcact aggcgttggc ggaactcagg 1440
aaggagtctg aatttcttcc cgtttcttcc ttctctctc attcctatc ttagcttctg 1500
tctctctttc ctctctctcg tttccccct tcttccctcc ctctctgtt cagggccaca 1560
gatggaccgg gagacctcaa gcatgtcaaa tcaactaact gctctaccac tcaaccacac 1620
cctcgctgc attgttacta ctactattat tatcttgata caggtctcca cattgagctc 1680
accctcacag tctccacatt gagctcacc tcacagtctc cacattgagc tcaccctcac 1740
agtctccaca ttgagctcac cctcacagtc tccacattga gctcaccctc acagtctcca 1800
cattgagctc accctgtggc tctggcaaac cttgaattct ctcatctctc ctgctcagc 1860
ctctgggggtc gtggggatta gccaaaccca cttgaggttt tcttcaatca gcaaattctt 1920
agcgttcaat taacacacac tcataactcc agtactttgg aaaccggaac aggagaattt 1980
ctgtgagctg gaggctagct tggactacag tatgagaccc tgtctctaaa taaatacaca 2040
aagaaatctc accaagggcc tccctctctc agcaagctct aactgtgggt ggagttctgg 2100
gttggtccag ttaacgggct cagaactcta ctgccagca catcagcccc tagacacagg 2160
tggtctctta catgtgaaca tgcagtcaca gaaatgaaat aaagtgaaaa ttttatttct 2220
tcagttgtat agcctcttcc gtgtgggctg tagttaactg cttgaatagg ataggctcag 2280
aatccttggg gctggaacca agagtttgat tccattagac gacaggggat ataatgcca 2340
atagggcatt cctcctccc gtcactagcg gtgcactttc tccgaatctt tgtcatgttg 2400
aattagaaaa gttagtatth tcttccatcc ctctccctcc tccctctctc cctcctccc 2460
ctcctcccct cctcctccc tctccccgc cctccctcc cctctgata ctcccccatc 2520
tatcaaatcc aagaattcca gtaaaaagag gaaaacaatc gaagtgattt cgttgattgt 2580
cagttccacc aaagcaagac ttgacttttag ttccgcgttt cggttcccgg catgaccac 2640
agccagcgag caccgtggaa ggatgctagc acggtcctcc ccccgcccc actagctgtc 2700
ttcagctccc cagtagaggg caaccgcaact ccagattctc aatggagagt gtttacacaa 2760
tcgttgcggg tttgtgtgag cgcgcccgt tccagagaca cttcttcttt ttcttttttc 2820
catttcatcc cagtggcaac gcagagtgcc agatcattca ggcggtttgc agggcaagcc 2880

```

gtgggagctt	ggcaagcaag	gccccatttc	ctaggggaacc	cgtgcctggc	gcttcaggaa	2940
agcacgggaa	cctggcactg	tgactctgcg	ggtattatnt	tgcagaactc	tttattaaac	3000
gggagtttca	agtccagctg	gagacgacca	ggcagcgcct	ttaaccccag	agtcacacac	3060
aggtgccttt	tcttggggcc	agattggggt	tgtgtggcag	acctgcgacc	agcttgacaa	3120
ctcttctgcc	aggccacaaa	atggtgttgg	ctgtaagagg	tgacaccagg	gacaggggag	3180
atcgctgcta	ttctcctgag	ctctccaaa	acccacacca	gtctgtcccc	ctttcctcct	3240
gctcttcccc	tgtatcgccc	cctcaccatc	tcccccaacg	agactccttg	catctcctcg	3300
gcacaaggat	ttgaaaatag	atgcttgggg	gtgagaagaa	gaagagagaa	agagagagaa	3360
ggaaggaagg	atatatatag	gatacagacg	catacagggt	acatgtagct	aatcattttt	3420
aattaaaaaa	taaaataaaa	gcaaatacaag	gatatatatg	atacccttag	agcaagtgtc	3480
tcatacacac	acaaacacac	acacacaata	tatatatata	tatatatata	tatatatata	3540
tatatatata	ttataacttg	aacaagtgtc	cagaagggct	ggggactcta	aagtgtttgt	3600
caaagccagg	ctcacatcag	taatcttctc	acctggtaga	ctgagacagg	aggattttga	3660
tgagttcagg	cccagcctga	gctgcagaat	gtgattctat	ccccaaaaag	taaaaataaaa	3720
taaaattcaa	aatacacgaa	aagagtatnt	gctgaacaaa	caagcctaaa	gccctggatc	3780
ccttccccca	tgtcctaaga	aaataagttt	cttgaagctg	gagggatggc	tcagaggtta	3840
agagccccag	ctgcacttgc	ggaacactaa	gacccagttc	ccagacccca	cactgtgggt	3900
cacaactgtc	tcaaacgcca	gctccggagg	atccatgccc	tctcctggcc	tccaccggca	3960
ccaagaacac	atacagtgcc	catacattta	tgcaagcaag	gtattcacgc	acataaaaact	4020
aaaagaatat	ttaataaaga	tataacaaaa	tagcatgaag	cccagctggg	acagagggtc	4080
aaactacatc	ccaggttcat	ccctctgcct	ttgctctcag	ttggccttgg	taggtctctt	4140
ctctgaactg	gcgccttgcg	ggttccacat	tgagaccctc	tcatttttaa	acctacttct	4200
tctgggcggg	gttaattgct	gccagggctc	aagccaacgc	ttcctcttct	ccacagcaat	4260
cttccaagtt	tcacgagata	accaggaact	gctaagttca	tgtgaacctt	agtgaagaac	4320
ctgagtcctc	ccatgtgatt	ggtgtgtgca	tgtgtgcata	cacaaatgta	tgtgtgtgct	4380
ctatgtgtgc	ctatgtatgt	gtgcatgcat	gtgtgcatat	acaaatgcat	atatgtctat	4440
gtagtgtgcg	tacacaaatg	tatgtgtgtg	ctcaatgtgt	gcctatgtgt	gtgtatgcat	4500
gtgtgcgtac	acaatgcatg	tgtgtgggtg	ctgtgtgcct	gtgtgtgtat	gcatgtatgc	4560
atacagaaat	gtatatgtgt	ggtgtgtgaa	tgtgtgccta	tgtatgtgtg	tgtgtgtgtg	4620
gggtgtggta	tgtgtgtgat	gtgtggaggg	gtgtgtatgt	gtgggtatga	taggtgatgc	4680
gtttgggggtg	taatatgcgt	atgtggtttg	tgaaatgtag	ttcgtgtgtg	tgcgtgtgtg	4740
cgtgcgtgcg	tgcgtgcgtg	cgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtt	ggatatagta	4800
tgtgtgaggt	gtgtgtactc	accatggcct	ccctcacttg	ggggagtga	gtcagcagcc	4860
tggaccactc	agggacatga	gatactcaga	cacatcttga	tttccacccc	tcttttctcg	4920
atcctccttc	acgtgtcact	ttcccaaaac	ctggacaaca	gtttgggggg	atctgattcc	4980
actaatgaca	gggacatcac	atgtctccag	agggaacacc	ttctgtgtca	catgtcatct	5040
gagaatgtag	cagagtcaca	gagaaatgtc	acagaaacca	aaatgcagag	taccaaggta	5100
tagctaggca	cagagcagag	gggaagccgc	tgaattttatt	aaaaatgtca	gaatcgtaaa	5160
agacagggga	cagcgggtgg	gacattcagg	gtccagtagc	acacaggcag	tccaaacctg	5220
atcactggaa	ggtagttagt	aaggaaaagg	tgacacacaga	ttattcacac	agttttatata	5280
tgtacacaga	ttattcacat	ggttttgtgt	tgtgcacaga	ttattcacac	agttttatata	5340
tgtgtggctt	cgtggtaact	ttgagcttac	tttcaattta	aaaggatctc	tctcacaagc	5400
tggggccggg	aatggctgca	gtcaaacactc	catcacttag	tcacactgtg	caaacagcac	5460
ctcctgactc	atggtgactt	gtagtaaaat	gaagaggcca	catttgcatc	caagacagct	5520
catcagtacc	tagtgaagaa	tctgtccctg	agtattttgca	tgaatggacc	cgggtccagg	5580
gcoctggctg	gagtcctccag	gtgttgacgc	cagaatgtca	ttgtgttttt	tcaggatccc	5640
agaagtttct	aaaatacacg	ccaagtactc	atttgtgtta	caaagtatct	gactaataga	5700
agtgatttag	taacacaaa	ccttttaaaa	accgagatca	cccttgtcat	gtccctggcc	5760
tcttagaaca	agatccaagc	ttttgtctgt	tgacaagtg	ggccatccag	tgcgtctccg	5820
ttcctgctac	ttcatctgga	agacctctcc	cactaacttg	cccctgaccc	ctcacacctg	5880
ctgtttcctt	tccacccgga	agtgccttgc	taggctttca	tggccatctg	actgagcatc	5940
taggcctcag	tccagtggtc	cctcagctct	ctctagtcat	tgtactaatg	gaaacggcca	6000
ctaactacat	tttcaatatg	gaagcctcct	cctcagggaac	ctccaagggc	agaagcctcc	6060
agagaaccac	tccctgacccc	ctggagttct	gagtgccttct	ggccctctct	gtgtctgcag	6120
gactattcac	cacttgtgtt	gaatggttca	gtcctcacct	cctctggcat	gtgtctcagtt	6180
ctcatctcat	tggggagttc	ttcccagggtc	actcttctct	cctgtctttg	aagtgttttt	6240
ttccttcatg	gtattttctgt	ctgggcacac	acacagacac	acatacacac	acatacacac	6300
ccatgcagta	tggcagatac	atcacctatg	tttcagatnt	ttattctacc	atcacccaat	6360

acctgaatcc	ccgaaaaagc	cttagaaagc	caggaatttg	tgtatTTTTTg	tcagcactcc	6420
accccagcac	ctgaagccaa	gcctgactta	atatttttgg	ttttgtttct	aga	6473

<211> 7045

<212> DNA

<213> *Cricetulus griseus*

<400> 4

(SEQ ID NO: 55)

caattgatta	tagatggatg	atagatagat	agatagatag	atagatagat	agatagatga	60
tggatagaca	gatgatggat	agttagagga	tagataatga	ctgaataata	agtacataaa	120
tagatgatag	agcggggcgt	tgggtggtgca	cgtcttttaac	cccagcacca	gagaggcaga	180
ggcagttgga	tctctgtgag	tttgaggaca	gcctgggttac	agaatgggtt	ccaggacagc	240
caaggctgtc	actcagagaa	atactgtctc	aaataaaaaa	agtaagtaaa	caaataaata	300
aatgataact	agttagaaga	tagatgattg	aatgataggt	agataaatag	aagatagata	360
gatagatgat	tgatagatga	tagacagata	gacagacaga	cagacagaca	gacagcagaa	420
agataatgca	cggtgaaaca	tgggtctgatt	tagttagcaa	gatcagagaa	gccttccttg	480
aaagtgacat	ttgagagcat	ttcaaacgct	gttcatgtca	ggcatgccaa	tggggagaga	540
agggccttgca	gaaagcaggc	ccggcaagcc	atggggagca	agctaggagg	cagcattcct	600
tgcattttgcc	tctgcctcag	ctgcttcctg	gagttccccg	gtttttatca	caacagtaga	660
aataaaaacca	ggacaatggt	gtttccatgc	atacatctgc	aagaacttac	tccggttcaa	720
tagacagacc	aaggcacctg	tgtttgctca	agaagcacgg	agggaggtgt	gtgcacctgc	780
tgggtgctgg	tgctctggct	gtgccagaca	gagagcaaga	caggaaagtt	cctgggtggcc	840
tagagcacac	agcccagccc	aggaagtcac	gtctctctct	gtctctgtct	ctgccccacc	900
cccaccccat	ttaggccaga	gaacagctgt	ggcaagcttt	gggtttgggt	gagtcattcc	960
tcaagagcca	agagccgccc	accttgatg	gggtagtttg	ttgtttgtgt	tgtttgtatt	1020
atltgtttgt	ttgtttgttt	ggtaaagggt	tttcaatagg	agttggaatt	tggcaattca	1080
gctaggctgg	ctgagcagcc	agctagcccc	gggcactcat	ccgtctctac	ctccccagtt	1140
ctgggatttc	gggtacatgc	tgccacatcc	gacttttttc	ccctgctcca	gttcttaaga	1200
ccaagtcttc	atgtcaaaaca	cttcaccacc	ttagccatct	ttctgggtca	gaagttagat	1260
cttcaggaag	acaaggagtg	tatcaggaca	tgagcgtgcc	ccaactctgc	tcagaccttc	1320
tgatagagaa	aatgggggga	gggggtgtcag	aggctgccgg	agaaagacaa	gtccagggtta	1380
aggaggacga	ccctgggctc	tgaatccaag	ggtgattccc	tcaccttgta	cacttggcat	1440
tttgggaagg	aagcatcaga	taaaagcagt	gcagacatag	tcaggaatat	ttacacgtgt	1500
gagtcaacct	gggagtgagt	ctgtgtacaa	ctgaacatga	agcaagtttt	gaagcttcat	1560
ttccagacta	ttcccagggt	gcaataactt	cctgttttct	ttgcagcctt	cccagtctct	1620
gccactgcca	tctctacttc	agtctggaat	ggtgggcaca	cagaaaaagt	ctatggcaat	1680
cctgcgagaa	gacaagtggg	cgctgacttt	cgggctcctg	ttacaagaga	ggaatccagg	1740
agttttatttt	gcagctgatt	cagtgttgac	caagagtcca	gctctggggg	agtgggaagc	1800
aaccaaagca	gagacaggtc	ccagcacaat	ttttgggtttt	caagacagca	cttctctgtg	1860
gctttgaagg	ctatcctaga	actgttcttt	gtatatcctt	ccttgcaact	agctcttata	1920
gaccaggctg	gtcttgaact	cacagagatc	catctgcctc	tgccctccaa	gtgctgggat	1980
taaaggcgtg	cacctcggtc	gccaccacc	agctacatac	ataattttaca	ataataaaaa	2040
taaaataactt	taaagtgtta	tagcagtttg	aatgtaattg	gccctgtcat	ctcataggga	2100
gtggcactat	taggaggtat	ggctttgttg	aaggaaatat	gtcactgtga	gggtgggctg	2160
tgaggttttcc	tatgctcagg	gtaccagcca	gtgtctcagc	tgaggctcctg	ttgcctgcaa	2220
gatgtaggac	tctcatccct	ttctccagca	ccatgtctgt	ctgcatgcca	tcatgtttccc	2280
agccatgatg	acaatgtact	aaaacctctg	aaactgccac	ccaactaaat	gttttccctt	2340
ataagagttg	ccatgctcat	ggtgtctctt	cacagcaata	gaaaccctaa	ctaagataag	2400
tgtattctcc	cctactcccc	atgattttaaa	atlttaggaag	gcaggtaggc	aggcaggcag	2460
gctggatatag	tgggttcattc	tagcacctga	gacctggaat	gggaggattg	tgagttagtt	2520
ctaggccatt	ctgggtgccta	gaaaccagag	ccggggggtg	gccaatgca	gagcacttgc	2580
tctacgtatg	gcccagcaca	ataagtcaat	ttcctcacct	taaaggcttg	acaattttaaa	2640
aacactggtt	tttaggttagt	ccgtgtctgc	tccacagatg	gagacagcta	atcacagatg	2700

catcaggggc	cttcctgagt	gctaaacatc	aaacagcctt	ctccccctct	gagcctttgt	2760
gtgcagaatg	tgtccatcgc	aagaagcaaa	cagtcttgc	tgccccacca	cttccttct	2820
gcatcagaag	agctgggtgc	aaactgcaag	agtagcctca	ccttagagat	gggtccatt	2880
gctctacatg	ggagcattac	cttccaagaa	ggcaaaaatg	tctcctgggt	gagctttttt	2940
tgtcacctgt	taaaggcaaa	tcaacagaga	ggctttgtct	caccactaa	catcttggaa	3000
acaaatacca	acgaacgctg	gggaggatgt	ggggaaagca	gagccctcat	gctctccgag	3060
ggaaaatcac	acccactgtg	gaacagtgtg	gaaacctcaa	agactgggat	tacaagcagc	3120
acacaagcca	gccacgctac	tcctgggtcac	acaccacaaa	gacgcttgca	cattcacgct	3180
tacgtcgca	acactagcaa	cgttcccact	gcctcctttg	agccccgccc	cccgccctg	3240
ccccccgccc	cgccctgtg	gtctatgttc	ctcttcccta	aagtcagctt	ccacttctct	3300
gtctccatct	tcgccccacc	ctccctcctc	gctacataat	tgtctctatt	ccatttctct	3360
gctttgaaac	agcttttttg	aaagcatcaa	atctattgtc	ctatgcccc	aatcaacctc	3420
cagtttcaca	agtgatacag	gaaatcgttt	tcctaattaa	aaatcccccc	tttgaccatt	3480
tattcccaact	cttggaacat	cttccccttg	aggaaaagtt	cagaatgagg	tggctctcct	3540
cttcctatct	gagggtgtttc	cttcagactt	tgtccgtgtc	taatcttttt	aactgttggc	3600
caggcctcca	ccacggcaca	gatgaactgt	ggggttcatt	tacctgaaac	tctatggaag	3660
gatgtttatt	tctccttcac	tttagcaaat	gataaagggc	accattcact	ctgtctatct	3720
tgcagggggc	attcctttct	ctaggccaga	tactgagaat	tgctcccaga	atcaatgtgg	3780
tatacatatt	tccccctcaa	cattgatagg	cattgatcac	acacacacac	acacacacac	3840
acacacacac	acacagtagc	acaaatgtat	tcccctagcc	cgcttccatc	ttgccacagg	3900
actccagagt	ggccctggat	agcaagcttc	ctgttttgtt	tctctgttcc	tgctgctttt	3960
ccaccctcca	gtctatcttt	tctaagtcct	tctgccattg	tcctcttccc	aactgtcctg	4020
agatgcagtc	attgtctggg	attcagacct	tctctctctg	cccaagtgag	tatattgacc	4080
cccacggttt	gtacaaccat	aacttcaggg	agccccgaca	aaactgtttt	atgagccaag	4140
tagtcccagg	acttgagagg	tagaggcggg	aagatcagca	gtttgaggcc	agcctggaga	4200
gcataagagc	cgggtctcaa	acaacaatgg	aaactagata	ctaagtaaaa	atcctggggg	4260
gtttcatcat	gaatgtctgt	tcttctagta	ccacgctgaa	ctccgtacac	agctccagct	4320
gttacggctt	tcttagaatc	catactcttt	tttttttttt	tttttttttt	tttttttttg	4380
tttttcgaga	cagggtttct	ctgtggcttt	ggaggctgtc	ctggaactag	ctcttataga	4440
ccaggctggg	ctcgaactca	cagagatcca	cctgcctctg	cctccagagt	gctgggatta	4500
aaggcgtgcg	ccaccaacac	ccggcagaat	ccatactctt	tttaaaaaaa	gatttatcaa	4560
tttactatgt	atacagcttt	ctgcctgcat	gtatccatgc	atgtcagaag	atggcaccag	4620
gtcgcattac	agatggttgt	gagccaccat	gtggttgctg	ggaattgaac	tcagaatgtc	4680
tagaagagca	accagttctc	ttaacctctg	agccatctct	ccggccccca	gaaatccata	4740
ttcttgagga	tttttttacac	cccccccacc	aaaagacgta	tatctaaatt	ttaatgtgag	4800
aattcacatt	ttcttaagag	ttgaacatag	attttagagga	aaatcagatc	ccacatgatt	4860
aacaaagcat	gcttgtgggc	aggtctgcta	ccaagagggtg	ggccgtagct	tctagctcag	4920
acaaactcac	tcccttctct	gtggcctctt	cgccctcaag	tcagaaactc	accctgtgat	4980
tctgccccag	aagttgctct	agagcacagt	gcatecttcc	gtcttccact	tgtggcttga	5040
attgtgtcca	tcgcttatga	ttacaacccc	tcacagagca	tcctaactgg	tttcttttgc	5100
tgcctatggg	cactcctcca	ttctagaaca	cccttgccat	caatactatg	aaaggagggg	5160
tggaggagga	agagcaggaa	gaggaggggg	aagcgaggga	agaggaagac	acggatggca	5220
atgaggaggg	gggagcacc	aagtccctcc	tggatgagag	tctcactggg	agacttaata	5280
ttaattataa	atgcttggtc	agcagctggg	caggataagg	ttaggcagga	gaaccagact	5340
aaggactctg	ggaagcagaa	gggcagagtc	agacaaggag	aggaaacagg	aagtacaagg	5400
taaagtacag	tggcagaatg	tagataatag	aaatgggttc	atttaagttg	gaagagttag	5460
ctagtaacaa	gcctgagcta	tcagccgagc	atttataatt	aatattgagt	ctccatattg	5520
gttatctggg	aattggcggg	cagaaaaaaa	aaagtctgcc	tacaagtcaa	tgtcatgtag	5580
ctcccaaagc	caaggtaoct	ttgttcagtg	cttgactgag	ccagcattat	aaattttctc	5640
cagatgtacc	gaatcacatt	tcatagcaac	atgcagacat	caagttttcc	ctgaagctct	5700
aaccagctgg	ttgcatgctg	tcgggagtct	cagctataac	ccagaagtga	cctgggtcgg	5760
ggaagagggtg	gtactttgcc	ttcttttgac	tctctgtgtt	gcctcaccca	ttcagcttca	5820
agcaatgtga	ctgcctgacc	ctgagggcgt	ttacaacgcc	tgacccacag	accacaagtc	5880
aaccagctgg	tgtgctcacg	atacctagtc	tgaaccatag	ccctgctccc	accctgcctc	5940
catctccacc	ctttcttcac	tgctcatcac	agctggctag	caaagactgc	ctcagacctg	6000
agcacaggct	ccactccaca	gccgtgactg	ttcgagccac	ttaaatcaaa	gagcgcttgt	6060
cttccgctca	gtaaatctct	cctcagctca	ctgatgacgt	tgactttctc	tagacagcac	6120
atttgggttt	aagacactgc	tacttgagct	cttcattcag	ttcctcagaa	tacctatttt	6180

gggtcagatt	cccaaagagg	aagataggg	tcttggcaga	cagacatgtc	tcatttcctt	6240
gaaatccttc	agagaaatgc	agtgactatg	gcaccttctt	aaaaagcaca	cacacaaata	6300
acacacacac	acacacacac	acacacacac	acacacacac	atatccccct	cactgtcatc	6360
cttgatatgt	atatgatata	tataaaatca	ttgtttttata	ctgtgataat	tgattatgaa	6420
taaaattttac	taaaatgaac	aattaaaaatt	atggggggggg	ctggagagat	ggctcatcag	6480
ttaagagaac	agttgctgct	cttgcagaac	acgagagttc	agttcccagc	accacatca	6540
ggcagctcat	aacctatgtgt	ggtgtcagtt	ccaggagatc	tggtgccctc	ttctggcctc	6600
ctccagcacc	tgctacatgt	ggttcacaca	cacacacaca	cacacacaca	cacacacaca	6660
cacacacaca	caaataaata	taaagattat	ttttttcaaa	actgagttaa	aaatagggtc	6720
tatctgattc	atactaaggc	ttttcacagt	ggttaagtct	attagatatg	tctagccata	6780
tcctttctcc	cttctttctt	gaggagaggc	ttttaaagct	acaagttaca	gccttctttg	6840
caaataagag	taccatttaa	caggcctctg	accaatgaga	tgccagaatc	ggttgcccag	6900
gagcttccca	aacagtccat	tatagggaaa	ggtggtacaa	accagtagat	taggcatggt	6960
ccacttccta	agtgccgtgc	caaataagga	aatggcctca	aatgtttgcc	ttttatcttc	7020
accacactct	gaattgcacg	ctagt				7045

<211> 13515

<212> DNA

<213> *Cricetulus griseus*

<400> 5

(SEQ ID NO:56)

tctagaaaca	aaacccaaaa	tattaagtca	ggcttggtt	caggtgctgg	ggtggagtgc	60
tgacaaaaat	acacaaattc	ctggctttct	aaggcttttt	cggggattca	ggtattgggt	120
gatggtagaa	taaaaatctg	aaacataggt	gatgtatctg	ccatactgca	tggtgtgtga	180
tgtgtgtgta	tgtgtgtctg	tgtgtgtgcc	cagacagaaa	taccatgaag	gaaaaaaca	240
cttcaaagac	aggagagaag	agtgacctgg	gaaggactcc	ccaatgagat	gagaactgag	300
cacatgccag	aggaggtag	gactgaacca	ttcaacacaa	gtggtgaata	gtcctgcaga	360
cacagagagg	gccagaagca	ctcagaactc	caggggggtca	ggagtgggtc	tctggagggt	420
tctgcccttg	gagggttctg	aggaggaggc	ttccatattg	aaaatgtagt	tagtggccgt	480
ttccattagt	acagtgacta	gagagagctg	agggaccact	ggactgagge	ctagatgtct	540
agtcagatgg	ccatgaaagc	ctagacaagc	acttccgggt	ggaaaggaaa	cagcaggtgt	600
gaggggtcag	gggcaagtta	gtgggagagg	tcttccagat	gaagtagcag	gaacggagac	660
gcaactggatg	gccccacttg	tcaaccagca	aaagcttggg	tcttggttcta	agaggccagg	720
gacatgacaa	gggtgatctc	ggttttttaa	aggctttgtg	ttacctaate	acttctatta	780
gtcagataact	ttgtaacaca	aatgagtact	tggcctgtat	tttagaaaact	tctgggatcc	840
tgaaaaaaca	caatgacatt	ctggctgcaa	cacctggaga	ctcccagcca	ggcctgggac	900
ccgggtccat	tcatgcaaat	actcagggac	agattcttca	ctaggtactg	atgagctgtc	960
ttggatgcaa	atgtggcctc	ttcatttttac	tacaagtca	catgagtcag	gagggtgctgt	1020
ttgcacagtg	tgactaagtg	atggagtgtt	gactgcagcc	attcccggcc	ccagcttgtg	1080
agagagatcc	ttttaaattg	aaagtaagct	caaagttacc	acgaagccac	acatgtataa	1140
actgtgtgaa	taatctgtgc	acatacacia	accatgtgaa	taatctgtgt	acatgtataa	1200
actgtgtgaa	taatctgtgt	gcagcctttc	cttacctact	accttccagt	gatcagggtt	1260
ggactgcctg	tgtgctactg	gacctgaat	gtccccaccg	ctgtcccctg	tctttttacga	1320
ttctgacatt	tttaataaat	tcagcggctt	cccctctgct	ctgtgcctag	ctataccttg	1380
gtactctgca	tttttggtttc	tgtgacattt	ctctgtgact	ctgtacatt	ctcagatgac	1440
atgtgacaca	gaagggtgtc	cctctggaga	catgtgatgt	ccctgtcatt	agtggaatca	1500
gatgccccca	aactgtttgt	cagtgttttg	gaaagtgaca	cgtgaaggag	gatcaggaaa	1560
agaggggtgg	aaatcaagat	gtgtctgagt	atctcatgtc	cctgagtggg	ccaggtgtgt	1620
gacttcactc	ccccaaagtga	gggaggccat	ggtgagtaca	cacacctcac	acatactata	1680
tccaacacac	acacacacac	acacacacac	acgcacgcac	gcacgcacgc	acgcacacac	1740
gcacacacac	gaactacatt	tcacaaaacca	catacgcata	ttacacccca	aacgtatcac	1800
ctatacatac	cacacataca	caccctcca	cacatcacac	acataccaca	cccacacaca	1860
gcacacacac	acattagcac	acattcacac	accacacata	tacatttgtg	tatgcataca	1920
tgacatacaca	cacagggcaca	cagacaccac	acacatgcac	tgtgtacgca	cacatgcata	1980

cacacacata	ggcacacatt	gagcacacac	atacatttgt	gtacgcacac	tacatagaca	2040
tatatgcatt	tgtatatgca	cacatgcatg	cacacatata	taggcacaca	tagagcacac	2100
acatacattt	gtgtatgcac	acatgcacac	accaatcaca	tgggaagact	caggttcttc	2160
actaagggttc	acatgaactt	agcagttcct	ggttatctcg	tgaaacttgg	aagattgctg	2220
tggagaagag	gaagcgttgg	cttgagccct	ggcagcaatt	aaccccgccc	agaagaagta	2280
ggttttaaaaa	tgagaggggtc	tcaatgtgga	acccgcaggg	cgccagttca	gagaagagac	2340
ctacccaagc	caactgagag	caaaggcaga	gggatgaacc	tgggatgtag	tttgaacctc	2400
tgtaccagct	gggcttcatg	ctatttttgt	atatctttat	taaatattct	tttagtttta	2460
tgtgcgtgaa	taccttgctt	gcataaatgt	atgggcactg	tatgtgttct	tggtgccggt	2520
ggaggccagg	agagggcatg	gacctccgg	agctggcggt	tgagacagtt	gtgaccaca	2580
gtgtggggtc	tgggaactgg	gtcttagtgt	tccgcaagtg	cagctggggc	tcttaacctc	2640
tgagccatcc	ctccagcttc	aagaaactta	ttttcttagg	acatggggga	agggatccag	2700
ggcttttaggc	ttgtttgttc	agcaaatact	cttttctgtg	attttgaatt	ttattttatt	2760
ttactttttt	gggatagaat	cacattctgc	agctcaggct	gggcctgaac	tcatacaaat	2820
cctcctgtct	cagtctacca	ggtgataaga	ttactgatgt	gagcctggct	ttgacaagca	2880
cttttagagtc	cccagccctt	ctggacactt	gttccaagta	taatataat	atatatata	2940
atatatata	atatatata	atatattgtg	tgtgtgtgtt	tgtgtgtgta	tgagacactt	3000
gctctaaggg	tatcatatat	atccttgatt	tgcttttaat	ttatttttta	attaaaaatg	3060
attagctaca	tgtcacctgt	atgcgtctgt	atcatctata	tatccttctc	tccttctctc	3120
tctttctctc	ttcttcttct	cacccccaa	catctatttt	caaatccttg	tgccgaggag	3180
atgccaagag	tctcgttggg	ggagatggtg	agggggcgat	acaggggaag	agcaggagga	3240
aagggggaca	gactggtgtg	ggtctttgga	gagctcagga	gaatagcagc	gatcttccct	3300
gtccctggtg	tcacctctta	cagccaacac	catttttgtg	cctggcagaa	gagttgtcaa	3360
gctggtcgca	ggtctgccac	acaaccccaa	tctggcccca	agaaaaggca	cctgtgtgtg	3420
actctgggg	taaaggcgct	gcctggtcgt	ctccagctgg	acttgaaact	cccgtttaat	3480
aaagagttct	gcaaaataat	acccgcagag	tcacagtgcc	aggttcccgt	gctttcctga	3540
agcgccaggc	acgggttccc	taggaaatgg	ggccttgctt	gccaagctcc	cacggcttgc	3600
cctgcaaacg	gcctgaatga	tctggcactc	tgcgttgcca	ctgggatgaa	atggaaaaaa	3660
gaaaaagaag	aagtgtctct	ggaagcggcg	gcgctcacac	aaaccgcgaa	cgatttgtga	3720
aacactctcc	attgagaatc	tggagtgcg	ttgccctcta	ctggggagct	gaagacagct	3780
agtgggggcg	gggggaggac	cgtgctagca	tccttccacg	gtgctcgctg	gctgtggtgc	3840
atgccgggaa	ccgaaacgcg	gaactaaagt	caagtcttgc	tttgggtggaa	ctgacaatca	3900
acgaaatcac	ttcgattgtt	ttcctctttt	tactggaatt	cttggatttg	atagatgggg	3960
gaggatcaga	gggggagggg	aggggcgggg	agacggaggg	aggaggggag	gaggggagga	4020
ggggaggagg	ggaggagggg	aagggatgga	ggaaaatact	aacttttcta	attcaacatg	4080
acaaagattc	ggagaaagt	caccgctagt	gaccgggagg	aggaatgcc	tattgggcat	4140
tatatccct	gtcgtcta	ggaatcaa	tcttggttcc	agcaccaagg	attctgagcc	4200
tatcctattc	aagacagtaa	ctacagccca	cacggaagag	gctatacaac	tgaagaaata	4260
aaattttcac	tttatttcat	ttctgtgact	gcatgttcac	atgtagagag	ccacctgtgt	4320
ctaggggctg	atgtgctggg	cagtagagtt	ctgagcccgt	taactggaac	aaccagaac	4380
tcccaccaca	gttagagctt	gctgagagag	ggaggccctt	ggtgagattt	ctttgtgtat	4440
ttatttagag	acagggtctc	atactgtagt	ccaagctagc	ctccagctca	cagaaattct	4500
cctgttccgg	tttccaaagt	actggagtta	tgagtgtgtg	ttaattgaac	gctaagaatt	4560
tgtgattga	agaaaacctc	aagtgggttt	ggctaatacc	cacgaccca	gaggctgagg	4620
caggaggaat	gagagaattc	aaggtttgcc	agagccacag	ggtgagctca	atgtggagac	4680
tgtgagggtg	agctcaatgt	ggagactgtg	agggtagct	caatgtggag	actgtgagg	4740
tgagctcaat	gtggagactg	tgagggtag	ctcaatgtgg	agactgtgag	ggtgagctca	4800
atgtggagac	ctgtatcaag	ataataatag	tagtagtaac	aatgcaggcg	aggggtggt	4860
tgagtggtag	agcagttagt	tgatttgaca	tgcttgaggt	ctcccggtcc	atctgtggcc	4920
ctgcaacagg	aagggaggga	ggaagggggg	gaacgagaga	gaggaaagag	agacagaagc	4980
taagataggg	aatgagagag	gaagggaagaa	acgggaagaa	attcagactc	cttctgagt	5040
tccgccaacg	cctagtga	tcctgtgcac	accctaaggt	ggcctttgtg	tggcactggc	5100
ttgggtggtc	gggaaaggca	ttttcagctt	gttgcagaac	tgccacagta	gcatgctggg	5160
tccgtgaaag	tttctgccc	ttaacaagaa	gtctctacta	cttgtgacct	caccagtga	5220
aatttcttta	attgtctcct	ggtgttctgg	gttttgcatt	tttgtttcta	aggatacatt	5280
cctgggtgat	gtcatgaagt	ccccaaagac	acagtggggc	tgtgttggat	tgggaaagat	5340
gatttatctg	gggtgtcaaa	aggaaaagaa	gggaaacagg	cacttgggaa	aatgtcctcc	5400
cgcccacccg	aatttttggt	tggcaaccgt	ggtggaggag	caagaaacac	gtggacgttt	5460

gaggaggcat	ggggtcctag	gaggacagga	agcagaagga	gagagctggg	ctgacagcct	5520
gcaggcattg	cacagtttca	gaaggagatt	acagcatgac	tgagttttta	gggatccaac	5580
agggacctgg	gtagagattc	tgtgggctct	gaggcaactt	gacctcagcc	agatggtatt	5640
tgaataacct	gctcttagag	ggaaaacaga	catagcaaac	agagccacgt	ttagtgatga	5700
aactctcact	ttgcctgagt	catgtgcggc	catgccaggg	ggtcaggctg	acactcaact	5760
caaaaacaag	tgagaaattg	aagacaatcc	gtgggtggcag	ctactggaag	ggccaccaca	5820
tccccagaaa	gagtggagct	gctaaaaagc	cattttgtgat	aggcacagtt	atcttgaatg	5880
catggagcag	agattacgga	aaaatcgaga	atgttaatga	ggcaacattc	gagttgagtc	5940
attcagtggt	ggaaacccag	acgcttccat	cccctaaaag	gaacatcttg	ctctcagtc	6000
aaatggaaat	aaaaattggg	gcttgaattt	ggcaaattgat	tcagaactct	gtgtaggtat	6060
tttcacacgc	acagtggata	attttcatgt	tggagtttat	ttgtgctaaa	aggcagaaaa	6120
gggtaaaaag	cacatcttaa	gagttatgag	gttctacgaa	taaaaataat	gttacttaca	6180
gctattcctt	aattagtacc	cccttccacc	tgtggtaatt	tcctgagata	gtcagtgggg	6240
aaaagatctc	tccttctctt	ctttctcccc	ctccccctct	ctccccctct	ccctccctcc	6300
ctccccctct	tccttccctc	ccccctttct	tctttctttg	ctccttctcc	tctgcctcct	6360
tctccccctt	ttcttcatth	attctaagta	gcttttaaca	gcacaccaat	tacctgtgta	6420
taacgggaaa	acacaggctc	aagcagctta	gagaagattg	atctgtgttc	actagcgtgc	6480
aattcagagg	tgggtgaaga	taaaaggcaa	acatttgagg	ccatttccct	atttggcacg	6540
gcacttagga	agtggaaacat	gcctaatact	ctgggtttgt	ccacctttcc	ctataatgga	6600
ctgtttggga	agctcctggg	caaccgattc	tggcatctca	ttgggtcagag	gcctgtttaa	6660
tggtactctt	atlttgcaaag	aaggctgtaa	cttgtagctt	taaaagcctc	tcctcaagaa	6720
agaagggaga	aaggatatgg	ctagacatat	ctaatagact	taaccactgt	gaaaagcctt	6780
agtatgaatc	agatagaacc	tatttttaac	tcagttttga	aaaaaataat	ctttatatth	6840
atlttgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gaaccacatg	6900
tagcagggtg	tggaggaggc	cagaagaggg	caccagatct	cctggaactg	acaccacaca	6960
tggttatgag	ctgcctgatg	tgggtgctgg	gaactgaact	ctcgtgttct	gcaagagcag	7020
caactgttct	cttaactgat	gagccatctc	tccagccccc	cccataattt	taattgttca	7080
tttttagtaaa	ttttattcat	aatcaattat	cacagtataa	aacaatgatt	ttatatatat	7140
catatacata	tcaaggatga	cagttagggg	gatatgtgtg	tgtgtgtgtg	tgtgtgtgtg	7200
tgtgtgtgtg	tgtgtgtgtg	gtgtgtgtgc	tttttaagaa	gggtgccatg	tcactgcatt	7260
tctctgaagg	atlttcaaagg	aatgagacat	gtctgtctgc	caggaacctc	atcttctctc	7320
ttgggaatct	gacccaaatg	aggtattctg	aggaactgaa	tgaagagctc	aagtagcagt	7380
gtcttaaaac	caaattgtgt	gtctagagaa	agtcaacgtc	atcagtgagc	tgaggagaga	7440
tttactgagc	ggaagacaag	cgctctttga	tttaagtggc	tcgaacagtc	acggctgtgg	7500
agtggagcct	gtgctcaggt	ctgaggcagt	ctttgctagc	cagctgtgat	gagcagtga	7560
gaaaggggtg	agatggaggc	aggggtgggag	cagggctatg	gttcagacta	ggtatcgtga	7620
gcacaccagc	tgggtgactt	gtgggtctgtg	gggtcaggcgt	tgtaaacgcc	ctcaggggtc	7680
ggcagtcaca	ttgcttgaag	ctgaatgggt	gaggcaacac	agagagtgca	aagaaggcaa	7740
agtaccacct	cttccccgac	ccaggctcact	tctgggttat	agctgagact	ccggacagca	7800
tgcaaccagc	tgggttagagc	ttcagggaaa	acttgatgtc	tgcattgttc	tatgaaatgt	7860
gattcgggtac	atctggagaa	aattttataat	gctgggtcag	tcaagcactg	aacaaaggta	7920
ccttggcttt	gggagctaca	tgacattgac	ttgtaggcag	actttttttt	ttctgcccgc	7980
caattcccag	ataaccaata	tggaggtctca	atattaatta	taaatgctcg	gctgatagct	8040
caggcttgtt	actagctaac	tcttccaaact	taaatgaacc	cattttctatt	atctacattc	8100
tgccacgtga	ctttaccttg	tacttctctg	ttctctctct	tgtctgactc	tgcccttctg	8160
cttcccagag	tccttagtct	ggttctctctg	cctaaccctta	tcctgcccag	ctgctgacca	8220
agcatttata	cttaatatata	agttctcccg	tgagactctc	atccagggag	gacttgggtg	8280
ctccccctct	ctcattgcca	tcctgtctct	cctcttccct	cgcttccccc	tcctcttctc	8340
gctcttctct	ctccaccctc	cctttcatag	tattgatggc	aagggtgttc	tagaatggag	8400
gagtgcccat	aggcatgcaa	agaaaccagt	taggatgtct	tgtgaggggt	tgtaatcata	8460
agcgatggac	acaattcaag	ccacagagtg	aagacgggaag	gatgcactgt	gctctagagc	8520
aacttctggg	gcagaatcac	aggggtgagtt	tctgacttga	gggcgaagag	gccacgagga	8580
agggagtggg	tttgtctgag	ctagaagcta	cggcccacct	cttggttagca	gacctgccc	8640
caagcatgct	ttgttaatat	tgtgggatct	gattttctct	taaatctatg	ttcaactctt	8700
aagaaaatgt	gaattctcac	attaaaattt	agatatacgt	cttttggtgg	gggggggtgt	8760
aaaaatcctc	aagaatatgg	atlttctggg	gccggagaga	tggctcagag	gttaagagaa	8820
ctgggtgtct	ttctagacat	tctgagttca	attcccagca	accacatggg	ggctcacaac	8880
catctgtaat	gcgacctggg	gcatcttctc	gacatgcatt	gatacatgca	ggcagaaagc	8940

tgtatacata	gtaaattgat	aaatcttttt	ttaaaaagag	tatggattct	gccgggtggt	9000
ggtggcgcac	gcctttaatc	ccagcactct	ggaggcagag	gcagggtgat	ctctgtgagt	9060
tcgagaccag	cctgggtctat	aagagctagt	tccaggacag	cctccaaagc	cacagagaaa	9120
ccctgtctcg	aaaaacccaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaga	gtatggattc	9180
taagaaagcc	gtaacagctg	gagctgtgta	cggagttcag	cgtgggtacta	gaagaacaga	9240
cattcatgat	gaaacacccc	aggattttta	cttagtatct	agttttccatt	gttggttttga	9300
gaccggctct	tatgctctcc	aggctggcct	caaactgctg	atcttcccgc	ctctacctct	9360
caagtcctgg	gactacttgg	ctcataaaac	agtttttgtc	gggctccctg	aagttatggg	9420
tgtacaaacc	gtgggggtca	atatactcac	ttgggcagag	agagaagggtc	tgaatcccag	9480
acaatgactg	catctcagga	cagttgggaa	gaggacaatg	gcagaaggac	ttagaaaaga	9540
tagactggag	ggtggaaaag	cagcaggaac	agagaaacaa	aacaggaagc	ttgctatcca	9600
ggggccactct	ggagtccctgt	ggcaagatgg	aagcgggcta	ggggaataca	tttgtgctac	9660
tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgat	caatgcctat	caatggtgaa	9720
ggggaaatat	gtataccaca	ttgattctgg	gagcaattct	cagtatctgg	cctagagaaa	9780
ggaatggccc	ctgcagaata	gacagagtga	atgggtgccct	ttatcatttg	ctaaagtgaa	9840
ggagaaataa	acatcccttc	atagagtttc	aggtaaataa	acccacaggt	tcattctgtgc	9900
cgtgggtggag	gcctggccaa	cagttaaaaa	gattagacac	ggacaaagtc	tgaaggaaac	9960
acctcgaata	ggaagaggag	agccacctca	ttctgtaact	ttcctcaagg	ggaagatggt	10020
ccaagagtgg	gaataaatgg	tcaaaggggg	gattttttaat	taggaaaacg	atttcctgta	10080
tcacttgtga	aactggaggt	tgatttgggg	cataggacaa	tagatttgat	gctttgcaaa	10140
aagctgtttc	aaagcagaga	aatggaatag	agacaattat	gtagcgagga	gggaggggtg	10200
ggcgaagatg	gagacagaga	agtggaagct	gactttaggg	aagaggaaca	tagaccacag	10260
gggcggggcg	gggggcaggg	gcggggggcg	gggctcaaag	gaggcagtg	gaacgttgct	10320
agtgttcgca	gcgtaagcgt	gaatgtgcaa	gcgtctttgt	ggtgtgtgac	caggagtagc	10380
gtggctggct	tgtgtgctgc	ttgtaatccc	agtctttgag	gtttccacac	tgttccacag	10440
tgggtgtgat	tttccctcgg	agagcatgag	ggctctgctt	tccccacatc	ctccccagcg	10500
ttcgttggtg	tttgtttcca	agatgttagt	gggtgagaca	aagcctctct	gttgatttgc	10560
ctttaacagg	tgacaaaaaa	agctcaacca	ggagacattt	ttgccttctt	ggaaggtaat	10620
gctcccattgt	agagcaattg	gacccatctc	taagtgagag	ctactcttgc	agttttgcacc	10680
cagctcttct	gatgcaggaa	ggaagttggg	gggcaagcaa	gactgtttgc	ttcttgccat	10740
ggacacattc	tgacacaaa	ggctcaggag	gggagaaggc	tgtttgatgt	ttagcactca	10800
ggaaggcccc	tgatgcatct	gtgattagct	gtctccatct	gtggagcaga	cacggactaa	10860
ctaaaaacca	gtgtttttta	attgtcaagc	ctttaagggtg	aggaaattga	cttattgtgc	10920
tgggccatac	gtagagcaag	tgctctgcat	tgggccaaac	cccggctctg	gtttctaggc	10980
accagaatgg	cctagaacta	actcacaatc	ctcccatctc	aggctctcagg	tgctagaatg	11040
aaccactata	ccagcctgcc	tgctgccta	cctgccttcc	taaattttta	atcatgggga	11100
gtaggggaga	atacacttat	cttagttagg	gtttctattg	ctgtgaagag	acaccatgag	11160
catggcaact	cttataaagg	aaaacattta	gttgggtggc	agtttcagag	gttttagtac	11220
attgtcatca	tggctgggaa	catgatggca	tgcagacaga	catggtgctg	gagaaagggg	11280
tgagagtcc	acatcttgca	ggcaacagga	cctcagctga	gacactggct	ggtaccctga	11340
gcataggaaa	cctcacagcc	caccctcaca	gtgacatatt	tccttcaaca	aagccatacc	11400
tcctaatagt	gccactccct	atgagatgac	agggccaaat	acattcaaac	tgctataaca	11460
ctttaaagta	ttttattttt	attattgtaa	attatgtatg	tagctgggtg	gtggcagccg	11520
aggtgcacgc	ctttaatccc	agcacttggg	aggcagaggc	agatggatct	ctgtgagttc	11580
aagaccagcc	tggtctataa	gagctagtgt	caaggaagga	tatacaaaga	acagttctag	11640
gatagccttc	aaagccacag	agaagtgtcg	tcttgaaaac	caaaaattgt	gctgggacct	11700
gtctctgctt	tggttgcttc	ccactcccc	agagctggac	tcttgggtcaa	cactgaatca	11760
gctgcaaaat	aaactcctgg	attcctctct	tgtaacagga	gcccgaagtc	aggcgcccac	11820
ttgtcttctc	gcaggattgc	catagacttt	ttctgtgtgc	ccaccattcc	agactgaagt	11880
agagatggca	gtggcagaga	ctgggaaggc	tgcaacgaaa	acaggaagtt	attgcaccct	11940
gggaatagtc	tggaaatgaa	gcttcaaaac	ttgcttcatg	ttcagttgta	cacagactca	12000
ctcccagggt	gactcacacg	tgtaaatatt	cctgactatg	tctgcaactgc	ttttatctga	12060
tgcttccctc	ccaaaatgcc	aagtgtacaa	ggtgagggaa	tcacccttgg	attcagagcc	12120
cagggtcgtc	ctccttaacc	tggacttgtc	tttctccggc	agcctctgac	acccctcccc	12180
ccattttctc	tatcagaagg	tctgagcaga	gttggggcac	gctcatgtcc	tgatacactc	12240
cttgtcttcc	tgaagatcta	acttctgacc	cagaaagatg	gctaagggtg	tgaagtgttt	12300
gacatgaaga	cttgggtctta	agaactggag	caggggaaaa	aagtcggatg	tggcagcatg	12360
tacccgaaat	cccagaactg	gggaggtaga	gacggatgag	tgcccggggc	tagctggctg	12420

ctcagccagc	ctagctgaat	tgccaaattc	caactcctat	tgaaaaacct	ttaccaaaaca	12480
aacaaacaaa	caaataataa	caacaacaac	aacaacaaac	taccccatac	aaggtgggcg	12540
gctcttggct	cttgaggaat	gactcaccca	aacccaaagc	ttgccacagc	tgttctcttg	12600
cctaaatggg	gtgggggtgg	ggcagagaca	gagacagaga	gagacatgac	ttcctgggct	12660
gggctgtgtg	ctctaggcca	ccaggaactt	tccctgtctg	ctctctgtct	ggcacagcca	12720
gagcaccagc	acccagcagg	tgcacacacc	tccctccgtg	cttcttgagc	aaacacaggt	12780
gccttgggtc	gtctattgaa	ccggagtaag	ttcttgcaga	tgtatgcatg	gaaacaacat	12840
tgtcctgggt	ttattttctac	tgttgtgata	aaaaccgggg	aactccagga	agcagctgag	12900
gcagaggcaa	atgcaaggaa	tgctgcctcc	tagcttgctc	cccatggctt	gccggggcctg	12960
ctttctgcaa	gcccttctct	ccccattggc	atgcctgaca	tgaacagcgt	ttgaaatgct	13020
ctcaaagtgc	actttcaaag	aaggcttctc	tgatcttgct	aactaaatca	gaccatgttt	13080
caccgtgcat	tatctttctg	ctgtctgtct	gtctgtctgt	ctgtctatct	gtctatcatc	13140
tatcaatcat	ctatctatct	atcttctatt	tatctaccta	tcattcaatc	atctatcttc	13200
taactagtta	tcattttattt	atttgtttac	ttactttttt	tatttgagac	agtatttctc	13260
tgagtgcacg	ccttggctgt	cctggaaccc	attctgtaac	caggctgtcc	tcaaactcac	13320
agagatccaa	ctgcctctgc	ctctctgggtg	ctgggggttaa	agacgtgcac	caccaacgcc	13380
ccgctctatc	atctatttat	gtacttatta	ttcagtcatt	atctatcctc	taactatcca	13440
tcctctgtct	atccatcatc	tatctatcta	tctatctatc	tatctatcta	tctatcatcc	13500
atctataatc	aattg					13515

<211> 14553

<212> DNA

<213> Mus musculus

<400> 6

(SEQ ID NO:57)

cttgaagaac	acatgttttc	caagagggag	cacccatggt	ggaatgacaa	tgtagttagt	60
gctcctctcc	tgtaggttag	tgctcctttg	ctataggtaa	gtgctcctct	cctatagggtc	120
agtgtccctc	tcctataggt	tagtgctcct	ctcctatagg	ttagtgctcc	tctcctacag	180
gttagtgctc	ctctgctcta	ggtagtcct	gctctcctat	agtacctaga	gagctagggc	240
aaatgggcta	ggcccgaagt	gcagagacaa	acagctatgg	aagactgggt	aagcacttcc	300
aagctacgaa	agagcagtgt	gaagggtcag	ggcttgtgca	gttagtaggg	gagatcttcc	360
agttgaagaa	acagaagaac	tgagagccac	tgggtatcat	cctcctgcgc	catgccttcc	420
tggatactgc	catgtcctcc	ccttgatgat	aatggaatga	acctctgaac	ctgtaagcca	480
gccccaatga	aataattgtt	ttatgagagt	tgcttgggtc	atgctgtctg	ttcacagcag	540
taaaacccta	aataaggcag	aagttgggtac	cagtattgct	gtgatagacc	tgaccatgct	600
ttcctttgaa	agaatgtgga	tttggtgact	ttggatttgc	aacacagtgg	aatgctttaa	660
atggagatta	atgggtcatc	aattcctagt	aggaatatgg	aagactttgt	tgctggggagt	720
atltgaaactg	tgttgacctg	gcctaagaga	tttcaaagga	gaagaatttc	agaatgtggc	780
ataaagacag	tttttgtggt	atlttgggtga	agaatgtggc	tacttttttgc	ccttgtctga	840
aaagtctgcc	tgagactaaa	gtgaagagaa	tcagattaat	tgcattgaca	agggaaagttt	900
gtggctgcgc	tatctggaaa	cttacagcca	gcctcttgga	cctcgggtga	cttacgcaaa	960
tactcaggga	cagagatgct	tgactctgta	ctgatgagtt	gtcttggatg	caaatatggg	1020
ctcttcattt	gactacatgt	cacgatgagt	caggagctgc	tctctccaga	gtgtgacaaa	1080
gcgaggggat	gctgacggtg	gctgttctag	ctttgaaggt	aagcctgcac	ttatgctaaa	1140
gtcacacata	cacgagccgg	gtggagaacc	tgtctgtgtg	gagacacctt	tcattacctg	1200
tggcatccag	cctctcaagc	ttggactgcc	tgtgtgctcc	tggactctgg	aggtccact	1260
gctctgtcct	ctgctgctta	tgatactgac	atltttaaag	aatccagtgg	ttccccctg	1320
tactcggtgt	ctacttctac	ctggatgttc	ctcatlttatg	ttctgtgaca	cttctctgtg	1380
actctgctgc	attcctgggt	gacatgtgga	caccctgtcc	ctttgcagac	catgatgtca	1440
ctgtcactag	tggaaatcaga	tgccccaaagt	gttgtcctgt	gtttgggaac	gtgacaggca	1500
gtacagaagc	agaagaggaa	gggtgaaaac	ggaaatgtca	cagcagcatc	tgatgtgtgc	1560
ctcagtcacg	catgctgctg	atttgaacta	ctcagcatga	gagagggcca	tgggtgaatac	1620
acaaccctat	acacactgtg	tccatttctc	tctctctctt	acacagagag	agagggagga	1680
gggggagggg	gagggcggag	ggggagggga	ggggagaggg	gtgggagagg	gagagggaga	1740
gggagagggg	gagggagagg	gagagggaga	gggagagttt	aatgtctgtg	aagagatacc	1800

atgaccaaag	caactcttat	aaaggacaac	atttaattgg	ggctggctta	caggttcaga	1860
aattcagtc	attctcacca	tgggtgggaag	catgcaggta	gatgtgggtgc	tggaggaacc	1920
aagagttcta	tatcctgac	tgaaggcagc	caggagaaga	ctgcctcttc	tgcacagggc	1980
agagcttgag	catagaacat	caaagccctt	ccccacactt	cctccaacaa	ggtcatacat	2040
acttcaacaa	agacacacct	cctaacggtg	ccactccctg	tggaccaacc	atttaaaccg	2100
atgagtctat	gaggggtcaaa	gctcttcaaa	ccaccacact	catgtacaca	cacacacaca	2160
cacacacaca	ctctcataca	cacacacaca	cacactcaca	cacacacaca	cacacacaca	2220
cacacacaca	ccacacacac	acacacacac	agagttctat	tttgcaactgt	ttcactgtca	2280
caaggttcta	cttatctcag	acacactgcc	aggaattgtg	tgggaagact	ttcagtttct	2340
ttgggttcac	atggacttag	cagttcttgg	tgatcctgaa	agatttctgc	agaaagaagc	2400
caaagtgttg	agcccaaggc	ctggccacac	attagtcctg	tctagatgaa	caggggttta	2460
aaaataaggg	ggcatcaagg	tgaagccagc	aggggctgac	ttagagagga	gaccaccca	2520
agccaactgc	togaagtcaa	aagcgatgaa	tccccatata	cagctgtgcc	cggtgtgtc	2580
ttgctacatc	tttagtaaat	gttcttttag	ttgtatgcgt	atgaatatatt	tgcttgcata	2640
tattttgtgta	caccataggt	gttcttaggg	cctatggagg	ccagaagagg	gcatcagatc	2700
ctttggaact	ggaattatag	acacttggtta	cccatagagt	agatttgtggg	aatgagcct	2760
ttagtcttcg	agagcggcca	gtgctcttaa	cctttggctg	tttctccagg	tctttgagac	2820
tttattttct	tggacatcag	gacaggatcc	agggctttga	gcttgtttct	tcagccagct	2880
ttcttttcat	gtatatataa	ttttatgtta	ttttgcttct	tttttcccca	agacagaatc	2940
acactctata	tagctcaggc	tgggtttgaa	ttcagtttcc	ctgtctcagt	ctaccgggta	3000
atatgattac	agatgtgagt	ctgacttttg	tatcaaagtc	cccagccctt	ctggatatgt	3060
gttttaagga	tatcagatat	atccttgatt	tgctttgaat	tttcttttta	gttacaacat	3120
aattagttcc	gtgtcacctg	aatatgtgta	tgtcacctac	atagtcttcc	ttcttctctt	3180
cttccctctc	ccaccttccc	aggtacctgt	ctgtcttcat	atccttgtgc	tgagagtctt	3240
gttgagggag	atgatgaccg	agacagagcc	actggggaag	ggagatgggc	tagtgcaggt	3300
cttcagagag	gagctcgtga	atattgtagc	cccttttagtc	cctggcatgt	cctcttgtat	3360
agccaccgcc	atgctgtggc	ctggcagaag	tgaataagtt	gtccagctgt	tgacaggcct	3420
gccctccaga	cccagtctga	tcccaagaaa	gggcatctgt	gtctgtctct	gaggccgtaa	3480
gtgctgcctg	gttgtctcca	gcttgacttg	acactccctc	cttaataaga	gtaccacaga	3540
acagggctga	cagagtcctc	gggccaagtc	cctgtgctgt	cctggaatgc	caggcgtgaa	3600
tttccgtgtg	agtaggactt	tgctcgccaa	gctcccacgg	cttgcccttc	agatagccag	3660
aattatctgg	tacctgcat	tgccgttcaa	tacgcagagt	atcactggaa	gcgcgcgcgc	3720
gcacacacac	acacacacac	acacacacac	acacacacac	acacgcccac	tccatcttta	3780
aacccccacc	cccagcaacg	gcggtgtaaa	cactctccat	caggaagctg	aaacgcagtt	3840
gccctctgct	ggggagatga	aggcagcttg	ctgggggcga	ggaccgtgct	agcaaccttc	3900
cctgggtgcac	acgggctctg	gtgcatgacg	ggaacggaaa	cgcggaacta	aagtcagtcc	3960
tgcttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	ggcgttggtg	4020
gtggactgag	tgacaatcag	tgaaatcact	taggttggtt	ttctcttctt	cgttggggtt	4080
gatagacggt	gggagagggg	cagaggagaa	ggggagggat	ggggagagag	ggaggagggg	4140
ggggcgggag	gcggggggcg	aggaaaacgt	gctaacttct	ccaatcctac	aagacaaagg	4200
tttgagaaaa	gccgcactga	gtgacccagc	agaaggaatc	caggaatgtc	cgctggaatc	4260
tgactgttga	ttccagcgcc	atgcagagaa	tctaggctgg	taggaacatt	ctttgtccta	4320
tccgacataa	taactccaac	caacacggaa	aagaaaggct	atacaagtga	agaaatggca	4380
ttttcacttt	catgactata	caatcacttc	caggtagtaa	cacgtgtcta	gcacagcggg	4440
tctcaacctg	ggggtcacga	tccccacttt	ttctgcata	cagacatttt	tacgttggtta	4500
ttcataacag	tagcaaaatt	gcagctatga	agtaacaatg	aaatgcattt	atggtgcgtg	4560
tgtgtgtgtg	tgggggggga	tcaccttaac	atttactgta	agaagggttg	gaatactgct	4620
ccagcagcta	gtgtgttggg	cttaggttct	gggtatatta	ttagcaatag	ccaaccagaa	4680
tccccaccca	ccacagcatt	gaggccccat	gcagggcttg	ctgggagagg	cactgataag	4740
acttctttat	gtattttatt	agagacgaat	actcattagg	taggccaagc	tagcgtcaaa	4800
ctcatggcaa	ttctctctct	ccagtttctt	aagtaactga	ctcaggagtg	tgttgccatc	4860
atatacagta	aggattttatt	gactgaagaa	aatctcaagt	ggcttttggt	aatccctact	4920
acgccagagg	ctgaggcagg	aggcgcgcaa	ggtcaaggct	tgcttgggct	acatatagag	4980
tgagctcaat	tttgacactt	ggtgcgggtg	tagtagtaat	agtaaagatg	aaggtgtggc	5040
tcaggtgggg	ccggtgattg	gacacacttg	gggtctctct	gtccatctgc	agctgtgcaa	5100
caggaaagag	ggagaatgag	aggaaagaga	gaaaagacag	aatgagagag	agggaggaag	5160
agagaaaaag	gaaaagagag	aggaaaggaa	aaaggaaaat	gaggaaaagc	agaaagaaga	5220
aatgagaaaag	aggaaaggga	gaaagaaatg	agagagagaa	aagaaaagac	agaatgcgag	5280

agaggggagga	agagagaaaa	aggaaaaagag	agaggaaaagg	aaaaaggaaa	atgaggaaaag	5340
cgagaaagaa	gaaatgagaa	agaggaaaagg	gagaaagaaa	tgagagagag	aaaagaaaag	5400
acagaatgcg	agagaggggag	gaagagagaa	aaaggaaaaag	agagaggaag	ggaaaaagga	5460
aatgaggaa	agcgagaaaag	aagaaatgag	aaagaggaaa	gggagaaaaga	aatgagagag	5520
agaaaagaaa	agacagaatg	cgagagaggg	aggaagagag	aaaaaggaaa	agagagagga	5580
aggggaaaaag	gaaaatgag	aaagcgagaa	agaagaaatg	agaaagagga	aagggagaaa	5640
gaaatgagag	agagaaaaaga	aaagacagaa	tgcgagagag	ggaggaagag	agaaaaagga	5700
aaagagagag	gaagggaaaa	tggaaaatga	ggaaaagcgag	aaagaagaaa	tgagaaagag	5760
gaaagggaga	aagaaatgag	cgagataaaa	gacagaat	gagagagggg	ggaagaaata	5820
ggaaaagaga	ggaaaaggatg	gagaaaagag	agaaagaaaag	agagatgaaa	gagagaaaag	5880
agaaatgaaa	tgagagagag	agagagacac	aaagagccag	agagagaaga	aaaaagggga	5940
aagagaaaaga	gaaagaggaa	ggctcctcct	ggacacatct	tcctttatct	ttccttgggg	6000
accgccaaaag	cctgggtggca	tactgtacat	tctgtacact	gttcattcaa	aacaggctct	6060
gtcttaaaga	tggctctgagc	ggtcagaaaa	gggtattggt	aacttggttg	caaaactgcc	6120
tcaggagagt	gctgagtgcg	tgaaagtgtg	tgcccgtaa	ggagaagtct	ctactacttg	6180
tgatctcacc	atcgaaaatt	tccttaattg	tctcctgggt	ttctgggttt	tgcatgtttg	6240
tttctaagga	tacattcctg	ggtgatgtca	caaagtcccc	aaagacacgg	tggagctgtg	6300
ttagatgggg	aaagacagtc	tgctgaggat	ttatctggaa	ctgtcagaag	gaaaagaagg	6360
taaatggggc	acttgggaaa	gtggcctcta	gtttgacttc	tggcttagca	aaggttgttg	6420
ggagataaag	catacacagt	agttagcagg	aggcaacagg	gtcctgggag	gacgcgaggc	6480
agaaggagag	gctgggctga	cagcatgcaa	tcattgcata	gtctccaaag	gagattgcaa	6540
catggctgag	ttttcagagg	tcctacagag	cccgtggtag	agattctgtg	ggttctgaga	6600
caacttgact	ttagccagat	ggtatattgag	taatctggga	gagagaaaac	agctacagca	6660
aacagggcca	catttagtga	cgaaactctc	actttgactg	ttgagtcatt	tgcatggggc	6720
cctgaggtca	ggctggccct	cagctcaaaa	acaagcgagg	aactgaagca	attactcaga	6780
taatccacag	ccacagccac	tggaaaaggc	cacatcccca	gagacagcac	agcaggggtg	6840
ggggtggggc	tatgagaaaag	ttagtgattg	tagcagttat	ctagaatgtg	cggagcagag	6900
gaggttacac	aaaaacctag	aatgtcattc	aatgtgggaa	accgagaggc	tcccaagccc	6960
taaaaggaac	agtttgcttt	cagccaaaaa	ggaaataaaa	tttggggcct	aaatctggca	7020
aatgattcag	acctctgtg	taggtgtcct	taaatgcaca	gcagattgat	tttcatgttg	7080
gagtttattt	gaactaaaag	acagaaatgg	tgaaaagcac	acctgaagaa	attgagatgc	7140
tatgaataaa	atcatttact	tacagctatc	acttaattag	tacctccttc	caccttgctg	7200
atttattggg	ctagtcaagg	aagaaaagat	cttccctcct	ccttctctcc	tcctccccct	7260
cctctcctcc	tccctccccc	tccttgacct	tcctctcctc	ccttctctcc	ctccccctct	7320
tcttctcttc	acccctcctc	cccctcccct	cctctgtact	cctccccctt	cctcccaatc	7380
tcttttttct	cccccttctt	ctctttctcc	cccctcctct	tcctcctctc	tcctccctcc	7440
ctccctcctc	ctcctcatcc	tcctcttctc	cttcactcct	ttctccttcc	tcctctcctc	7500
cctcctcctt	ttccagccct	acctaccttc	cctttcttct	tcattttattc	aaagtagctt	7560
tgaacagcac	tactcggttt	agttgtgtat	aaaaggaaaa	tgcaggtcca	agcagcttgg	7620
ggaagattgc	tttttgctct	ctggaggcag	atgatgacag	ttcaagatca	ttccttttgc	7680
tccatgtcac	aggaaggggg	acatgccgaa	tctaccagtt	tgcagccacc	tacacaggat	7740
ccaccttcac	ttctaaggaa	atgtttggga	agctacctac	caaccacttc	tggcatctca	7800
tgggctagag	gactcttaaa	tggcaactct	atttgtttta	taaaggagg	tgtgacgtgt	7860
agtttttaaat	cccttcacaca	caacaattgc	tactctctga	ccaaaaaaga	agggagacag	7920
gatacggtca	ggtgtctagt	agactttacc	actttgaaaa	gccttaatat	aaatcaggta	7980
gatacatctt	tttaacttat	tcttgtaaa	acaaaaacaa	aactttattt	ttattttgtg	8040
gtatgcttgt	gtgtgtgtgc	ctgtgtgtat	accacatgtc	gctgggtgcc	gagaacacca	8100
gaagagggga	cctgatctcc	tggagctaaa	gctatccatg	gttctgagct	gcctgatgtg	8160
ggtgctggga	acagaactct	ggtcttctgc	aagagcaaca	agcctcctct	taactacgaa	8220
tctcctcccc	atccccccaa	atacatttaa	ttattcattt	tagcagcttt	atctcgtaac	8280
tacttatcac	agcataaaaac	aaggatttta	tatatattac	atgcaatcga	ggataagagt	8340
tgaggggaga	tgcgtgtgct	ccttctgggt	gtctgtgctt	ttgaagaatg	taagcagtgc	8400
acaagggacc	gaggcggtgc	tgtctgccag	gagctgtctt	cttcccttgg	actctgagct	8460
gagtgcagtg	ctccgaagaa	gtaaaagacg	acctcatgaa	gcaatgtctt	caaccctaac	8520
atgctgtcca	gacaaagtcc	agcttcatta	gtgctctgag	gagagactta	ctgagcctca	8580
ggaaagcccc	cctcagcatg	gcgaaagtcc	actttgattg	aagtgactcg	aaagccatgg	8640
cagtgcggcg	gcggccgcgt	ggagcttgtg	ctcgagtcgg	aagcggcatc	tttgtcaggc	8700
ggctgtgatt	agcacgggga	ggcaggactg	gagtgaagga	agagttgggg	gcggggctta	8760

gcgctctggt	ctcctaagct	gtagtcagcg	cctcaagatt	tgtaacctgc	cttctgcctt	8820
cccagccagg	cagtcaagtg	gctccaagct	gaagactgca	aagtgccctt	aaccttttgg	8880
ttatagcgag	gctgaagaca	ccgtgctctt	tcatgaaagc	cggatgtctg	aaatccgatt	8940
tgataaatat	ggataaaacg	tataacgctc	gatcaatcga	atcgaaggag	ctcacgattg	9000
gcaccacggc	tttggggaca	acagagtact	gactcggttg	gaggacttgg	atacttcccc	9060
tcctcttcca	tctcttcccc	tttccctcact	tcctcctcct	tccttctcca	ttttctccct	9120
cttcactggt	tcttactatt	tttacaaaag	atttttattta	tttattttatt	tattttattta	9180
tttattttatt	tattttattta	tttattttaat	gtatgcgagt	acactgtagc	tgtcttcaga	9240
cacaccagaa	gagggcgctca	agttccatta	gagatggttt	cgagccacca	tgtgggttgc	9300
ggggcctctg	gaaggaccgc	cagtgcctct	aacccctgag	ccattttctcc	agtacccttc	9360
tcaccgtttc	tcttcaatct	tcttccctct	ccttctccac	tttcccttgc	ttcttgggtt	9420
cattactctt	ctccctttct	tcctcttctc	cccttcttcc	tcctccactg	tagttttctc	9480
tcctactctt	tcttctgcct	ccctcctcct	ccctctcat	tcctccctct	cttctctcct	9540
tctccctcct	cctccttccct	tctccctctc	ccctctcccc	tctcccttct	ccctctctcc	9600
cctcctcttc	ctcttttctcc	ttctccaccc	ctcctgtcac	agtatcaatg	gcaagggtgt	9660
tctagaatgg	aggagtgtcc	cctaggcact	aacgaaagcc	agttaggatg	ctctgagacg	9720
ggtacaattc	agggagggcc	gtggggatgg	aagggttgtg	ctgcgattca	ttctggagca	9780
acccccaggc	agaatcatga	ggttgggttcc	ggattcgcag	ggcacaattc	agaagaggaa	9840
ggtttcagga	aggacgagtt	tgtctgagat	aggagttaca	tctgatgtct	tggcagcaga	9900
gccactgtac	aagcgtgctt	tattaaccac	gtgggattaa	atcttctttt	aaattttatt	9960
tcaactctta	aggaaacgtg	aactttcaca	ttcaaattta	gacttgcagc	tcttatgggg	10020
aaaaaaaggg	gatcttaaga	atattaagca	taggcggctg	gagagatggc	tcagcgggta	10080
agagcactct	ctgctctccc	agaggctcct	agttcaattc	ctagcaacca	cataatagtt	10140
aacaacagtc	tttaaatgaat	tctaattgcc	tcttctgggt	tgtctgaaga	cagttacagt	10200
gtactcatat	aaataaaata	aagaaattta	aaaaaatgaa	tattaggcat	agattcctgg	10260
atcctaagaa	agccatcaga	gctggagcca	tgtgtgggat	cctgcttggg	gctggagggg	10320
cagagttcat	gccccggggg	tttttactta	ttatcacatt	ttcatcggtg	ttttgaaaca	10380
gggtcttgtg	tgggtccaggc	tggccttgaa	ctcatcttcc	agcctctacc	tcacagggtc	10440
tgggattact	tggttcctaa	aagtatctcc	gtcaagctcc	ctggtgttat	ggctgtgcca	10500
accagagagg	tctatacact	cgctcaggtg	gagggagaag	atccgaatct	ctgacaggga	10560
ctgctgcctc	tcggggcaaa	tggagtgaag	gacagcggca	gaaggattta	ggaaagatgg	10620
acgggagagt	ggaaatgctg	cagaagccag	aaaacaaagc	aggaagcctg	ctgtccagtg	10680
gggctcaaga	gcgaggggat	gcgagggggc	tgcgcaggaa	catttagcgt	ctgcgtctat	10740
gggggtaggg	gcggggtgcc	agcacctagt	cacctgaagg	ggaaatgctt	gcccagggag	10800
cagggtctcag	tagctgacct	agagaaagga	gcggccctca	cagaggagac	acgggtcact	10860
gtttgttaaa	gtgaaggaga	aataaatatt	ctttcaaaga	atcttaggtg	agcccagttc	10920
atctgcgctg	tggaggcctg	gggaacagtt	aaaaagaccc	tgacacacac	ccaaggcaaa	10980
caagcaacac	acggctcctt	ccgtaagggt	ccatgattct	ctgaagaatc	agccccggaa	11040
tcagccccgg	aatcaggtag	tccgtaaaca	caatgagtgt	ttactctgc	agaagtccag	11100
cctgctggcg	tctcccatta	ccaaaataga	gggatagtca	cgtgagctca	ccggctcgat	11160
ttaaggcacg	tggttttcca	gggtagatga	gctttggctt	ctggaaccat	tatggggcac	11220
gaaggatgga	gccaggattt	tttttttttt	tttttttttc	tattagcaat	tgattttgctt	11280
gggcttggct	ggacttgccc	agttcttagg	cccagtcttc	ttaaactgcc	atctgaagtc	11340
tgtcatggag	tcagcctagc	cttctcactt	cccttcagct	cgaataggaa	gaggaggtgc	11400
acaccagatg	gtctgagagc	agggataaat	ggtgtgcctt	tgtctttcag	tatttcgtta	11460
ttttaagtag	gaagatgctt	ttctgtatta	cattgcttgt	gaaaccggaa	gttgattcgg	11520
ggcacaggac	aatggatttt	gtgttttgca	aggactgttt	cagaagagag	aggagtggaa	11580
gggtggttag	agtgaggagt	ggggtgggac	ggagtggggg	aagagaagga	agggccagac	11640
aggctaggta	gggctgagag	gaggcgggtg	gaacttcttg	agttagcgca	gcagtaaact	11700
tggatgtgcg	tgtatctttg	tgatatatga	cccggagccg	tgtagctggc	tccgatagta	11760
ctgctaattg	cagtgtcggg	gggggggggt	cccatactgt	tccacagggg	ctgcacattc	11820
ccatcgagag	caggagggct	cctctctcca	tacatcctcg	ccagcattcc	ttgtttgttt	11880
tgtgatgaca	gggggtggga	tgaaatctct	ctgttgggtt	gagagaccgt	gaagaagctc	11940
aaccccagga	catttttgag	tcttggaagg	cagtgcctcc	atgtggagcc	gtggagccca	12000
tctctgagtc	caggtcactc	ttgcagttcg	cactcagctc	ttcagatgca	ggagagacgt	12060
tgggtgggaaa	gcaagattgt	ttgcttggtg	agatagacac	attctccaca	caaaggctca	12120
cgtggggcaa	aggctgattg	acgtacagcg	ttcaggaacg	cctgtggtag	agctatgatt	12180
agctgtctcc	atctatgaag	cagacaaaga	gttataaaaa	aaatcaatgt	tttcaaattg	12240

tcaaaactttt	aacccgacag	caagcgctct	gtccctgggc	taatccctag	ccctgggttc	12300
ttgagatggg	gtcttttgtg	cactagactg	gcctagaact	cacgatctta	gtgttccagc	12360
ctcccagctg	ctgggatgag	ccgctataac	cagtctgcct	gccttcctaa	attttaagtg	12420
atgggaagtg	ggggagaata	cagtttaaag	tatgcagatc	tgagagcagg	aacctggcaa	12480
agccaagggg	ccggagttac	aggcggctaa	catgggtgct	gggaactgac	ccaggtcctt	12540
gagaggagca	gtgtgtactc	ttgaccaaac	aggtccgtct	ctccagtccc	cgtagtatta	12600
aaaataggta	ctacgggcat	ggtggtgcac	acctttaatc	ccagcactag	ggaggcagag	12660
gcaggtggat	ttctgagttt	gaggccagcc	tggtctacaa	aatgagttcc	aggacagcca	12720
cggctatata	gagaaaccct	gtcttgaaaa	caaaacaaca	acaaaatagg	tactacaaag	12780
cgatgtaatt	gtgctcaaac	atgcaaaccg	aggggactgt	atgcataaga	aagagaaaag	12840
cggccacact	ggttctatct	gggtgacagg	aaatcagtat	ttttattttt	cacattcatt	12900
tttttgttgt	tggtgttgac	acagtgattt	ttctatcaaa	aacattattt	cttttatagt	12960
tccctgagg	agctgttttt	aaagccgtgc	tttgaaaaac	cattgaagga	gcagaggcag	13020
ggagactcct	gtgtggcagt	cgggtgaagca	ggccctctgc	aggcaggctg	gccctggact	13080
tgggagtctc	tttccctccc	tcctgtgctc	aaatagcaaa	tgtcaggctt	caatgtagct	13140
agaaggttct	agaatgatta	agtttccaag	gctgaagagc	ttccctgttt	gcctttcact	13200
tccctggaga	ggtcgttgtg	tgttccggag	tctgcaaggt	gcctttggtg	atgcgggtgg	13260
ttcatctcgg	gagattccgc	ctggaggacc	caagttcaag	ccctgcctga	gctacagagt	13320
gactttcagg	tcttctgcgc	aattcagtga	gacccagtct	acaaataaaa	agtaaaaaga	13380
aggctgtgga	tggaaactcg	tggtagagtt	ctgggtttac	tccctagagg	aggggagaag	13440
gaggaggagg	gaggagggaag	aggaagaaaag	aagaagagaa	gggaagagga	gaagggaagg	13500
aggggaaggg	ctgacaagaa	gagagaagag	ggaggaggag	gagggaaagg	aaggggaaag	13560
gaaggaggag	aaggggctga	caagaagaga	gaagagggag	ggaggggagg	gaaagggaagg	13620
ggaaagaaga	gaagggttaag	aagaaactgt	tccaatggtc	tggggccacag	agtgatggcc	13680
ttttgtggtg	atcagctgta	atccttgatt	tgacacaacc	tagaatctgg	gaagcgagtt	13740
tctgtgaagg	agcatttcaca	ctggctggcc	tgtgggcgtg	catgtgggag	actgtcataa	13800
ttaggttcat	taatacacgga	agtcccagcc	cactacaaat	ggcttcgttc	catacccaag	13860
agatgctaac	tgtagacggg	tggagaaagc	aagcaagctg	tggatacccc	acgctctttc	13920
acctcggtc	ctgggggggtg	ggtgcactgt	gtctcttggt	attttaaaagt	cctgccttga	13980
cgtccctgct	gtgacagact	gtaactggaa	ttgtgagctt	tagtccttta	gttttctacg	14040
ttggtttttc	tcaggatatt	ttatcgcagt	aacagaaaaca	agaccaggac	acttgatctc	14100
ctctgatcaa	cactgaagag	ttacaaaaca	ggctgaggaa	acaaactttc	ttctccctct	14160
cccccttctg	tccctccctt	tccttctcgc	tccctccctt	gccccctctc	tccctgtctc	14220
tgtctctgtc	tctgtctctg	tctctgtctc	tgtctctgcc	tctccccctc	cctccccctc	14280
ctctgtctct	gtctctgtct	ctgtctctgt	ctctgtctct	gtctctgtcc	ctttctcctc	14340
tatctcctaa	atggctggag	gccatgctag	ctcaatggtg	aactttgaac	acgtatttag	14400
gaaatctttg	ttcttaacag	ttctgaagtg	ctgaagtggg	ggtttagtct	ctcggcctga	14460
caagctcact	tcctctcact	ctgtcttaat	gaccaaactc	gccatttccc	taaaacagca	14520
caggtccag	ctccagggtg	ctccggagcg	gag			14553

EXAMPLE 15 – CHO Stable Site 2 Sequences – U.S. Patent No. 9,816,110

<211> 4001

<212> DNA

<213> *Cricetulus griseus*

<400> 1

(SEQ ID NO:58)

ccaagatgcc	catcaactga	ttaatagatg	ataaaattat	tgtacatttc	agtgtaatat	60
tattcagttt	ttaagaaaaa	tgaaattatg	taataagcat	gtaaatggat	atatcttgaa	120
acaaccattc	cccattatat	tacctaaaca	ttgaaagtcc	aaaatcatat	gatcttttta	180
gtggatctac	taatcttttg	ctatatgtat	tttattgaac	tacccatgga	tgtgagataa	240
ttggttaacaa	cagcacatgg	gagagcatgg	gatcattcaa	ggaagattag	agagaatgca	300

tttttttagga	gataatggag	gagcaataga	aaggattaaa	tgaggttact	gatgaaagtg	360
atgggttagag	aaggcaatat	gaggagggat	aactagcact	tagggccttt	tgaaaaagac	420
atagagaaaa	tactattgta	gaaacttcct	ataattgggtg	tatagttata	tacaccaaaag	480
agctcagatg	gagttaccct	ataatggaaa	tattaactac	tttttatcac	tgtgataaaa	540
catcctgaac	agagcaacat	agattgggaa	gcattttactt	tggtttacag	ttctaacggg	600
ataaaaaattc	atgatgaaa	aatgaatatg	tcagcaaaca	gcagtagcaa	tggcctgaga	660
agcagggtgag	agctcacatc	ttgaagtgtg	agaatgtagc	agagagaaca	aactgcaa	720
gaccagaaaa	tgtttttgga	tcagagccca	taccctctg	actgacttct	ccagaaattc	780
tgaacaaata	aaactcccca	aacagagcca	taactgaagg	tccagtgtct	gagactacta	840
ggggtatttc	ttattcaaac	cactacaatg	gggtgggggg	agcaatcctc	caagtaggca	900
ctacacacag	acaaataaaa	actctagtaa	ctggaatgga	ttgacttatt	tgaattactt	960
gccagtggag	ctacatagag	cacaattatt	gtattttaat	taccctttat	gatcttacaa	1020
aacttgacag	taagatcata	ttgctaaaga	aaccacatat	ttgaatcagg	gaacatgggtg	1080
atatctagtt	gttcttcaac	tggaaacttc	atgctttctg	cccagcattc	atgttgctgg	1140
aaagagcaat	gtacactacc	agtgtagaaa	ttaaatcatc	aatcttatca	agatgtggat	1200
cctataagtt	acaataaaaa	ttagcctgat	aagatatccc	caccagaaga	atattcacat	1260
aatgctatg	ggagcaacaa	gctattttct	aaattagctt	taatcctatt	ctacaagaga	1320
gaatccatat	ctagaatagt	tatagggatc	aagaacccat	ggcttgattg	gtcataggcc	1380
caatgggaga	tcctaataatt	attgttctac	aaaatgaaaa	taactcctaa	tgacttggtg	1440
ctgcagtaat	aagttagtat	gttgtctcaac	tctcacaaga	gaagttttgt	cttacaataa	1500
atggcaatta	aagcagcccc	acaagattta	tatcataccg	atctcctcat	ggcctatgca	1560
tctagaagct	aggaaacaaa	gaggacccta	agagagacat	acatgggtccc	cctggagaag	1620
gggaaggggg	caagacctcc	aaagctaatt	gggagcatgg	gggaggggag	agggagttag	1680
aagaaagaga	aggggataaa	aggagggaga	ggaggacaag	agagagaagg	aagatctagt	1740
caagagaaga	tagaggagag	caagaaaaga	gataccatag	tagagggagc	cttgatggtt	1800
taaatagaaa	actggcacta	gggaattgtc	caaagatcca	caaggtccaa	ctaataatct	1860
aagcaatagt	cgagaggcta	ccttaaaaagc	ctttctctga	taatgagatt	gatgactacc	1920
ttatatacca	tcctagagcc	ttcatccagt	agctgatgga	agcagaagca	gacatctaca	1980
gctaaacact	gagctagtgtg	cagacaggga	ggagtgatga	gcaaagtcaa	gaccaggctg	2040
gagaaacaca	cagaaacagc	agacctgaaa	aaaatgttgc	acatggacct	cagctgata	2100
gctgggagtc	cagcatagga	cttttctaga	aaccttgaat	gaggatatca	gtttggagggt	2160
ctgggttaatc	tatggggaca	ctggtagtgg	atcaatattt	atccctagtt	catgactgga	2220
atltgggtac	ccattccaca	tggaggaatt	ctctgtcagc	ctagacacat	gggggagggtt	2280
ctaggtcctg	ctccaaataa	tgtgttagac	tttgaagaac	tcccttgaga	agactcacc	2340
tccttgggga	gcagaaagg	gatgggatga	gggttgggtga	gggacaggag	aggaggggag	2400
ggtgagggaa	ctgggattga	caagtaaatg	atgcttggtt	ctaattttaa	tgaataaagg	2460
aaaagtaaaa	gaagaaaaga	aaacaggcca	aaagattata	aaagacagag	gtggtgggtg	2520
actataaaga	aacactatta	tctaaataaa	aacatgtcag	aagcacacat	gaacttatag	2580
tgtttatgaa	agtatgtata	ataactacat	aatctcaagc	caagaaaaaa	atatcatctt	2640
tcagtgatga	aggtgatttt	atttctccca	gaattaaagc	caaagaccta	atgaaagtaa	2700
ttatcttcaa	aaggttgaaa	atacatactt	tgcaatacac	agatctgcct	agaaatctca	2760
tgttcacaat	acacatgatg	ctcaattgaa	ttccattcaa	tgttacagtt	tagataaaca	2820
gtttgtagat	aaactcacia	tgtatcattt	ctttttattt	tttgaccaa	cagcttctca	2880
tctgttattc	agaataattc	ctcgatggca	ggatatccat	cccaattggg	ggaaggggag	2940
aatttgaaga	aaacctagac	cacatacata	tttgccattg	ggaaacaaag	tctaaaatga	3000
tgttgttcac	atcttctcta	ctagtcctct	ccccgtccca	agaaccttg	gtatatgtgc	3060
ctcatttttac	agagagagga	aagcagggaac	tgagcatccc	ttacttgcca	tcctcaacct	3120
aaaatttgca	tcattgctca	gctctgcct	tctcatatga	cagttacaag	tcaaggcttc	3180
caaagtccct	ctgtcatggt	tggtgtcaat	agttttataca	gatgacttca	tgtcttcata	3240
tctaattgtct	tatatagatt	aatattaaac	aatgttattt	ctctaaccac	atltttaaatt	3300
aatttaaaaa	tccattaatt	gtgtctataa	aatgcagaca	gagtgtgag	acacaatata	3360
agcctgatga	tctgaatttg	aaactcacac	ccaccacatg	gagaatcaac	ttccaaaaat	3420
tttccattta	cttccacact	tacaccattg	tacaaacaca	ataataatga	acaaaatgaa	3480
atgaaataaa	aaattaagtc	tctgtaggta	atgctactgt	gcagcaaaag	taaaaatggc	3540
agcttaagct	tgttttatgg	ttacacttta	ccatcttcca	tttaattata	ggacttcaat	3600
catggcagaa	ctatgctggt	attgtctcag	tgtaacctaa	ccaggtgttc	cagatgttct	3660
taatgtggac	acctaaacta	tttgatattt	gggttaagat	ctttccctct	ttcagaagaa	3720
acctcaggac	agaggggaatc	ttgtctttta	atltttgagtc	tgtagacttt	ttccatttca	3780

aatatacatg	aaacaagtga	tgaagaaaat	taatcaaaag	gtgggaattg	caatgatatt	3840
aggttcaata	ttaagcttca	atattatcat	ggaatcgctt	gttatacact	gagtgtttgg	3900
caataaggga	tttttagaag	aaggagtttt	tattctcaac	aggttcctta	agtttagctc	3960
aaataaatct	aagcaatcca	ctctagaatt	aaatagtttc	c		4001

<211> 14931

<212> DNA

<213> Cricetulus griseus

<220>

<221> misc_feature

<222> (2176)..(2239)

<223> n is a, c, g, t or nucleotide is missing

<400> 4

(SEQ ID NO:59)

catgtacact	tatgcaagta	tgatatggcc	caacacagta	ttttacacca	atTTTTatct	60
ataaaatata	catgtacatc	aaaatatatt	attaataata	acatcattat	tctttctttc	120
caagtaataa	acacatacac	tgaaatTTtg	gttcttggtg	ataatTTTaa	tgaaacagga	180
aatgcaaatt	tatcttagca	tgtttacttc	actttctttg	catagataac	cagtaatcac	240
attgatggat	catgtagtga	aatgtatttt	taggtatcta	aggaatTTtg	gcttcgTTTT	300
gtgcttggtg	acactgaatt	ctattcctaa	caacagtggt	taaggattct	gtctgatttc	360
ttttaccagt	atttgtccat	ttgcattttc	tttattattc	atggctgctg	ttctagaaaag	420
tggaaggtag	tgtgtcaagt	ctgtttaaca	tgtttccctg	atgatcagtg	tcttaacacc	480
tctctgagta	catgttggcc	aatgtcgttt	ctagaccctt	ctattcttgc	ttgactttatc	540
ctggtacatg	cctgccaaga	aattttctct	catcctttct	gtctcttcac	tgattttactt	600
gatgtgtgga	tttcacattg	atcatatgga	aatagaagat	acaatTTTct	ttattccacag	660
tttggaagac	tttcaatctc	atagatcatc	attattTTTT	gctactgttc	cctatgctat	720
ggtgaaattt	ccattttgaat	aattgcttaa	acaatttaaca	agaaagaatc	tattttttact	780
tgcaataaact	tccattttcag	aacattttact	acactgttac	tatatccaaa	aactagtTTT	840
atatatcatg	tgagaaatga	ctaattcata	atttggccat	gacattTTTT	tcagaaacag	900
aaaaagtgac	caatacacat	acaatgctat	aaatattaag	acttcagcaa	attaaatatt	960
tattcatgat	atcacataaa	attcattttat	tatgtTTTTat	ttaaatgtgt	ttttaaaaca	1020
gtgggtatcac	taaatattaa	gttagatgtg	tttatgtgct	taatgaattt	atatttttaga	1080
atgtttataag	ttgtatatag	tcaaatatgt	aataaatttt	atTTTTtagg	tctttctcat	1140
taagggtattt	taattttggg	tcccttttcc	agagtgactc	tagctcatga	tgagttgaca	1200
taaaaactaa	acagtacaaa	atgtacattg	cattcagtat	tgcacttgat	ctttgcactg	1260
aagtttgagt	cagttcatac	atttagtact	tgggaagtac	attaagctaa	ctttcattgc	1320
tctggcaaaa	tgctcgataa	gataagagtc	tattgtggaa	agccatggca	gcaggaaagt	1380
aagactgctg	atgatgttta	atccatagtc	aagacgcaga	aggagatgaa	tgctggtatc	1440
caacattttt	tgctgttcat	tttctctaga	accctagtcc	ataaagatgt	atgacttgca	1500
ttcaaaaatgc	gtccccttca	gttgttcaac	ttttctgtaa	atatcctttc	aggcatgtct	1560
agaagattgt	ttcgcaaata	cttctcaatc	cattcaagtt	gatagtgcag	attaatcact	1620
gcagaataaa	agcctgtaac	ttggctcacg	tgccaaggaa	tatgcacact	cctgacacat	1680
caataagtaa	atcaaagtgt	agcttttgcc	tttaacattg	ccagacttat	gtaatgttct	1740
gcacgttctt	cctccatcac	tttttattct	aatggtgttt	ccttgacatt	gaatcacgct	1800
gtggaagctg	cttagaatta	acattgaaat	ctactgatat	atztatgatg	cagcaattta	1860
gatttactat	tttacttaga	atTTTTtata	attgagagaa	tataatatatt	tcacagttat	1920
ctatctgctg	taaatagagg	atTTTTaaaaa	aatctctat	aactTTTTTT	tacaacacac	1980
agtaaaaatta	agttaaaatt	taataaagtc	actatgttga	tttcaaagtg	tgctacgccc	2040
acgggtggtca	cgcagggtgta	gcagaagatg	ccactaaggt	gggctaaggc	cgatgggttg	2100
gggtctgcgc	tccctggaga	tgagccccag	gcggttccct	ggcaatcagc	tgcgatcatg	2160
atgccccgatg	agccannnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	2220
nnnnnnnnnn	nnnnnnnnnc	tgggtgactt	tatggaaaga	atttgataga	tttcatgatg	2280

tagaagaatt	ttatttaggt	tattttacag	gagactaaga	ccctgggacc	taaagatatc	2340
tgggtcctga	gaatcaggaa	atgggtagag	acgtggttga	tggtagaga	cagatttttag	2400
agaactctta	gatcatgggc	aatgaccgca	atctgatgct	tagaatagat	catctataaa	2460
caattatgct	gttctttttc	tttctgttgt	atgatctgat	gatgtagccc	ccttgccaag	2520
ttccctgac	ccccttgcca	agtccctga	ttgtaacagt	atataagcat	tgcttgagag	2580
catattcaac	tacattgagt	gtgtctgtct	gtcatttcct	cgccgattcc	tgatttctcc	2640
ttgagccttt	tcccttggtc	tccctcggtc	ggtggtctcc	acgagaggcg	gtccgtggca	2700
aaagtgtata	aatgttctaa	aacatttgaa	ctctaaaaca	tgcaaaatga	aaaattaaaa	2760
taaaataaca	tgaaaattaa	aatatattag	ctgctaaaag	ttaaacaata	ctatataata	2820
ttttgttatt	agaattcaaa	atcacattag	ttggatttaa	tttgaacatt	gcattctttc	2880
aataataatt	tcaataaaaa	aagtttcccc	atgatatag	aaaataataa	catatgtatc	2940
tatctattta	tttaactaca	catatatagc	atttgtttca	actaaaataa	atgaatgagc	3000
aaagcaccta	agtaattggg	gtctattata	tttatgaagc	caatagtttc	aaataaatta	3060
tcatgcataa	ggaggtattg	caaagtgtta	accttttttg	aaacagatat	tcccagttac	3120
agaaattata	atttctaata	tttctataaa	gtagaatgat	gataaattaat	ataggccatt	3180
tgtaataaat	gttcagatta	aaatattctc	tatttcacta	gagaagaatg	atattaaatg	3240
tattatattt	tatttcccat	tttgtttgca	ccactattct	atatccctca	gcagtttaaa	3300
tttgtttcac	catatgtgtg	tgtgtttgta	tcttaaatat	ggcactaaaa	ttagaataat	3360
ttaatataaa	tctttaggag	aaaagatatt	gaattatttt	atgttgatag	gaaaatatct	3420
tttaattgtc	caagaatact	ttttcttcta	ttttaggact	gatcagaccc	aggactaata	3480
ttttatatgt	actaattcta	tgtacaaaa	tatgttatta	tctcatgaat	tctgtctcaa	3540
tattgaggta	ataaaaatag	tccatcatga	actttaaaat	taaaataatg	attaattaat	3600
ttttattcat	attttgtttg	tatgaatggg	tatacatcac	atgtgtgcct	ggtgactgtg	3660
aatgtcagga	gaaggtatga	aagccactgg	aattggaata	agagataata	tttgagatgt	3720
tatgtgggtg	ctgagaatta	gacgcaagcc	atcttcaaga	atagccagca	tactatacca	3780
ctgagtaatc	cattcatccc	tcaataatta	tctttgtaga	cagtaaatat	atttctaaac	3840
tataaatgac	cagaaaaatt	aatgtattat	taatgaagac	attcatctca	tgtgacacac	3900
ttcacctgtc	taaatcagta	acactctctc	cactaattaa	gatttttctaa	gtgcatgaca	3960
tctactattt	ctaaagctgt	ccaatggggg	ccagtcacca	gtcagcacc	agtgcagataa	4020
tccatgaatg	catttatatac	ttaggaaaaa	ttcttatcta	tgtagtattt	agaacatttt	4080
catgtgaggg	gataaacaag	gaagcacaga	tgctttctga	tagaaaacttt	ctctttaatt	4140
catctagaaa	aaaaaaacct	ctcaggaaaa	tctctcttgc	tctcctccca	atgctctatt	4200
cagcatcttc	tccctactta	attctagatc	tttttctcta	tgccctcctg	ctgctgccct	4260
gotggotctg	ctctatgcct	ccccatgtca	cttttctttg	ctatctcacc	gttaccttct	4320
ctgcctcact	ctctgccttc	ttctctgctt	ctcacatggc	caggctctgg	acaattatag	4380
ttatatgtta	cattctcata	acacatgata	tgtcacatag	tttctctcag	gctagggata	4440
tcacaatgac	tggccaatga	gcaagtggcc	ttgcatgtag	ctctaagttg	gtgatggttc	4500
ccagacagta	agtagccatt	tgggtgaaat	ttgaggttgg	gtagtacatg	aagactgaat	4560
tttcttcaaa	ctctggcctt	gaaatagtaa	aacaacacct	atgaaaatga	cgacctgtat	4620
ttgtcttttag	aggcaaccac	atattgtctg	cagggcctgc	tttgaatttg	ctctgaagtt	4680
agcttgtttg	tgtaaaagga	agaatcctat	atcagcctga	gaaatgtaaa	atatacctagc	4740
atttcaagtc	atcaaaatta	tatggagagt	ataaatcatc	cttctgacta	ttcatagtca	4800
tatttggtgc	caccaagtat	aaaacacact	accaaagggc	tgtggaaaaa	atcgccataa	4860
ctgttcttat	tagggaggca	tagcagtggg	acctgaggaa	gttacagcaa	caaccagtca	4920
tccagtcatt	aaccccatgg	ctttgccact	tggaggtacc	caataatggt	tggctttgcc	4980
gagtaggact	ccaacaaatt	cagagggtca	atttttaaat	gctggttgtc	actgctgaac	5040
agtcocattg	cctctgcatt	aattccacaa	tggaaagctt	tttacctga	ttgccaatca	5100
ttaaacagcc	tactcagcat	aaacaggtat	gatattattc	tgcattttgt	tacattacta	5160
gatgaattcc	tatttcttcc	tacaatagt	gaactgaaaa	aagatacaca	atcatactac	5220
ccctctacta	atcttatgac	ttatatcatt	tcaattttca	gaccataatg	caaactattg	5280
acaaaaacat	gtgaagatga	aaaatagaaa	tgtagaataa	tattacatat	aaaaagaaaa	5340
ggcggactta	ttttgtttta	tttcttagca	tgcatagcaa	tacatgattt	gagggtttata	5400
taataaaagg	acaataaatc	ttcaagaaac	ttacccttac	tgaattaaaa	tattaaagaa	5460
ggtcacacat	ttactcaaat	atattagact	actgggcaaa	tagacatgaa	aagtagagtt	5520
aatattgagg	taggccttct	gtgaaatgtc	taaggaaatt	atgtttcata	cagtgtgtaa	5580
ccaagtggga	atcatatcag	aaagcagtca	aaagcttata	ttacaagtaa	cagatgcttg	5640
gttatatgac	ctcccagagc	ttgactgtct	atacacaaaa	agtggtgtta	ataaaactgt	5700
aatttggggt	atgttttttt	aaatggcttc	accaacatga	aaggaaggga	atgagcatgt	5760

catggatgct	tagagattat	gcttccagca	agaagaattg	agctttggct	cttattacag	5820
aaacatgaca	aggtgtgagt	tttattttatt	agaaattata	taatatTTTT	agctggggac	5880
taaaaaTTTT	attgaaacaa	acaggcaagg	gataggcatg	tactagaagc	aaaaatagga	5940
tgtcaatgct	gtaatgttat	tttttggacc	aaaatagtat	ttcctataga	aatgacaatg	6000
atccttaggtt	attattcttc	ataaagatga	caagttcaca	agatatccta	gttcattaaa	6060
atcgTTTTag	tcatTTaata	gagtgtgtgtg	atagattaca	caaaggaaaag	cacttacgat	6120
gagaaataat	gatatccaca	attatttttct	taattcttag	aaacatttcta	ttgttatatc	6180
tcaatctcag	aagccactta	ttgcttttatt	attgaaacat	atgaaattgt	aagttatata	6240
ttgtctatgg	tgacatttca	aagaacatgt	gacgtacagt	gtagcacaga	taaagaacat	6300
aactgcagct	gaatcagtaa	ctaaacttac	atacattaaa	tctgccatgt	tggcaacagt	6360
gtgtgcacta	ccaaaggatg	tactaatgct	cacgacactc	ccctatgtca	ccctttgttc	6420
atcattacat	cataggtcta	ttttgtttgc	ttttgaaatc	tagaccaagt	cttttgtgtc	6480
tttccaagca	cagagctcat	taatttacct	catagacttg	ttaaacttct	tctggttcat	6540
caattgaata	gaaatactca	ctactaatta	tgtgagacct	tgccagtacc	atagcacatg	6600
gataaTTTT	acataaaaaca	tgcatacaag	taagattatt	cagactgaac	atgaatttta	6660
gagaaatcag	gaaggagtat	atgggagtgg	ttggagttag	actagagaaa	tgtaatTTaa	6720
ctataatctc	aatacaaaaga	tctactaagc	aaaaaacatg	aaacattgtc	attcaagtga	6780
aacatcagtc	ttcaaattgg	aaagatatTT	ttactaggaa	aatgtctggt	agatggttat	6840
tatctagaaa	acacaaaaat	tagaaaacgg	taaactTTaa	taaaaaagaat	aatacaatga	6900
gactacatga	aaagtTctta	actaatgaaa	caaatatctt	gaaactTTTT	tcttaaaagt	6960
ttaatatcaa	taaccatcat	ggaaattcaa	attaaaaacta	tttacaatatt	acccctgaaa	7020
taataactaa	tacccaataa	aaataatata	aacaaaaaat	ggcaatgcat	gccatcatgg	7080
atTTgggaga	gagaatgttc	attgcagttc	tgaatggata	ctggtgccac	cacggtgaaa	7140
atctctgtat	aggtccttcc	aaaagctgaa	aatagacata	tcacaagacc	tgccacacat	7200
ttttcaagca	aatacccaaaa	ggactctacc	tgactgcaga	gacactttct	cataaaatat	7260
tattgtttgat	ctattcataa	tatctggaaa	atagaaacag	ccaagatgcc	catcaactga	7320
ttaatagatg	ataaaaattat	tgtacatttc	agtgtaatat	tattcagttt	ttaaagaaaa	7380
tgaaattatg	taataagcat	gtaaatggat	atatcttgaa	acaaccattc	cccattatat	7440
tacctaaaca	ttgaaagtcc	aaaatcatat	gatctTTTT	gtggatctac	taatcttttg	7500
ctatatgtat	tttattgaac	tacccatgga	tgtgagataa	ttggtaacaa	cagcacatgg	7560
gagagcatgg	gtatcttcaa	ggaagattag	agagaatgca	tttttttagga	gataatggag	7620
gagcaataga	aaggattaaa	tgaggttact	gatgaaagtg	atggtttagag	aaggcaatat	7680
gaggagggat	aactagcact	tagggccttt	tgaaaaagac	atagagaaaa	tactattgta	7740
gaaacttct	ataattgggtg	tatagttata	tacaccaaaag	agctcagatg	gagttaccct	7800
ataatggaaa	tattaactac	tttttatcac	tgtgataaaa	catcctgaac	agagcaacat	7860
agattgggaa	gcatttactt	tggcttacag	ttctaacggg	ataaaaaattc	atgatgaaag	7920
aatgaatatg	tcagcaaaca	gcagtagcaa	tggcctgaga	agcaggtgag	agctcacatc	7980
ttgaagtgtg	agaatgtagc	agagagaaca	aactgcaaat	gaccagaaaa	tgctttttgga	8040
tcagagccca	taccctctctg	actgacttct	ccagaaattc	tgaacaaata	aaactcccca	8100
aacagagcca	taactgaagg	tccagtgtct	gagactacta	ggggtatttc	ttattcaaac	8160
cactacaatg	gggtgggggg	agcaatcctc	caagtaggca	ctacacacag	acaaataaaa	8220
actctagtaa	ctggaatgga	ttgacttatt	tgaattactt	gccagtggag	ctacatagag	8280
cacaattatt	gtattttaaat	taccctttat	gatcttacia	aacttgacag	taagatcata	8340
ttgctaaaga	aaccacatat	ttgaatcagg	gaacatggtg	atatctagtt	gttcttcaac	8400
tggaaacttc	atgctttctg	cccagcattc	atgttgctgg	aaagagcaat	gtacactacc	8460
agtgtagaaa	ttaaatcatc	aatcttatca	agatgtggat	cctataagtt	acaataaaaa	8520
ttagcctgat	aagatatccc	caccagaaga	atattcacat	aaatgctatg	ggagcaacaa	8580
gctattttct	aaattagctt	taatcctatt	ctacaagaga	gaatccatat	ctagaatagt	8640
tatagggatc	aagaaccocat	ggcttgattg	gtcataggcc	caatgggaga	tcctaataatt	8700
attgttctac	aaaatgaaaa	taactcctaa	tgacttgttg	ctgcagtaat	aagttagtat	8760
gttgctcaac	tctcacaaaga	gaagttttgt	cttacaataa	atggcaatta	aagcagcccc	8820
acaagatttta	tatcataccg	atctcctcat	ggcctatgca	tctagaagct	aggaaacaaa	8880
gaggacccta	agagagacat	acatggtccc	cctggagaag	gggaaggggg	caagacctcc	8940
aaagctaatt	gggagcatgg	gggaggggag	agggagttag	aagaaagaga	aggggataaa	9000
aggagggaga	ggaggacaag	agagagaagg	aagatctagt	caagagaaga	tagaggagag	9060
caagaaaaga	gataccatag	tagagggagc	cttgtagttt	taaatagaaa	actggcacta	9120
gggaattgtc	caaagatcca	caaggtccaa	ctaataatct	aagcaatagt	cgagaggcta	9180
ccttaaaaagc	ctttctctgga	taatgagatt	gatgactacc	ttatatacca	tcctagagcc	9240

ttcatccagt	agctgatgga	agcagaagca	gacatctaca	gctaaacact	gagctagttg	9300
cagacagggga	ggagtgatga	gcaaagtcaa	gaccaggctg	gagaaacaca	cagaaacagc	9360
agacctgaaa	aaaatgttgc	acatggaccc	cagactgata	gctgggagtc	cagcatagga	9420
cttttctaga	aacctgaat	gaggatatca	gtttggaggt	ctggttaatc	tatggggaca	9480
ctggtagtgg	atcaatat	atccctagtt	catgactgga	atgtgggtac	ccattccaca	9540
tggaggaatt	ctctgtcagc	ctagacacat	gggggaggtt	ctaggtcctg	ctccaaataa	9600
tgtgttagac	tttgaagaac	tcccttgaga	agactcaccc	tccctgggga	gcagaaaggg	9660
gatgggatga	gggttggtga	gggacaggag	aggaggggag	ggtgagggaa	ctgggattga	9720
caagtaaatg	atgcttgttt	ctaatttaaa	tgaataaagg	aaaagtaaaa	gaagaaaaga	9780
aaacaggcca	aaagattata	aaagacagag	gtggtgggtg	actataaaga	aacactatta	9840
tctaaataaaa	aatatgtcag	aagcacacat	gaacttatag	tgtttatgaa	agtatgtata	9900
ataactacat	aatctcaagc	caagaaaaaa	atatcatctt	tcagtgatga	agggtgattt	9960
atattctccc	gaattaaagc	caaagacctt	atgaaagtaa	ttatcttcaa	aagggtgaaa	10020
atacatactt	tgcaatacac	agatctgcct	agaaatctca	tgttcacaat	acacatgatg	10080
ctcaattgaa	ttccattcaa	tgttacagtt	tagataaaca	gtttgtagat	aaactcacaa	10140
tgtatcattt	ctttttat	tttgaccaa	cagcttctca	tctgttattc	agaataattc	10200
ctcgatggca	ggatatccat	cccaattggg	ggaaggggag	aatttgaaga	aaacctagac	10260
cacatacata	tttgccattg	ggaaacaaag	tctaaaatga	tgttgttcac	atcttctcta	10320
ctagtccctc	ccccgtccca	aagaaccttg	gtatatgtgc	ctcattttac	agagagagga	10380
aagcaggaac	tgagcatccc	ttacttgcca	tcctcaaccc	aaaatttgca	tcattgctca	10440
gctctgccct	tctcatatga	cagttacaag	tcaaggcttc	caaagtccct	ctgtcatggt	10500
tgggtgtcaat	agtttataca	gatgacttca	tgtcttcata	tctaattgtc	tatatagatt	10560
aatattaaac	aatgttattt	ctctaaccac	attttaaat	aatttaaaaa	tccattaatt	10620
gtgtctataa	aatgcagaca	gagtgtctgag	acacaatata	agcctgatga	tctgaatttg	10680
aaactcacac	ccaccacatg	gagaatcaac	ttccaaaaat	tttccattta	cttccacact	10740
tacaccattg	tacaaacaca	ataataatga	acaaaatgaa	atgaaataaa	aaattaagtc	10800
tctgtaggta	atgctactgt	gcagcaaaag	taaaaatggc	agcttaagct	tgttttatgg	10860
ttacacttta	ccatcttcca	ttaattataa	ggacttcaat	catggcagaa	ctatgctggt	10920
attgtctcag	tgtaacctaa	ccaggtgttc	cagatgttct	taatgtggac	acctaaacta	10980
tttgatattt	gggttaagat	ctttccctct	ttcagaagaa	acctcaggac	agagggaaac	11040
ttgtctttta	attttgagtc	tgtagacttt	ttccatttca	aatatacatg	aaacaagtga	11100
tgaagaaaaat	taatcaaaag	gtgggaattg	caatgatatt	aggttcaata	ttaagcttca	11160
atattatcat	ggaatcgctt	gttatacact	gagtgttttg	caataaggga	tttttagaag	11220
aaggagtttt	tattctcaac	aggttccctt	agtttagctc	aaataaatct	aagcaatcca	11280
ctctagaatt	aaatagtttc	ctaagggcac	agctatgaat	agagctcaat	ttacatataa	11340
aattttgttc	accatttatg	tcattccagt	tttcattagt	acaaggaaaa	tacaaaatat	11400
ttagatgtca	atatcaagtg	aatagttcat	ctcctttttt	aatatatatc	acctaaatca	11460
ccattttctc	agaaaaatct	ggcctgaagt	tctgtctgga	acttcaacat	gaaaaatatg	11520
cacagcttgc	tattataaat	cctagtgtgat	ttttaagatt	catgtctggt	gtctgactca	11580
gagggggccag	aggctagaca	aatatttttt	gaatcttcat	tgtgaagatt	tttaattgatt	11640
attttaatat	aaataacaaa	gatgatggat	aatgtaactt	tgtacagttc	atagacgctg	11700
aactactttg	tgttataaat	gttagttccc	tatcataaat	gatagggtgat	aagtgtatgt	11760
ttataacttt	ccctctgagc	tatattcatg	tactagagaa	ttatttttaa	catgaaaaga	11820
ctgtgtttat	agtctcagct	cctgagaact	ggtccaacct	taggcagggtg	aatgccagga	11880
gcaacgtttt	tcttctacag	aggatgcttt	gctgccaaagc	aacctgggttg	tgtggaaatg	11940
ttcctttttt	aatcaagttt	aaagggtctt	catcatgctg	ttgctccaca	tattttcagg	12000
ttagagcttg	gtccttggag	tattatcttt	taccagaaaa	ttcatagtat	tctttcaata	12060
actaacaact	aaacttttctg	ataaaaaaga	attggaattt	caatttttaa	gcctgagtaa	12120
aattcttgtg	aatcaggata	ttttattttta	agtcttatct	tttaaaaagt	tattttat	12180
tttaaaaaat	tataatatac	tttcataatt	tccctccttc	acttttcttt	acaaacactt	12240
ctatagatca	ccatgtgttt	ttttttttac	atttatggcc	tctttctggt	cattgttatt	12300
acatacaaat	agtcttgctt	atagaagaac	accacaattt	gttacctgat	aacaaattat	12360
caacctttaa	aacctacaaa	ctattgatat	tactgaaaag	actatactta	tagatgtaaa	12420
gatatatgtg	tgtgcacata	tatagataca	catatatgta	ggatttttta	tttttagattt	12480
tagacatcaa	aattatttat	atgactgaga	aactagacac	tataaatgag	cattcagtat	12540
tcaacaccgt	gatttttagat	attgtcacaa	tgacagaaaa	ttttcttata	gaaaattttta	12600
agttttgtga	ttgctctgtg	cacttagtga	agtctcacag	aaaaagaatc	atagtatttt	12660
tagttttataa	taaaaagtac	atataattaa	aatggttggc	acaaaacaac	atgttagcat	12720

ttttcctatt	tactatcaag	tagtatcatt	ttgaaataat	aatttgacta	gtttcaaaaa	12780
tgaaaacaaa	atttaaacta	aatgccta	ctagcctgat	aacattttta	tgaatgaaat	12840
tattcaatag	tgttatcaat	taggggcccc	aaacttttcc	taaaataaaa	cttttaattt	12900
ttttccattt	ttattttaat	tagaaacaaa	attgttttac	atgtaaatca	gagtttcctc	12960
accctcccct	tctccctgtc	cctcactaac	accctacttg	tcccatacca	tttctgctcc	13020
ccagggaggg	tgaggccttc	catggggaaa	cttcagagtc	tgtctatcct	ttcggatagg	13080
gcctaggccc	tcacccattt	gtctaggcta	aggctcacaa	agtttactcc	tatgctagt	13140
ataagtactg	atctactaca	agagacacca	tagatttcct	aggettccct	actgacaccc	13200
atgttcatgg	ggtctggaac	aatcatatgc	tagtttccta	ggtatcagtc	tggggaccat	13260
gagctcccc	ttgttcaggt	caactgtttc	tgtgggtttc	accaccctgg	tcttgactgc	13320
tttgctcatc	actcctccct	ttctgtaact	gggttcagtc	acaattccgt	gtttagctgt	13380
gggtgtctac	ttctactttc	atcagcttct	gggatggagc	ctctaggata	gcatacaatt	13440
agtcacatc	tcattatcag	ggaagggcat	ttaaagtagc	ctctccattg	ttgcttggt	13500
tgttagttgg	tgctactctt	gtagatctct	ggacatttcc	ctagtgccag	atatctcttt	13560
aaacctacaa	gactacctct	attatggtat	ctcttttctt	gctctcgtct	attcttccag	13620
acaaaatctt	cctgctccct	tatatcttcc	tctccctctc	tcttctcccc	ttctcattct	13680
cctagatcca	tcttcccttc	ccccatgctc	ccaagagaga	tgttgctcag	gagatcttgt	13740
tccttaaccc	ttttcttggg	gatctgtctc	tcttaggggt	gtccttggtt	cctagcttct	13800
ctggaagtgt	ggattgtaag	ctggtaatca	tttgctccat	gtctaaaatc	catatatgag	13860
tgatgtttgt	ctttttgtga	ctgggttacc	tactcaaaa	tgggttcttc	catatgtctg	13920
tggatttcaa	tagcacaaac	aacatacagt	atcttggggc	aacactaacc	aaacaagtga	13980
aagaccagta	tagcaagaac	tttgagttaa	aagaaagaaa	ttaaagaaga	taccagaaaa	14040
tggaaagatc	tcccatgctc	tttgataggc	agaatcaaca	tagtaaaaat	ggcaatcttg	14100
ccaaaatcca	tctacagact	caatgcaatc	cccattaaat	accagcacac	ttcttcacag	14160
acctgaaaga	ataatactta	actttatatg	gagaaacaaa	agaccagga	taggccaaac	14220
aacctgtac	aatgaaggca	cttccagagg	catccccatc	cctgaacttc	agctctatta	14280
tagagtaata	atcctgaaaa	cagcttggtg	atggcacaaa	aatagacagg	tagaccaatg	14340
gaattgagtt	gaaaaccctg	atattaaccc	acatatctat	gaacacctga	ctttgacaaa	14400
gaagctaagg	ttatacaatg	taagaaagaa	agcatcttca	acaaatcgtg	ctggcataac	14460
tggatgctgg	catgtagaag	actgcagata	gatccatgct	taatgccatg	cacaaaactt	14520
aagtccaaat	ggatcaaaaa	cctcaacata	aatccagcca	cactgaacct	catagaagag	14580
aaagtgggaa	gtatccttga	ataaattggt	acaggagacc	acatcttgaa	cttaacacca	14640
gtagcacaga	caatcagatc	aataatcaat	aatgggacc	tctgaaact	gagaagcttc	14700
tgtaggcaa	tggataagtc	aacaggacaa	aatggcagcc	cacggaatgg	gaaaagatat	14760
tcaccaatcc	tatatctgac	agagggctgc	tctctatttg	caaagaacac	aataagctag	14820
tttttaaaac	accaattaat	ccgattataa	agttgggtag	agaactaaat	aaagaattgt	14880
taacagagca	atctaacttg	gcagaaagac	acataagaaa	gtgctcacca	t	14931

It is to be understood that the description, specific examples and data, while indicating exemplary embodiments, are given by way of illustration and are not intended to limit the present inventions. Various changes and modifications within the present invention, including combining embodiments in whole and in part, will become apparent to the skilled artisan from the discussion, disclosure and data contained herein, and thus are considered part of the inventions.

What is Claimed Is:

1. A eukaryotic cell from a eukaryotic cell line, wherein the eukaryotic cell comprises a polynucleotide comprising (i) a promoter, (ii) an intron, (iii) a first internal ribosome entry site, (iv) a first AAV Cap gene, (v) a second internal ribosome entry site, (vi) a second AAV Cap gene, and (vii) a polyadenylation site, wherein (i) to (vii) are operably linked.

2. The eukaryotic cell according to claim 1, wherein the polynucleotide is integrated into a cell genome.

3. A method of producing adeno-associated virus (AAV) Cap protein in cell culture, wherein the method comprises the steps of:

providing the cell culture that comprises eukaryotic cells, wherein a eukaryotic cell of the cell culture comprises a polynucleotide comprising (i) a promoter, (ii) an intron, (iii) a first internal ribosome entry site, (iv) a first AAV Cap gene, (v) a second internal ribosome entry site, (vi) a second AAV Cap gene, and (vii) a polyadenylation site, wherein (i) to (vii) are operably linked; and

culturing the eukaryotic cells in a culture medium to allow the eukaryotic cells to produce AAV VP1, VP2 and VP3 protein.

4. The method according to claim 3, wherein the polynucleotide is integrated into a cell genome.

5. The method according to claim 3, wherein the cell is a CHO cell or a BHK cell.

6. The method according to claim 3, wherein the cell is a HEK 293 cell or a human amniotic cell.

7. The cell according to claim 1, further comprising an operator.

8. The cell according to claim 7, wherein the promoter is a CMV promoter and the operator is a Tet operator.

9. The cell according to claim 1, further comprising:
a polynucleotide encoding AAV Rep;
a polynucleotide encoding Ad E1A;
a polynucleotide encoding Ad E1B;
a polynucleotide encoding Ad E2A or E2A orf;
a polynucleotide encoding Ad E4 or E4 orf 6;
a polynucleotide encoding VA RNA; and
a polynucleotide encoding AAV ITRs and a protein of interest.

10. The cell according to claim 1, wherein the cell is a CHO cell or a BHK cell.

11. The cell according to claim 1, wherein the cell is a HEK 293 cell or a human amniotic cell.

12. The method according to claim 3, wherein the cell further comprises an operator.

13. The method according to claim 12, wherein the promoter is a CMV promoter and the operator is a Tet operator.

14. The method according to claim 3, wherein the cell further comprises:
a polynucleotide encoding AAV Rep;
a polynucleotide encoding Ad E1A;
a polynucleotide encoding Ad E1B;
a polynucleotide encoding Ad E2A or E2A orf;
a polynucleotide encoding E4 or E4 orf 6;
a polynucleotide encoding VA RNA; and
a polynucleotide encoding AAV ITRs and a protein of interest, wherein the cell can produce recombinant AAV.

15. The method according to claim 3, wherein the cells in the cell culture can express VP1, VP2 and VP3 proteins in a ratio of 1:2:9.3 of VP1 to VP2 to VP3.