

**MAMMALIAN CELLS COMPRISING INTEGRATED CAS9 GENES TO PRODUCE
STABLE INTEGRATION SITES, AND MAMMALIAN CELLS COMPRISING
STABLE INTEGRATION SITES AND OTHER SITES**

This Application claims priority to U.S. Application Serial No. 63/256,675, filed October 18, 2021, which is hereby incorporated by reference in its entirety.

FIELD OF THE INVENTIONS

[0001] The present inventions provide mammalian cells (including cell lines), including human and rodent cells (including cell lines), that comprise multiple Stable Integration Sites (SIS), which can be produced using integrated Cas9 genes. The inventions provide Stable Integration Sites (1) introduced genomically into Genomic Safe Harbors (GSH), for example AAVS1 (Adeno-Associated Virus Integration Site 1) and AAVS1-like, and (2) introduced genomically outside of that particular Genomic Safe Harbor, such as a different Genomic Safe Harbor or other region that is not a Genomic Safe Harbor. Polydeoxyribonucleotides of interest that encode polypeptides or RNAs of interest can be inserted into the Stable Integration Sites provided according to the inventions.

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 7, 2022, is named "135975-97402.xml" and is 709,205 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] Mammalian cell lines are the preferred approach for producing commercial quantities of therapeutic proteins, such as antibodies. However, it has been reported that modified mammalian cells often exhibit production decreases due to genetic and epigenetic instability. Hilliard and Lee, *Biotech. Bioeng.* 118: 659-75 (2021).

[0004] Integration of polynucleotides is the preferred approach for creating and maintaining transformed cells. Integration of particular sequences into human AAVS1 is discussed in Liu *et al.*, *BMC Research Note*, 7: 626 (2014) and Ramachandra *et al.*, *Nucl. Acids Res.* 39: e107 (2011). Human AAVS1 is known as a Genomic Safe Harbor. Papapetrou *et al.*, *Molecular Therapy* 24: 678-84 (2016). Gaidukov *et al.*, *Nucl. Acids Res.* 46: 4072-86 (2018) have disclosed sites for DNA integration into landing pads.

[0005] Chinese hamster ovary (CHO) cells and baby hamster kidney cells (BHK) are used in the production of therapeutic proteins, and hamster genomes have been extensively studied. Hamaker and Lee have reported on CHO chromosomal loci as potential sites for stable integration and refers to them as “genomic hot spots”. *Curr. Op. Chem. Eng.* 22: 152-60 (2018) at 153. At Table 1, Hamaker and Lee identify 30 hot spot loci, of which 17 are identified by gene and 13 are unannotated. *Curr. Op. Chem. Eng.* 22: 152-60 (2018) at 154. This work was followed by Hilliard and Lee, who sought to identify safe harbor regions in CHO using an epigenome analysis. Hilliard and Lee, *Biotech. Bioeng.* 118: 659-75 (2021). The authors determined that 10.9% of the CHO genome contained chromatin structures with enhanced genetic and epigenetic stability. The authors

further determined that of the 30 hot spots identified Table 1 by Hamaker and Lee, five of which overlapped with stable regions determined by high throughput chromosome conformation capture (Hi-C). The closest genes to the regions were *ALDH5A1*, *SMAD6* and *CLCN3*, and two other regions were unannotated. Hilliard and Lee, *Biotech. Bioeng.* 118: 659-75 (2021) at Supplementary Table 3 (S3). Gaidukov *et al.*, *Nucl. Acids Res.* 46: 4072-86 (2018) at Table 1 also identifies loci for integration in CHO cells. Lee *et al.*, *Scientific Reps.* 5: 8572 (2015) identifies the *COSMC* locus.

[0006] The present inventions advantageously employ an integrated Cas9 gene to efficiently create mammalian cell intermediates that are further modified to provide mammalian cells having multiple Stable Integration Sites for stable integration of multiple DNA cassettes and other polydeoxyribonucleotides of interest. According to the inventions, a Stable Integration Site can be located in a Genomic Safe Harbor or other regions, including newly-identified Genomic Safe Harbors.

SUMMARY OF THE INVENTIONS

[0007] The inventions provide mammalian cells, wherein any cell thereof can comprise a first Stable Integration Site located in a Genomic Safe Harbor and a second Stable Integration Site that is not located in the Genomic Safe Harbor, wherein the first Stable Integration Site comprises a first reporter gene encoding a first reporter protein and the second Stable Integration Site comprises a second reporter gene encoding a second reporter protein, wherein the first reporter protein and the second reporter protein are different. The first and second Stable Integration Sites can comprise recombinase recognition sites (RRSs). The first and

second reporter genes can be under the control of SV40 promoters. The first and second reporter genes can be fluorescent proteins. The cells can further comprise a polynucleotide encoding a repressor protein under the control of a CMV promoter. The cells can be a Human Amniotic Epithelial, HEK 293, CHO or a BHK Cell. The polynucleotide encoding a protein of interest can be inserted into the first Stable Integration Site or the second Stable Integration Site. The second Stable Integration Site can be located in a second Genomic Safe Harbor that is different from the first Genomic Safe Harbor or in a region that is not a Genomic Safe Harbor.

[0008] The inventions also provide mammalian cells, wherein any cell thereof can comprise a first Stable Integration Site located in a Genomic Safe Harbor and a second Stable Integration Site that is not located in the first Genomic Safe Harbor, wherein the first Stable Integration Site comprises first polynucleotide encoding a first protein and the second Stable Integration Site comprises a second polynucleotide encoding a second protein. The first and second proteins can be viral proteins, such as an adenovirus associated virus protein or an adenovirus protein. For example, mammalian cells can comprise a polynucleotide encoding an adeno-associated virus protein and a polynucleotide encoding an adenovirus protein. Other polynucleotides encoding proteins include, but are not limited to, antibody genes, for example. Cells can have the second Stable Integration Site located in a second Genomic Safe Harbor that is different from the first Genomic Safe Harbor that the first Stable Integration Site is located in, or in a region that is not a Genomic Safe Harbor.

[0009] The inventions further provide a mammalian cells, wherein any cell thereof can comprise a first Stable Integration Site located in a Genomic Safe

Harbor and a second Stable Integration Site that is not located in the Genomic Safe Harbor, wherein the first Stable Integration Site comprises a polynucleotide encoding a first reporter gene encoding a first reporter protein and the second Stable Integration Site comprises a polynucleotide encoding Cas9 and a polynucleotide encoding a second reporter gene encoding a second reporter protein, wherein the first reporter protein and the second reporter protein are different. The second Stable Integration Site can further comprise a selection marker gene and an internal ribosome entry site (IRES). The first and second Stable Integration Sites can comprise recombinase recognition sites. The first and second reporter genes can be under the control of SV40 promoters. The first and second reporter genes can be fluorescent proteins. The cell can further comprise a polynucleotide encoding a repressor (for example TetR) under the control of a promoter (for example, CMV). The cell can be a Human Amniotic Epithelial Cell, HEK293, CHO or a BHK Cell. The polynucleotide encoding a protein of interest can be inserted into the first Stable Integration Site or the second Stable Integration Site. The selection marker protein can confer drug resistance. The second reporter gene, the selection marker gene, the IRES and an SV40 promoter can be arranged on a DNA cassette. The cell can further comprise a polynucleotide encoding a repressor protein under the control of a promoter (for example, CMV). The second Stable Integration Site can be located in a second Genomic Safe Harbor that is different from the first Genomic Safe Harbor that the first Stable Integration Site is located in or in a region that is not a Genomic Safe Harbor. The first reporter gene can be flanked by a 5' genomic safe harbor homology arm and a 3' genomic safe harbor homology arm. The 5' genomic safe harbor homology arm can comprise a

CRISPR sgRNA target site and the 3' genomic safe harbor homology arm can comprise a CRISPR sgRNA target site.

[0010] The inventions further provide methods for making at least one protein of interest, wherein any method thereof can comprise: (a) providing mammalian cells comprising a first Stable Integration Site located in a Genomic Safe Harbor and a second Stable Integration Site that is not located in the first Genomic Safe Harbor, wherein the first Stable Integration Site comprises a first reporter gene encoding a first reporter protein and the second Stable Integration Site comprises a second reporter gene encoding a second reporter protein, wherein the first reporter protein and the second reporter protein are different, and wherein the first and second Stable Integration Sites comprise recombinase recognition sites; (b) introducing a polynucleotide encoding the protein of interest into a Stable Integration Site by recombinase mediated cassette exchange, and (c) culturing the mammalian cell of under conditions that allow expression of the polynucleotide encoding the polynucleotide of interest. The first and second reporter genes can be under the control of SV40 promoters. The first and second reporter genes can be fluorescent proteins. The cell can further comprise a polynucleotide encoding a repressor protein under the control of a CMV promoter. The cell can be a Human Amniotic Epithelial, HEK 293, CHO or a BHK Cell. The polynucleotide encoding a protein of interest can be inserted into the first Stable Integration Site or the second Stable Integration Site. The second Stable Integration Site can be located in a second Genomic Safe Harbor that is different from the first Genomic Safe Harbor that first Stable Integration Site is located in, or in a region that is not a Genomic Safe Harbor. The first Stable Integration Site comprises a first polynucleotide encoding a first protein and the second Stable Integration Site comprises a second

polynucleotide encoding a second protein. The first and second proteins can be viral proteins, such as an adenovirus associated virus protein or an adenovirus protein. For example, the mammalian cell can comprise a polynucleotide encoding an adeno-associated virus protein and a polynucleotide encoding an adenovirus protein. Other polynucleotides encoding proteins include, but are not limited to, antibody genes, for example. The second Stable Integration Site also can be located in a region that is not a Genomic Safe Harbor.

[0011] The inventions further provide methods of creating mammalian cells with multiple Stable Integration Sites, wherein any method thereof can comprise: (A) providing a mammalian cell comprising a first DNA cassette comprising in 5' to 3' order a polynucleotide encoding the first lox site, a promoter, a selection marker gene encoding a selection marker protein, an IRES, a first reporter gene encoding a first reporter protein, a promoter operably linked to an operator, a Cas9 gene and the second lox site; (B) integrating a second DNA cassette comprising in a 5' to 3' order a polynucleotide comprising a first Genomic Safe Harbor homology arm containing a CRISPR sgRNA target site, a third lox site, a second reporter gene encoding a second reporter protein, a forth lox site and a second Genomic Safe Harbor homology arm containing an CRISPR sgRNA target site, wherein the first lox site, the second lox site, the third lox site and the forth lox site are different, wherein the first and second guide arms can contain a region with alterations (if needed to avoid recreating a targetable site), and wherein the second reporter protein is different from the first reporter protein; (C) exchanging the first DNA cassette with a third DNA cassette, wherein the third DNA cassette comprises in a 5' to 3' order a polynucleotide encoding the first lox site, a promoter, a third reporter gene encoding a third reporter protein, and the second lox site, wherein the

third reporter protein is different from the second reporter protein, thereby providing the mammalian cell with multiple Stable Integration Sites. The mammalian cells can be Human Amniotic Epithelial Cells, HEK 293 Cells, CHO Cells or BHK Cells.

Reporter genes for use can fluorescent proteins. The cell of step (A) can further comprise a polynucleotide encoding a repressor (for example, TetR) under the control of a promoter (for example, CMV). The cell of step (B) can further comprise a polynucleotide encoding a repressor (for example, TetR) under the control of a promoter (for example, CMV). The cell of step (C) can further comprise a polynucleotide encoding a repressor (for example, TetR) under the control of a promoter (for example, CMV). The selection marker protein can confer drug resistance. Lox sites are the most commonly used type of RRS; however, different RRSs can be used as well.

[0012] The inventions also provide methods of creating a mammalian cell with multiple recombinase-mediated cassette exchange sites, wherein any method thereof can comprise: (A) randomly integrating a promoter and polynucleotide encoding a repressor into the cell genome, wherein the repressor can bind to a ligand; (B) randomly integrating into the cell genome a first DNA cassette comprising in 5' to 3' order a polynucleotide encoding a first lox site, a promoter and optionally an operator, a first reporter gene encoding a first reporter protein, an IRES, a first selection marker gene encoding a first selection marker protein and a second lox site, wherein the first lox site and the second lox site are different; (C) exchanging the first DNA cassette with a second DNA cassette, wherein the second DNA cassette comprises in 5' to 3' order a polynucleotide encoding the first lox site, a promoter, a second selection marker gene encoding a second selection marker protein, an IRES, a second reporter gene encoding a

second reporter protein, a promoter and an optional operator, a Cas9 gene and the second lox site, wherein the first and second selection marker proteins are different and the first and second reporter proteins are different; (D) integrating a third DNA cassette comprising in 5' to 3' order a polynucleotide comprising a first Genomic Safe Harbor (GSH) homology arm containing an sgRNA (single guide RNA) target site, a third lox site, a third reporter gene encoding a third reporter protein, a fourth lox site and a second GSH homology arm containing an sgRNA target site, wherein the first lox site, the second lox site, the third lox site and the fourth lox site are different, wherein the first and second guide arms can contain at least one region with alterations (if needed to avoid recreating a targetable site), and wherein the third reporter protein is different from the second reporter protein and can be the same or different from the first reporter protein; and (E) exchanging the second DNA cassette with a fourth DNA cassette, wherein the fourth DNA cassette comprises in a 5' to 3' order a polynucleotide encoding the first lox site, a promoter, a fourth reporter gene encoding a fourth reporter protein, and the second lox site, wherein the fourth reporter protein is different from the third reporter protein and the second reporter protein and preferably different from the first reporter protein, thereby providing the cell with multiple Stable Integration Sites. Lox sites are the most commonly used type of RRS; however, different RRSs can be used as well.

[0013] The inventions further provide mammalian cells comprising a modified genome, wherein a given genome is modified by insertion of at least three DNA cassettes within different regions of the genome, wherein the modified genome comprises (1) a first deoxyribonucleic acid sequence that is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%

identical to at least one selected from the group consisting of SEQ ID NOS: 1 and 2 prior to modification; (2) a second deoxyribonucleic acid sequence that is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to at least one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification; and (3) a third deoxyribonucleic acid sequence that is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification, wherein the first deoxyribonucleic acid sequence is modified by insertion of a first DNA cassette, the second deoxyribonucleic acid sequence is modified by insertion of a second DNA cassette, and the third deoxyribonucleic acid sequence is modified by insertion of a third DNA cassette. The mammalian cells can each have (a) the first DNA cassette comprise a promoter and at least one selected from the group consisting of a selectable marker gene and a reporter gene; (b) the second DNA cassette comprise a promoter and at least one selected from the group consisting of a selectable marker gene and a reporter gene; and (c) the third DNA cassette comprise a promoter and at least one selected from the group consisting of a selectable marker gene and a reporter gene. Moreover, the mammalian cells can each have (a) the first DNA cassette comprise a promoter, a selectable marker gene and a reporter gene; (b) the second DNA cassette comprise a promoter, a selectable marker gene and a reporter gene; and (c) the third DNA cassette comprises a promoter, a selectable marker gene and a reporter gene. The first deoxyribonucleic acid sequence comprises a Stable Integration Site, and a gene of interest inserted therein. The gene of interest can encode a polypeptide of interest selected from the group consisting of antibodies, antibody chains, receptors, Fc-containing proteins,

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. The mammalian cells can be human cells and the first deoxyribonucleic acid sequence is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 1. Alternatively, the mammalian cell can be a CHO cell and the first deoxyribonucleic acid sequence is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 2. The first deoxyribonucleic acid sequence can comprise a Stable Integration Site produced using a guide sequence selected from the group consisting of SEQ ID NOS: 13 to 419. Additionally, the first deoxyribonucleic acid sequence can comprise a Stable Integration Site produced by using a guide sequence that binds to and/or is complementary to target sequences in SEQ ID NO:2 at nucleotide position ranges selected from the group consisting of: (a) 1 to 2000; (b) 2001 to 4000; (c) 4001 to 6000; (d) 6001 to 8000; (e) 8001 to 10,000; (f) 10,001 to 12,000; (g) 12,001 to 14,000; (h) 14,001 to 16,000; (i) 16,001 to 18,000; (j) 18,001 to 20,000; (k) 20,001 to 22,000; (l) 22,001 to 24,000; (m) 24,001 to 26,000; (n) 26,001 to 28,000; (o) 28,001 to 30,000; (p) 30,001 to 32,000; (q) 32,001 to 34,000; (r) 34,001 to 36,000; (s) 36,001 to 38,000; (t) 38,001 to 40,000; (u) 40,001 to 42,000; and (v) 42,001 to terminus (44,232).

[0014] Additionally, there are provided mammalian cells comprising a modified genomes, wherein a modified genome comprises a deoxyribonucleic acid sequence comprising an AAVS1-like region modified by insertion of at least one

DNA cassette, and wherein a guide sequence selected from the group consisting of SEQ ID NOS: 13 to 419 that binds to and/or is complementary to a sense or antisense strand of the AAVS1-like region. The mammalian cells can further comprise a second deoxyribonucleic acid sequence that is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to at least one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification; and a third deoxyribonucleic acid sequence that is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification, wherein the first deoxyribonucleic acid sequence is modified by insertion of a first DNA cassette, the second deoxyribonucleic acid sequence is modified by insertion of a second DNA cassette, and the third deoxyribonucleic acid sequence is modified by insertion of a third DNA cassette. The second deoxyribonucleic acid sequence is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to at least one selected from the group consisting of one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification; and the third deoxyribonucleic acid sequence is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification. The first deoxyribonucleic acid sequence comprises a Stable

Integration Site produced by using a guide sequence that binds to and/or is complementary to at least one target sequence having at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO:2 at nucleotide positions: (a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

[0015] There are also provided mammalian cells comprising a modified genome, wherein a modified genome comprises a Stable Integration Site in a AAVS1-like region, wherein the Stable Integration Site is produced by using a guide sequence that binds to and/or is complementary to at least one target sequence having at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO:2 at nucleotide positions: (a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

[0016] There further provided mammalian cells according to the preceding paragraph, further comprising a second deoxyribonucleic acid sequence that is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to at least one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification; and a third deoxyribonucleic acid sequence that is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification, wherein the first deoxyribonucleic acid sequence is modified by insertion of a first DNA cassette, the second deoxyribonucleic acid sequence is modified by insertion of a second DNA cassette, and the third deoxyribonucleic acid sequence is modified by insertion of a third DNA cassette. The mammalian cell can have the second deoxyribonucleic acid sequence at least is 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to at least one selected from the group consisting of one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification; and the third deoxyribonucleic acid sequence is at least 50%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification.

[0017] Additionally, there are provided methods of producing proteins of interest, wherein the method comprises the steps of: (1) culturing the above

mammalian cells; and (2) harvesting the protein of interest. There also are provided cells made according to any of the above methods, as well as methods of using the disclosed cells.

BRIEF DESCRIPTION OF THE FIGURES

[0018] The below figures illustrate an exemplary progression and creation of intermediate cells useful for creating cells having Stable Integration Sites in different regions of the genome, and thereafter creating cells having Stable Integration Sites in different regions of the genome. These figures illustrate embodiments of the invention, and do not limit the inventions in any manner.

[0019] **FIGURE 1** schematically depicts the modification of a cell that has a polynucleotide encoding a repressor protein and a polyadenylation signal under transcriptional control of a promoter, wherein the polynucleotide is randomly inserted in the cell genome.

[0020] **FIGURE 2** schematically depicts the modification of the cell of Figure 1 after a DNA cassette (1) is randomly or site-specifically inserted into the cell genome. DNA cassette (1) comprises flanking lox sites (1 and 2), a promoter, a reporter gene (1), an IRES, selection marker gene (1) and a polyadenylation signal. Instead of lox sites, other RRSs can be used as well.

[0021] **FIGURE 3** schematically depicts the modification of the cell of Figure 2 wherein DNA cassette (1) is replaced by recombinase mediated cassette exchange with DNA cassette (2). DNA cassette (2) comprises flanking lox sites (1 and 2), a promoter, selection marker gene (2), an IRES and reporter gene (2) and polyadenylation signal, and a Cas9 gene with a second polyadenylation signal

under control of a second promoter (an operator is optional). Instead of lox sites, other RRSs can be used as well.

[0022] **FIGURE 4** schematically depicts the modification of the cell of Figure 3 that has a DNA cassette (3) comprising flanking Genomic Safe Harbor (GSH) homology arms, lox sites (3 and 4) and a reporter gene (3) with a polyadenylation signal under the control of a promoter inserted into the Genomic Safe Harbor. The insertion is a site-specific integration and creates a Stable Integration Site between Lox3 and Lox4. Instead of lox sites, other RRSs can be used as well.

[0023] **FIGURE 5** schematically depicts the modification of the cell of Figure 4, wherein DNA cassette (2) is replaced by recombinase mediated cassette exchange with DNA cassette (4). DNA cassette (4) comprises flanking lox sites (1 and 2), reporter gene (4) and a polyadenylation signal under the control of a promoter. This exchange removes the Cas9 gene. Instead of lox sites, other RRSs can be used as well.

[0024] **FIGURE 6** schematically depicts an sgRNA plasmid used in Example 6.

[0025] **FIGURE 7** depicts plots of Example 6 showing green fluorescent protein positive populations (Q1) for No HDR Template (control), 104mer HDR Template, 401mer HDR Template and 1030mer HDR Template. GFP positive is the vertical axis and CFP positive is the horizontal axis.

[0026] **FIGURE 8** schematically depicts a mammalian cell (HEK293, for example) with stably integrated Cas9 gene flanked by Lox sites 3 and 4. The Cas9 gene is under the control of at least a promoter (not depicted). AAVS1 also is

schematically depicted. Instead of lox sites, other RRSs can be used as well.

Promoters are present 5' of genes, but are not depicted.

[0027] **FIGURE 9A and FIGURE 9B** schematically depict targeting plasmids containing sgRNA target site, left homology arm (here a GSH homology arm) for insertion into a region, such as a Genomic Safe Harbor (here AAVS1), Lox 1 site, a reporter gene (color 1), Lox 2 site, a right homology arm (here a GSH homology arm) for insertion into a region, such as a Genomic Safe Harbor (here AAVS1). At the 3' end, **Figure 9A** schematically depicts a reporter gene (Color 2), and **Figure 9B** schematically depicts at the 3' end a negative selection gene (Negative Selection 1). Promoters and optionally other moieties (such as operators) are represented by arrows pointed in a 5' to 3' direction. Both plasmids insert color 1 into a region, such as a Genomic Safe Harbor. Instead of lox sites, other RRSs can be used as well.

[0028] **FIGURE 10** schematically shows the results after Cas9 mediated integration into the Genomic Safe Harbor (AAVS1) of the mammalian cell (HEK293, for example). Color 1 is flanked by Lox 1 and Lox 2. A gene of interest can replace color 1 via RMCE. When a targeting plasmid according to **FIGURE 9A** is properly integrated, the cell will be color 1 positive and color 2 negative. When a targeting plasmid according to **FIGURE 9B** is properly integrated, the cell will be color 1 positive and will be able to propagate because the negative selection gene is removed. Instead of lox sites, other RRSs can be used as well. Promoters and optionally other moieties (such as operators) are represented by arrows pointed in a 5' to 3' direction.

[0029] **FIGURE 11** schematically depicts the insertion of **FIGURE 10** in greater detail. The cellular genome, including AAVS1, flanks the insert and the 5'

and 3' ends. Color 1 is flanked by Lox 1 and Lox 2. **FIGURE 11**, left side indicates the location of 5' genome primer and 3' insertion primer used with 5' junction PCR. **FIGURE 11**, right side indicates the location of 5' insertion primer and 3' genome primer used with 3' junction PCR. Instead of lox sites, other RRSs can be used as well. A promoter 5' of the color 1 gene is depicted as a 5' arrow.

[0030] **FIGURE 12** shows that correct size fragments are amplified in HEK 293 cells by the junction PCR schematically depicted in **FIGURE 11**. Stable Cas9 targeted HEK293 cells and the 5' junction and the 3' junction are obtained and detected, which establish correct insertion.

[0031] **FIGURE 13** shows that correct size fragments are amplified in CHO cells by the junction PCR schematically depicted in **FIGURE 11**. Stable Cas9 targeted CHO cells and the 5' junction and the 3' junction are obtained and detected, which establish correct insertion. Instead of lox sites, other RRSs can be used as well.

[0032] **FIGURE 14** schematically depicts an exemplary cell comprising three cassettes integrated into regions of the genome with flanking RRSs (here lox 1 and lox 2). Depending on the cell type, each of the three cassettes can be integrated into different Stable Integration Sites (for example, AAVS1-like) schematically depicted at position A, and other available sites (such as Stable Site 1 and Stable Site 2) schematically depicted at positions B and C. The reporter genes can be the same or different. The negative selection genes can be the same or different, but preferably the same. The cell can contain additional Stable Integration Sites and integrated cassettes according the teachings contained herein. Promoters are present 5' of genes, but are not depicted.

[0033] **FIGURE 15** schematically depicts the modification of the cell of **FIGURE 14** at schematically depicted positions A, B and C. Three cassettes each comprise flanking RRSs (here lox 1 and lox 2), a gene of interest, a positive selection marker gene, and a reporter* gene. The positive selection marker genes can be the same or different, but preferably the same. The reporter* genes can be the same or different, but each must be different from any of the reporter genes in the cell of **FIGURE 14**. The genes of interest can be the same or different. The cassettes of **FIGURE 14** are replaced by the cassettes of **FIGURE 15** by RMCE. The cell can contain additional Stable Integration Sites and integrated cassettes according the teachings contained herein. Promoters are present 5' of genes, but are not depicted.

[0034] **FIGURE 16** is a bar graph comparing protein produced by a three-site CHO-K1 cell (A, B, and C) compared to a two-site CHO-K1 cell (B and C).

[0035] **FIGURE 17** schematically depicts an exemplary cell comprising four cassettes integrated into regions of the genome with flanking RRSs (here lox 1 and lox 2, or lox 3 and lox 4). Depending on the cell type, each of the four cassettes can be integrated into different Stable Integration Sites (and other available sites (such as Stable Site 1 and Stable Site 2), and schematically depicted as positions A and B (SISs) and C and D (Stable Sites 1 and 2). The reporter genes can be the same or different. The negative selection genes can be the same or different, but preferably the same. The cell can contain additional Stable Integration Sites and integrated cassettes according the teachings contained herein. Promoters are present 5' of genes, but are not depicted.

[0036] **FIGURE 18** schematically depicts the modification of the cell of **FIGURE 17** at schematically depicted positions A, B, C and D. Four cassettes each

comprise flanking RRSs (here lox 1 and lox 2, or lox 3 and lox 4), a gene of interest, a positive selection marker gene, and a reporter* gene. The positive selection marker genes can be the same or different, but preferably the same. The reporter* genes can be the same or different, but each must be different from any of the reporter genes in the cell of **FIGURE 17**. The genes of interest can be the same or different. In this figure, there are two copies of Gene of Interest 1 and two copies of Gene of Interest 2. The cassettes of **FIGURE 17** are replaced by the cassettes of **FIGURE 18** by RMCE. The cell can contain additional Stable Integration Sites and integrated cassettes according the teachings contained herein. Promoters are present 5' of genes, but are not depicted.

DETAILED DESCRIPTION OF THE INVENTIONS

Definitions

[0037] 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 this invention belongs.

[0038] 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 invention 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.

[0039] “AAVS1” can be a Genomic Safe Harbor and refers to Adeno-associated virus integration site 1, and is reported to be located on human chromosome 19 in nature and contains approximately 4.7 kilobases. The AAVS1 locus can be used according to the inventions.

[0040] “AAVS1-like” refers to an AAVS1 homolog found in CHO cells, and is disclosed herein. An AAVS1-like region containing an AAVS1-like Genomic Safe Harbor (GSH) can be used according to the inventions. SEQ ID NO:2 is an example of an AAVS1-like region.

[0041] 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. A cassette can comprise one or more gene and promoters, enhancers, operators, repressors, transcription termination signals, ribosomal entry sites, introns and polyadenylation signals.

[0042] “COSMC” has reportedly been found in hamster cells. Homologs of a partial or whole COSMC locus are candidates for use according to the inventions.

[0043] “CCR5” refers to C-C chemokine receptor type 5 gene, and has been reportedly found in human, mouse and rat cells. Homologs of a partial or whole CCR5 locus are candidates for use according to the inventions.

[0044] “Genomic Safe Harbors” or “GSH” refers to sites in the cell genome that can accommodate insertions of polynucleotides, such as DNA cassettes, and permit the inserted polynucleotide to function and not pose an undue

burden on a transformed cell. Accordingly, Genomic Safe Harbors are ideal locations for creating Stable Integration Sites for the insertions of DNA cassettes through the practice of the inventions. Genomic Safe Harbors that can be utilized herein include, but are not limited to, AAVS1 and AAVS1-like. Reported loci that are candidates include, but are not limited to, CCR5, COSMC and Rosa26.

[0045] “Genomic Safe Harbor homology arm” or “GSH homology arms” is derived from Genomic Safe Harbors, and have homology to the Genomic Safe Harbor. Preferably, the Genomic Safe Harbor homology arm comprise about 100 to 2000 bases, more preferably about 300 to 1800 bases, more preferably about 400 to 1600 bases, more preferably about 500 to 1500 bases, more preferably about 500 to 1300 bases, more preferably about 500 to 1100 bases, more preferably about 500 to 1000 bases, more preferably about 600 to 1000 bases, more preferably about 700 to 1000 base, more preferably about 800 to 1000 bases, and still more preferably about 900 to 1000 bases. Typically, a polynucleotide to be inserted into a Genomic Safe Harbor will be flanked by a 5’ GSH Homology Arm and a 3’ GSH Homology Arm. For example, see Figures 4 and 5 showing a lox site-flanked DNA cassette that is further flanked by GSH Homology Arms.

[0046] “hRosa26” refers to the human homolog of the murine Rosa26 locus (“Reverse Orientation Splice Acceptor”). “Rosa26” refers to a partial or whole Rosa26 locus, and has been reportedly found in hamster cells in addition to mouse and human cells. Homologs of a partial or whole Rosa26 locus are candidates for use according to the inventions.

[0047] An “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).

[0048] 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.

[0049] “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.

[0050] “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 *Lad* and *trpR* genes of *E. coli*. The bacteriophage operators from the lambda *Pi* and the phage P22 *Mnt* and *Arc*. Preferred operators are the Tet (tetracycline) operator (TetO or TO) and the Arc operator (ArcO or AO). 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: 4253-4266 (1986).

[0051] The Tet operator is preferred, and can be used to control transcription using a repressor, such as the Tetracycline repressor (TetR). Appropriate ligands for the repressor are tetracycline (tet), doxycycline (dox) and derivatives thereof. When the ligand binds to TetR, the affinity of the Tet repressor for the Tet operator is lessened and the Tet repressor separates from the operator, and thereby the operator becomes permissive for transcription. Other repressors can be paired for usage with their own respective operators.

[0052] 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.

[0053] A "homologous sequence" in its various grammatical forms 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%-99%, 75%-99%, 85%-99%, 90%-99%, 95%-98%, 98%-99%, 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.

[0054] "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.

[0055] "Polypeptide" and "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.

[0056] "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.

[0057] 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.

[0058] 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.

[0059] “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.

[0060] “Recombinase recognition sites” (RRS), also known as “heterospecific recombination sites,” are used in recombinase mediated cassette exchange (RMCE). Cre/Lox, Dre/Rox, Vre/Vlox, SCre/Slox and Flp/Frt are suitable RRS systems, for example. Suitable RRSs for use according to the inventions include Lox P, Lox 66, Lox 71, Lox 511, Lox 2272, Lox 2372, Lox 5171, Lox M2, Lox M3, lox M7 and Lox M11. These sites 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. Cre/Lox is most commonly used RRS, but other RRSs can be used instead of Cre/Lox according to the inventions.

[0061] “Reporter proteins” as used herein, refers to any protein capable of generating directly or indirectly 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.” Reporter genes and proteins 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. 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.

[0062] A “repressor protein”, also referred to as a “repressor,” is a protein that can bind to DNA in order to repressor transcription, and is encoded by a polynucleotide, also referred to herein as a “repressor gene” or a “repressor proteins gene.” 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, CalR, CprB, FarA, JadR*, JadR2, MphB, NonG, PhIF, TylQ, VanT, TarA, TyIP, BM1P1, Bm3R1, ButR, CampR, CamR, DhaR, KstR, LexA-like, AcnR, PaaRR, Psbl, Th1R, UidR, YDH1,

BetI, McbR, MphR, PhaD, Q9ZF45, TtK, Yhgd, YixD, CasR, IcaR, LitR, LuxR, LuxT, OpaR, Orf2, SmcR, HapR, Ef0113, HlyIIIR, 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,

[0063] “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 (selectable marker 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. In the figures, the selectable markers are positive selectable markers unless otherwise specified as a negative (neg.) marker.

[0064] “Single guide RNA” or “sgRNA” is used for targeting Cas9 to a site, and is usually 17-24 nucleotides long.

[0065] A “Stable Integration Site” or “SIS” is a region for site-specific integration of DNA polynucleotides of interest, including cassettes that comprise genes and/or other open reading frames, promoters and optionally other elements. Stable Integration Sites comprise an exogenously-sourced DNA cassette, and can be created according to the methods of the inventions described and depicted herein, preferably in a GSH. Constructs can be inserted into an SIS by a variety of approaches. Multiple Stable Integration Sites can be created and located on different chromosomes, different regions of the same chromosome or different positions in a same region of a chromosome.

[0066] A “Tetracycline Response Element” or “TRE” comprises seven copies of the 19 nucleotide TetO spaced apart by spacers comprising 17-18 nucleotides, and are commercially available. TetO sequences can vary and nucleotide substitutions are known. For example, altered sequences based on the Tet operator are disclosed in Wissmann *et al.*, *Nucleic Acids Res.* 14: 4253-66 (1986). The spacers are not sequence specific. The spacers can be similar, but all should not be identical. A TRE is considered a type of operator as used herein.

[0067] 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. 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. Thus, a recitation of ranges of values herein are merely intended to serve as a

shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein.

Detailed Description

[0068] The inventions provide mammalian cells with multiple Stable Integration Sites, and are suitable for production of proteins of interest, including viral proteins, and the production of viral vectors, including adeno-associated virus vectors (AAV). One or more Stable Integration Sites can be within the Genomic Safe Harbor and one or more Stable Integration Sites can be outside of the particular Genomic Safe Harbor. Multiple Stable Integration Sites can be created and located on different chromosomes, different regions of the same chromosome or different positions in a same region of a chromosome.

[0069] Genomic Safe Harbors are discussed in Pellenz *et al.*, *Hum. Gene Therapy* 30: 814-28 (2019); Papapetrou *et al.*, *Molecular Therapy* 24: 678-84 (2016).

[0070] Preferably, the Stable Integration Sites contain recognition sites to allow for Recombinase-Mediated Cassette Exchange (RMCE). Stable modification of cellular genomes can be undertaken with known approaches employing heterospecific recombination sites (also known as RRSs), 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; Sajjo *et al.*, *PLoS ONE* 9: e91435 (2014); Suzuki *et al.*, *Nucl. Acids. Res.* 39: e49 (2011) Integration using Bxb1 integrase in human,

mouse and rat cells also can be undertaken. Russell *et al.*, *Biotechniques* 40: 460-64 (2006).

[0071] Recombinase recognition sites, also known as heterospecific recombination sites, are 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. Suitable Lox sites for use according to the inventions include, but are not limited to, Lox P, Lox 66, Lox 71, Lox 511, Lox 2272, Lox 2372, Lox 5171, Lox M2, Lox M3, Lox M7 and Lox M11. Other RRSs can be used as well. Lox sites are the most commonly used type of RRS; however, different RRSs can be used as well.

[0072] Homology arms preferably start within about 10 to 20 bases, more preferably 10 to 15 bases, of the cut site. A greater distance can be used as well, but with lower efficiency. In order to ensure that the DNA cassette(s) inserted into the Genomic Safe Harbor(s) maintain stability in the event that the homology repair could possibly recreate a targetable site, as determined by the skilled person, the guide arm region of the DNA cassette can be made to contain alterations (for example, base mismatches) that disrupt the function of CRISPR target site. There are two approaches that can be employed independently or together. The first approach is to insert base substitutions to create base mismatches in the CRISPR twenty base target site or the protospacer adjacent motif (PAM), which is usually 2 to 6 bases. The second approach is to create a donor plasmid where insertion divides the CRISPR target site or divides the CRISPR target site from the PAM.

[0073] Human cell lines include amniotic cells (such as Human Amniotic Epithelial cells), Hela cells, Per.C6 cells and HEK 293 cells. Examples of HEK 293 cells include, but are not limited, to HEK 293, HEK 293A, HEK 293E, HEK

293F, HEK 293FT, HEK 293FTM, HEK 293H, HEK 293MSR, HEK 293S, HEK 293SG, HEK 293SGGD, HEK 293T and mutants and variants thereof. Rodent cell lines, such as Sp2/0 cells, BHK cells and CHO cells and mutants and variants thereof, also can be used according to the inventions. 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.

[0074] The mammalian cells of the inventions are produced by advantageously producing and utilizing a cell intermediate that has a cassette comprising a Cas9 endonuclease gene flanked by recombinase recognition sites and integrated into the genome via RCME. Without being bound by any theory, the inventive use of an integrated Cas9 gene when expressed appears to increase the efficiency of homology arm integration into Genomic Safe Harbors by increasing the occurrence of cuts in genomic DNA caused by the Cas9 endonuclease. The use of stably integrated Cas9 gene of the inventions provides 10 , 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , or 10^{10} greater HDR efficiency than HDR without a stably integrated Cas9 gene. Ultimately, this intermediate cell can be further subjected to RMCE to remove the cassette containing the Cas9 gene.

[0075] As a starting point for engineering of cells, polynucleotide sequences of interest, as well as the operably linked promoter and optional operators, may be introduced into the cell by transfection of a plasmid containing said polynucleotide sequences and elements. Accordingly, the inventions include the generation of cells as described.

[0076] Suitable plasmid constructs can be made by those of skill in the art. Useful regulatory elements, described previously or known in the art, can also be included in the plasmid constructs used to transfect the cells. Some non-limiting

examples of useful regulatory elements include, but are not limited to, promoters, enhancers, sequences encoding suitable mRNA ribosomal binding sites, and sequences that control the termination of transcription and translation. Suitable plasmid constructs also may comprise non-transcribed elements such as an origin of replication, other 5' or 3' flanking non-transcribed sequences, and 5' or 3' non-translated sequences such as splice donor and acceptor sites. One or more selectable marker genes may also be incorporated. Useful selectable marker proteins and reporter proteins for use with the present inventions are known and can be readily identified by those of skill in the art.

[0077] A plasmid construct encoding a gene of interest may be delivered to the cell using a viral vector or via a non-viral method of transfer.

[0078] Non-viral methods of nucleic acid transfer include naked nucleic acid, liposomes, and protein/nucleic acid conjugates. A plasmid construct that is introduced to the cell may be linear or circular, may be single-stranded or double-stranded, and may be DNA, RNA, or any modification or combination thereof.

[0079] A plasmid construct may be introduced into the cell by transfection. Those of skill in the art are aware of numerous different transfection protocols, and can select an appropriate system for use in transfecting cells. Generally, transfection methods include, but are not limited to, viral transduction, cationic transfection, liposome transfection, dendrimer transfection, electroporation, heat shock, nucleofection transfection, magnetofection, nanoparticles, biolistic particle delivery (gene gun), and proprietary transfection reagents such as Lipofectamine, Dojindo Hilymax, Fugene, jetPEI, Effectene, or DreamFect.

[0080] The inventions are further described by the following Examples, which are illustrative of the many embodiments and aspects of the invention, but do not limit the inventions in any manner. In the Examples, the selectable markers are positive selectable markers unless otherwise specified as a negative (neg.) marker.

EXAMPLE 1

[0081] This example concerns the creation of mammalian cells comprising a repressor, such as TetR, under control of a promoter, such as a CMV promoter. See Figure 1. The cell is transfected with a polynucleotide comprising the promoter and the repressor gene. The polynucleotide is randomly inserted into the cell genome. Western blots and Taqman can be used in the cell pool to identify transformants and determine average copy number. The integration of a repressor, such as TetR, allows for control of transcription of polynucleotides that under control of a promoter and an operator.

EXAMPLE 2

[0082] This example concerns further engineering of the cells of Example 1. DNA cassette 1 is schematically depicted in Figure 2 and comprises flanking lox sites (1 and 2) and further comprises in 5' to 3' order a promoter, reporter gene (1) encoding reporter protein (1), an IRES and selection marker gene (1) encoding selection marker protein (1) and a polyadenylation signal. DNA cassette (1) optionally can include an operator operably linked to the promoter. DNA cassette (1) is randomly or site-specifically inserted into the cell genome. The first lox site and the second lox site on DNA cassette (1) are different.

[0083] Where the tet operator is used in DNA cassette (1), multiple rounds of -ligand/+ligand sort and single cell sort will identify Lox-site stable cells for

dox-regulated expression. Thus, when the ligand, such as doxycycline or tetracycline, is present, TetR will not bind to the operator, and thereby conditions are permissive for transcription of reporter gene (1) and selection marker polynucleotide (1).

EXAMPLE 3

[0084] In this example, RMCE is performed to replace DNA cassette (1) with DNA cassette (2) in the cells of Example 2. As schematically depicted in Figure 3, DNA cassette (2) comprises flanking lox sites (1 and 2), and further comprises in 5' to 3' order a promoter, selection marker gene (2) encoding selection marker protein (2), an IRES and reporter gene (2) encoding reporter protein (2), and a Cas9 gene under control of a second promoter (optionally operably linked to an operator).

[0085] In an embodiment, a CMV promoter is operably linked to a tet operator to control transcription of the Cas9 gene. When the cells are in the presence of doxycycline or tetracycline, TetR is no longer able to bind the tet operator, and thus allow transcription of the Cas9 gene to occur. Reporter protein (1) is different from reporter protein (2), and selection marker protein (1) is different from selection marker protein (2).

EXAMPLE 4

[0086] This example concerns the integration of DNA cassette (3) into a Genomic Safe Harbor. See Figure 4. DNA cassette (3) comprises in 5' to 3' order a polynucleotide comprising a first Genomic Safe Harbor homology arm containing an sgRNA target site, lox site (3), a promoter operably linked to reporter gene (3) encoding reporter protein (3), a polyadenylation signal, lox site (4) and a

second Genomic Safe Harbor homology arm containing an sgRNA target site, wherein the first and second guide arm target sites each can contain a region with alterations if needed to avoid recreating a targetable site. Lox site (1), lox site (2), lox site (3), and lox site (4) are different from one another. Reporter protein (3) is different from reporter protein (2). Reporter protein (3) and reporter protein (1) can be the same or different. Homology arms of about 1000 bases are used in this example.

[0087] When the Cas9 endonuclease is expressed, the efficiency of DNA cassette 3 integration is increased. Without being bound by any theory, the inventive use of an integrated Cas9 gene appears to increase the efficiency of integration by increasing the occurrence of cuts in genomic DNA caused by the Cas9 endonuclease. The use of stably integrated Cas9 gene of the inventions provides 10 , 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 , 10^8 , 10^9 , or 10^{10} greater HDR efficiency than HDR without a stably integrated Cas9 gene.

[0088] If needed, alterations in the first and second Genomic Safe Harbor homology arms ensue that the DNA cassette (3) will stay integrated by avoiding recreation of a targetable site. The smaller cassette therein, namely the region between lox site (3) and lox site (4), is available for RMCE and is referred to as a Stable Integration Site.

EXAMPLE 5

[0089] This Example concerns the final form of the cell line, and is schematically depicted in Figure 5. To ensure stability of a cell line over time, it is preferred to remove the Cas9 gene. Accordingly, DNA cassette (2) is replaced by RMCE with DNA cassette (4), and removes the Cas9 gene. DNA cassette (4)

comprises flanking lox sites (1 and 2) and reporter gene (4) encoding reporter protein (4) under the control of a promoter. Reporter protein (4) is different from reporter protein (2) and reporter protein (3) and preferably different from reporter protein (1).

[0090] The resulting cells will have two integration sites within the genome, one integration site within a Genomic Safe Harbor (for example, a Stable Integration Site) and one integration site outside of that particular Genomic Safe Harbor. It is possible to create still further integration sites by applying the approaches described above, including the use of an integrated Cas9 gene and the use of additional and different GSH homology arms.

EXAMPLE 6

[0091] This example is a comparison of the efficiency of using Cas9 with homology directed repair (HDR) as disclosed herein compared to conventional HDR. As reported in the literature, HDR is precise, but desired recombinational events occur infrequently: 1 in 10^6 - 10^9 cells (0.0001% to 0.0000001%). Hsu *et al.*, *Cell* 157: 1262-78 (2014).

[0092] In order to assess the advantages of a stably integrated Cas9 gene, a CHO cell having the sites disclosed in U.S. Patent Nos. 7,771,997 ("Stable Site 1") and 9,816,110 ("Stable Site 2") was modified. 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, and each is referred to as an "enhanced expression locus" therein. Sequences set forth in these patents and Examples 11 and 12 can be used according to the inventions described and depicted herein.

[0093] A CHO cell was modified to include a cyano fluorescent protein reporter gene under control of a promoter in Stable Site 1, and a selection marker gene and a yellow fluorescent protein reporter gene under the control of the same promoter in Stable Site 2. Additionally, a Cas9 gene under control of a second promoter with an operator also was inserted into Stable Site 2. The Cas9 gene can be eventually removed in accordance with the teachings contained herein.

[0094] The cyano fluorescent protein can be change to fluoresce green by changing the tyrosine residue at position 66 to tryptophan. The sgRNA Delivery Plasmid comprise a selection marker (Ampicillin resistance), a POL III promoter (RNA Polymerase III promoter), a target sequence and gRNA scaffold, a POL III Terminator and Digest Sites 1 and 2. Pol III promoters include H1 and U6.

[0095] As depicted in Figure 6, sgRNA delivery plasmids were constructed containing HDR Templates: a 104mer insert (having a 57 bp arm and a 45 bp arm), a 401mer insert (having a 198 bp arm and a 201 bp arm) or a 1030mer (having a 524 bp arm and a 504 bp arm) insert containing homology arms and the sequence to effect the change from cyano to green, which in this example was composed of 2 nucleotides ("repair nucleotides"). The HDR templates were inserted into the Digest Sites (for example, NotI and/or other appropriate sites) of the sgRNA delivery plasmid to form a sgRNA target plasmid. A sgRNA delivery plasmid without an insert (No HDR Template) was used as a control

[0096] Figure 7 shows that the control exhibited no green positives in Q1. The cells with HDR Template exhibited green positives in Q1, and the green positive population in Q1 consistently increased with the increased size of the HDR Template (left to right). The cells with the 1030mer HDR Templates showed the greatest efficiency in repair, which was about 6.5 percent.

[0097] The cells of this example possess Stable Site 1 and Stable Site 2 and the SIS created in a GSH according to the inventions. Thus, this cell possess three sites for stable integration of genes of interest.

EXAMPLE 7 – Generation of an Intermediate Human Cell comprising a Stable Integration Site in a Genomic Safe Harbor (AAVS1)

[0098] In this example, the starting point is HEK293 cell with stably integrated Cas9 gene flanked by Lox sites 3 and 4. The Cas9 gene is under the control of at least a promoter (not depicted). AAVS1 also is schematically depicted. See Figure 8. This cell can be made according to Examples 1-4 and Figures 1-4.

[0099] Targeting plasmids containing sgRNA target site, left homology arm (here a GSH homology arm) for insertion into a region, such as a Genomic Safe Harbor (here AAVS1), Lox 1 site, a reporter gene (color 1), Lox 2 site, a right homology arm (here a GSH homology arm) for insertion into a region, such as a Genomic Safe Harbor (here AAVS1). See Figures 9A and 9B for alternative targeting plasmids. At the 3' end, one targeting plasmid has reporter gene (Color 2), See Figure 9A. The other targeting plasmid has at the 3' end a negative selection gene (Negative Selection 1). See Figure 9B. Promoters and optionally other moieties (such as operators) are represented by arrows pointed in a 5' to 3' direction in Figure 9A and Figure 9B. Both plasmids insert color 1 into a region, such as a Genomic Safe Harbor (here AAVS1).

[00100] Cas9 mediated integration of a targeting plasmid (for example, Figure 9A or Figure 9B) into the Genomic Safe Harbor (AAVS1) of the HEK293 cell is schematically depicted in Figure 10. Color 1 is flanked by Lox 1 and Lox 2. A gene of interest can replace color 1 via RMCE.

[00101] When a targeting plasmid according to Figure 9A is properly integrated, the cell will be color 1 positive and color 2 negative. When a targeting plasmid according to FIGURE 9B is properly integrated, the cell will be color 1 positive and will be able to propagate because the negative selection gene is removed. This cell is considered an intermediate. Ultimately, the cell can be further subjected to RMCE at lox sites 3 and 4 to remove the cassette containing the Cas9 gene, as shown in Figure 8. See, for example, Example 5.

[00102] The precision of this inventive methodology is shown in Figures 10 and 11. FIGURE 11 depicts the insertion of FIGURE 10 in greater detail. The cellular genome, including AAVS1, flanks the insert and the 5' and 3' ends. Color 1 is flanked by Lox 1 and Lox 2. Figure 11, left side identifies the location of 5' genome primer and 3' insertion primer used with 5' junction PCR. Figure 11, right side identifies the location of 5' insertion primer and 3' genome primer used with 3' junction PCR.

[00103] Junction PCR shows that correct size fragments are amplified and labeled as "Stable Cas9 targeted cells." See Figures 12 and 13. Stable Cas9 targeted cells and the 5' junction and the 3' junction are obtained and detected, which establish correct insertion. Positive and negative controls are at the right hand columns of each gel.

EXAMPLE 8 – CHO Regions and Sequences

[00104] For CHO cells, the sequences set forth in U.S. Patent Nos. 7,771,997 (Stable Site 1) and 9,816,110 (Stable Site 2) can be utilized. The sequences and homologous sequences within the percent identity values of U.S. Patent Nos. 7,771,997 and 9,816,110 are hereby incorporated by reference. An

AAVS1-like region disclosed herein can be used to create Stable Integration Sites according to the inventions.

[00105] Candidate loci for use according to the inventions are reported in the literature. Hamaker and Lee, *Curr. Op. Chem. Eng.* 22: 152-60 (2018) identify 30 hot spot loci. Hilliard and Lee, *Biotech. Bioeng.* 118: 659-75 (2021) sought to identify safe harbor regions in CHO using an epigenomic analysis for Hi-C stable regions, and found an overlap with 5 of the 30 regions identified by Hamaker and Lee. See Supplementary Table 3 of Hilliard and Lee. Gaidukov *et al.*, *Nucl. Acids Res.* 46: 4072-86 (2018) also identifies loci for integration in CHO cells, including a putative Rosa26. Lee *et al.*, *Scientific Reps.* 5: 8572 (2015) reported a COSMC locus in hamster cells. In sum, these papers identify several unannotated regions and gene regions in CHO, and the gene regions are set forth below:

<i>BMP5</i>	<i>SSBP2</i>	<i>TRMT6</i>	<i>CLCC1</i>	<i>FAM114A1</i> (<i>NOXP20</i>)
<i>LRBA</i>	<i>DCN</i>	<i>CEP128</i>	<i>AACS</i>	<i>ALDH5A1</i>
<i>SMAD6</i>	<i>PTPRQ</i>	<i>ROSA26</i>	<i>ADGRL4</i>	<i>GPM6A</i>
<i>KIAA1551</i> (<i>C12ORF35</i>)	<i>HPRT</i>	<i>CLCN3</i>	<i>FER1L4</i>	<i>COSMC</i>

EXAMPLE 9 – CHO Cells with Three or More Insertion Sites

[00106] CHO cells containing multiple insertion sites using the cells disclosed in US 2019/0233544 A1. Stable Site 1 and Stable Site 2 can be used initially in accordance with the teachings contained herein that utilize an integrated Cas9 gene. Once one or more Stable Integration Sites are created in Genomic Safe Harbors, such as in the AAVS1-like region (see, for example, SEQ ID NO:2) and counterpart guide sequences (see, for example, SEQ ID NOS:13 to 419). Guide sequences can bind to target sequences in SEQ ID NO:2 at nucleotide

position ranges selected from the group consisting of: (a) 1 to 2000; (b) 2001 to 4000; (c) 4001 to 6000; (d) 6001 to 8000; (e) 8001 to 10,000; (f) 10,001 to 12,000; (g) 12,001 to 14,000; (h) 14,001 to 16,000; (i) 16,001 to 18,000; (j) 18,001 to 20,000; (k) 20,001 to 22,000; (l) 22,001 to 24,000; (m) 24,001 to 26,000; (n) 26,001 to 28,000; (o) 28,001 to 30,000; (p) 30,001 to 32,000; (q) 32,001 to 34,000; (r) 34,001 to 36,000; (s) 36,001 to 38,000; (t) 38,001 to 40,000; (u) 40,001 to 42,000; and (v) 42,001 to 44,232.

[00107] Stable Site 1 and Stable Site 2 of U.S. Patent Nos. 7,771,997 and 9,816,110 can be used for expression of genes of interest to encode proteins of interest. Cells with SISs ultimately can have 3, 4, 5, 6, 7, 8, 9, 10 or more sites for expressing genes of interest.

[00108] Preferably, a CHO cell comprising Stable Sites 1 and 2 is modified to create a third site in a Genomic Safe Harbor, namely a Stable Integration Site. Preferred Genomic Safe Harbors for creation of such a CHO cell are in the AAVS1-like region. Other CHO cell types can be used to create multiple sites according to the teachings contained herein.

[00109] Figure 14 schematically depicts an exemplary cell comprising three cassettes integrated into regions of the genome with flanking RRSs (here lox 1 and lox 2). Depending on the cell type, each of the three cassettes can be integrated into different Stable Integration Sites and other available sites (such as Stable Site 1 and Stable Site 2) schematically depicted as positions A, B and C. The reporter genes can be the same or different. The negative selection genes can be the same or different, but preferably the same. The cell can contain additional Stable Integration Sites and integrated cassettes according the teachings contained herein.

[00110] Figure 15 schematically depicts the modification of the cell of Figure 14 at schematically depicted positions A, B and C . Three cassettes each comprise flanking RRSs (here lox 1 and lox 2), a gene of interest, a positive selection marker gene, and a reporter* gene. The positive selection marker genes can be the same or different, but preferably the same. The reporter* genes can be the same or different, but each must be different from any of the reporter genes in the cell of Figure 14. The genes of interest can be the same or different. The cassettes of Figure 14 are replaced by the cassettes of Figure 15 by RMCE. The cell can contain additional Stable Integration Sites and integrated cassettes according the teachings contained herein.

[00111] The combination of negative and positive selection assures isolation of cells that underwent recombination in all sites. If the gene of interest is the same in each of the three cassettes, the cell can result in high yield protein expression. For example, 7, 8, 9, 10 or more grams per liter (g/l) of protein production is possible.

[00112] Figure 16 shows the results from five different human IgG antibodies that were stably integrated using Cre-lox recombination into CHO K1 derived hosts engineered with either 2 integration sites (Stable Site 1 and 2) or 3 integration sites (Stable Site 1, Stable Site 2 and AAVS1-like (see SEQ ID NO:2)). Isogenic cell lines (ICLs) were isolated using flow cytometry. Fed batch production of ICLs were inoculated into chemically defined production media, and production cultures were carried out for 13 days. Antibody titer in conditioned media was determined using a protein A HPLC based method, and each three-site cell expressing a given antibody (1, 2, 3, 4, or 5) expressed a greater amount of protein than the comparison two-site cell. The three-site cell can provide increases of 10%,

20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 110%, 120%, 130%, 140%, 150% or more over the two-site cell.

[00113] Alternatively, different genes of interest can be used in the cassettes. For example, heavy chain and light chain sequences of an antibody can be gene of interest.

[00114] Turning to a four-site cell, preferably a CHO cell comprising Stable Site 1 and 2 is modified to create a third and fourth site in a Genomic Safe Harbor, namely a Stable Integration Site. Preferred Genomic Safe Harbors for creation of such a CHO cell are in the AAVS1-like region, which can be the third site. A fourth site can be created in other loci, including but not limited to:

<i>BMP5</i>	<i>SSBP2</i>	<i>TRMT6</i>	<i>CLCC1</i>	<i>FAM114A1</i> (<i>NOXP20</i>)
<i>LRBA</i>	<i>DCN</i>	<i>CEP128</i>	<i>AACS</i>	<i>ALDH5A1</i>
<i>SMAD6</i>	<i>PTPRQ</i>	<i>ROSA26</i>	<i>ADGRL4</i>	<i>GPM6A</i>
<i>KIAA1551</i> (<i>C12ORF35</i>)	<i>HPRT</i>	<i>CLCN3</i>	<i>FER1L4</i>	<i>COSMC</i>

[00115] Other CHO cell types can be used to create multiple sites according to the teachings contained herein.

[00116] Figure 17 schematically depicts an exemplary cell comprising four cassettes integrated into regions of the genome with flanking RRSs (here lox 1 and lox 2, or lox 3 and lox 4). Depending on the cell type, each of the four cassettes can be integrated into different Stable Integration Sites and other available sites (such as Stable Site 1 and Stable Site 2) schematically depicted as positions A, B, C and D. The reporter genes can be the same or different. The

negative selection genes can be the same or different, but preferably the same. The cell can contain additional Stable Integration Sites and integrated cassettes according the teachings contained herein.

[00117] Figure 18 schematically depicts the modification of the cell of Figure 17 at schematically depicted positions A, B, C and D . Four cassettes each comprise flanking RRSs (here lox 1 and lox 2, or lox 3 and lox 4), a gene of interest, a positive selection marker gene, and a reporter* gene. The positive selection marker genes can be the same or different, but preferably the same. The reporter* genes can be the same or different, but each must be different from any of the reporter genes in the cell of Figure 17. The genes of interest can be the same or different. In this figure, there are two copies of Gene of Interest 1 and two copies of Gene of Interest 2. The cassettes of Figure 17 are replaced by the cassettes of Figure 18 by RMCE. The cell can contain additional Stable Integration Sites and integrated cassettes according the teachings contained herein.

[00118] The combination of negative and positive selection assures isolation of cells that underwent recombination in all sites. A four-site cell is useful for making bispecific antibodies, wherein two distinct heavy chain/light chain plasmids can be targeted into distinct sites.

EXAMPLE 10 – Genomic Safe Harbor Sequences

[00119] Genomic Safe Harbors Sequences and the like are described herein, and many are in the literature and are publically available. Exemplary sequences are set forth below.

Human AAVS1 sequence

Human AAVS1 (Native RBS and guide RNA site for safe harbor insertion *indicated*)

(SEQ ID NO:1)

GAATTCCTAACTGCCCCGGGGCAGTCTGCTATTTCATCCCCCTTTACGCGGTGCTACACACACT
TGCTAGTATGCCGTGGGGACCCCTCCGGCCTGTAGACTCCATTTCAGCATTCCCCGGAGG
AGGCCCTCATCTGGCGATTTCCACTGGGGGCCTCGGAGCTGCGGACTTCCAGTGTGCATCG
GGGCACAGCGACTCCTGGAAGTGGCCACTTCTGCTAATGGACTCCATTTCAGGCTCCCGC
TACCTGCCCAGCACACCCTGGGGCATCCGTGACGTCAGCAAGCCGGGCGGGGACCGGAGATC
CTTGGGGCGGTGGGGGGCCAGCGGCAGTTCAGGCGGCCCCCGGGGCGGGCGGGCGGGCGG
GTGGTGGCGGC GGTTGGGGCTCCGGGCGCGTCGCTCGCTCGCTCGCTCGGGCGGGCGGGCGGT
GCGATGTCCGAGAGGATGGCCGGCGGTGGCCCCGGGGCGGCGGCGGGCTGCCCGGAGC
GGCGACGGGAGCAGCTGCGGCAGTGGGGCGCGGGCGGGCGCCGAGCCTGGCCCCGAGAGCG
CCGCGCCCCGACCGTCCGCTTCGAGCGCGCCGCCGAGTTCTTGGCGGCCTGTGCGGGCGGCG
ACCTGGACGAGGCGCGTCTGATGCTGCGCGCCGCCACCCTGGCCCCGGCGCCGAGCTCGA
CCCCGCCGGCCGCCGCCGCCGCCGCGCGTGTGGACTCCACCAACGCCGACGGTATCAGCGC
CCTGCACAGGTCAGCGCCCCCGCGGCGTCTCCCGGGGCCAGGTCCACCCTCTGCGCCACC
TGGGGCATCCTCCTTCCCCGTTGCCAGTCTCGATCCGCCCGTCGTTACTGGCCCTGGGTTT
NCACCCTATGCTGACACCCCGTTCCAGTCCCCCTTACCATTCCCTTCGACCACCCCACTTCG
AATTGGAGCGCTTCAACTGGCTGGGCTAGCACTCTGTGTGACACTCTGAAGCTCTACATTCC
CTTCGACCTACTCTCTTCGATTGGAGTCGCTTTAACTGGCCCTGGCTTTGGCAGCCTGTGCT
GACCCATCGAGTCCTCCTTACCATCCCTCCCTCGACTTCCCCCTCTTCCGATGTTGAGCCCCT
CCAGCCGGTCTTGACTTTGTCTCCTTCCCTGCCCTGCCCTCTCCTGAACCTGAGCCAGCTC
CCATAGCTCAGGTCTGGTCTATCTGCCTGGCCCTGGCCATTGTCACCTTTGCGCTGCCCTCCT
CTCGCCCCCGAGTGCCCTTGCTGTGCGCCGGAACCTCTGCCCTCTAACGCTGCCGTGCCGTC
TCTCTCCTGAGTCCGGACCACTTTGAGCTCTACTGGCTTCTGCGCGCCTCTGGCCCACTGTT
TCCCCTTCCAGGCAGGTCTGCTTTCTCTGACCAGCATTCTCTCCCCTGGGCTGTGCCGC
TTTCTGTCTGCAGCTTGTGGCCTGGGTCACCTCTACGGCTGGCCCAAGATCCTTCCCTGCCG
CCTCCTTCAGGTTCCGTCTTCTCCACTCCCTCTTCCCCTTGCTCTCTGCTGTGTTGCTGCC
CAAGGATGCTCTTTCGGAGCACTTCCTTCTCGGCGCTGCACCACGTGATGTCTCTGAGCG
GATCCTCCCCGTGTCTGGGTCTCTCCGGGCATCTCTCCTCCCTCACCCAACCCCATGCCGT
GTTCACTCGCTGGGTTCCCTTTTCTTCTCCTTCTGGGGCTGTGCCATCTCTCGTTTCTTA
GGATGGCCTTCTCCGACGGATGTCTCCCTTGCGTCCCGCCTCCCCTTCTTGTAGGCCTGCAT
CATCACCGTTTTTCTGGACAACCCCAAAGTACCCCGTCTCCCTGGCTTAGCACCTCTCCATC
CTCTTGCTTTCTTGCCTGGACACCCCGTTCTCCTGTGGATTGGGTACCTCTCACTCCTT
TCATTTGGGCAGTCCCCCTACCCCCCTTACCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCC
TGTCATGGCATCTTCCAGGGGTCCGAGAGCTCAGTAGTCTTCTTCTCCAAACCCGGGCCCT
ATGTCCACTTCAGGACAGCATGTTTGTGCTGCCCTCCAGGGATCCTGTGTCCCCGAGCTGGGACC
ACCTTATATTCCCAGGGCCGGTTAATGTGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTC
CACCCACAGT GGGGCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCCATCCTTAGGCCT
CCTCCTTCTAGTCTCCTGATATTCTGTCTAACCCCCACCTCCTGTTAGGCAGATTCCCTATC
TGGTGACACACCCCCATTTCTTGGAGCCATCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCT
TACGATGGAGCCAGAGAGGATCCTGGGAGGGAGACTTGGCAGGGGGTGGGAGGGAAGGGGGG
GATGCGTGACCTGCCCGGTTCTCAGTGGCCACCCTGCGCTACCCTCTCCCAGAACCTGAGCT
GCTCTGACGCGGTGTCTGGTGCCTTCACTGATCCTGGTGCTGCAGCTTCCTTACACTTCC

CAAGAGGAGAAGCAGTTTGGAAAAACAAAATCAGAATAAGTTGGTCCTGAGTTCTAACTTTG
 GCTCTTCACCTTTCTAGNCCCCAATTTATATTGTTCCGTGCGTCAGTTTTACCTGTGAG
 ATAAGGCCAGTAGCCACCCCGTCCTGGCAGGGCTGTGGTGAGGAGGGGGGTGTCCGTGTGG
 AAAACTCCCTTTGTGAGAATGGTGCGTCCTAGGTGTTACCAGGTCGTGGCCGCCTCTACTC
 CCTTTCTCTTTCTCCATCCATCCTTCTTTCCTTAAAGAGCCCCCAGTGCTATCTGGACATAT
 TCCTCCGCCCAGAGCAGGGTCCGCTTCCCTAAGGCCCTGCTCTGGGCTTCTGGGTTTGAGTC
 CTTGCAAGCCCAGGAGAGCGCTAGCTTCCCTGTCCCCCTTCCCTCGTCCACCATCTCATGCCC
 TGGCTCTCCTGCCCTTCCCTACAGGGGTTCCCTGGCTCTGCTCTTCAGACTGAGCCCCGTTCC
 CCTGCATCCCCGTTCCCCTGCATCCCCCTTCCCCTGCATCCCCCAGAGCCCAGGCCACCTA
 CTTGGCCTGGAACCCACGAGAGGCCACCCAGCCCTGTCTACCAGGCTGACCTTTTGGGTG
 ATTCTCCTCCAACGTGTGGGGTGACTGCTTGGGCAAACCTCACTCTTCGGGGTATCCCAGGAGG
 CCTGGAGCATTGGGGTGGGCTGGGGTTTCAGAGAGGAGGGATTCCCTCCAGGTTACGTGGCCA
 AGAAGCAGGGGAGCTGGGTTTGGGTCAGGCTGGGTGTGGGGTGACCAGCTTATGCTGTTTGC
 CCAGGACAGCCTAGTTTTAGCGCTGAAACCCTCAGTCCTAGGAAAACAGGGATGGTTGGTCA
 CTGTCTCTGGGTGACTCTTGATTCCCGGCCAGTTTCTCCACCTGGGGCTGTGTTTCTCGTCC
 TGCATCCTTCTCCAGGCAGGTCCCCAAGCATCGCCCCCTGGCTGTTCCCAAGTTCTTAGGT
 ACCCCACGTGGGTTTATGAACCACTTGGTGAGGCTGGTACCCTGCCCCATTCTGACCCCC
 AATTGCCTTAGTGGCTAGGGGGTTGGGGGCTAGAGTAGGAGGGGCTGGAGCCAGGATTCTTA
 GGGCTGAACAGAGCCGAGCTGGGGGCTGGGGCTCCTGGGTTTGAGAGAGGAGGGGCTGGGGC
 CTGGACTCCTGGGTCCGAGGGAGGAGGGGCTGGGGCTGGACTCCTGGGTCTGAGGGTGGAG
 GGACTGGGGGCTGGACTCCTGGGTCCGAGGGAGGAGGGGCTGGGGCTGGACTCGTGGGTCT
 TGAGGGAGGAGGGGTGCGGGGCTGGACTTCTGGGTCTTAGGGAGGCGGGGCTGGGCCCTGGA
 CCCCTGGGTCTGAATGGGGAGAGGCTGGGGGCTGGACTCCTTCATCTGAGGGCGGAAGGGC
 TGGGGCTGGCCTCCTGGGTTGAATGGGGAGGGGTTGGGCCTGGACTCTGGAGTCCCTGGTG
 CCCAGGCCTCAGGCATCTTTACAGGGATGCCTGTAC

CHO AAVS1-Like Region Sequence

(Guides for Insertion are shown further below in Example 13)

(SEQ ID NO:2)

CCAGCACCCACATGGTGGCTCACAACTGTCCGTAACCTCCAGTTCCAGAGGATCTGATGCCCTCTTCTG
 TCTCCCGCGAGCACCTGGCACACACGTGATGCACACTTAAACACATGCAAGCAAACCATCAGACACAT
 AACTTTTTTTTCCAATTTTTTAAAGATTTAGTTATTATTATTTACTTAATAAAATATTTATTATATTTA
 TTACATATACAGTTTCTGCCTACATGCCAGCAGAGGGCACCAGATTGAATTGTAGATGGTTGTGAGCC
 ACCATGTGGTTGCTGGGAATTGAACTCAGGACCCCTGGAAGAGCAGTCAGTGCTCTTAACCTCTGAGC
 CATCTCTCCAGCCCCCTCCATTTTTTTTTTTTTTAAATAAAGAAATGTAATGTCTAAGTGGGGCTTAGA
 GAGTGGAAAGCAGATAAAGAAAGATGGAGTTAAGAATTTAAGAAGCCAGTTGGCGGTTGTGCATGCCA
 GCACTCAGGAGGCAGAGGCAGGTGGATGGATCTCTATGAGTTCGAGGCCAGCCTGGTCTACAGAGAGA
 GAGTTCAGGACAGACTTCTCCAAAGCTACAGAGAAACCCTGTCTGAACCCACCACGACCACCACAAA
 GAAAAAAAGGATTTCAAGAGGAGAGCCAGGTTTATAGCAAGAGAGAAAGTTGTGAACTAATGCCCAGG
 GCTTAGTGTGGCCTACCTCTGGGCTGGGTCTCTCTCTGAACACAGGGTGGAGCTGCCCCGGGAGGAAG
 AAGCGGCTCCGTACAGTCCCGAATTCTACAGTGGCTGGGAGCCTCCCGCCACTGACCCGCAGGGCCGC

GCCTGGGAGGACCCGGTGGAAAAACAGCTACAGCATGAGAAGAGGCGCAGGCAGGTGAGGCAGGGTTG
CCGGGGGAGCACTGGGCTCCCCGTTTCTGCACAACATGGGCGAGCAGGACGTCTGAGGTCTAGCCTGC
CTGACCCCAAGCTCTCTCTCTTCCCGCAGCAAAGCGCCCCCAGATCGCTGTCAATGGGTGAGTGACC
GCTGCAGGGTGGCCAGGGATGGGGTTGGGAGGACTGAGTCCCGGGGTACCCCGGCTCTGACTCCGAC
CCTCCCCCTTTTTCTTGTCTTTTTTTTTTTTTTTTTTTTTTTTTTAAACCTCTGCCTTCCCGGCTCTT
TGCAGGTGGGTGAGGTGGTGAGGAGGCGGGCTGGGGTGGGGTGGGGGAGGAGCCAGGAGGGAGGGG
GGGAGGAGCCAGAACTCTGGGTCCAAGGGAAGAGGGAAAGGAGGCTTAGTTTGCTGAAGCTATGAGA
GTTAGGGGCTGAAAGTGGGTGGGTCTAAAGGCTTGGACCCACACCCCCACCCCGGCATCCTCAAAA
GATTGAAAAGGTGCAGTTTGGTGTTCTAGGACCTGGGAGAGCACCATGCTTGAGTCCCCAGAGCACAG
AGCACTGGGTGTGAGAGAAAAAATGAGACCAAAAAGCAGGGTTGGGACTTCCGAGGATTCA
GGGACAAGTTTGGAGAAACGTGAGAAAGTGCTGGCATCCCTGGACCACTAACTGAGGTGGGACTTCCG
GCTTCCCTAATGCGCAAAGGAATAGCACGTACTGAGCAAACCTGGAATGCTCCAGGGCTGAAAGAATGG
AGGAAATTGAAGGTCAAGGCACGGAATCCTGCCTAGGTCCCTGGGAAGGAAAGAACTAGGGACCTAAA
TTTACAGTTCTACCAAACCTATGGAAGCTGAGGGCTGCAGGTCCAGGTGAGGAAGTGATGGAGAGGGGG
TCACAGCCCTAGGATCCTTGGGGAAATAGGGGCCAGGAGTGGAGGGCGTGATGTGGCTTGAGAACAA
AATGATAGACTTGAGGAGAGGAATTGGGGCCTAGGTGAGAGCCCCAGCAGAGGGTCTCAGCAGGGA
CGGCATACTGGGAGCTGTGAGTCCACACATGGGGCGCCGAGGCCCTGAAGAGTCCCCTCCTCCTTC
CACAGGTAGGCTGATCCGGGATGAGGTCTCTCTTGCTGGGGCGCCAGAGCTAATCGTCCCCAGGC
TGCTGGTGCTGCAGGGCCCTCTTGTCTGTCTGTCTGCTTCTGAATCTTGGGCTCAGCACCTGCAAGC
TGTTTACTCGCCTTCTCTGGCTGTAATTTCTTTCCTGGAAGGGTGAGGACTCTCTGGCGCTGTAAGG
GGCTTGCAAAGAGCTCAGTGCCGTGACTCAGCCTGAGTTCAAATCCAGCTGCATGAAGAACAGTACAG
AGTGACCTTGACAAGGGCAGCCTAGGGCCAGCTCAGTCACACCTTTCTCTTTCTTGTGCACTGGCCGT
TACTACAGTATCCCTCGGTTCCCTTCATATAGAAAGAGAAATAGTGAGCCGGGCAGTGGTGCGCACAC
CTTTAATCCCAGCACTTGGGAGGCAAAGGCAGGTGGACCTCTGTGAGTTCAAGACCAGCCTGGTCTAC
AAGAGCTAGTTCCAGGATAGTCTCCAAAGCCACAGAGAAACCCTGTCTCGAAAAACAAAAAGAAAA
AAGAAAGAAAGAGAAATAGTGAGACCGGCAGTGGTGGTGACGTCTTTAGTCCAGCACTGGGGAGGC
AGAGGCAGCCGATTTCTGTGAGTTCAAGGATAGACTGGTCTACAGAGTGAGTTCCAAGACAGCCAGA
ACTAAACAGTGAAACCCCTGTCTTGGAATAAAGTGAATAATGGCCATATTCTGGTGATGGTG
TAGGCCGTGGTCCAGCTACTCAGAGACATGAAGCAGGAGAATAAAATCAAGGCCTGCTTTGACTA
CAAAGTGAGCTTCAAAGGCCAGCCTGGGCAAAGCAACAAGGCCTTGCCCTCAAAATGAAAAAATAAAAA
TAAAGAGGCTGGAGAAATGGCTTAGTGTTAAGAGTACTGGCCGCTCTTCCAGGGGACAGGGTTCA
ATCCCAGCACCCAGACATACAGCAGCTCACAACCTCCAGTTTCAGGGAATCCGGTGTTCTCTCTGGTC
TCTGTAGGCACCAGGCACTCAAGTTGTGCAGACATAAAATAACACAGAGGGCTGGGCTGGGGCTCAGT
GGCAGGCATTTGCCAGAATCCCCAGTAAAGACATAGCTCAGTGAATCCAGAGCTGAGGGGCTGGGC
GTATATTAATGGTGGAATCCTTGCCTAGAATTCAACCAGCGAAGGGCTGTGGCCGTGGCTCGGCTGTA
GAACCCTGTCTGGTATCTACCATGAAGGGCTGGGACATGGCTCAGAGATAAAACACTTGCCTAGACT
CTACCGCTGAGAGCCTGGGGTGTGGATCAGTGACAGTGGCCGCTAGCATGCACAAGGCCCTGGGT
TCAATCCCCTGTACCACAAAAAAGGGGGGTGGAGGGAGGGTAAGAGTGAGATCTCAGGAGAAGGA
AGGAACCAAAATTCATGGAATAACAAGGGAATCCAGGAGAATCGAAGCGTTTCTGGCGTACGTTGCTG
TGTAAGCACAAAGGGTCGGCTATTTTGCACCCTGTTTATTATCCTAGCGGGTGATGGGAATAGATCTG
CTGTCTCTAGCCGATTCTCATGATCCTCACTGATGAAAATGCAGGTGAGGGGCTGGAGAGATTAAGA
ACACTGTCTGCTCTGGCACTGGACCTAGGTTTATTAGCTCCCCACAGCACATGGTGGCCACAAAATA
TCTGTAACCTCAGCTCTAAGAACCAGGTCTAGGACACCCTCTCCTGGACTCTGTGGCTACTGCACAC
AGGTGATGCACATACACACATGCATGCAGGCAACACACACACACACACACACACACACACACACAC
ACACACAATGCATGTGAACGACTGGGGATGAAGCTCGGAAGCTAAGCACTTCCCTGGCATGCACGGGC
CCTGGGTTCAATCCCCAGCACCCCATGAATTAAATCGTTATCATGATACGGTGTGTTTACTGCAT
GGTGCCAGGCAAGGAAATGAGCTAACTCCATTCAAGCTGTGACTCCAGTGTCAGCCCTGTATTAACAT
ATTAACCTGGGCCTCTGCTCTGACCCCTGCTTGGCTCTAACCCACCTCACACCTTAGAGTCCAGAC

CAGCAGGGCTGGCTACCTCCTAATCTCCTGCTGGTTTCTTTCTCCCCAGTCATCAAGATCCAGACCTG
GAAGCCGCCGAGCTAGAAGAGAGAGCCAGAAAGTGGGTTCTGTGTAAGTATGACTTCCAGGCCCGAAA
TAGCAGCGAGCTGTCTGTCAAGCACGGAGATGTGTTGGAGGTTAGCGGTGTGGGGGGCCTGAGACCTT
GAAATTGGTCAATTTAGCCCTAGGTATAGAACCGGAGCGTGAATTCTCTCCTTATACGCCACCTAGGT
CCTGGATGACAGGCGCAAGTGGTGAAGGTTTCGGGACCATCAGGGACAGGAGGGTTATGTACCCTATA
ACATCCTGACACCCACCCCTGGACCTCAGGTGCACCGCAGCCAAAGTCCTGCAGGAAACCTAGTAAGT
CGGCGTGTCTTCTGCTTCTTCGGGGAGAAAGGGGGGCAAGATCCTAGGTCCTGGGGATGAGGACAGAGA
AAATCAGGTGTGAAGGTTGCTGTTTGGAAAGGGGGGGGGTGGTCAGATGTTTATTGGGAAAGGAGCT
GGAAGCCTCTCTTCAATTCCTTCCAGGAGACGAGTACTCCTCCTCCCCACCCGCACCAGCTCCAGCC
CCTGCTCAGGTGCGACCCCACTGGGACAGTTGCGACAGTCTCAACAATTTGGACCCACGCGAGAAGGG
TGAGTGGTGGAGCGTCACTCTGGGAAGTGATCCTTGTCTTCGCTTTTCAGGCTCCACCCTGGGCACCC
TAGCGGTCCCAGCCCCCTGACCCAGAACCCCTGAGCGCGCACTCCCTCCGCCCCCCCCCTCACG
GTTTCGCTTCTGCAGAGAAATTTCTCCAGATGCTCAGTGTCAATGAGGAGCTGCAGGCGCGCCTTGCG
CAGGGCCGTTTCGGTCCCAGCCGGGTAGCCCCGGGACCCCGCGCCCCGGAGCCTCAGCTCAGCCCGCG
CTCTGAGGCTCGGTGGTCCGTGCCTGGCTGCAGACCAAGGGCTTTAGCTCGGGGTGAGTGGGGCTCC
CCCCGGGGCTAGTCTGAAGAGACCTGTGCTTGAAGTGAAGGCGAGGTTCCCATTTGGTCCAGGGGTGG
GGGCGTGGAACTGTGGAGCAGGCCCAAATTGCAACGCCCAATGCCAGGGACAGGCTCCAAACGGAG
GCCACAGGAAAGGAAGTCCCATCCCTTTCCGAAGCCCCAAATCTCCAAGAGTTTGAACATCCCCCCC
TCCCCCAGCTTCTTGTGTTTGAAGTCTGATTGCACAAGCAGCTAGGTAGGTGTGGCGTGATTGGTG
GAGGGCCGAGGGAGCTTGATGAGCTGTGATGGCCCTGCTGCCTCGCTCAGGACTGTGGACGCGCTCG
GCGTGCTGACCGGAGCACAGCTCTTCTCGCTGCAAAAGGAAGAGTTGCGGGCGGTGTGCCCCGAGGAA
GGGGCGCGGTGTACAGCCAAGTCACCGTGACGCGCGCTGCTGGAGGTGAGCGAATCCTTGGGGCC
GGACAAGGCGACGGAGGGTAGGGTGGGGATGGGGACCTGGGGGGAGGGGGTCTCCAGGGTTACAT
ACTAAGATCTTGATTTCTACCCCGCTCTGCAGGACAGAGAAAAAGTGTGCGAGCTGGAGCCGTGATGG
AGAAGCAAAAGAAAAAGTGAAGGCGAGACCAAAACAGAAGTTATTTGATCCTTCTGACTCGGTCA
CAAAACGTGATGGCATGGCGGGCTCCCAGCGCCCCCTAGGACAACAGTCGCCAGACTCCTCCCCGTG
ACCGGGACAGTAGATGTCCGAAGGATCGCCACCCCTCATCTCCCGGCTCACTCGCTCGCTCGCTCT
CCTGGCGGGCAGGCTGCGCTGACAGTGCCGGCTGGAATCCTTCCGGGGGACCTCAGACTGACGGGGAC
GGGGACGGGGACGGGGACGGGGACGGAGCATAACAGACACTACCAGAGAGGCACGCCCAAGAGGCGCAC
GGAGGGAGGGCCCTGGGCGTCGTGACGTGCTATAAACAGCCTCCTTTCTAGACCATGCGTGTACCTG
CTGTCCCTTCTCTCGCCGGCTACCCAGGAGCCAGGAATCTGAGAGATGCCCCACGCTTCTCCCCAT
AAACCTGGAGAGTCCAGCCCAGGCTTCTAATCACCAGTCTATCCTCGCACTGGCCCCATCTACATCC
CTTCTCCTGTTCAAAACCCCTCGCCTGGCTGGCTCCTCGTTGTTCTCAGTCTGTCTCCTGGTGTAA
GGCCTGGGCTTTTCTCATTGTCTCCGCCACCCTGCATTTTCGGCCCAGCCGCTCCAGACCACAAGCGG
TTTGCACTTAACGCTTCTGAGGGTTGGAGCGGCCCCCATCACCTGGCTCGGCTCTCCTAGCCACACC
GTGGACACCCGTGTCCAGCCTCTAAGGACCGGCCATGCAGATCTGGACGCTCCCGGGCATGCCACGG
GCTCTTGGTTCTTCTGGCCCCCTCAACAACCTTCTCCCTGCCAAGCCCTGCAACTGTCCAGGTTATG
CAGGTGGATGGTAAGAGCCGGTTTTCTCATCCGCGCTAGGTTTATCTAAGGCCTTTCTTTTCCCTGCA
TCCTTGAACACTCCCAAGAGTCCCACCGTTGCAGTCGGCCTCTGCTCCCCGCGCAGCTCAGTCCTTA
CCTGGGCCACCAGGTGGCGCACCTCGAATCTGACCCAGGAGGGCCAGCCTTGGGCTGACTTCACTAAG
CCCCCTTCTCTTCTGGAACACTGTAGCGTTCCAGTAAGCCTTTAGTGTCCATTCCCTTGGTTTTCTCT
GGTACATGAGATAAAACCTAACTCCAGCATGACAGCCGATGGCCTGTGACCCCTATGGGCTCAGGTGCG
CCCTTCTCTCTGTTTCGGGACTCCAGGCACTGGTCCATGCTGTTGGTTCTGTTGGGATGTCTTGGCTC
CATGGTGTCTTATCACTGCCTGGGGCGTCATTTCTTATGTGCGGCTTGGTTGGTTTGGTGGAGCCGT
CTGGGTACAGCCCCAACTCTCGGTCTCTCAGTTTTCAGTTTCTGCAATGTGGGGATATTGGCAGGCGC
CCTGCTGCCACCTCTTTTCTAATCGAGAAACCAAAAGTACAAGCAGTTGCCAAGCTGTTTTGATTC
CGGCAGTGAGGTCCAGACTACAGACTGAAATGCCAGCAGGAGCCATCTGGCTGTGGGACATCAGG
TGATCAGGTGCTGTGGCTGGCTCTCTGTGGTTTGGAGTCTGACCTTTTCATCCTGACTTGACCTCT

GTTCGATCACTTTGTCCATCCATCACTCCCCAAGTCTACATCCAGCCAGGGGCACCTGTGAGAGCTCAA
GCCGGATGGTAACCTGGTGGTCAGGCCTCCCAGCTCAGGTGGAGCTCAAGTTCTTAACAGAGCCATGA
TCACACACAAAGCCATCACCTCAGCGCCACAGCACGCCAGGCCTGCTCTACCCCACGCTGCACACGGT
TCTCATCATCATGCAAAAGGTGCTTCCTTCAGATACAGGGCTCACCGTCACCTTCTAGCATCTGTCTG
TGCAGCTTGTCTATGGGGCCTACTTTTGAAGTGTCTATAAACACCACACACGCACATATATATACACACCA
GATACACACACACCACACACATGCCCAATACACTGTGCATGCGCACACACAAACACACACACATACCT
CATACACCATACACCTATAAACCACACCAGCCATACCACACACCACATATACACAGTTACCTCAGA
CAGCATGGCACACCACACACACACACACACACACACACACGCGCGCGCGCGCACACACACACACACAC
ACACACACACACTCCGCACTCTCCCTTCTCCACAGCACTGTAGCTGAAATCCACACAGTGGCAACCT
TCCTCAGTGTACTGGCTGCTGGACCAAGCTGTTCACTCCTGTGACGCCAGCTGGCAGAACAGCCCATT
CCTGACTGTGAGGATGGAGGAGGCACCGATCCATCTCAAGACTGATTCTGGCTCTGCCCCAGT
CACTGTGGCCACGAAGGACTACTTACCATCACCTACTCCTTTCCCAGAAAACCTAGACTTGCAGTTTC
CTATGTTGGCCATCCTACCTTTTCAATGTTAAGCCACTGACTCCGCTCACTTCCAAAGCACTGAGGGT
CAATGTGAGCACCCGGATCAGGTACAGGCTTCCTTCTGACCCCCCTACCTACCTGGGGCTCTTTTC
TCTCCAGCTGCTCACTCGAGCAAGCTCCCCTCCCCACACCTGTGAGCAAGCTCCAGCCACCCACTGG
CCCTCATCCAAATGGATGAGCGGTTTCAGTCAGATACACAGGCTGAGTATACAAGCAGGAACCACTGC
CCCACACCCAGGGGGAGACAAGTCACTGAGTGGCAATGTACGACTTTATTTGTGGTGCCTGTGCTTT
GTCTCAAAAATACCTTCTCCCCCTCCCCAGACAATGGGTGGGAAGGAGGCAGCAAAAATAGAAGACAA
CCCTCCCTATTGCACACGGACCCTATATACAGGCCACCTGGCAGAGGCCAGTGGGGCTCTTGGCACA
TTCTTGATCCCTGCTGGGGAGGGAAGGGATACTGGGTAGCATCACACGTGAGGTGGGCCCCGGGGCAG
CCACTCTGCTCCTGGATACTGATCCTGGCTTCCTTGGTCTTCTGCTTCTTCTGCTGGTCCCATCTCTGGT
GCCTGCCCACTCTCGGCAACATTTCCCTACCTGGCTCAGCCTCCCACCTCCACCCTGGTTCTGGGGAC
TCTGTGCTTTCTCCGGGTTCTGAGGTCCCAGAGGAGGTTATGGCTTCTCAACAACCTTCCCCCGGAG
CCCTGTCACTCATGTTCACTCGGGGAAGGGGTGCGTGTGTCAAAAGCAGCTGTATAAATACGGTGC
GGAGCCCCCTCAGAGTCACTTGGAGAGCTTGCTAATGACGCGGATCAGTGCTGCATTCTCATCCTTGA
GCCGCTGGTTGTGACGCGGAGGTGCGACAGGGCCTAGGGGGCAGGGTGGAGTCAGCTGGGCAGGGCG
GGGCAGGGTGGGCTCTGGCCACCGCCCTTACAAGCTCGTTACCTTCAGCTCCTCCTCCAGCTCTGCG
GCCTTGCCTCCAGGGCCCTGCGCTCCTGCAGGAATGGGTGGGCTCAGAAGCAGGGTAAGGGCAGGG
GACAGGGCAAGGGCGGGACACACCCAGCGGCCCAAACCTCACGAATCTCTCCAGCTCCAGGAGGGCG
GGCTCTCGGCAAGCGTTTCTGCGCTGGTAAGGGCAGAGAAGACTGGGCGTCAGGAGCTGCTTCTT
ACCCCTAGGACATCAGAGCCCTGCCCCCCCCCCCCGAGTGGGGGACCTCCAACCTCCCAGCCACGGCC
AGGCCCCCTTGCCACTGGGGCTCTGACTCCCCTGCCCCAACAGCTGGTTCTTAGGTCTCAGTATCTGC
ACCTGCGTGGCCCGCTCAAGCTCCACCTTGAGCTGTGCCAGCCGAGGGTGGTCTCTGTGAGGGCCTC
ACGAAGCCGCTCGTTCTCCCTCCGAAGCTCCATGTACAGCTAGGGACACAGAGGAAGCAGGCAGGCTC
AGAAGGGCCCCGGAAGGGGCCAGGACAGGGTGGGGTGGGGCAGGAGGTAGCATGCGGCACCTTCCGGA
AGCTTCCATCGGGTCTTCTGTTCTGCTTGGATTCTGGATTGAGGTCTCTGTCAAACGCTGTCTA
CGGGCAGTGGAGCCGCCATCCACGGTGTGACAGAAATTCAGGCCTTAGGGCCAGGCCCTGCCCGA
GGGGTGCCCCAGCCCCACGCATGACCCGGCCTACCTGCACTCCAGGCTCCGTTCTGCGGGCCCCGCC
TCCTCCCCCTGCAGAAGAGCCCTGAGAGTTCACTCTCCATGCAACGTCTCCTCCAGCCCGCCGGC
CTCCACACAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCCCGCCGGCCTCCACACAGCATCCTCA
CCTCCGCGGGCCCCCTCCCTCCAGCCCGCCGGCCTCCACACAGCATCCTCACCTCCGCGGCCCTCC
TCCAGCCCGCCGGCCTCCACACAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCCCGCCGGCCTC
CACACAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCCCGCCGGCCTCCACACAGCATCCTCACC
CTCCGCTGCCCTCCCTCCAGCCCGCCCGCCTCCACACAGCATCCTCACCTCCGCGGCCCTCCCTC
CAGCCCGCCGGCCTCCACACAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCCCGCCGGCCTCCA
CACAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCCCGCCGGCCTCCACACAGCATCCTCACC
TCCGCGGCCCTCCCTCCAGCCCGCCGGCCTCCACACAGCATCCTCACCTCCGCGGCCCTCCCTCC
AGCCCGCCCCGCTCCACACAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCCCGCCGGCCTCCA

CACAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCCCGCCCCGCTCCACACAGCATCCTGACCTCC
GCGGCCCTCCCTCCAGCCCGCCCGGCTCCACACAGCATCCTCACCTCCGCGGGCCCTCCCTCCAGCC
CCCCCCCCCGCTCCACACAGCAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCCCGCCCCGGC
CTCCACACAGCAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCTCCGCGGCCCTCCACACAGCAT
CCTCACCTCCTCGGCCCTCCCTCCAGCCCGCCCCCCCCCGCTCCACACAGCATCCTCACCTCCGCG
GCCCTCCCTCCAGCCCGCCCGGCTCCACACAGCAGCATCCTCACCTCCGCGGGCTCCCTCCAGCCC
GCCCGGCTCCACACAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCCCGCCCGGCTCCACACA
GCATCCTCACCTCCGCGGCCCTCCCTCCCGCCCGGCTCCACACAGCATCCTCACCTCCGCGGGCC
CTCCCTCCAGCCCGCCCGGCTCCACACAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCCCGGCC
CCCCCCCCCACCCTCCGCTCCACACAGCATCCTCACCTCCGCGGGCTCCCTCCAGCCCGCCCGC
CTCCACACAGCATCCTCACCTCCGCGGGCTCCCTCCAGCCCGCCCGCTCCACACAGCAGCATCT
CACCTCCGCGGCCCTCCCTCCAGCCCGCCCGGCTCCACACAGCAGCATCCTCACCTCCGCGGGCC
CTCCCTCCAGCCCGCCCGCTCCACACAGCAGCATCCTCACCTCTGCGGCCCTCCCTCCAGCCCGC
CCGGCTCCACACAGCATCCTCACCTCCGCGGCCCTCCCTCCAGCCCGCCCGGCTCCACACAGCA
TCCTCACCTCTGCGGCCCTCCCTCCAGCCCGCCCGGCTCCACACAGCATCCTCACCTCCGCGGGCC
CCCTTCGCTCGTGGCGACCTTTCGATGCTCCCTGGCGCTGTGGTCCCTGGCCGTGCCCGTGGGC
GCCTCTGTGGGAACAGTGGGAATCAGCTCAGACACCACATAGGGGCCCTGTCTACTGTGCAGG
GAACCTGACTTAGCCCCAGTGAACAAAGACACTTTATGGGGAGACAGGATGGCTCCCTGGGGAGCGA
CTTCCAGAAAGCCGACCTCACCTCTCTGGGAGGGCCCTCAGCGTTCTCCACGCCAGGGACCCGAGG
TCTCCGAGAAGGGTCTGCGGGGAGGAGCAGTCAGAAACAGGGAGACGGGTCCACCCGCCAGT
TCACCTGACCCTGCTCACCAGGCTGGGAGGGCAGGCTGCTCTGGCTCTGGAGCCTTCCCTGCCACCT
TCTCTGCTTCTTCAAGTCTGTCAAGGTACGCCCTAGTGTAGAGAACTCGGTCAAGGAAAAAGGGCC
TAGACTTCCACAGGACTCAAGTTCAAGACCCCGGCCCTCCCTCCCTCAGACCCGGGAGCACAGCCCCAG
CCCCATCCACACCTGTGTAGACCTCCGAGACTGGCGCATAAGGCGGGAGCGAGCCTTCCGCTGAGAC
TCGGACTCTTCATCACGCACAGGCATCTGGTAGGACCTGAGTGGAGAATGTCCCTTGGGTGCTGTG
CACTGTGAATGACACCTTAGGGGAGTGACATTCTGGCAGAGAACGTGTCAACTGGGCAAACAGGACC
CCAGGAGCCTACCCAGAGCCCCAGAGACCCCTAAACACTGTCTTCCCTAGCCTCTTACCTCCGCCG
ATCCCTGGAGTCAGGTAGGGCTGCTGCAGAGGCTGCGAGGGCATTGGCTTCACTGGGGATCCAGGCT
CCCTCTGGAGGATGGGGCGGAGTGATCCGAAGAAGGAGGCATGGCAGCCTCACACCTGTATGGATT
ATTCAATCATCAGCAAATATTCTCAAGCCCGCATTTCTGTGTAGGCATAGGAGAGACCACAGAGAAG
GAGCCAATCATGGCTGCTGATGAGCCATTTCTGGGCAAAACAGATAAAACAAAACAGCAGCCAAAGAGA
CCAGTGTGGAGCTTGGGGAGAAAAGGTGCTTGGAAAAATAAAGAGAATAAGCAATTATTTGATGCAC
CCTAAGGGCTTTCTCAGATCTCAAATGCCAGGATGGCACCAGACCTGTCCCTTGCCCCAGCCACTGG
TACTTACAGGGTAGAGGGCTCTGGCACTTTCTGGGAGGGGTAGGGTTATTCTGGCAAGACGGGGTT
CCCTGGCTGTGGAGACAGGAGAGAAGCAAAGGAGGCACTGTCTGCCCCAAGGCAGGAGCCTGTACCC
CACACACTTACGGCACCTACCTGAGAGGAGGCCCTTTCTAGGAGGGAGGAGGAGGCTGAGCGCTGCA
GACCGAGAACCCCTCTGCACCCCTCCTCTCTGAGGGACCCAGGGCACCAGAGCTTCTGTCTTCTGG
AGACCGCCGCGCTGGAGAAGGGAGCCTCTTCTGGCGGCTGGGAGAGGAAGAAGGTCTTCACTACTGA
GCAAAGCAATGACCCTTCTCCTCAGAGCCTACGCGTGTAACCTCAGGGGAATTACAGTAAACCACAGC
CAAAGCAATGACCCTCCTCCTCAGAGCCTACGCGTGTAACCTCAGGAGAATCACAGTAAACCACAGCC
AAAGCAATGGCCCTTCTCCTCAGAGCCTACGCGTGTAACCTCAGGGGAATCACAGTAAACCACAGCCA
AAGCAATGACCCTTCTCCTCAGAGCCTACGCGTGTAACCTCAGGGGAATCACAGTAAACCACAGCCAA
AGCAATGGCCCTTCTCCTCAGAGCCTACGCGTGTAACCTCAGGGGAATCACAGTAAACCACAGCCAAA
GCAATGACCCTTCTCCTCAGAGCCTACGCGTGTAACCTCAGGAGAATCACAGTAAACCACAGCCAAAG
CAATGGCCCTTCTCCTCAGAGCCTACGCGTGTAACCTCAGGAGAATCACAGTAAACCACAGCCAAAGC
AATGGCCCTTCTCCTCAGAGCCTACGCGTGTAACCTCAGGAGAATCACAGTAAACCACAGCCAAAGCA
ATGGCCCTTCTCCTCAGAGCCTACGCGTGTAACCTCAGGGGAATCACAGTAAACCACAGCCAAAGCAA
TGACCTTCTCCTCAGAGCCTACGCGTGTAACCTCAGGAGAATCACAGTAAACCACAGCCAAAGCAAT

GACCCCTTCTTTCAGAGCCTAAGAGTGTAACCTCCAGGAGAATCACAGTAAACCACAGCCCAGGCAGGT
GCCACCAAAAAAAAAAAAAAAAAAAAAACATTACTTCTTGGTCCACAAGGACCTAAGAACCAAGTCAAAA
AGCCACTTTTCCTCAGCGGAAGCAGAAGTATTTACCGTATCCCACCCGCTGCCCCAAACCTCACATCTG
CTCAGGGCGCTCAGGCTCACCACAGGGCTCTTGGGGCTGGAGGACACAGGAGAAGACACGCCATTGAG
GGCTCTTGGTTGCACAGGAGGGTGATCTGTGTGCAGGAACAGGAGAGGGGGGTACAGGAGAGGCCGG
CCGCCTCTGAGATTGGGGACCCACAAGTCCAGCTCCTTCCTCAGACCCAGGGTCCAGCATCCCTACCA
GCTGCCTCTTCTTCTCCCTCATCTCATCCCCAAGAGAGGGGGCCCGCGCCCCACCAGGCCGGCGCTC
CTTCGACAGATCCTGCAGGGAGATCTTCTCACGGCTGCTCAAACGACACACGGAGCTCCTAGGAGGAC
AGGGTGTCCGTGTCCAAGTCTGGGGGCGAGTCCGACCCACCCAGGCCCTAGGCATCTCTTACCTTCTG
TGCTTGCTGTTGGAGGGCACCTGTGGCTCTTGGCCTCGGCTCTGAGAGGCTTCCTTTTGGTTCCGAAG
CTACAAGGATGGAAGGGGGCAACTGGGGAGGGGCAGAGAGCACAAAGCCCTCCAGGGTCTCCTGGCCGC
CCCCCTCTGTGCCACCTCTCCACCTCGAGGGCCATCACGCATAACTGGGCTAGTCACACTTTATGCAGG
GTCTTGCAAAACATGGGGGACTCAGTAACCCGGCAGCACACTGGCTCTGGGGCTTATTCAGGCTCTCCC
AGGCTTGCCCTGGTCCAGCTGTCACTGCCTCCAGCCTCATTCCCAGGGGGATTCTGTCTTCTTCCCAGG
AGCGAGCACCTTGCTCAGACTTCCCCCTACCCTCCAGCACATCCAGGGCAGGACAGGGCAGGTGGCTC
TTTCTGGTTATCACAGGCCAGCTCTCAGCTCAAGGACAACGGCCACCGTCCCATATAAGCAGTCTGG
TGTCGTAACCCAGGAACACCTCTTGCCCATGCCCTCCTTGTCATCCAGTGTGCCACGGGACTCCTCT
CTGGACAATGTTCCCGATGGTTCCACGAGGCCCGGGCCACCTCACTAAATAATGGAATTGCAGCCATG
CCGTCTGCTTGGGGCCACACCCATGATGCCTCACTCTCCACTTTCCTAGCAAAAGTGCTAACTAGAGT
GGGGGGGGGTAGATACAGGTTCAACCTGTGTACACACAGCTGTCTTCCCAAGCGAGCAGGCAGGAA
ACTCTGGGCATAGCCTCAAGTCTCCAGATATGGAGGTGCCTCTGTTCTTAGCCCTCCACCAGAGCTG
GGCTGACAGGTGGGAATAGCGGGTCTCAGTACTGAGGGTGTCAAGGGACAAAGACTGTGAGCCCTCCC
GGTTACTGTTACCTCCTCAGAGCTGCCAAGTAAAGAGGCAAACCTAGAGTCGAGACTCACGTCTCTCTG
TTTCTGGGCCAGTTCTCCAAAAGGTTTCATCACTTCTCATCAGCCAGGTACAGGGCCCGCTGCCCCCT
GAGTAGGAGAAGGAGGCAGATGACGGTGATGGTGGTGGTGTAGTAGGGGCTCCCCCGCCACCTGCC
CACCATCTGAGATGGCCCTTACCGCATGGGTGAGCGAATCCATGCCCCACCGTGCTCAGCCAGGAGA
CGGCAGGCGTCTCCACACCCCAAGTGGGCTGCTGCGTGCAACGGTGTCCAGCCATCTCCATCCCGGAG
CTCTGTGTCGTAGCCAGCTTGGAGTAGCAGCCTAAGGGCCAGGGAGGCTTGGGTGAGATGGCAAGCTA
GGCCAATGGCTGATCTCAACTTCTGTTCTGTGGCCACAGGACTACTGATCAATAACCAAGCGTTACTA
GTTTTACCAGCAACCAGCCCCACCCCAAGCTCAACTGAGCCCTCCCTTGGACCAGCAGCTACTAATGA
AAAAGCTCCCTCATACCACAGGGATCCCCTCCTCAGGCCCCAGGGTAAAGGGTTAGGGCAGTGGTGA
GGCGATGAGGTGGATGCAGGACTCCCCACTAACGCAAGCCCATGGAGAGGATGGACCCTGAAGGGGCT
GTGATGCTGGAACCACTGGAACCACGCGTTTTAGGACACGGATCCTCAACAGTGTCAAGCAGCTCTC
ACACCTCTCTACAACCTGGAGACATCACCCTAGAAATCCTAACTTACGGGTACAAGCAGGAAGCACCA
GTGTGTGGGAGCTGGAGAGGCTGCTCAACCCCTCCACGCACAGGACAGCCCTACCACAGCACGGTA
AGACCCCAAACATCACAGTGCCGGAGGAGAGCGAGCCTGGCTCAGCCTTCCAGAAGGTAACAACCTGG
AGCTCTCAAACTCAGCATGGCACGAGGCGAGGCCTCTTTGGAAGCAGTGTGATGAGGTCTGTGTG
AGTGAGGAAGGCTTCAAGCCCAGGGAGGCAGAGGTACAAGGCACAAGGTGCTGTGTGGCCCTGGGACC
CTCCTCCCTCACACTTCCCAAGATTCCCCTGTCCCCTTGACAGCAGGGCACGCTGGGCTTCTTGTTACA
TTCCACATGCCAGGGTCTCTAGCCAGCTGTGCGCTCCTTCTGGTCAGTATCCTAGGAGCCTGAAGCG
TGCCACCCAGCCACACCCCTAGTCCATCAGCACTTCTCACCTGGCAGTTTCTTACCACCATCTCT
GCCAGGGGGCCTCCCTACTGCCCACTAGTTATAGCCTCCCAAGGCCAAGGTTTTCTTTGTATAAGCTT
AGTGTATTATTTACCATAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT
TGAAGATCAGAGGACATATTACGATTCACTTTCTCCTTCCACCACCGTTCCAGGACCTAACACAAG
TCACCAGGCTCTTGTGTGGCCAACACTTTTACCTCTGAGCTATCTCACTGGTCTAGAAGCCAACGTTT

GCAGCTGGACCCCTGCTACTCCCCAGAGGACCTGTGGCAATGTCTACAGTCATCACACAACCTGGGTCAG
AGGTGCTGCAATGGACTGGACAGCCATCAGAATAGAATGACCCAGCCCATCAAGTCTCTCATTGGCTA
CGGTGGGTACACATCTGAAACACCACGACCAGCCAGGAGGCTAGCCCCTAACAGACACCAATATTTA
CCTGTACTTCAATGAGTACAATCATAGAAGACTTTTAAATACAGTCAGAAACAATAGATAACTATAAAT
TCAGTGAACAGGAGTCTAAACGCAAACTCACACAAAGGGGGCCATCACAAAATTACAAAATTCAGTAT
GATGGCTCACACCTGCAATCCCAGAACACAGAAGCTGAGGCAGGAGGACAGCTGTGAGTGCAAGGCCA
ACCTAGGCTATCTATCCAGTACCAGGCTAGTCAGGACTACATAGCAAGACCTTGTCTCCATTAGAAAA
GAAAGAAGCCAGAGGGGAGGGAGGCAAGCATGGTGGCTCTCACCTCTATCCCACAGGAAGGTGAAGGA
ACAAAGAGTAGAAAATCAAGACCAGTGAAGTAGAGGCGATCATGACCGACATGAGCTATTTATGGAAG
AGGCCAAATAAAACAAACACAAAAGTTGTCTCATCAGTGCAATTTTTTTTTTTCAGGGCTGGGACTGGAACCC
AGAACGCTAGGCAAGTGCTCTATCCCTGAGGCACCCCCCTTCCCTCACGGGTAGACACCAGGGAAG
CATCTATCTACCTATGGCCTGCGACCACAGCCAGTGCTTCAGTTCTGGGACAAGTATTGGCTCACTT
TCTCTACTAACTAGCCCCCGGACCTATGCAGGTGACACCGGGGAAAGCATTTAAGCACAAAGACAGG
AAGGAGTTCTGATCACCAGAATCCACTTAAAACTCAGTGGATAGCTGTTATAAAAAAATGACATCAG
GGTGGAGAGAGATAGATGGCTCTGCTCTTCCAGAGACCCGGGTTCAATTCCCAGCACCCACACGGCAG
CTCCAGGGGTCTGACCCCTCACACTGACATAACACAGACAGGCAAAGCACTAATTAATGCACATTAA
AAAAATAACATCATGAAATCTGCAGGCAAATGGATGGAAGTGGAAAAAAAAAAAAAAAAAAAAACA
TCCTGGGTGAGGTAACCCAGCCCCAGAAAGACAAACATGGTGTGTACTCATTTACAAGTGACATTAG
CTGTTCAAGTGAAGGACAATCGTGCTACAATCCACAGACCCAGAGAGGCTAGGTAACAAGGAGGGCTCC
GGGGAGGGACGGTGACGGATGCCCCAGGGAAAGGGAAAGAGAAAAGACTTTGCAGATGGAATGGGCA
GGTAGGGATGGAACAGGAGAGGTGGGGAGAGGGAGTGAGGGGAAATACTGGGGGGGTGGCTGCAA
TGGGGGCTCACTTTGGGGGTGTTAAGGAAACCCAGCACAGTGGAAGTCCCTGGACTCTGCAAGGGTGG
ACCTAGCCAAGTAACGAGGGACACAGAGTCTGAACCGGCTACTTTGGGTAACAGGCAAGGCTCCCAGC
AGTGGGACATCAACCCGGCCACAAAACCTTTTGACCTACGATGTGCCCTGCCCTGCAAGGTGTGCTGAGG
TAATGGTGGCGCAGAGCTTGTGGGAGTGGCCAACCAATGACAGGTCCAGCTTGAGGTCCATGCCACAA
GAGGGAGCCCACGCCTGACACAGCCTTGATGGCCAGGAGCCTGGATAGCCCAGACCTGGGGTAGAAC
CAATACAATTGGGGGAAAAGAAAAAAGGCAAGAAACAATTTCTTAATGATATTCTGCTGTTCTCATG
GATCTGTGGCTAGCCCACTGTCTGTCAGAGAGCTTTTTCCAGCAGTTGACGGGAGCAGATGCAGAGAC
CCACAGCTCAGGGAACCCACAGGAAGGATTATGGGGGGGGGGGCGCAGGACACCAGGAGAACAAA
GCCCACAGAAATCAACTAAGCAGGGCTCCTTGGGGCTCATGGAGACTGAAGGAGCTAGCCATCAGGACC
TGATGGGTCTGCGCTGGGTCTCGCCTGGTGTCTTTCGCGGACTCCTTAACACTGGGACTGGAGCTG
TCGCTGACTCTTGTGCCTGTTTGGGGACCCAGACAAGCATAACTGGTTACGCTGTGCTTGGCTGTCTAT
CTCTGAGATGCCTGTTCTTTTCTGAAGGGAAACAGAGGACTGGATCTGGAGGAGGGGTGGAGGGGAAC
AGGGCAGAGGGGAGGGAGGAATGTAATATGAGAGGAAAAACAACAACCTACAATTATTGAGTGGACAT
GGCAGCCCATCTGCAGAGACAGGCCACCCTCAGACGGAGATGGCAGCTAACTTGCCAAAAAGGCAAG
CTGAGGGATCGGCCAGAGGCCCTGCCTCAATATTAGAGTGGAGAGCAACCAGAGAAAGTACTACATGC
CAACACACACACGAGTGTGAACACACACACACACAAGTCATACCCATACACATGCACACGCGCGCG
CGCACACACACACACACACCACAACCGTTAACCAGACATATAGTTGTGTGGAAACAAACCTAGTTTTTC
CTTGCAACTAGGACTGGCCAATGGTGAGAACTGGGTTAATGGAACACAGATATTAAATATGCACACTT
CTGGAATGTTCTCCTGAAAAGGAATAGACATTGCTCCCTTTGCCTCTGCTTCCCACCAACTTGAGAT
ATAGACGCAAAGGCAGGTGAGGCAAGTCACCCTCAAGTGAGAGGCACCGCTAGAGCAGGGCGCAAGCT
CTGCACTCGGAGATTTAGGGCATCCTGTCCCCAAAAGGAATGGGCTCAGAGCGCACTGGGACTCATG
CTGTAACCTACAGAGACTGATGCCCCTCCCCAGGAGCACAACTATGCAGGCAGGCTGTAAAGTCTGGGG
GTGGCACGAGGTCTTAAATCCTGCTGGAGAAAACCTGCCTGCAACCTTACCAGTATGAAAAGCAGAGA
GGTTCATCTTAATTCATTTGGGTCTTTGTTTTTTTTTGTGTGTTTTTTTTTACACAGGATCCCTCTATA
AAGCACTAGCCTCACACTCAGTATATAGACAAATCTATCCTGGAATTCAGTAATCCTCCTGCCTCTG
ATTCTCAAGTGTAATTATAGACATATAACACCGTATCAAGCAAGCAAGTGACACACGACGACGACACG
CTCTTGTTACATAGCCTGGGCTAGCCTACAACCTCACAGCAATCCTGCCTCGACCTCCCAAGTGAGGAA

ATTAAAAGCGTATACCACCATGCCTGGCTTAATGCCATTTTTTTAGGTTGGTATTATTTTTATGCGTA
TATGTTTTGCCTACATGTATGTATGCATACAAATACACACAGACACAGAGATAAAATAAATGTAATTTT
TAAACCTCTTTGGCTTTAGGTATGTAAACCAGGAGAAGAAAAGGACAAGAGCCCCGAAAAGCTTCCAG
ACACAAAACAATCACTCTGGCCTCGCTCACCTCATCACCTCGATGTAGCCCTTGGCGGCAGCCACATG
CAGGGCAGAGGCCCGGTCCGGGGGTGGCGGGCCTCTGGCATGGCACCCCATTCAGCCAGCACCTTG
TGTCATGAAGCAGCAGTTCTTCTTCAGCCCGCTTGGCTGCCTCGACATCCACACCTGGGAGAATGAGA
GGTGACAGGTGGACTCACACAGGTGGCCTAGGAAACCCCGGTGCGGTCTCAACTAGTCACAGCCCG
GCCCCGTGACTCATCAAGTCTCTGGACCACTCAGGAGACCGGGACTGCCCCAGTGTTCCTCAACTGTG
CTCCCTGAAGACCTGGGCACCAACCGAGGGGGCCAAGACAGGCCAGGAATGGAAACCACAGGTCTGAC
CCCTGTGGGTGAGTATCCTCTTTATGTTTTTCTAATAGAAAACCCACACCGGATTCCATCTAGGTTT
TCCTACCCCTCCAGCTATAAGCTAAAGCCAGCGCCTTCACACAATGTCACTGCTGGTTCTTCTCCCTT
TGAAGTACGATAGGCCAAACAAAACCTTCACTACGGCGTTGTACGTGGTGGCTCCGGCCCTCTATTCCAG
ATCTCAGCCTTGGCAGGATGACCGGTGGCCTCGAATCTGAGAACAGCCTGAGTACATACATGGTGTG
AAGCCAACCAGGACTAGAGAGACAGACTCTGTCTTAGACAACAGTAAAACTAAAACTCAAAAGCTTC
TGGGCGGTGGTGACACCATTAATCCAGCACTCGGGAGGCAGAGGCAGGCGGATCTCTGTGAGTTTCG
AGACCAGCCTGGTCTCCAGAGTGAGTGCCAGGATAGGCTCCAAAGCTACACAGAGAAAACCTGTCTCG
AGAAAAAAAAAAAAAAAAAAAAAAAAAGCTATTTCCCAAACCTATTTGCATGCATAGTTTCATTCTTGCC
CAGATGTCCAGGCATTTGACACCTCGCTGGCCACGACAGAAGTGAGAAGTGAGTGAAGTGCCTTGGCA
CTTTGTGCTTATGCGGGTATGCTGCATGCCTGTGACCCCAACACAGGCAAGAGGCAAGAGACCAGCAG
GGCTCCACAGAGACCTGAGTCAAAAGACAAACAGAGGGGGAGGGGCTGGAGAGATGGTTTTAGAGGAT
GAAGTGCCAAAGCCCGATGACCCAAGTTCAATCCTGGGAACCTCATGAGGCAAAGGAAAGAATCAAGTTG
CACAAGGGGTTTTCCCTTTGTGAACCCAGCTTGGCCTCAAACCTCACAGCAATCCTCAGTCTCTGGAAAG
CTGAGATTAAGAGGGGGTTTTTTTTGTTGTTTATTTGTTTTTTTTTGGTTTTGGTTTTACGAGACAGGGT
TTCTCTGTGTATCTTTGGAGCCTATCCTGGCACTCGCTCTGGAGACCAGGCTGGTCTCCAACCTCACAG
AGATCCACCTGCCTCTGCCTCCCGAGTGCTGGGATTAAGGCGTGCACCACCAACACCTGGCTAAGAT
TAAGAGTTACACACCACACCTGTACTCCCTGCACTCAAAGAGGCTGAGGCAGGAGGATTGCTCCAAGT
CCAAGGTCAGCCTGGGCGCCAGCATGAGAGCCTGTCTCAAACACCTCAGGGGGGAAACAGAAAGCCAG
GCAGACTAGCTGAGGCTGAAGCACTCCAGCCCTCACTGTGACCCTCATCCCTTAAAGCACCCCTAACT
CACTGAGACCACAGCAAAATGGCCTCTGCTGAATAACTTCCTCCTGGGAAGGTTATTACTGCCCATGC
TTTTGCAGTTGTGAAACTCTTGACTTGCCGAAGTTCCCTCAGAGTTGAGCTGTTGTATCCAGTAGCCGG
CAGCTATGTGGAAGTGGCGAGCGAGCACTCGAAACTGAGACATGCTGTGAACGTCAAGTGCACTCGGG
ATTTCAAAACACAGGAAAAGGTAAGGGCTCTCGTGGACAGTTGTCTCTAAACTGTTTTGTGCACACGT
GCATGGCAGCGTGGTCAGGGGGCAATTCTGAGGAGCCGAGTCCTGCTTCCCACCTTGCTGAGGCTGGG
TCTCTGGATTCTGCGGCTGTGCTGTGTACTCTAGGCTACCCGCCCCACAGGGCGTCCCCACAGTTCTC
CCACCTTCCTCTAGGACCTGGGAGTACGGACGTGCACCAGTGCCGCCGCCGGCTTTTTACAAGGGTTC
TAGAGGTGGTAACCTGGGTGCATCTAACATCTTTACTGGCTGAGCTATCTCCCAGTTCCCCTTACTG
TGTTGATTGCACCCAACGGAATACTGGGTTTGTTTTTTGTCTGGAGCGTGCGTGAGAGAGAGAGAGA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATACAGACACAGAATCT
TAACAGAGAGACAGAGTCTTACATTATATAGTGGAGGCTGAATATCTAGTCTTCTGCCTCCCAAGTA
CTATGGGATAGTCTTCTGTACGCTACGAATATGGACTCTTCTCATTGGTTAATAATAAAGCTGAGTT
GGCCCAACAGCCAGGCAGAATAAGGGTAGGCGGGAAAGCCAAACAGAGATACAGGGAGAAAAGAGGGCA
GAGTTGAGTGAGACGTAAGCAGCCACCAGGGAAGCAAGATGCCAGGTGACAGGTAAAGCCACGAGCCA
TGTGGCAAACACAGGCTAATAGAAATGGGTTGATTTAAGTTGTAAGAGCTAGTTAGTAATAAGCCTG
AGCTATAAGCCGAGCATTCCGTAATCAATACGAGCTCTTGTGTATTTATTTGGGGCTGGCGATTGGA
ACTAAAGGGAAGCTTAGACTACGGACTTGACCTCCATGCTTAGTTTATGGGTTTCAAGGACCATGCT
AATGGATAAGCACTCTACCAACTAAGCTACACCCCAGCCTATGGCTTGCAAGTTTCAAACCTACATCT
GTGGCTCATTGAGATTTCCTGAGGCTGCTGAGGCAAGGCTTCAGGTCCCACCTGGAGCGCTGGC
TCAGCCATTAGAGCCATTAATGGTAGACTCACAACTACACAAGAGACAAAAACCCACACAAGGGGTG

GAATGCAGAGACTCAACCAGTTCCAAGCCAGCCGGGACTAACAAAGCAAGATCCTGGCTCATAAACCC
AGGAGCAGGGTTTAGCCAGTGGTGGTATGCCTGCCTGGAAAGGGATGGCCCCAGGTTTCAGGCCTCTA
CACAGAGGGTGCTTTCTCACCACACTCCCTCTTAACCAAGGTGAGCAGCCGCTCCCCCTCAGCACAC
ACATTGTACTGCCCACCATAAAGCTTTACATGGGACCCAAGAAACAGTCCTGAAAGCTGGTTCGGGA
TGTTCTTTCTCATTGCAAGGCAAGGCCAACTCCATGCGGACACCGGCTGCAGCTTGGTGCTACCTGGC
GGCAGCCGGGTCTAGCTCCTTGTGTCTCCTGGCCAACTAGGGTTTCCCTTGTGGTGGCAGAGTTCAA
GAATGCATGGCGAAAGTCCACCCGCAGCACAGTCACAGGGAACAGGGCAGGGAGGGGCCAGGCCCCGCC
TCGTCTCCAGACTCCTGCTTCCTTAAAGGGAGTCTCCACAGTTCCACCTACTGTGGGGGGAAGGGG
AAGGGGAAGGGCGGAGCTCCCTTGCTGTTCTTCAACCACCAGCCAGTACTCCCGTGCAGGCTCAAGGG
CAGCCTGTGCTCTCCACACAGCCAAGACCTGCTTGCTTGTTACTCAGTTTTTCTTACACAGGCCGTTA
GCTGATTAATTGGGTTTTTATTTTATGTGTATGGATGTGTTGCCTGTGTGCATGAATGTATACATGTG
TGCTGGTGCCCGAAAAGGCCAGAAGAGGGGTGTGAGATTCTCTGGAAATGGAGTTACAGGTGTGATGT
GGGTGCTAGAGTTGAACCCGAGTCTTAACCACCGAGGCATAGCCACTGATTACGCCAGACTGACCAGC
CTGCAAACCCAGGGATCCTCTGTATCGGCCTTGCCCCACACCTGCTAGGATTACAGGTGGTGGTGGG
CTTGGCTTTGTGGTTGCTGGGGAAGTGAAGTAAAGACCTCAGTATGTGCACCAAACCTTCTACTGAC
TTAGCAACATTCCCGTGGAAGTCTGAAAATGAAGGACGGGGGGAAGGATACTGACTCTGCTGTAAGA
AAATTTCTACATTTATGTTATTGTGTGGATGTGTGTGCACTCAAGCACAAATGAGAGCTAGAGAGGGC
CTGCAGAAAGTCACTTCTCTCCTCCACCAAGCGGATCCAGGGACTGAGCCAGGGGTGAGGCTTGGT
GGTGAGCGCTTCGCCCAGTGAAGCACCTCACCAGCCTGAAAATAAAGCTCTCATGGCACCCAGCCAC
GTCTGCTTTTTCTCATCCCGTCGCTGGCTCATTTCCCCGATAGCAGCTGGTAGAGTCATTATAGCAA
GACCGTGCACCTGTAAAGCCTGAAACACATACTGACCAGCCCTCCACAGGTCCCAGCTGACTCCTGC
TGGGACCACTGAGTTATAAATCAGAGCGTCATCTACCGGCTGCAGAGGCGACAGCTTTTTTGGTGTAC
CAACAGCAAACACTGTGCTGTATTCTGTGCACTCACCACCTGTGAGAAAATGCACCAGGGCAGGAGC
TCAGGCCTGCAGCTTCAGCCAGGTTAAGGCCGGCCTGAGCTCTCCTTCCAGGCACAGCCTTGCCTAC
TCTAGCTGGCATCTGTAACCTACCGCAGTCCACTGTGCCATCTCTGCATGCTACAGCCCTCACTGTCC
TTCCTGGATGTGAGTTTCCATGGGAGACAGCTTCGCCTTTCTCAAAGCACATCCTAAGTCTCCTTCC
TACCCTGTCCCCCAAGGGGGGCTCCTCCTCCACGGACACTGTGCTTTTCACTCTTTGCCAGGGGTCTG
ACCTAGGCCTGGGCCGACCAACACTGCTAGGACCTGGCAGCACCCACTCTTCCTCCTTCAGGAACCA
CGTCTGACTTTCTCGCCACAGGGCTTCAGTGTGACACTTGTGACAAGGTGACAAGTCTCCTGCA
CACTGGTGGGCTCTGGGACTGACATGAGATCATGTGCCAGTGTGACACAGAGAGCTGTGGCTCAGCCA
CTGAGGGGGGTAGGGCTACTGTACCCACCTAGGCATTAGCCCTGCTAAGCACCAAGGGAGAGACCA
GACCCACCAAGGAGGCCAAAGCGGCTGGTGGCATTAGGGATCCACTGCCAAAGACTGGAAGTCTAGGG
CTTGGACAGACAGATGGCTCAGTTGGTAAAGTGTTTACCACACAAGCTGCAAACCTGAGTTCAACCC
CAGCACCCATGTAAATGCCAGGCATGGTGGAGCATGTGTAATCCCAGTACTGGGGAGGTAGAGACTG
GAGGACAACCGGGGTTCACTGGCCAGCCAACTAGCCCAATCGTGGAAGCTACTGAGACACCCTGACT
CAAAAATCAAGGTAGACGGCTCCTGACGAACATTGATTGACCTCTGGTCTCAAACACACCTGTGCA
CGCACACATGCACACACAAACATATGAAGGACTGAGATTCTGCATACGTTAAACAGCAACTCTCTCCT
CCCCCTTTTTATTGCATTCAATTTACTGGTGTGTGTGGCTGAGAACAATCTATGGGGTCAGTTCTCTGG
TCCACCAAGTGGGTCCAAGGAATCAGACTCAGTTGTAGGCTTGGCAGCAAGCACCTTGACCCACTGAG
CCATCCAGCCAACTCTCTTTTTTGTGTTGTTAGTTTGTGTTGGTTCAGAACTCTCTGTGTAGTCTGGC
TGTCTGGAAGCCACTCTGTAGACCAGGCTGGCCTCAAACCTCAGAGATTACCTGTCTCTGCCTCCAG
TGCTGGGAATAAGGCATGTGATAGGGTTAAACCCACCACACCCAGCTCTACTCCCAGGGTTGTTAGCG
CCTGGAACCCACATTGGATGCTCTGTGGCTATGACTTTGGGAGCGCTATTATTTTTCATTACGCTTGGT
GTCTCACACAAATGGAACCTCGCACTTGTCTTTGAGTGGCTGACTTGCTCCATGTAGGTCCTTAAGT
CTCATCCAAATCTGGCAGGCGGTGGTGGCGCAGCCTTTAGTCCAGTACTCAGGAGGCAGAGGCAGG
CGGAGTTTGAAGCCAGCCTGGTCTACAGAGACAGTTCCAGGACAGGCTCCAAAGCAATACAGAGAAAC
CCTGCCTCAAAAAACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAGAAAAAG
TTTCATCCACATCGTTAGACGTGTTGGTTTCTTCCCTGTTAAGGATGAGCAGCATCTTTTATCTGT

58

GCCTCCAGCCCTCGTTGTCCGCTGGTTACAGTAGCACCCCTGCTCCACCAGGAAGCGCACCCACCTCC
AGGTTCTCGTCTATGCAGGCCTGGGGACGGGGACAGGCCCATCAGCTCCCGGCCGGGCCAATGAGAGG
TGTGGAAGCAACGCCGATGGGCTGCAGCACAGATTCCAGGGGCCCTCTGGTCAGTGGCCGCCCTAAAA
TATGCCTCGTTACCCATGCTTGGGTAAATCTATGCATGCAGAGCTCATGGAGACTAGAGCAGGCTCCAA
AAGGCAGATTGAAAAGGCGACCAGGGAAGAGGCGGAGCTGCCATCCCTGCATGTGACTGCTGAACATA
CCCTATGAGGCAGAGGAACCCCAAGAGCCAGCCATGTTCTTCCAAGGGGCAGGGCAAGGCTAGGTTGA
GGCAAAACGCTCACCTAGCCCTGGGTTCATCCACAACACAGGAAAAAGAGAGATCACACAAAGAGA
CACACGCACATCCCAGAGTTGAGGGCTTGGGGCACAGTTCCCAAAAGGGATGAGTAGGCTATGTTCCC
GGTGTCCAGGGATCCAAGCAGACCAAGCTCTGGGTCACTGAGGGCCTACCGTGCACAGGTCTCCTCAG
AACTTCTTTTCTAAACACCCACACCACTAATCCCCACCTCCTCACCCCTCGAACCAACAGCTCA
GGAGAGGAAGGCCTCACCTAGCCAGCACAGCACCCAGGGCCACAAGAGAGCTGGTCTCAATCCCTGA
TGACACTGCTTGGACACTGGAACCACTGAGTCTAGGGAGGGGGTTAGGTCCGGACTCCTGGGTCTCTCA
TGAAAAATTAACCCCTTCTACAAGCGCGACCATCTGGAGAAAGAGGGAAGGAAAGCTACGAGGGGCCAAG
TGCATGAAGTCATGGAAATTTAGGCTGGGGGGGGGGGCACGTGCCCTGGGAACGGGATGAACCTCTGGG
CTTCACTCTGGGCTCAGTTTATTTCCACCCTGTTGTCTGATGGTGTGAGGGAGGGGGGCAAGGAGGCAGA
TGGGCTTTTCCCTTTCAAGGACCTGGCCGGGTACGGGCATCCATGTGAAAGATGCCTGAGGCTGGGCA
CTGGGGACCAAGAATCCTCCTCCTCAGATGTAGAATCTAGCCAATCCTCTTCTCTAGACCCAGG
GATCCAGACTTGGCCCTCCTCCTCAGGCCAGGTGCTAGGGCTCCCCATCTCTCCCCTGCTCAAACC
TAGGACTCTTAACTCCCAGCCCTACCTACTCCAGACCCAACTCATAGCCATTATTGGACAAGGCAATT
ATTGGACAAGGGAAGAGGAAGGAATGTCCCTGCCTTGCTAAGGCAGAGGCTGGGGCTTAGGAAATGT
CATTGCAGGAGGCTGATGCCCCAAGGAGGGTCTAGAACCGGAAACACTAAAAAGTCTGAGGTGTAGAA
ATCACACAGACTGGGTGGCTCAATGCCCTGCTTTCTGGGACTGAAACTAGTTTCAGGAGTTTTCA
CTGCTGAAGCCAGGGCAGTGGTACTAGGAGGTGATGCTACGTACGCACCACTCCAAACCCAGCCCC
TCTGCGTTCTGGCCCTGAAAGCCAAATGATCTCACTGAATCTGATCTCCAGTCTCCCAAGCCTCCTGC
AAAGGCCTGAAGAGTCAGGTCACCAAGGTGTCTGCATGGCGGGAGGAGAGTCCACCTGGAAGGCTGA
CACGTCAGGCCTGAGGTCACAGGTTCTGTCTAGAGAGGATGCTCTAGGGACCTCCAGCAGATGCAGAG
GAAGGGGATGCAGTTGGGAGGGAATCTTGGGAGGGCCAGGGACTTTGGTGATCATGTGAGCAGCCTG
AGCTGATCTCCTGGACTGGTCAAAGACGCTGACACCCCTGAGTGTGGCCTCGGGAAAACAGGACCCCTGC
TATATATAGAGGACGATGTCCACACAGCTCACGCCGGCCCCATAAAGGAAGTTTTCCACAGGACGCC
TCTCACCATAGAGTCCCTCTGGGGACAGGGGTGACCACTGGTCCCATTCTACAGGTAAAAAACTAAG
GCGCACCGAGAAAAAGACACTCAAGATACAAGACACAAGAAGCAGACTGACACAAAAAGTCAGCACGGA
CTTTTTTTGTTTGGTCAAGATTTTGAACCTGGGTCTCATGTGGGCTATCAAGATGGCCTCAAAGTCAC
TGTGTAGATGAAGGTGACACTGAATTTCAACCCCTCCTGCCTCTACTTTCCAAGCACCTGATTTCTGTG
GTATTGGGGTTGGAACCTGAGGCTTCCTGCACTCTAGGCAAGCACTCTGTCTAAAAGACAGCCCAGCC
CAGGACAGACGGATTCTGTTTTTCTCTGCCTGGATGAGTGAAACACTGAACCTTTATTCCCCACCTC
CACGAGCATCCTAGCAAGAGGACGACAACCCAGGAGATGGAAGTTGCCATGAAAGACTGAAAGTGAAC
CAACACTGTGGCCAGGAGGAAGAAGAACGGGGATGGGGCTCTGCTGTGACTAATCTTGTCCCTGACA
ATGCCAGCTTTTGGATGACGGGGAGATAAAAGCATCCCGAATCCAGAAGGATCCCGGCAATACAAGAT
GGTCTCACTCTCGGGGCACAGATCACTGGAAAAAGATAATCACAGTGTCTGAGTCGCCAGGGTCTCT
GGTGGGGTAGGTTCTAGAAGGTGACAGGGTGGAAATCTAAGAGACAGGGCATAGTTTTTAAAGCAGGA
TGCTGCCCCAAATATAGTCCATGGGGTGGTAGGTGGAGTGGGCATGCCTGTAATCCACAAATGAGGGG
GTCAGAGACACAGGACCTGTATTTGAGGTTGGCCTGGGCTACACAGGGGAAAAAACAACAA
CAAAAACAAAACAGAGGAGAGAGAAATAGGGCTTGAAGACCGAATGCTTAGAAGTTTCCAGAAAGG
CAAATCAATGTGGACACAGAGAGAGAAAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
GAGAGAGAGAGAGAGATTGAGGAGAAACCTATGGGGGGTGGGGTGGGTGGCAAGATCACCAAAAGG
GGATGAGCCGAGAATTGAATTAATAGGACATGGGGAGGGGAGGAAGTTATGGGATGGGGCCCAACAG
ACGGTGAGCGCCTCTTCTCCAGGGGAACAAAGGGGTACACTGCCTTGAGGGGCAAAGGACCCCTCCT
GAGGCCACAGCGACAGCACGGGTACAGGAAGTGGGGTAGGGAACAAGGTGGACCCCCAAAAGAAG

TGACACTGAAGGGCCTGGGCCTGGCTAGCCTCAGAGGAGGGAGTGGGGGATTGGGGGGGGGGCGTCA
AGTCAGAGCTGGGCCCTGGAAGCCTGCGGCACAGCCAGGGCAGCCACCAGCCTGGAAGGCACGGGGT
GTAAGCCATCCGTGTGCGGAAGACGCCGCCGGGAGAGCGGTGACAGCGCGGATGACAGGGGCGAGGC
GGCCCCCTGCAGGGCAGGAGGCGGGGAGGAGGAGGGGTGGCTCGGGGGGCCCCGGGAGGGAGGATGC
TCGGGGGCCGCTGACCTGGTGCAGGGCGCTGATGCCGTGCGCGTTGGTGGAGTCGAGCACCGCTCGGG
CAGGCGGCGGGACGCCGGCGTCCGATTCCCCGCTGGCGCCGGGATCGGGGCCCCGGGTCCCCGGGA
TCCCCGGGGTTCGGCTGCGCGAAGCATCAGGCGAGCCTCATCCAGATCGCCGCCCGCACAGGCCGCCAG
GAACTCGGCGGCGCGCTCGAAGCGCACGGTGCAGACGACGCTCCCCGGGGCCCCGGCTCAGCCCCCG
CCCGCGCCCCCACC GCCCAGCTGCTCCTGCCGCCGCTCGCGGGCTGCCGCCGCCGCCAAGACGAC
GACGACGACGACGACGACGACGACGCCGCCCGGGGCCGCTCCTCGCCCCGACATCGCGCCCCACACCGG
GCCGCTCGCCCCGCTCACCCACCGAGCGAGCGAGCGAGCGAGCTGAGCGAGCGCCCCGCCGAAGCCGG
CCGGCGACGAACAGCCGCCACCCGCCCGCTCGCTCGCTCGCCCCGCCGCCGCCAGCCCCGGGGGCCG
CCGGGAACCGCCGCCGCCGCCGCCGCCGCCCAAGCACCGCCCCGAGGCTCAGGCTGGGCCCCAC
CCCTCCCCCACGGACGGGCGTTGACGTACGACGCTGCCCCACAGCCCTCTGGGAAATGGAGTCCTC
CGTTGAGAAGCCCGCAGGGTTTTTTTTCAGCAGACTCGCTAACTGCTGAGGGAACGGTCGGGGTGGCAG
GAAGCCGCCAGCAGGCGCGCCTACAGCCCCCAGCACCTGAGAGGCAAACTGCTCTCTCGAGTTCGAGG
TCAGCCTGGGCTACAGAGGCAGTGCCAGGGTAGCACC AATTGCCTAAAGCAGGACACGCCCCCCCCCG
GGAATGCTGGAATCTGAGTTTAGAGGCGGGACGGGATGCCGGGGGGATGCTGGGAGATGTAGTTTT
TTTGGTAAAGCGGCGCAAAGGATGGCGCGTGGGAAATGATGGCGTGTAGCGGAACCCGAGAGACGCAG
AAATAAGACTCGCGTACTTTTTCAGTTGTGTTTTTGTCTGTGAGATGGGTTTGCCCTCGAGCTCGCTGTGT
GACTGCGATTGTCTGTTTTTAACTCCCGACCTTCCTGCCTCCGCTCTCCTAATTGCTGGGGTTGCAGAC
GTTTTTTTGGTTTTTGTGGTTTTGGTTTTGGTTTTGGTTTTTTTCTTGGGGCGGGGGTATTTTTGTTGTT
TTGTTTTTGTTTTTTGTGTTTTGAGACAGCGTCTCACTATGTAGCCCTGCCTGGCCTGAAACTCGCTACGT
AGACCAGGCTGGCCTCGAACTCATAGAGACTCCTCCCCCACACTTCTGCCTGGTATGAAGGGGGCGC
CACCAGTCCCGCTTGTTTTGTTTTGGAATCTGCCCTCCCTCCCTCCCATCAACACCCGATGAAG
GACAAGGATTTGTGAATGAATGAATGAATGCATGAGTGCATGAATGAATGGGCTCCCCAAGACGTCGG
GGAGACCAGGGGCCACGGGAAACTGAGTCTTGAACCAGATTAAACACCAATCGCCGCCAACTCCT
CTGGGTAACTAAGGTTCCCGTGCAAAATCCAAGGGTATCGGGTAGCATGGGGCAAGCTGGGAAATGTA
GTCCCAGGGCCACGCTCCTAAAGAGTTTCAGCCCCCAGACTTCCAAAACGCTGAGATGCCAAGGTA
CCCCGGAAGTCAGTTTCAGATGAAGACAAGCCTCCGGTCTCCAGCGGTAATCCCTTGAGCACCCGG
GAAGAAGGGTCCCCAAAGAACACACATTTCTCCTTAGCCCACTCGGGGCTGCGGGGGACGCTAGGAG
ATGCTCTCCCGGCTGCATCAATGCTCTCCTGGAATTCTGGGATCGGTAGCACAAAATGTGATGCTCCG
ATAGTTTTGGAAGTTTTTGTAGTAGACCCAACAGATAAAAGAACACCTTGATCTTTCAAGAATCTTCC
CCCCACCCCCACCCCCACCCCCACCTCCACCCCAAAATTTGCAATTTGAGAAGGACAGAAACACTTTT
GAGACAGGAACACAGACTCACACACACACACAAAAAAGTAGAACAGAAAGCTGTCAAGTTTATAGA
GAGAAAACACGTCTTCCTAAGGGTCGTTAGGGCAGCCCCGTTCACTGTGACCCTTGATTGTTGAA
TGAGAGATAAATTACAGACCCTGGCAGAGTCTAGGGAATAACGACCATAAATCCAAAAGGATAACCT
GTGGTTTTTAAGATGTGAGATCACACACACACACACACACACACACACACACACACACACACAC
CATCTTCCCCAAGGCAAGAAATCAGATATTTCAACCCCTGGGGTCCAGAAGGAAGGAGGTGCTGAC
TCCAAAACGTTCTTCTGATTTCCACCATGGATTTCCACACACACACACCCCTATCAACACACACACTA
AATAGACGTTTATAAAATGATCCACAAAATAAGGCTACACCAACACACAGAGGTAAGACTGTTGTTAG
ACAGTTTTTGGTCTGGTTGGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGTAGCCTTCTCCTGTCCAT
TTCTCATGCCCTTACACACACCTGGCCTCTGGGTGTGTTATTTTTAAAACATCCTTAGAAGAATTAATG
ACCTTGTACAACAGTTTTAAATGCAAGAGGCAATTAATTTGTTTTGTTTTGTTTTTCGAGACAGGGT
TTCTCTGTGTAGCTTTGGAGCCTGTCCTGGCACTCGTTCTGTAGACCAGGCTGACCTCGAACTCACAG
AGATCCCCCTGCCTCTGCCTCCCGAGTGCTGGGATTAAAGGCGAGCCCGGCAGCACTGGAGATTTAAC
TCAAGGTCTCCTGAGTGCTCGACAAGCTACTCCCAGCCATGAACCTTGATATCTCTTAATGGCAGCTG
ATGCTCTCCCGGGCAACATGGAGCTGTCCAGCCAAGCCGCACAGCCAGCCACGCATAATGACAACAC

GGAAGAACTCAAGCGGATGTCTGGAGGGCCTTTATTTTGGAGTTACAGATGGGGGACACACTCCAGAGG
CTCCCAGGCTCCATGCAGTGGGGCGTGTCTGGCAGTCTCACTTCCAGCGGCCTCCAACCTCGACCCCTT
CCCAGCCCCCTTTCGGCTGTGGGAGAAGAAGGTGGAGTCAGGAAGAAGCCCGGAGCCTCCGAGATAAG
CTTAACACAGTCCCTTTAAAATTAAGGAAGTCCACCAAATACCCACCCCCACCCAGAGGGAAGAGAGA
GCAGAGGTGAGCAGAGCTGTTTTTTTTTGTGTTTTTTGGGTTTTTTTTTGCAGTAGTGAGCATAA
AGTCAAGGCCTCACACGTGCTAAGTATGTTCTGTACACTGAGCCACGCCCTTGCTCTCACTGGCGA
TTCTAAGCAAGGGCTCTACCACTGAGCCACATCCCAGCCCCCTCACTGGGGGATTCCAGGCAGGGGCT
CTACCACTGAGCCACGCCCCAGCCCCCTCTCACTGGGGGGATTCTAGGTAGGGGCTCCACCACTGAG
CCACACCCCCAGCCCCCTCTCACTGGGGGGATTCTAGGCAGGGGCTCCACCACTGAGCCACGCCCCCA
GCCCCCTCTCACTGGGGGGACTCTAGGCAGGGGCTCTACCACTGAGCCACGCCCCAGCCCCCTCTCA
CTGGGGGGACTCTAGGCAGGGGCTCTACCACTGAGCCACGCCCCAGCCCCCTCACTGGGGGGATTCTAG
GCAGGGGCTCCACCACTGAGCCACACCCAGCCCCCTCACTGGGGGATTCTAGGCAGGGGCTATACCACT
GAGCCACACCCCCAGCCCCCTCACTGGGGGATTCTAGGCAGGGGCTCCACCACTGAGCCACGCCCCAG
CCCCCTCACTGGGGGATTCTAGGCAGGGGCTCTACCACTGAGCCACACCCCCAGCCCCCTCACTGGGGGA
TTCTAGGCAGGGGCTCCACCACTGAGCCACGCCCCCATTAAAGGGCATCTCTTCAGATAATTCCCAGT
AGGGGGTTGTTGGCCATGTTGGAGTTGACTTTCTTGGGTTAGTTCGAGAACACATGCAAATTTATGA
GTAAGGGGCTGAGGGAGAAGGAAGGGTGAGCTGGAGTTGGTGACTTGCATGCAACAATGTTGAGTGA
GGCTGGAACAGTACAGAAAATGCTAGAAAAAGGCAGAGACTGAGCAGTGAAGCGCCTGGATACGGTGG
AGCACATCGGTAAATCTCTGCACTCTGAAGGGGATGAGGCAGGAGGATCACCGAGAGTTTGAGGACAAC
CTGGGCTATATAGCAAGAGCCTGACTCAAATGAAAACAACAACAGCAAAAAAGTCGGGTATGATG
GCTCTGTAATCCCTGAACTTGGAAGCAGAGGCAGGAAAGTGTGAGGAGTTCAAGGACACCCTCAACT
ACAAATGGAGTTCAAGGTCATTACGCTTACAGGAGACCTTGTCTTAAAGCAAGAAAATAGAAGAAAA
GGGGCAGGAAGTGAGACAGACAGATGGAGAAGGGGGGAGGGGGGAAAGAAAGGAAGAAAGAGAGAGAGA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
AGCAGGTAAAGGAGCTTGAGCCAAAGCCTAGGCCCTGAGTTTCAACTCTGGGACCCACATGATAGAAG
GAGAAAACCGACTTGTTTCGAGTCATCCTTCGGCCACATCTGCACCATAACAGCACACACACACACACA
CACACACACACACACACACACGCACGCACGCACGCACGCACGCACGCACGCACGCACGCACGCACACACA
CACTATGCGGTGTGATATGATACAAAAAAAGTGTAAGAAAATGTACTCAGAAAGAAAGGGTTGGA
GGGAGGCAGAGAGGCGGGGAGATTGAGACCAAAGAGTTGATAAAGAGAAGCAAGAGATTGGAGTGAG
GCCAATAAACACAGCACTCAGCAGGCTGAAGCCAGGGGACCAGGAGGAGTTCAAGGTCAGCCTCAGCT
ACCTAGTGAGACTGGGCTGCATGAAACCTTGCTTAAAAATAAATAGACAGAGCCGGGCAGTGATGCG
CACGCCCTTAAATCCAGCACTTGGGAGGCAGAGGCAGGCGGATCTCTGTGAGTTTCGAGACCAGCCTGG
TCTACAGAGCTAGTTCCAGGACAGCCTCCAAAGCCACAGAGAAACCCTGTCTCGAAAAAAACCAAAAAA
TAAATGAATAAATAAATAAATAAATAAATAGACATACCAAAAAAAAAAAAAAAAAACAGGAACAGTGAG
TCATGCCAATCATCCCCACATACATGGGATTAAAGCAAGAGGATCTCCTACAGGTTCAAAGAAAGCCT
GGTCTACATAGTGAGTACCAGGCCAGCCTGGGCTACAAAGTAAGACTTCCTCGAAATAATAAACAAC
TAAACAAACAAACAAACAAATAAATAAACAACCCGAGAGAACAGATACAGAAAGGATGTCTCAGGGA
GCAAGGAACAAAGACATATAAGATGCCAAAAGGAGGGCTGGAGAGATGGCTCAGCAGTTAAGAGCACT
GGCTGCTGTTCTGGAGGTCTGAGTTCAATTCCCAGCAACCATATGGTGGCTCACAGCCATCTATAAT
GAGATCTGGTGCCCTCTTCTGGCCTGCAGGCAGGCATACAAGCTGGCAGAACTGCATACATAAATA
AATAAATCAAAAAAAGATAACACTTTAAAGAAAATGATACTTTGAGAATTCTATGTATAGAGCCAGGC
GGTGGTGGCGCACGCCCTTAAATCTAGGTCTCAGGAGGCAGAGGCAGGAGGATCTCTGTGAGTTTCGAG
GCCAGCCTGGTCTACTGAGCAAGTTCCAGGACAGGCTCCAAAGCTACAGAGAAACCCTGTCTCAGAAG
AAGAAAAAATAAAGATCCCAAAGGGCAGTGGTATGCAGAAGACAGGGAGGAAGGGAGGGAGGGAGGG
ACAGAGGGAGGGACAGAGGGAGGACAGCAGGCCTTTTGTGGAAGCAGCACTTACAATTTCTGGGCATG
GCTGATTCGGTTGTACAGCACATTGATCTGTAGAACGAGAAGCCAGGCTAGGTGCAGATGTCCAACCA
AAGCCCTGCCCTGCCTATCACCCCTGTACCCACGCTGGACCCCAACAGAGGCAGGTCCCACCTCGT
ATTTCTGTTGCTTCAGTTTCTCCATCAGATCAAATTTCTCTGACTCGAGCTGGTGGATCCATTCCGAC

AGCTCCTGGGCCTTCTCCCTGGCAGAATGAGAATAACTGGGATGCAGCGAGACTATGTTCTGGGCCCA
GAAAGTTGAGACACCTACCCCAAGCCTCAAGGCAAGTCTCTCTGAGCTTGATACAGTTGGTATGCTA
AGCCACTATGAGCCTGATACAGTTGGTATGTTGTGGAATGTTACTTTAACTATGTAAAGATGCCTTAC
ATTTGTTTACTTTGTGGAATGTTACTTTAACTATGTAAAGATGCGTTACATTTGTTTACTTTGTGGAA
TGTTACTTTAACTATGTAAAGATGAGTTACATTTGTTTACGCTGTGGAATGTTACTATAACTATGTAA
AGATGCGTTACATTTGTTTATATTGTGGAATGTTACTTTAACTATGTAAAGATGCGTTACATTTGTTT
ACGTTGTGGAATGTTACTATAACTATGTAAAGATGCGTTACATTTGTTTAGGTTGTGGAATGTTACTT
TAACTATGTAAAGATGCGTTACATTTGTTTACTTTGTGGAATGTTACTCTAACTATGGGAAGGTGTGT
TGCATTTGTTTCTGCTGCATTTGTTGAGTTGGATAAAGGTGTGTTGCTGTTTCACCTTGCTGCCTAA
GGCACCTGATTGGTCTAATAAAAAGCCGAACAGCCAATAGCTAGGCAGGAGAGGGATAAGCGGGGCTG
GCAGGCAGAGAGAATAAGTAGGAGGAGGAATCTAGGATGCAGGGAGGGAGACCAAGGGAGAAAGAGAG
GGAGATGCCTGGAGCCAACAGACATGGAGTAGTCAAAATGCAGATGAAGAGAAAACAGGTAAATTTAA
GTTATAAGAGCTGGTAGGACAAGCATAAGCTAAGGCCAAGCTTTCATAACTAAATTATCTCTCCACGT
CTTGATTTGCGAACCGGTTGGTGGCCCGAAAGGAAGGCAGCTACACTGGTAAATCTTGTTAAGATTG
GAGGCTGAAGTTTAAATATGGCCACTTTCTGAAACGGAGGTCTCCCAAAGGAGAGAAGGAAGTGAT
GACTGGGAGCCCCAGAATCGGAAAGATGTTTGGTTTTTATTTATTGCTTATGTAGTGTGTGTGTGT
GTGCGTGTGTGTGTGTTAATTTTGATTTCTGAGACAAGGTCTCGTGTAAACCAAGCTTCAATATATA
GGAGAGGATGACCCAGAACTTCTGATCCTCCTGCCTCCCCCTCCTGAGTGCTAAGATTCCACCTAAAT
GAGGGAGAGAGCACTGTCAATGCCAGACTCCACAGGCCATGCCAGGCACAGGACAGATTGCACCCGAG
TGACATTTTGAACAGAGAAAAGCAACAGTGGCCAGAAAAGGAATGTCACTTGCTTAAAGTGACACAGC
ACCAAGGCTCACACCTGGAACATCACCCACGGAACCTTAGGAGAGCCACAGGTGTTGCTGTAGTTCC
TTGTCCACCAGGTCCCTTCGTCTTCTCAGCCCCAAATAAACACACAGATACTTATATTAATTATAA
AACTGTTGGCTGATGGCTAGGGCTTCTTATTGGCCAGCTCTGTCTTAATTGACCCATTTCTATAACTC
TATGTATCTCCACGTGGTCTTGGCTTACCGGAGAATGGCCGACCTGTTACTCCTTTTGGCAGCTACA
TGGTGTCTTCCCTGTGGCCCCCTCTCTACCTACCTTTCCAGAATCCTCCTCGTCTCCTAGCCCCGCT
ATCTTGCTGCCTTTATTGGCCAAGCAGTGTTTCATTTCATCAACCAATAAGAGAAAACATATACAGAA
AGGCATCCCCCATCACACAGGCACTTACCGGAGTTGGTCTCCTCCCATGTAGTCGATGTTCAAGGGTT
TTTTCTCTCAGACAGGATCCTGAGCTTCATCTCCGGCCGGTCTGCCGCTTCCCACGTTTCTGCTCA
GCCTGGAGGGGGAGGAATCCCAAGCCAGGGATGGGACCTGGAGGCCAAATCCACTCAGGGTCTACAG
TCATGGCCAGGGCTCCACCACTTGCAAGGGGGCCAGGCTGCAGCCCTCCTCCCCCAGACCCAGGAAT
CCAAGCTCCATGCCTCCTCCATCAGACACGGGAGTACAGGCCCACTCTTCCCTTGACCCAGGCACCA
CCAGAAATGTTATCTAAAGCGACAGCTGCCCCCTTACCCAAACCCCAACGGGATCCAGGCTCAGTGC
CCTTCTGAGGACCAGCATTCTGCAGCTCCCTTCCCTCCTCCTTGTAACGCAGACACCCCCCCCCC
AGGGCTAGACTCACCTTGACCAGGTAGCCCCAAAATGAGCCCCATGTTGGAGAGAACCTTCTTCTT
CTTGGCATCGTCTCGGCTCGCTTCTTGGCTTCTCCTCCTCTTTGCGCATCTTCTCTTCTGTGAAA
ACAGAGGGGTTCCTCCATGTGGCCCTACTAAGGAAGGCACGAGCCTGGGTAGTGCATGGCTAGGCTC
CATAGACGGGGCCGAGAGGGATCCACTCCTAACTTGATATATAGTTCACCGAGCCCTCAGGTTGGC
AGAAATCCTCCTGCCTCAGCTTCTCCAGGGCTGGGATCACAGGTCTGAGCCACCGCAGGCAGCAAAG
CTCAGTCATTGTGCTATGGTTTTGAGTCTCTTGATCCAGCCATGCCTGAACTTTTTAAATTTATATG
CAGTTCATCCTTTCACTAGGAAAAACAAAAACAAAAACAAAAAACCTTGCGCTGTACATT
ACAGGGACTTTCTGAGAAACACAGGGGACATCAACGGCAGGAAAGGAAAGTATTTCGTCTCCAAGAAAG
GGGCAGACACCTACCGCCAGCTTGGCCTGCCGCTCCCGCTCCTTTTCGTTTCTGAATCGCTGCTGCTC
AGCTCTCTCTGCACGACGCTCTCCTGGGGTGGGGTGGGGATGGAAGAGAGGAAGATAGCGGAGGGG
GATTGGAGGCTACCTCTCCCCAGTTTATCCCTCCCTCCACCAGTAGACAGCAAGAGAGGTAAGTGC
GGTTTCTTTGTTTTCTTTTCTTTTGTAGATAGTCTCAGAGAGCCCAGGCTGGTCTCAAATTCGTG
GCAATCTACCTTCTTGTCCCTCTAAGAGCACGCTGGCCGAAATACGCATTGGAGTGCAACCACCCG
GTGTCCCCATTGTTTCTCCCGGAGACCCCAAAGCACTTCACGGAGGAAGGGACGAGGAGGGACCGAAG
GACCCGGCCTCCCCGGCCTCCCCGGCCTCCCAGGCTCACGATGCGATCTTTCAGCGCAATGAGCTCCT

CCR5 sequence

GTAACACAGAGTCTCTGTAATGCAAGGTCCGGCCTTGGCAGCCCCAGCCTGGAGCCACAGTGAG
ATGTGAGCCGAGGGTTATGCTGGGAAAAACCTCTCCCTCCCAGCACCTGAAAGGCTCTGCAG
GCCCAGCAGCTCAGCAAGCAAGGGTAAGGGCATGGACTAACATCTTATTTTCACTATCCCT
TATAACACATCCTAATGTAATCAGCTCACAATATGAAATTATTTTCACTTCTCTCCAGTCATT
GTTTCAATGGGGCCTTAGGGTTGACTGGATTCTGGAGGGGCCCTGCCTAGAGGAGGGGGTGCAT
TTCTGTCCCTATGTCCCCCTCCTGCTCCATCCTCCACAGCACGTGCCTAGTGGTCTACCTTGT
GGGGAATTCTTGACCTCCCTCTTCTAGGCATGGACTAGCATTGAGAAGTGGGAGAGGAGTG
TTAGGAAAAAGGGCAAATATAGACATACCTTGTCTTATTGTGCTTTACAGATATTGTTTTTG
TTGTTGTTGTTGTTGTTTACAAATTGAAGGTTTGTGGCAACCCTGCCTCGAGCAAGTCTATT
GGTGCTGTTTTTCCAACAGCATGTGCTTGTTTTACATCTCTGTGTCACATTTTTGGTAATTCT
CCCAATATTTCAAACCTTTGTCATTATTTCTATATCTGTTATGGTAATCTGTGATCAGTGATC
TTTGATGTCACTATTGTAGTTGTTTTGGGGCACCATGAAGTGCACCCATGTAAGATGGCAA
CAATCAATAAATGTTGTGTGTGTTCTGACTGCTCCATGGACTGCCTGTTCTCTGAGACACAAT
AATGTATATATAACAATTATATATATATATATTTTATAACAATTATATATATATATATATATA
TTTTTTTTTTTGAGGCAGAGTCGCACTCTGATTGCCAGGCTGGAGTGCAATGATGTGATTTT
AGCTCACTGCAACCTCTGCCTCCCCAGGCTCAGGTGATTCTCCCACTTCAGCCTCCCAAGCT
GGGACTACAGGTGTGCACCATCACACCCGGCTAATTTTTTTTTTTGATTTTTTAGGAGAGACA
GGGTTTTTGCCATGTTGCCCAGGCTGGCCTTAAACTCCTAGACTCAAACAATCCACCTGCCTC
AGCTTCCCAAAGGGCTGGGATTACAGGCATGAGCCACTGTGCCCAGCCCAAGACACAATAAT
ATTGAAATTAAGCCAATTAATAACCCTACAATGGCCTCTAAGTGTTCAAGTGAAGGGAAAAG
TCCCACGTCTCTCACTTTAAATCAAATCTAGAAATGATTAAGCTTAGTAAGGAGGACATAT
TGAAAGTCAAGGCCAAAAGCTCACCTCTGCACCAGTTAGCCAAATTGCGACTTCACAGGAAA
AGTTCTTGAAGGATATTTAAGCTCTACTCCAGGGAACATGCAAATGAAGAGAAAACAAAGCA
GCCATATTGCTAATATGGAGAAAAGTTGAGTGGTCTGGAGAAAAGATCCAACCAGCCACAAC
ATTTCTTAAAGTCAAAAGCCTAATCCAGAGCAAGACTCTAACTCTCTTCAATGCTATGAAGG

GGAGAGAGGTGAGGAAGCTGCAGAAGAAAAGTTTGAAGCTAGCGGAGGTTGGTTTGTGAGGT
TTAATGAAAGACAACATCTCCATAACATAAAAAATGCAAGATGAAGCAGCAAGTGCAAAGGGA
GAAGCTGTGGCAAGTTATCCAGAAAATCTAGATAAGATAATTGATGAAAGTGTCTACACGAA
ACAACAGATTTTCAGTGTAGACAAAACAGTCTTATGTTGGAAGAAGATGCCATCCAGGACTT
TCACAGCTAGAGAGGAGATGTCAAGGCAAGCTGCAAAGCTCCACAGGACAGGCTGACTCTCT
TTTTAGAGGTGAATGCAGCTGATGACTTTAAGTTGAAGTAAATGTTCAATTTACTATTTTGTA
AATCCTGGTGTCTATTAAGAATTATGCGAAATCTACTCTATCTGTGCTCCATAAATGGAACAA
TAAAGCCTGGATGACAACACATCTGTTTACAGCATGGTTTACTGAATATTTCAAGCCCACTA
TTGAGAACTATTGCTCAGAAAAAAGATTCCCTTTCAAATATTACTGCTCTGCACCATGTCTG
ATCAAGAGCTGTGTTGGAGATGTACGAGAATATTATGTTGTTTTCATCCCTGCTAACACAA
ACATCCATTCTGCAGTCCATGGACCAAGACTTTCAAGTCTTATTAAGAAATATATTTTCATAA
GGCTATTAAGAAAATAGCTATATATATATATATAGCCTTATATAGTTTATATAGCTACCATTG
ATAGTGATTCCATTGATGGATCTGAGCAAAGCAAATTGAAAAGCTTCTGGAAAGTAGTCATT
ATTCTAGATGCCATTAGGAACATTTGTAATTCATGGGAGGAGGTCAAATACCAACATTAAC
AGGAGTGTGAAAGACATTGATTCCAACCCCCATAGATGACTTTTCAGGGGTTTCACGTCTTCAG
TGGAGGAAGTCGCTGTAGATGTGGTGGAAACAGCAAGAGAACTAGAACTAGAAAGTGGAGCCT
GAAGTTGTGACTGAATTGCCGCACTCTCATGATCAAACCTGAACAGATGAAGAGTTGCTTCT
TACATATGAGCAGTGAAAGTGGTCTCTTGAGATGGAATCTCCTCCTGGTGAAGATGCTGTGA
ACACGGTTAAAATGACAACAATCGATTTAGAATATTACATAAAATTTAGTTAATAAAGCAGTG
GCAGGGTTTGGAGGATTGACTCCAATTTTGAAAGAAGTGGGTAAAATGCTATCAAATAGCA
TCACATGGTATGGAGAAATCTTTTGTGAAGGGAAGAGTCGACCAAGGTGGCAAATTGCATTG
TCATCTTATTTTAAGAAATTGCCACAGCCACCCCCAGCTTTAGCAACCACCACCCTGATCAG
TAAGCAGCCATCAACATCAAAACAAGACCGCCATCCTCTTCAGCAAAAACACTATGACTTGC
TGAAGGCTCAGATGATGGTTAGCATTTTGTAGCAATACAATATTTTAAATTAAGGTATGCACA
TTGGTTTTTCTGACATAATACTATTGCATACTTAATAGACTACAGTATAGGATAAAACACAAC
TTTTATATGCACTGGGAAACCAAAAAGGTTATTTTTTGAGATATTTGCTTTACTGTGGTGGTC
TGAAGCTGAACTCACAATCTCACCAAGGTGTGCCTGAACCTCTTTAGCTAACTGGCCACTGC
CACAGTCCACTCTGTGTTGGTCAAGATGCCCCAGAGTGGCAGGCACACTGTGTGGTCAATC
CAAGGGCCTAGATATGGTGGGGGCTCCAAATGGATCTAGATATGTGAGATCTCTCTTTGATT
TGACTTCTTCAACCCACCATTCTTCTGGGTGCTGGGCTCATCTCACCCAGAAAGTAGGACCC
AATGTGACAGTTCCCTGCCCAGTTCCCTCCTGTGGTAGCCACTTGACCCAGGGGCACTCTTGA
TCCTTGACAGCTCACTTACACACCCTATCTCTACCCCTATTAACCTCTCTCCAATCCCCACTC
CCCCTGCTCAGCTTGTCTGCTGCCCAGTGGGGGCCCCACCCATGCTGGCCTCTCCTTTTGCA
AGTCCCCATTCCCTCATATGGTTTCTTCAGAGCCCCCTTCTTTGGCTTTGAGGAGAGATGCCC
TCACTCGCTTCCCCACCAATCCTGCCCCTTCTACAATCCATTCAATATCCTAATTGCCTCC
GTATACAGACTGGAGTGAGAGGAGTTGATGTGATGGGTGTGGATACAGGGCTGGTGCTGTCA
TCTTCTAGTAAGCCCTGGGAGAGGTGTCTGAGCCCAGGTGTCAGTGGTTTTCTTTGGAACGT
TGAGTGATAACACTTCTTTGCCTTCAGCCTTAGGCCATAGTTGCTAGTTCTGGGACAACCA
GAAAAGCCCTACATAATCTCGTGTTATGTGCAGAGCTGAGTATAGAGCTCCAGGTATGATCT
GACTCACTTAAGATCACAGTGAGTCTATTGTATTGTTGAACTGTTAGCTTAGACATCTGTTA
CTGTACCTACATGGCACTAGCCTCACGCCTAGACACCGATCTGAAAGAAATCCCCTAAATGC
ATAGAGAAGACTTCTCAGCTGAGCTAAGGGGCTCCCACCAGGTTTGAGCCTATCTAATGAAT
CCATGAGGTAGACAGCCTGCACATGTCCACTTGGTTTGTGTAATTGCACAAATCCCTATGGG
GGATGTGGTTCATGGGCTGGGAAGTGGGTTACCCTGGGAAAGGTCTACAGGACAGAGGCAGG

GATGGAGACAACAGCATGGTGAGTTCCCAACCCACCCACGATGATAGGTGTCTGAGGCAGAA
GGTAAAGAGGCTGTCACCTGGTGGGTGTCATAAGACTCAAGTGTCAATTGTTGAGGCACATGG
GTAACAAAGCGTGGCACTGGATGGGGGTAGATTCTTCCTATTTCTGTGAGGATCAGGGGGAC
TCCCTGGCTCTCCTGCTAAAGGTGGCTCTAGGGACAGGAAGAGTGTACTTCTTGACAGGGAT
GTCAGAGCACTGATGGTGACAATCAGTGTGACACTGCTCACATGACTGAACAACCGAGAAGA
GCCCCACTGTCTACTGAACAACGGGAAGAGCCCCGACTGTCAATGACGGAGCTCTGTTAAATA
TAGTTAAGGCTATTTTGTGAATGAATGAAGCCAGACAGGAAAAGAGGACAGTATCTTTAATC
CATTTATAGAAGTTAAAGACAGGCTTATTTAATCTCTATGAAGACAGAGTGGCCCTTACCTC
TGGGTGGAGCAAAAGGCACCTTCTGAAGTGATAGGGATGTTCCCTATCATCTTGATCCGGAG
TGGTAGTTACATGCATGTGTGCATATCAAACTCACCAAGCTGTACCACTAAGTGTGTTCTT
CCTCAATAAAAAATAATAAAGAACTACACTTATAAAGAATTTTTTAATAATATAGGAAAAATGT
CTACACTATAATCTTTAGCTAAAAAAGAGCCGCCTACAGAATGGTATATGC
ATGAGAACAATTAATCGAAAAGTGCATGGGAAAAGTCAGGATTGAAACATCATGTTTTAAAA
GACATTGTTTTGATACTGTGAGAATGTACCTAAGTTTTTCTTTTTCTGTTTTTCCCAATT
TTATACAATGAGCATGTGTTGGTTTTATAATTAGACATTTTGTGTTGTTGGTTTTGGTTTTGA
GACACAGCTTGCTGTCACCCAGGTTGGAGTGCAATGGCCCAATCTTGGTTCACTGCAACCTC
CATCTCCTGGGTTCAAGAGATTCTCCCACTTCAGCCTCCTGAGTAGCTGGGACTATAGGGGC
GCACCACCACATCCAGCTAATTTTGTGTATTTTTAGTAGAGATGGGGTTTCACCATGCTGGC
CAGGTTGGTCTCAAACCTCCTGACCTCAAGTTATCCACTCGCCTTGGCTTCCCAAAGTGCTGG
GATTATAGGCATGAGCCACCGCACTTGGCCTAGACATTTGTTTTTAAAAATAAAAGATTTCAT
TTGCTCTTTTTACAGCCCGTCTCACTGTTGACTGATATTGACCAGGAGTCAACTCAGGCCCC
AGGGATTTTTCACAACAGCTGCTGTATGGCAGGGTTTCTGCTCACTGTGCTCATGTAGTTGGC
CCTTGCAACCAAAGTGAATAATTAACATTCTCCCATCCTGTTGACGATGCTCTGAAAATAT
GGTCCAGAAATGGTGTGAGCAAGGAGACAGCAAAGCAATGCTTGGAACATAGGTGCAGTGAC
TAGACATGGGGCAGCTGTTTAAAGACAAAAGGCCCAAAAAGGAGGGATGGCACGAAACAC
CCTCCAATATGGGCATGGAGTCTAGAGTGACAAAGTGATCAAAGTTCATTTCTATGGGGT
GTCCGAATGTACTTAATAATAAAAAGAGAACAAGAGCCATGCAAACCTGAGAGGGACAAAGTA
GAAAGAGTAGCAGACACCAAGCAACTAAGTCACAGCATGATAAGCTGCTAGCTTGTTGTCAT
TATTGTATCCAGAACAACATTTTCAATTTAAATGCTGAAGAATTTCCCATGGGTCCCCACTTTC
TTGTGAATCCTTGGGCTGAACCCCCCTGTCCTGAGTGGTTACTAGAACACACCTCTGGACCA
GAAACACAAAAGTGGAGTAACGCACACTGCAAAGCTGTGCTTCCTTGTTTCAGCCTGTGAAT
CCTCACCTTGTTTCCCATCTAGCCTATATTTTTCAAACCTAAGTGGCCATAGAATCATGTAG
TATTTAGGGTGGAAGCTGCCCCAGGTCTAGCACGTCATTTAACAGATGAGGAAATGGAAGCT
TGGGCAGTGGAAGTATCTTGCCGAGGTACACAGCAAGTCAGCAGCACAGCGTGTGTGACTC
CGAGCCTGCTCCGCTAGCCACATTGCCCTCTGGGGGTGAGTATGTCTTACATCCTCCAAT
ACCCTAATGACAGACAAACAGAACATGGCAAAGCCTCAGCTCTGCATGGTGAAAGTAAGAAC
CAGCAATTGCCACAAACAGAAATACAGTGTGGTCCGGCAGCCTCCGGGGGTCTGCACAAG
TGGATTACCAGTGAATACAAGGCTATCTATCTTTCGAAAAACCAAAGTTGTATTTATGCTAT
CTATTTTCTATAAAATTTTATATTAATTTATTTGTTACCTATTTTTGAACTCTTTCAAAAGC
ACACTTTATATTTCCCTGCTTAAACAGTCCCCCGAGGGTGGGTGCCCAAAGGCTCTACACT
TGTTATCATTCCCTCTCCACCACAGGCATATTGAGTAAGTTTGTATTTGGGTTTTTTTAAAA
CCTCCACTCTACAGTTAAGAAAATAAGGCACAGAGCTTCAATAATTTGGTCAGAGCCAAGT
AGCAGTAATGAAGCTGGAGGTTAAACCCAGCAGCATGACTGCAGTTCTTAATCAATGCCTTT
TGAATTGCACATATGGGATGAACTAGAACATTTTCTCGATGATTGCTGTCTTGTATGAT

TATGTTACTGAGCTCTGTTGTAGCACAGACATATGTCCCTATATGGGGCGGGGTGGGGGTG
TCTTGATCGCTGGGCTATTTCTATACTGTTCTGGCTTTTCCCAAGCAGTCATTTCTTTCTAT
TCTCCAAGCACCAGCAATTAGCTTTACCTTTTTCAGCTTCTAGTTTGCTGAAACTAATCTGCT
ATAGACAGAGACTCCGGTGAACCAATTTTATTAGGATTTGATCAAATAAACTCTCTCTGACA
AAGGACTGCTGAAAGAGTAACCTAAGAGTTTGATGTTTACTGAGTGCATAGTATGTGCTAGAT
GCTGGCCGTGGATGCCTCATAGAATCCTCCCAACAACCTCATGAAATGACTACTGTCATTCAG
CCCAATACCCAGACGAGAAAGCTGAGGGTAAGACAGGTTTCAAGCTTGGCAGTCTGACTACA
GAGGCCACTGGCTTAGCCCCCTGGGTTAGTCTGCCTCTGTAGGATTGGGGGCACGTAATTTTG
CTGTTTGGGGTCTCATTTCCTTCTTAGAGATCACAAGCCAAAGCTTTTTATTCTAGAGCCA
AGGTACACGGAAGCCCAGAGGGCATCTTGTGGCTCGGGAGTAGCTCTCTGCTGTCTTCTCAGC
TCTGCTGACAATACTTGAGATTTTTCAGATGTCACCAACCGCCAAGAGAGCTTGATATGACTG
TATATAGTATAGTCATAAAGAACCTGAACTTGACCATATACTTATGTCATGTGGAATAATTC
TCATAGCTTCAGATAGATTATATCTGGAGTGAAGAATCCTGCCACCTATGTATCTGGCATAG
TGTGAGTCTTCATAAATGCTTACTGGTTTGAAGGGCAACAAAATAGTGAACAGAGTGAAAAT
CCCCACTAAGATCCTGGGTCCAGAAAAAGATGGGAAACCTGTTTAGCTCACCCGTGAGCCCA
TAGTTAAACTCTTTAGACAACAGGTTGTTTCCGTTTACAGAGACAATAATATTGGGTGGT
GAGCATCTGTGTGGGGGTGGGGTGGGATAGGGGATACGGGGAGAGTGAGAGAAAAGGGGAC
ACAGGGTTAATGTGAAGTCCAGGATCCCCCTCTACATTTAAAGTTGGTTTAAAGTTGGCTTTA
ATTAATAGCAACTCTTAAGATAATCAGAATTTTCTTAACCTTTTAGCCTTACTGTTGAAAAG
CCCTGTGATCTTGTACAAATCATTTGCTTCTTGGATAGTAATTTCTTTTACTAAAATGTGGG
CTTTTACTAGATGAATGTAAATGTTCTTCTAGCTCTGATATCCTTTATTCTTTATATTTTC
TAACAGATTCTGTGTAGTGGGATGAGCAGAGAACAAAACAAAATAATCCAGTGAGAAAAGC
CCGTAAATAAACCTTCAGACCAGAGATCTATTCTCTAGCTTATTTTAAAGCTCAACTTAAAAA
GAAGAACTGTTCTCTGATTCTTTTCGCCCTTCAATACACTTAATGATTTAACTCCACCCCTCT
TCAAAAGAAACAGCATTTCCTACTTTTATACTGTCTATATGATTGATTTGCACAGCTCATCT
GGCCAGAAGAGCTGAGACATCCGTTCCCCTACAAGAACTCTCCCCGGTAAGTAACCTCTCA
GCTGCTTGGCCTGTTAGTTAGCTTCTGAGATGAGTAAAAGACTTTACAGGAAACCCATAGAA
GACATTTGGCAAAACCAAGTGCTCATACAATTATCTTAAAAATATAATCTTTAAGATAAGGA
AAGGGTCACAGTTTGGAATGAGTTTTCAGACGGTTATAACATCAAAGATACAAAACATGATTG
TGAGTGAAAGACTTTAAAGGGAGCAATAGTATTTTAATAACTAACAATCCTTACCTCTCAAA
AGAAAGATTTGCAGAGAGATGAGTCTTAGCTGAAATCTTGAAATCTTATCTTCTGCTAAGGA
GAACTAAACCTCTCCAGTGAGATGCCTTCTGAATATGTGCCCAAGAAGTTGTGTCTAAG
TCTGGTTCTCTTTTTTCTTTTTCTCCAGACAAGAGGGAAGCCTAAAAATGGTCAAAATTAA
TATTAAATTACAAACGCCAAATAAAATTTTCTCTAATATATCAGTTTCATGGCACAGTTAG
TATATAATTCTTTATGGTTCAAAATTAATAATGAGCTTTTCTAGGGGCTTCTCTCAGCTGCC
TAGTCTAAGGTGCAGGGAGTTTGAGACTCACAGGGTTTAATAAGAGAAAATTCTCAGCTAGA
GCAGCTGAACTTAAATAGACTAGGCAAGACAGCTGGTTATAAGACTAACTACCCAGAATGC
ATGACATTCATCTGTGGTGGCAGACGAAACATTTTTTATTATATTATTTCTTGGGTATGTAT
GACAACTCTTAATTGTGGCAACTCAGAACTACAAACACAACTTCACAGAAAATGTGAGGA
TTTTACAATTGGCTGTTGTCTATGACCTTCCCTGGGACTTGGGCACCCGGCCATTTTCAC
TCTGACTACATCATGTACCAAACATCTGATGGTCTTGCCTTTTAATTCTCTTTTCGAGGAC
TGAGAGGGAGGGTAGCATGGTAGTTAAGAGTGCAGGCTTCCCGCATTCAAAATCGGTTGCTT
ACTAGCTGTGTGGCTTTGAGCAAGTTACTCACCCCTCTCTGTGCTTCAAGGTCCTTGTCTGCA
AAATGTGAAAAATATTTCTGCCTCATAAGGTTGCCCTAAGGATTAAATGAATGAATGGGTA

TGATGCTTAGAACAGTGATTGGCATCCAGTATGTGCCCTCGAGGCCTCTTAATTATTACTGG
CTTGCTCATAGTGCATGTTCTTTGTGGGCTAACTCTAGCGTCAATAAAAAATGTTAAGACTGA
GTTGCAGCCGGGCATGGTGGCTCATGCCTGTAATCCCAGCATTCTAGGAGGCTGAGGCAGGA
GGATCGCTTGAGCCCAGGAGTTTCGAGACCAGCCTGGGCAACATAGTGTGATCTTGTATCTAT
AAAAATAAACAAAATTAGCTTGGTGTGGTGGCGCCTGTAGTCCCCAGCCACTTGGAGGGGTG
AGGTGAGAGGATTGCTTGAGCCCGGGATGGTCCAGGCTGCAGTGAGCCATGATCGTGCCACT
GCACTCCAGCCTGGGCGACAGAGTGAGACCCTGTCTCACAACAACAACAACAACAACAAAAA
GGCTGAGCTGCACCATGCTTGACCCAGTTTCTTAAATTGTTGTCAAAGCTTCATTCACTCC
ATGGTGCTATAGAGCACAAGATTTTATTTGGTGAGATGGTGCTTTCATGAATTCACCAACA
GAGCCAAGCTCTCCATCTAGTGGACAGGGAAGCTAGCAGCAAACTTCCCTTCACTACAAAA
CTTCATTGCTTGGCCAAAAAGAGAGTTAATTCAATGTAGACATCTATGTAGGCAATTA AAAA
CCTATTGATGTATAAAACAGTTTGCATTTCATGGAGGGCAACTAAATACATTCTAGGACTTTA
TAAAAGATCACTTTTTATTTATGCACAGGGTGGACAAGATGGATTATCAAGTGTCAAGTCC
AATCTATGACATCAATTATTATACATCGGAGCCCTGCCAAAAAATCAATGTGAAGCAAAATCG
CAGCCCGCCTCCTGCCTCCGCTCTACTCACTGGTGTTCATCTTTGGTTTTGTGGGCAACATG
CTGGTCATCCTCATCCTGATAAACTGCAAAAGGCTGAAGAGCATGACTGACATCTACCTGCT
CAACCTGGCCATCTCTGACCTGTTTTTCTTCTTACTGTCCCTTCTGGGCTCACTATGCTG
CCGCCAGTGGGACTTTGGAAATACAATGTGTCAACTCTTGACAGGGCTCTATTTTATAGGC
TTCTTCTCTGGAATCTTCTTCATCATCCTCCTGACAATCGATAGGTACCTGGCTGTCTGCCA
TGCTGTGTTTGCTTTAAAAGCCAGGACGGTCACCTTTGGGGTGGTGACAAGTGTGATCACTT
GGGTGGTGGCTGTGTTTGCCTCTCTCCAGGAATCATCTTTACCAGATCTCAAAAAGAAGGT
CTTCATTACACCTGCAGCTCTCATTTTTTCATACAGTCAGTATCAATTCTGGAAAGAATTTCCA
GACATTAAAGATAGTCATCTTGGGGCTGGTCTGCGCTGCTTGTGTCATGGTCATCTGCTACT
CGGGAATCCTAAAACTCTGCTTCGGTGTGCAAATGAGAAGAAGAGGCACAGGGCTGTGAGG
CTTATCTTCACCATCATGATTGTTTATTTTCTTCTTGGGCTCCCTACAACATTGTCCCTTCT
CCTGAACACCTTCCAGGAATTCTTTGGCCTGAATAATTGCAGTAGCTCTAACAGGTTGGACC
AAGCTATGCAGGTGACAGAGACTCTTGGGATGACGCACTGCTGCATCAACCCCATCATCTAT
GCCTTTGTGCGGGGAGAAGTTCAGAACTACCTCTTAGTCTTCTTCCAAAAGCACATTGCCAA
ACGCTTCTGCAAAATGCTGTTCTATTTTCCAGCAAGAGGCTCCCGAGCGAGCAAGCTCAGTTT
ACACCCGATCCACTGGGGAGCAGGAAATATCTGTGGGCTTGTGACACGGACTCAAGTGGGCT
GGTGACCCAGTCAGAGTTGTGCACATGGCTTAGTTTTTCATACACAGCCTGGGCTGGGGGTGG
GGTGGGAGAGGTCTTTTTTAAAAGGAAGTTACTGTTATAGAGGGTCTAAGATTTCATCCATTT
ATTTGGCATCTGTTTAAAGTAGATTAGATCTTTTAAAGCCATCAATTATAGAAAGCCAAATC
AAAATATGTTGATGAAAAATAGCAACCTTTTTATCTCCCCTTCACATGCATCAAGTTATTGA
CAAACCTCTCCCTTCACTCCGAAAGTTCCTTATGTATATTTAAAAGAAAGCCTCAGAGAATTG
CTGATTCTTGAGTTTAGTGATCTGAACAGAAATACCAAAATTATTTAGAAATGTACAACCTT
TTTACCTAGTACAAGGCAACATATAGGTTGTAAATGTGTTTAAAACAGGTCTTTGTCTTGCT
ATGGGGAGAAAAGACATGAATATGATTAGTAAAGAAATGACACTTTTCATGTGTGATTCCC
CTCCAAGGTATGGTTAATAAGTTTTCACTGACTTAGAACCAGGCGAGAGACTTGTGGCCTGGG
AGAGCTGGGGAAGCTTCTTAAATGAGAAGGAATTTGAGTTGGATCATCTATTGCTGGCAAAG
ACAGAAGCCTCACTGCAAGCACTGCATGGGCAAGCTTGGCTGTAGAAGGAGACAGAGCTGGT
TGGGAAGACATGGGGAGGAAGGACAAGGCTAGATCATGAAGAACCCTTGACGGCATTGCTCCG
TCTAAGTCATGAGCTGAGCAGGGAGATCCTGGTTGGTGTTCAGAAGGTTTACTCTGTGGCC
AAAGGAGGGTCAGGAAGGATGAGCATTTAGGGCAAGGAGACCACCAACAGCCCTCAGGTGAG

GGTGAGGATGGCCTCTGCTAAGCTCAAGGCGTGAGGATGGGAAGGAGGGAGGTATTTCGTAAG
GATGGGAAGGAGGGAGGTATTTCGTGCAGCATATGAGGATGCAGAGTCAGCAGAACTGGGGTG
GATTTGGGTTGGAAGTGAGGGTCAGAGAGGAGTCAGAGAGAATCCCTAGTCTTCAAGCAGAT
TGGAGAAACCCCTGAAAAGACATCAAGCACAGAAGGAGGAGGAGGAGGTTTAGGTCAAGAAG
AAGATGGATTGGTGTAAAAGGATGGGTCTGGTTTGAGAGCTTGAACACAGTCTCACCCAGA
CTCCAGGCTGTCTTTCACTGAATGCTTCTGACTTCATAGATTTCCTTCCCATCCCAGCTGAA
ATACTGAGGGGTCTCCAGGAGGAGACTAGATTTATGAATACACGAGGTATGAGGTCTAGGAA
CATACTTCAGCTCACACATGAGATCTAGGTGAGGATTGATTACCTAGTAGTCATTTTCATGGG
TTGTTGGGAGGATTCTATGAGGCAACCACAGGCAGCATTTAGCACATACTACACATTCAATA
AGCATCAAACCTCTTAGTTACTCATTACAGGGATAGCACTGAGCAAAGCATTGAGCAAAGGGGT
CCCATAGAGGTGAGGGAAGCCTGAAAACTAAGATGCTGCCTGCCAGTGACACAAGTGTA
GGTATCATTTTTCTGCATTTAACCGTCAATAGGCAAAGGGGGGAAGGGACATATTCATTTGGA
AATAAGCTGCCTTGAGCCTTAAACCCACAAAAGTACAATTTACCAGCCTCCGTATTTTCAGA
CTGAATGGGGGTGGGGGGGCGCCTTAGGTACTTATTCCAGATGCCTTCTCCAGACAAACCA
GAAGCAACAGAAAAAATCGTCTCTCCCTCCCTTTGAAATGAATATACCCCTTAGTGTTTGGG
TATATTCATTTCAAAGGGAGAGAGAGAGGTTTTTTTTCTGTTCTGTCTCATATGATTGTGCAC
ATACTTGAGACTGTTTTGAATTTGGGGGATGGCTAAAACCATCATAGTACAGGTAAGGTGAG
GGAATAGTAAGTGGTGAGAACTACTCAGGGAATGAAGGTGTCAGAATAATAAGAGGTGCTAC
TGACTTTCTCAGCCTCTGAATATGAACGGTGAGCATTGTGGCTGTCAGCAGGAAGCAACGAA
GGGAAATGTCTTTCTTTTGCTCTTAAGTTGTGGAGAGTGCAACAGTAGCATAGGACCCTAC
CCTCTGGGCCAAGTCAAAGACATTCTGACATCTTAGTATTTGCATATTCTTATGTATGTGAA
AGTTACAAATTGCTTGAAAGAAAATATGCATCTAATAAAAAACACCTTCTAAATAATTCAT
TATATTCTTGCTCTTTTCAGTCAAGTGACATTTAGAGAATAGCACATAAACTGCCAGAGCA
TTTTATAAGCAGCTGTTTTCTTCTTAGTGTTGTGTGCATGTGTGTGTGATGTATACAAAGAG
AGAGATAATTGTATTTTTGTATTTCTTTTAAATAATTTTTTAAATTGACCCTTTTCTGAG
ACAAATTGCCAGAATAGTTTGTATTTAGAGATGGTACCTCTAAGAGTAAGGTTGCTGGTTGC
TGAGCAATTGACTTGAAAACTTTTAAATTTCAAATTTAATTCCACTACTCAAAGAATTGC
CATGTTTTTAAAAAAGAGAATTGGTGCCATAAGTTAGTTGTCTATGTTTGAAAATGAAGAAGA
TATGCAACGTCATGGCCTGGTCACTTACCCGCAGCCCTGAGTTGTAGGCACATCATATGTGA
GAATGAGGATGCTTTTCTTTCAATTTAAATCCCTCCCCAAAACCTGGCTCTAATTGCAGTCA
TGACAATCATGTACATTTGGATTTATGTGCACGAGTCTCTTACCCTGAGAGAGGACAGGTGC
TACAGGTGGAGGGGACCCGTCTGGGTACGTTACATTTTGAACATGCTGGTTTTTCAGTCAC
TGCACACTCATCTCCCAGCACAGGTGATGGGCAGCAGATGCAAAAGCTGCCCGTGGTCCTAT
TTGGAGGTGCATGAAATGAGCAGAAGACAGAACAGCTTGATCTGACTAGAAGGGCAGCTTGT
CCCTACCAAGACTTGAAGGATTGCCTTTTCATCTGTTAGGGTAAAAGGTAGAATGAACCAAGG
AAGGGCAGGAGGGGGCTGGGGTTAGGGTAGAAGGAAGGGGCCATGGAGAAGGGAGATCCATC
CCATAGGAGGAAGGCAGTGCGGCAGGGAGGTTTGAAGGTATCAGCTTTTGTGGCTGACATAC
ATGCAGTCATGTCAATTGCTCGTTTTTCTTTTCCATCTTATTAAATGTCTTCCAACGTTAG
CACGAAGAAAAGCTATTTGCAGTGTTGCCAGCCTTTCCAGAGCCCGTCCCCATTACCTCCCC
AGGCCCATGCCTTTACTCCTTGAGTTTCAACTCACGACCTTCAGGATCTGACTTTTATTCAC
CAACTCTGGGGTGAACGTACCTTCTGTCTCCACCCAGAGGTCTCTATCAAAGAGGAGATTGC
ATGCCATGGATAAAGTCAAAGTAGAGGTGACTGTCCTTAGGAAGAGTAATGTGAAAATTCAT
AAACTGGGATTCTGTTTACATTTTGTACTCCAGGGGTTCTTAGTTTAAATCGCTCTGAATAA
ATTAAGATGCAATGGCATTTCAACTGTTATGATTAAATTTACAAATCATTTATTTTCTATCA

CGGGGAGAGATAGAGCTCCAAATGCAAACATAACTGCTCAAGTGTTAACACTTATAATGAAA
ACATAAGAATTACCACCAACTACCCTGGGGGCTAGAAGCAGAAATGTGAACCAGAAAAACAAA
TCATGAACTTTCTTTTTTTTTTTGAGATGGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCA
ATGGTGCGATCTCGGCTCACTGCAACCACTGCCTCCCGGGTTCAAGCAATTCTCCTGCCTCA
GCCTCCTGAGTAGCTGGGACTACAGGCATGCACCACCACGCCTGGGTAATTTTTTGTATTTT
TAGTAGAGACAGGGTTTCACCGTATTAGCCAGGATGCTCTCGATCTCCTGACCTCGTGATCT
GCCCCGCTCGGCCCTCCACCGAAGTGCTGGGATTACAGGCATGAGCCACTGTGCCCCGCCAA
CAAATCATGAACTTTCTAACTGCAGTTTCTTGTAGCTTGTAAACACATCCACTTACTTATTG
TCAGAGTACGTGGAGATTTTCCACAACCCTCGGGGATAAGGCTGAACAGAAGAGGCCAAAAAC
GTGAAAACATTTGATAGCTCCTATACTTTGAAATAAAATTCAGTGTAAAAGTTGCTTGTAT
TTTTCCAAAACAGAGTCAACCCTTAATATTTAAGATTCTGTATACAAATACATATTTTTTATA
TAATTAATATATATTGTCATATGACATATATCTTTATATTAATATGCATGCATATAATATAT
ATTTCTTCTTAATTTTCTATAAGCAATTTTACAAGACTGACTTCTATTTGCCTCCTTATTG
TTACTACGTGGTTTGATAATCCGTTTTGTGTCATTGTGATTCTGTATGTTTTGGGGACTTA
TTTTTGTTTTCTCTGGGTGGTCACTAGTTTTTTTTAAAGCATTTCATGGAAGAGTGTGAATCTTT
TACAAGCTAGGAAGCCATGGCAAGCCTTGGGTCACTGCCCCCGCGAGGCCACATTGGCAA
ACCAGCAAGGGTGTTCAACTTCCAGACTTGGCCATGGAGAAGACACACGAGGAGGCTTTTCA
CATTACAGCTCTTTAATGTTTGTCTCTGCCGGCACCATCCAGTTGTGAAAAGAGGTATTTTC
CACAGCGGCTCAGGGTAGGTAGTGCACAGCTCACATTCATCATTTCTGAAAACCGAGAGGAG
TCTCCATTGCGGGTACAGGTTGATGCCTGTCGTGGAATGAAGGTTCCAACACCCAGACCAAT
CTCTGCAGTGTGCTGCTCTCATGAGCTTGCAACAAGATCAGAAAATGTTTTGTGACTAAGCA
TTTTTCATATTGCATAAAATGCTTCAAGCTCCTCCCTTGTCTCTCTATAATCCTGTATAT
CTGATGATTGTGGGTACCAAGTGTTTGAAATAATCAAATGTGATTGATGTTGGTAAATTTTC
TTTTTTTTTTTTTTTTTACTTCTATTTTTTTTTATTATACTTTAAGTTTTAGGGTACATGTGC
ACATTGTGCAGGTTAGTTACATATGTATACATGTGCCATGCTGGTGCGCTGCACCCACTAAC
TCGTATCTAGCATTAGGTATATCTCCCAATGCTATCCCTCCCCCTCCCCCACCCACCA
CAGTCCCCAAAGTGATATTCCCCTTCCTATGTCCATGTGATCTCATTGTTCAATTCCCAC
CTATGAGTGAGAATATGCGGTGTTTGGTTTTTTGTTCTTGCGATAGTTTACTGAGAATGATG
GTTTCCAATTTTCATCCATGTCCCTACAAAGGACATGAACATAGCAAAGACTTGGAACCAACC
CAAATGTCCAACAATGATAGACTGGATTAAGAAAATGTGGCACATATACACCATGGTAAATT
TCTTTATCATTGCACTCTCCTTTCTCTATTATTGTTATTGTAAGTGAACCGCAGATTAGTC
ACTCATTGCTTGCAGAAATCCAATTAACAAGAGCGAGGTGAGATATAAAGAAAATGATTTATT
CCAAACCTCCTTCAGGGAAGAGGTGCAGCCTCCTGCCTCTAAATGCACTGCTTCGCCAGGCG
TGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGACCGAGGAGGGCAGATCACTTAAGGT
CAGGAGTTCAAGACCGGCCTGGCCAATATAGTGAACCCCTGCCTCTACTAAAAATACAAAA
AATTAGCCAGACGTGGTGGCGGGTGCTTGTAAATCCCAGCTACTCGGGAGGCTGAGGCAGGAG
AATCGCTTGAACCTGGGAGGTGGAAGTTGCAGTGAGCTGACATCTAGCCACTGCACTCCAGC
CTGGGTGACAGAGTGAGACTCTGTCTCAAATAAATAAATAAATAAATAAATAAATAAATAA
ATAAATAGTAAATGCACTGCTTTGCTTTTGGAGCAGAAAGCAGGCACCTTTGAAAAGGCAGGG
GAGGAAGTGAGCAAGGGCAGGGGTCTGCACACTGGCATGGTGCCTGATCTATCCAGGCAGT
TGAATTGGCACTTTTCATAGGCAGAAATAAGTTGAAAAAGTGGCCTAAACTCTCTAGGTGGG
AGTGGATAGTGGGCATGCCTTCAACCTGCCTTTCTGGAGGGTGAGTTCCATGGCAACCCCT
GAAGGGTGAGAGTTCCATGGAGATCATGCTTTGGTCTGTAAATCAGCTGTTAACTCTCTAGA
AAGTTCTGTCTTGAGCATATAGTTAGATGAACTTGCCCTGTAAAGAATGTCTGGTGAAGGG

GAAGTAAAAGGTGAGATTTGCATTTCTAAAGGGCTAAGTAGAAAAGTGGGGTACAAGAGGAAA
GGAGAAAAGAGAAAATAATTTAAAAAATAATTGTAACCTATTCCCTTTTACTTAGAAAAAAG
GGAATACTCAGTTACATTATCACCTCGTTTACATCAAACCTCTTATGGAATCCTATGGTTT
GAAAACAAAAAGGTTGTTGAGGACCAGTGAGCCCAACCCCTTTGCTTTATAAATGAAGAGCA
TTGCCTGCCCTAAGCCCCAGAGACTCTGATGTCGTGGGTCTGGAGTGGGCTCCAACAGCGGC
ATGTTTTGATGGTGCTTCCCAGTGGCACGCCAGCGATGAGCCTTTGAGTAGGGAAAGTAGGA
GCACTCGTGA CTCCCTTACGATCAGCACCTGTGTGCTAATAAATTCACAAAAGCCAACATA
TTGGAGTCACTCAGGGAGTTTTACAAATAGTGAGGTTAAATCCAACCTCAAATAGTTCTGAT
TCGATCTGCCTGCATTGCTGCCCTGTGGTTCCCCACTGTAGAAGCTCCCCAGGTGATTCTAA
GTGTAGCCAAGTCTGAGAAATACTGCCTAAAGCCTGTTGGACTGACAGCAAGGGCTGTTGTC
TGAGCAAGACTTTGCCTGGCCTGGGGTGGCATGTGCACCAGGAAGAGTCTCAACTTTCATAA
CAGAACATTCCCCAAGCTGGTTTTTTTTTAAAGCATGTGAATCTAGACTTCATTGGCAATACCA
AAGATCTGTATTTGAGGCTCCAAGTATTTCACTTTTCAATTTTGGTTTTGGGTTATGTTTTCA
CCCTTCCTTTTCCAAGTGAAAAGTAAACAGAAGTGGGATGTCTGGCGCCCATGCTGAGCTTGG
CAACTTCAAATTCAATAGAGAAGAAGTCTCTTGATAGAAAAGGGCCTGTCTGAGATGTTTC
TCAAATAAATATAGATTTTGCTTATGTGGCTAAAGGATTCTTCTCCCCCATTTCTTATCC
CTGCAGTGAGCCATCCTTCTTAACTCTTTCCATGAAAGCATTATTCCTGAAGAACTGGGAAC
TCATGCCAGCCCTGATCAGGCAATGATAATTCTGCAGAGAATTAGAATTTAGATTTAAATTG
TCAACTCTTATACATCCTGGCATATGGTTTTAAACACATGTACACACACACAAACACCTCCTA
CTATTTACTGAAGAGCAGATATCTGATAACTTAATCTTTTTGGTTTTGAGTCAAGACAATTC
CTCCTTTTGAACTGCATACCGCTGAATATAATAAATGTAATTAAGATTAAAAATAAGAAA
CTAATGGGAGAATTTCAATATTGTCTATGTTCACTTTAAATTCCTCTACTTAGGTTTACTG
CCATTACCAAAGACTATTCAAAAATCCTTTTTAGGAGAATCCTAATGGTTTCCTGACATATA
ATCAAATAAGGACTCTGTTGATTGGCTAACTCAATCTTCTGTGCCAAAAAGCAGAGCCCAG
CAGAGAAGAGGGCAGGGACTTGAAAGTCAGACTGACTCGAGTTCCAGCCTTGGGGCTGTGGG
AGCTTGGGCAAGTGACTTAACGTCTCTGGCTCTCAGGATCTAAAAGGATTTCCAGTAGTAAT
TTGGGGTGTTACTGATACAGGAGCTAAAAAGAAATTATTTAGGTGGTTAGTGAGGGTCAGAG
AGTCCTCGGTAAAGATTTGCCTTTTAAACAAAAAGCAGCCCCAAAATCATTTGTTTGCTAACAA
AGAGAAGCCTGTAAAATTGAGCTGCAGACATAGATAAGCAAGCTGGAAGCTTGCACGGGTGA
ATGCCGGCAGCTGTGCCAATAGGAAAAGGCTATCTGGGGGCCAGGCATGTTCAACATGGATT
CTCCATCTTCCCTTTTCTTGTCAACCAAGTGTACAGTAAAGGAACAGGCAACATGGCACGG
GCCAGGTAGAGAACCCTTCTGCATAATAAAAGATTAGGGTGAGATGGCCAGCTTCTTCCCGT
GCTATGTAAATGGCATACTGGTCCAACAGTCTTTTTGGGCCCTGTGTAAATCAGACACCGC
CTCCTCAAGTTAGTCTATAAAACCCCATGCATTTTACCGTGAAACTGGGAGATCCACTCGGA
ACCCCTCCTGCACGAGAGACCTTTTCTCTTTTGCCTATTACACTTCCGCTCTTAACTCAC
TGCTCATGTGTTAGCATCCTTGATTTCTTGGCATGAGGCAACGAACCTTGTGTATTACCCC
ATACAAATGATGCTGCTTCATTACTAATAGCAACCTGACAGGGTTGTGTTGGGGTATAAATT
ATCTAGACCAGGGAGATCCAATATAATTTTTTTGTAATGACGGGAATGCTTTGTATCTGCAT
CATCCAAAATGGTAGCCACCAGGCCAGGGTGAATGTGGCCAGTGTGACTGAGGAACTGAAT
GTTTTCCATGATTTAATTTAAATGTGGCCAATGGCTACTGTAGGAGACAGTGTGAGTCTGGC
ATATTATAAATAAATAATATTAATATAATTTGAACTTTGGCATCAGTGTTCCTAGATTTGA
ATTACTATGCAAGTTGCTTACTGTTTCCAAGCCTCAGCTTTCTAATCTGTAATTGGGGCTAA
TAATAGTATCTGCCTTACAGGTTTGTTCAGAGGATAAATGAGAAATTGCATGTTGAGGGCTT
AACACAGTGCTTGGCACATAAAAGCTCTGGTAACAGTTAGCCACTTTAATAATTTGCTAATA

ATGGCTATTTCTTCTTCAGATTAGGATGTGCTCCCCCAAACAGTGCACCTTAGACATAGCGGG
CAATCCAGCTCACTCTCTGCAGTGAGAGAGAAGCACTGGCCGACCAGAGTCAGCCAGGGGCT
CATGGGTATGAAATCAACAGCATGATTTTGTAAGTAATGGATGGAAAGGGCCTCACAACTTT
ATGGCACTGTGTTCAATTTGCTTGGTCTTCTGTAGCTCCTTTTGAAAGCCTTTTAGGGTGGA
TTAACCTGCTACCAATAATTCTGGTCAGATGTAGACTCCATAGCTCAAAGCAAACCTGAGAGA
GTGAGGGCAGCAGGCCAATTCCCCACCCCTTCCTTCTGGACTCTGACAGAAGCTTACACTCA
AGGAAGAGCAAGTAGGAATTAACGTGTTAAGAGCTAGGTAAGCAAAACCAATGAGAAAGTTC
TGGCAAAGCCCCATGGGCAGGGGTGGCTTAGGCACAGGAAACAAGTAGGATTTTCATACCACG
CGCCTCAGTCTACTTCCGGGGCCCTCATCCTCAGCTGTGCCTATGCAAAGGAGAGCAACCAA
TAAACCCACCGCCACTCTCCTACTGTGGAGGCCAGGGATGGCCAGGGGTAAGAGAGGGATG
GGAAGTGTTCCTCCAGCCGTCTCTGAGAAGGAGAGGAAACTGGGCAGAGCTTCTGTCTCTC
CTTCAAGCAGAAACAGAAACAAAAGAAACCCCTAAGGGGGTTTCTTACTTCCCCTCTAGTTCA
GTTGTGCTACTAACCATCTGCAGCTCAACATTTCAGCATTTCATTGATTTCAGCAAACATTG
AAGGAGGGCCAGCTATGTGCCAGATGCCAACTCATGCCATGAAAGAGAGTCCCTGTCTTAT
GAAATTCATCTATTTAGAGAGAAAAGCAAGCAAAAGGCAAAAGTTTGAAAAGTACTGTTGAAG
TGGCATCATTGTCTGGGGTGAATACCTGAGGTTTGTGGTCTCACGCCAAGGGAATCAAGGAC
TCAGGCACACAAGAAGTGAGTTTAAGAGCAGAGGTTTAATAGGCAAAAGAAAGAGAAAAAAG
AATAGCTCTCTTGCCTGCACAGAGAGAGGGGCACCTGAGTGGATCTTCCTGTTTTGTGGTGA
AATGCAAGGCATTTTATAGACGAGCTTGAGGAAGTGGTGTCTGATTTACTTAGGACCCGAGA
GATTGGTCAGACCAGGTGTGATGTTTACATAGCATACAAAGAAGCTGGCCATCCCATCCTAA
TCTTTTATTACGCAGACGGGGTCTATACCTGGCTGGTGCCATGTTGTCTGTTCTTACTGTGA
CACGTGGTTGACAAAGAAAAGGGAAGATGAAGAATCCATGTTGAACATGCCTGGCCCCCAGA
TAGTCTTTTCTATTGGCACAGCTGCCGGCATTCACTCTTGCAAGCTTCCAGATTGCTTATC
TATGTCTGCAGCCCAATTTTACAGGTTGCTCTTTGCTAGAAAAGAAATGATTTGGGGGCTGC
TTTTCATTA

hRosa26 sequence
(SEQ ID NO:4)

(Putative guides for insertion of a integration site are indicated)

ACCATTTAAACCTCAAATTAAGCAACCCACAGAACCAGGAAGTTCAAGGACCATGTCTGTTT
TCACCACGATGTCTTTCCCCACCCCCCACCCTCCACCCCCCACTAAGGGCAGGGTA
TTGTATCTGCCAGACTGGGTATTTGTTGAACAAGCGAGTATTTTCGCCTATTAGCTTAGTTT
TTAAGGAAATCATTTTTTACTTGATTTCATCATAGCTTTAATTCTATTACATACTACAATAAA
AATTTGACAAGACTGATACAAATATGTAGTGGGCAATAGTTTGCCGTCTTCTTCCCTAGTAT
GGTGTTTTTTCAATCTGGTGACTAGAATAGGCAGTGGGCTATAAGCAGGATTTCATAAGGCCTG
GAGCTGAGTTATATGTGACACTGCCACCTATTCATTGTGTGACCTTGGTTTTAACCTTCAAA
GTGGGTCTCCTGGACTAAAAGAATGTGAAAAGATGGGGAAATAAATCTGTAATCTGAACATG
GAATGACTTAGTTACAGACCAGACATATTGTTACTGGGAATGAAAAAGTCAATATATTTGAG
GGGAAAAAATGTAAATAAATATTGAGAAAGATTTTACAAATCTAATTAGGGGAAGATAGTT
ATCTCCCAATACTAGAGGGTACCAGAGTGCTTTTAAGGGGAACATTTGGTTACCCTTATTTT
TTTAAAAAATGGCAGTTTAGGAAATCTGCCCTAACTGTAGTCCCAATGTCAGATAGGACTC
AGGTCTCCACTGCAAGGACCAAAATGTTAAGTTGAAGACTGAAAATGGGAAAATTTGGAAAT
GTCTTTGGAACCTCAAGTACATAAAAGCCTGTAAGTGCTTCATACTCATTAAACAACATAGGC
ATAGAAAAAAGATATCCTTATTCTCAAGCATAGCCTTTTCTAATAAGTTTCATGTTAGATGTC
ATGAAGTTTAGTGAGGAGTGAAAATCTATGAGGAAAAACATGAACCATTCTACTCTGGCAAA

AGTTCAGGACAAACACCATAGGCCTGTATACCAAATTTTAAACCATGTTGAATAATGTGAAA
AAAAGCATCACATTGCTTATGAAAGGCTTTTCCTGTCGCCCCCTAATACTTCTGTCTCAGGCT
AACATGTTTTGTTAATGAGTTACAGTGGTGAAGTTAAGGAAATCTGCTTCCTGTCTCAGCATG
CCCATTATCCCAGCCATACAGATTTAATACCAGGAGTCACTTTAACTCCATGAAGTCATTCA
ACAGGTACTTGAGTATTTACTATGTGCGTTTGTGCTAGAGTAGCCATTTCTTAAACTTTGTG
GCCTCAGAGAATCCCTTCACATTCTTAAAGATTGAGGGCCCCAAATCACTTTACTAGTATCT
TACCATATTAGAAAGTAAATATTTAAATATTATAGCAACAAATGCATATTTTTAAAAAAT
AGGTGCATTGTTTTACATTTTGGCAATTGTTTCCCTAAATGTCTGACTTAAATAGAAGACAA
CTGAATTGTTTTCTTTTGCATTCAATTGGTTACAATGTTACTAGCAGTTTTCTATAATCTCAC
GTATTAGTCATTAGGAAAATATAGCTTCACTGAGTTATGTCAATCTCCCAAATGTTGGCCCA
TTTTACTGTATACTACCTAAATATCATTGGCCGGGCACAGTGGCTTAAACCTGTAATCCTA
GCACCTTTGGGAGGCCAAGGTGGATCACCTGAGGTGAGGAGTTCAAGACCACCCTGGCCAACA
TGGTGAAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGTACAGTGGTACACACCTGT
AGTCCCAGTTATGCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCCAGGAGACAGAGGTTGC
AGTCAGCCAGATGTCCCAAAAAAAAAAAAAAAAAATTATTTTCAATATCACTATCTCATGAAGTA
TTGGGAAGCTGTCAAGTTCTGATATTAGACACAAGTTTTCCCGAAATTCTGATTTTGACTC
AACGTTTGGATTTTATCATTGAAAACAATTGCTGTGAGTTGTTAGGCTCCAAGGAAATAGCA
GATAATTCAGCTTACATGAGTGCTTTTTCTTGAGACAACCATTTCAAAAAGTTATGTACTGT
AGGGTTTTAAGATTTAACAATGTCACTGCTTTCACAAGGACATTCTTGAGTGAAACTGGTT
TTTTTCTTTTGTGGGGTTTACACCACAAATGCATGGCAGTGAAAAATAACTTAGAGTTTGA
TGCCACTGCCACAGTTTGTGCCAAGGTGCCTGAAATTTTACTTTTACCTACTGTTGCCTTAT
CACCCTCTTATGTCAACATATAGTTTAGCATAAACCATGAGATTTTAAAAAGTTATTCTAC
ACTTGCATTATTTTCAAGGACATGTGTTTGTGCAAGCTTTTACGTAAGAGTATCTTTAACTA
GTTGGTGCTGATGCCTGGCAAATACAAGCCAAGTAACAAGTCCAGCCATGTTTATGCACGCA
TCCATTGATAACGTATTAGCACAGTCAGCCCTCCATATCCTCAGGTTCTGCATCTGCAGATT
CAACCAAATGTGCTTGAAATATTTGAGAATATTTAATACAATAGTACAAATATAAGTACAC
AATGTAACAACATTTTATATAGCATTACATTGTATGAAGTATAAGAAATCTGGAAATGATT
TAAAGTATATGGGAGGATGTGTGTAAGTTGTATGCAGATACAGCGCCATTTTATAAAAGGGA
ATTCAACATCCTTGGGTTTTGGTGTCTTGGCAAATGGCCAGCAAGGGTGGGGGAGTGCTGT
CCAGAAACCAATCCCAAAGCAAGGGACAAATATATACTTCTACAGATGAGATATTAACCTCAG
AATCCATGTCTGCACACACATCTTTAATGACAAGTTGCTTTTACTCAACAGCGGCACTAC
TATGATCTTTTACTTACACTTCAGCCAGCAGAGACATTCAAGCAGTGTGAGATTTTTTAAAGTT
TTTCAGCTATTGTGTATGTTTCTAGTGTATAATAAAGTAACTTATCCTTTAAGATACTTAAG
TAGCTTTTCACTTCTAGCTTTTAAACCTGTTTTTTTTTTTTTTTCCCAGTAGTGGCATACTGC
ATTAATAATAATGCCTTACCAAAAAAGCACTCTGAATGATTGGTTTCAAATGATGCCAC
AACATAGGTGGCACCAACACTATTCAAGATCATTCCATTCCCATCTCTAAAAAATTTTTGG
CTGGGTATGGTGGCTCACGCCTATTAACCTAACATTTTGAGAGGCCCAAAGCAAGATCACTC
AGGGCTAGGAGTTGAAGACTAGCCTGAGCAACATGGCAAGATCCTGTCTCAAAATTCTTTTT
AAAATTTTTTTTAAAAGCCAGGCTTGGTGGCGCATGCCTATAGTTCCAGCTACTCAGGAGGC
TGAGGCAGAAGGATCTCTTGAACCCAGGAGTTTCAGGCTGTAGTTCACTATGATGGCAGCTG
TGAATAGCCTGGGCAACACAGCAAGACCTCATCTCCAAAAAAGACAAAAGAATAAATTAT
TCTACTGCAGAACATGATTAGGTAAATATCTCCAAAGCAGAAAGACAGTTTCATATTTTCG
TTAGTTTGTAGTCAGTCCTTCCAAATCAAATCTTGTTTTTTATTAGTATACAGATGGTATAGC
CAGTAAGTAAATGAGAAGCAGTCTTTTTAAGCCGATCCATTCTTAAATGAAAAATATATAA
ATATTTTAGAATAAATTTATTAAATCTAAAGTTGTAGAATTTTTTAAATTTGGATATTTTGG
GAAATATTTTAAACCACTATTGCAAAACAAACAAATGTACTTATGTTTATACTTAGGC
ACAAAGAAAACCTACAGTATTTTAAAGTAACCATTACACAATATTGAGGTTGCAAAGATTACT
GAAGGCATAACCTAAAAATGAGTTGATTTCTAAAAATGGGAAAAAGGAAAAAATAATTC
TAAAAACAAGTATGCATACCTAAACCTACCTAATGACACCTTAGAAAATTCAGGTATAGCAC
CATTCATTAACATCAATGAGGATGTCATCACACATCATGTAGCCTCTGCACACCGTGAGAAT
AAATGAAAAAGACAGGCATCTTGCTATCATGACAATAGTTTTGACCTCGCAGACCTCTCTGT

GCTTACGCAACGGATAAAGCCATAAAGAACTGTCCTGCCCTCAAGGAGCAACCTAAAAGTAGGA
AAAAAAAAAACAAAATTACACAATTATTATTTACAATTGTGAGAAGAGCTCTTGACAACATTG
AAATGGAGGATACAGTGTAGTAAGGGTGAGGGTATCAAGGCTTCCTTGAGAAGTGATGTTTT
GAGGCCATTCTTTCTTCTCAATAAACTGGTATTTGGTTCTGAATCCTTTAACTTCCTTACCA
TTGTCACTCCTAAGCCAAATCTCATTACGTATGTCTAGACTACTGTTAAGAGAACCCTTA
AGTGGTCTCTGCAGCCCTCAATTTATTGGTGTATCTATGGGAAAATTGCTCAAACTCTAAG
CCTTAGTTTTCTACCCTATAAAATGGGGTTTTATATACAAGGAACATACTAAATACACAGGT
ATACCTCAGAAAACACGGCAGGCTCAATTCCAGAGCACTACAATAAAGCGAATCTCATGAATT
TGTTGGTTTTCCCAGTGCATAAATTATGGTTACACTATAACCATAGTCTATTAAGAAGTGTA
ATAGCATTATGTATAAATTGATAAATACATCACTGCTAAAAAAATGCTAACAATCTTTTTGCT
GGTGGAGGGTCTTACGCCAACGTTAACGATGGCTACTGACTCATCAGGGTGGTGGTGGTGA
AGATTACGATAGCTGTGGCAATTTCTTAAAAGACAATGAAGTTTGCCACACTGACATCCTTT
CACAAGACTTCTCTGTAGCATGTGATGCTGTTTGATAGCATGATAGCATTTTACCCACAGTA
GAACTTTTTTTTCTTTTCTTTTTTTTTTTTTTTTTTAAACGCAAGGTCTCACTCTGTACCCCA
GGCTGGAGTGCAGGGGCGCCATCTCGGCTTACTGCAACCTCCTCCTCCCTGGTTCAAGAGAT
TCTCCTGCCTCAGCTTCCCAAGTAGCTGGGACTACAGGTGTGCACCACACCTGGCTAATTTG
GTAGAGGTGGGGTTTTACCATGTTGGCAAGGCTGGTCTTGAACTCCTGACCTCAAATGATCT
ACCAGTCTCGGCCCTCTAAAGTACTGGGATTGCAGGTGTGAGCCACCACACCCAGCCAGTGG
AACTTATTTCAAATTTGAAGTCAACTCTCTCACACCCTGGCACTGCTTTATCAACCAGGTTT
CTGTAATTCCTAAATCCTTTATTGGCATTTTAACAATGTTTACAGCAACTTCACCAGTAGAT
TCCATCTCGAGAAACCCTTTCTTTGCTCATCCCTAAAAAGCAACTCCTCATCCATTCAAAT
TTGATCATGAGATTGCAGCAATTCAGTCACATCTTCAATGCTTTACTTCCAGTTCTAGTTCT
CTTCTGTTTCCACACCTGCAGTACACAAAAAGCATTCAATAACTATTACTTCTATTCTTCT
ACCTATGTTTCCATTAGCTTTTGCCTATAGTACGCACTAGAGTATGTTACCATTATTTGTTA
TAAGTAGTACCTCATTATTACACTATTCGTAAGCAATACCTCAAGGTCTAAGATTAGATTTT
AAATCAAGGTGAGTAAAAATAGAAAAGGCTGTGAAGACTGTTGACTGACTTTACCAGAATCC
ATACACTAGAGGTGAGATTAGTTAGGTGATGAAATAACCATTCTATAAACATGATCTGAAAC
TCTGTTACTGTTGTCAGCAGGAAAAGCCAATGTTACATATGTTTAAAAAAGAAAAAAGAAC
CCAAAACCAGAAAACAAAAGGTGACAAAGTATCAAGACAAAAGGTCACTGATGACTGATCTC
TAGGAAAAGCTGGAAAGCAGGATTATTAATGTAACCACGACTAAGATAAAAAATCAGAGACA
GAAAAGTCTTTGTCACCAAGAAGATATACTCCATGAGAGAGCAGAAACAATTCATCAGGTTT
AACCCTGCTCTAGATAAAATAAACTATCTGATTCAATACTCACACTTCTCTAATAATCCAA
TACATTATCCCATCTCAAGAAGAGAGAGTCACAGATAAGAAAAAAGGCTTCTTGAGAAGT
ATGTGCTCTAATATAAACTAATATGCCACTAAGAAAGCAACCTGCAAAGTCCAGTACCAGAC
TTCTGGATTTGTGACCTAACAAGGTGCTCTACAATTAACCTAACAGTCAAACCAGAGTGTTG
TAAAAGAGAATTATGTAATTATGCCAAACCTCCACTCACAAAAATATATGGAAGTAACCTA
AGTTTACATTTTGCAAATCTCACACACACACTAGCCCTGACAAAAGTTTACCAGCTTTCTC
ATCCAAGTACAAGCGTGTAATATACTTAATAAATTTGTCTTATAAGGGTAAGAAATAGTATG
TAATACTTTGAAAAGGAGATAGGTAGCTGGTTAATTTAAACAAAAAGCCCAAGGAAGTAAGG
TGCAGGAAAAGGATAACTGCAATGATTAGTACAGGAAACCCAAAGAAGAACTGAATGGTGGG
ATAGATGTACTCAGAGACCATGAGGCATCAGTTTCTCTATGAATAGAATATTAGGAGATGT
AGGTAAATGGGACCCTGAAGTCTCTCCCAAAAAGCCTTGTTTATATGTTTTCTGAGCTTAA
CTATTACTTGAGAATCAATTTACGTATAAACCAACAAAATAACATTTATTGAGCTTCCAG
CTCTGTGCTTAGGCACTGAAAAATCACTTTCTTAAAGGATTGCAATTAAGCAGGAGAAACAC
AAATAAGGTGAACTTCTCTTGTTGCAAAGAATATATTTCAACATTCCTTTTAAAAGGAAAAC
CTGACCTGCAAGTTTCCAAAAATATTAATTACTATTCTCTTTGCCTCTCAAAATTTCCATT
CTGTTATTTTTTAGGAGGAGGAAAAACAGTTTCAATTTGAGGAAAAATTGAGGGTCACATACT
ATACAATTGAGAAGAGTTTCTCTGAACTGTAATCATTTTTTGGCAGGTAAATAGGCATATCC
GAGTCAGCAATGAACTTGAAGATACTGAGTTATACTGCCTGCCCTGTGGGGTTCCACCTTC
CCCAAGAAGTTCAGAAATTTTTGGGTGATCTGAGAATCTACATTAAGACAACCTGTCTCCACA
CACAGGAGGCCTGAAGATCGCTGACATAAGGGTCTTTTTAAAAAGTATATTTAATGGCCTAG

GGCGGTGGCTCACACCTGTAATCCCAGGACTTTGGGAAGCTTAGGGCAGGAAGATCACTTGA
GCCCAGGAGTTCTAACCTGTGCAGCACAGCAAAACCCATCTCTACAAAAAACAACAA
AAAAATTAGCTGGGCATGGAAGCGTGTGCCTGTAGTTCCAGCTACTCAGGAGGCTGAGGCAG
GAGGATCACTTGAGCCCAGGAAGTCAAGGCTGCGTGAGCCATGATCATGCCATTGCAATCCA
GTATGTGACACTAAGACTCCGTCTCAAAAAAAGATAATTAATTTGTTAAGATAC
TGTATTAGCAATATAAAAAAGCATTGTTGTTAATTTGTTGTTATTATAATTCCTCAGGATAA
AACTTACTTTGTGATTGTTTTCTATAACTCAAGATATGATGCTTAGAGCTCCTCCAATCAAG
TGTTTCCAGGAAGTGAAGACTTGTAGGACAGAAATTTAGGCTGGGTTTCAATTTGTATCACACA
GACCTATTCTTCATTCAAGTTCTGATATATTTAACTATGTAGCTCCTGTAACAGTTTAATGG
AATCTCACCTCCCTAAAATTCATTATGCATTTTTTTTTTGAATCCAACTCATTAAACGCTTG
CTTTCCTGTTGTCCAAGGCAGGCACATCTTTAAAAATGGTTTGTGGACTTAGCTTTTCAGC
TAAATATATAATAAATAAAACAAAAACAAGCAGTTAAATGAAATGTAATGGGCCAGAGAGCTT
CAGCTTTTATTTCTTACTGCTCAGTAAAAAGAGAAAACCATCAATGTCCACGTATTCTGTA
ATCCACAGAACAAGTCCGGGGCTACAGCTATACTGTCCACAGTTGCAATTCAAATTAGATAA
AAAATAAAAAATTCAGTTCTTTAGTCATACCAGCCACTTTTCCAATGCTCAAGATTAATAAAA
TGTCAAACCATAAAGACATTTACATGTGCTCCTCCTTAAAGTTGGCTAGACATC
AGAGTATACTAGGAGCTCAGGAGTACAAGACACTATTCTTCAAAAAGCTCAGAATAGTTAA
GGTAATTTAAATCAGCAATGACAACAACCCAGAAATTACTATGACCCACGCAGTACAACTG
CTCAGGAGTCAGAAGAAAAGTCTTTTTTAAAGGGCAGTTGGGTCATAGAACAACAGACC
ATGGAAGGCATGACCAAGGGGAGATGACATTTGAATCTGCAGGATTAAAAGCAGCAAGGGT
AGCATTCCAAAAAGAACCACCCACAAAAGATATATGACGTCTCTATGATTTGGGTAACTGCA
ATTCATTCCATGTGACTTCAGGAGAGAGGTCAATTTGTGTGTGTAGTATGTGGAAAAAGT
GAAAAATGAAAAAGCTGTTAAATTGAGGAAAGTCTATCCAGGGACCTTATGCATCACATTCA
CGAGAACAGAAATTCATCCTGTAAACCAGGGGTGTCCAATCTTTCGGCTTCCCTGGGCCACAC
TGCAAGAACTGTCTTGGGCCACATATAAAGGACAGCTGATGAGCAAAAAAACAACAGACA
ACAACAACAAAAAACAACCCCGCAAAAAAAGCTCTTAAACTTTAAGAAAGTTTACGAATT
TGTGTTGGGTGCGATTCAAAGCTGTCTTGGGTCCCATGCGGCCCGCGGGTTAGACAAGTGC
TGTAACAGTACAAGCCAGTAATGGAGTTTACCTGTCAATTTTCATGCTCTATCTTCTTTA
GGACAATCATCCTAACAAGATGTAAGATGGATCAAAAGATAACACTAAAGACAGAGACAGCA
ATTTGGAAGCTATCACACAGGCATCTGAGATCAGTTACTAACTGGTAAGAACAGAAATGAGA
GGTATTTAGAGGAAGAAAAAGGGAGATGTTGCCTAACCTCAGATCCAATTCTCTGTAAAGCA
GTAGTCAAGATCACCTGGACTGTGAAGACGGTCAGGGACAGAATCCCAGCTAAGGAAAAAGG
ATAAAATGAAAATCAAGATAAACATTTAAGAACGTGAACTAGGGAGGAATAAAGCACTGCT
GGGTAAGAGTCAAGCCCCAGCTCAAGCCTTAATTTGTGGTGAACCAATCTGTCTGGTTTCG
CGAGACACCAGGCTACCCAAGATCAAGAGAGGGAGAAAGCTAGTGCTATGTCTGAATACTAG
AGGAGCAAGTACAACAATGGAAAATGGGATCAAGTATGAGTGAGAGTTGCTAAGATGCCTG
GTAGGGATGCAAAGGGGTAGAGAGCCTGGGGAGAGAGGGTGAGGGAGGGAAGCACTGGTTTC
TCAAGCAAAAGCTAAAATTTTTCTATTAAGATTTAACCTGATGCTACACTTTGGTGGTGCAG
CAAGGGTCTCAAATGGTATAAACTCAGGTGATCATGCTTTATGTCTGTCTCTAGAAAAATG
CTCCAAAAATGATAAGTAGTGATAATCCGCAGTCTCGTTGCATAAAATCAGCCCCAGGTGAA
TGACTAAGCTCCATTTCCCTACCCACCCCTTATTACAATAACCTCGACACCAACTCTAGTCC
GTGGGAAGATAAACTAATCGGAGTCGCCCCCTCAAATCTTACAGCTGCTCACTCCCTGCAGG
GCAACGCCCAGGGACCAAGTTAGCCCCCTTAAGCCTAGGCAAAAGAATCCCGC**CCATAATCGA**
GAAGCGACTCGACATGGAG**GGCGATGACGAGATCACGCGAGG**AGGAAAGGAGGGAGGGCTTCT
TCCAGGCCCAGGGCGGTCTTTACAAGACGGGAGGCAGCAGAGAACTCCCATAAAGGTATTGC
GGCACTCCCCCTCCCCCTGCCCAGAAGGGTGGCGCCTTCTCTCCACCTCCTCCACCGCAGCTC
CCTCAGGATTGCAGCTCGCGCCGGTTTTTGGAGAACAAGCGCCTCCACCCACAAACCAGCC
GGACCGACCCCCGCTCCTCCCCACCCCCACGAGTGCCTGTAGCAGGTGCGGCTTGTCTCGC
CCTTCAGGCGGTGGGAACCCGGGGCGGAGCCGCGGCCGCCATCCAGAAGTCTCGGCCGG
CAGCCCGCCCCCGCTCCAGCGCGCGCTTCTGCCACGTTGCGCAGGGGCGCGGGGCCAGAC
ACTGCGGCGCTCGGCCTCGGGGAGGACCGTACCAACGCCCCGCTCCCCGCCACCCCCGCGCC

CCGCGCAGTGGTTTCGCTCATGTGAGACTCGAGCCAGTAGCAAGGGCCCCGGTCCACAGCTT
CGACAGCCAATCAGGTGTGGAAGACAAGCAGGCGGGTAAACCGACTCCCCGAAGGAAG
GGGAGGGTGGGAGGACGCCCCGCCAGAGCCGATTTCACTGACCCTCCCCTCCCGCCGAGG
AGGCCGGCCGCGCCCGCACACCCAGCATCTCTACACCCACCTACCTACCCGCCCCACCCAG
GGGGCAACGCGAGAGTCGCTAAGCGGCTGCGTACTCCCGACGGCGTAAGTGACAGGAGCTTT
ACTCCAACCAGAATACGCCATTTGTGTTTTTACACACGCGGGAGGAGAAACGGCCAATCGG
CGACAAGAGGCTAGCCGGAAGCGCTCCTCCCTCTGCGAGAGCAATGGCTCCGTCCGGTTTTCG
AGCATTTTTCCGCTCCCTTCTCCCTCCCCCTCCGGTTGCCGACGGGCGGGCCTCCCTCCCGCC
TGCATCCAGCCACCCCTTTCCCTCCCAACGTAACAAACATTATGTTCCCGACTTTCCACGGG
AAAGGCAACCCCCGCAAGCCACCAGACGGCCCCCTAGCCACCCATCCCCCAGTGTAACGC
ACCTCCCCCTCCACCCAGAGTTCCGCTCCCTACCTAGCCGAGGCTCTCTGAGGAGCCGGAGC
GCCGAAGCACAGCCTCTTCTCTAGGCGGCCCGCGGGCTTCCGCTGATTGGCGGCGAGTGGG
CCAATGGGTGCGGGGCGGTGGGCGGAGAGGCCAATGGCGCGGCGGGAGGGGGCGTGTCCCGG
GTGCCCTGGCGCCGGCGCTGGGAATCCCCGTGCGGTGAGTGGCGTTTCCGCTCGGGCAGCG
GGCTGAGTGAGCTGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCTGCCGGGGAGGGG
CGGCCGCCGCCGCCCTGCGCTCAGAGACTCACGCAGCCCCAGTCCCGCCAGTCCGCCAACAC
AGTAGTGCCGGCCCCCTCTTTCCCTGGCCCTGCCCCCTCCCCGCCTTTGGCTCGCTCCG
CCTTTCTGCCCCCACCACCTCACGGGTACGGGCCATTCCCGGCCAGGAAACGCCGTGG
CGCCGCGTTGGGCCTAACTCGAGTCTGCCGCTCCCGGGAGTGCCGTGCGCCGAGCCCGG
GCCCAGGCCCCGGCAGCGCTGGGACAAGGTAAGGGTCCGACAGAAAAGAGACCGAACCTCA
CGATCGGGCCCCAGGGGAGGGAAGGGTCACCTCCTCCGTCTCCCGCGCTCGCTCTCCTTGG
GCTTGCACACACCGTTGCAACACCCCATTTTTCCCAGGGAGAGAGATCCCCCTCTAATCTAG
GCGACCCAATCCCCCTTTCATGTTTTTCTGGGTGAGGACGCTTCCCCCTCCCCAACGCCT
CTTACCCCCCTTCTGGGAAGTGCCTACTCCACGTTTACCTTTCCCTTGAGGAGAGGCCTC
TTGCTGCCCTCCGCTCGAAATACACAGGCATACTTTTTTTCTCTCCCGATCCCCACTCCC
TACCCCCGTTCTCGCGGCCTTGTGACAGACAAGTCTGATCGCTCTGGGGGCGCGATCTCCC
CTCCGTAATCTTCTGGACGCCTTCCCTCTCGTTTTTCTGGCTTCCCACCTCAGATGGCTGCT
TCCCAAAGGCATTACCTTCGCCACCCCCACCACACGTTCTCTGGCTCCCGTGGCGTGTGCC
ACAGCGTGTCTGAGATAGCCTCGTTGAATGTGTAGGGTTCGAGCCTGGAGTTGAGCCAGATT
GTGTGTTTTTACTTGCCTTGGGCGTGAGAAACGATCTTGTGAGAATATCTTCAAAGGCAGAA
AAATATTCCCTTTATGAATTCTCTTTCCCTCTGCGTGTAAGTCGGGAATGTGAAGAGGAGTG
TAGGAAAGAGCCCTGGTTCAAGTAGGTAAATCGCATGAGAGGGAAAGTTAACTGTTGGGAA
AGCCCCCTTCTATGCTAATTGATTCTATAGAGTCCTTGCTTGTCTCACTTCTTGGGCGTCAGT
GGTCTTTCTCTTGGATATGGATGCTGCAGTCAGCTCTGCTGGTCTGGGTGAGGGGTGCGTGT
ATGACCTGCATTTTCTGCTTTCTCATGTTACTTGTGCAATGTATTCACCGGTAAGTCAATTC
TTTCCCAGACCTCTGGGTTCCACTGGGCTTTGTCTATATTTAAGTTCATTTCTCCAGTTTCC
TTCTTGCACATAGGTACTGAACGAATCCCCAAGTTCTGTGCTAATTACCTTCATCAGTTGAC
TAAACAAGTTTTTAGATGACATATTTGTGACCAAGGTCATATTTACATTTCTTTGTTGGACA
GATGTTACATAGCTATACTTGTGATTGGGGAGGATCCAGCTGAGTGGAGTGTGCTGAGCTTT
TTAGGAGAGTGTGTAAGTCCCTATTTGAAATTATTTTTTGGTTGTTAATTTTATATTATTAAT
GTTTTTAGGTACACAGAAAGTTCTAAGTGGTAATTTTAGATGTGTGGGATCTGAGCTAGGACT
AAAGCAGAGAATACCCACGTAATCAGAGGTTTCTGGGCTCCATAGAGGACGTAGGGCTTTTT
TTTTTCTATTGGATTTCTTCCAGTTTTCTCAGGATCATTAGTTCTCTTCTGTAGCCAAAAAT
TCTGGCCTGTTATGGGATTAGAGTCTTTAAGGTTTACTCAGACTGTCATTATGTGTAGAAAA
ATGAATTATGCCCTTTGGTAGGACATGACACAAGGCTCTGTTTCTAGCTGCAAAATTTAAATT
AGATTGTAGAGTGCTTGGGAAATTGGCTTTCAAAGACCAAAGCTTAATCTTCACTCCTAAA
CTGCTGGCTTAATTTAAATGGATATTTAGAATTTGGTAAATGTTGATTTTTCTAATAAAAGG
CCTTGGTTTTAAAGGGTGACCTTAGGATTGTTTTCTTTCTTAAAGCATAATTCCAGCCCTTC
TGGCATGGAGCACTGGTCCAAAAAAGTGTGTGTAAGGAGTGGGGGTGGGGT
AAAGAGAAGGTTGTTCCCTTTGGGTTGGATCACAGGGGTGAGTATACAAGGCAGCAGCAGCTG

CTGGCTCTGGAGCTCTGGTTGCTACGTGAGAAGCTTGAGTAGTGCTGGCTGCTGTCTCCAGG
GAAGGACAGCAGTGCAGCGTCCATTAATGCTGCTGGCTGCAGGGAGCAGCACTTAGGCGATG
GCTGCTTCAGGACTAAGAAGAAACCTTGCTTTTCTGGGAATTTTCACTGCTGAGCTGGTTTG
CTTTTTATTGGTGGGGAGATGGGAATTAGTAATTCATAATCTCCTACCCATTTATGGATATT
GGCATCTGGAACCTGGATCATGGTTAAAGCCTTTCTTTTTTTGTTTGTTTGATTTGATTTTT
GTTTTTTGGCAGATTTTTGTTTTTTATCTAGACATTTGTGCTTGGATAGGACTAAAAAGTTCC
ATTAGAGTTTTAATTTTTCAATCAGTTTTAAAAACCCAAGTAATAATTTAAGAATCTTTCTG
ATAACCACAATAGGAAGAAAATAACAGGAATTTTTTCTGCAGCTCACATATCATGCCCTTCC
TCCATCTCTTTAATCATAGAATCAATTCTTATTATTTGTTATGTGTCTCCATCCTTTTCGAT
TAGACCACATTTACCTTATAGACGATTTGCTAAACATTTTACTAAGCTTGAACCTCTTAACT
CTAAAAAGGTGCCATTTTGGAGTGGTTTCTAAATAAATATTTTTAATTTGTATATTAGTAAT
AAACTTCTCCAGATTAGATATTTTTCTTTGGAGTTTGACTTATAAGATTGATTCATTATATAC
ATGTTGGATATAGCCTTCTGACATCACAAATATATGTCTTTGGCCATAATCCATCTGAAATG
TAGGACAGACCAGAAGAAATATGCAGAAATCGAATAAGTCTAGTTCAGGATACTGAGAAGAT
GGCCTCTGAGCCCCCTTAGGTGATCTCCCCCTCCCCCACAACCTCCTGAACATTAGGATGATCTC
TGATTAAGCAAAAACAGTCTGAGCGTGGAAAACTTGAAGGAGAACCACCACCACCAATTATA
TGCAATACTGGACATATTCCTGTGTGCTGTTTTTCTTCCCCAAGACTCGTGTATCCTATACT
TTTTTCTCTCAGAATTTTGATTTGTTCATTTTCGTGTAATGTACTTAAATCTCACAAACAT
CTATAATTTGTAGTATCACTCTGGCATTGTGGCAGAGAACCAAAAAGAATGGAAATGAGTT
TTGTCATTACAAATGTGGCTCACATTGTTTTCCCAGTAATAAAAAGCAGACCAATGAAACAG
AACCTTTAATGGATACTATTTTAGGAGGTTCCAATTCTTATTACTATCACATAGATAAGATG
CAATAGCAGATAAATATGATTTTCATGTATACTGGCTGTTTGACATACTTAGGGTTTTAAGATA
AAAATGTTTGTAGTTTTTTACTCTGTGGCTTAAGTTGCTATATAAAATAATTGCTTTTACAC
TCGAATTTCTGTGTTGTTTGAACCTTTTGTGCTCTTGATATTATCATTTTTTTAGAGGATCAT
ACAGGCCCTTTTCATAGAAGGATTTACTTAAGTTATACCCTTGAAAACCTTTTTTATATCTTT
TGATACTGTTTTGTGTCCAGGAACCTGACTTTCTGAAATTATTCTGGCTTTTCTGGGGAGAAT
GACTATTTTCATTTTTACCTTTGAATGGGGAAATAATAAAGTGCAAAGTACAGATTTGCAGAT
AATTACTTTTGCTTTATCCTCTCCATGTTGAAATAACTTATGAAAAATTAGGCCATAGTTAA
CAGCAGTCAATGACTATTGGATACATTTTATCAGAGGGGAACTGGATCATGAATAAAATAAA
ATTTTAAAAATAATTTTTGGCTGAACTCTGGTGATTTCATCAGTTTAATTTGAAGTCAGAAGG
TCTAGCAGTGAATTTTTATTTATAAAAAATTGTATTTCAAGTGTTGAAAACCTGAAACTTCTTGA
CCAGTATATTTGTTTGAGGCATCAAACTTTGCAAAATGTGCATCGTATATTTAGTGATATA
ACTGGTAGTCATTTGTAATTTAAAGTATTCTTTCAAAGGCACTCTTTAGAAAGTAATGTAGT
GTACCCGTGATGGGCAGGGATTGGTACCATTCCCTTACTGCCAAAAATTCAAAAATATGTGGC
AAAATGATTGATTTATCTTGTGGGTGGGATTCTGGGAAGTTCATGAAAGGTGGAGAGAATAT
AGTTTCTCTTCACTTGTCTATATACATTTTGTTAAATAAGTCTTAGGAAAACCTGTTTTATTGT
ATCTTTAATTATGAATTGCGTAAAAGATACCCAGTAACCTTTGGGGGGAGGTGCTGTTAGAAA
GCATTACATTGGAGAGAATTCCCCTACCCTGGGACAAAATGCATTCTGTCTTTAATACTTAG
CGAAGGGAACCTATGGGATAAAATAAACAAATGAAGGTAAGCTCAGTCTGCTTTATATGTGCCC
TCACTGAGCAAGGAATTTGTAATCGCATCGTGCCTCATTGTTTATACCATCATATTGATTT
TGTTTGCTGAGTACCTGAGGGAATACCTTACTTAATGTAAGGTCACATTAAGTATGTTTGAT
ATGAAGACAGGGAAAGGAATTTCTGCTTCTTGAGTAATGTCTTAGTATTTTTAAACACT
TAAGTTTTTACATCAGGCCAGTTTTGCCTGATGCTCATGTCTGTTGCTTTGGTTGGGCTGCT
GCTTTCTCTTCTGTGTTCTTATGGGTTTCGTTGTGGTATAAGGATTCCCACAGCTTTTCATGGC
AGTATGAAGTAATGAGAAGCATTGCCTTAGCCATGTTAGTTACATGTATACTTTTGGCCTAT
GTTATGAATCACAAAAGCGGTAGCTATAGGAATGTATACAAAATAGATTTCTGTCTGGGGA
ATCAAGTTTTTTGATTTGTGCTACCTAATGGAGGGGAAAATGCTGAATTTCTTGCTGCTCTGT
TTGAGAAATAGATGGAAGCATGGGAGGAGCCAGAGACCTCTGCAGCAGGATTTGGTCTAAGT
AGAAAAGGAAGATTTTTGTTTCAAATTGCCAGCTGCTTATGTCAGACTGACTCCCTTATTAT
GCCTCCAGTAGGCCTGTCAATATGGCCAAACAGCTAGATAAGTGCGGGGCAGGACAAAGGGC
TCTTTCACAGCAGGGAGGCAATGTTGGTGGGGGAGGGGCAGGAGGTAGGAAAGGCAAGAGG

AGGAGGTTCTTTTCCCTGGGAGATTATTCAGTTTGGCATAACAATTAAAGAAATCATTTTTGTAG
TTCCCACTCAAGCATTGAATTTTTGCCAACCACATACTATTAACCCCAAATTTGATACATTT
CAGAATATCTTGTAGGGATCCATTCTCGCCAAGGAAAAATAAAAAATAAATAAAGCTCTGT
ATAGGTTAAAAATAAAATAAATCCCACACTCTGCACCCTCCTAGGTGCAAGTCACCTCCCGAG
GAGACCCGTTCTAGAGCTGAATTCTCATTAAAGAAATGGAAAAAGAACTCTATCTGAATAAA
AACACATTGTAATACAATGTGTTTATTTGGGTTGGGATTGGACCTGAACATGTAGAATAATT
TGTTTTCCCTTTATGAAATAGTTGCTCGTAGTTGTCTACAATTTTATTTTATTAAAGATAGGTA
GCACATTACAGCTTTTCATGTGTTGGGTTGCCATATGTAAAAATGCTAACTGAAGAAAAGGCTAC
TTTTTAATTTTACGCCTCATCCTTAGTTTCTGGGAGAACCTGATATTTCTGGAGATTACTCCC
TCCCCACCTTTTATGTTTAGGCAACCTCTTTTGTATACATTTGTGTTTCTAGCTCGCATACAAGT
GGGATAGTTGCATCCAGTTTATTAAGACTTAGTATGAATCATAGAGTTGGAAAAGATCTGTT
GGTTATCTGGTCCTTTAAACCAAAATCATAATGAAATATTTTGAATTTGGGTCCCTATTGA
AGTTTTTATTAAATGTTAAAGGATCGGTGTTCTGAACAACATTTTATGTTACTTTTAAAT
AAATGTTTTGCGTCAGTTCTTTTTTAAAAATAAAGAATTTTATTATAGGCAAATTAGCTG
GCAATTATTTGAATTGTGATAGGATTTCTCTTTTATGAAGGAATATATGACAAGGTTTTTCA
AAATGCTTAATATATTTTAAAGACTTTAATTTTTAGAAAATAATTGGTTTGAACAGTTTTCC
AAGAGCACATTTGTTGCTTGGGTTGAGGTACCACCTATATTGCAATGTTACTAACTAGCCT
TAAAGTTTTCCCTTCTGTCTATACTGCATGCAACAATAAAGGGAAGTGAATGTTAATTTCC
ATTTATGGATTAGCAGAGGAGATGTTTTAACCGATTAATAACCAAAAAACTGCCTTTCTGTAC
ACGTAATATTAAGCAAGCCTGACCAAGTTTTGTGTTATTTCTCTGTTAAAGAAAACTGGA
TGTGTTACTACTTAACATTATATTGTTATTTAATGGTCTTGGCAGTAATGATATAATATTTT
GACCAAAAGAAATTTTGAGTAATTAATTATTATTGTAATTAGTTGGAAGTTTCTCATCAGTA
AAATAGCAACAGCATTAAACACAAAATCTAGTGAGCTATATTTTATATTACTACAGAAATTTA
GGGTAGTCATTTCTTTCTTTATAATTTATTCACATGGATTATTTCCATAAATTTGTGGGACT
AAAATAGAAGCCATCTAGTCAAGCACCAGTCTCCATACCAGACAGTTTTCTCTGCATGTGCT
ATGACCCACATTGCCAGTATTAAACATCCTTTACACCCTCCCCCTTCCAGATAATTAGAAA
TCTCTTCAGGGTAGCTTCCATTGCTCCTATTACCTGGATCTTGCTAGAGGCTCTAAGAAGTT
CCTGGTAAAAGTGAGACAGTAAGGGACCACATTTTGATTCCAAAGGTTTTGATAACTGTTAG
GGCTCCCCAAACAGCTAATCTCATTTTACCAAGACTTAGCCAGCAGAGGGCTGGAATGGAG
GTGAAACACAAGCACTGTACCTCATCTTGCTGTGCAGCTGCTCCACCTTATTTCTGTCTAT
TATTATCTCACAACGCCTCCTCCCATCAAAAAGAACTAGGACAAAGGGGGAAAATTGGATG
GGCTAATGTGATTTTTATTATGCTAGGTTGTGGGCTTGTATTATGTACTTAAATACAAAGC
TAATTTGCCCATTTCTTAAAGTCTTTAGTGATAGAGATTTTGTAAGTTCTGTATCTTCTAC
TTTCTTTCTTGATAAACCATTTTCAGATTCTCAGCCTTACAGAAAGAAAGGTTTTAAGCATA
TTAATTTTCTGTGGCCGTTTACAGTCATTATTACCACCAGATGCCACTGTATTATTAGCTTG
AAGAAAGGTGGGCTCTCTCTGTACATAATATCTGCAATTTGTTTTGGAAAATACTAATTTG
TATAAATCTGATTTATGACTAAAATAAGGTTAAAAATTAGACCTCTATGTATGTTTACCCTA
TTACCTTAGTGGGGGTGAAATTAATTAGCTCTTTGAACATAAAATTTTTCATGTCTTAGAGTT
CTTTTTTCAAGCTGCATAATTTATGTTCTTCAAGCCATTTTATCCCATACCACCCCAACAA
AGGGGGAAATTTTATTTTTTATCATTTTTTATTGTCTTTCAATGGTGAGATTTTCGCCACCC
ACTCCTGAAATGTGAAGACTCAAATAAACTGAGTAATCTAATAAGGTATATGCGTTGCTGA
ATGTAGTAAGATGATTGTTTCATCATTTCTTAGATATTATGATCTAGTTTGAATCTGGTTTCC
AGTATCATGTTAGCATATTTAATACTGTTGATATGTTAATTTAATACATGCCAGGTGGAT
CTCCTTGCTTTCTATTTGTGCCCCCTGTTTGTGCTTTTGTATGAAGGGGGTTTTTGTGTTG
GATTTTCTTCCCATCTCTGTGTCTGTTATGTTCTTTGGCTTATGTTTCAAAAATTCTGTT
TCCTACCACCAACCTCTGTACATGCCACAACACATAACAATTTGTACTTTTACAGTTTCTGTG
AAGTAGGATGATCTGCAGTTAATAATCAACTGTTTGGGCATTCTTGGTATCCAAGGAAGGTT
TTACTTAGAAGGAAGAACCTGGAAGGACCTGTTGGCAATTAGACTACTTCTGCGTTTATTTT
ACATTTTCCCTTATTAACGTAGGCTGTTGAGAGTTGACTTGTTTTATAAGAGAAACCAGATT
GACAGAGAAGACCCCAATCAGATAGAGTTATTTTAAAAATAAATGTGTTTATTATGGTAAC

ATTTGGGGTAGAATCTAAAGGGCATATTTTTAAAAAACTTTTAGTTCTAAAGACAAAAGAG
TTTAACCTAAACAGAACAAAGAGAAGGGCCTTTGAAGCAGTATGATTGATTATAT

EXAMPLE 11 – 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:5)

tctagaaaca	aaacccaaaa	tattaagtca	ggcttggcct	caggtgctgg	ggtggagtgc	60
tgacaaaaat	acacaaattc	ctggctttct	aaggcttttt	cggggattca	ggtattgggt	120
gatggtagaa	taaaaaatctg	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	cagggggtca	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	ggtttttaaa	aggctttgtg	ttacctaatc	acttctatta	780
gtcagatact	ttgtaacaca	aatgagtact	tggcctgtat	tttagaaaact	tctgggatcc	840
tgaaaaaaca	caatgacatt	ctggctgcaa	cacctggaga	ctcccagcca	ggccctggac	900
ccgggtccat	tcattgcaaat	actcagggac	agattcttca	ctaggtactg	atgagctgtc	960
ttggatgcaa	atgtggcctc	ttcatttttac	tacaagtcac	catgagtcag	gaggtgctgt	1020
ttgcacagtg	tgactaagtg	atggagtgtt	gactgcagcc	attcccggcc	ccagcttgtg	1080
agagagatcc	ttttaaatg	aaagtaagct	caaagttacc	acgaagccac	acatgtataa	1140
actgtgtgaa	taatctgtgc	acatacacaa	accatgtgaa	taatctgtgt	acatgtataa	1200
actgtgtgaa	taatctgtgt	gcagcctttc	cttacctact	accttccagt	gatcaggttt	1260
ggactgcctg	tgtgctactg	gacctgaat	gtcccaccg	ctgtcccctg	tcttttacga	1320
ttctgacatt	tttaataaat	tcagcggctt	cccctctgct	ctgtgcctag	ctataccttg	1380
gtactctgca	ttttggtttc	tgtgacattt	ctctgtgact	ctgtctacatt	ctcagatgac	1440
atgtgacaca	gaaggtgttc	cctctggaga	catgtgatgt	ccctgtcatt	agtggaaatca	1500
gatgccccca	aactgttgtc	cagtgttttg	gaaagtgaca	cgtgaaggag	gatcaggaaa	1560
agaggggtgg	aaatcaagat	gtgtctgagt	atctcatgtc	cctgagtggg	ccaggctgct	1620
gacttcactc	ccccaaagtga	gggaggccat	ggtgagtaca	cacacctcac	acatactata	1680
tccaacacac	acacacacac	acacacacac	acgcacgcac	gcacgcacgc	acgcacacac	1740
gcacacacac	gaactacatt	tcacaaacca	catacgcata	ttacacccca	aacgtatcac	1800
ctatacatatac	cacacatac	cacccctcca	cacatcacac	acataccaca	cccacacaca	1860
gcacacacat	acataggcac	acattcacac	accacacata	tacattttgtg	tatgcataca	1920
tgcatacac	cacaggcac	cagacaccac	atacatgcat	tgtgtacgca	cacatgata	1980
cacacacata	ggcacacatt	gagcacacac	atacatttgt	gtacgcacac	tacatagaca	2040
tatatgcatt	tgtatatgca	cacatgcatt	cacacataca	taggcacaca	tagagcacac	2100
acatacatatt	gtgtatgcac	acatgcacac	accaatcaca	tgggaagact	caggtttcttc	2160
actaagggttc	acatgaactt	agcagttcct	ggttatctcg	tgaactttgg	aagattgctg	2220
tggagaagag	gaagcgttgg	cttgagccct	ggcagcaatt	aaccccgccc	agaagaagta	2280
ggttttaaaaa	tgagaggggtc	tcaatgtgga	acccgcaggg	cgcagtttca	gagaagagac	2340
ctacccaagc	caactgagag	caaaggcaga	gggatgaacc	tgggatgtag	tttgaacctc	2400
tgtaccagct	gggcttcatg	ctattttgtt	atatctttat	taaatattct	tttagtttta	2460
tgtgcgtgaa	taccttgctt	gcataaatgt	atgggcactg	tatgtgttct	tgggtgccgtg	2520
ggaggccagg	agagggcatg	gacccctcgg	agctggcgtt	tgagacagtt	gtgaccacaca	2580
gtgtgggggtc	tgggaactgg	gtcttagtgt	tccgcaagtg	cagctggggc	tcttaacctc	2640

tgagccatcc	ctccagcttc	aagaaaactta	ttttcttagg	acatggggga	agggatccag	2700
ggcttttaggc	ttgtttgttc	agcaaatact	cttttcgtgt	attttgaatt	ttatttttatt	2760
ttactttttt	gggatagaat	cacattctgc	agctcaggct	gggcctgaac	tcatacaaat	2820
cctcctgtct	cagctctacca	ggtgataaga	ttactgatgt	gagcctggct	ttgacaagca	2880
cttttagagtc	cccagccctt	ctggacactt	gttccaagta	taatataat	atatataat	2940
atatataat	atatataat	atatattgtg	tgtgtgtgtt	tgtgtgtgta	tgagacactt	3000
gctctaagg	tatcatatat	atccttgatt	tgcttttaat	ttatttttta	attaaaaatg	3060
attagctaca	tgtcacctgt	atgcgtctgt	atcatctata	tatccttctt	tccttctctc	3120
tctttctctc	ttcttctctt	cacccccaa	catctatttt	caaatacctt	tgccgaggag	3180
atgccaagag	tctcgttggg	ggagatggtg	agggggcgat	acaggggaag	agcaggagga	3240
aaggggggaca	gactggtgtg	ggtctttgga	gagctcagga	gaatagcagc	gatcttccct	3300
gtccctgggtg	tcacctctta	cagccaacac	cattttgtgg	cctggcagaa	gagttgtcaa	3360
gctgggtcgca	ggtctgccac	acaaccccaa	tctggcccca	agaaaaggca	cctgtgtgtg	3420
actctgggggt	taaaggcgct	gcctggtcgt	ctccagctgg	acttgaaact	cccgtttaat	3480
aaagagttct	gcaaaataat	accgcagag	tcacagtgcc	aggttcccgt	gctttcctga	3540
agcgccaggc	acgggttccc	taggaaatgg	ggccttgctt	gccaaagctcc	cacggccttg	3600
cctgcaaacg	gcctgaatga	tctggcactc	tgcgttgcca	ctgggatgaa	atggaaaaaa	3660
gaaaaagaag	aagtgtctct	ggaagcgggc	gcgctcacac	aaacccgcaa	cgattgtgt	3720
aacactctcc	attgagaatc	tggagtgcgg	ttgccctcta	ctggggagct	gaagacagct	3780
agtggggggcg	gggggaggac	cgtgctagca	tccttccacg	gtgctcgctg	gctgtggtgc	3840
atgccgggaa	ccgaaacgcg	gaactaaagt	caagtcttgc	tttgggtgaa	ctgacaatca	3900
acgaaatcac	ttcgattgtt	ttcctctttt	tactggaatt	cttggaattg	atagatgggg	3960
gaggatcaga	gggggagggg	aggggcgggg	agacggaggg	aggaggggag	gaggggagga	4020
ggggaggagg	ggaggagggg	aagggatgga	ggaaaatact	aacttttcta	attcaacatg	4080
acaaagattc	ggagaaagt	caccgctagt	gaccgggagg	aggaatgcc	tattgggcat	4140
tatattccct	gtcgtcta	ggaatcaaac	tcttggttcc	agcaccagg	attctgagcc	4200
tatcctattc	aagacagtaa	ctacagccca	cacggaagag	gctatacaac	tgaagaaata	4260
aaattttcac	tttatttcat	ttctgtgact	gcatgttcac	atgtagagag	ccacctgtgt	4320
ctaggggctg	atgtgctggg	cagtagagtt	ctgagccctg	taactggaac	aaccagaaac	4380
tcccaccaca	gttagagctt	gctgagagag	ggaggccctt	ggtgagattt	ctttgtgtat	4440
ttattttagag	acagggtctc	atactgtagt	ccaagctagc	ctccagctca	cagaaattct	4500
cctgttcccg	tttccaaagt	actggagtta	tgagtgtgtg	ttaattgaac	gctaagaatt	4560
tgctgattga	agaaaacctc	aagtgggttt	ggctaattcc	cacgacccca	gaggctgagg	4620
caggaggaat	gagagaattc	aaggtttgcc	agagccacag	ggtgagctca	atgtggagac	4680
tgtgaggggt	agctcaatgt	ggagactgtg	agggtagact	caatgtggag	actgtgaggg	4740
tgagctcaat	gtgagagctg	tgagggtgag	ctcaatgtgg	agactgtgag	ggtgagctca	4800
atgtggagac	ctgtatcaag	ataataatag	tagtagtaac	aatgcaggcg	aggggtgtggt	4860
tgagtggtag	agcagttagt	tgatttgaca	tgcttgaggt	ctcccggtcc	atctgtggcc	4920
ctgcaacagg	aagggaggga	ggaagggggg	gaacgagaga	gaggaaagag	agacagaagc	4980
taagataggg	aatgagagag	gaaggaagaa	acgggaagaa	attcagactc	cttccctgagt	5040
tcgccaacag	cctagtaca	tcctgtgcac	accctaaggt	ggcctttgtg	tgccactggc	5100
ttgggtggtc	gggaagggca	ttttcagctt	ggtgcagaac	tgccacagta	gcatgctggg	5160
tccgtgaaag	tttctgcccg	ttaacaagaa	gtctctacta	cttgtgacct	caccagtga	5220
aatttcttta	attgtctcct	ggtgttctgg	gttttgcat	tttgtttcta	aggatacatt	5280
cctgggtgat	gtcatgaagt	ccccaagac	acagtggggc	tgtgttggt	tggaagagat	5340
gattttatctg	gggtgtcaaa	aggaaaagaa	gggaaacagg	cacttgggaa	aatgtcctcc	5400
cgccacccg	aattttggct	tgccaaccgt	ggtggaggag	caagaaacac	gtggacgttt	5460
gaggaggcat	ggggtcctag	gaggacagga	agcagaagga	gagagctggg	ctgacagcct	5520
gcaggcattg	cacagtttca	gaaggagatt	acagcatgac	tgagttttta	gggatccaac	5580
agggacctgg	gtagagattc	tgtgggtctc	gaggcaactt	gacctcagcc	agatgggtatt	5640
tgaataacct	gctcttagag	ggaaaacaga	catagcaaac	agagccacgt	ttagtgtatga	5700
aactctcact	ttgcctgagt	catgtgcggc	catgcccagg	ggtcaggctg	acactcaact	5760
caaaaaacaag	tgagaaattg	aagacaatcc	gtggtggcag	ctactggaag	ggccaccaca	5820
tccccagaaa	gagtggagct	gctaaaaagc	catttgtgat	aggcacagtt	atcttgaatg	5880
catggagcag	agattacgga	aaaatcgaga	atgttaatga	ggcaacattc	gagttgagtc	5940
attcagtggtg	ggaaacccag	acgcttccat	cccctaaaag	gaacatcttg	ctctcagtc	6000
aaatggaaat	aaaaattggg	gcttgaattt	ggcaaatgat	tcagaactct	gtgtaggtat	6060
tttcacacgc	acagtgata	attttcatgt	tggagtttat	ttgtgtctaa	aggcagaaaa	6120
gggtaaaaag	cacatcttaa	gagttatgag	gttctacgaa	taaaaataat	gttacttaca	6180
gctattcctt	aattagtacc	cccttccacc	tgtggtaatt	tcctgagata	gtcagtgagg	6240
aaaagatctc	tccttctctt	ctttctcccc	ctccctcct	ctccctcct	ccctccctcc	6300

ctccctctctc	tccctccctc	cccctttctc	tctttctttg	ctccttctcc	tctgcctcct	6360
tctccctttc	ttcttcattt	attctaaagta	gcttttaaca	gcacaccaat	tacctgtgta	6420
taacgggaaa	acacaggctc	aagcagctta	gagaagattg	atctgtgttc	act	6473

<211> 7045

<212> DNA

<213> *Cricetulus griseus*

<400> 2

(SEQ ID NO:6)

actagcgtgc	aattcagagg	tgggtgaaga	taaaaggcaa	acatttgagg	ccatttccctt	60
atttggcacg	gcacttagga	agtggaaacat	gcctaactcta	ctgggtttgta	ccacctttcc	120
ctataatgga	ctgtttggga	agctcctggg	caaccgattc	tggcatctca	ttggtcagag	180
gcctgttaaa	tggtagctctt	atttgcaaaag	aaggetgttaa	ctttagactt	taaaagcctc	240
tcctcaagaa	agaagggaga	aaggatatgg	ctagacatat	ctaatagact	taaccactgt	300
gaaaagcctt	agtatgaatc	agatagaacc	tatttttaac	tcagttttga	aaaaataaat	360
ctttatatatt	atttgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	420
gaaccacatg	tagcagggtc	tggaggaggc	cagaagaggg	caccagatct	cctggaactg	480
acaccacaca	tggttatgag	ctgcctgatg	tgggtgctgg	gaactgaact	ctcgtgttct	540
gcaagagcag	caactgttct	cttaactgat	gagccatctc	tcagccccc	cccataattt	600
taattgttca	ttttagtaaa	ttttattcat	aatcaattat	cacagtataa	aacaatgatt	660
ttatatatat	catatacata	tcaaggatga	cagtggaggg	gatattgtgtg	tgtgtgtgtg	720
tgtgtgtgtg	tgtgtgtgtg	tgtgttattt	gtgtgtgtgc	tttttaagaa	gggtgccatag	780
tcactgcat	tctctgaagg	atttcaaagg	aatgagacat	gtctgtctgc	caggaaccct	840
atcttcctct	ttgggaatct	gacccaaatg	aggtattctg	aggaactgaa	tgaagagctc	900
aagtagcagt	gtcttaaaacc	caaatgtgct	gtctagagaa	agtcaacgct	atcagtgagc	960
tgaggagaga	tttactgagc	ggaagacaag	cgctctttga	tttaagtggc	tcgaacagtc	1020
acggctgtgg	agtggagcct	gtgctcagg	ctgaggcagt	ctttgctagc	cagctgtgat	1080
gagcagtga	gaaaggggtg	agatggaggc	aggggtgggag	cagggtatg	gttcagacta	1140
ggtatogtga	gcacaccagc	tggttgactt	gtggtctgtg	ggtcaggcgt	tgtaaacgcc	1200
ctcagggtca	ggcagtcaca	ttgcttgaag	ctgaatgggt	gaggcaacac	agagagtga	1260
aagaaggcaa	agtaccacct	cttcccgcac	ccaggctact	tctgggttat	agctgagact	1320
ccggacagca	tgcaaccagc	tggttagagc	ttcaggga	actgatgtc	tgcatgttgc	1380
tatgaaatgt	gattcggtag	atctggagaa	aatttataat	gctggctcag	tcaagcactg	1440
aacaaaggta	ccttggcttt	gggagctaca	tgacattgac	ttgtaggcag	actttttttt	1500
ttctgcccgc	caattcccag	ataaccaata	tggaggctca	atattaatta	taaatgctcg	1560
gctgatagct	caggcttgtt	actagctaac	tcttccaaact	taaatgaacc	catttctatt	1620
atctacattc	tgccacgtga	ctttaccttg	tacttctctg	ttcctctcct	tgtctgactc	1680
tgcccttctg	cttcccagag	tccttagtct	ggttctcctg	cctaaccctta	tcctgccag	1740
ctgctgacca	agcatttata	attaatatta	agtctcccag	tgagactctc	atccaggagg	1800
gaattgggtg	ctccccctc	ctcattgcca	tccgtgtctt	cctcttccct	cgcttcccc	1860
tcctcttccct	gctcttccct	ctccacccct	cctttcatag	tattgatggc	aagggtgttc	1920
tagaatggag	gagtgcccat	aggcatgcaa	agaaaccagt	taggatgctc	tgtgaggggt	1980
tgtaatcata	agcgatggac	acaattcaag	ccacagagtg	aagacggaag	gatgcactgt	2040
gctctagagc	aacttctggg	gcagaatcac	agggtaggtt	tctgacttga	gggcgaagag	2100
gccacgagga	agggagttag	tttgtctgag	ctagaagcta	cggcccacct	cttggtagca	2160
gacctgcccc	caagcatgct	ttgttaatca	tgtgggatct	gattttctct	taaatctatg	2220
ttcaactctt	aagaaaatgt	gaattctcac	attaaaaatt	agatatacgt	cttttggttg	2280
gggggggtgta	aaaaatcctc	aagaatatgg	atttctgggg	gccggagaga	tggctcagag	2340
gttaagagaa	ctgggttgctc	ttctagacat	tctgagttca	attcccagca	accacatggg	2400
ggctcacaac	catctgtaat	gcgacctggg	gccatcttct	gacatgcatg	gatacatgca	2460
ggcagaaagc	tgtatacata	gtaaatgtat	aaatcttttt	ttaaaaagag	tatggattct	2520
gccgggtggt	gggtggcgac	gcctttaatc	ccagcactct	ggaggcagag	gcagggtgat	2580
ctctgtgagt	tcgagaccag	cctgggtctat	aagagctagt	tcaggagacag	cctccaaagc	2640
cacagagaaa	ccctgtctcg	aaaaaccaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	2700
gtatggattc	taagaaagcc	gtaacagctg	gagctgtgta	cggagttcag	cgtggtagta	2760
gaagaacaga	cattcatgat	gaaacacccc	aggattttta	cttagtatct	agtttccatt	2820
gttgttttga	gaccggctct	tatgctctcc	aggctggcct	caaactgctg	atcttccccg	2880

ctctacctct	caagtcctgg	gactacttgg	ctcataaaac	agtttttgtc	gggctccctg	2940
aagttaatggt	tgtacaaacc	gtgggggtca	atatactcac	ttgggcagag	agagaaggctc	3000
tgaatcccag	acaatgactg	catctcagga	cagttgggaa	gaggacaatg	gcagaaggac	3060
ttagaaaaga	tagactggag	ggtggaaaag	cagcaggaaac	agagaaacaa	aacaggaagc	3120
ttgctatcca	gggccactct	ggagtcctgt	ggcaagatgg	aagcgggcta	ggggaatata	3180
tttgtgtac	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgat	caatgcctat	3240
caatggtgaa	ggggaaatat	gtataccaca	ttgattctgg	gagcaattct	cagtatctgg	3300
cctagagaaa	ggaatggccc	ctgcagaata	gacagagtga	atggtgccct	ttatcatttg	3360
ctaaagtga	ggagaaataa	acatccttcc	atagagtttc	aggtaaatga	accccacagt	3420
tcatctgtgc	cgtgggtggag	gcctggccaa	cagttaaaaa	gattagacac	ggacaaagtc	3480
tgaaggaaac	acctcgaata	ggaagaggag	agccacctca	ttctgtaact	ttcctcaagg	3540
ggaagatggt	ccaagagtgg	gaataaatgg	tcaaaggggg	gatttttaat	taggaaaacg	3600
atttcctgta	tcacttgtga	aactggagggt	tgatttgggg	cataggacaa	tagatttgat	3660
gctttgcaaa	aagctgtttc	aaagcagaga	aatggaatag	agacaattat	gtagcgagga	3720
gggaggggtg	ggcgaagatg	gagacagaga	agtggaaagt	gactttaggg	aagaggaaca	3780
tagaccacag	gggcggggcg	gggggcaggg	gcggggggcg	gggctcaaag	gaggcagtg	3840
gaacgttgct	agtgttcgca	gcgtaagcgt	gaatgtgcaa	gcgtctttgt	gggtgtgtgac	3900
caggagtagc	gtggctggct	tgtgtgctgc	ttgtaatccc	agtccttgag	gtttccacac	3960
tgttccacag	tgggtgtgat	ttccctcgg	agagcatgag	ggctctgctt	ttcccatc	4020
ctcccacg	ttcgttggtg	tttgtttcca	agatgttagt	gggtgagaca	aagcctctct	4080
gttgatttgc	ctttaacagg	tgacaaaaaa	agctcaacca	ggagacattt	ttgccttctt	4140
ggaaggtaat	gctcccatgt	agagcaatgg	gacccatctc	taagggtgag	ctactcttgc	4200
agtttgacc	cagctcttct	gatgcaggaa	ggaagttggt	gggcaagcaa	gactgtttgc	4260
ttcttgcat	ggacacattc	tgacacaaaa	ggctcaggag	gggagaaggc	ttgttgatgt	4320
ttagcactca	ggaaggcccc	tgatgcattc	gtgattagct	gtctccatct	gtggagcaga	4380
cacggactaa	ctaaaaacca	gtgtttttta	attgtcaagc	ctttaagggtg	aggaaattga	4440
cttattgtgc	tgggccatac	gtagagcaag	tgctctgcat	tgggccaaac	cccggtctctg	4500
gtttctaggg	accagaatgg	cctagaacta	actcacaatc	ctcccatctc	aggtctcagg	4560
tgctagaatg	aaccactata	ccagcctgcc	tgcctgccta	cctgccttcc	taaattttaa	4620
atcatgggga	gtaggggaga	atacacttat	cttagttagg	gtttctattg	cttgaagag	4680
acaccatgag	catggcaact	cttataaagg	aaaacattta	gttgggtggc	agtttcagag	4740
gttttagtac	attgtcatca	tggctgggaa	catgatggca	tgacagacaga	catggtgctg	4800
gagaaaggga	tgagagtcct	acatcttgca	ggcaacagga	cctcagctga	gacactggct	4860
ggtaccctga	gcataggaaa	cctcacagcc	caccctcaca	gtgacatatt	tccttcaaca	4920
aagccatacc	tcctaatagt	gccactccct	atgagatgac	agggccaaat	acattcaaac	4980
tgctataaca	ctttaaaagta	ttttattttt	attattgtaa	attatgtatg	tagctgggtg	5040
gtggcagccg	aggtgcacgc	ctttaatccc	agcacttggg	aggcagaggc	agatggatct	5100
ctgtgagttc	aagaccagcc	tggctctataa	gagctagtgt	caagggaagg	tatacaaaga	5160
acagttctag	gatagccttc	aaagccacag	agaagtgtctg	tcttgaaaac	caaaaattgt	5220
gctgggacct	gtctctgctt	tggttgcttc	ccactccccc	agagctggac	tcttgggtcaa	5280
cactgaatca	gctgcaaaat	aaactccttg	attcctctct	tgtaacagga	gcccgagtc	5340
aggcgccac	ttgtcttctc	gcaggatttg	catagacttt	ttctgtgtgc	ccaccattcc	5400
agactgaagt	agagatggca	gtggcagaga	ctgggaaggc	tgcaacgaaa	acaggaagtt	5460
attgcaccct	gggaatagtc	tggaaatgaa	gcttcaaaaac	ttgcttcatg	ttcagttgta	5520
cacagactca	ctcccagggt	gactcacacg	tgtaaatatt	cctgactatg	tctgcactgc	5580
ttttatctga	tgttctcttc	ccaaaatgcc	aagtgtacaa	ggtgagggaa	tcacccttgg	5640
attcagagcc	cagggtcgct	ctccttaacc	tggacttgtc	tttctccggc	agcctctgac	5700
acccctcccc	ccattttctc	tatcagaagg	tctgagcaga	gttggggcac	gctcatgtcc	5760
tgatacactc	cttgtcttcc	tgaagatcta	acttctgacc	cagaaagatg	gctaagggtg	5820
tgaagtgttt	gacatgaaga	cttggtctta	agaactggag	caggggaaaa	aagtcggatg	5880
tggcagcatg	taccgaaat	cccagaactg	gggaggtaga	gacggatgag	tgcccggggc	5940
tagctggctg	ctcagccagc	ctagctgaat	tgccaaattc	caactcctat	tgaaaaacct	6000
ttaccaaaaa	aacaaacaaa	caaataataa	caacaacaaac	aacaacaaac	tacccatac	6060
aagggtggg	gctcttggct	cttgaggaat	gactcaccca	aacccaaagc	ttgccacagc	6120
tgttctctgg	cctaaatggg	gtgggggtgg	ggcagagaca	gagacagaga	gagacatgac	6180
ttcctgggct	gggctgtgtg	ctctaggcca	ccaggaaactt	tcctgtcttg	ctctctgtct	6240
ggcacagcca	gagcaccagc	acccagcagg	tgacacaccc	tcctccctg	cttcttgagc	6300
aaacacaggt	gccttggctt	gtctattgaa	cgggagtaag	ttcttgacga	tgtatgcatg	6360
gaaacaacat	tgtcctggtt	ttattctctac	tttgttgata	aaaaccgggg	aactccagga	6420
agcagctgag	gcagaggcaa	atgcaaggaa	tgctgcctcc	tagcttgctc	cccatggctt	6480
gccgggcctg	ctttctgcaa	gcccttctct	ccccattggc	atgcctgaca	tgaacagcgt	6540

ttgaaatgct	ctcaaagtgc	actttcaaag	aaggcttctc	tgatcttgct	aactaaatca	6600
gaccatgttt	caccgtgcat	tatctttctg	ctgtctgtct	gtctgtctgt	ctgtctatct	6660
gtctatcatc	tatcaatcat	ctatctatct	atcttctatt	tatctaccta	tatttcaatc	6720
atctatcttc	taactagtta	tcattttatt	atttgtttac	ttactttttt	tatttgagac	6780
agtattttctc	tgagtgcacg	ccttggtgtg	cctggaaccc	attctgtaac	caggctgtcc	6840
tcaaactcac	agagatccaa	ctgcctctgc	ctctctggtg	ctggggttaa	agacgtgcac	6900
caccaacgcc	ccgctctatc	atctatttat	gtacttatta	ttcagtcatt	atctatcctc	6960
taactatcca	tcatctgtct	atccatcatc	tatctatcta	tctatctatc	tatctatcta	7020
tctatcatcc	atctataatc	aattg				7045

<211> 6473

<212> DNA

<213> *Cricetulus griseus*

<400> 3

(SEQ ID NO: 7)

agtgaacaca	gatcaatctt	ctctaagctg	cttgagcctg	tgttttcccg	ttatacacag	60
gtaattgggtg	tgctgtttaa	agctacttag	aataaatgaa	gaagaaaggg	agaaggaggc	120
agaggagaag	gagcaaagaa	agaaggaaa	ggggaggagg	ggagaggagg	gaggggaggga	180
gggaggggagg	gagaggaggg	gagggggaga	aagaagagaa	ggagagatct	tttccccact	240
gactatctca	ggaaattacc	acaggtggaa	gggggtacta	attaaggaat	agctgtaagt	300
aacattattt	ttattcgtag	aacctcataa	ctcttaagat	gtgcttttta	cccttttctg	360
ccttttagca	caaataaaact	ccaacatgaa	aattatccac	tgtgcgtgtg	aaaataccta	420
cacagagttc	tgaatcattt	gccaaaattca	agccccaatt	tttattttcca	ttttgactga	480
gagcaagatg	ttccttttag	gggatggaag	cgctcgggtt	tcccacactg	aatgactcaa	540
ctcgaatgtt	gcttcattaa	cattctcgat	ttttccgtaa	tctctgctcc	atgcattcaa	600
gataactgtg	cctatcacaa	atggcttttt	agcagctcca	ctcttttctg	ggatgtgggtg	660
gcccttccag	tagctgccac	cacggattgt	cttcaatttc	tcaattgttt	ttgagttgag	720
tgtcagcctg	acccctgggc	atggccgcac	atgactcagg	caaagtgaga	gtttcatcac	780
taaacgtggc	tctgtttgct	atgtctgttt	tcctcttaag	agcaggttat	tcaaatacca	840
tctggctgag	gtcaagttgc	ctcagagccc	acagaatctc	taccagggtc	cctgttggtg	900
ccctaaaaac	tcagtcatgc	tgtaatctcc	ttctgaaact	gtgcaatgcc	tgcaggctgt	960
cagcccagct	ctctccttct	gcttcctgtc	ctcctaggac	cccatgcctc	ctcaaacgtc	1020
cacgtgtttc	ttgctcctcc	accacgggtg	ccaagccaaa	attcgggtgg	gcggggaggac	1080
atttttccaa	ctgcctgttt	cccttctttt	ccttttgaca	ccccagataa	atcatctttc	1140
ccaatccaac	acagcccac	tgtgtctttg	gggacttcac	gacatcaccc	aggaatgtat	1200
ccttagaaac	aaaaatgcaa	aaaccagaa	accaggagac	aattaaagaa	attttctactg	1260
gtgaggtcac	aagtagtaga	gacttcttgt	taacgggcag	aaactttcac	ggaccacagca	1320
tgctactgtg	gcagttctgc	aacaagctga	aatgccttt	cccagaccac	caagccagtg	1380
ccacacaaa	gccaccttag	ggtgtgcaca	ggatgtcact	aggcgttggc	ggaactcagg	1440
aaggagtctg	aattttcttc	cgtttcttcc	ttcctctctc	attccctatc	ttagcttctg	1500
tctctctttc	ctctctctcg	ttccccccct	tcctccctcc	cttctgtgtg	cagggccaca	1560
gatggaccgg	gagacctcaa	gcatgtcaaa	tcaactaact	gctctaccac	tcaaccacac	1620
cctcgctctg	attgttacta	ctactattat	tatcttgata	caggctctca	cattgagctc	1680
accctcacag	tctccacatt	gagctcacc	tcacagtctc	cacattgagc	tcaccctcac	1740
agtctccaca	ttgagctcac	cctcacagtc	tcacattga	gctcacccctc	acagtctcca	1800
cattgagctc	accctgtggc	tctggcaaac	cttgaattct	ctcattcctc	ctgcctcagc	1860
ctctgggggc	gtggggatta	gccaaaacca	cttgagggtt	tcttcaatca	gcaaattctt	1920
agcgttcaat	taacacacac	tcataactcc	agtactttgg	aaaccggaac	aggagaattt	1980
ctgtgagctg	gaggctagct	tggaactacg	tatgagaccc	tgtctctaaa	taaatacaca	2040
aagaaatctc	accaagggcc	tcctctctct	agcaagctct	aactgtgggtg	ggagttctgg	2100
gttgttccag	ttaacgggct	cagaactcta	ctgcccagca	catcagcccc	tagaacacag	2160
tggtctctca	catgtgaaca	tgcagtcaca	gaaatgaaat	aaagtgaata	ttttatttct	2220
tcagttgtat	agcctcttcc	gtgtgggctg	tagttactgt	cttgaatagg	ataggctcag	2280
aatccttggt	gctggaacca	agagtttgat	tccattagac	gacagggaat	ataatgccca	2340
atagggcatt	cctcctcccg	gtcactagcg	gtgcactttc	tccgaatctt	tgtcatgttg	2400

aattagaaaa	gttagtattt	tcctccatcc	cttccccctcc	tccccctctc	ccctccctccc	2460
ctcctcccc	cctccccctcc	tctccccgc	cctccccctcc	ccctctgata	ctcccccatc	2520
tataaatcc	aagaattcca	gtaaaaagag	gaaaacaatc	gaagtgattt	cggttgattgt	2580
cagttccacc	aaagcaagac	ttgacttttag	ttccgcgttt	cggttcccgg	catgcaccac	2640
agccagcgag	caccgtggaa	ggatgctagc	acgggtctcc	ccccgcccc	actagctgtc	2700
ttcagctccc	cagtagagg	caaccgact	ccagattctc	aatggagagt	gtttacacaa	2760
tcgttgccgg	tttgtgtgag	cgcgcccgt	tccagagaca	cttcttcttt	ttcttttttt	2820
catttcatcc	cagtggcaac	gcagagtgcc	agatcattca	ggcgttttgc	agggcaagcc	2880
gtgggagctt	ggcaagcaag	gccccatttc	ctagggaaac	cgtgcctggc	gcttcaggaa	2940
agcacgggaa	cctggcactg	tgactctgcg	ggtattattt	tgagaactc	tttattaaac	3000
gggagtttca	agtccagctg	gagacgacca	ggcagcgcc	ttaaccccag	agtcacacac	3060
aggtgccttt	tcttggggcc	agattggggt	tgtgtggcag	acctgcgacc	agcttgacaa	3120
ctcttctgcc	aggccacaaa	atggtgttgg	ctgtaagagg	tgacaccagg	gacagggaag	3180
atcgctgcta	ttctcctgag	ctctccaaag	acccacacca	gtctgtcccc	ctttcctcct	3240
gctcttcccc	tgtatcgccc	cctcaccatc	tcccccaacg	agactcttgg	catctcctcg	3300
gcacaaggat	ttgaaaatag	atgcttgggg	gtgagaagaa	gaagagagaa	agagagagaa	3360
ggaagggaag	atatatagat	gatacagacg	catacaggtg	acatgtagct	aatcattttt	3420
aattaaaaaa	taaattaaaa	gcaaatacaag	gatataatag	atacccttag	agcaagtgtc	3480
tcatacacac	acaaacacac	acacacaata	tatatatata	tatatatata	tatatatata	3540
tatatatata	ttataacttg	aacaagtgtc	cagaagggct	ggggactcta	aagtgttgt	3600
caaagccagg	ctcacatcag	taatcttata	acctggtaga	ctgagacagg	aggattttga	3660
tgagttcagg	cccagcctga	gctgcagaat	gtgattctat	cccaaaaaag	taaaataaaa	3720
taaaattcaa	aatacacgaa	aagagtattt	gctgaacaaa	caagcctaaa	gccctggatc	3780
ccttccccca	tgtcctaaga	aaataagttt	cttgaagctg	gagggatggc	tcagaggtta	3840
agagccccag	ctgcacttgc	ggaacactaa	gaccagttc	ccagacccca	caactgtgggt	3900
cacaactgtc	tcaaacgcca	gctccggagg	atccatgcc	tctcctggcc	tccaccggca	3960
ccaagaacac	atacagtgcc	catacattta	tgcaagcaag	gtattcacgc	acataaaaact	4020
aaaagaatat	ttaataaaga	tataacaaaa	tagcatgaag	cccagctgg	acagaggttc	4080
aaactacatc	ccaggttcac	ccctctgcct	ttgctctcag	ttggcttggg	taggtctctt	4140
ctctgaactg	gcgcctcg	ggttccact	tgagaccctc	tcatttttaa	acctacttct	4200
tctgggcggg	gttaattgct	gccagggtc	aagccaacgc	ttcctcttct	ccacagcaat	4260
cttccaagtt	tcacgagata	accaggaact	gctaagttca	tgtgaacctt	agtgaagaac	4320
ctgagctctc	ccatgtgatt	ggtgtgtgca	tgtgtgcata	cacaaatgta	tgtgtgtgct	4380
ctatgtgtgc	ctatgtatgt	gtgcatgcat	gtgtgcata	acaaatgcat	atatgtctat	4440
gtagtgtgcg	tacacaaatg	tatgtgtgtg	ctcaatgtgt	gcctatgtgt	gtgtatgcat	4500
gtgtgcgtac	acaatgcatg	tgtgtggtgt	ctgtgtgcct	gtgtgtgtat	gcatgtatgc	4560
atacacaaat	gtatatgtgt	ggtgtgtgaa	tgtgtgccta	tgtatgtgtg	tgtgtgtgtg	4620
gggtgtggta	tgtgtgtgat	gtgtggagg	gtgtgtatgt	gtggatgta	taggtgatac	4680
gtttgggggtg	taatatgcgt	atgtggtttg	tgaaatgtag	ttcgtgtgtg	tgcatgtgtg	4740
cgtgcgtgcg	tgcgtgcgtg	cgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtt	ggatatagta	4800
tgtgtgaggt	gtgtgtactc	accatggcct	cctcacttg	ggggagtga	gtcagcagcc	4860
tggaccactc	aggacatga	gatactcaga	cacatcttga	tttccacccc	tcttttctctg	4920
atcctccttc	acgtgtcact	ttcccaaaaca	ctggacaaca	gtttgggggc	atctgattcc	4980
actaatgaca	gggacatcac	atgtctccag	agggaaacacc	ttctgtgtca	catgtcatct	5040
gagaatgtag	cagagtcaca	gagaaatgtc	acagaaacca	aaatgcagag	taccaaggta	5100
tagctaggga	cagagcagag	gggaagccgc	tgaatttatt	aaaaatgtca	gaatcgtaaa	5160
agacagggga	cagcgggtgg	gacattcagg	gtccagtagc	acacagggcag	tccaaacctg	5220
atcactggaa	ggtagttagt	aaggaaaggc	tgacacacaga	ttattcacac	agtttatata	5280
tgtacacaga	ttattcacat	ggtttgtgta	tgtgcacaga	ttattcacac	agtttatata	5340
tgtgtggctt	cgtggtaact	ttgagcttac	tttcaattta	aaaggatctc	tctcacaaagc	5400
tggggccggg	aatggctgca	gtcaacactc	catcacttag	tcacactgtg	caaacagcac	5460
ctcctgactc	atggtgactt	gtagtaaaat	gaagaggcca	catttgcata	caagacagct	5520
catcagtacc	tagtgaagaa	tctgtccctg	agtatttgca	tgaatggacc	cgggtccagg	5580
gcctggctgg	gagtcctcag	gtgttgccagc	cagaatgtca	ttgtgttttt	tcaggatccc	5640
agaagtctct	aaaatacagg	ccaagtactc	atltgtgtta	caaagtatct	gactaataga	5700
agtgattagg	taacacaaa	ccttttaaaa	accgagatca	cccttgatct	gtccctggcc	5760
tcttagaaca	agatccaagc	ttttgtgtgt	tgacaagtgg	ggccatccag	tgcgtctccg	5820
ttcctgtctac	ttcatctgga	agacctctcc	cactaaactg	cccctgacc	ctcacacctg	5880
ctgtttcctt	tccaccggga	agtgtgtgtc	taggttttca	tggccatctg	actgagcatc	5940
taggcctcag	tccagtggtc	cctcagctct	ctctagtcac	tgtactaatg	gaaacggcca	6000
ctaactacat	tttcaatatg	gaagcctcct	cctcaggaa	ctccaagggc	agaagcctcc	6060

agagaaccac	tcctgacccc	ctggagttct	gagtgccttc	ggccctctct	gtgtctgcag	6120
gactattcac	cacttggtgt	gaatggttca	gtcctcacct	cctctggcat	gtgctcagtt	6180
ctcatctcat	tggggagtc	ttcccaggtc	actcttctct	cctgtctttg	aagtgttttt	6240
ttccttcatg	gtatttctgt	ctgggcacac	acacagacac	acatacacac	acatacacac	6300
ccatgcagta	tggcagatac	atcacctatg	tttcagattt	ttattctacc	atcacccaat	6360
acctgaatcc	ccgaaaaagc	cttagaaagc	caggaatttg	tgtatttttg	tcagcactcc	6420
acccagcac	ctgaagccaa	gcctgactta	atatttttgg	ttttgtttct	aga	6473

<211> 7045

<212> DNA

<213> *Cricetulus griseus*

<400> 4

(SEQ ID NO: 8)

caattgatta	tagatggatg	atagatagat	agatagatag	atagatagat	agatagatga	60
tggatagaca	gatgatggat	agttagagga	tagataatga	ctgaataata	agtacataaa	120
tagatgatag	agcggggcgt	tgggtggtgca	cgtctttaac	cccagcacca	gagaggcaga	180
ggcagttgga	tctctgtgag	tttgaggaca	gcctggttac	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	gccttctttg	480
aaagtgacat	ttgagagcat	ttcaaacgct	gttcatgtca	ggcatgccaa	tggggagaga	540
agggtcttga	gaaagcaggc	ccggcaagcc	atggggagca	agctaggagg	cagcattcct	600
tgcattttgcc	ctgcctcag	ctgcttctctg	gagttccccg	gtttttatca	caacagtaga	660
aataaaaacca	ggacaatggt	gtttccatgc	atacatctgc	aagaacttac	tccggttcaa	720
tagacagacc	aaggcacctg	tgtttgctca	agaagcacgg	agggaggtgt	gtgcacctgc	780
tgggtgctgg	tgtctgtggt	gtgccagaca	gagagcaaga	caggaaagtt	cctgggtggc	840
tagagcacac	agcccagccc	aggaagtcac	gtctctctct	gtctctgtct	ctgccccacc	900
cccaccccat	ttaggccaga	gaacagctgt	ggcaagcttt	gggtttgggt	gagtcattcc	960
tcaagagcca	agagccggcc	acottgtatg	gggtagtttg	ttgttgttgt	tgttgtttatt	1020
atttgtttgt	ttgtttgttt	ggtaaagggt	tttcaatagg	agttggaatt	tggcaattca	1080
gctaggctgg	ctgagcagcc	agctagcccc	gggcactcat	ccgtctctac	ctccccagtt	1140
ctgggatttc	gggtacatgc	tgccacatcc	gacttttttc	ccctgctcca	gttcttaaga	1200
ccaagtcttc	atgtcaaaca	cttcaaccacc	ttagccatct	ttctgggtca	gaagttagat	1260
cttcagggaag	acaaggagtg	tatcaggaca	tgagcgtgcc	ccaactctgc	tcagaccttc	1320
tgatagagaa	aatgggggga	gggtgtcag	aggctgccgg	agaaagacaa	gtccaggtta	1380
aggaggacga	ccctgggctc	tgaatccaag	gggtgattccc	tcaccttgta	cacttggcat	1440
tttggaagg	aagcatcaga	taaaagcagt	gcagacatag	tcagggaatat	ttacacgtgt	1500
gagtcaacct	gggagtgagt	ctgtgtacaa	ctgaacatga	agcaagtttt	gaagcttcat	1560
ttccagacta	ttcccagggt	gcaataaact	cctgttttcg	ttgcagcctt	cccagtctct	1620
gccactgcca	tctctacttc	agtctggaat	gggtgggcaca	cagaaaaagt	ctatggcaat	1680
cctgcgagaa	gacaagtggg	cgctgactt	cgggctcctg	ttacaagaga	ggaatccagg	1740
agtttatttt	gcagctgatt	cagtgttgac	caagagtcca	gctctggggg	agtgggaagc	1800
aaccaaagca	gagacaggtc	ccagcacat	ttttggtttt	caagacagca	cttctctgtg	1860
gctttgaagg	ctatcctaga	actgttcttt	gtatatcctt	ccttgcaact	agctcttata	1920
gaccaggctg	gtcttgaact	cacagagatc	catctgcctc	tgctcccaa	gtgctgggat	1980
taaaggcgtg	cacctcggt	gccaccaccc	agctacatac	ataattttaca	ataataaaaa	2040
taaaataactt	taaagtgtta	tagcagtttg	aatgtaattg	gccctgtcat	ctcataggga	2100
gtggcactat	taggaggtat	ggctttgttg	aaggaaatat	gtcactgtga	gggtgggctg	2160
tgaggtttcc	tatgctcagg	gtaccagcca	gtgtctcagc	tgaggctcctg	ttgcctgcaa	2220
gatgtaggac	tctcatccct	ttctccagca	ccatgtctgt	ctgcatgcc	tcatgttccc	2280
agccatgatg	acaatgtact	aaaacctctg	aaactgccac	ccaactaaat	gttttctctt	2340
ataagagttg	ccatgctcat	ggtgtctctt	caccgaata	gaaaccctaa	ctaagataag	2400
tgtattctcc	cctactcccc	atgattctaaa	atcttaggaag	gcaggtaggc	aggcaggcag	2460
gctggtatag	tggttcattc	tagcacctga	gacctggaat	gggaggattg	tgagttagtt	2520
ctaggccatt	ctgggtgccta	gaaaccagag	ccgggggttg	gccaatgca	gagcacttgc	2580

tctacgtatg	gcccagcaca	ataagtcfaat	ttcctcacct	ttaaaggcttg	acaattttaa	2640
aacactgggt	tttagttagt	ccgtgtctgc	tccacagatg	gagacagcta	atcacagatg	2700
catcaggggc	cttcctgagt	gctaaacatc	aaacagcctt	ctccctcct	gagcctttgt	2760
gtgcagaatg	tgtccatcgc	aagaagcaaa	cagtcttgct	tgcccaccaa	cttccttcct	2820
gcatcagaag	agctgggtgc	aaactgcaag	agtagcctca	ccttagagat	gggtcccatt	2880
gctctacatg	ggagcattac	cttccaagaa	ggcaaaaatg	tctcctgggt	gagctttttt	2940
tgtcacctgt	taaaggcaaa	tcaacagaga	ggctttgtct	caccactaa	catcttgga	3000
acaaatacca	acgaacgctg	gggaggatgt	ggggaaagca	gagccctcat	gctctccgag	3060
ggaaaatcac	accactgtg	gaacagtgtg	gaaacctcaa	agactgggat	tacaagcagc	3120
acacaagcca	gccacgctac	tcttggtcac	acaccacaaa	gacgcttgca	cattcacgct	3180
tacgtcgca	acactagcaa	cgttcccact	gcctcctttg	agccccgccc	ccccccctg	3240
ccccccgccc	cgccccgtg	gtctatgttc	ctcttcctca	aagtcagctt	ccacttctct	3300
gtctccatct	tcgccccacc	ctccctcctc	gctacataat	tgtctctatt	ccatttctct	3360
gctttgaaac	agctttttgc	aaagcatcaa	atctattgtc	ctatgcccc	aatcaacctc	3420
cagtttcaca	agtatacag	gaaatcgttt	tcctaattaa	aaatcccccc	tttgaccatt	3480
tattcccact	cttggaacat	cttccccttg	aggaaagtta	cagaatgagg	tggtctctct	3540
cttcctatct	gaggtgtttc	cttcagactt	tgtccgtgtc	taatcttttt	aactgttggc	3600
caggcctcca	ccacggcaca	gatgaactgt	ggggttcatt	tacctgaaac	tctatggaag	3660
gatgtttatt	tctccttcac	tttagcaaat	gataaagggc	accattcaact	ctgtctattc	3720
tgcagggggc	attcctttct	ctaggccaga	tactgagaat	tgctcccaga	atcaatgtgg	3780
tatacatatt	cccccttcaa	catgatagg	cattgatcac	acacacacac	acacacacac	3840
acacacacac	acacagtatg	acaaatgtat	tcccctagcc	cgcttccatc	ttgccacagg	3900
actccagagt	ggccctggat	agcaagcttc	ctggtttgtt	tctctgttcc	tgctgctttt	3960
ccaccctcca	gtctatcttt	tctaagtcct	tctgccattg	tctcttccc	aactgtcctg	4020
agatgcagtc	attgtctggg	attcagacct	tctctctctg	cccaagttag	tattattgacc	4080
cccacgggtt	gtacaacat	aacttcaggg	agcccgacaa	aaactgtttt	atgagccaag	4140
tagtcccagg	acttgagagg	tagaggcggg	aagatcagca	gtttgaggcc	agcctggaga	4200
gcataagagc	cgggtctcaa	acaacaatgg	aaactagata	ctaagtaaaa	atcctggggg	4260
gtttctcat	gaatgtctgt	tcttctagta	ccacgctgaa	ctcgtacac	agctccagct	4320
gttacggctt	tcttagaatc	ctactctttt	tttttttttt	tttttttttt	tttttttttg	4380
tttttgcaga	cagggtttct	ctgtggcttt	ggaggctgtc	ctggaactag	ctcttataga	4440
ccaggctggg	ctcgaactca	cagagatcca	cctgcctctg	cctccagagt	gctgggatta	4500
aaggcgctgc	ccaccaacac	ccggcagaat	ccatactctt	tttaaaaaaa	gatttatcaa	4560
tttactatgt	atacagcttt	ctgcctgcac	gtatccatgc	atgtcagaag	atggcaccag	4620
gtcgcattac	agatggttgt	gagccacct	gtgggtgctg	ggaattgaac	tcaaatgtc	4680
tagaagagca	accagttctc	ttaacctctg	atcactctct	ccggccccca	gaaatccata	4740
ttcttgagga	tttttttacac	ccccccccac	aaaagacgta	tatctaaatt	ttaatgtgag	4800
aattcacatt	ttcttaagag	ttgaacatag	atthagagga	aaatcagatc	ccacatgatt	4860
aacaaagcat	gcttggtggc	aggtctgcta	ccaagagggt	ggccgtagct	tctagctcag	4920
acaaaactcac	tcccttctct	gtggcctctt	cgcctcaag	tcagaaaactc	accctgtgat	4980
tctgccccag	aagttgtctc	agagcacagt	gcactcctcc	gtcttcactc	tgtggcttga	5040
attgtgtcca	tcgcttatga	ttacaacccc	tcacagagca	tcctaactgg	tttcttttgc	5100
tgcttatggg	cactcctcca	ttctagaaca	cccttgccat	caatactatg	aaaggagggg	5160
tggaggagga	agagcaggaa	gaggaggggg	aagcgaggga	agaggaaagc	acggatggca	5220
atgaggaggg	gggagcacc	aagtcctccc	tggatgagag	tctcactggg	agacttaata	5280
ttaattataa	atgcttggtc	agcagctggg	caggataaag	ttaggcagga	gaaccagact	5340
aaggactctg	ggaagcagaa	gggcagagtc	agacaaggag	aggaaacagg	aagtacaagg	5400
taaagtcaag	tggcagaatg	tagataatag	aatgggttc	atttaagttg	gaagagttag	5460
ctagtaacaa	gcctgagcta	tcagccgagc	atttataatt	aatattgagc	ctccatattg	5520
gttatctggg	aattggcggg	cagaaaaaaa	aaagtctgcc	tacaagtcaa	tgtcatgtag	5580
ctcccaaagc	caaggtacct	ttgttcagtg	cttgactgag	ccagcattat	aaattttctc	5640
cagatgtacc	gaatcacatt	tcatagcaac	atgcagacat	caagttttcc	ctgaagctct	5700
aaccagctgg	ttgcatgctg	tccggagtct	cagctataac	ccagaagtga	cctgggtcgg	5760
ggaagagggtg	gtactttgcc	ttctttgcac	tctctgtgtt	gcctcaccca	ttcagcttca	5820
agcaatgtga	ctgcctgacc	ctgaggcggt	ttacaacgcc	tgaccacacg	accacaagtc	5880
aaccagctgg	tgtgctcacg	atccctagtc	tgaaccatag	cctgctccc	accctgcctc	5940
catctccacc	ctttcttcac	tgtcatcac	agctggctag	caaagactgc	ctcagacctg	6000
agcacaggct	ccactccaca	gccgtgactg	ttcgagccac	ttaaatacaa	gagcgcttgt	6060
cttcogctca	gtaaatctct	cctcagctca	ctgatgacgt	tgactttctc	tagacagcac	6120
atgttggttt	aagacactgc	tacttgagct	cttcattcag	ttcctcagaa	tacctcattt	6180
gggtcagatt	cccaaagagg	aagatagggt	tcctggcaga	cagacatgtc	tcattccttt	6240

gaaatccttc	agagaaatgc	agtgactatg	gcaccttctt	aaaaagcaca	cacacaaata	6300
acacacacac	acacacacac	acacacacac	acacacacac	atatccccct	cactgtcatc	6360
cttgatatgt	atatgatata	tataaaatca	ttgtttttata	ctgtgataat	tgattatgaa	6420
taaaattttac	taaaatgaac	aattaaaatt	atggggggggg	ctggagagat	ggctcatcag	6480
ttaagagaac	agttgctgct	cttgacagaac	acgagagttc	agttcccagc	acccacatca	6540
ggcagctcat	aacctgtgt	ggtgtcagtt	ccaggagatc	tggtgccttc	ttctggcctc	6600
ctccagcacc	tgctacatgt	ggttcacaca	cacacacaca	cacacacaca	cacacacaca	6660
cacacacaca	caaataaata	taaagattat	ttttttcaaa	actgagttaa	aaatagggttc	6720
tatctgattc	atactaaggc	ttttcacagt	ggttaagtct	attagatatg	tctagccata	6780
tcctttctcc	cttctttctt	gaggagaggc	ttttaaagct	acaagttaca	gccttctttg	6840
caaataagag	taccatttaa	caggcctctg	accaatgaga	tgccagaatc	ggttgcccag	6900
gagcttccca	aacagtcct	tatagggaaa	ggtgggtacaa	accagtagat	taggcatgtt	6960
ccacttccca	agtgccgtgc	caaataagga	aatggcctca	aatgtttgcc	ttttatcttc	7020
accacctct	gaattgcacg	ctagt				7045

<211> 13515

<212> DNA

<213> *Cricetulus griseus*

<400> 5

(SEQ ID NO:9)

tctagaaaca	aaacaaaaaa	tattaagtca	ggcttggtct	cagggtgctgg	ggtggagtgc	60
tgacaaaaat	acacaaattc	ctggctttct	aaggtctttt	cgggggattca	ggtattgggt	120
gatggttagaa	taaaaatctg	aaacataggt	gatgtatctg	ccatactgca	tgggtgtgta	180
tgtgtgtgta	tgtgtgtctg	tgtgtgtgcc	cagacagaaa	taccatgaag	gaaaaaaaca	240
cttcaaagac	aggagagaag	agtgacctgg	gaaggactcc	ccaatgagat	gagaactgag	300
cacatgccag	aggaggtgag	gactgaacca	ttcaacacaa	gtggtgaata	gtcctgcaga	360
cacagagagg	gccagaagca	ctcagaactc	caggggggtca	ggagtgggtc	tctggaggct	420
tctgcccttg	gaggttcctg	aggaggaggc	ttccatattg	aaaatgtagt	tagtggccgt	480
ttccattagt	acagtgaacta	gagagagctg	agggaccact	ggactgaggc	ctagatgtc	540
agtcagatgg	ccatgaaagc	ctagacaagc	acttccgggt	ggaaaggaaa	cagcagggtg	600
gaggggtcag	gggcaagtta	gtgggagagg	tcttccagat	gaagtagcag	gaacggagac	660
gcactggatg	gccccacttg	tcaaccagca	aaagcttgga	tcttggttcta	agaggccagg	720
gacatgacaa	gggtgatctc	ggttttttaa	aggctttgtg	ttacctaatc	acttctatta	780
gtcagatact	ttgtaacaca	aatgagtact	tggcctgtat	tttagaaaact	tctgggatcc	840
tgaaaaaaca	caatgacatt	ctggctgcaa	cacctggaga	ctcccagcca	ggccctggac	900
ccgggtccat	tcatgcaaat	actcaggggac	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	gacaggttt	1260
ggactgcctg	tgtgtctactg	gaccctgaat	gtccccaccg	ctgtccctcg	tcttttacga	1320
ttctgacatt	tttaataaat	tcagcggctt	cccctctgct	ctgtgcctag	ctataccttg	1380
gtactctgca	ttttggtttc	tgtgacattt	ctctgtgact	ctgtacatt	ctcagatgac	1440
atgtgacaca	gaaggtgttc	cctctggaga	catgtgatgt	ccctgtcatt	agtggaatca	1500
gatgccccca	aaactgtgtc	cagtgtttgg	gaaagtgaca	cgtgaaggag	gatcaggaaa	1560
agaggggtgg	aatcaagat	gtgtctgagt	atctcatgtc	cctgagtggt	ccaggctgct	1620
gacttcactc	ccccagtgca	gggaggccat	ggtgagtaca	cacacctcac	acatactata	1680
tccaacacac	acacacacac	acacacacac	acgcacgcac	gcacgcacgc	acgcacacac	1740
gcacacacac	gaactacatt	tcacaaacca	catacgcata	ttacacccca	aacgtatcac	1800
ctatacatatc	cacacatata	cacccctcca	cacatcacac	acataccaca	cccacacaca	1860
gcacacacac	acataggcac	acattcacac	accacacata	tacatttgtg	tatgcataca	1920
tgcatatacaca	cacaggcaca	cagacaccac	acacatgcat	tgtgtacgca	cacatgcata	1980
cacacacata	ggcacacatt	gagcacacac	atacatttgt	gtacgcacac	tacatagaca	2040
tatatgcatt	tgtatatgca	cacatgcatt	cacacatata	taggcacaca	tagagcacac	2100
acatacatatt	gtgtatgcac	acatgcacac	accaatcaca	tgggaagact	caggttcttc	2160
actaaggttc	acatgaactt	agcagttcct	ggttatctcg	tgaacttgg	aagattgctg	2220

tggagaagag	gaagcggttg	cttgagccct	ggcagcaatt	aaccccgccc	agaagaagta	2280
ggtttaaaaa	tgagaggggtc	tcaatgtgga	acccgcaggg	cgccagttca	gagaagagac	2340
ctacccaagc	caactgagag	caaaggcaga	gggatgaacc	tgggatgtag	tttgaacctc	2400
tgtaccagct	gggcttcatg	ctattttggt	atatctttat	taaatattct	tttagtttta	2460
tgtgcgtgaa	taccttgctt	gcataaatgt	atgggcaactg	tatgtgttct	tgggtgccggt	2520
ggaggccagg	agagggcag	gatcctccgg	agctggcggt	tgagacagtt	gtgaccacca	2580
gtgtggggtc	tgggaactgg	gtcttagtgt	tccgcaagt	cagctggggc	tcctaacctc	2640
tgagccatcc	ctccagcttc	aagaaactta	ttttcttagg	acatggggga	agggatccag	2700
ggcttttaggc	ttgtttgttc	agcaaatact	cttttcgtgt	attttgaatt	ttatttttatt	2760
ttactttttt	gggatagaat	cacattctgc	agctcaggct	gggcctgaac	tcataaaaat	2820
cctcctgtct	cagcttacca	ggtgataaga	ttactgatgt	gagcctgggt	ttgacaagca	2880
cttttagagtc	cccagccctt	ctggacactt	gttccaagta	taatatatat	atatatatat	2940
atatatatat	atatatatat	atatattgtg	tgtgtgtgtt	tgtgtgtgta	tgagacactt	3000
gctctaaggg	tatcatatat	atccttgatt	tgtttttaat	ttatttttta	attaaaaatg	3060
attagctaca	tgtcacctgt	atgcgtctgt	atcatctata	tatccttcct	tccttctctc	3120
tctttctctc	ttcttctctc	cacccccaa	catctatttt	caaatccttg	tgccgaggag	3180
atgccaagag	tctcgttggg	ggagatggtg	agggggcgat	acaggggaag	agcaggagga	3240
aagggggaca	gactgggtg	ggtctttgga	gagctcagga	gaatagcagc	gatcttccct	3300
gtccctggtg	tcacctctta	cagccaacac	cattttgtgg	cctggcagaa	gagttgtcaa	3360
gctggtcgca	ggtctgccac	acaaccccaa	tctggcccca	agaaaaggca	cctgtgtgtg	3420
actctggggg	taaaggcgct	gctgggtcgt	ctccagctgg	acttgaaact	cccgtttaat	3480
aaagagttct	gcaaaaataat	acccgcagag	tcacagtgcc	aggttcccg	gctttcctga	3540
agcgccaggc	acgggttccc	taggaaatgg	ggccttgctt	gccaagctcc	cacggcttgc	3600
cctgcaaacg	gcctgaatga	tctggcactc	tgcgttgcca	ctgggatgaa	atggaaaaaa	3660
gaaaaagaag	aagtgtctct	ggaagcgggc	gcgctcacac	aaacccgcaa	cgatttgtga	3720
aacactctcc	attgagaatc	tggagtgcgg	ttgccctcta	ctggggagct	gaagacagct	3780
agtggggggc	gggggaggac	cgtgctagca	tccttccacg	gtgctcgctg	gctgtggtgc	3840
atgccgggaa	ccgaaacgcg	gaactaaagt	caagtcttgc	tttggtggaa	ctgacaatca	3900
acgaaatcac	ttcgattggt	ttcctctttt	tactggaatt	cttggaattg	atagatgggg	3960
gaggatcaga	gggggagggg	aggggcgggg	agacggaggg	aggaggggag	gaggggagga	4020
gggggaggag	ggaggagggg	aagggatgga	ggaaaatact	aacttttcta	attcaacatg	4080
acaaagattc	ggagaaagt	caccgctagt	gaccgggagg	aggaatgcc	tattgggcat	4140
tatatccct	gtcgtcta	ggaatcaaac	tcttggttcc	agcaccagg	attctgagcc	4200
tatcctattc	aagacagtaa	ctacagccca	cacggaagag	gctatacaac	tgaagaaata	4260
aaattttcac	tttatttcac	ttctgtgact	gcactgttcac	atgtagagag	ccacctgtgt	4320
ctaggggctg	atgtgctggg	cagtagagtt	ctagaccctg	taactggaac	aacccgaaac	4380
tcccaccaca	gttagagctt	gctgagagag	ggaggccctt	ggtgagattt	ctttgtgtat	4440
ttattttagag	acagggtctc	atactgtagt	ccaagctagc	ctccagctca	cagaaattct	4500
cctgttcccg	tttccaaagt	actggagtta	tgagtgtgtg	ttaatgaa	gctaagaatt	4560
tgtgattga	agaaaacctc	aagtgggttt	ggctaattccc	cacgacccca	gaggctgagg	4620
caggaggaat	gagagaattc	aaggtttgcc	agagccacag	ggtgagctca	atgtggagac	4680
tgtgaggggtg	agctcaatgt	ggagactgtg	agggtagct	caatgtggag	actgtgagg	4740
tgagctcaat	gtggagactg	tgagggtgag	ctcaatgtgg	agactgtgag	ggtgagctca	4800
atgtggagac	ctgtatcaag	ataataatag	tagtagtaac	aatgcaggcg	aggggtgtgt	4860
tgagtggtag	agcagttagt	tgatttgaca	tgcctgaggt	ctcccggtcc	atctgtggcc	4920
ctgcaacagg	aagggagggg	ggaagggggg	gaacgagaga	gaggaaagag	agacagaagc	4980
taagataggg	aatgagagag	gaaggaagaa	acgggaagaa	attcagactc	cttccctgag	5040
tccgccaacg	cctagtgaca	tcctgtgcac	accctaaggt	ggcctttgtg	tggcactggc	5100
ttgggtggtc	gggaaaggca	ttttcagctt	gttgacagaa	tgccacagta	gcatgctggg	5160
tccgtgaaag	tttctgcccg	ttacaagaa	gtctctacta	cttgtgacct	caccagtga	5220
aattttcttta	attgtctcct	ggtgttctgt	gttttgcaat	tttgtttcta	aggatacatt	5280
cctgggtgat	gtcatgaagt	ccccaaagac	acagtggggc	tgtgttggat	tgggaaagat	5340
gatttatctg	gggtgtcaaa	aggaaaagaa	gggaaacagg	cacttgggaa	aatgtcctcc	5400
cgccaccccg	aatttttgct	tggcaaccgt	ggtggaggag	caagaaacac	gtggacgttt	5460
gaggaggcat	ggggtcctag	gaggacagga	agcagaagga	gagagctggg	ctgacagcct	5520
gcaggcattg	cacagtttca	gaaggagatt	acagcatgac	tgagttttta	gggatccaac	5580
agggacctgg	gtagagattc	tgtgggtctc	gaggcaactt	gacctcagcc	agatgggtatt	5640
tgaataacct	gctcttagag	ggaaaacaga	catagcaaac	agagccacgt	ttagtgtatga	5700
aactctcact	ttgctcggtg	catgtccagg	cagtcaccag	ggtcaggctg	acactcaact	5760
caaaaacaag	tgagaaattg	aagacaatcc	gtggtggcag	ctactggaag	ggccaccaca	5820
tccccagaaa	gagtggagct	gctaaaaagc	catttgtgat	aggcacagtt	atcttgaatg	5880

catggagcag	agattacgga	aaaatcgaga	atgttaatga	ggcaacattc	gagttgagtc	5940
attcagtggtg	ggaaacccag	acgcttccat	cccctaaaag	gaacatcttg	ctctcagtc	6000
aaatggaaat	aaaaattggg	gcttgaattt	ggcaaatgat	tcagaactct	gtgtaggtat	6060
tttcacacgc	acagtggata	attttcatgt	tggagtttat	ttgtgctaaa	aggcagaaaa	6120
gggtaaaaag	cacatcttaa	gagttatgag	gttctacgaa	taaaaaataat	gttacttaca	6180
gctattcctt	aattagtaac	cccttccacc	tgtggtaatt	tcctgagata	gtcagtgagg	6240
aaaagatctc	tccttctctt	ctttctcccc	ctccctcct	ctccctccct	ccctccctcc	6300
ctccctcctc	tccttccctc	cccctttcct	tctttctttg	ctccttctcc	tctgcctcct	6360
tctccctttc	ttcttcattt	attctaagta	gcttttaaca	gcacaccaat	tacctgtgt	6420
taacgggaaa	acacaggctc	aagcagctta	gagaagattg	atctgtgttc	actagcgtgc	6480
aattcagagg	tgggtgaaga	taaaaggcaa	acatttgagg	ccatttcctt	atttggcacg	6540
gcacttagga	agtggaaacat	gcctaatact	ctggtttgta	ccacctttcc	ctataatgga	6600
ctgtttggga	agctcctggg	caaccgattc	tggcatctca	ttggtcagag	gcctgttaaa	6660
tggtagctct	atttgcaaag	aaggctgtaa	ctttagctt	taaaagcctc	tcctcaagaa	6720
agaagggaga	aaggatatgg	ctagacatat	ctaatagact	taaccactgt	gaaaagcctt	6780
agtatgaatc	agatagaacc	tatttttaac	tcagttttga	aaaaataaat	ctttatattt	6840
atttgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gaaccacatg	6900
tagcaggtgc	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	gatattgtgtg	tgtgtgtgtg	tgtgtgtgtg	7200
tgtgtgtgtg	tgtgttattt	gtgtgtgtgc	tttttaagaa	ggtgccatag	tcactgcatt	7260
tctctgaagg	atttcaaagg	aatgagacat	gtctgtctgc	caggaaacct	atcttccctc	7320
ttgggaatct	gacccaaatg	aggtattctg	aggaactgaa	tgaagagctc	aagtagcagt	7380
gtcttaaac	caaagtgtct	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	tggttgactt	gtggtctgtg	ggtcaggcgt	tgtaaacgcc	ctcagggtca	7680
ggcagtcaca	ttgcttgaag	ctgaatgggt	gaggcaacac	agagagtga	aagaaggcaa	7740
agtaccacct	cttccccgac	ccaggtcact	tctgggttat	agctgagact	ccggacagca	7800
tgaacaccagc	tggttagagc	ttcagggaaa	acttgatgtc	tgcatgttgc	tatgaaatgt	7860
gattcggtac	atctggagaa	aattttataat	gctggctcag	tcaagcactg	aacaaaggta	7920
ccttggcttt	gggagctaca	tgacattgac	ttgtaggcag	actttttttt	tctgcccgc	7980
caattcccag	ataaccaata	tggaggctca	atattaatta	taaatgctcg	gctgtagct	8040
caggcttggt	actagctaac	tcttccaact	taaatgaacc	catttctatt	atctacattc	8100
tgccacgtga	ctttaccttg	tacttctctg	ttctctcct	tgtctgactc	tgcccttctg	8160
cttcccagag	tccttagtct	ggttctcctg	cctaacctta	tcctgccag	ctgctgacca	8220
agcatttata	attaatatta	agtctcccag	tgagactctc	atccaggagg	gacttgggtg	8280
ctccccctc	ctcattgcc	ctcgtgtctt	cctcttccct	cgcttcccc	tctctgtcct	8340
gctcttctc	ctccaccct	cctttcatag	tattgatggc	aagggtgttc	tagaatggag	8400
gagtgcccat	aggcatgcaa	agaaaccagt	taggatgtct	tgtgaggggt	tgtaatcata	8460
agcgatggac	acaattcaag	ccacagagtg	aagacggaag	gatgcactgt	gctctagagc	8520
aacttctggg	gcagaatcac	agggtaggtt	tctgacttga	ggcggaagag	gccacgagga	8580
agggagttag	tttgtctgag	ctagaagcta	cggcccacct	cttggtagca	gacctgccc	8640
caagcatgct	ttgttaatca	tgtgggatct	gattttctc	taaatctatg	ttcaactctt	8700
aagaaaatgt	gaattctcac	attaaaaatt	agatatacgt	cttttggtgg	ggggggtgta	8760
aaaaatctc	agaatatg	atttctgggg	gccggagaga	tggctcagag	gttaagagaa	8820
ctgggtgctc	ttctagacat	tctgagttca	attcccagca	accacatggt	ggctcacaac	8880
catctgtaat	gcgacctggt	gccatcttct	gacatgcatg	gatacatgca	ggcagaaagc	8940
tgtatacata	gtaaattgat	aaatcttttt	ttaaaaagag	tatggattct	gccgggtgtt	9000
ggtggcgcac	gcctttaatc	ccagcactct	ggaggcagag	gcagggtgat	ctctgtgagt	9060
tcgagaccag	cctggcttat	aagagctagt	tccaggacag	cctccaaagc	cacagagaaa	9120
ccctgtctcg	aaaaaccaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	gtatggattc	9180
taagaaagcc	gtaacagctg	gagctgtgta	cggagttcag	cgtgggtact	gaagaacaga	9240
cattcatgat	gaaacacccc	aggattttta	cttagtatct	agtttccatt	gttgttttga	9300
gaccggctct	tatgctctcc	aggctggcct	caaactgctg	atcttcccgc	ctctacctct	9360
caagtctctg	gactacttgg	ctcataaagc	gtgttctc	gggtccctg	aagttatggt	9420
tgtacaaacc	gtgggggtca	atatactcac	ttgggcagag	agagaaggtc	tgaatcccag	9480
acaatgactg	catctcagga	cagttgggaa	gaggacaatg	gcagaaggac	ttagaaaaga	9540

tagactggag	ggtggaag	cagcaggaac	agagaaacaa	aacaggaagc	ttgctatcca	9600
gggccactct	ggagtcctgt	ggcaagatgg	aagcgggcta	ggggaataca	tttgtgtctac	9660
tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgat	caatgcctat	caatgttgaa	9720
ggggaaatat	gtataccaca	ttgattcttg	gagcaattct	cagtatctgg	cctagagaaa	9780
ggaatggccc	ctgcagaata	gacagagtga	atgggtgccct	ttatcatttg	ctaaagtga	9840
ggagaaataa	acatccttcc	atagagtttc	aggtaaatga	accccacagt	tcctctgtgc	9900
cgtggtggag	gcctggccaa	cagttaaaaa	gatttagacac	ggacaaagtc	tgaaggaaac	9960
acctcgaata	ggaagaggag	agccacctca	ttctgttaact	ttcctcaagg	ggaagatgtt	10020
ccaagagtgg	gaataaatgg	tcaaaggggg	gatttttaaat	taggaaaacg	atttcctgta	10080
tcacttgtga	aactggaggt	tgatttgggg	cataggacaa	tagatttgat	gctttgcaaa	10140
aagctgtttc	aaagcagaga	aatggaatag	agacaattat	gtagcgagga	gggaggggtg	10200
ggcggaagatg	gagacagaga	agtggaaagt	gacttttaggg	aagaggaaca	tagaccacag	10260
gggcgggggcg	gggggcaggg	gcggggggcg	gggctcaaa	gaggcagtgg	gaacgttgct	10320
agtgttcgca	gcgtaagcgt	gaatgtgcaa	gcgtctttgt	ggtgtgtgac	caggagtagc	10380
gtggctggct	tgtgtgctgc	ttgtaatccc	agtcttttag	gtttccacac	tgttccacag	10440
tgggtgtgat	tttccctcgg	agagcatgag	ggctctgctt	tccccacatc	ctccccagcg	10500
ttcgttggtg	tttgtttcca	agatgttagt	gggtgagaca	aagcctctct	gttgatttgc	10560
ctttaacagg	tgacaaaaaa	agctcaacca	ggagacattt	ttgccttctt	ggaaggtaat	10620
gtccccatgt	agagcaatgg	gacccatctc	taaggtagag	ctactcttgc	agtttgacc	10680
cagctcttct	gatgcaggaa	ggaagttggt	gggcaagcaa	gactgtttgc	ttcttgcat	10740
ggacacattc	tgacacaaaa	ggctcaggag	gggagaaggc	tgtttgatgt	ttagcactca	10800
ggaaggcccc	tgatgcatct	gtgattagct	gtctccatct	gtggagcaga	cacggactaa	10860
ctaaaaacca	gtgtttttaa	attgtcaagc	ctttaagggtg	aggaaattga	cttattgtgc	10920
tgggccatac	gtagagcaag	tgctctgcat	tgggcccaacc	cccggctctg	gtttctaggg	10980
accagaatgg	cctagaacta	actcacaatc	ctcccattcc	aggtctcagg	tgctagaatg	11040
aaccactata	ccagcctgcc	tgctgcctta	cctgccttcc	taaattttaa	atcatgggga	11100
gtaggggaga	atacacttat	cttagtttag	gtttctattg	ctgtgaagag	acaccatgag	11160
catggcaact	cttataaagg	aaaacattta	gttggggtggc	agtttcagag	gttttagtac	11220
attgtcatca	tggctgggaa	catgatggca	tgacagacaga	catggtgctg	gagaaagga	11280
tgagagctct	acatcttgca	ggcaacagga	cctcagctga	gacactggct	ggtaccctga	11340
gcataggaaa	cctcacagcc	caccctcaca	gtgacatatt	tccttcaaca	aagccatacc	11400
tcctaatagt	gccactccct	atgagatgac	agggccaat	acattcaaac	tgctataaca	11460
ctttaaagta	ttttattttt	attattgtaa	attatgtatg	tagctgggtg	gtggcagccg	11520
aggtgcacgc	ctttaatccc	agcacttggg	aggcagaggc	agatggatct	ctgtgagttc	11580
aagaccagcc	tggctctata	gagctagtgt	caagggaagga	tatacaaaga	acagttctag	11640
gatagccttc	aaagccacag	agaagtgcgt	tcttgaaaaa	caaaaattgt	gtctgggacg	11700
gtctctgctt	tgggttgctt	ccactcccc	agagctggac	tcttggtcaa	cactgaatca	11760
gctgcaaaat	aaactcctgg	attcctctct	tgtaacagga	gcccgaagtc	aggcgccccc	11820
ttgtcttctc	gcaggattgc	catagacttt	ttctgtgtgc	ccaccattcc	agactgaagt	11880
agagatggca	gtggcagaga	ctgggaaggc	tgcaacgaaa	acaggaagtt	attgcaccct	11940
gggaatagtc	tggaaatgaa	gcttcaaaac	ttgtctcatg	ttcagttgta	cacagactca	12000
ctcccagggt	gactcacacg	tgtaaatatt	cctgactatg	tctgcaactg	ttttatctga	12060
tgcttccctc	ccaaaatgcc	aagtgtacaa	ggtgagggaa	tcacccttgg	attcagagcc	12120
cagggtcgtc	ctccttaacc	tggacttgct	ttctccggc	agcctctgac	acccctcccc	12180
ccattttctc	tatcagaagg	tctgagcaga	gttggggcac	gctcatgtcc	tgatacactc	12240
cttgtcttcc	tgaagatcta	acttctgacc	cagaaagatg	gctaaggtgg	tgaagtgttt	12300
gacatgaaga	cttggcttta	agaactggag	caggggaaaa	aagtcggatg	tggcagcatg	12360
taccgaaat	cccagaactg	gggaggtaga	gacggatgag	tgcccggggc	tagctggctg	12420
ctcagccagc	ctagctgaat	tgccaaattc	caactcctat	tgaaaaacct	ttaccaaaca	12480
aacaaacaaa	caaataataa	caacaacaac	aacaacaaac	taccccatac	aaggtggggc	12540
gctcttggct	cttgaggaat	gactcaccca	aacccaaagc	ttgccacagc	tgttctctgg	12600
cctaaatggg	gtgggggtgg	ggcagagaca	gagacagaga	gagacatgac	ttcctgggct	12660
gggctgtgtg	ctctaggcca	ccaggaaact	tcctgtcttg	ctctctgtct	ggcacagcca	12720
gagcaccagc	acccagcagg	tgacacaccc	tcctctcgtg	cttcttgagc	aaacacaggt	12780
gccttgggtc	gtctattgaa	ccggagtaag	ttcttgacaga	tgtatgcatg	gaaacaacat	12840
tgtcctgggt	ttatttctac	tgttgtgata	aaaaccgggg	aactccagga	agcagctgag	12900
gcagaggcaa	atgcaaggaa	tgtgcctcc	tagcttgctc	cccatggctt	gccgggctctg	12960
ctttctgcaa	gcccttctct	ccccattggc	atgcctgaca	tgaacagcgt	ttgaaatgct	13020
ctcaaatgct	accttcaaa	aagcttctc	tgatcttgct	aactaaatca	gacctgtttt	13080
caccgtgcat	tatctttctg	ctgtctgtct	gtctgtctgt	ctgtctatct	gtctatcatc	13140
tatcaatcat	ctatctatct	atcttctatt	tatctacct	tcattcaatc	atctatcttc	13200

taactagtta	tcatttattt	atttgtttac	ttactttttt	tatttgagac	agtattttctc	13260
tgagtgcacag	ccttggtgt	cctggaaccc	attctgtaac	caggctgtcc	tcaaactcac	13320
agagatccaa	ctgcctctgc	ctctctggtg	ctgggggttaa	agacgtgcac	caccaacgcc	13380
ccgctctatc	atctatttat	gtacttatta	ttcagtcatt	atctatcctc	taactatcca	13440
tcatctgtct	atccatcatc	tatctatcta	tctatctatc	tatctatcta	tctatcatcc	13500
atctataatc	aattg					13515

<211> 14553

<212> DNA

<213> Mus musculus

<400> 6

(SEQ ID NO:10)

cttgaagaac	acatgttttc	caagagggag	cacccatggt	ggaatgacaa	tgtagttagt	60
gctcctctcc	tgtaggttag	tgctcctttg	ctataggtaa	gtgctcctct	cctataggtc	120
agtgtcctct	tctataggt	tagtgctcct	ctcctatagg	ttagtgctcc	tctcctacag	180
gttagtgctc	ctctgctcta	ggttagtcc	gctctcctat	agtacctaga	gagctagggc	240
aaatggcgta	ggcccgaagt	gcagagacaa	acagctatgg	aagactgggt	aagcacttcc	300
aagctacgaa	agagcagtgt	gaagggtcag	ggcttggtgca	gttagtaggg	gagatcttcc	360
agttgaagaa	acagaagaac	tgagagccac	tgggtatcat	cctcctgccc	catgccttcc	420
tggatactgc	catgctccca	ccttgatgat	aatggaatga	acctctgaac	ctgtaagcca	480
gccccaatga	aatattgttt	ttatgagagt	tgcttggtgc	atgctgtctg	ttcacagcag	540
taaaacccta	aataaggcag	aagttggtac	cagtattgct	gtgatagacc	tgaccatgct	600
ttcctttgaa	agaatgtgga	tttggtgact	ttggatttgc	aacacagtgg	aatgctttta	660
atggagatta	atgggtcatc	aattcctagt	aggaatatgg	aagactttgt	tgctgggagt	720
atttgaactg	tggtgacctg	gcctaagaga	tttcaaagga	gaagaatttc	agaatgtggc	780
ataaagacag	ttttgtggt	attttggtga	agaatgtggc	tactttttgc	ccttgctctga	840
aaagtctgcc	tgagactaaa	gtgaagagaa	tcagattaat	tgcatgaca	agggaaagttt	900
gtggctgccc	tatctggaaa	cttacagcca	gcctcttgga	cctcggtga	cttacgcaaa	960
tactcaggga	cagagatgct	tgactctgta	ctgatgagtt	gtcttggtg	caaatatggg	1020
ctcttcaattt	gactacatgt	cacgatgagt	caggagctgc	tctctccaga	gtgtgacaaa	1080
gcgaggggat	gctgacggta	gctgttctag	ccttgaaagt	aagcctgcac	ttatgctaaa	1140
gtcacacata	cacgagccgg	gtggagaacc	tgtctgtgtg	gagacacctt	tcattacctg	1200
tggcatccag	cctctcaagc	ttggactgcc	tgtgtgctcc	tggactctgg	aggtcccact	1260
gctctgtcct	ctgctgctta	tgatactgac	attttaaaaag	aatccagtgg	ttccccctg	1320
tactcgggtg	ctacttctac	ctggatgttc	ctcatttatg	ttctgtgaca	cctctctgtg	1380
actctgctgc	attcctgggt	gacatgtgga	caccctgtcc	ccttgccagc	catgatgtca	1440
ctgtcactag	tggaatcaga	tgccccaaagt	gttgctcctgt	gtttgggaac	gtgacaggca	1500
gtacagaagc	agaagaggaa	gggtgaaaac	ggaaatgtca	cagcagcatc	tgatgtgtgc	1560
ctcagtcacg	catgctgctg	attggaacta	ctcagcatga	gagagggcca	tggtgaatac	1620
acaaccctat	acacactgtg	tccattttctc	tctctctctt	acacagagag	agagggagga	1680
gggggagggg	gaggcgagg	gggaggggga	gggagagggg	gtgggagagg	gagagggaga	1740
gggagagggg	gaggagagg	gagagggaga	gggagagttt	aatgtctgtg	aagagatacc	1800
atgaccaaag	caactcttat	aaaggacaa	atttaattgg	ggctggctta	caggttcaga	1860
aattcagtc	attctcacca	tggtgggaag	catgcaggta	gatgtggtgc	tgagggaacc	1920
aagagttcta	tatcctgac	tgaaggcagc	caggagaaga	ctgcctcttc	tgacacaggc	1980
agagcttgag	catagaacat	caaagccctt	ccccacactt	cctccaacaa	ggtcatacat	2040
acttcaacaa	agacacacct	cctaaccgtg	ccactccctg	tggaaccaac	atttaaacgc	2100
atgagtcctat	gagggtcaaa	gctcttcaaa	caccacact	catgtacaca	cacacacaca	2160
cacacacaca	ctctcataca	cacacacaca	cacactcaca	cacacacaca	cacacacaca	2220
cacacacaca	ccacacacac	acacacacac	agagttctat	tttgactgtg	ttcactgtca	2280
caagggttcta	cttatctcag	acacactgcc	aggaattgtg	tggaagact	ttcagtttct	2340
ttgggttcac	atggacttag	cagttcttgg	tgactcctgaa	agatttctgc	agaaagaagc	2400
caaagtgttg	agcccaaggc	ctggccacac	attagtcctg	tctagatgaa	cagggtttta	2460
aaaataagg	ggacccaagg	tgaagccagc	aggggctgac	ttagagagga	gacccacca	2520
agccaactgc	tcgaagtcaa	aagcgatgaa	tccccatata	cagctgtgcc	cgggtgctgtc	2580
ttgctacatc	tttagtaaat	gttcttttag	ttgtatgcgt	atgaatattt	tgcttgcata	2640
tatttggtga	caccataggt	gttcctaggg	cctatggagg	ccagaagagg	gcatcagatc	2700
ctttggaact	ggaattatag	acacttggtta	cccatagagt	agattgtggg	aaatgagcct	2760

ttagtcttcg	agagcggcca	gtgctcttaa	cctttggctg	tttctccagg	tctttgagac	2820
tttattttct	tggacatcag	gacaggatcc	agggctttga	gcttgtttct	tcagccagct	2880
ttcttttcat	gtatattaaa	ttttatgtta	ttttgctttc	tttttcccca	agacagaatc	2940
acactctata	tagctcaggc	tgggtttgaa	ttcagtttcc	ctgtctcagt	ctaccgggta	3000
atatgattac	agatgtgagt	ctgactttgg	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
gcctccaga	cccagctctg	gtcccatctgt	gggcatctgt	gtctgtctct	gaggccgtaa	3480
gtgctgctcg	gttgtctcca	gcttgacttg	acactccctc	cttaataaga	gtaccacaga	3540
acagggctcg	cagagtcctt	gggccaggtc	cctgtgctgt	cctggaatgc	caggcgtgaa	3600
tttctgtgta	agtaggactt	tgctcgccaa	gctcccacgg	cttgcccttc	agatagccag	3660
aattatctgg	taccctgcat	tgccgttcaa	tacgcagagt	atcactggaa	gcgcgcgcgc	3720
gcacacacac	acacacacac	acacacacac	acacacacac	acacgcccac	tccatcttta	3780
aaccccaccc	cccagcaacg	gcggtgtaaa	cactctccat	caggaagctg	aaacgcagtt	3840
gccctctgct	ggggagatga	aggcagcttg	ctgggggcga	ggaccgtgct	agcaaccttc	3900
cctggtgcac	acgggctctg	gtgcatgacg	ggaacggaaa	cgcggaacta	aagtcagtc	3960
tgcttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	ggcgttggtg	4020
gtggactgag	tgacaatcag	tgaatcact	taggttgttt	ttctcttctt	cgttgggttt	4080
gatacagcgt	gggagagggt	cagaggagaa	ggggagggat	gggagagag	ggaggaggga	4140
ggggcgggag	cggggggcgc	aggaaaacgt	gctaacttct	ccaatcctac	aagacaaagg	4200
tttgagaaaa	gcccactga	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
tctcaactcg	ggggtcacga	tccccactt	ttctgcatat	cagacatttt	tacgttgcta	4500
ttcataacag	tagcaaaatt	gcagctatga	agtaaccaatg	aaatgcattt	atggcgctg	4560
tgtgtgtgtg	tgggggggta	tcaccttaac	atttactgta	agaaggttga	gaatactgct	4620
ccagcagcta	gtgtgttgga	cttaggttct	gggtatatta	ttagcaatag	ccaaccagaa	4680
tccccaccca	ccacagcatt	gaggccccat	gcagggcttg	ctgggagagg	cactgataag	4740
acttctttat	gtattttatt	agagacgaat	actcattagg	tagggccaagc	tagcgtcaaa	4800
ctcatggcaa	ttctcctcct	ccagtttctt	aagtaactga	ctcaggagtg	tgttgccatc	4860
atatacagta	aggatttatt	gactgaagaa	aatctcaagt	ggctttgggt	aatccctact	4920
acgccagagg	ctgaggcagg	aggcgcgcaa	ggtcaaggct	tgccctgggct	acatatagag	4980
tgagctcaat	tttgacactt	ggtgcggtgt	tagtagtaat	agtaaagatg	aagggtgtggc	5040
tcagggtggg	ccggtgattg	gacacacttg	gggtctcctg	gtccatctgc	agctgtgcaa	5100
caggaagagc	ggagaatgag	aggaaaagaa	gaaaagacag	aatgagagag	agggaggaag	5160
agagaaaaag	gaaaagagag	aggaaaaggaa	aaaggaaaat	gaggaaaagc	agaaaagaag	5220
aatgaaaaag	gaaaagagg	gaaaagaaatg	agagagagaa	agaaaagac	agaatgcgag	5280
agagggagga	agagagaaaa	aggaaaagag	agaggaaaag	aaaaaggaag	atgaggaaaag	5340
cgagaaaagaa	gaaatgagaa	agaggaaaag	gagaaaagaa	tgagagagag	aaaagaaaag	5400
acagaatgcg	agagagggag	gaagagagaa	aaaggaaaaag	agagaggaag	ggaaaaagga	5460
aatgaggaa	agcgagaaaag	aagaaatgag	aaagaggaaa	gggagaaaag	aatgagagag	5520
agaaaagaaa	agacagaatg	cgagagaggg	aggaagagag	aaaaaggaag	agagagagga	5580
agggaaaaag	gaaaatgagg	aaagcgagaa	agaagaaatg	agaaagagga	aaggagagaa	5640
gaaatgagag	agagaaaaaga	aaagacagaa	tgcgagagag	ggaggaagag	agaaaaagga	5700
aaagagagag	gaagggaaga	tggaaaatga	ggaaagcgag	aaagaagaaa	tgagaaagag	5760
gaaaagggaga	aagaaatgag	cgagataaaa	gacagaattt	gagagagggg	ggaagaaata	5820
ggaaaagaga	ggaaaggatg	gagaaaagag	agaaaagaaag	agagatgaaa	gagagaaaag	5880
agaaatgaaa	tgagagagag	agagagacac	aaagagccag	agagagaaga	aaaaagggga	5940
aagagaaaaga	gaaagaggaa	ggctcctctt	ggacacatct	tcctttatct	ttccctgggg	6000
accgccaaag	cctggtggca	tactgtacat	tctgtacact	gttcattcaa	aacaggctct	6060
gtcttaagaa	tggtctgagc	ggtcagaaaa	gggtattgtt	aacttggttg	caaaactgcc	6120
tcaggagagt	gctgagtcgc	tgaagtttgc	tgcccgttaa	ggagaagtct	ctactacttg	6180
tgatctcacc	atcgaaaaatt	tctttaattg	ttccttggtg	ttctgggttt	tgagtttttg	6240
ttcttaagga	tacattcttg	ggtgatgtca	caaagtcctc	aaagacacgg	tgagagctgtg	6300
ttagatgggg	aaagacagtc	tgctgaggat	ttatctggaa	ctgtcagaag	gaaaagaagg	6360
taaatggggc	acttgggaaa	gtggcctcta	gtttgacttc	tggttagca	aaggttgtgg	6420

ggagataaagg	catacacagt	agttagcagg	aggcaacagg	gtcctggggag	gacgcgaggc	6480
agaaggagag	gctgggctga	cagcatgcaa	tcattgcata	gtctccaaag	gagattgcaa	6540
catggctgag	ttttcagagg	tcctacagag	cccgtggtag	agattctgtg	ggttctgaga	6600
caacttgact	ttagccagat	ggtatttgag	taatctggga	gagagaaaac	agctacagca	6660
aacaggggcca	catttagtga	cgaactctc	actttgactg	ttgagtcatt	tgcagtgggc	6720
cctgaggtca	ggctggccct	cagctcaaaa	acaagcgagg	aactgaagca	attactcaga	6780
taatccacag	ccacagccac	tggaaagggc	cacatcccca	gagacagcac	agcaggggtg	6840
ggggtggggc	tatgagaaag	ttagtgattg	tagcagttat	ctagaatgtg	cggagcagag	6900
gaggttacac	aaaaacctag	aatgtcattc	aatgtgggaa	accgagaggc	tcccaagccc	6960
taaaaggaac	agtttgcttt	cagccaaaat	ggaaataaaa	tttggggctt	aaatctggca	7020
aatgattcag	accttctgtg	taggtgtctt	taaatgcaca	gcagattgat	tttcatgttg	7080
gagttttattt	gaactaaaa	acagaaatgg	tgaaaagcac	acctgaagaa	attgagatgc	7140
tatgaataaaa	atcatttact	tacagctatc	acttaattag	tacctccttc	caccttgctg	7200
atttattggg	ctagtcaagg	aagaaaagat	cttccctcct	ccttctctcc	tcctccccct	7260
cctctcctcc	tcccctcccc	tccttgacct	tcctctcctc	cttttccctc	ctccccctct	7320
tcttctcttc	acccccctcct	cccctcccct	cctctgtact	cctccccctt	cctcccaatc	7380
tcttttttct	cccccttctt	ctctttctct	cccctcctct	tccctcctct	tcctccctcc	7440
ctccctcctc	ctcctcatcc	tcctctcctc	cttcatcctc	ttctccttcc	tcctctcctc	7500
cctcctcctt	ttccagccct	acctaccttc	cctttcttct	tcatttatcc	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
tgggcttagag	gactcttaaa	tggcactctt	atttgtttaa	taaaggagg	tgtgacgtgt	7860
agttttaaat	cccttccaca	caacaattgc	tactctctga	ccaaaaaaga	agggagacag	7920
gatacggcta	ggtgtctagt	agactttacc	actttgaaaa	gccttaatat	aaatcaggta	7980
gatacatctt	tttaacttat	tcttgtaaa	acaaaaacaa	aactttatct	ttatttgtgt	8040
gtatgcttgt	gtgtgtgtgc	ctgtgtgtat	accacatgtc	gctggtgccg	gagaacacca	8100
gaagagggga	cctgatctcc	tggagctaaa	gctatccatg	gttctgagct	gcctgatgtg	8160
ggtgctggga	acgaactct	ggtcttctgc	aagagcaaca	agcctcctct	taactacgaa	8220
tctcctcccc	atccccccaa	atacatttaa	ttattcattt	tagcagcttt	atttcgtaac	8280
tacttatcac	agcataaaac	aaggatttta	tatatattac	atgcaatcga	ggataagagt	8340
tgaagggaga	tgcgtgtgct	ccttctgggt	gtctgtgctt	ttgaagaatg	taagcagtgc	8400
acaagggacc	gaggcgtgcc	tgtctgccag	gagctgtctt	cttcccttgg	actctgagct	8460
gagtgacgtg	ctccgaagaa	gtaaaagacg	acctcatgaa	gcaatgtctt	caacccaaac	8520
atgctgtcca	gacaaagtcc	agcttcatta	gtatgctgag	gagagactta	ctgagcctca	8580
ggaaagcccc	cctcagcatg	gcgaaagtcc	actttgattg	aagtgactcg	aaagccatgg	8640
cagtgcggcg	gcggccgcgt	ggagcttggt	ctcgagtcgg	aagcggcatc	tttgtcaggc	8700
ggctgtgatt	agcacgggga	ggcaggactg	gagtgaaagg	agagttgggg	gcggggctta	8760
gcgctctggg	ctcctaagct	gtagtcagcg	cctcaagatt	tgtaacctgc	cttctgcctt	8820
ccagcccgag	cagtcaagtg	gctccaagtc	gaagactgca	aagtgcacct	aaccttttgg	8880
ttatagcgag	gctgaagaca	ccgtgctctt	tcataaaaag	cggatgtctg	aaatccgatt	8940
tgataaaatat	ggataaaaacg	tataacgctc	gatcaatcga	atcgaaggag	ctcacgattg	9000
gcaccacggc	tttggggaca	acagagtact	gactcgttgg	gaggacttgg	atacttcccc	9060
tcctcttcca	tctcttcccc	tttccctcact	tcctcctcct	tccttctcca	ttttctccct	9120
cttcaactgtt	tcttactatt	tttacaaaag	attttattta	tttatttatt	tattttattta	9180
tttattttatt	tattttattta	tttattttaat	gtatgctgag	acactgtagc	tgtcttcaga	9240
cacaccagaa	gagggcgtca	agttccatta	gagatgggtt	cgagccacca	tgtgggtgct	9300
ggggcctctg	gaaggaccgc	cagtgtctct	aacccctgag	ccatttctcc	agtacccttc	9360
tcaccgtttc	tcttcaatct	tcttccctct	ccttctccac	tttccctgtc	ttcttggttt	9420
cattatcttt	ctccctttct	tcctcttctc	cccttcttcc	tcctccactg	tagttttcct	9480
tcctactctt	tttccctgct	ccctcctcct	ccctctcat	tccccctcct	ctttcctcct	9540
tctccctcct	cctccttctc	tctccctcct	ccctctcccc	tctcccttct	cccttctccc	9600
cctcctcttc	ctctttctcc	ttctccaccc	ctcctgtcac	agtatcaatg	gcaaggggtg	9660
tctagaatgg	aggagtgtcc	cctaggcact	aacgaaagcc	agttaggatg	ctctgagacg	9720
ggtacaattc	agggaggggc	gtggggatgg	aagggttggt	ctgcgattca	ttctggagca	9780
acccccaggc	agaatcatga	ggttggttcc	ggattcgcag	ggcacaattc	agaagaggaa	9840
ggtttcagga	aggacgagtt	tgtctgagat	aggagttaca	tctgatgtct	tggcagcaga	9900
gcactgttac	aagcgtgctt	tattaaaccac	gtgggattaa	aatctctttt	aaatttattt	9960
tcaactctta	aggaaaactg	aactttcaca	ttcaaattta	gacttgcagc	tcttatgggg	10020
aaaaaaaggg	gatcttaaga	atattaagca	taggcggctg	gagagatggc	tcagcgggta	10080

agagcactct	ctgctctccc	agaggctcctg	agttcaattc	ctagcaacca	cataatagtt	10140
aacaacagtc	tttaatgaat	tctaatagccc	tcttctgggtg	tgtctgaaga	cagttacagt	10200
gtactcatat	aaataaaaata	aagaaatttta	aaaaaatgaa	tattaggcat	agattcctgg	10260
atcctaagaa	agccatcaga	gctggagcca	tgtgtgggat	cctgcttggg	gctggagggg	10320
cagagttcat	gcccccgggg	tttttactta	ttatcacatt	ttcatcggtg	ttttgaaaca	10380
gggtcttgtg	tggccagggc	tggccttgaa	ctcatctttc	agcctctacc	toacaggttc	10440
tgggattact	tggttcctaa	aagtatctcc	gtcaagctcc	ctgggtgttat	ggctgtgcca	10500
accaggaggg	tctatacact	cgctcaggta	gagggagaag	atccgaatct	ctgacaggga	10560
ctgctgcctc	tcggggcaaa	tggagtgaag	gacagcggca	gaaggattta	ggaaagatgg	10620
acgggagagt	ggaaatgctg	cagaagccag	aaaacaaagc	aggaagcctg	ctgtccagtg	10680
gggctcaaga	gcggagggat	gcgagggggc	tgcgcaggaa	catttagcgt	ctgcgtctat	10740
gggggtaggg	gcgggggtgcc	agcacctagt	cacctgaagg	ggaaatgctt	gcccagggag	10800
cagggtctcag	tagctgacct	agagaaagga	gcggccccta	cagaggagac	acgggtcact	10860
gtttgttaaa	gtgaaggaga	aataaatatt	ctttcaaaga	atcttaggtg	agcccagttc	10920
atctgcgctg	tggaggcctg	gggaacagtt	aaaagaccc	tgacacacac	ccaaggcaaa	10980
caagcaacac	acggctcctt	ccgtaagggt	ccatgattct	ctgaagaatc	agccccggaa	11040
tcagccccgg	aatcaggtag	tccgtaaaaca	caatgagtg	tttactctgc	agaagtcag	11100
cctgctggcg	tctccatta	ccaaaataga	ggatagtc	cgtagctca	ccggctcgat	11160
ttaaggcacg	tggttttcca	gggtagatga	gctttggctt	ctggaaccat	tatggggcac	11220
gaaggatgga	gccaggattt	tttttttttt	tttttttttc	tattagcaat	tgatttgctt	11280
gggcttggct	ggacttgccc	agttcttagg	cccagtcctc	ttaactgccg	atctgaagtc	11340
tgtcatggag	tcagcctagc	cttctcactt	cccttcagct	cgaataggaa	gaggaggtgc	11400
acaccagatg	gtctgagagc	agggataaat	gggtgacctt	tgtctttcag	tatttcgtta	11460
ttttaagtag	gaagatgctt	ttctgtatta	cattgcttgt	gaaaccggaa	gttgattcgg	11520
ggcacaggac	aatggatttg	gtgttttgca	aggactgttt	cagaagagag	aggagtggaa	11580
gggtgggttag	agtgaggagt	gggggtgggac	gggatggggg	aagagaagga	agggccagac	11640
aggctaggta	gggctgagag	gaggcgggtg	gaacttcttg	agttagcgca	gcagtaaaact	11700
tggatgtgcg	tgtatctttg	tgatatatga	cccggagccg	tgtagctggc	tccgatagta	11760
ctgctaattg	cagtgctggg	gggggggggtg	cccatactgt	tccacagggg	ctgcacattc	11820
ccatcgagag	caggagggct	cctctctcca	tacatctcgc	ccagcattcc	ttgttttttc	11880
tgtgatgaca	gggggtggga	tgaatctct	ctgttggttt	gagagaccgt	gaagaagctc	11940
aaccccagga	cattttgcag	tcttggaagg	cagtgccctc	atgtggagcc	gtggagccca	12000
tctctgagtc	caggtcactc	ttgcagttcg	cactcagctc	ttcagatgca	ggagagacgt	12060
tgggtgggaaa	gcaagattgt	ttgcttgttg	agatagacac	attctccaca	caaaggctca	12120
cgtggggc	aggctgattg	acgtacagcg	ttcaggaacg	cctgtggtag	agctatgatt	12180
agctgtctcc	atctatgaag	cagacaaaaga	gttataaaaa	aaatcaatgt	tttcaaattg	12240
tcaaactttt	aacccgacag	caagcgctct	gtccctgggc	taatccctag	ccctgggtttc	12300
ttgagatggg	gtcttttgtg	cactagactg	gcctagaact	cacgatctta	gtgttccagc	12360
ctcccagctg	ctgggatgag	ccgctataac	cagctgcctc	gccttcctaa	atttttaagt	12420
atgggaagtg	ggggagaata	cagtttaaa	tatgcagatc	tgagagcagg	aacctggcaa	12480
agccaaaggg	cgggagttac	aggcggctaa	catgggtgct	gggaactgac	ccaggtcctt	12540
gagaggagca	gtgtgtactc	ttgaccaaac	aggtccgtct	ctccagtcctc	cgtagtatta	12600
aaaataggta	ctacgggcat	ggtggtgcac	acctttaatc	ccagcactag	ggaggcagag	12660
gcaggtggat	ttctgagttt	gaggccagcc	tggctctaaa	aatgagttcc	aggacagcca	12720
cggctataca	gagaaaccct	gtcttgaaaa	caaaacaaca	acaaaatagg	tactacaaag	12780
cgatgtaatt	gtgctcaaac	atgcaaaccg	aggggactgt	atgcataaga	aagagaaaga	12840
cggccacact	ggttctatct	gggtgacagg	aaatcagtat	ttttattttt	cacattcatt	12900
tttttgttgt	tgttgttgac	acagtgtttt	ttctatcaaa	aacattattt	cttttatagt	12960
tcccctgagg	agctgttttt	aaagccgtgc	tttgaaaaac	cattgaagga	gcagaggcag	13020
ggagactcct	gtgtggcagt	cgggtgaagca	ggcctctctg	aggcaggctg	gccctggact	13080
tgggagttct	tttccctccc	tcctgtgctc	aaatagcaaa	tgtcaggctt	caatgtagct	13140
agaaggttct	agaatgatta	agtttccaag	gctgaagagc	ttccctgttt	gcctttcact	13200
tccttgagga	ggtcgttgtg	tgttccggag	tctgcaaggt	gcctttgggtg	atgogggtgg	13260
ttcatctcgg	gagattccgc	ctggaggacc	caagttcaag	ccctgcctga	gctacagagt	13320
gactttcagg	tcttctgcgc	aattcagtga	gacccagtct	acaaataaaa	agtaaaaaga	13380
aggctgtgga	tggaaactcg	tggtagagtt	ctgggtttac	tccttagagg	aggggagaag	13440
gaggaggagg	gaggaggaag	aggaagaaag	aagaagagaa	gggaagagga	gaagggaagg	13500
aggggaaggg	ctgacaagaa	gagagaagag	ggagggaagg	gagggaaagg	aaggggaaag	13560
gaaggaggag	aaggggctga	caagaagaga	gaaggggag	ggaggggagg	gaaggaagg	13620
ggaaagaaga	gaagggttaag	aagaaactgt	tccaatgggtc	tgggccacag	agtgatggcc	13680
ttttgtgggtg	atcagctgta	atccttgatt	tgacacaacc	tagaatctgg	gaagcgagtt	13740

tctgtgaagg	agcattcaca	ctggctggcc	tgtgggcgtg	catgtgggag	actgtcataa	13800
ttaggttcat	taatacagga	agtcccagcc	cactacaaat	ggcttcgttc	catacccaag	13860
agatgctaac	tgtagacggt	tggagaaagc	aagcaagctg	tggatacccc	acgctctttc	13920
acctcggtc	ctggggggtg	ggtgcactgt	gtctcttggt	attttaaagt	cctgccttga	13980
cgccccgtc	gtgacagact	gtaactggaa	ttgtgagctt	tagtccttta	gttttctacg	14040
ttgggttttc	tcaggatatt	ttatcgagc	aacagaaaca	agaccaggac	acttgatctc	14100
ctctgatcaa	caactgaagag	ttacaaaaca	ggctgaggaa	acaaactttc	ttctccctct	14160
cccccttctg	tccctccctt	tccctctctc	tccctccctt	gccccctctc	tccctgtctc	14220
tgtctctgtc	tctgtctctg	tctctgtctc	tgtctctgce	tctccccctc	cctccccctc	14280
ctctgtctct	gtctctgtct	ctgtctctgt	ctctgtctct	gtctctgtcc	ctttctcctc	14340
tatctcctaa	atggctggag	gccatgctag	ctcaatgttg	aactttgaac	acgtatttag	14400
gaaatccttg	ttcttaacag	ttctgaagtg	ctgaagtggg	ggtttagtct	ctcggcctga	14460
caagctcact	tcctctcact	ctgtcttaat	gaccaaatct	gccatttccc	taaaacagca	14520
caggctccag	ctccaggttg	ctccggagcg	gag			14553

EXAMPLE 12 – CHO Stable Site 2 Sequences – U.S. Patent No. 9,816,110

<211> 4001

<212> DNA

<213> Cricetulus griseus

<400> 1

(SEQ ID NO:11)

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	tacctatgga	tgtgagataa	240
ttggtaacaa	cagcacatgg	gagagcatgg	gatcattcaa	ggaagattag	agagaatgca	300
tttttttagga	gataatggag	gagcaataga	aaggattaaa	tgaggttact	gatgaaagtg	360
atggtttagag	aaggcaatat	gaggagggat	aactagcact	tagggccttt	tgaaaaagac	420
atagagaaaa	tactattgta	gaaacttcct	ataattgggtg	tatagttata	tacaccaaaag	480
agctcagatg	gagttaccct	ataatggaaa	tattaactac	tttttatcac	tgtgataaaa	540
catcctgaac	agagcaacat	agattgggaa	gcatttactt	tggcttacag	ttctaacggg	600
ataaaaaattc	atgatgaaag	aatgaatatg	tcagcaaaaca	gcagtagcaa	tggcctgaga	660
agcaggtgag	agctcacatc	ttgaagtgtg	agaatgtagc	agagagaaca	aactgcacaa	720
gaccagaaaa	tgtttttgga	tcagagccca	tacctctctg	actgacttct	ccagaaattc	780
tgaacaaata	aaactcccca	aacagagcca	taactgaagg	tccagtgtct	gagactacta	840
ggggtatttc	ttattcaaac	cactacaatg	gggtgggggg	agcaatcctc	caagtaggca	900
ctacacacag	acaaataaaa	actctagtta	ctggaatgga	ttgacttatt	tgaattactt	960
gccagtggag	ctacatagag	cacaattatt	gtattttaa	taccttttat	gatcttacaa	1020
aacttgacag	taagatcata	ttgctaaga	aaccacatat	ttgaatcagg	gaacatggg	1080
atatctagtt	gttcttcaac	tggaaacttc	atgctttctg	cccagcattc	atgttgctgg	1140
aaagagcaat	gtacactacc	agtgtagaaa	ttaaatcatc	aatcttatca	agatgtggat	1200
cctataagtt	acaataaaaa	ttagcctgat	aagatatccc	caccagaaga	atattccat	1260
aaatgctatg	ggagcaacaa	gctattttct	aaattagcct	taatcctatt	ctacaagaga	1320
gaatccatat	ctagaatagt	tatagggatc	aaagaacccat	ggcttgattg	gtcataggcc	1380
caatgggaga	tcctaataat	attgttctac	aaaatgaaaa	taactcctaa	tgacttggtg	1440
ctgcagtaat	aagttagtat	gttgctcaac	tctcacaaga	gaagttttgt	cttacaataa	1500
atgggaatta	aagcagcccc	acaagattta	tatcataccg	atctcctcat	ggcctatgca	1560
tctagaagct	aggaaacaaa	gaggacccta	agagagacat	acatggtccc	cctggagaag	1620
gggaaggggg	caagacctcc	aaagctaatt	gggagcatgg	gggaggggag	agggagttag	1680
aagaaaagaga	aggggataaa	agggaggaga	ggaggacaag	agagagaagg	aagatctagt	1740
caagagaaga	ctagggagag	caagaaaaga	gataccatag	tagagggagc	cttgatgtt	1800
taaatagaaa	actggcacta	gggaattgtc	caaagatcca	caaggtccaa	ctaataatct	1860
aagcaatagt	cgagaggcta	ccttaaaagc	ctttctctga	taatgagatt	gatgactacc	1920
ttatatacca	tcctagagcc	ttcatccagt	agctgatgga	agcagaagca	gacatctaca	1980

gctaaacact	gagctagttg	cagacagggg	ggagtgatga	gcaaagtcaa	gaccaggctg	2040
gagaaacaca	cagaaacagc	agacctgaaa	aaaatgttgc	acatggaccc	cagactgata	2100
gctgggagtc	cagcatagga	cttttctaga	aaccctgaat	gaggatatca	gtttggagggt	2160
ctgggttaatc	tatggggaca	ctggtagtgg	atcaatat	atccctagtt	catgactgga	2220
atttgggtac	ccattccaca	tggaggaatt	ctctgtcagc	ctagacacat	gggggagggt	2280
ctaggtcctg	ctocaaataa	tgtgttagac	tttgaagaac	tcccttgaga	agactcacc	2340
tccctgggga	gcagaaaggg	gatgggatga	gggttggtga	gggacaggag	aggaggggag	2400
ggtgagggaa	ctgggattga	caagtaaatg	atgcttgttt	ctaatttaaa	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	atctctcca	gaattaaagc	caaagacctt	atgaaagtaa	2700
ttatcttcaa	aagggtgaaa	atacatactt	tgcaatacac	agatctgcct	agaaatctca	2760
tgttcacaat	acacatgatg	ctcaattgaa	ttccattcaa	tgttacagtt	tagataaaca	2820
gtttgtagat	aaactcacaa	tgtatcat	ctttttattt	tttgaccaaa	cagcttctca	2880
tctgttattc	agaataattc	ctcgatggca	ggatatccat	cccaattggg	ggaaggggag	2940
aatttgaaga	aaacctagac	cacatacata	tttgccattg	ggaaacaaag	tctaaaatga	3000
tgttggtcac	atcttctcta	ctagtcctct	ccccgtcca	aagaaccttg	gtatatgtgc	3060
ctcattttac	agagagagga	aagcaggaac	tgagcatccc	ttacttgcca	tcctcaaccc	3120
aaaatttgca	tcattgctca	gctctgccct	tctcatatga	cagttacaag	tcaaggcttc	3180
caaagtcctc	ctgtcatgtt	tgggtgtcaat	agtttataca	gatgacttca	tgtcttcata	3240
tctaattgtct	tatatagatt	aatattaaac	aatgttattt	ctctaaccac	attttaaatt	3300
aatttataaaa	tccatttaatt	gtgtctataa	aatgcagaca	gagtgtctgag	acacaatata	3360
agcctgatga	cttgaatttg	aaactcacac	ccaccacatg	gagaatcaac	ttccaaaaat	3420
tttccattta	cttccacact	tacaccattg	tacaaacaca	ataataatga	acaaaatgaa	3480
atgaaataaaa	aaattaagtc	tctgtaggta	atgctactgt	gcagcaaaag	taaaaatggc	3540
agcttaagct	tgttttatgg	ttacacttta	ccatcttcca	ttaattataa	ggacttcaat	3600
catggcagaa	ctatgctgtt	attgtctcag	tgtaacctaa	ccaggtgttc	cagatgttct	3660
taatgtggac	acctaacta	tttgatattt	gggttaagat	ctttccctct	ttcagaagaa	3720
acctcaggac	agagggaatc	ttgtctttta	attttgagtc	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:12)

catgtacact	tatgcaagta	tgatatggcc	caacacagta	ttttacacca	atttttatct	60
ataaaatata	catgtacatc	aaaatatatt	attaataata	acatcattat	tctttctttc	120
caagtaataa	acacatacac	tgaaattttg	gttcttgttg	ataattttta	tgaaacagga	180
aatgcaaaat	tatcttagca	tgtttacttc	actttctttg	catagataac	cagtaatcac	240
attgatggat	catgtagtga	aatgtatttt	taggtatcta	aggaattttg	gcttcgtttt	300
gtgcttggtg	acactgaatt	ctattcctaa	caacagtgtg	taaggattct	gtctgatttc	360
ttttaccagt	atgtgtccat	ttgcattttc	tttattattc	atggctgctg	ttctagaaag	420
tggaaaggtg	tgtgtcaagt	ctgttttaaca	tgtttccctg	atgatcagtg	tcttaacacc	480
tctctgagta	catgttggcc	aatgtcgttt	ctagaccoat	ctattcttgc	ttgacttata	540
ctggtagatg	cctgccaaaga	aatttctcct	catectttct	gtctcttcac	tgatttactt	600
gatgtgtgga	tttcacattg	atcatatgga	aatagaagat	acaattttct	ttattcacag	660

tttgaagac	tttcaatctc	atagatcatc	attatntttt	gctactgttc	cctatgctat	720
ggtgaaat	ccatttgaat	aattgcttaa	acaattaaca	agaaagaatc	tatttttact	780
tgcaataact	tccatttcag	aacatttact	acactgttac	tatatccaaa	aactagtttt	840
atatatcatg	tgagaaatga	ctaattcaca	atttggccat	gacatttttt	tcagaaacag	900
aaaaagtgc	caatacacatc	acaatgctat	aaatattaag	acttcagcaa	attaaatatt	960
tattcatgat	atcacataaa	attcattttat	tatgtttttat	ttaaatgtgt	ttttaaaaca	1020
gtggtatcac	taaatattaa	gttagatgtg	tttatgtgct	taatgaattt	atattttaga	1080
atgttataag	ttgtatatag	tcaaataatgt	aataaatttt	atttttttagg	tctttctcat	1140
taaggatatt	taattttggg	tcccttttcc	agagtgactc	tagctcatga	tgagttgaca	1200
taaaaactaa	acagtacaaa	atgtacattg	cattcagtat	tgcaattgat	ctttgcactg	1260
aagttttagt	cagttcatalc	atttagtact	tggaagtagc	attaagctaa	ctttcattgc	1320
tctggcaaaa	tgctcgataa	gataagagtc	tattgtggaa	agccatggca	gcaggaaagt	1380
aagactgctg	atgatgttta	atccatagtc	aagacgcaga	aggagatgaa	tgctggtatc	1440
caacattttt	tgctgttcat	tttctctaga	accctagtcc	ataaagatgt	atgacttgca	1500
ttcaaaatgc	gtcccttcca	gttggtccac	ttttctgtaa	atatcccttc	aggcatgtct	1560
agaagattgt	ttcgcaataa	cttctcaatc	cattcaagtt	gatagtgcag	attaatcact	1620
gcagaataaa	agcctgtaac	ttggctcacg	tgccaaggaa	tatgcacact	cctgacacat	1680
caataagtaa	atcaaagtgt	agcttttggc	tttaacattg	ccagacttat	gtaatgttct	1740
gcacgttctt	cctccatcac	tttttattct	aatggtgttt	ccttgacatt	gaatcacgct	1800
gtggaagctg	cttagaatta	acattgaaat	ctactgatat	atttatgatg	cagcaattta	1860
gatttactat	tttacttaga	attttttata	attgagagaa	tataatattt	tcacagttat	1920
ctatctgctg	taaatagagg	attttaaaaa	aaatctctat	aacttttttt	tacaacacac	1980
agtaaaatta	agttaaaaat	taataaagtc	actatgttga	tttcaaagtg	tgctacgccc	2040
acggtggtca	cgagggtgta	gcagaagatg	ccactaaggt	gggctaaggc	cgatgggttg	2100
gggtctgctg	tccctggaga	tgagccccag	gggttccct	ggcaatcagc	tgcatcatg	2160
atgcccgatg	agccannnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	2220
nnnnnnnnnn	nnnnnnnnnc	tggtgacttt	tatggaaaga	atttgataga	tttcatgatg	2280
tagaagaatt	ttattaggct	tattttacag	gagactaaga	ccctgggacc	taaagatata	2340
tggttctga	gaatcaggaa	atgggtagag	acgtggttga	tggtatgaga	cagatttttag	2400
agaactctta	gatcatgggc	aatgaccgca	atctgatgct	tagaatagat	catctataaa	2460
caattatgct	gttctttttc	tttctgttgt	atgatctgat	gatgtagccc	ccttgccaag	2520
ttccctgatc	ccccttgcca	agttccctga	ttgtaacagt	atataagcat	tgcttgagag	2580
catattcaac	tacattgagt	gtgtctgtct	gtcatttccct	cgccgattcc	tgattttctcc	2640
ttgagccttt	tcccttggtc	tccctcggtc	gggtggtctcc	acgagaggcg	gtccgtggca	2700
aaagtgtata	aatgttctaa	aacatttgaa	ctctaaaaaca	tgcaaaatga	aaaattaaaa	2760
taaaataaaca	tgaaaattaa	aatatattag	ctgtaaaaag	ttaaaacaata	ctatataata	2820
ttttgttatt	agaattcaaa	atcacattag	ttggatttaa	tttgaacatt	gcattctttc	2880
aataataaatt	tcaataaaaa	aagtttcccc	atgatagtag	aaaataataa	catatgtatc	2940
tatctatttta	tttaactaca	catatatagc	atttggttca	actaaaataa	atgaatgagc	3000
aaagcaccta	agtaattggg	gtctattata	tttatgaagc	caatagtttc	aaataaatta	3060
tcatgcatata	ggagggtattg	caaagtgtta	accttttttg	aaacagatat	tcccagttac	3120
agaaaattata	atttctaata	tttctataaa	gtagaatgat	gataattaat	ataggccatt	3180
tgtaaaataat	gttcagatta	aaatattctc	tatttcaacta	gagaagaatg	atattaaatg	3240
tatttatattt	tatttcccat	tttgtttgca	ccactattct	atatccctca	gcagtttaaa	3300
tttgtttcac	catatgtgtg	tgtgtttgta	tcttaaatat	ggcactaaaa	ttagaataat	3360
ttaatataaaa	tctttaggag	aaaagatatt	gaattatttt	atggtgatag	gaaaaatatc	3420
tttaattgtc	caagaataact	ttttcttcta	ttttaggact	gatcagaccc	aggactaata	3480
ttttatatgt	actaattcta	tgtacaaaaa	tatgttatta	tctcatgaat	tctgtctcaa	3540
tattgaggta	ataaaaaatag	tccatcatga	actttaaaaa	taaaataatg	attaattaat	3600
ttttattcat	attttgtttg	tatgaatggg	tatacatcac	atgtgtgcct	ggtagctgtg	3660
aatgtcagga	gaagggtatga	aagccactgg	aattggaata	agagataata	tttgagatgt	3720
tatgtgggtg	ctgagaatta	gacgcaagcc	atcttcaaga	atagccagca	tactatacca	3780
ctgagtaatc	cattcatccc	tcaataatta	tctttgtaga	cagtaaataat	atttctaatac	3840
tataaatgac	cagaaaaatt	aatgtattat	taatgaagac	attcatctca	tgtgacacac	3900
ttcacctgtc	taaatcagta	acactctctc	cactaattaa	gattttctaa	gtgcatgaca	3960
cttactatatt	ctaaagctgt	ccaatggggg	ccagtcoccca	gtcagcaccc	agtgagataa	4020
tccatgaatg	catttatatc	ttaggaaaaa	ttcttatcta	tgtagtattt	agaacatttt	4080
catgtgaggg	gataaacaag	gaagcacaga	tgctttctga	tagaaacttt	ctctttaatt	4140
catctagaaa	caaaaaaacct	ctcaggaaaa	tctctcttgc	tctctcccca	atgctctatt	4200
cagcatcttc	tccctactta	attctagatc	ttttctctca	tgccctcctg	ctgctgccct	4260
gctggctctg	ctctatgcct	ccccatgtca	ctttcttttg	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	tggttgaaat	ttgaggttgg	gtagtacatg	aagactgaat	4560
tttcttcaaa	ctctggcctt	gaaatagtaa	aacaacacct	atgaaaaatga	cgacctgtat	4620
ttgtcttttag	aggcaaccac	atattgtctg	cagggcctgc	tttgaatttg	ctctgaagtt	4680
agcttgtttg	tgtaaaagga	agaatcctat	atcagcctga	gaaatgtaaa	atatcctagc	4740
atttcaagtc	atcaaaatta	tatggagagt	ataaatcatc	cttctgacta	ttcatagtca	4800
tatttggtgc	caccaagtat	aaaacacact	accaaagggc	tgtggaaaaa	atcgccataa	4860
ctgttcttat	tagggaggca	tagcagtggg	acctgaggaa	gttacagcaa	caaccagtca	4920
tccagtcaat	aaccccatgg	ctttgccact	tggaggtacc	caataatgtt	tggctttgcc	4980
gagtaggact	ccaacaaatt	cagagggtca	atttttaaat	gctggttgtc	actgctgaac	5040
agtcccattg	ccctctgcat	aattccacaa	tggaaagctt	tttacctga	ttgccaatca	5100
ttaaacagcc	tactcagcat	aaacaggtat	gatattattc	tgcattttgt	tacattacta	5160
gatgaattcc	tatttcttcc	tacaatagtg	gaactgaaaa	aagatacaca	atcactactac	5220
ccctctacta	atcttatgac	ttatatcatt	tcaattttca	gaccataatg	caaactattg	5280
accaaacaat	gtgaagatga	aaaatagaaa	tgtagaataa	tattacatat	aaaaagaaaa	5340
ggcggactta	ttttgtttta	tttcttagca	tgcatagcaa	tacatgattt	gaggtttata	5400
taataaaggg	acaataaatc	ttcaagaaac	ttacccctac	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	agtgggtgta	ataaaactgt	5700
aatttgggct	atgttttttt	aaatggcttc	accaacatga	aagggaaggga	atgagcatgt	5760
catggatgct	tagagattat	gcttccagca	agaagaattg	agctttggct	cttattacag	5820
aaacatgaca	aggtgtgagt	tttattttatt	agaaattata	taatatttta	agctggggac	5880
taaaaatttt	attgaaacaa	acaggcaagg	gataggcatg	tactagaagc	aaaaatagga	5940
tgtcaatgct	gtaatgttat	tttttgacc	aaaatagtat	ttcctataga	aatgacaatg	6000
atcttaggtt	attattcttc	ataaagatga	caagttcaca	agatatccta	gttcattaaa	6060
atcgtttttag	tcattttaata	gagtgctgtg	atagattaca	caaaggaaaag	cacttacgat	6120
gagaaataat	gatatccaca	attattttct	taattcttag	aaacattcta	ttgttatatc	6180
tcaatctcag	aagccactta	ttgctttatt	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	cttttggtgc	6480
tttccaagca	cagagctcat	taattttacct	catagacttg	ttaaacttct	tctggttcat	6540
caattgaata	gaaatactca	ctactaatta	tgtgagaccc	tgccagtacc	atagcacatg	6600
gataattttt	acataaaaaca	tgcatacaag	taagattatt	cagactgaac	atgaatttta	6660
gagaaatcag	gaaggagtat	atgggagtgg	ttggagtgg	actagagaaa	tgtaatataa	6720
ctataatctc	taatacaaga	tctactaagc	aaaaaacatg	aaacattgtc	attcaagtga	6780
aaacatcagtc	ttcaaatggg	aaagatat	ttactaggaa	aatgtctggg	agatggttat	6840
tatctagaaa	acacaaaaat	tagaaaaacg	taaactttta	taaaaagaat	aatacaatga	6900
gactacatga	aaagttctta	actaatgaaa	caaatatctt	gaaacttttt	tcttaaaagt	6960
ttaatataca	taaccatcat	ggaaattcaa	attaaaaacta	tttacctatt	acccctgaaa	7020
taataactaa	tacccaataa	aaataatata	aacaaaaaat	ggcaatgcat	gccatcatgg	7080
atgtgggaga	gagaatgttc	attgcagttc	tgaatggata	ctgggtccac	cacggtgaaa	7140
atctctgtat	aggctccttc	aaaagctgaa	aatagacata	tcacaagacc	tgccacacat	7200
ttttcaagca	aatacccaaa	ggactctacc	tgactgcaga	gacactttct	cataaaatat	7260
tattgttgat	ctattcataa	tatctggaaa	atagaaacag	ccaagatgoc	catcaactga	7320
ttaatagatg	ataaaattat	tgtacatttc	agtgtaatat	tattcagttt	ttaagaaaaa	7380
tgaaattatg	taataagcat	gtaaatggat	atatcttgaa	acaaccattc	cccattatat	7440
tacctaaaca	ttgaaaagtc	aaaatcatat	gatcttttta	gtggatctac	taatcttttg	7500
ctatatgtat	tttattgaac	tacccatgga	tgtgagataa	ttggtaacaa	cagcacatgg	7560
gagagcatgg	gatcattcaa	ggaagattag	agagaatgca	tttttttagga	gataatggag	7620
gagcaataga	aaggattaaa	tgaggttact	gatgaaagtg	atggtttagag	aaggcaatat	7680
gaggagggat	aactagcact	tagggccttt	tgaaaaagac	atagagaaaa	tactattgta	7740
gaaacttcct	ataattgggtg	tatagttata	tacacaaaag	agctcagatg	gagttaccct	7800
ataatggaaa	tattaactac	tttttatcac	tgtgataaaa	catcctgaac	agagcaacat	7860
agattgggaa	gcattttactt	tggcttacag	ttctaacggg	ataaaaaatc	atgatgaaag	7920
aatgaatatg	tcagcaaaaca	gcagtagcaa	tggcctgaga	agcaggtgag	agctcacatc	7980

ttgaagtgtg	agaatgtagc	agagagaaca	aactgcaaat	gaccagaaaa	tgctttttgga	8040
tcagagccca	taccctctctg	actgacttct	ccagaaattc	tgaacaaata	aaactcccca	8100
aacagagcca	taactgaagg	tccagtgtct	gagactacta	gggggtatttc	ttattcaaac	8160
cactacaatg	gggtgggggg	agcaatcctc	caagtaggca	ctacacacag	acaaataaaa	8220
actctagtaa	ctggaatgga	ttgacttatt	tgaattactt	gccagtggag	ctacatagag	8280
cacaattatt	gtattttaa	taccctttat	gatcttataa	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	aatgctatg	ggagcaacaa	8580
gctattttct	aaattagctt	taatcctatt	ctacaagaga	gaatccatat	ctagaatagt	8640
tatagggatc	aagaacccat	ggcttgattg	gtcataggcc	caatgggaga	tcctaatatt	8700
attgttctac	aaaatgaaaa	taactcctaa	tgacttggtg	ctgcagtaat	aagttagtat	8760
gttgctcaac	tctcacaaga	gaagttttgt	cttacaataa	atggcaatta	aagcagcccc	8820
acaagatttta	tatcataccg	atctcctcat	ggcctatgca	tctagaagct	aggaacacaa	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
ccttaaaagc	ctttctctga	taatgagatt	gatgactacc	ttatatacca	tcctagagcc	9240
ttcatccagt	agctgatgga	agcagaagca	gacatctaca	gctaaacact	gagctagttg	9300
cagacaggga	ggagtgatga	gcaaagtcaa	gaccaggctg	gagaaacaca	cagaaacagc	9360
agacctgaaa	aaaatggttc	acatggaccc	cagactgata	gctgggagtc	cagcatagga	9420
cttttctaga	aacctgaat	gaggatatca	gtttggaggt	ctggttaatc	tatggggaca	9480
ctggtagtgg	atcaatatct	atccctagtt	catgactgga	atttgggtac	ccattccaca	9540
tggaggaatt	ctctgtcagc	ctagacacat	gggggaggtt	ctaggtcctg	ctccaaataa	9600
tgtgttagac	tttgaagaac	tcccttgaga	agactcacc	tccctgggga	gcagaaaggg	9660
gatgggatga	gggttggtga	gggacaggag	aggaggggag	ggtgagggaa	ctgggattga	9720
caagtaaatg	atgcttggtt	ctaattttaa	tgaataaagg	aaaagttaaa	gaagaaaaga	9780
aaacaggcca	aaagattata	aaagacagag	gtggtgggtg	actataaaga	aacactatta	9840
tctaaataaa	aatatgtcag	aagcacacat	gaacttatag	tgtttatgaa	agtatgtata	9900
ataactacat	aatctcaagc	caagaaaaaa	atatcatcct	tcagtgatga	aggtgatttt	9960
atttctccca	gaattaaagc	caaagacctc	atgaaagtaa	ttatcttcaa	aaggttgaaa	10020
atacatactt	tgcaatacac	agatctgcct	agaaatctca	tgttcacaa	acacatgatg	10080
ctcaattgaa	ttccattcaa	tgttacagtt	tagataaaca	gtttgtagat	aaactcacaa	10140
tgtatcattt	ctttttattt	tttgaccaa	cagcttctca	tctgttatct	agaataattc	10200
ctogatggca	ggatatccat	cccaattggg	ggaaggggag	aatttgaa	aaacctagac	10260
cacatacata	tttgccattg	ggaaacaaa	tctaaaatga	tggtgttcac	atcttctcta	10320
ctagtectct	ccccgtccca	aagaaccttg	gtatattgtc	ctcattttac	agagagagga	10380
aagcagaac	tgtcatccc	ttacttgcca	tccctaaccc	aaaatttgca	tcattgtctc	10440
gctctgccc	tctcatatga	cagttacaag	tcaaggcttc	caaagtcctc	ctgtcatggt	10500
tggtgtcaat	agtttataca	gatgacttca	tgtcttcata	tctaattgtc	tatatagatt	10560
aatattaaac	aatgttattt	ctctaaccac	attttaaat	aatttaaaaa	tccatttaatt	10620
gtgtctataa	aatgcagaca	gagtgtgag	acacaatata	agcctgatga	tctgaatttg	10680
aaactcacac	ccaccacatg	gagaatcaac	ttccaaaaat	tttccattta	cttccacact	10740
tacaccattg	tacaaacaca	ataataatga	acaaaatgaa	atgaaataaa	aaattaagtc	10800
tctgtaggta	atgctactgt	gcagcaaaa	taaaaatggc	agcttaagct	tgctttatgg	10860
ttacacttta	ccatcttcca	tttaattata	ggacttcaat	catggcagaa	ctatgctggt	10920
attgtctcag	tgtaacctaa	ccaggtgttc	cagatgttct	taatgtggac	acctaaacta	10980
tttgatattt	gggttaagat	ctttccctct	ttcagaagaa	acctcaggac	agagggaatc	11040
ttgtctttta	attttgagtc	tgtagacttt	ttccatttca	aatatacatg	aaacaagtga	11100
tgaagaaaat	taatcaaaa	gtgggaattg	caatgatatt	aggttcaata	ttaagcttca	11160
atattatcat	ggaatgcctc	gttatacact	gagtgtttgg	caataaggga	tttttagaag	11220
aaggagtttt	tattctcaac	agggttcctt	agtttagctc	aaataaatct	aagcaatcca	11280
ctctagaatt	aaatagtttc	ctaagggcac	agctatgaat	agagctcaat	ttacatataa	11340
aattttgttc	accattttatg	tcattccagt	tttcattagt	acaagggaaa	tacaaaatat	11400
ttagatgtca	atatcaagtg	aatagttcat	ctcctttttt	aatatatatc	acctaaatca	11460
ccattttctc	agaaaaatct	ggcctgaagt	tctgtctgga	acttcaacat	gaaaaaatag	11520
cacagcttgc	tattataaat	octagttgat	ttttaagatt	catgtctggt	gtctgactca	11580
gagggggccag	aggctagaca	aatatttttt	gaatcttcat	tgtgaagatt	tttaatgatt	11640

atTTtaatat	aaataacaaa	gatgatggat	aatgtaactt	tgtacagttc	atagacgctg	11700
aactactttg	tgtttaaata	gttagttccc	tatcataaat	gataggtgat	aagtgtatgt	11760
ttaatacttt	cctctgagc	tatattcatg	tactagagaa	ttattttaaa	catgaaaaga	11820
ctgtgtttat	agtctcagct	cctgagaact	gggtccaact	taggcagggtg	aatgccagga	11880
gcaacgtttt	tcttctacag	aggatgcttt	gctgccaaagc	aacctgggtg	tgtggaaatg	11940
ttcctttttt	aatcaagttt	aaagggctct	catcatgctg	ttgctccaca	tattttcagg	12000
ttagagcctt	gtccttggag	tattatcttt	taccagaaaa	ttcatagtat	tctttcaata	12060
actaacaact	aaacttttctg	ataaaaaaga	attggaattt	caatttttaa	gcctgagtaa	12120
aattccttgtg	aatcaggata	ttttattttta	agtcttatct	tttaaaaagt	tattttatatt	12180
tttaaaaaat	tataatatac	tttcataatt	tccctccttc	acttttcttt	acaaacactt	12240
ctatagatca	ccatgtgttt	ttttttttac	atttatggcc	tctttctgtt	cattgttatt	12300
acatacaaat	agtcttgcct	atagaagaac	accacaattt	gttacctgat	aacaaattat	12360
caacctttaa	aacctacaaa	ctattgatat	tactgaaaag	actatactta	tagatgtaaa	12420
gatatatgtg	tgtgcacata	tatagataca	catatatgta	ggatttttaa	tttttagattt	12480
tagacatcaa	aattattttat	atgactgaga	aactagacac	tataaatgag	cattcagtat	12540
tcaacaccgt	gatttttagat	attgtcacaa	tgacagaaaa	ttttcttata	gaaaatttta	12600
agttttgtga	ttgctctgtg	cacttagtga	agtctcacag	aaaaagaatc	atagtatttt	12660
tagtttataa	taaaaagtac	atataattaa	aatggttggc	acaaaaacaac	attttagcat	12720
ttttcctatt	tactatcaag	tagtatcatt	ttgaaataat	aatttgacta	gtttcaaaaa	12780
tgaaaacaaa	attttaaacta	aatgcctaata	ctagcctgat	aacattttta	tgaatgaaat	12840
tattcaatag	tgttatcaat	taggggcca	aaacttttcc	taaaataaaa	cttttaattt	12900
ttttccattt	ttattttaa	tagaaacaaa	attgttttac	atgtaaatca	gagtttcctc	12960
accctcccc	tctccctgtc	cctcactaac	accctacttg	tcccatacca	tttctgtctc	13020
ccagggagg	tgaggccttc	catggggaaa	cttcagagtc	tgtctatcct	ttcggtatag	13080
gcctaggccc	tcacccattt	gtctaggcta	aggctcacaa	agtttactcc	tatgctagt	13140
ataagtactg	atctactaca	agagacacca	tagatttcct	aggcttcctc	actgacacc	13200
atgttcatgg	ggctctggaac	aatcatatgc	tagtttccta	ggatcagtc	tggggaccat	13260
gagctcccc	ttgttcaggt	caactgtttc	tgtgggtttc	accaccctgg	tcttgactgc	13320
tttgctcatc	actcctccct	ttctgtaact	gggttccagt	acaattccgt	gttttagctgt	13380
gggtgtctac	ttctactttc	atcagcttct	gggtaggagc	ctctaggata	gcatacaatt	13440
agtcatcatc	tcattatcag	ggaagggcat	ttaaagtagc	ctctccattg	ttgcttggat	13500
tgtagtttgg	tgtcatcttt	gtagatctct	ggacatttcc	ctagtgcag	atatctcttt	13560
aaacctacaa	gactacctct	attatggtat	ctcttttctt	gctctcgtct	attcttccag	13620
acaaaatctt	cctgtccct	tatattttcc	tctccctcc	tcttctcccc	ttctcattct	13680
cctagatcca	tcttcccttc	ccccatgctc	ccaagagaga	tgttgctcag	gagatcttgt	13740
tccctaaacc	ttttcttggg	gatctgtctc	tcttaggggt	gtccttgttt	cctagcttct	13800
ctggaagtgt	ggattgttaag	ctggtaatca	tttgctccat	gtctaaaatc	catatatgag	13860
tgatgtttgt	ctttttgtga	ctgggttacc	tcactcaaaa	tggtttcttc	catatgtctg	13920
tggatttcaa	tagcacaaaac	aacatacagt	atcttggggc	aacactaacc	aaacaagtga	13980
aagaccagta	tagcaagaac	tttgagttaa	aagaaagaaa	ttaaagaaga	taccagaaaa	14040
tggaaagatc	tccatgctc	tttgataggc	agaaatcaaca	tagtaaaaat	ggcaatcttg	14100
ccaaaatcca	tctacagact	caatgcaatc	ccattaaat	accagcacac	ttcttcacag	14160
acctgaaaga	ataatactta	actttatatg	gagaaacaaa	agaccagga	taggccaaaac	14220
aacctgttac	aatgaaggca	cttccagagg	catccccatc	cctgacttca	agctctatta	14280
tagagtaata	atcctgaaaa	cagcttggtta	atggcacaaa	aatagacagg	tagaccaatg	14340
gaattgagtt	gaaaaccttg	atattaacct	acatatctat	gaacacctga	ctttgacaaa	14400
gaagctaagg	ttatacaatg	taagaaagaa	agcatcttca	acaaatcgtg	ctggcataac	14460
tggatgctgg	catgtagaag	actgcagata	gatccatgtc	taatgccatg	cacaaaactt	14520
aagtccaaat	ggatcaaaaa	cctcaacata	aatccagcca	cactgaacct	catagaagag	14580
aaagtgggaa	gtatccttga	ataaattggg	acaggagacc	acatcttgaa	cttaacacca	14640
gtagcacaga	caatcagatc	aataatcaat	aaatgggacc	tccctgaaact	gagaagcttc	14700
tgttaaggcaa	tggataagtc	aacaggacaa	aatggcagcc	cacggaatgg	gaaaagatat	14760
tcaccaatcc	tatatctgac	agagggctgc	tctctatttg	caaagaacac	aataagctag	14820
ttttttaa	accaattaat	ccgattataa	agttgggtag	agaactaaat	aaagaattgt	14880
taacagagca	atctaacttg	gcagaaagac	acataagaaa	gtgctcacca	t	14931

EXAMPLE 13 – Guide Sequences for AAVS1-Like Region Sequences in CHO

(The below guides can be sense guide sequences or antisense guide sequences)

	SEQ ID NO
CCCCGCTGGCGCCGGGATCGGGG	13
GAGTCGAGCACCGCTCGGGCAGG	14
TTCCCGCTGGCGCCGGGATCGG	15
GTGTGCGGAAGACGCCGCCGGG	16
CGGTGACAGCGCGGATGACAGGG	17
CAGCGCGGATGACAGGGGCGAGG	18
GCCGGCGTCCGATTCCCCGCTGG	19
CGTGTGCGGAAGACGCCGCCGGG	20
GAGGCGCTCCACCGTCTGTTGGG	21
GTCCGATTCCCCGCTGGCGCCGG	22
GACCCCGGGGGCCCCGATCCCGG	23
CGGCGTCTCCGCACACGGATGG	24
TCGAGCACCGCTCGGGCAGGCGG	25
AGCTACGCCGGCCCCATAAAGG	26
CATCGTCCTATATATAGCAGG	27
AGAGGCGCTCCACCGTCTGTTGG	28
GGTCGGCTGCGGAAGCATCAGG	29
TGCTTCGCGCAGCCGACCCCGG	30
GGCCCCGATCCCGGCCAGCGG	31
TCCGATTCCCCGCTGGCGCCGGG	32
TCCCGGCGCCAGCGGGGAATCGG	33
TGGTGGAGTCGAGCACCGCTCGG	34

CAAGATGGTCCTCACTCTCGGGG	35
GCTTCGCGCAGCCGACCCGGGG	36
GTGGAGCGCCTCTTCTCCAGGGG	37
GACGTGTCAGCCTTCCAGGTGGG	38
GCCAGCGGGGAATCGGACGCCGG	39
TCTCCCCGTCATCCAAAAGCTGG	40
GGTGGAGTCGAGCACCCTCGGG	41
GCTGCCCAAATATAGTCCATGGG	42
ATGCTTCGCGCAGCCGACCCGG	43
GCGGTGACAGCGCGGATGACAGG	44
ATGCTCGGGGGCCGCTGACCTGG	45
TTGTATTGCCGGGATCCTTCTGG	46
TCCCCGCTGGCGCCGGGATCGGG	47
CTCGACTCCACCAACGCCGACGG	48
GGTGGCAAGATCACAAAAGGGG	49
GGCGCTGATGCCGTCGGCGTTGG	50
TCCACGAGCATCCTAGCAAGAGG	51
AGGCTGACACGTGAGGCCTGAGG	52
GGGTGTAAGCCATCCGTGTGCGG	53
AGGATCCCGGCAATACAAGATGG	54
GGATGGGGCCCAACAGACGGTGG	55
GAGGACCATCTTGTATTGCCGGG	56
AGTCGCCCAGGGTCTGGTGGGG	57
GGCGGGAGGAGAGTCCACCTGG	58
ACCTACCCACAGGACCCTGGG	59
TCAGCGTCTTTGACCAGTCCAGG	60

CGTCCCGCCGCTGCCCGAGCGG	61
GTCCCGGGATCCCGGGGTCGG	62
AGCACCGCTCGGGCAGGCGGCGG	63
CGGTGGAGCGCCTCTTCTCCAGG	64
GCCCCGATCCCGGCGCCAGCGGG	65
ATCGGGGCCCCGGGGTCCCCGG	66
ATTCTCGGCTCATCCCCTTTTGG	67
ACCACCCCATGGACTATATTGG	68
TAGCAAGAGGACGACAACCCAGG	69
GCTGATGCCGTCGGCGTTGGTGG	70
TACAAGATGGTCTCACTCTCGG	71
TGCCGGGATCCTTCTGGATTCGG	72
CCCAAATATAGTCCATGGGGTGG	73
TACCTGTAGAATGGGACCACTGG	74
TCTTGCTAGGATGCTCGTGGAGG	75
ATGCCAGCTTTTGGATGACGGGG	76
ATAGTCCATGGGGTGGTAGGTGG	77
CAGGACCCTGCTATATATAGAGG	78
GGGGCCGGCGTGAGCTGTGTGGG	79
ACCTGGAAGGCTGACACGTCAGG	80
CAGCGGACAGCACGGGTACAGG	81
CGCCGGGATCGGGGCCCCGGGG	82
GCGCCGGGATCGGGGCCCCGGGG	83
GCCGGGATCCTTCTGGATTCGGG	84
GCTGTCACCGCTCTCCCCGGCGG	85
GGTGACAGCGCGGATGACAGGGG	86

GAAGACGCCCGGGGAGAGCGG	87
AAGGGGTAACTGCCTGGAGGG	88
CTTCTAGAACCTACCCACCAGG	89
GGGGCCGCTGACCTGGTGCAGGG	90
TCCCGAATCCAGAAGGATCCCGG	91
TCACAGTGTCTGAGTCGCCAGG	92
GCTAGGATGCTCGTGGAGGTGGG	93
CTCGTGGAGGTGGGGAATAAAGG	94
TGATCTGTCCCCGAGAGTGAGG	95
TGAATTAATAGGACATGGGGAGG	96
ACTCTCGGGGCACAGATCACTGG	97
GACCACTGGTCCCATTCTACAGG	98
TGAGGACCATCTTGATTGCCGG	99
AGAATCCGTCTGTCCTGGGCTGG	100
TGACGTGTCAGCCTTCCAGGTGG	101
AAAAGCATCCCGAATCCAGAAGG	102
TTTTCCCGAGGCCACACTCAGGG	103
AGCGCCCTGCACCAGGTCAGCGG	104
TGAGTCGCCCAGGGTCCTGGTGG	105
CACGTCAGGCCTGAGGTCACAGG	106
GTTTTCCCGAGGCCACACTCAGG	107
GAGTCGCCCAGGGTCCTGGTGGG	108
GAATCCGTCTGTCCTGGGCTGGG	109
CCGGGGAGGGAGGATGCTCGGGG	110
GAGCCGAGAATTGAATTAATAGG	111
GTGACCCGTGCTGTCCGCTGTGG	112

GGGGGCGTCAAGTCAGAGCTGGG	113
CGCGCTGTACCGCTCTCCCCGG	114
TGCTGCCCAAATATAGTCCATGG	115
AGTTGAGGAGAAACCTATGGGGG	116
ATCGTCTCTATATATAGCAGGG	117
GAGTTGAGGAGAAACCTATGGGG	118
AGGGGTACACTGCCTTGAGGGG	119
ATAGAGTCCCTCTGGGGACAGGG	120
CTAGGATGCTCGTGGAGGTGGGG	121
CGCGGATGACAGGGGCGAGGCGG	122
AAACAGAATCCGTCTGTCCTGG	123
TGCTAGGATGCTCGTGGAGGTGG	124
GAGTGTGGCCTCGGGAAAACAGG	125
CGAGGCGGCCCTGCAGGGCAGG	126
CTTACACCCCGTGCCTTCCAGG	127
CTGCCCAAATATAGTCCATGGGG	128
AGGGGCAAAGGACCCTCCTGAGG	129
TCTCACCATAGAGTCCCTCTGGG	130
AGTGTACCCCTTTGTTCCCTGG	131
GTTGAGGAGAAACCTATGGGGGG	132
CACAGTGTCTGAGTCGCCAGGG	133
TCCTCTTGCTAGGATGCTCGTGG	134
CACATGATCACCAAAGTCCCTGG	135
AAACAGAATCCGTCTGTCCTGGG	136
GTGCAGGGCGCTGATGCCGTCGG	137
CAGCACGGACTTTTTTTGTTTGG	138

CAGCCGACCCCGGGGATCCCGGG	139
TTGGTCAAGATTTTGCAACTGG	140
TTGGTCAAGATTTTGCAACTGGG	141
CCTGTCAGAGAGGATGCTCTAGG	142
TGGGTTGTCGTCCTCTGCTAGG	143
CTCACCATAGAGTCCCTCTGGGG	144
AGACGCTGACACCCTGAGTGTGG	145
GGTGGAGCGCCTCTTCTCCAGGG	146
AACAAAGGGGTACACTGCCTTGG	147
CTGTCCCAGAGGGACTCTATGG	148
GTCTCTGACCCCTCATTTGTGG	149
GTCTGAGTCGCCCAGGGTCCTGG	150
GCCTGACGTGTCAGCCTTCCAGG	151
ATGTCCCACACAGCTCACGCCGG	152
GGCTGTGCCGCAGGCTTCCAGGG	153
CACCCCGTGCCTTTCCAGGCTGG	154
TCACCAAGGTGTCTGCATGGCGG	155
GAATTGAATTAATAGGACATGGG	156
AATATAGTCCATGGGGTGGTAGG	157
CAGGGCATAGTTTTAAAGCAGG	158
CGGCATCAGCGCCCTGCACCAGG	159
GCAGCCGACCCCGGGGATCCCGG	160
ACAGGACCTGTATTTGAGGTTGG	161
GGGGCGAGGCGGCCCTGCAGGG	162
AACCTACCCACAGGACCCTGG	163
GCACCGCTCGGGCAGGCGGCGGG	164

TGGAAGCCTGCGGCACAGCCAGG	165
GAACCAACACTGTGGCCAGGAGG	166
AGCAGGGTCCTGTTTTCCCGAGG	167
TGGGTGGCAAGATCACCAAAGG	168
AGCCGACCCCGGGGATCCCGGGG	169
CATGGCAACTTCCATCTCCTGGG	170
ACAGCACGGGTCACAGGAAGTGG	171
GGGGGCCGCTGACCTGGTGCAGG	172
AGGGGCGAGGCGGCCCTGCAGG	173
GTGGTCACCCCTGTCCCCAGAGG	174
TGGGGCCGGCGTGAGCTGTGTGG	175
GGAAGCCTGCGGCACAGCCAGGG	176
CCTGAGCTGATCTCCTGGACTGG	177
ATCCAAAAGCTGGCATTGTCAGG	178
CTCTACCATAGAGTCCCTCTGG	179
GGGGGGCGTCAAGTCAGAGCTGG	180
GGGTGGAAATCTAAGAGACAGGG	181
CGGGGAGAGCGGTGACAGCGCGG	182
AGAATTGAATTAATAGGACATGG	183
CCATAAAGGAAGTTTTCCACAGG	184
AGTGAACCAACACTGTGGCCAGG	185
GTTGGGAGGGAACTCTTGGGAGG	186
CGGGTCACAGGAAGTGGGGTAGG	187
GGCCTGGCTAGCCTCAGAGGAGG	188
GGTAGGTTCTAGAAGGTGACAGG	189
ACAAGATGGTCCTCACTCTCGGG	190

CAAGGTGTCTGCATGGCGGGAGG	191
TGTTTCACTCATCCAGGCAGAGG	192
TGTGGAAACTTCCTTTATGGGG	193
GAGGACGACAACCCAGGAGATGG	194
GGGTGGCAAGATCACCAAAGGG	195
CTGGTGGGGTAGGTTCTAGAAGG	196
ACTCTTCAGGCCTTTGCAGGAGG	197
CAGAGGGACTCTATGGTGAGAGG	198
AGCACGGGTCACAGGAAGTGGGG	199
CTATGGTGAGAGGCGTCCTGTGG	200
CAGCACGGGTCACAGGAAGTGGG	201
ATGGGATGGGGCCCAACAGACGG	202
CTCCCGCCATGCAGACACCTTGG	203
GGGGATCCCGGGGACCCGGGGG	204
TGTCCGCTGTGGCCTCAGGAGGG	205
GCAGTTGGGAGGGAACCTTGGG	206
GCCTGGCTAGCCTCAGAGGAGGG	207
GAGAGTTGAGGAGAACTATGG	208
ACCTGTATTTGAGGTTGGCCTGG	209
TCGGGCAGGCGGCGGGACGCCGG	210
TAGAGTCCCTCTGGGGACAGGGG	211
GAAGTGACACTGAAGGGCCTGGG	212
AGCAGCCTGAGCTGATCTCCTGG	213
TCATGGCAACTCCATCTCCTGG	214
GTCACAGGTTCTGTCAGAGAGG	215
CACCAAGGTGTCTGCATGGCGGG	216

GTAGGTTCTAGAAGGTGACAGGG	217
AATTGAATTAATAGGACATGGGG	218
TTCCAAGCACCTGATTTCTGTGG	219
CTGTCAGAGAGGATGCTCTAGGG	220
GTGCTGTCCGCTGTGGCCTCAGG	221
CCCGGGGAGGGAGGATGCTCGGG	222
GGGGTGGCTCGGGGGGCCCCGGG	223
GTCAAGTCAGAGCTGGGCCCTGG	224
GAACTGAAGGGCCTGGGCCTGG	225
CTGAAAGTGAACCAACTGTGG	226
AAAGGGGTAACTGCCTTGGAGG	227
TCTGGAACTTCTAAGCATTGG	228
AGAGTTGAGGAGAAACCTATGGG	229
CTCTGAGGCTAGCCAGGCCAGG	230
AGGGGTGGCTCGGGGGGCCCCGG	231
TCCACATTGATTTGCCTTTCTGG	232
AGGGTGGAAATCTAAGAGACAGG	233
CTGTCCGCTGTGGCCTCAGGAGG	234
AGACACAGGACCTGTATTTGAGG	235
GGGTGGCTCGGGGGGCCCCGGGG	236
TGGCTGTGCCGAGGCTTCCAGG	237
TCCAGAAAGGCAAATCAATGTGG	238
CACCAGCCTGGAAAGGCACGGGG	239
CGGGGTCCCCGGGATCCCCGGGG	240
AACTGCCTTGGAGGGGCAAAGG	241
TGACAGGAACCTGTGACCTCAGG	242

TTCATGTGGGCTATCAAGATGG	243
GGGAACTCTTGGGAGGGCCAGGG	244
GAGGCCACAGCGGACAGCACGGG	245
TGTCCTATTAATTCAATTCTCGG	246
AGGGAACCTTGGGAGGGCCAGG	247
GCACCTGATTCTGTGGTATTGG	248
GTATCTTGAGTGTCTTTCTCGG	249
CTGTGGAAAACCTCCTTTATGGG	250
GGGACCTCCAGCAGATGCAGAGG	251
CTGGGGACAGGGGTGACCACTGG	252
TTCAGTGTCACTTCTTTGGGGG	253
TCCCTCCTCTGAGGCTAGCCAGG	254
GGCGCCGGGATCGGGGCCCCCGG	255
GGTCTAGAAGGTGACAGGGTGG	256
TTGGGAGGGAACCTTGGGAGGG	257
ATCAGGTGCTTGAAAGTAGAGG	258
CGGGGAGGGAGGATGCTCGGGGG	259
TGAGGCCACAGCGGACAGCACGG	260
TGGTCACCCCTGTCCCAGAGGG	261
TCCCCGCCTCCTGCCCTGCAGGG	262
GGCTGCCCTGGCTGTGCCGAGG	263
TCCAAAAGCTGGCATTGTCAGGG	264
CAATGCCAGCTTTTGATGACGG	265
CTGGGCCTGGCTAGCCTCAGAGG	266
GGTTCACCTTCAGTCTTTCATGG	267
AGGAGAAACCTATGGGGGGTGGG	268

TCTAAAAGACAGCCCAGCCCAGG	269
CGGGGATCCCGGGGACCCCGGGG	270
TCTTCTCCAGGGGAACAAAGGGG	271
ACTGACACAAAAAGTCAGCACGG	272
CCTGAAGAGTCAGGTCACCAAGG	273
GGAGGAGAGTCCACCTGGAAGG	274
GGGCAGCCACCAGCCTGGAAAGG	275
AGCCCTATTCTCTCTCTCTGG	276
GCCACCAGCCTGGAAAGGCACGG	277
TGACACCCTGAGTGTGGCCTCGG	278
AATTAATAGGACATGGGGAGGGG	279
GGCTCGGGGGGCCCCGGGGAGGG	280
TAATAGGACATGGGGAGGGGAGG	281
CTCTTCTCCAGGGGAACAAAGGG	282
TCGGGGCCCCCGGGGTCCCCGGG	283
TCCCTGACAATGCCAGCTTTTGG	284
GAATTAATAGGACATGGGGAGGG	285
AATGAGGGGGTCAGAGACACAGG	286
GAAAACTTCCTTTATGGGGCCGG	287
CTTGGGAGGGCCAGGGACTTTGG	288
CCCCTGCAGGGCAGGAGGCGGGG	289
TCAGTGCACTTCTTTTGGGGGG	290
ATCCCCGTTCTTCTCCTCCTGG	291
CTTCCTCCTGGCCACAGTGTTGG	292
TGCAGTTGGGAGGGAACTCTTGG	293
TGGCTCGGGGGGCCCCGGGGAGG	294

CTGCAAAGGCCTGAAGAGTCAGG	295
CCGTGTGCGGAAGACGCCCGG	296
CCCCGGGGAGGGAGGATGCTCGG	297
TTCCAGGCTGGTGGCTGCCCTGG	298
AGGTCACCAAGGTGTCTGCATGG	299
GGCGGCCCTGCAGGGCAGGAGG	300
TCGGGGGGCCCCGGGGAGGGAGG	301
CCAAAAGAAGTGACACTGAAGGG	302
GGCCAGGAGGAAGAAGAACGGGG	303
GCCCAGGGTCCTGGTGGGGTAGG	304
GCTAGCCTCAGAGGAGGGAGTGG	305
GAGGGTCCTTTGCCCTCCAAGG	306
CCACCAGCCTGAAAGGCACGGG	307
GATTTCTGTGGTATTGGGGTTGG	308
CTAGCCTCAGAGGAGGGAGTGGG	309
CCCGGGGTCCCCGGGATCCCCGG	310
CATGGGGTGGTAGGTGGAGTGGG	311
AATGCCAGCTTTTGGATGACGGG	312
GCCCCTGCAGGGCAGGAGGCGGG	313
GAGGAGAAACCTATGGGGGGTGG	314
AGAAGTGACACTGAAGGGCCTGG	315
CCTCCAGCAGATGCAGAGGAAGG	316
CCTCTTCTCAGGGGAACAAAGG	317
CCGGGGTCCCCGGGATCCCCGGG	318
TAGCCTCAGAGGAGGGAGTGGGG	319
CAGAGGAAGGGGATGCAGTTGGG	320

CTCCAGCAGATGCAGAGGAAGGG	321
GATTCTGTTTTCTCTGCCTGG	322
CTTCAGTGTCACTTCTTTGGGG	323
CATAGAGTCCCTCTGGGGACAGG	324
GGACCCTCCTGAGGCCACAGCGG	325
CCATGGGGTGGTAGGTGGAGTGG	326
GACACCCTGAGTGTGGCCTCGGG	327
ATGCTTAGAAGTTCCAGAAAGG	328
AGCTGGGCCCTGGAAGCCTGCGG	329
TACCACAGAAATCAGGTGCTTGG	330
ACCCAATACCACAGAAATCAGG	331
TTCTACAGGTAAAAAACTAAGG	332
GGCCCCTGCAGGGCAGGAGGCGG	333
CTCCCCGCTCCTGCCCTGCAGG	334
TCTCTGACCCCTCATTTGTGGG	335
GGAGAAACCTATGGGGGGTGGGG	336
ACAGCCCAGCCCAGGACAGACGG	337
CCTGTATTTGAGGTTGGCCTGGG	338
AGCCAGGGCAGCCACCAGCCTGG	339
AGCCTCAGAGGAGGGAGTGGGGG	340
G TTCAGTGTTCACATCCAGG	341
CTGACTCTTCAGGCCTTTGCAGG	342
ATCCCCACTCCCTCCTCTGAGG	343
CCCAAAGAAGTGACACTGAAGG	344
TGGCCAGGAGGAAGAAGAACGGG	345
GGGAGGAAGGTTATGGGATGGGG	346

CCTGAGGCTTCCTGCACTCTAGG	347
TAGTTTTTTTACCTGTAGAATGG	348
AAGTGGGGTAGGGAACAAGGTGG	349
GGGTCACAGGAAGTGGGGTAGGG	350
CACCTGATTTCTGTGGTATTGGG	351
TTTGCAACTGGGTCTCATGTGGG	352
GAGAAACCTATGGGGGGTGGGGG	353
GAGGGAGGAGGGGTGGCTCGGGG	354
GCCTGTAATCCCACAAATGAGGG	355
GCAGAGGAAGGGGATGCAGTTGG	356
ACCTGATTCTGTGGTATTGGGG	357
AAACCAGAGGAGAGAGAAATAGG	358
AACCAGAGGAGAGAGAAATAGGG	359
GAGGAGAGAGAAATAGGGCTTGG	360
CTGCAGGGCAGGAGGCGGGGAGG	361
AGGAAGGGGATGCAGTTGGGAGG	362
GGTATTGGGGTTGGAACCTGAGG	363
TTTTGCAACTGGGTCTCATGTGG	364
GGGAGGAGGGGTGGCTCGGGGGG	365
TCCCCTTCCTCTGCATCTGCTGG	366
AGGAAGTGGGGTAGGGAACAAGG	367
AGGGGAGGAAGGTTATGGGATGG	368
CCTGTAATCCCACAAATGAGGGG	369
AGTTTTTTTACCTGTAGAATGGG	370
AGGAAGAAGAACGGGGATGGGGG	371
TGCAGGGCAGGAGGCGGGGAGGG	372

GGAAGGGGATGCAGTTGGGAGGG	373
CAAAGTCACTGTGTAGATGAAGG	374
GTGGCCAGGAGGAAGAAGAACGG	375
TCCAGCAGATGCAGAGGAAGGGG	376
AGGGAGGAGGGGTGGCTCGGGGG	377
AACCTATGGGGGGTGGGGGTGGG	378
GGGGAGGGGAGGAAGGTTATGGG	379
CACCCACCCCCACCCCCATAGG	380
AAACCTATGGGGGGTGGGGGTGG	381
GGAGGAAGAAGAACGGGGATGGG	382
GAGGAAGAAGAACGGGGATGGGG	383
GGAGGGAGGAGGGGTGGCTCGGG	384
GGTTGGCCTGGGCTACACAGGGG	385
AGAGGAGGGAGTGGGGGATTGGG	386
GAGGTTGGCCTGGGCTACACAGG	387
AGGTTGGCCTGGGCTACACAGGG	388
TGGGGAGGGGAGGAAGGTTATGG	389
GAAAGTAGAGGCAGGAGGGTTGG	390
GAGGAGGGAGTGGGGGATTGGGG	391
GGGGAGGAAGGTTATGGGATGGG	392
AGAGTGCTTGCCTAGAGTGCAGG	393
AGGAGGGAGTGGGGGATTGGGGG	394
CAGAGGAGGGAGTGGGGGATTGG	395
AGGAGGAAGAAGAACGGGGATGG	396
TTTTTCCCCTGTGTAGCCAGG	397
GGTGCTTGGAAGTAGAGGCAGG	398

CTTGAAAGTAGAGGCAGGAGGG	399
CTGTAATCCACAAATGAGGGGG	400
AGGACATGGGGAGGGGAGGAAGG	401
TCATCTACACAGTGACTTTGAGG	402
GAGGGAGTGGGGGATTGGGGGGG	403
GGGAGGGAGGAGGGGTGGCTCGG	404
AGGGAGTGGGGGATTGGGGGGGG	405
AACAACAAAAACAAAACCAGAGG	406
CTATGGGGGGTGGGGGTGGGTGG	407
GGAGGGAGTGGGGGATTGGGGGGG	408
TGCCTGTAATCCACAAATGAGG	409
GAGTGGGGGATTGGGGGGGGGGG	410
AGGGCAGGAGGCGGGGAGGGAGG	411
CAGGAGGCGGGGAGGGAGGAGGG	412
AGGAGGCGGGGAGGGAGGAGGGG	413
GGAGTGGGGGATTGGGGGGGGGG	414
GGGAGTGGGGGATTGGGGGGGGG	415
GCAGGAGGCGGGGAGGGAGGAGG	416
AGGCGGGGAGGGAGGAGGGGTGG	417
GCTTGAAAGTAGAGGCAGGAGG	418
GAGAGAGAGAGAGAGAGTTGAGG	419

[00120] 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 inventions, 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 claim is:

1. A mammalian cell comprising a first Stable Integration Site located in a Genomic Safe Harbor and a second Stable Integration Site that is not located in the Genomic Safe Harbor, wherein the first Stable Integration Site comprises a first reporter gene encoding a first reporter protein and the second Stable Integration Site comprises a second reporter gene encoding a second reporter protein, wherein the first reporter protein and the second reporter protein are different.

2. The mammalian cell according to claim 1, wherein the first and second Stable Integration Sites comprise recombinase recognition sites.

3. The mammalian cell according to claim 1, wherein the first reporter gene is under the control of an SV40 promoter.

4. The mammalian cell according to claim 1, wherein the second reporter gene is under the control of an SV40 promoter.

5. The mammalian cell according to claim 1, wherein the first reporter gene encodes a fluorescent protein.

6. The mammalian cell according to claim 1, wherein the second reporter gene encodes a fluorescent protein.

7. The mammalian cell according to claim 1, wherein the cell further comprises a polynucleotide encoding a repressor under the control of a CMV promoter.

8. The mammalian cell according to claim 1, wherein the mammalian cell is a human cell.

9. The mammalian cell according to claim 8, wherein the human cell is a Human Amniotic Epithelial Cell.

10. The mammalian cell according to claim 8, wherein the human cell is a HEK293 Cell.

11. The mammalian cell according to claim 1, wherein the mammalian cell is a CHO cell.

12. The mammalian cell according to claim 2, wherein the recombinase recognition sites are lox sites.

13. The mammalian cell according to claim 1, wherein a polynucleotide encoding a protein of interest is inserted into the first Stable Integration Site or the second Stable Integration Site.

14. The mammalian cell according to claim 1, wherein the second Stable Integration Site is located in a second Genomic Safe Harbor that is different from the first Genomic Safe Harbor.

15. The mammalian cell according to claim 1, wherein the second Stable Integration Site is located in a region that is not a Genomic Safe Harbor.

16. A mammalian cell comprising a first Stable Integration Site located in a Genomic Safe Harbor and a second Stable Integration Site that is not located in the Genomic Safe Harbor, wherein the first Stable Integration Site comprises first polynucleotide encoding a first protein and the second Stable Integration Site comprises a second polynucleotide encoding a second protein.

17. The mammalian cell according to claim 16, wherein the first protein is a viral protein.
18. The mammalian cell according to claim 16, wherein the second protein is a viral protein.
19. The mammalian cell according to claim 16, wherein the first protein is an adenovirus associated virus protein.
20. The mammalian cell according to claim 17, wherein the second protein is an adenovirus associated virus protein.
21. The mammalian cell according to claim 17, wherein the first protein is an adenovirus protein.
22. The mammalian cell according to claim 17, wherein the second protein is an adenovirus protein.
23. The mammalian cell according to claim 16, wherein the mammalian cell comprises a polynucleotide encoding an adeno-associated virus protein and a polynucleotide encoding an adenovirus protein.
24. The mammalian cell according to claim 16, wherein the second Stable Integration Site is located in a second Genomic Safe Harbor that is different from the first Genomic Safe Harbor.
25. The mammalian cell according to claim 16, wherein the second Stable Integration Site is located in a region that is not a Genomic Safe Harbor.
26. The mammalian cell according to claim 16, wherein the mammalian cell is a human cell.

27. The mammalian cell according to claim 16, wherein the mammalian cell is a CHO cell.

28. A mammalian cell comprising a first Stable Integration Site located in a Genomic Safe Harbor and a second Stable Integration Site that is not located in the Genomic Safe Harbor, wherein the first Stable Integration Site comprises a polynucleotide encoding a first reporter gene encoding a first reporter protein and the second Stable Integration Site comprises a polynucleotide encoding Cas9 and a polynucleotide encoding a second reporter gene encoding a second reporter protein, wherein the first reporter protein and the second reporter protein are different.

29. The mammalian cell according to claim 28, wherein the second Stable Integration Site further comprises a selection marker gene and an internal ribosome entry site (IRES).

30. The mammalian cell according to claim 28, wherein the first and second Stable Integration Sites comprise recombinase recognition sites .

31. The mammalian cell according to claim 28, wherein the mammalian cell is a Human Amniotic Epithelial Cell.

32. The mammalian cell according to claim 28, wherein the mammalian cell is a HEK293 Cell.

33. The mammalian cell according to claim 28, wherein the mammalian cell is a BHK Cell.

34. The mammalian cell according to claim 28, wherein the mammalian cell is a CHO Cell.

35. A method for making at least one protein of interest, comprising:
- (a) providing a mammalian cell comprising a first Stable Integration Site located in a first Genomic Safe Harbor and a second Stable Integration Site that is not located in the first Genomic Safe Harbor, wherein the first Stable Integration Site comprises a first reporter gene encoding a first reporter protein and the second Stable Integration Site comprises a second reporter gene encoding a second reporter protein, wherein the first reporter protein and the second reporter protein are different, and wherein the first and second Stable Integration Sites comprise recombinase recognition sites;
 - (b) introducing a polynucleotide encoding the protein of interest into a Stable Integration Site by recombinase mediated cassette exchange, and
 - (c) culturing the mammalian cell under conditions that allow expression of the polynucleotide encoding the polynucleotide of interest.
36. The method according to claim 35, wherein the mammalian cell is a Human Amniotic Epithelial Cell.
37. The method according to claim 35, wherein the mammalian cell is a HEK293 cell.
38. The method according to claim 35, wherein the mammalian cell is a CHO cell.
39. The method according to claim 35, wherein the first Stable Integration Site comprises a first polynucleotide encoding a first protein and the second Stable Integration Site comprises a second polynucleotide encoding a second protein.
40. The method according to claim 39, wherein the first protein is a viral protein.

41. The method according to claim 39, wherein the second protein is a viral protein.

42. The method according to claim 35, wherein the mammalian cell comprises a polynucleotide encoding an adeno-associated virus protein and a polynucleotide encoding an adenovirus protein.

43. The method according to claim 35, wherein the second Stable Integration Site is located in a second Genomic Safe Harbor that is different from the first Genomic Safe Harbor.

44. A method of creating a mammalian cell with multiple Stable Integration Sites, wherein the method comprises:

(A) providing a mammalian cell comprising a first DNA cassette comprising in 5' to 3' order a polynucleotide encoding the first lox site, a promoter, a selection marker gene encoding a selection marker protein, an IRES, a first reporter gene encoding a first reporter protein, a promoter operably linked to an operator, a Cas9 gene and the second lox site;

(B) integrating a second DNA cassette comprising in a 5' to 3' order a polynucleotide comprising a first Genomic Safe Harbor homology arm containing a CRISPR sgRNA target site, a third lox site, a second reporter gene encoding a second reporter protein, a forth lox site and a second Genomic Safe Harbor homology arm containing an CRISPR sgRNA target site, wherein the first lox site, the second lox site, the third lox site and the forth lox site are different, wherein the first and second guide arms can contain a region with alterations, and wherein the second reporter protein is different from the first reporter protein;

(C) exchanging the first DNA cassette with a third DNA cassette, wherein the third DNA cassette comprises in a 5' to 3' order a polynucleotide encoding the first lox site, a promoter, a third reporter gene encoding a third reporter protein, and the second lox site, wherein the third reporter protein is different from the second reporter protein, thereby providing the mammalian cell with multiple Stable Integration Sites.

45. The method according to claim 44, wherein the alterations prevent recreating a targetable site.

46. The method according to claim 44, wherein the mammalian cell is a HEK293 Cell.

47. The method according to claim 44, wherein the mammalian cell is a CHO Cell.

48. The method according to claim 44, wherein the cell of step (A) further comprises a polynucleotide encoding a repressor under the control of a CMV promoter.

49. The method according to claim 44, wherein the cell of step (B) further comprises a polynucleotide encoding a repressor under the control of a CMV promoter.

50. The method according to claim 44, wherein the cell of step (C) further comprises a polynucleotide encoding a repressor under the control of a CMV promoter.

51. A mammalian cell comprising a modified genome, wherein the genome is modified by insertion of at least three DNA cassettes within different regions of the genome, wherein the modified genome comprises

- (1) a first deoxyribonucleic acid sequence that is at least 90% identical to at least one selected from the group consisting of SEQ ID NOS: 1 and 2 prior to modification;
- (2) a second deoxyribonucleic acid sequence that is at least 90% identical to at least one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification; and
- (3) a third deoxyribonucleic acid sequence that is at least 90% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification,

wherein the first deoxyribonucleic acid sequence is modified by insertion of a first DNA cassette, the second deoxyribonucleic acid sequence is modified by insertion of a second DNA cassette, and the third deoxyribonucleic acid sequence is modified by insertion of a third DNA cassette.

52. The mammalian cell according to claim 51, wherein

- (1) the first deoxyribonucleic acid sequence is at least 95% identical to at least one selected from the group consisting of SEQ ID NOS: 1 and 2 prior to modification;
- (2) the second deoxyribonucleic acid sequence is at least 95% identical to at least one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification; and

(3) the third deoxyribonucleic acid sequence is at least 95% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification.

53. The mammalian cell according to claim 51, wherein

(1) the first deoxyribonucleic acid sequence is at least 98% identical to at least one selected from the group consisting of SEQ ID NOS: 1 and 2 prior to modification;

(2) the second deoxyribonucleic acid sequence is at least 98% identical to at least one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification; and

(3) the third deoxyribonucleic acid sequence is at least 98% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification.

54. The mammalian cell according to claim 51, wherein

(1) the first deoxyribonucleic acid sequence is at least 99% identical to at least one selected from the group consisting of SEQ ID NOS: 1 and 2 prior to modification;

(2) the second deoxyribonucleic acid sequence is at least 99% identical to at least one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification; and

(3) the third deoxyribonucleic acid sequence is at least 99% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification.

55. The mammalian cell according to any one of claims 51-54, wherein

- (a) the first DNA cassette comprises a promoter and at least one selected from the group consisting of a selectable marker gene and a reporter gene;
- (b) the second DNA cassette comprises a promoter and at least one selected from the group consisting of a selectable marker gene and a reporter gene; and
- (c) the third DNA cassette comprises a promoter and at least one selected from the group consisting of a selectable marker gene and a reporter gene.

56. The mammalian cell according to claims 55, wherein

- (a) the first DNA cassette comprises a promoter, a selectable marker gene and a reporter gene;
- (b) the second DNA cassette comprises a promoter, a selectable marker gene and a reporter gene; and
- (c) the third DNA cassette comprises a promoter, a selectable marker gene and a reporter gene.

57. The mammalian cell according to claims 55 or 56, wherein the first deoxyribonucleic acid sequence comprises a Stable Integration Site.

58. The mammalian cell according to claim 57, wherein a gene of interest is inserted into the Stable Integration Site.

59. The mammalian cell according to claim 57, wherein the gene of interest encodes a polypeptide of interest selected from the group consisting of antibodies, antibody chains, receptors, Fc-containing proteins, 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.

60. The mammalian cell of claims any one of 51-59, wherein the mammalian cell is a human cell and the first deoxyribonucleic acid sequence is at least 90% identical to SEQ ID NO: 1.

61. The mammalian cell of any one of claims 51-59, wherein the mammalian cell is a CHO cell and the first deoxyribonucleic acid sequence is at least 90% identical to SEQ ID NO: 2.

62. The mammalian cell of claim 60, wherein the mammalian cell is a human cell and the first deoxyribonucleic acid sequence is at least 95% identical to SEQ ID NO: 1.

63. The mammalian cell of claim 61, wherein the mammalian cell is a CHO cell and the first deoxyribonucleic acid sequence is at least 95% identical to SEQ ID NO: 2.

64. The mammalian cell of claim 62, wherein the mammalian cell is a human cell and the first deoxyribonucleic acid sequence is at least 98% identical to SEQ ID NO: 1.

65. The mammalian cell of claim 63, wherein the mammalian cell is a CHO cell and the first deoxyribonucleic acid sequence is at least 98% identical to SEQ ID NO: 2.

66. The mammalian cell of claim 64, wherein the mammalian cell is a human cell and the first deoxyribonucleic acid sequence is at least 99% identical to SEQ ID NO: 1.

67. The mammalian cell of claim 65, wherein the mammalian cell is a CHO cell and the first deoxyribonucleic acid sequence is at least 99% identical to SEQ ID NO: 2.

68. The mammalian cell of any one of claims 61, 63, 65 or 67, wherein the first deoxyribonucleic acid sequence comprises a Stable Integration Site produced using a guide sequence selected from the group consisting of SEQ ID NOS: 13 to 419.

69. The mammalian cell of any one of claims 61, 63, 65 or 67, wherein the first deoxyribonucleic acid sequence comprises a Stable Integration Site produced by using a guide sequence that are complementary to target sequences in SEQ ID NO:2 at nucleotide position ranges selected from the group consisting of: (a) 1 to 2000; (b) 2001 to 4000; (c) 4001 to 6000; (d) 6001 to 8000; (e) 8001 to 10,000; (f) 10,001 to 12,000; (g) 12,001 to 14,000; (h) 14,001 to 16,000; (i) 16,001 to 18,000; (j) 18,001 to 20,000; (k) 20,001 to 22,000; (l) 22,001 to 24,000; (m) 24,001 to 26,000; (n) 26,001 to 28,000; (o) 28,001 to 30,000; (p) 30,001 to 32,000; (q) 32,001 to 34,000; (r) 34,001 to 36,000; (s) 36,001 to 38,000; (t) 38,001 to 40,000; (u) 40,001 to 42,000; and (v) 42,001 to 44,232.

70. A mammalian cell comprising a modified genome,
wherein the modified genome comprises a deoxyribonucleic acid sequence comprising an AAVS1-like region modified by insertion of at least one DNA cassette,
and
wherein a guide sequence selected from the group consisting of SEQ ID NOS: 13 to 419 that is complementary to a sense or antisense strand of the AAVS1-like region.

71. The mammalian cell according to claim 70, further comprising
a second deoxyribonucleic acid sequence that is at least 90% identical to at least one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification;
and

a third deoxyribonucleic acid sequence that is at least 90% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification,

wherein the first deoxyribonucleic acid sequence is modified by insertion of a first DNA cassette, the second deoxyribonucleic acid sequence is modified by insertion of a second DNA cassette, and the third deoxyribonucleic acid sequence is modified by insertion of a third DNA cassette.

72. The mammalian cell of claim 71, wherein the second deoxyribonucleic acid sequence is at least 90% to 99%, 95% to 99% or 98% to 99% identical to at least one selected from the group consisting of one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification; and the third deoxyribonucleic acid sequence is at least 90% to 99%, 95% to 99% or 98% to 99% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification.

73. The mammalian cell of any one of claims 70-72, wherein the first deoxyribonucleic acid sequence comprises a Stable Integration Site produced by using a guide sequence that is complementary to at least one target sequence having at least 50%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

74. The mammalian cell of claim 73, wherein the first deoxyribonucleic acid sequence comprises a Stable Integration Site produced by using a guide sequence that is complementary to at least one target sequence having at least 75%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

75. The mammalian cell of claim 74, wherein the first deoxyribonucleic acid sequence comprises a Stable Integration Site produced by using a guide sequence that is complementary to at least one target sequence having at least 85%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to

16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

76. The mammalian cell of claim 75, wherein the first deoxyribonucleic acid sequence comprises a Stable Integration Site produced by using a guide sequence that is complementary to at least one target sequence having at least 90%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

77. The mammalian cell of claim 76, wherein the first deoxyribonucleic acid sequence comprises a Stable Integration Site produced by using a guide sequence that is complementary to at least one target sequence having at least 95%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to

30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

78. The mammalian cell of claim 77, wherein the first deoxyribonucleic acid sequence comprises a Stable Integration Site produced by using a guide sequence that is complementary to at least one target sequence having at least 98%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

79. A mammalian cell comprising a modified genome, wherein the modified genome comprises a Stable Integration Site in a AAVS1-like region, wherein the Stable Integration Site is produced by using a guide sequence that is complementary to at least one target sequence having at least 50%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to

30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

80. A mammalian cell of claim 79, wherein the modified genome comprises a Stable Integration Site in a AAVS1-like region, wherein the Stable Integration Site is produced by using a guide sequence that is complementary to at least one target sequence having at least 75%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

81. A mammalian cell of claim 80, wherein the modified genome comprises a Stable Integration Site in a AAVS1-like region, wherein the Stable Integration Site is produced by using a guide sequence that is complementary to at least one target sequence having at least 85%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or

(s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

82. A mammalian cell of claim 81, wherein the modified genome comprises a Stable Integration Site in a AAVS1-like region, wherein the Stable Integration Site is produced by using a guide sequence that is complementary to at least one target sequence having at least 90%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

83. A mammalian cell of claim 82, wherein the modified genome comprises a Stable Integration Site in a AAVS1-like region, wherein the Stable Integration Site is produced by using a guide sequence that is complementary to at least one target sequence having at least 95%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

84. A mammalian cell of claim 83, wherein the modified genome comprises a Stable Integration Site in a AAVS1-like region, wherein the Stable Integration Site is produced by using a guide sequence that is complementary to at least one target sequence having at least 98%-99% identity to SEQ ID NO:2 at nucleotide positions:

(a) 1 to 2000; or (b) 2001 to 4000; or (c) 4001 to 6000; or (d) 6001 to 8000; or (e) 8001 to 10,000; or (f) 10,001 to 12,000; or (g) 12,001 to 14,000; or (h) 14,001 to 16,000; or (i) 16,001 to 18,000; or (j) 18,001 to 20,000; or (k) 20,001 to 22,000; or (l) 22,001 to 24,000; or (m) 24,001 to 26,000; or (n) 26,001 to 28,000; or (o) 28,001 to 30,000; or (p) 30,001 to 32,000; or (q) 32,001 to 34,000; or (r) 34,001 to 36,000; or (s) 36,001 to 38,000; or (t) 38,001 to 40,000; or (u) 40,001 to 42,000; or (v) 42,001 to 44,232.

85. A mammalian cell according to any of claims 78 to 84, further comprising
a second deoxyribonucleic acid sequence that is at least 90% identical to at least one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification;
and
a third deoxyribonucleic acid sequence that is at least 90% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification,

wherein the first deoxyribonucleic acid sequence is modified by insertion of a first DNA cassette, the second deoxyribonucleic acid sequence is modified by insertion of a second DNA cassette, and the third deoxyribonucleic acid sequence is modified by insertion of a third DNA cassette.

86. The mammalian cell of claim 85, wherein the second deoxyribonucleic acid sequence at least is 90% to 99%, 95% to 99% or 98% to 99% identical to at least one selected from the group consisting of one selected from the group consisting of SEQ ID NOS: 5 to 10 prior to modification; and the third deoxyribonucleic acid sequence is at least 90% to 99%, 95% to 99% or 98% to 99% identical to at least one selected from the group consisting of SEQ ID NO: 11 and SEQ ID NO: 12 prior to modification.

87. A method of producing protein, wherein the method comprises the steps of:

- (1) culturing mammalian cells according to claims 51-86; and
- (2) harvesting the protein.

88. A cell according to any of the above methods.

89. A method using any of the above cells.