



US 20250263695A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2025/0263695 A1
(43) Pub. Date: Aug. 21, 2025

(54) NOVEL SITES FOR SAFE GENOMIC INTEGRATION AND METHODS OF USE THEREOF

(71) Applicant: BlueRock Therapeutics LP, Cambridge, MA (US)

(72) Inventors: Chew-Li SOH, Cambridge, MA (US); Mark J. TOMISHIMA, Cambridge, MA (US); Conor B. MCAULIFFE, Cambridge, MA (US); Dan Charles WILKINSON, Jr., Cambridge, MA (US); Benjamin BURNETT, Cambridge, MA (US)

(21) Appl. No.: 18/857,388

(22) PCT Filed: Apr. 28, 2023

(86) PCT No.: PCT/US2023/066396

§ 371 (c)(1),

(2) Date: Oct. 16, 2024

Related U.S. Application Data

(60) Provisional application No. 63/336,248, filed on Apr. 28, 2022.

Publication Classification

(51) Int. Cl.

C12N 15/11 (2006.01)

C12N 5/074 (2010.01)

C12N 5/078 (2010.01)

C12N 9/22 (2006.01)

C12N 15/86 (2006.01)

C12N 15/90 (2006.01)

(52) U.S. Cl.

CPC C12N 15/11 (2013.01); C12N 5/0634

(2013.01); C12N 5/0696 (2013.01); C12N

9/224 (2025.05); C12N 9/226 (2025.05);

C12N 15/86 (2013.01); C12N 15/907

(2013.01); C12N 2310/20 (2017.05); C12N

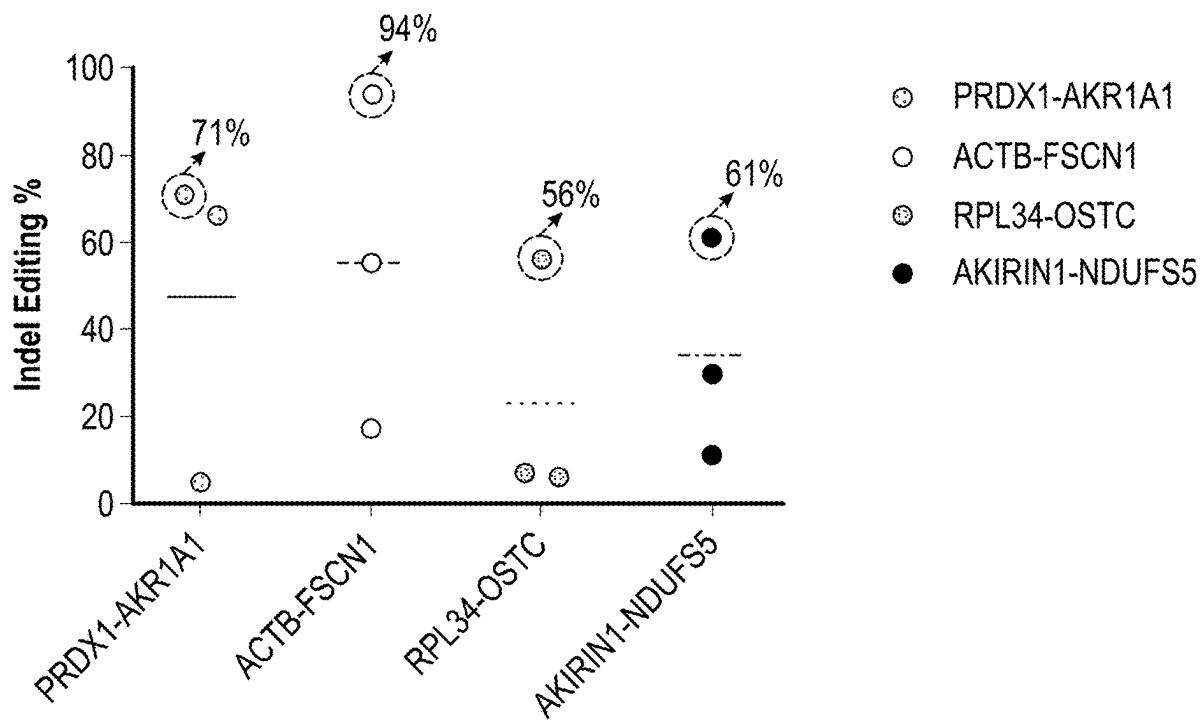
2510/00 (2013.01); C12N 2740/10043

(2013.01); C12N 2740/15043 (2013.01); C12N

2750/14143 (2013.01)

(57) **ABSTRACT**

The present disclosure is directed to genetically modified cells that express one or more transgenes at a sustained expression level from a site for safe genomic integration and stable expression. Also provided are methods of making the cells and nucleic acid vectors that can be used to make the cells.

Specification includes a Sequence Listing.

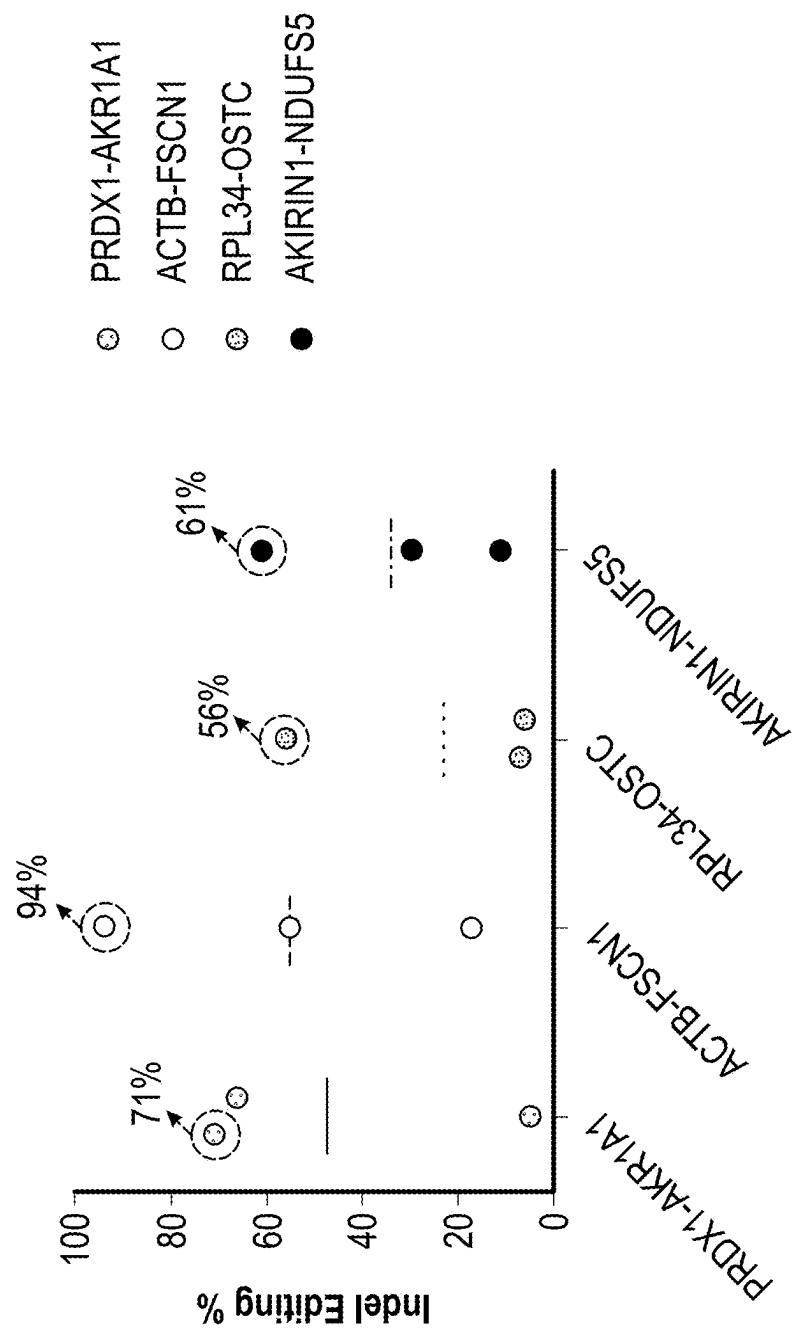


FIG. 1

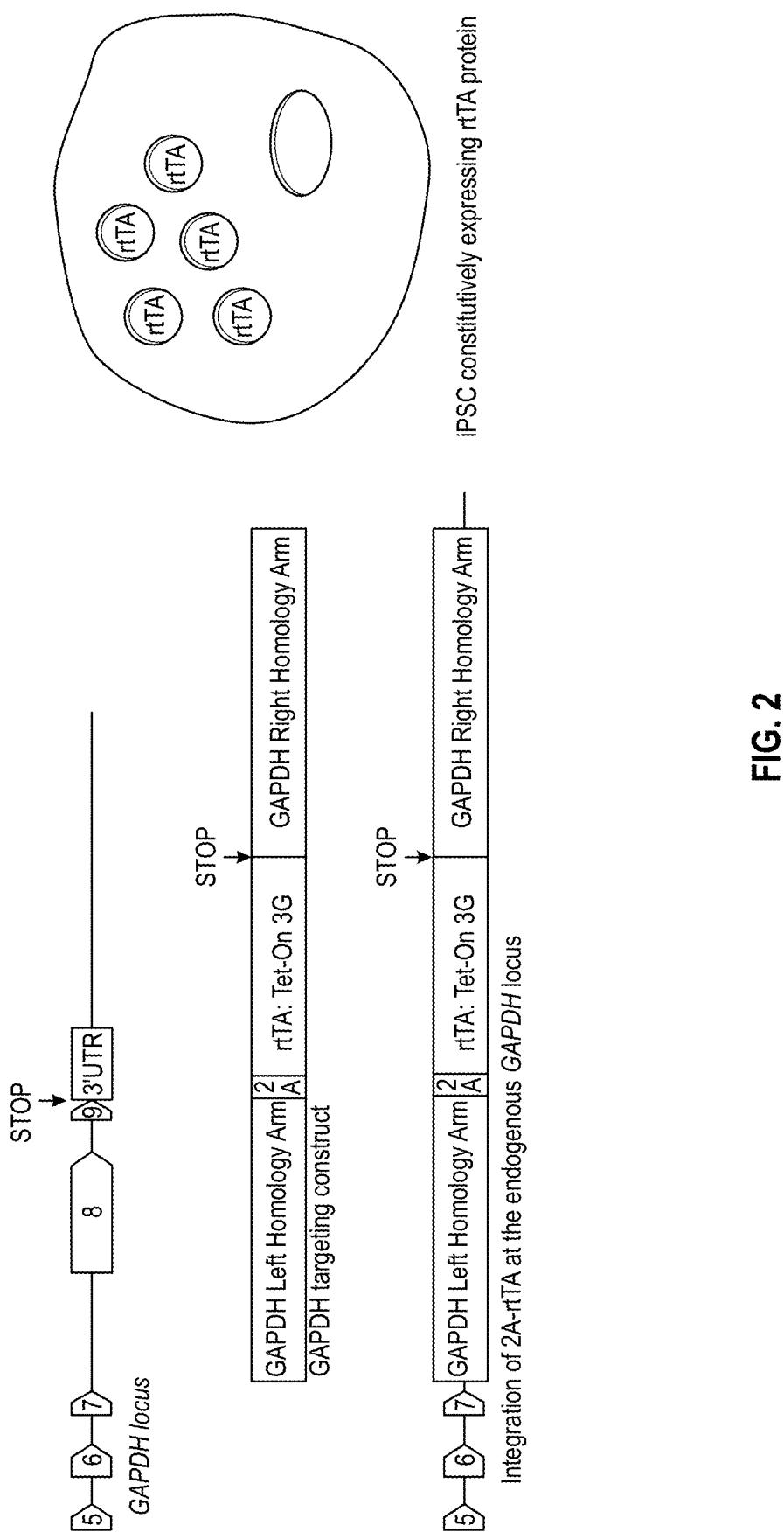


FIG. 2

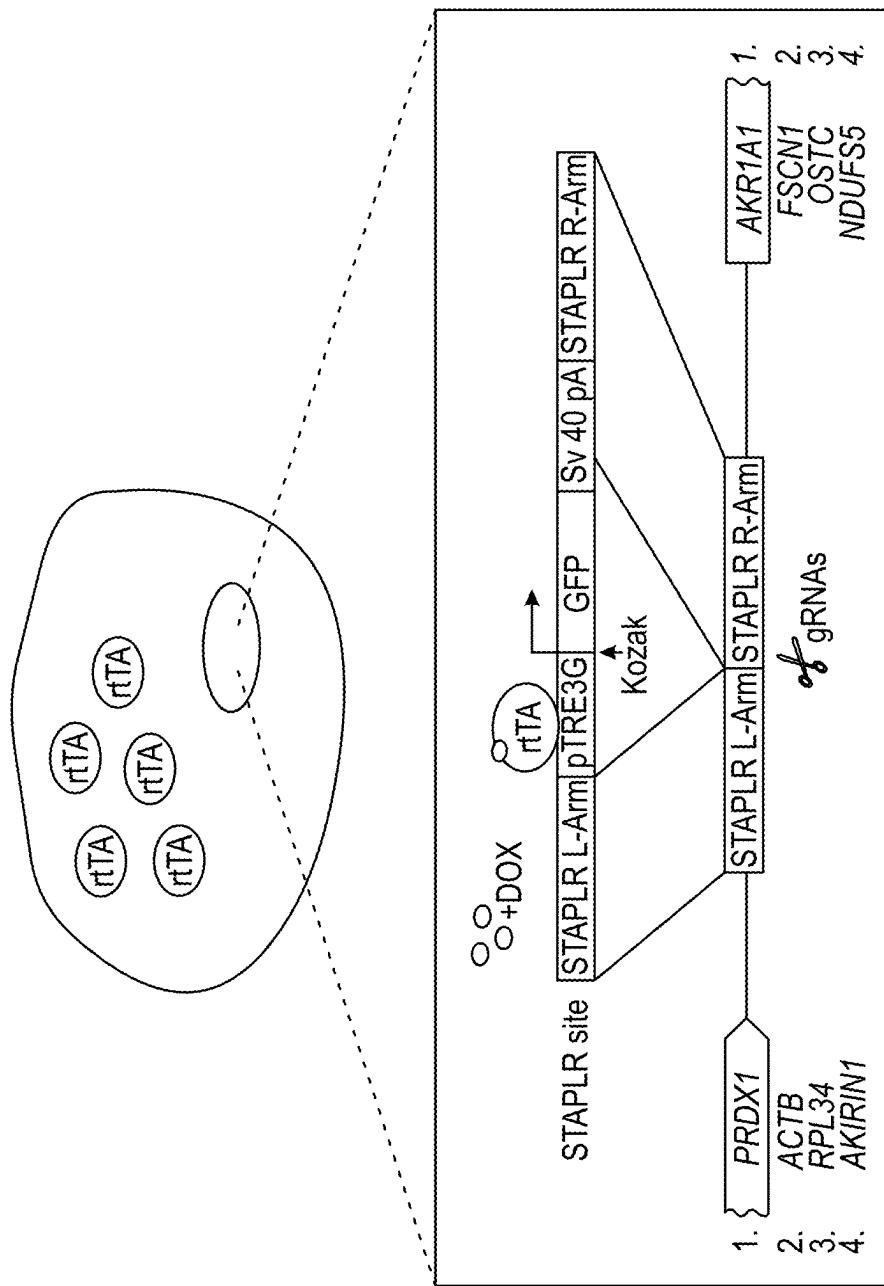


FIG. 3

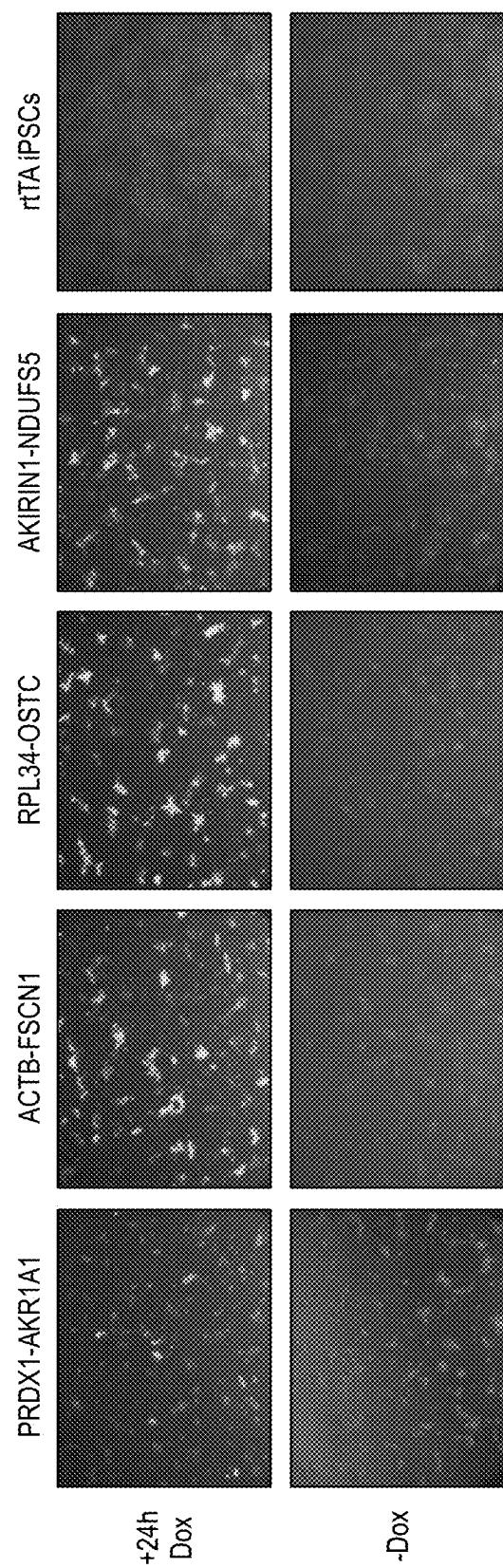


FIG. 4

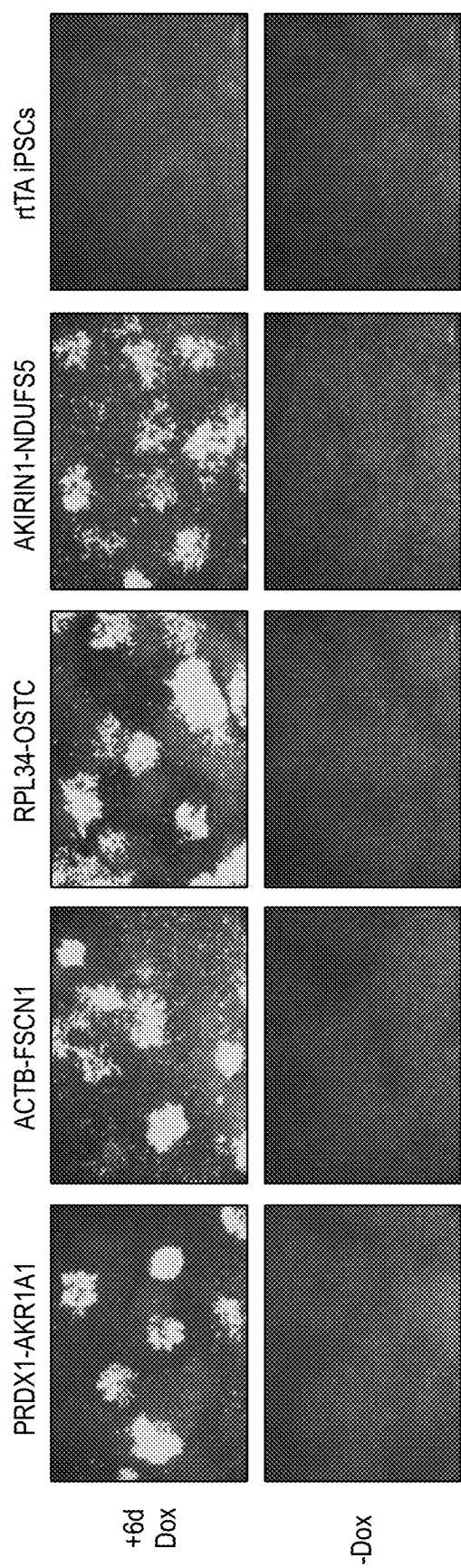


FIG. 5

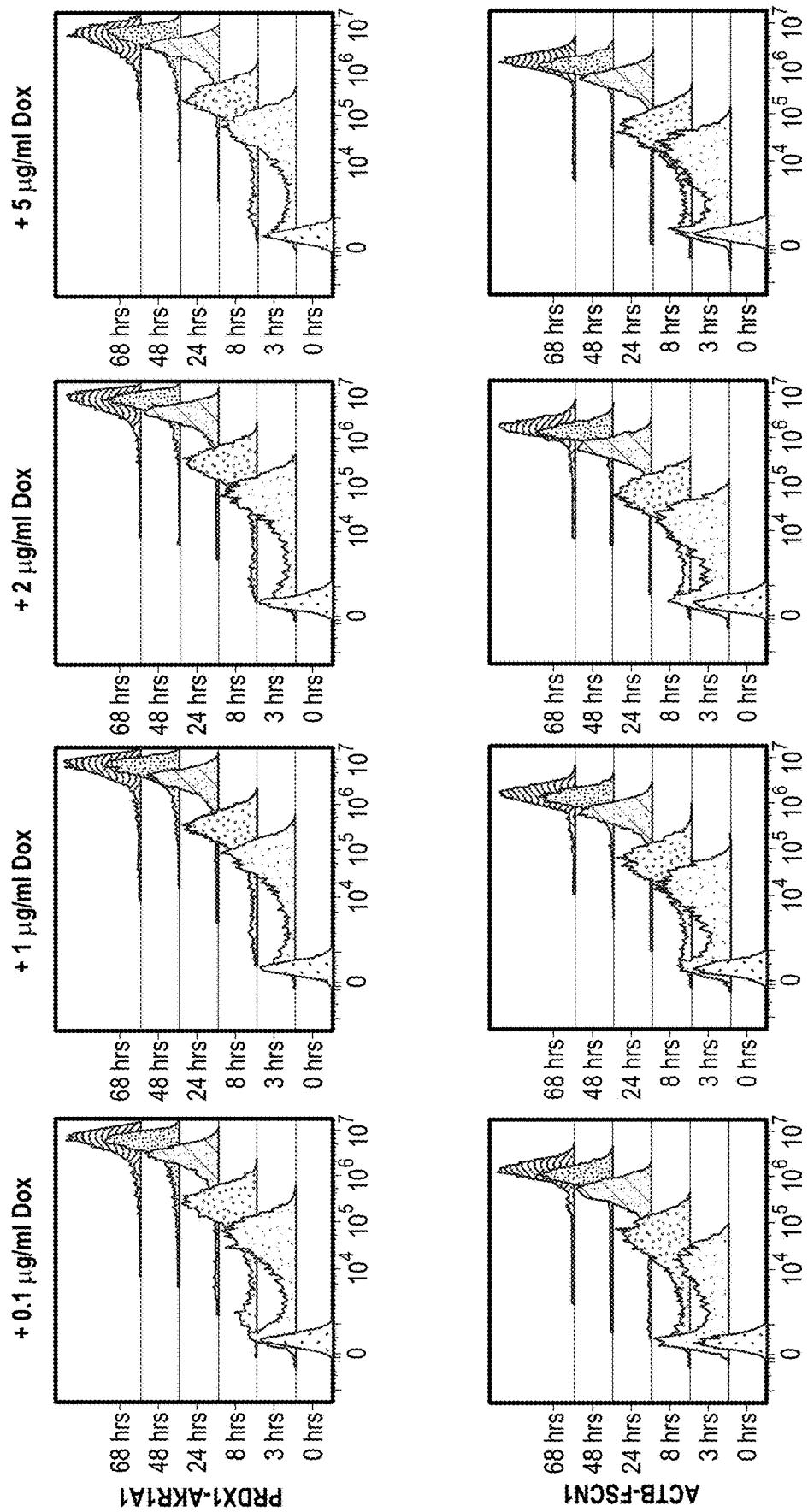


FIG. 6 (cont...)

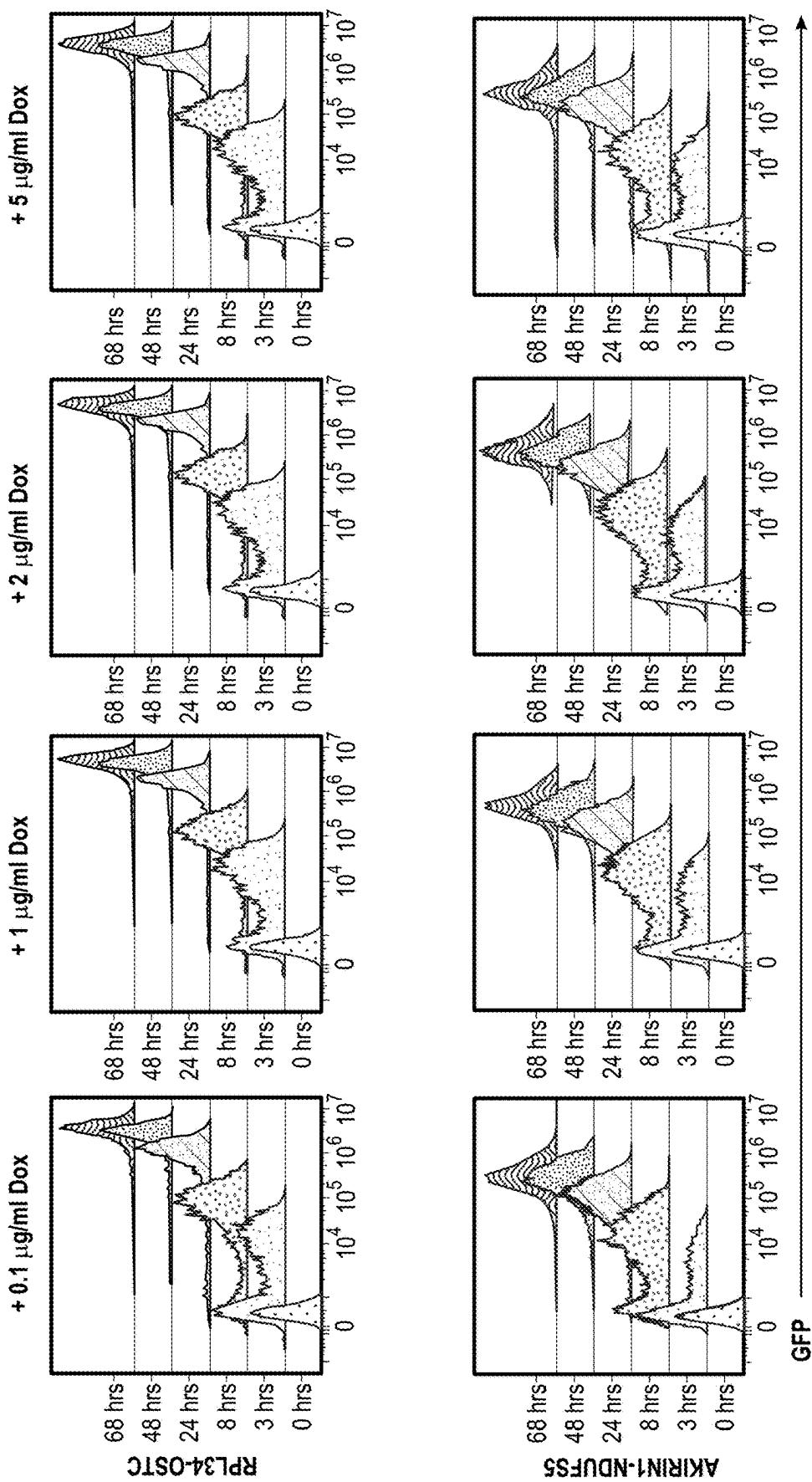


FIG. 6

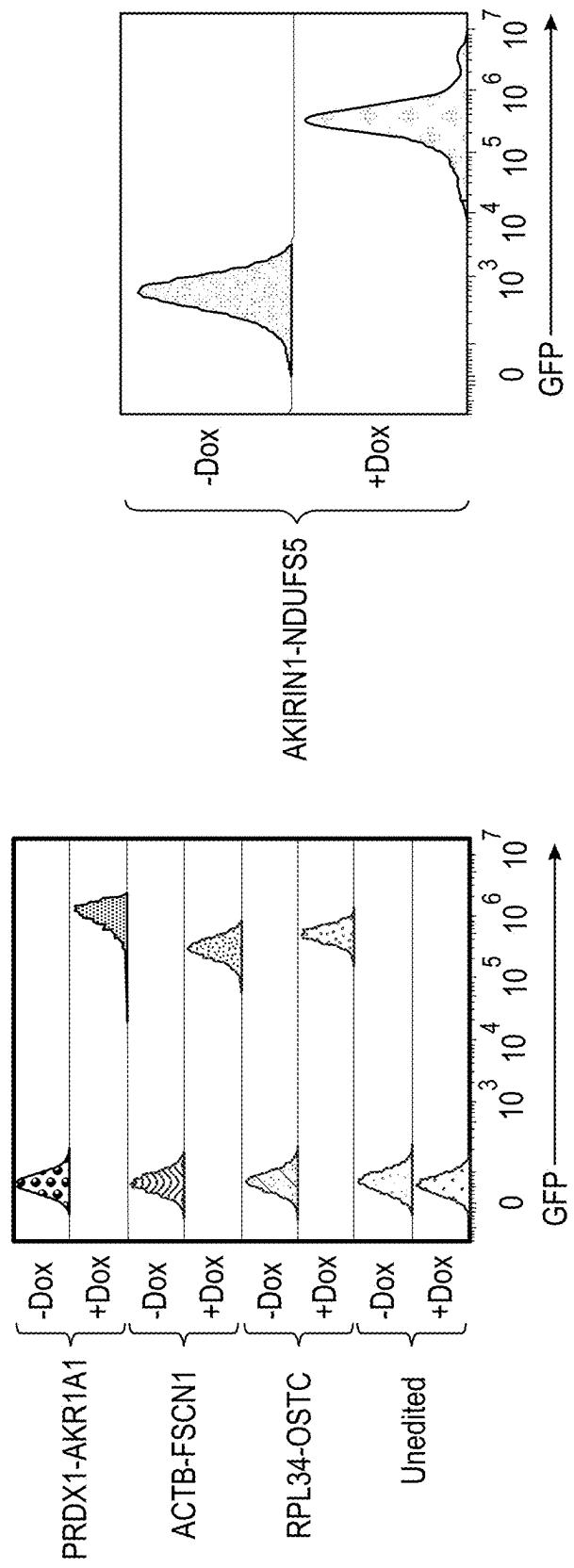


FIG. 7

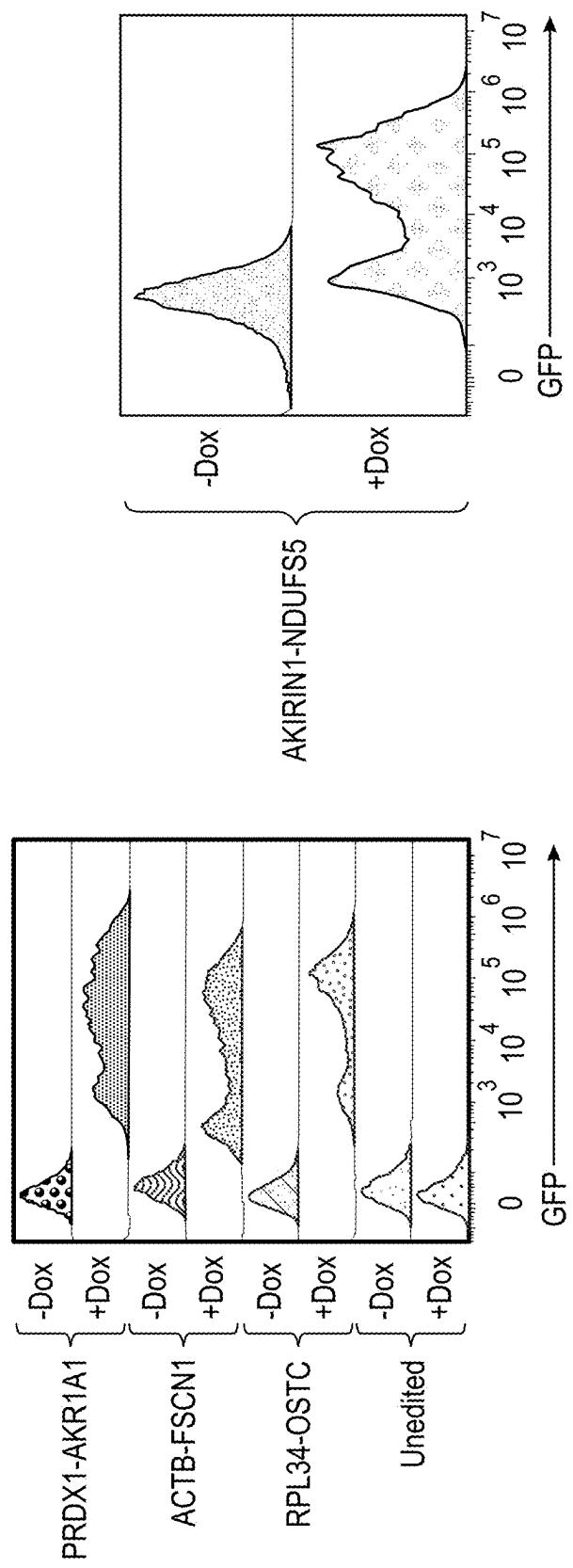


FIG. 8

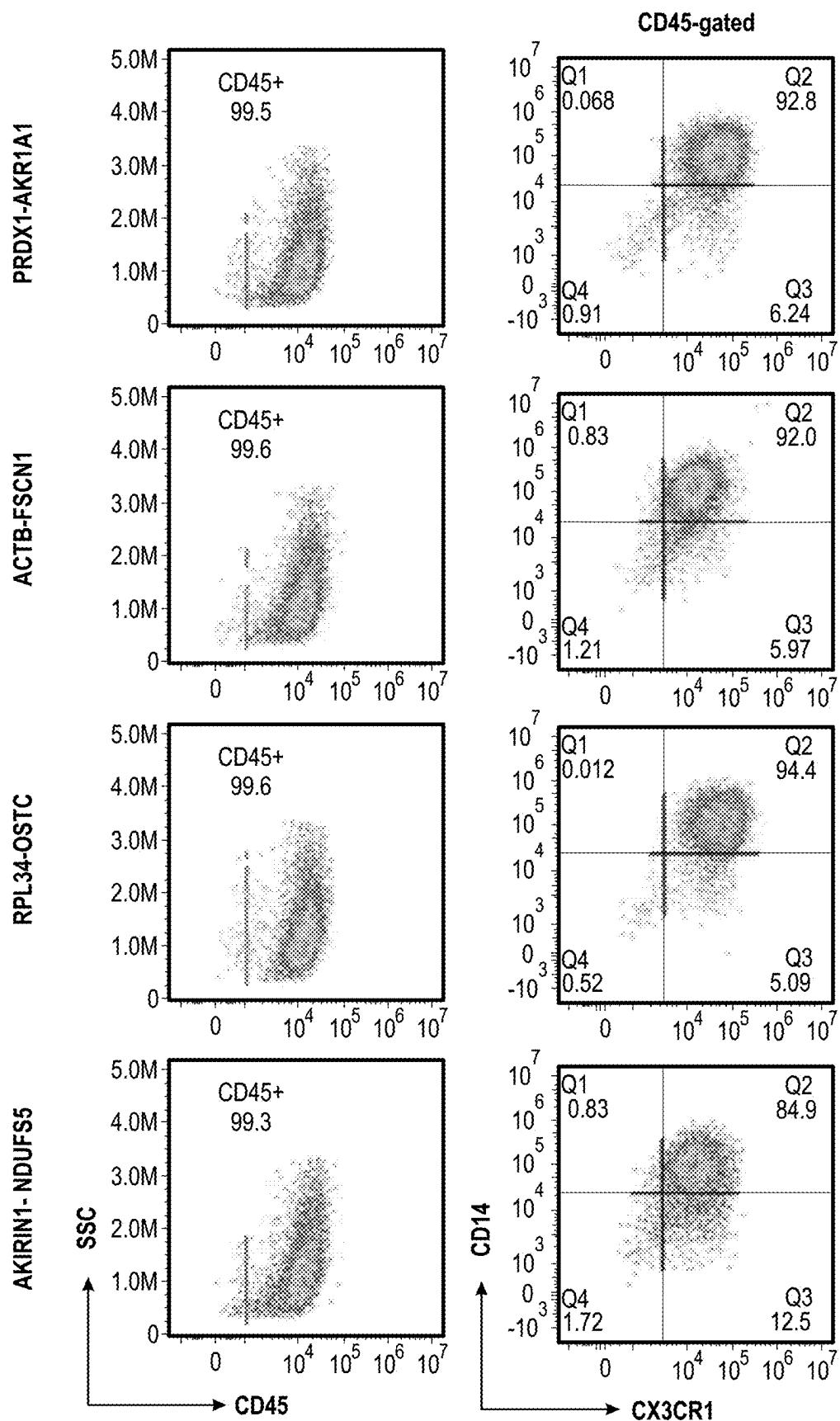


FIG. 9

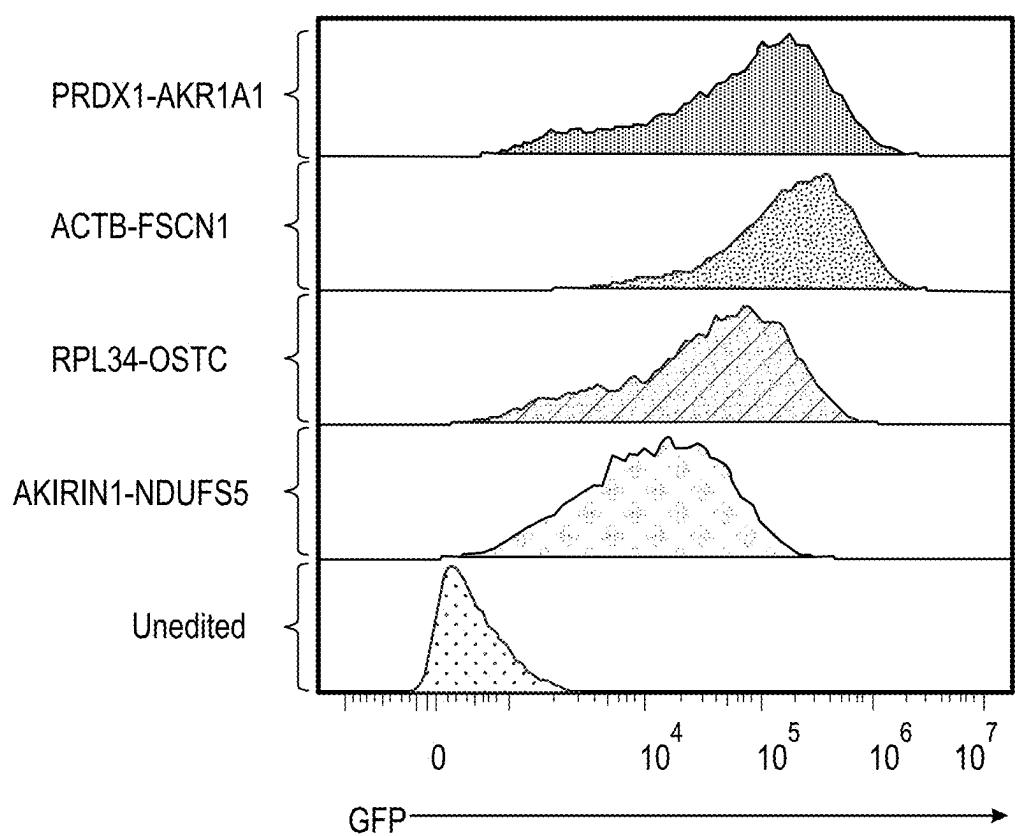


FIG. 10

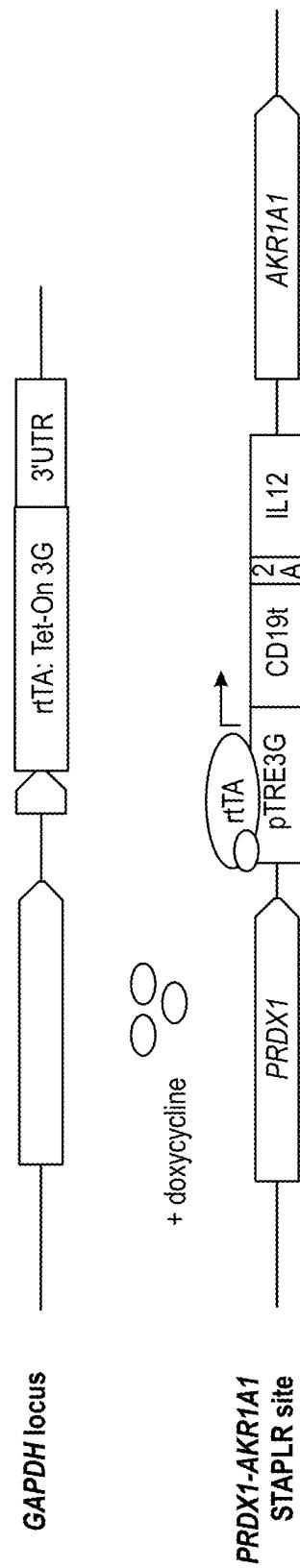


FIG. 11

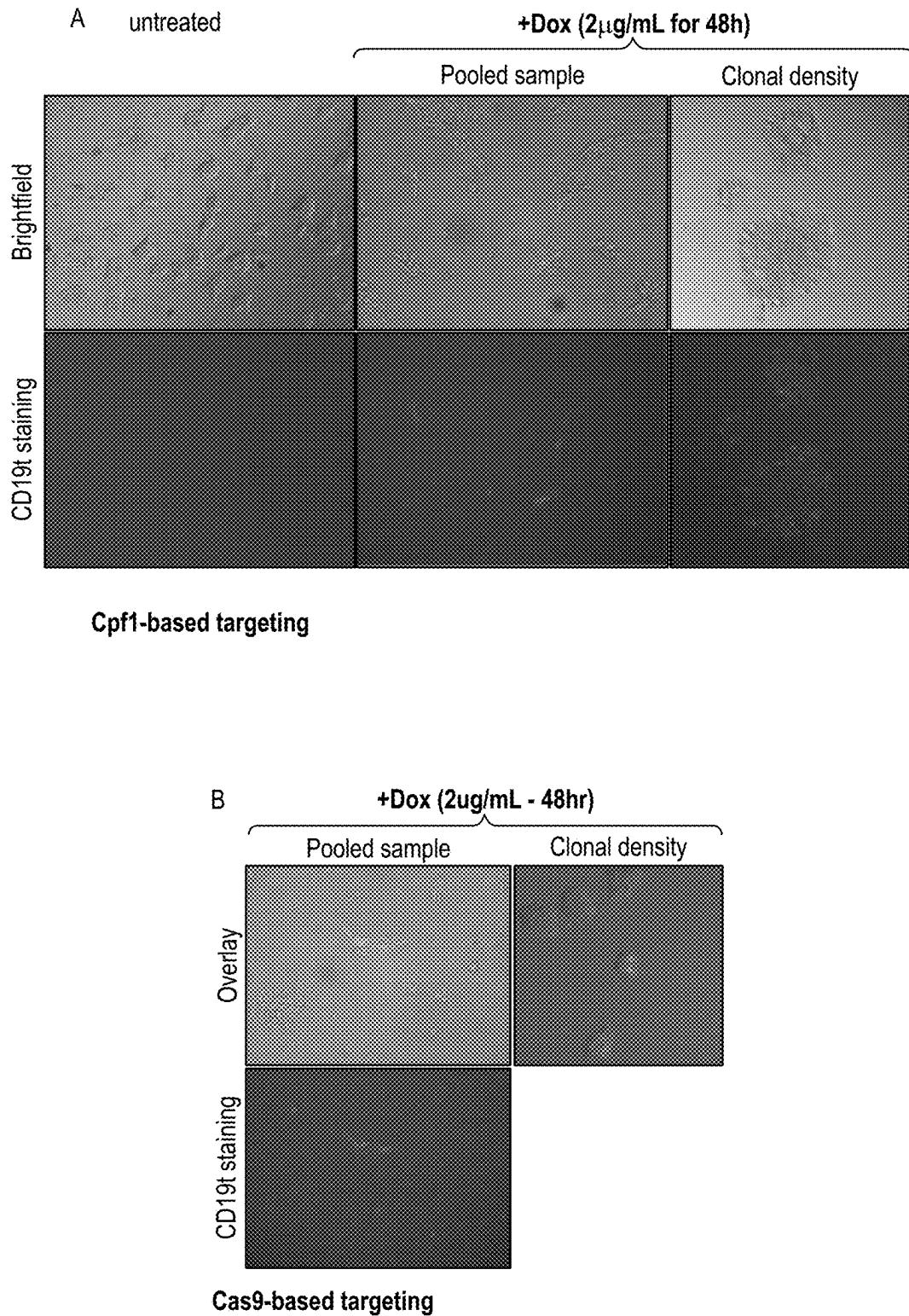


FIG. 12

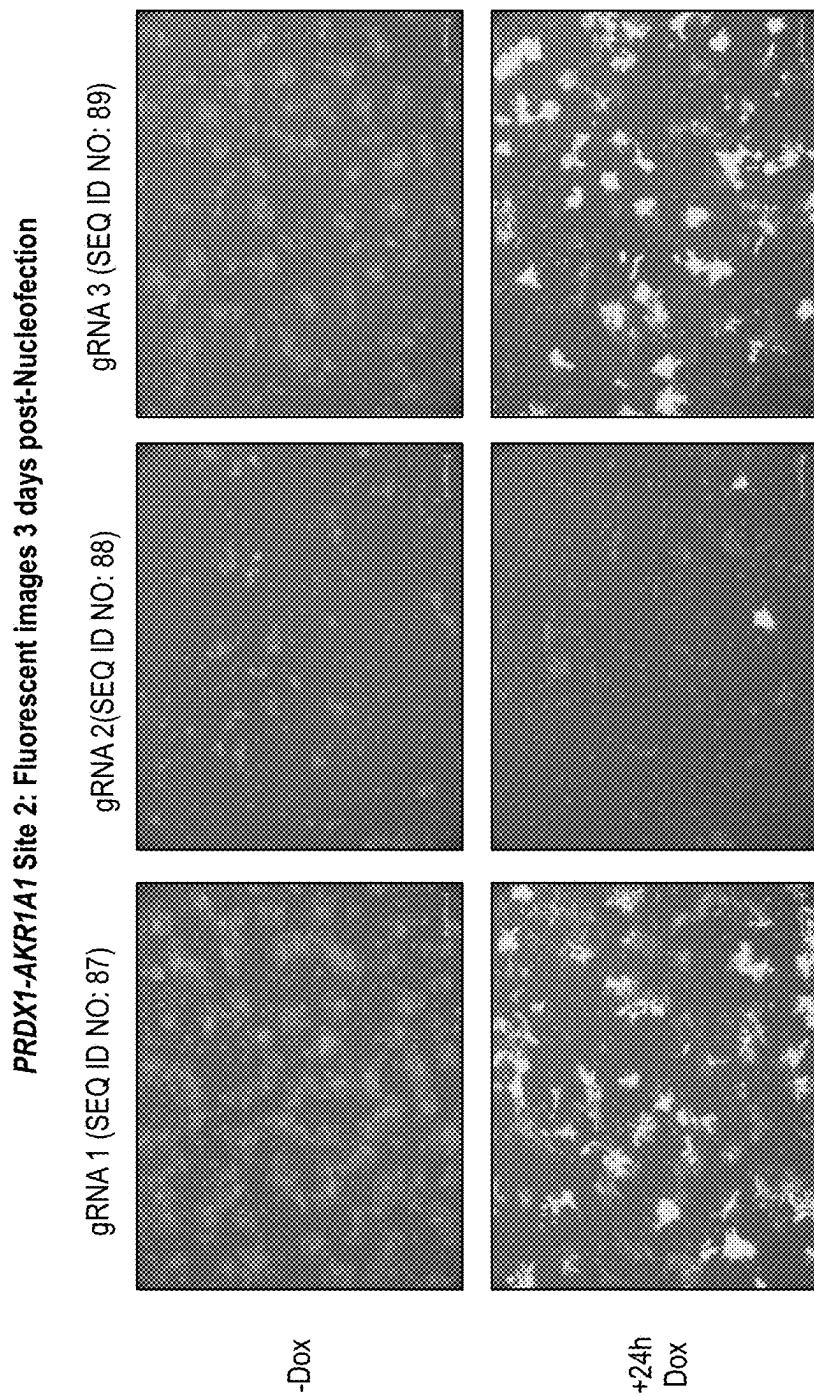


FIG. 13

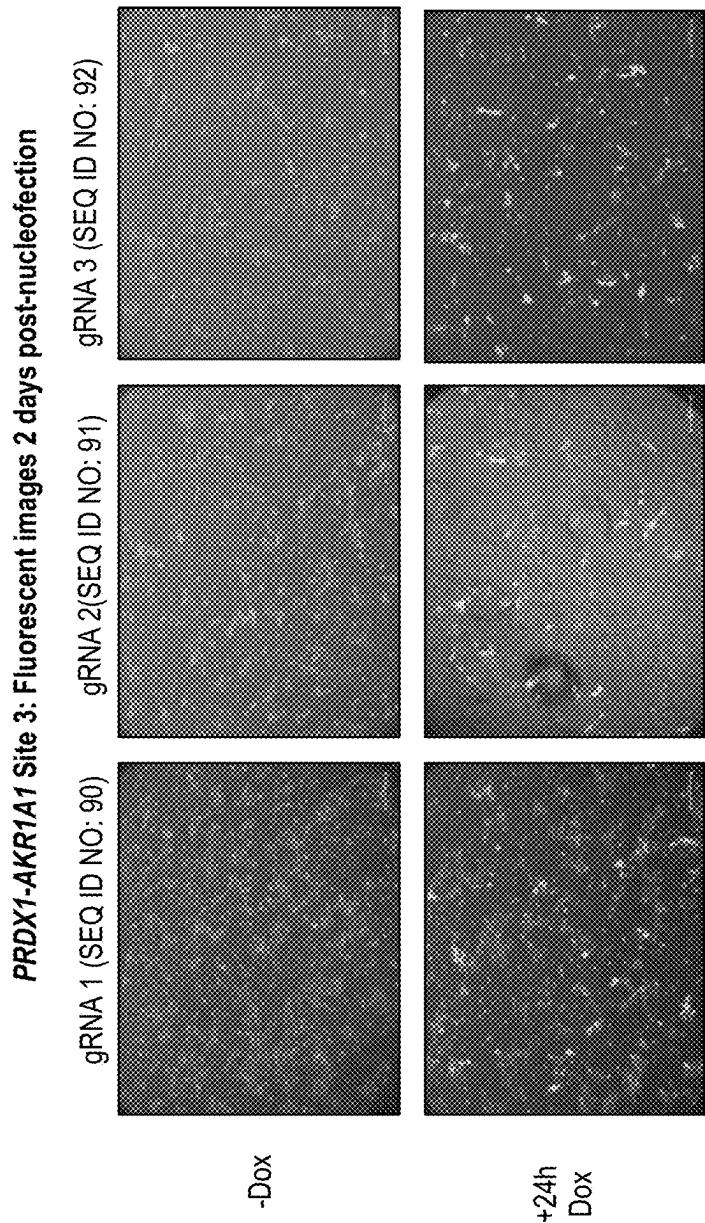


FIG. 14

PRDX1-AKR1A1 Site 2:
Flow cytometric analysis 7 days post-Nucleofection

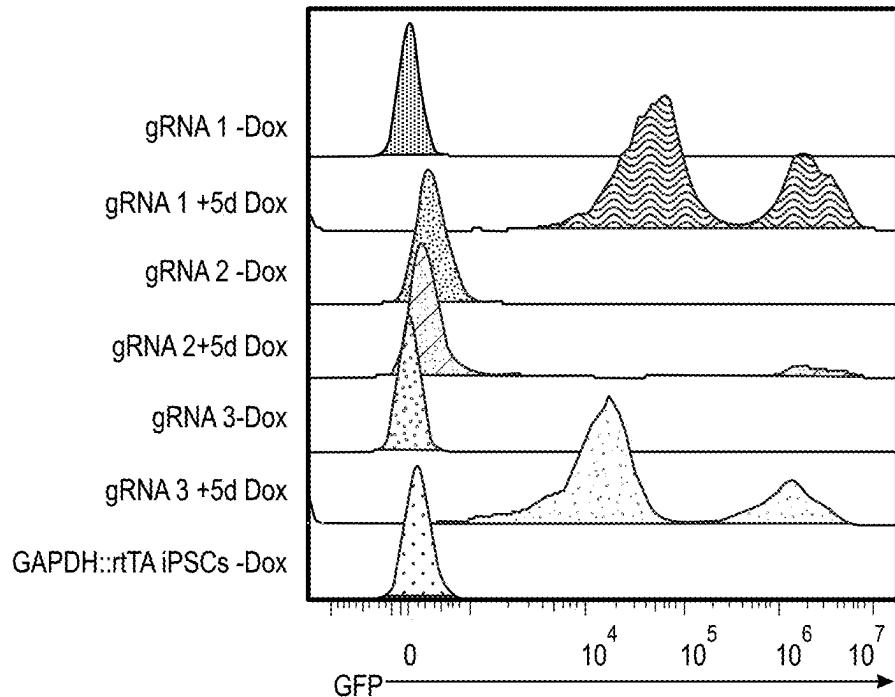


FIG. 15

PRDX1-AKR1A1 Site 3:
Flow cytometric analysis 7 days post-Nucleofection

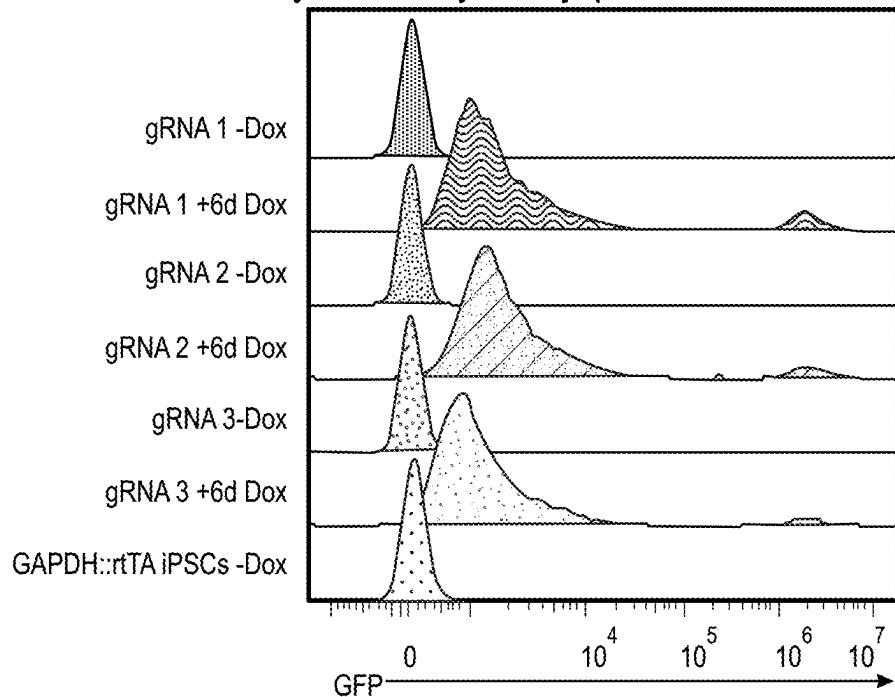


FIG. 16

**NOVEL SITES FOR SAFE GENOMIC
INTEGRATION AND METHODS OF USE
THEREOF**

**CROSS-REFERENCE TO RELATED
APPLICATION**

[0001] This application claims priority from U.S. Provisional Application No. 63/336,248, filed Apr. 28, 2022, the content of which is incorporated herein by reference in its entirety.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing submitted electronically IN XML format and is hereby incorporated by reference in its entirety. The electronic copy of the Sequence Listing, created on Apr. 27, 2023, is named 025450_WO017_SL.xml and is 379,876 bytes in size.

BACKGROUND

[0003] Many efforts to safely integrate transgenes into a genome have been made at so-called “genomic safe harbor” sites. Safe harbor sites in the genome are those where a nucleic acid (e.g., an exogenous gene) can be introduced without disrupting the expression or regulation of adjacent genes, and therefore the normal functioning of the cell. Three genomic sites—AAVS1, CCR5, and ROSA26—are traditionally considered safe harbor sites and have been used in most targeted transgene integrations. AAVS1 is a region for the rare genomic integration of AAV genome and has been found to allow robust expression without disrupting cell function. CCR5 was serendipitously identified because a naturally-occurring CCR5-delta-32 mutation results in an HIV-resistant phenotype; the disposability of the gene makes it an ideal integration site. The ROSA26 locus was originally identified in mouse embryonic stem cells through a lentiviral gene trap approach.

[0004] While these genomic safe harbor sites allow robust transgene expression under a given cell context, they may not support faithful transgene expression in other cell lineages or after a change in cell state. This is because reciprocal interactions between a transgene and the host cell’s genomic context can affect the expression of the transgene, leading to attenuation or complete silencing of transgene expression (e.g., through DNA methylation). More critically, these sites of genomic integration may also affect the expression of endogenous genes in the vicinity of the insertion site, thus affecting normal host cell function.

SUMMARY OF THE DISCLOSURE

[0005] The present disclosure is based, at least in part, on the identification of intergenic sites in the genome that remain transcriptionally active in different cell types and under different cell states, including maturation phases, such that an exogenous nucleotide sequence of interest (e.g., a transgene encoding a protein or an RNA) integrated therein remains expressed and functional as the cell undergoes proliferation and cell state changes.

[0006] Accordingly, in one aspect, the present disclosure provides a genetically modified cell, e.g., a mammalian (e.g., human) cell, comprising an exogenous nucleotide sequence integrated in a sustained transcriptionally active payload region (STAPLR) in the genome of the cell, wherein the STAPLR is selected from the group consisting of the

intergenic region between the RPL34 gene and the OSTC gene; the intergenic region between the ACTB gene and the FSCN1 gene; the intergenic region between the AKIRIN1 gene and the NDUFS5 gene; the intergenic region between the PRDX1 gene and the AKR1A1 gene; the intergenic region between the PTGES3 gene and the NACA gene; the intergenic region between the MLF2 gene and the PTMS gene; the intergenic region between the RAB13 gene and the RPS27 gene; the intergenic region between the JTB gene and the RAB13 gene; the intergenic region between the AKR1A1 gene and the NASP gene; the intergenic region between the NDUFS5 gene and the MA CF1 gene; the intergenic region between the SRSF9 gene and the DYNLL1 gene; the intergenic region between the MYL6B gene and the MYL6 gene; the intergenic region between the GPX1 gene and the RHOA gene; the intergenic region between the HNRNPA2B1 gene and the CBX3 gene; the intergenic region between the ROMO gene and the RBM39 gene; the intergenic region between the PA2G4 gene and the RPL41 gene; and the intergenic region between the NDUFB10 gene and the RPS2 gene.

[0007] In some embodiments, the intergenic region between the RPL34 gene and the OSTC gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 1, or a nucleotide sequence sufficiently similar to SEQ ID NO: 1 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0008] In some embodiments, the intergenic region between the ACTB gene and the FSCN1 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 2, or a nucleotide sequence sufficiently similar to SEQ ID NO: 2 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0009] In some embodiments, the intergenic region between the AKIRIN1 gene and the NDUFS5 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 3, or a nucleotide sequence sufficiently similar to SEQ ID NO: 3 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0010] In some embodiments, the intergenic region between the PRDX1 gene and the AKR1A1 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 4, or a nucleotide sequence sufficiently similar to SEQ ID NO: 4 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0011] In some embodiments, the intergenic region between the PTGES3 gene and the NACA gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 5, or a nucleotide sequence sufficiently similar to SEQ ID NO: 5 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0012] In some embodiments, the intergenic region between the MLF2 gene and the PTMS gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 6, or a nucleotide sequence sufficiently similar to

SEQ ID NO: 6 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0013] In some embodiments, the intergenic region between the RAB13 gene and the RPS27 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 7, or a nucleotide sequence sufficiently similar to SEQ ID NO: 7 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0014] In some embodiments, the intergenic region between the JTB gene and the RAB13 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 8, or a nucleotide sequence sufficiently similar to SEQ ID NO: 8 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0015] In some embodiments, the intergenic region between the AKR1A1 gene and the NASP gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 9, or a nucleotide sequence sufficiently similar to SEQ ID NO: 9 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0016] In some embodiments, the intergenic region between the NDUFS5 gene and the MACF1 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 10, or a nucleotide sequence sufficiently similar to SEQ ID NO: 10 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0017] In some embodiments, the intergenic region between the SRSF9 gene and the DYNLL1 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 11, or a nucleotide sequence sufficiently similar to SEQ ID NO: 11 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0018] In some embodiments, the intergenic region between the MYL6B gene and the MYL6 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 12, or a nucleotide sequence sufficiently similar to SEQ ID NO: 12 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0019] In some embodiments, the intergenic region between the GPX1 gene and the RHOA gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 13, or a nucleotide sequence sufficiently similar to SEQ ID NO: 13 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0020] In some embodiments, the intergenic region between the HNRNPA2B1 gene and the CBX3 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 14, or a nucleotide sequence sufficiently similar to SEQ ID NO: 14 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0021] In some embodiments, the intergenic region between the ROMO gene and the RBM39 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 15, or a nucleotide sequence sufficiently similar to SEQ ID NO: 15 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0022] In some embodiments, the intergenic region between the PA2G4 gene and the RPL41 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 16, or a nucleotide sequence sufficiently similar to SEQ ID NO: 16 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0023] In some embodiments, the intergenic region between the NDUFB10 and the RPS2 gene comprises a nucleotide sequence at least 95% (e.g., at least 96%, at least 97%, at least 98%, at least 99%, or 100%) identical to SEQ ID NO: 16, or a nucleotide sequence sufficiently similar to SEQ ID NO: 97 as necessary for the function of the sequence to remain intact (e.g., without adverse effects on the cell).

[0024] Also provided herein are methods of generating these genetically modified mammalian cells, as well as DNA constructs for introducing nucleotide sequences of interest into the novel genomic integration sites herein. Accordingly, in one aspect, the present disclosure provides a method for modifying a mammalian cell, comprising integrating a nucleotide sequence of interest (i.e., an exogenous nucleotide sequence) into a STAPLR described herein. In some embodiments, the integrating step is performed by using a CRISPR/Cas system; a Cre/Lox system; a FLP-FRT system; a TALEN system; a ZFN system; homing endonucleases; random integration; homologous recombination; a transposase; or a non-nuclease-dependent viral vector, optionally selected from a retroviral vector, an adeno-associated viral (AAV) vector, and a lentiviral vector. In further embodiments, the CRISPR/Cas system comprising a guide RNA, and wherein the STAPLR is the intergenic region between (i) the RPL34 gene and the OSTC gene and the gRNA is selected from SEQ ID NOs: 25-32, (ii) the ACTB gene and the FSCN1 gene and the gRNA is selected from SEQ ID NOs: 33-54, (iii) the AKIRIN1 gene and the NDUFS5 gene and the gRNA is selected from SEQ ID NOs: 55-70, or (iv) the PRDX1 gene and the AKR1A1 gene and the gRNA is selected from SEQ ID NOs: 71-92.

[0025] In some embodiments, the CRISPR/Cas system comprises a gRNA-dependent nuclease of type I, type II, type III, type IV, or type V, or a variant thereof. In further embodiments, the CRISPR/Cas system comprises a gRNA-dependent nuclease selected from the group consisting of Cas9, Cpf1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas12, Cas13, Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, CasX, CasY, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, CasPhi, MAD7, and Csf4.

[0026] In another aspect, the present disclosure provides a DNA molecule comprising a nucleotide sequence of interest flanked by a 5' homologous region (HR) and a 3' HR, wherein the 5' and 3' HRs are at least 85% (e.g., at least 90,

95, 96, 97, 98, or 99%) homologous, or 100% identical, to a first genomic region (GR) and a second GR, respectively, in a STAPLR described herein. In some embodiments, each of the 5' and 3' HRs is independently about at least 50, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, at least 500, at least 550, at least 600, at least 650, at least 700, at least 750, at least 800, at least 850, at least 900, at least 950, at least 1000, at least 1100, at least 1200, at least 1300, at least 1400, at least 1500, at least 1600, at least 1700, at least 1800, at least 1900, or at least 2000 base pairs long. In some embodiments, the HRs are each 200 to 2000 (e.g., 300 to 2500, 400 to 2000, or 500 to 1500) base pairs long. In further embodiments, the 5' and 3' HRs are at least 90% (e.g., at least 95%) homologous to SEQ ID NOs: 17 and 18, SEQ ID NOs: 19 and 20, SEQ ID NOs: 21 and 22, SEQ ID NOs: 23 and 24, SEQ ID NOs: 93 and 94, or SEQ ID NOs: 95 and 96, respectively.

[0027] In some embodiments, the exogenous nucleotide sequence or the nucleotide sequence of interest comprises a transgene. In further embodiments, the transgene comprises a coding sequence (e.g., for a protein or an RNA) and one or more regulator elements. In some embodiments, the one or more regulator elements include a constitutive or inducible promoter directing the transcription of the coding sequence. In some embodiments, the transgene encodes a therapeutic protein (e.g., a protein the deficiency or defectiveness of which leads to a disease such as a genetic disease; a cytokine; or a recombinant antigen receptor); a cellular marker; or a protein that regulates the differentiation state or activity of the cell (e.g., a reprogramming factor). In some embodiments, the transgene encodes SOX10, IL-10, IL-12, CD19t, or ThPOK.

[0028] In some embodiments of the present disclosure, the mammalian cell is a human cell. In some embodiments, the mammalian cell (e.g., human cell) is a pluripotent stem cell (PSC; e.g., an induced PSC (iPSC) or an embryonic stem cell (ESC)). In some embodiments, the mammalian cell (e.g., human cell) is a) a cell in the immune system (e.g., a T cell, a natural killer cell, a dendritic cell, a macrophages/monocyte, or a hematopoietic progenitor or precursor cell thereof); b) a cell in the cardiovascular system (e.g., a ventricular cardiomyocyte, a nodal cell, or a cardiac progenitor or precursor cell thereof); c) a cell in the metabolic system (e.g., a hepatocyte or a pancreatic beta-cell, or a progenitor or precursor cell thereof); d) a cell in the central nervous system (e.g., a sensory neuron, a motor neuron, an interneuron, a microglial cell, an oligodendrocyte, or a progenitor or precursor cell thereof); e) a muscle cell (e.g., a skeletal muscle cell or a smooth muscle cell, or a progenitor or precursor cell thereof); f) an adipose cell or a progenitor or precursor cell thereof; or g) a cell in the ocular system (e.g., a retinal pigment epithelium cell, a photoreceptor cell, or a progenitor or precursor cell thereof). Additional cell types of the present disclosure include those described below.

[0029] Also provided herein are pharmaceutical compositions comprising the genetically engineered cells herein and a pharmaceutically acceptable carrier, and gene editing systems comprising the DNA molecule as disclosed herein and the requisite gene editing system for incorporating the nucleotide sequence of interest on the DNA molecule (e.g., a nuclease and gRNA) into the STAPLR.

[0030] In another aspect, the present disclosure provides a method for identifying a sustained transcriptionally active payload region (STAPLR) in the genome of a mammalian cell, the method comprising: (i) performing single cell RNA sequencing analysis on a set of two or more mammalian cell types, wherein the sequencing analysis assigns a unique transcriptome to each cell type; (ii) assigning a Prevalence Score to a constituent gene in the transcriptome, wherein the Prevalence Score represents the fraction of the mammalian cell types containing at least one transcript of the gene in the set of mammalian cell types; (iii) identifying the constituent gene's neighboring gene(s) in the mammalian cell's genome, wherein the neighboring gene(s) do not overlap with the constituent gene; (iv) determining a Neighbor Score for pairs of non-overlapping genes or for regions comprising three or more genes identified in step (iii), wherein the Neighbor Score is the product of the Prevalence Scores of the individual genes in a pair or in a region; (v) ranking the Neighbor Scores; and (vi) selecting a pair of non-overlapping genes or a region comprising three or more non-overlapping genes based on a high ranking, thereby identifying the intergenic region between genes of the selected pair or region as a STAPLR. In some embodiments, the method further comprises (vii) selecting a targetable intergenic subregion in the STAPLR; and (viii) inserting a transgene at the selected subregion, wherein transcription of the transgene or gene circuit is sustained. In some embodiments, the targetable subregion comprises: no known promoter or enhancer regions, a minimal number of conserved regions, repetitive regions, epigenetic marks, and/or enzymatic hypersensitivity regions, and/or the nuclease is a CRISPR nuclease. In some embodiments, wherein the intergenic region is at least 30 (e.g., at least 40, at least 50, at least 75, or at least 100) base pairs in length, and/or does not comprise or comprises a minimal number of promoter regions, a CpG Island, an H3K4MeT epigenetic mark, an H3K4Me3 epigenetic mark, an H3K27Ac epigenetic mark, a DNase I hypersensitivity region, a conserved region, or a repetitive region.

[0031] Other features, objectives, and advantages of the invention are apparent in the detailed description that follows. It should be understood, however, that the detailed description, while indicating embodiments and aspects of the invention, is given by way of illustration only, not limitation. Various changes and modification within the scope of the invention will become apparent to those skilled in the art from the detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0032] FIG. 1 is a dot plot showing the indel editing percentage obtained after Sanger sequencing examination using Synthego's ICE Analysis Tool. For each different STAPLR site, three different gRNAs were tested and the gRNA with the highest indel editing percentage is encircled. The solid horizontal line indicates the mean indel editing percentage of three different gRNAs per STAPLR site.

[0033] FIG. 2 is a diagram illustrating integration of a sequence coding for a 2A peptide and a sequence coding for the Tet-On 3G version of rtTA at the GAPDH locus. Left and right homology arms were designed to enable in-frame integration of the transgene immediately 5' to the STOP codon of GAPDH. This permits expression of rtTA under

endogenous GAPDH promoter control. iPSCs that have been edited with the targeting construct constitutively express the rtTA protein.

[0034] FIG. 3 is a diagram illustrating integration of each of the four STAPLR targeting constructs comprising the pTRE3G-eGFP-Sv40 transgene flanked by left and right homology arms at each STAPLR site in iPSCs constitutively expressing the rtTA protein. The addition of doxycycline allows binding of the rtTA protein and activation of GFP expression from the TRE3G promoter.

[0035] FIG. 4 is a panel of fluorescent microscope images depicting the expression of GFP in a pooled population of cells that had received doxycycline for 24 hours. The doxycycline was added to media 24 hours after Nucleofection of iPSCs with the STAPLR targeting construct and corresponding RNP. No GFP was observed in cells that did not receive doxycycline. Control iPSCs constitutively expressing rtTA that were treated with doxycycline but were not nucleofected with the STAPLR targeting construct and RNP also did not express GFP.

[0036] FIG. 5 is a panel of fluorescent microscope images depicting the expression of GFP in a pooled population of cells that had received doxycycline for 6 days. The doxycycline was added to media 24 hours after Nucleofection of iPSCs with the STAPLR targeting construct and corresponding RNP. No GFP was observed in cells that did not receive doxycycline. Control iPSCs constitutively expressing rtTA that were treated with doxycycline but were not nucleofected with the STAPLR targeting construct and RNP also did not express GFP.

[0037] FIG. 6 is a panel of flow cytometric histograms depicting induction of GFP expression in four different clonally-derived STAPLR iPSC lines over time under different concentrations of doxycycline. Cells were collected for analysis after 0, 3, 8, 24, 48 and 68 hours of doxycycline administration.

[0038] FIG. 7 is a panel of flow cytometric histograms depicting induction of GFP expression after treatment with 2 µg/ml doxycycline in four different clonally-derived STAPLR iPSC lines over time. The left panel shows the PRDX1-AKR1A1, ACTB-FSCN1, and RPL34-OSTC STAPLR lines and a wildtype unedited iPSC control line either without doxycycline treatment or with doxycycline treatment for 72 hours. The right panel shows the AKIRIN1-NDUFS5 STAPLR line either without doxycycline treatment or with doxycycline treatment for 6 days.

[0039] FIG. 8 is a panel of flow cytometric histograms depicting induction of GFP expression after treatment with 2 µg/ml doxycycline in four different clonally-derived STAPLR iPSC lines differentiated into myeloid progenitor cells. Doxycycline was added to the culture medium at day 12 of differentiation. The left panel shows the PRDX1-AKR1A1, ACTB-FSCN1, and RPL34-OSTC STAPLR lines and a wildtype unedited iPSC control line after 15 days of myeloid differentiation either without doxycycline treatment or with doxycycline treatment for 72 hours. The right panel shows the AKIRIN1-NDUFS5 STAPLR line after 18 days of myeloid differentiation either without doxycycline treatment or with doxycycline treatment for 6 days.

[0040] FIG. 9 is a panel of flow cytometric dot plots showing expression of the myeloid progenitor markers CD45, CD14 and CX3CR1 in the non-adherent myeloid population of STAPLR-targeted iPSC lines that had been

differentiated past 30 days. The CD14 and CX3CR1 panel of cells was gated on CD45-positive cells.

[0041] FIG. 10 is a panel of flow cytometric histograms depicting induction of GFP expression in non-adherent myeloid progenitor cells after treatment with 2 µg/ml doxycycline in four differentiated clonally-derived STAPLR iPSC lines and a wildtype unedited iPSC control line. Doxycycline was added to the culture medium after day 30 of differentiation for six days.

[0042] FIG. 11 is a diagram illustrating integration of a targeting construct comprising the pTRE3G-CD19t-IL12 transgene flanked by left and right homology arms to allow integration at the PRDX1-AKR1A1 STAPLR site. This construct was transfected in iPSCs constitutively expressing the rtTA protein from the GAPDH endogenous promoter.

[0043] FIG. 12 is a panel of photographs showing live cell imaging of CD19t (truncated to prevent intracellular signal transduction) staining after 48 h of treatment with 2 µg/mL doxycycline either in a pooled sample of cells post-targeting with the PRDX1-AKR1A1 pTRE3G-CD19t-IL12 donor template, or in a clonal population of cells after single cell clonal density seeding compared to untreated cells. Panel A shows cells after targeting with a Cpf1-based RNP. Panel B shows cells after targeting with a Cas9-based RNP.

[0044] FIG. 13 is a panel of fluorescent microscope images depicting the expression of GFP in a pooled population of cells that had received doxycycline for 24 hours. The doxycycline was added to media 24 hours after Nucleofection of iPSCs with the PRDX1-AKR1A1 Site 2 targeting construct and three different RNPs which comprise three different gRNAs targeting Site 2. No GFP was observed in cells that did not receive doxycycline.

[0045] FIG. 14 is a panel of fluorescent microscope images depicting the expression of GFP in a pooled population of cells that had received doxycycline for 24 hours. The doxycycline was added to media 24 hours after Nucleofection of iPSCs with the PRDX1-AKR1A1 Site 3 targeting construct and three different RNPs which comprise three different gRNAs targeting Site 3. No GFP was observed in cells that did not receive doxycycline.

[0046] FIG. 15 is a panel of flow cytometric histograms depicting induction of GFP expression in a pooled population of cells after treatment with 2 µg/ml doxycycline. The doxycycline was added to media 48 hours after Nucleofection of iPSCs with the PRDX1-AKR1A1 Site 2 targeting construct and three different RNPs which comprise three different gRNAs targeting Site 2. Flow cytometric analysis was performed 5 days after doxycycline treatment. No GFP was observed in cells that did not receive doxycycline and in parental GAPDH::rtTA iPSCs that did not receive the targeting construct and RNP.

[0047] FIG. 16 is a panel of flow cytometric histograms depicting induction of GFP expression in a pooled population of cells after treatment with 2 µg/ml doxycycline. The doxycycline was added to media 24 hours after Nucleofection of iPSCs with the PRDX1-AKR1A1 Site 3 targeting construct and three different RNPs which comprise three different gRNAs targeting Site 3. Flow cytometric analysis was performed 6 days after doxycycline treatment. No GFP was observed in cells that did not receive doxycycline and in parental GAPDH::rtTA iPSCs that did not receive the targeting construct and RNP.

DETAILED DESCRIPTION

[0048] Genetically engineered cells are important tools for cell therapy. But artificial gene circuitry in engineered cells is often subverted by transgene silencing over time, as the cells undergo proliferation, or changes in cell states or in vivo environment. Thus, there is a need for identifying genomic regions that are safe for transgene integration and also provide a chromatin landscape that remains open for transcription across cell types, cell states, and in vivo milieus. Integration of a transgene into such a site would allow the transgene to remain transcriptionally active during the life time of a cell therapy product.

[0049] Provided herein are compositions (e.g., of nucleic acid molecules and cells) and methods for genetically (genetically) engineering cells to achieve expression of a transgene across various cell or differentiation states, without affecting endogenous gene expression that may be detrimental to the cell or the therapeutic purpose of the cell in a cell therapy. The provided compositions and methods are based, at least in part, on the identification of chromatin landscapes comprising sustained transcriptionally active payload regions (STAPLRs) that remain transcriptionally active across cell types and differentiation cell states.

I. STAPLRs

[0050] The present inventors have discovered that certain intergenic regions in the mammalian genome allow consistent levels of expression of transgenes integrated therein, regardless of cell type and/or even as the cell undergoes changes in its state (e.g., differentiation state, maturation, or activity state). This discovery greatly expands the repertoire of genomic sites where transgenes can be stably integrated and their expression can be maintained over changing cell states. The discovery thus solves a long-standing problem in transgene expression, for example, in the context of cell therapy. These intergenic regions are termed “sustained transcriptionally active payload region” (STAPLR) herein, where “payload” or “genomic payload” refers to one or more exogenous or heterologous nucleotide sequences introduced to the region. A STAPLR comprise an open chromatin landscape for landing genomic payloads. The chromosomal DNA in the STAPLR is in a conformation that is accessible to components of gene editing machinery and that allows integration of genetic material. In some instances, a STAPLR is in the vicinity of transcriptionally active genes.

[0051] One application of this discovery is the efficient generation of cells (e.g., therapeutic cells) that are first genetically modified and then made to change cell states, e.g., by differentiating or dedifferentiating. For example, the present genetic engineering method can be applied to iPSCs that are then differentiated into various cell types. In the past, when iPSCs are engineered to incorporate a transgene into their genome and then differentiated into the desired cell types, the transgene can become inactive upon iPSC differentiation. However, transgenes integrated into the STAPLRs as disclosed herein do not become inactive upon iPSC differentiation. Thus, the STAPLRs provide universal “landing pads” for transgene expression.

[0052] This stability in transgene expression is also advantageous after the therapeutic cells in a cell therapy are administered to a subject in need thereof (e.g., a human

patient), where they may encounter different and varying milieus that would have shut down transgenes integrated elsewhere.

[0053] Furthermore, integrating transgenes within intergenic regions, rather than within genes, will cause minimal disruption to the expression or regulation of adjacent genes and therefore allow the normal functioning of the genetically engineered cell. Transgene integration at the STAPLRs also reduces the risk of causing unwanted effects in the cells (e.g., activating an oncogene or disrupting an essential gene such as a tumor suppressor gene). Furthermore, the STAPLRs, with their constantly transcriptionally active status, will allow for the testing and use of a wider range of regulatory elements (e.g., promoters and enhancers).

[0054] As used herein, an “intergenic region” is a stretch of nucleotide sequence located between two neighboring genes. An intergenic region can be of various sizes. For example, the intergenic region can be at least 30, 40, 50, 75, or 100 base pairs in length. In some embodiments, the intergenic region can be at least 150, 200, 300, 400, 500, 750, or 1000 base pairs length. In some embodiments, the intergenic region can be at least 1500, 2000, 2500, 3000, 3500, 5000, or 10000 base pairs in length. In some embodiments, the intergenic region can be at least 15000, 20000, 30000, 40000, 50000, 75000, or 100000 base pairs in length. In some embodiments, the intergenic region is 30 base pairs to 100000 base pairs in length. In some embodiments, the intergenic region is 50 base pairs to 75000 base pairs in length. In some embodiments, the intergenic region is 75 base pairs to 70000 in length.

[0055] STAPLRs of the present disclosure include, without limitation (with the NCBI Gene IDs for the human genes shown in parentheses): the intergenic region between the RPL34 gene (Gene ID: 6164) and the OSTC gene (Gene ID: 58505), the intergenic region between the ACTB gene (Gene ID: 60) and the FSCN1 gene (Gene ID: 6624), the intergenic region between the AKIRIN1 gene (Gene ID: 79647) and the NDUFS5 gene (Gene ID: 4725), the intergenic region between the PRDX1 gene (Gene ID: 5052) and the AKR1A1 gene (Gene ID: 10327), the intergenic region between the PTGES3 gene (Gene ID: 10728) and the NACA gene (Gene ID: 4666), the intergenic region between the MLF2 gene (Gene ID: 8079) and the PTMS gene (Gene ID: 5763), the intergenic region between the RAB13 gene (Gene ID: 5872) and the RPS27 gene (Gene ID: 4840565), the intergenic region between the JTB gene (Gene ID: 10899) and the RAB13 gene (Gene ID: 5872), the intergenic region between the AKR1A1 gene (Gene ID: 10327) and the NASP gene (Gene ID: 4678), the intergenic region between the NDUFS5 gene (Gene ID: 4725) and the MAF1 gene (Gene ID: 23499), the intergenic region between the SRSF9 gene (Gene ID: 8683) and the DYNLL1 gene (Gene ID: 8655), the intergenic region between the MYL6B gene (Gene ID: 140465) and the MYL6 gene (Gene ID: 4637), the intergenic region between the GPX1 gene (Gene ID: 2876) and the RHOA gene (Gene ID: 387), the intergenic region between the HNRNPA2B1 gene (Gene ID: 3181) and the CBX3 gene (Gene ID: 11335), the intergenic region between the ROMO gene (Gene ID: 140823) and the RBM39 gene (Gene ID: 9584), the intergenic region between the PA2G4 gene (Gene ID: 5036) and the RPL41 gene (Gene ID: 6171), and the intergenic region between the NDUFB10 (Gene ID: 4716)

and the RPS2 gene (Gene ID: 6187). In some embodiments, the genes herein refer to human genes and the mammalian cells are human cells.

[0056] The start and end genomic coordinates and the sizes of the aforementioned STAPLR intergenic regions in the human genome are listed in Table 1 below. The coordinates are as defined by information available at NCBI's RefSeq database.

TABLE 1

Intergenic Regions Between Select Genes				
SEQ ID NO	STAPLR	Start coordinate	End coordinate	size
1	Intergenic region between the RPL34 gene and the OSTC gene	chr4: 108,630,485	chr4: 108,650,584	20100
2	Intergenic region between the ACTB gene and the FSCN1 gene	chr7: 5,530,602	chr7: 5,592,815	62214
3	Intergenic region between the AKIRIN1 gene and the NDUFS5 gene	chr1: 39,006,066	chr1: 39,026,294	20229
4	Intergenic region between the PRDX1 gene and the AKR1A1 gene	chr1: 45,522,891	chr1: 45,550,778	27888
5	Intergenic region between the PTGES3 and the NACA gene	chr12: 56,688,285	chr12: 56,712,426	24142
6	Intergenic region between the MLF2 and the PTMS gene	chr12: 6,753,142	chr12: 6,766,362	13221
7	Intergenic region between the RAB13 and the RPS27 gene	chr1: 153,986,340	chr1: 153,990,761	4422
8	Intergenic region between the JTB and the RAB13 gene	chr1: 153,977,675	chr1: 153,981,649	3975
9	Intergenic region between the AKR1A1 and the NASP gene	chr1: 45,570,050	chr1: 45,584,040	13991
10	Intergenic region between the NDUFS5 and the MACF1 gene	chr1: 39,034,616	chr1: 39,084,166	49551
11	Intergenic region between the SRSF9 and the DYNLL1 gene	chr12: 120,469,749	chr12: 120,469,841	93
12	Intergenic region between the MYL6B and the MYL6 gene	chr12: 56,157,983	chr12: 56,158,358	376
13	Intergenic region between the GPX1 and the RHOA gene	chr3: 49,358,354	chr3: 49,359,144	791
14	Intergenic region between the HNRNPA2B1 and the CBX3 gene	chr7: 26,200,747	chr7: 26,201,442	696
15	Intergenic region between the ROMO and the RBM39 gene	chr20: 35,700,985	chr20: 35,701,346	362
16	Intergenic region between the PA2G4 and the RPL41 gene	chr12: 56,113,911	chr12: 56,116,632	2722
97	Intergenic region between the NDUFB10 and the RPS2 gene	chr16: 1,961,976	chr16: 1,962,057	81

[0057] Due to variations between humans and variations between mammalian species, the intergenic regions between the aforementioned gene pairs may differ to some degree from the corresponding SEQ ID NOs shown in Table 1.

[0058] In some embodiments, the intergenic region between the RPL34 gene and the OSTC gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 1 or is sufficiently similar to SEQ ID NO: 1 so that the intergenic region retains the functionality of SEQ ID NO: 1, i.e., the functions (e.g., transcription regulation) of the intergenic region between the RPL34 gene and the OSTC gene remain intact (e.g., without adverse effects on the cell).

[0059] In some embodiments, the intergenic region between the ACTB gene and the FSCN1 gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 2 or is sufficiently similar to SEQ ID NO: 2 so that the intergenic region retains the functionality of SEQ ID NO: 2, i.e., the functions (e.g. transcription regulation) of the intergenic region between the ACTB gene and the FSCN1 gene remain intact (e.g., without adverse effects on the cell).

[0060] In some embodiments, the intergenic region between the AKIRIN1 gene and the NDUFS5 gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 3 or is sufficiently similar to SEQ ID NO: 3 so that the intergenic region retains the functionality of SEQ ID NO: 3, i.e., the functions (e.g., transcription regulation) of the intergenic region between the AKIRIN1 gene and the NDUFS5 gene remain intact (e.g., without adverse effects on the cell).

[0061] In some embodiments, the intergenic region between the PRDX1 gene and the AKR1A1 gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 4 or is sufficiently similar to SEQ ID NO: 4 so that the intergenic region retains the functionality of SEQ ID NO: 4, i.e., the functions (e.g., transcription regulation) of the intergenic region between the PRDX1 gene and the AKR1A1 gene remain intact (e.g., without adverse effects on the cell).

[0062] In some embodiments, the intergenic region between the PTGES3 gene and the NACA gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 5 or is sufficiently similar to SEQ ID NO: 5 so that the intergenic region retains the functionality of SEQ ID NO: 5, i.e., the functions (e.g., transcription regulation) of the intergenic region between the PTGES3 gene and the NACA gene remain intact (e.g., without adverse effects on the cell).

[0063] In some embodiments, the intergenic region between the MLF2 gene and the PTMS gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 6 or is sufficiently similar to SEQ ID NO: 6 so that the intergenic region retains the functionality of SEQ ID NO: 6, i.e., the functions (e.g., transcription

regulation) of the intergenic region between the MLF2 gene and the PTMS gene remain intact (e.g., without adverse effects on the cell).

[0064] In some embodiments, the intergenic region between the RAB13 gene and the RPS27 gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 7 or is sufficiently similar to SEQ ID NO: 7 so that the intergenic region retains the functionality of SEQ ID NO: 7, i.e., the functions (e.g., transcription regulation) of the intergenic region between the RAB13 gene and the RPS27 gene remain intact (e.g., without adverse effects on the cell).

[0065] In some embodiments, the intergenic region between the JTB gene and the RAB13 gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 8 or is sufficiently similar to SEQ ID NO: 8 so that the intergenic region retains the functionality of SEQ ID NO: 8, i.e., the functions (e.g., transcription regulation) of the intergenic region between the JTB gene and the RAB13 gene remain intact (e.g., without adverse effects on the cell).

[0066] In some embodiments, the intergenic region between the AKR1A1 gene and the NASP gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 9 or is sufficiently similar to SEQ ID NO: 9 so that the intergenic region retains the functionality of SEQ ID NO: 9, i.e., the functions (e.g., transcription regulation) of the intergenic region between the AKR1A1 gene and the NASP gene remain intact (e.g., without adverse effects on the cell).

[0067] In some embodiments, the intergenic region between the NDUFS5 gene and MCAF1 gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 10 or is sufficiently similar to SEQ ID NO: 10 so that the intergenic region retains the functionality of SEQ ID NO: 10, i.e., the functions (e.g., transcription regulation) of the intergenic region between the NDUFS5 gene and the MCAF1 gene remain intact (e.g., without adverse effects on the cell).

[0068] In some embodiments, the intergenic region between the SRSF9 gene and DYNLL1 gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 11 or is sufficiently similar to SEQ ID NO: 11 so that the intergenic region retains the functionality of SEQ ID NO: 11, i.e., the functions (e.g., transcription regulation) of the intergenic region between the SRSF9 gene and the DYNLL1 gene remain intact (e.g., without adverse effects on the cell).

[0069] In some embodiments, the intergenic region between the MYL6B gene and MYL6 gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 12 or is sufficiently similar to SEQ

ID NO: 12 so that the intergenic region retains the functionality of SEQ ID NO: 12, i.e., the functions (e.g., transcription regulation) of the intergenic region between the MYL6B gene and the MYL6 gene remain intact (e.g., without adverse effects on the cell).

[0070] In some embodiments, the intergenic region between the GPX1 gene and RHOA gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 13 or is sufficiently similar to SEQ ID NO: 13 so that the intergenic region retains the functionality of SEQ ID NO: 13, i.e., the functions (e.g., transcription regulation) of the intergenic region between the GPX1 gene and the RHOA gene remain intact (e.g., without adverse effects on the cell).

[0071] In some embodiments, the intergenic region between the HNRNPA2B1 gene and CBX3 gene comprises a nucleotide sequence at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 14 or is sufficiently similar to SEQ ID NO: 14 so that the intergenic region retains the functionality of SEQ ID NO: 14, i.e., the functions (e.g., transcription regulation) of the intergenic region between the HNRNPA2B1 gene and the CBX3 gene remain intact (e.g., without adverse effects on the cell).

[0072] In some embodiments, the intergenic region between the ROMO gene and RBM39 gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 15 or is sufficiently similar to SEQ ID NO: 15 so that the intergenic region retains the functionality of SEQ ID NO: 15, i.e., the functions (e.g., transcription regulation) of the intergenic region between the ROMO gene and the RBM39 gene remain intact (e.g., without adverse effects on the cell).

[0073] In some embodiments, the intergenic region between the PA2G4 gene and RPL41 gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 16 or is sufficiently similar to SEQ ID NO: 16 so that the intergenic region retains the functionality of SEQ ID NO: 16, i.e., the functions (e.g., transcription regulation) of the intergenic region between the PA2G4 gene and the RPL41 gene remain intact (e.g., without adverse effects on the cell).

[0074] In some embodiments, the intergenic region between the NDUFB10 and the RPS2 gene comprises a nucleotide sequence at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 16 or is sufficiently similar to SEQ ID NO: 16 so that the intergenic region retains the functionality of SEQ ID NO: 97, i.e., the functions (e.g., transcription regulation) of the intergenic region between the NDUFB10 and the RPS2 gene remain intact (e.g., without adverse effects on the cell).

[0075] The percent identity of two nucleotide sequences can be determined by, e.g., BLAST® using default parameters (available at the U.S. National Library of Medicine's National Center for Biotechnology Information website). In some embodiments, the length of a reference sequence

aligned for comparison purposes is at least 30%, (e.g., at least 40, 50, 60, 70, 80, or 90% of the reference sequence.

II. Integration of Exogenous Sequences into STAPLRs

A. Integration Sites

[0076] An exogenous nucleotide sequence of interest may be integrated at any site within a STAPLR. For example, the integration site, or the junction between the exogenous sequence and the adjacent endogenous sequence, may be located in the first half or the second half of the STAPLR; in the 5', middle, or 3' third of the STAPLR; or in the first, second, third, or fourth quarter of the STAPLR. In some embodiments, the integration site of the exogenous sequence, or the junction between the exogenous sequence and the adjacent endogenous sequence, is located within the STAPLR and at least 10, 20, 30, 40, 50, 80, 90, 100, 200, 300, 400, 500, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 5000, 10000, 15000, or 20000 base pairs away from the nearest gene, i.e., from the 5' or 3' boundary of the STAPLR (e.g., from the start or end coordinate shown in Table 1).

[0077] In a single genome, one or more exogenous nucleotide sequences may be integrated into one or more STAPLRs. In some embodiments, one or more (e.g., two, three, or four) exogenous nucleotide sequences may be integrated into one or more sites within a single given STAPLR. In some embodiments, more than one STAPLR in a single genome is targeted for integration of exogenous nucleotide sequences.

[0078] In some embodiments, exogenous sequences are introduced into at least one STAPLR and at least one sustained transgene expression locus (STEL) as described in WO 2021/072329. A STEL site is the locus of an endogenous gene that is robustly and consistently expressed in the pluripotent state as well as during differentiation (e.g., as examined by single-cell RNA sequencing (scRNAseq) analysis). While a STAPLR can be associated with a STEL site, it does not need to be associated with a STEL site. STEL sites may be identified from single cell RNA sequence data. A defining characteristic of a desirable STEL site is the ubiquity of expression. STEL sites may be identified by analyzing a candidate gene locus's expression across diverse cell types and cell maturity states such as PSCs and PSC-derived dopamine neurons (and select progenitor states), microglia (and select progenitor states), and cardiomyocytes (and select cardiomyocyte progenitor states). Adding publicly available single cell RNA sequencing data of adult human tissue allows for the refining of such a STEL analysis. STEL include, without limitation, certain housekeeping genes that are active in multiple cell types such as those involved in gene expression (e.g., transcription factors and histones), cellular metabolism (e.g., GAPDH and NADH dehydrogenase), or cellular structures (e.g., actin), or those that encode ribosomal proteins (e.g., large or small ribosomal subunits, such as RPL13A, RPLP0 and RPL7). Examples of STEL are genes encoding ribosomal proteins such as RPL genes (e.g., RPL13A, RPLP0, RPL10, RPL13, RPS18, RPL3, RPLP1, RPL15, RPL41, RPL11, RPL32, RPL18A, RPL19, RPL28, RPL29, RPL9, RPL8, RPL6, RPL18, RPL7, RPL7A, RPL21, RPL37A, RPL12, RPL5, RPL34, RPL35A, RPL30, RPL24, RPL39, RPL37, RPL14, RPL27A, RPLP2, RPL23A, RPL26, RPL36, RPL35, RPL23, RPL4, and RPL22) and RPS genes (e.g., RPS2, RPS19, RPS14, RPS3A, RPS12, RPS3, RPS6, RPS23, RPS27A, RPS8, RPS4X, RPS7, RPS24, RPS27, RPS15A,

RPS9, RPS28, RPS13, RPSA, RPS5, RPS16, RPS25, RPS15, RPS20, and RPS11); genes encoding mitochondria proteins (e.g., MT-CO1, MT-CO2, MT-ND4, MT-ND1, and MT-ND2); genes encoding actin proteins (ACTG1 and ACTB); genes encoding eukaryotic translation factors (e.g., EEF1A1, EEF2, and EIF1); and genes encoding histones (e.g., H3F3A and H3F3B). Additional STELs are those that encode proteins involved in focal adhesion, cell-substrate adherens junction, cell-substrate junction, cell anchoring, extracellular exosome, extracellular vesicle, intracellular organelle, or anchoring junction. Additional examples of STELs are FTL, FTH1, TPT1, TMSB10, GAPDH, PTMA, GNB2L1, NACA, YBX1, NPM1, FAU, UBA52, HSP90AB1, MYL6, SERF2, and SRP14.

[0079] In some embodiments, in a single mammalian (e.g., human) genome, exogenous sequences are introduced into a STAPLR such as the RPL34-OSTC or PRDX1-AKRIA1 STAPLR and a STEL such as the GAPDH locus. In some embodiments, exogenous sequences are introduced in multiple STAPLRs in a single genome, such as the RPL34-OSTC and PRDX1-AKRIA1 STAPLRs.

[0080] The integration site of an exogenous nucleotide sequence may be within the STAPLR or in gene sequences adjacent to the STAPLR (e.g., in exon, intron, or UTRs of a gene). In some embodiments, an endonuclease generates DNA breaks within a STAPLR. In other embodiments, an endonuclease generates DNA breaks in a gene adjacent to a STAPLR such that after integration, the exogenous nucleotide sequence is still integrated within the STAPLR. In some embodiments, screening of improper integration events may be performed in accordance with methods described in WO 2021/226151, wherein a DNA break is introduced in an exon of a gene that is adjacent to a STAPLR and is necessary for cell survival, and those cells in which integration is not properly achieved do not survive.

B. Methods of Integration

[0081] Any method of genomic integration can be used to take advantage of the STAPLRs described herein. In some embodiments, integration of the exogenous nucleotide sequence in the STAPLR is achieved by using a genomic editing system selected from the group consisting of a CRISPR/Cas system, a Cre/Lox system, a FLP-FRT system, a Transcription Activator-Like Effector Nuclease (TALEN) system, a zinc finger nuclease (ZFN) system, a homing endonuclease, a sequence-specific endonuclease, random integration (e.g., through transposons), a meganuclease, homologous recombination, transposases, and non-nuclease dependent viral vectors (e.g., retroviral, AAV, or lentiviral vectors). In some embodiments, the integration causes no deletion of the endogenous sequence in the region, and/or no addition of nucleotide sequences other than the exogenous donor sequence to be integrated. In some embodiments, the integration causes insertions (of non-donor sequence) and/or deletions (indels) at the integration site.

[0082] In some embodiments, the exogenous sequence may be incorporated into a STAPLR site via homologous recombination at DNA breaks generated by a suitable endonuclease such as a CRISPR-associated endonuclease, which may be, for example, a Cas endonuclease selected from, without limitation, a type I (e.g., subtype I-A, I-B, I-C, I-C variant, I-D, I-E, I-F, I-F variant 1, or I-F variant 2), type II (e.g., subtype II-A, II-B, II-B, or II-C), type III (e.g., subtype III-A, III-B, or III-B variant), type IV, or type V Cas protein,

or a variant thereof. In some embodiments, the nuclease is selected from the group consisting of Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas12 (e.g., Cas12a or Cpf1, or Cas12b), Cas13, Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, CasX, CasY, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, CasPhi, MAD7, Csf4, and homologs thereof, or modified versions thereof (e.g., truncated versions or variants of a wildtype Cas protein with a nuclease activity).

[0083] In some embodiments, the Cas endonuclease is a Cpf1 (Cas12a) endonuclease, or a variant, derivative, or fragment thereof, such as, for example, Cpf1 derived from *Francisella novicida* U112 (FnCpf1), Acidaminococcus sp. BV3L6 (AsCpf1, including improved variants such as enA-

sCpf1), *Lachnospiraceae bacterium* ND2006 (LbCpf1), *Lachnospiraceae bacterium* MA2020 (Lb2Cpf1), *Lachnospiraceae bacterium* MC2017 (Lb3Cpf1), *Moraxella bovoculi* 237 (MbCpf1), or *Prevotella disiens* (PdCpf1).

[0084] In some embodiments, the Cas endonuclease is a Cas9 protein or a variant, derivative, or fragment thereof. In some embodiments, the Cas9 protein is SaCas9, SpCas9, SpCas9n, Cas9-HF, Cas9-H840A, FokI-dCas9, or DGTA nickase.

[0085] In some embodiments, the Cas endonuclease is a Type V RNA programmable nuclease, as disclosed in WO 2022/258753.

[0086] In some embodiments, the Cas endonuclease is a MAD nuclease, such as MAD7 nuclease, as disclosed in U.S. Pat. No. 10,337,028.

[0087] Non-limiting examples of suitable endonucleases are set forth in Table A below.

TABLE A

Exemplary Endonucleases	
Enzyme	Sequence
Cas12a	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQQFIEEDKARNNDHYKELKPI IDRIYKTYADQCLQLVQLDWENLSAIDSYRKEETEETRNALIEEQATYRNAl HDYFIGRTDNLTDAINKRHAEIYKGLFAEFLNGKVLKQLGTVTTEHENALL RSFDKFTTYFSGFYENRKNVFAEDISTAIPHRIVQDNFPKFKENCHIFTRLl TAVPSLREHFENVKKAIGIFVFSTSIEEVFSFPFYQNLLTQTQIDLYNQLGGI SREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSF ILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFSHKKl ETISSALCDHWDTLNALYERRISELTGKITSAKEKVQRSLKHEDINLQEII SAAGKELSEAFQKQTSEILSHAHAALDQPLPTTLKQEEKEILSQLDSLGL YHLLDWFADVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEK FKLNQFQPTLASGWDVNKEKNNGA1FVKNGLYLGIMPQKGRYKALSFEPT EKTSEGFDKMYDYPFDAAKMI PKCSTQLKAVTAHFQTHTPILLSNNFIEPL EITKEIYDLNNPEKEPKKFQTYAKKTDQKGYREALCKWIDFRDFLSKYTK TTSDILSSLRPSSQYKDLGEYYAELNPPLYHISFORIAKEIMDAVETGKLYL FOIYNKDFAKGHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAEFLYRPKSRM KRMARLGEMLNKKLDQKTPIPDLYQELYDVNHRLSHDLSDEARALLPN VITKEVSHEIIKDRRTSDKFFFHVPITLNYQAANSPSKFNRQVRNAYLKEHPE TPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAA RQAWSVVGTIKDLKGQYLSQVIHEIVDLMIHQAVVLENLNPGFKSKRTGIA EKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGLNPYQLTDQFTSFAKMGTQSG FLFVVPAPYTSKIDPLTFGVDPFWKTIKNHESRKHFLEGDFLHYDVKTGDF ILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVEN HRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALI RSVLQMRNSNAATGEDYINSPVRLNGVCFDSRFQNPEWPMADANGAYHIAL KGQLLNHLKESKDLKLONGTISNODWLAYIQELRN (SEQ ID NO: 98)
Cas12a variant 1	MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQQFIEEDKARNNDHYKELKPI IDRIYKTYADQCLQLVQLDWENLSAIDSYRKEETEETRNALIEEQATYRNAl HDYFIGRTDNLTDAINKRHAEIYKGLFAEFLNGKVLKQLGTVTTEHENALL RSFDKFTTYFSGFYENRKNVFAEDISTAIPHRIVQDNFPKFKENCHIFTRLl TAVPSLREHFENVKKAIGIFVFSTSIEEVFSFPFYQNLLTQTQIDLYNQLGGI SREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSF ILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFSHKKl ETISSALCDHWDTLNALYERRISELTGKITSAKEKVQRSLKHEDINLQEII SAAGKELSEAFQKQTSEILSHAHAALDQPLPTTLKQEEKEILSQLDSLGL YHLLDWFADVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEK FKLNQFQPTLASGWDVNKEKNNGA1FVKNGLYLGIMPQKGRYKALSFEPT EKTSEGFDKMYDYPFDAAKMI PKCSTQLKAVTAHFQTHTPILLSNNFIEPL EITKEIYDLNNPEKEPKKFQTYAKKTDQKGYREALCKWIDFRDFLSKYTK TTSDILSSLRPSSQYKDLGEYYAELNPPLYHISFORIAKEIMDAVETGKLYL FOIYNKDFAKGHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAEFLYRPKSRM KRMARLGEMLNKKLDQKTPIPDLYQELYDVNHRLSHDLSDEARALLPN VITKEVSHEIIKDRRTSDKFFFHVPITLNYQAANSPSKFNRQVRNAYLKEHPE TPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAA RQAWSVVGTIKDLKGQYLSQVIHEIVDLMIHQAVVLENLNPGFKSKRTGIA EKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGLNPYQLTDQFTSFAKMGTQSG FLFVVPAPYTSKIDPLTFGVDPFWKTIKNHESRKHFLEGDFLHYDVKTGDF ILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVEN HRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALI RSVLQMRNSNAATGEDYINSPVRLNGVCFDSRFQNPEWPMADANGAYHIAL KGQLLNHLKESKDLKLONGTISNODWLAYIQELRN (SEQ ID NO: 98)

TABLE A-continued

Exemplary Endonucleases	
Enzyme	Sequence
Type V RNA programmable nuclease example 1	KGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRNGRSSDDEATADSQHAAPP KKKRKV (SEQ ID NO: 99)
Type V RNA programmable nuclease example 1	MTIRSMKLKLKIYSGRSAPQLRQLWRLHRLLNEGTAYYMDWLVHMRQEALPG KSKEIIRAEERRVRQQEKNGVQNDQVPMDEVLISALRQLYELLVPSAVNNSG DAQTLSRKFLSPVDPNSEGGKGTSNAGAKPGWRKKQEAGDPSWEKDAYERWLK RKQADPTAEILGKLETAGLKPLFPLYTNTEVKDIRWMPPLTSKQYVRNWDRDMFO QAIETHLLSWETWNRKVNEERAKLKTETVRRFEEQHLLANGKDWLSPHQAYEANRE QALRDMAISPSSDRFIRTRRQIKGWSELERYWNKLAPTAASVEAYMQEVRHVQKK LGGTFGDADLYRFLAKPENVHIWRDHQERLHYAAAYNDLHKRRLMSAKEQAAFT LPDPVAHPLWVRFDA RDGNLFTYLQADSKQRSSRYNFNSRFLWPVEDGYFE ETENVKVELALSQKFYRQVIVHDNPTGKQKITFQDYSSKEILEGHGLGGAKLQL DRNPLRKSGRDFETGDFGP AFLNVVLDLKPQEVKNGRLQSPLQGALLVKSRP NDIPK VGYKPDAALAWLEQASGEESLQGFRVMSIDLGVRSAAAISV FSVKGEKTRREGDKV CYPVGETGLFLEVDRSFLRLPGE SSEKRVNVERDKRT ERMQIYHIRT LARVLRLANKATPM DRIKAVQDV LNDIESTRFMNDHDHVYN HALETLR TYAPDHQI WEEQVIAAHRQLEHHVVGIVG EWRKNWGKDRGTVGL SMDNIEELDEM RRLI WSRR ARY PREAKP FQVNE SNPV HLLRHL QNL KEDRL KQLANLIVMT ALGYY DSKEKKWKAAYPACQ LILFEDLQ RYR FHLD R SARENS QLMKWAHRSI PKVW M QGE PY GLQ IGDV WAGFTS RY HAKT G A P G I R C K A L T E K DFQ Q R L E S L V A E G M F T L Q E V G T L K P G D I V P A R G G E L F V T L A D D S G D R I V I T HADINAAQNVQKRFWLANSERFRVACRSVQIASQECFIPSSES VAKKMKG VGF VRDFSHKDMEVYHWNQVKLTAKNVP TDHSDDLQDLQDYQ A I L E A R E S S S S YKTLP RDPSGFFF DDVV VPQNI YW REV KKT IT ALL R K R I M S T (SEQ ID NO: 100)
Type V RNA programmable nuclease example 1	MPIRSFKLKLVTHNGDSTYMDKLRRGLWKTHVIINRGIAYYMNTLALMRQE PY GSKSREEVRLDLLSTLREQQRNNWSEQTGTDDDELLSLSRRVYELLVPSAIGE KGDAQMLSRKFLSPVDPNSEGGRTAKSGRKPRWKKMMEEGHPDWEKEKEKD AAKKAEDPTASILADEAVGLLPLFPLFSDEQKEIRWLPKKKQFVRTWDRDM FQQALERMLSWESWNRVVAEYELKLQAQRDEVYAKYLEDAGSWLNDLQTFEKQ REELAEVS FEPN S E Y L I T R R Q I R G W K E V Y E K W S K T S E N A S Q E Q L W R M V A D V Q TAMAGAFGDPK VYQFLSQPKHHHIWREHPNRLFYYSKYNEVREKLNRAKKQAA FTL PDPV E HPLWTRFDARGN IHDYE I SKV GKQYHVT FSS L L P E A Q S W V E I E NVT VGIGNSLQLK Q I R L D G Y A D K K Q K V K Y Y D S S R F E L T G V L G G A K I Q F D R K HLKKAHRLAEGETGP IF LN V V D V E P F L E V K N G R L R T P L G Q V L Q V N T R D W P K VVDYKAKELSVLM T Q I G N E N G V S T I E A G M R I M S I D L G Q R T A A V S I F E V I S KKPDEKETKL F Y P I A D T D L Y A V H R R S L L R L P G E E I S S K K M I E K R K E R A R I R S LVRYQIRLLSEVRLHHTQGTAEQRFFKLDELLVSIQKLELDQSEWI SELEKL FDYIDESAEKWKEALVVAHRTLEPIVVEAVRNWKKSLSKENKDRRIAGISIW SIEELEETRKL LIAWSKHSREPGIPKRLEKEETFAPEHLQH I QNVK D D R L K Q M ANLFVMTALGKYD E G N K R W V E A Y P A C Q V I L F E D L S R Y R F A L D P R R E N N R L M KWAHRSI P R L T Y M Q A E L F G I Q V G D V Y S A Y T S R F H A K T G A P G I R C H A L T E A D L Q SNSYVNQ L I K D K F I Q D N Q T E I L K A G Q I V P W Q G G E L F V T F A D R S G A S L A V I H A DINAQNLQKRFWQHNSEVERVPCCKVVKGGLVPVYEKMRKLFKGGLFVNIDDP ESKEVYRWEHSTKMKSKTPVDLES EDIEHEELS DEWEDM Q E G Y K T L L R D P S G FFWSSDSWIPOQKDFWIRVKS RIG KSL REQ IR (SEQ ID NO: 101)
Type V RNA programmable nuclease example 3	MPIRSFKLKLVTHNGDSTYMDKLRRGLWKTHVIINRGIAYYMNTLALMRQE PY GSKSREEVRLDLLSTLREQQRNNWSEQTGTDDDELLSLSRRVYELLVPSAIGE KGDAQMLSRKFLSPVDPNSEGGRTAKSGRKPRWKKMMEEGHPDWEKEKEKD AAKKAEDPTASILADEAVGLLPLFPLFSDEQKEIRWLPKKKQFVRTWDRDM FQQALERMLSWESWNRVVAEYELKLQAQRDEVYAKYLEDAGSWLNDLQTFEKQ REELAEVS FEPN S E Y L I T R R Q I R G W K E V Y E K W S K T S E N A S Q E Q L W R M V A D V Q TAMAGAFGDPK VYQFLSQPKHHHIWREHPNRLFYYSKYNEVREKLNRAKKQAA FTL PDPV E HPLWTRFDARGN IHDYE I SKV GKQYHVT FSS L L P E A Q S W V E I E NVT VGIGNSLQLK Q I R L D G Y A D K K Q K V K Y Y D S S R F E L T G V L G G A K I Q F D R K HLKKAHRLAEGETGP IF LN V V D V E P F L E V K N G R L R T P L G Q V L Q V N T R D W P K VVDYKAKELSVLM T Q I G N E N G V S T I E A G M R I M S I D L G Q R T A A V S I F E V I S KKPDEKETKL F Y P I A D T D L Y A V H R R S L L R L P G E E I S S K K M I E K R K E R A R I R S LVRYQIRLLSEVRLHHTQGTAEQRFFKLDELLVSIQKLELDQSEWI SELEKL FDYIDESAEKWKEALVVAHRTLEPIVVEAVRNWKKSLSKENKDRRIAGISIW SIEELEETRKL LIAWSKHSREPGIPKRLEKEETFAPEHLQH I QNVK D D R L K Q M ANLFVMTALGKYD E G N K R W V E A Y P A C Q V I L F E D L S R Y R F A L D P R R E N N R L M KWAHRSI P R L T Y M Q A E L F G I Q V G D V Y S A Y T S R F H A K T G A P G I R C H A L T E A D L Q SNSYVNQ L I K D K F I Q D N Q T E I L K A G Q I V P W Q G G E L F V T F A D R S G A S L A V I H A DINAQNLQKRFWQHNSEVERVPCCKVVKGGLVPVYEKMRKLFKGGLFVNIDDP ESKEVYRWEHSTKMKSKTPVDLES EDIEHEELS DEWEDM Q E G Y K T L L R D P S G FFWSSDSWIPOQKDFWIRVKS RIG KSL REQ IR (SEQ ID NO: 102)

[0088] In some embodiments, the CRISPR/Cas system comprises a gRNA-dependent nuclease (or a coding sequence thereof) targeting a selected intergenic region, a gRNA (or a coding sequence thereof), and a donor DNA comprising the exogenous nucleotide sequence.

[0089] In some embodiments, the STAPLR is the intergenic region between the RPL34 gene and the OSTC gene, and the gRNA is selected from SEQ ID NOS: 25-32.

[0090] In some embodiments the STAPLR is the intergenic region between the ACTB gene and the FSCN1 gene, and the gRNA is selected from SEQ ID NOS: 33-54.

[0091] In some embodiments, the STAPLR is the intergenic region between the AKIRIN1 gene and the NDUFS5 gene, and the gRNA is selected from SEQ ID NOS: 55-70.

[0092] In some embodiments, the STAPLR is the intergenic region between the PRDX1 gene and the AKR1A1 gene, and the gRNA is selected from SEQ ID NOS: 71-92.

C. Exogenous Nucleotide Sequences

[0093] In some embodiments, the exogenous nucleotide sequence of interest for integration may comprise a transgene encoding a protein (as used herein, including a peptide) or an RNA. The transgene may comprise a coding sequence for the gene product and optionally one or more transcription regulatory elements. In some embodiments, the transgene comprises one or more regulatory elements wherein the one or more regulatory elements may be optionally linked operably to the coding sequence.

[0094] Nonlimiting examples of regulatory elements are promoters, enhancers, silencers, chromatin insulators, intronic sequences, Kozak sequences, ubiquitous chromatin opening elements (UCOE), transcription activator binding elements, sequences that enhance gene expression or RNA stability (e.g., a WPRE element), polyadenylation signal sequences (e.g., SV40 polyA signal), and the like.

[0095] In some embodiments, the promoter directing the expression of the transgene is a constitutive promoter, including, without limitation, EF1 α , EFS, UBC, PGK, CAGGS, CMV, SV40, B2M, and ROSA26 promoters. In some embodiments, the promoter is a cell type-specific, tissue-specific or lineage-specific promoter. For example, the promoters may be a tyrosine hydroxylase promoter for dopaminergic neurons; a Hb9 promoter for motor neurons; a SIRPA promoter for cardiomyocytes; a CD14, CD33, CD45, or CD11b promoter for cells of myeloid lineages; or a CD3, FOXP3, CD25, CD8, or CD4 promoter for T lymphocytes. In some embodiments, the expression of the transgene is under the control of an inducible promoter (e.g., lac operon, which can be triggered by Isopropyl β -D-1-thiogalactopyranoside (IPTG); TRE promoter, which can be triggered by tetracycline and its derivatives).

[0096] In some embodiments, the exogenous sequence comprises one or more regulatory elements that respond to factors expressed from another site (e.g., from an endogenous gene or a transgene integrated at a STEL or STAPLR). A non-limiting example of such a regulatory element is a transcription factor binding site. In some embodiments, such a regulatory element is integrated at a STAPLR site in the vicinity of other one or more regulatory elements and/or the coding sequence of a transgene. For example, a cell may be modified with a DNA molecule as disclosed herein comprising an exogenous nucleotide sequence comprising a transgene and a transcription factor binding site, where the transcription factor that can bind to the transcription factor

binding site is expressed from an endogenous gene, or another transgene in any part of the genome (e.g., in a STAPLR, STEL, or another safe harbor site) or ectopically.

[0097] In some embodiments, the transgene encodes an RNA (e.g., a small interfering RNA or a micro-RNA) or a protein of interest. The protein of interest (as used herein, including a peptide) may be, for example, a globular protein (e.g., an albumin, a globulin, a glutelin, a prolamine, a histone, a globin, or a protamine), a fibrous protein (e.g., a scleroprotein such as a collagen, an elastin, a keratin, or a fibroin), or an intermediate protein. In some embodiments, the protein of interest is a complex protein such a metalloprotein, a chromoprotein, a glycoprotein, a mucoprotein, a phosphoprotein, a lipoprotein. In some embodiments, the protein of interest is a therapeutic protein (e.g., a protein that can improve or prevent symptoms of a disease or condition). Nonlimiting examples of therapeutic proteins include proteins that are deficient or defective in genetic diseases such as hemophilia and lysosome storage diseases, hormones, enzymes, cytokines that regulate immunity, recombinant antigen receptors (e.g., chimeric antigen receptors), antibodies, proteins that regulate differentiation or activity of the modified cells (e.g., transcription factors or proteins maintaining cells in M1 or M2 polarity), and the like. In some embodiments, the protein of interest is a cellular marker, a protein used for immune evasion, or a safety or kill switch used in cell therapy. Examples of proteins of interest are, without limitation, SOX10, IL-10, IL-12, CD19t, and ThPOK.

D. Targeting Vectors

[0098] The present disclosure provides targeting vectors for integrating exogenous nucleotide sequences into the STAPLRs. As used herein, a “targeting vector” is a nucleic acid comprising an exogenous nucleotide sequence of interest and sequences homologous to endogenous chromosomal nucleotide sequences that flank the desired integration location in the genome. These flanking homology sequences are referred to as “homology arms.” Homology arms direct the targeting vector to a specific chromosomal location within the genome by virtue of the homology existing between the homology arms and the corresponding endogenous nucleotide sequences. In some embodiments, the targeting vector is a DNA molecule comprising a nucleotide sequence of interest, flanked by a 5' nucleotide sequence (a left homology arm or homology region) and a 3' nucleotide sequence (a right homology arm or homology region), wherein the 5' nucleotide sequence and the 3' nucleotide sequence are homologous to the nucleotide sequences flanking the integration site in the genome of the cell and mediate integration of the nucleic acid of interest through homology recombination into the integration site.

[0099] The 5' and 3' sequences are sufficiently similar to the endogenous nucleotide sequences being targeted for homology recombination such that the homology arms if integrated (either wholly or partially) do not cause adverse effects on the genetic environment of the integration (e.g., not impact the neighboring genes' functions). In some embodiments, the homology arms are at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the nucleotide sequences in the targeted STAPLR.

[0100] For example, in some embodiments, the intergenic region between the RPL34 gene and the OSTC gene comprises a nucleotide sequence at least 80% identical to SEQ ID NO: 1 in that the functions of the intergenic region between the RPL34 gene and the OSTC gene remains intact after integration. In some embodiments, the intergenic region between the RPL34 gene and the OSTC gene comprises a nucleotide sequence at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 1 so that the functions of the intergenic region between the RPL34 gene and the OSTC gene remains intact after integration. The same may be said for the intergenic region between the ACTB gene and the FSCN1 gene and its identity to SEQ ID NO: 2, the intergenic region between the AKIRIN1 gene and the NDUFS5 gene and its identity to SEQ ID NO: 3, and the intergenic region between the PRDX1 gene and the AKR1A1 gene and its identity to SEQ ID NO: 4, the intergenic region between the PTGES3 gene and the NACA gene and its identity to SEQ ID NO: 5, the intergenic region between the MLF2 gene and the PTMS gene and its identity to SEQ ID NO: 6, the intergenic region between the RAB13 gene and the RPS27 gene and its identity to SEQ ID NO: 7, the intergenic region between the JTB gene and the RAB13 gene and its identity to SEQ ID NO: 8, the intergenic region between the AKR1A1 gene and the NASP gene and its identity to SEQ ID NO: 9, the intergenic region between the NDUFS5 gene and the MA CF1 gene and its identity to SEQ ID NO: 10, the intergenic region between the SRSF9 gene and the DYNLL1 gene and its identity to SEQ ID NO: 11, the intergenic region between the MYL6B gene and the MYL6 gene and its identity to SEQ ID NO: 12, the intergenic region between the GPX1 gene and the RHOA gene and its identity to SEQ ID NO: 13, the intergenic region between the HNRNPA2B1 gene and the CBX3 gene and its identity to SEQ ID NO: 14, the intergenic region between the ROMO gene and the RBM39 gene and its identity to SEQ ID NO: 15, the intergenic region between the PA2G4 gene and the RPL41 gene and its identity to SEQ ID NO: 16, and the intergenic region between the NDUFB10 and the RPS2 gene and its identity to SEQ ID NO: 97.

[0101] In the methods of the present disclosure, the homology arms vary in length. In some embodiments, each of the homology arms is independently about at least 50, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, at least 500, at least 550, at least 600, at least 650, at least 700, at least 750, at least 800, at least 850, at least 900, at least 950, at least 1000, at least 1100, at least 1200, at least 1300, at least 1400, at least 1500, at least 1600, at least 1700, at least 1800, at least 1900, or at least 2000 base pairs long. In some embodiments, each of the homology arms is independently 50-2000, 50-1500, 100-1900, 150-1800, 200-1700, 250-1600, 300-1500, 350-1400, 400-1300, 450-1200, 500-1100, 550-1000, 600-950, 650-900, 700-850, or 750-800 base pairs in length.

[0102] In the methods of the disclosure, the homology arms (i.e., the 5' and 3' nucleotide sequences) can be designed to target anywhere within the disclosed intergenic region. The homology arms can be designed based on genomic sequences available in sequence databases (e.g., the NCBI database).

[0103] In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar

to SEQ ID NO: 17 as necessary for the function of the sequence to remain intact and the 3' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar to SEQ ID NO: 18 as necessary for the function of the sequence to remain intact. In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 17 for the function of SEQ ID NO: 17 to remain intact. In some embodiments, the 3' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 18 for the function of SEQ ID NO: 18 to remain intact.

[0104] In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar to SEQ ID NO: 19 as necessary for the function of the sequence to remain intact and the 3' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar to SEQ ID NO: 20 as necessary for the function of the sequence to remain intact. In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 19 for the function of SEQ ID NO: 19 to remain intact. Similarly, in some embodiments, the 3' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 20 for the function of SEQ ID NO: 20 to remain intact.

[0105] In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar to SEQ ID NO: 21 as necessary for the function of the sequence to remain intact and the 3' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar to SEQ ID NO: 22 as necessary for the function of the sequence to remain intact. In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 21 for the function of SEQ ID NO: 21 to remain intact. Similarly, in some embodiments, the 3' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 22 for the function of SEQ ID NO: 22 to remain intact.

[0106] In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar to SEQ ID NO: 23 as necessary for the function of the sequence to remain intact and the 3' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar to SEQ ID NO: 24 as necessary for the function of the sequence to remain intact. In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 23 for the function of SEQ ID NO: 23 to remain

intact. Similarly, in some embodiments, the 3' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 24 for the function of SEQ ID NO: 24 to remain intact.

[0107] In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar to SEQ ID NO: 93 as necessary for the function of the sequence to remain intact and the 3' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar to SEQ ID NO: 94 as necessary for the function of the sequence to remain intact. In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 93 for the function of SEQ ID NO: 93 to remain intact. Similarly, in some embodiments, the 3' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 94 for the function of SEQ ID NO: 94 to remain intact.

[0108] In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar to SEQ ID NO: 95 as necessary for the function of the sequence to remain intact and the 3' nucleotide sequence comprises a nucleotide sequence that is sufficiently similar to SEQ ID NO: 96 as necessary for the function of the sequence to remain intact. In some embodiments, the 5' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 95 for the function of SEQ ID NO: 95 to remain intact. Similarly, in some embodiments, the 3' nucleotide sequence comprises a nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to SEQ ID NO: 96 for the function of SEQ ID NO: 96 to remain intact.

[0109] In some embodiments, the homology arms completely fall within the targeted STAPLR. In other embodiments, the homology arms may overlap with a portion of a neighboring gene without disrupting its function after integration and the exogenous sequence still is integrated within the STAPLR.

[0110] In some embodiments, the targeting vector is a circular vector. In some embodiments, the targeting vector is linear vector. In some embodiments, a targeting vector as provided herein comprises one or more endonuclease targeting sequences, e.g., to linearize the vector when being used with an endonuclease-guide combination. In some embodiments, the target vector is a viral vector (e.g., an AAV vector, an adenoviral vector, a lentiviral vector, a herpes simplex viral vector), or a plasmid vector.

[0111] The present disclosure provides a STAPLR-targeting system that comprises the targeting vector herein and an appropriate gene editing system such as those described herein, for incorporating the nucleotide sequence of interest on the targeting vector into the STAPLR.

III. Genetically Modified Mammalian Cells

[0112] Provided herein are genetically modified cells comprising modifications in one or more of the STAPLRs disclosed herein. The mammalian cells targeted for STAPLR integration may be of any cell type or in any cell state of interest. For example, the cells may be pluripotent cells (e.g., pluripotent stem cells) or differentiated cells. The cells, such as human cells, may be engineered in vitro, in vivo, or ex vivo by gene editing methods such as those described herein. The cells may also be non-human cells, such as cells from laboratory animals (e.g., non-human primates, mice, rats and rabbits), farm animals (e.g., cattle and horses), and pets (e.g., dogs and cats).

A. Stem Cells

[0113] In some embodiments, the mammalian cells targeted for modification at their STAPLRs are stem cells, particularly pluripotent stem cells (PSCs) such as induced pluripotent stem cells (iPSCs; e.g., human iPSCs) or embryonic stem cells (ESCs; e.g., human ESCs). Engineered stem cells can be subsequently induced to differentiate into a desired cell type, referred to herein as PSC-derivatives, PSC-derivative cells, or PSC-derived cells. Stem cells can be the starting point for the potential generation of large numbers of cells of a specific cell type that are delivered for regenerative medicine in patients with different diseases.

[0114] As used herein, the term “pluripotent” or “pluripotency” refers to the capacity of a cell to self-renew and to differentiate into cells of any of the three germ layers: endoderm, mesoderm, or ectoderm. “Pluripotent stem cells” or “PSCs” include, for example, ESCs derived from the inner cell mass of a blastocyst or derived by somatic cell nuclear transfer, and iPSCs derived from non-pluripotent cells.

[0115] As used herein, the terms “embryonic stem,” “ES” cells, and “ESCs” refer to pluripotent stem cells obtained from early embryos. In some embodiments, the term excludes stem cells involving destruction of a human embryo; that is, the ESCs are obtained from a previously established ESC line.

[0116] The term “induced pluripotent stem cell” or “iPSC” refers to a type of pluripotent stem cell artificially prepared from a non-pluripotent cell, such as an adult somatic cell, partially differentiated cell or terminally differentiated cell, such as a fibroblast, a cell of hematopoietic lineage, a myocyte, a neuron, an epidermal cell, or the like, by introducing or contacting the cell with one or more reprogramming factors. Methods of producing iPSCs include, for example, inducing expression of one or more genes (e.g., POU5F1/OCT4 (Gene ID: 5460) in combination with, but not restricted to, SOX2 (Gene ID: 6657), KLF4 (Gene ID: 9314), c-MYC (Gene ID: 4609), NANOG (Gene ID: 79923), and/or LIN28/LIN28A (Gene ID: 79727)). Reprogramming factors may be delivered by various means (e.g., viral, non-viral, RNA, DNA, or protein delivery); alternatively, endogenous genes may be activated by using, e.g., CRISPR tools to reprogram non-pluripotent cells into PSCs. See, e.g., WO 2013/177133 and WO 2022/204567.

[0117] Methods for inducing differentiation of PSCs into cells of various lineages are known in the art. For example, methods for inducing differentiation of PSCs into dendritic cells are described in Slukvin et al., *J Imm.* (2006) 176: 2924-32; and Su et al., *Clin Cancer Res.* (2008) 14(19):

6207-17; and Tseng et al., *Regen Med.* (2009) 4(4):513-26. Methods for inducing PSCs into hematopoietic progenitor cells, cells of myeloid lineage, and T lymphocytes are described in, e.g., Kennedy et al., *Cell Rep.* (2012) 2:1722-35.

[0118] The recombinant PSCs can be differentiated into cells suitable for therapy, including the cells in the endoderm (e.g., lung, thyroid, or pancreatic cells, or progenitors thereof), ectoderm (e.g., skin, neuronal, or pigment cells, or progenitors thereof) and mesoderm (e.g., cardiac cells, skeletal muscle cells, red blood cells, smooth muscle cells, or progenitors or precursors thereof) lineages.

[0119] In some embodiments, the recombinant PSCs are differentiated into cells in the endoderm (e.g., lung, thyroid, or pancreatic cells, or progenitors or precursors thereof), ectoderm (e.g., skin, neuronal, or pigment cells, or progenitors or precursors thereof) or mesoderm (e.g., cardiac cells, skeletal muscle cells, red blood cells, smooth muscle cells, or progenitors or precursors thereof) lineages.

[0120] In some embodiments, a recombinant PSC of the disclosure is differentiated into a cardiac cell. In various embodiments, the cardiac cell is a cardiac progenitor cell or a mature or immature (atrial or ventricular) cardiomyocyte. In other embodiments, the cardiac cell is a cardiac endothelial cell or a nodal cell.

[0121] In some embodiments, a recombinant PSC of the disclosure is differentiated into a human immune cell, optionally selected from a T cell, a T cell expressing a chimeric antigen receptor (CAR) or recombinant TCR, a regulatory T cell, a myeloid cell, a dendritic cell, and/or a macrophage/monocyte (e.g., an immunosuppressive macrophage), or a progenitor or precursor thereof.

[0122] In some embodiments, a recombinant PSC of the disclosure is differentiated into an oligodendrocyte progenitor cell or precursor cell, or an oligodendrocyte. In some embodiments, a recombinant PSC of the disclosure is differentiated into a microglial progenitor cell or precursor cell, or a microglial cell.

[0123] In some embodiments, a recombinant PSC of the disclosure is differentiated into a neural lineage cell, for example a neural crest cells, an astrocyte, a dopaminergic neuron progenitor cell, a dopaminergic neuron, a midbrain dopaminergic neuron progenitor cell, a midbrain dopaminergic neuron, an authentic midbrain dopamine (DA) neuron, a dopaminergic neuron precursor cell, a floor plate midbrain progenitor cell, a floor plate midbrain DA neuron, or a progenitor or precursor thereof.

[0124] In some embodiments, a recombinant PSC of the disclosure is differentiated into a cell of the ocular system, such as a photoreceptor cell, a photoreceptor progenitor or precursor cell, a retinal pigmented epithelium (RPE) cell or a progenitor or precursor thereof, a neural retinal cell or a progenitor or precursor thereof. In other embodiments, an unedited PSC is differentiated into a cell of the ocular system, which is then engineered with a targeting construct of the disclosure.

[0125] In further embodiments, a recombinant PSC of the disclosure is differentiated into a microglial cell or a microglial progenitor or precursor cell.

[0126] In further embodiments, a recombinant PSC of the disclosure is differentiated into a cell in the human metabolic system, optionally selected from a hepatocyte, a cholangiocyte, and a pancreatic beta cell, or a progenitor or precursor thereof.

[0127] In further embodiments, a recombinant PSC of the disclosure is differentiated into an enteric progenitor or precursor cell or an enteric cell.

B. Differentiated Cells

[0128] In still other embodiments, the cells to be engineered are differentiated cells (e.g., partially or terminally differentiated cells). Partially differentiated cells may be, for example, tissue-specific progenitor or stem cells, such as hematopoietic progenitor or stem cells, skeletal muscle progenitor or stem cells, cardiac progenitor or stem cells, neuronal progenitor or stem cells, and mesenchymal stem cells.

[0129] Exemplary differentiated cell types that can be engineered at one or more of their STAPLRs include the cells in the endoderm (e.g., lung, thyroid, or pancreatic cells, or progenitors thereof), ectoderm (e.g., skin, neuronal, or pigment cells, or progenitors or precursors thereof) and mesoderm (e.g., cardiac cells, skeletal muscle cells, red blood cells, smooth muscle cells, or progenitors or precursors thereof) lineages. Alternatively, PSCs can be differentiated into cells in these lineages and then engineered with a targeting construct of the disclosure.

[0130] In some embodiments, a cardiac cell is engineered. In some embodiments, the cardiac cell is a cardiac progenitor cell or a mature or immature (atrial or ventricular) cardiomyocyte. In other embodiments, the cardiac cell is a cardiac endothelial cell or a nodal cell.

[0131] In some embodiments, a human immune cell is engineered. The human immune cell is optionally selected from a T cell (e.g., a CD4+ T cell, a CD8+ T cell, or a Treg cell), a T cell expressing a chimeric antigen receptor (CAR) or recombinant TCR, a regulatory T cell, a myeloid cell, a dendritic cell, and/or a macrophage (e.g., an immunosuppressive macrophage), or a progenitor or precursor thereof such as a hematopoietic stem or progenitor cell.

[0132] In some embodiments, an oligodendrocyte progenitor cell or precursor cell or an oligodendrocyte is engineered.

[0133] In some embodiments, a neural lineage cell is engineered. In various embodiments, the neural lineage cell is a neural crest cell, an astrocyte, a dopaminergic neuron progenitor cell, a dopaminergic neuron cell, a midbrain dopaminergic neuron progenitor cell, a midbrain dopaminergic neuron, an authentic midbrain dopamine (DA) neuron, a dopaminergic neuron precursor cell, a floor plate midbrain progenitor cell, a floor plate midbrain DA neuron, or a progenitor or precursor thereof.

[0134] In some embodiments, a cell of the ocular system is engineered. In various embodiments, the cell of the ocular system is a photoreceptor cell, a photoreceptor progenitor or precursor cell, a retinal pigmented epithelium cell or a progenitor or precursor thereof, a neural retinal cell or a progenitor or precursor thereof.

[0135] In further embodiments, a microglial cell or a microglial progenitor or precursor cell is engineered.

[0136] In further embodiments, a cell in the human metabolic system is engineered. In various embodiments, the cell in the human metabolic system is optionally selected from a hepatocyte, a cholangiocyte, and a pancreatic beta cell, or a progenitor or precursor thereof.

[0137] In further embodiments, an enteric progenitor or precursor cell or an enteric cell is engineered.

[0138] Additional cell types that can be engineered herein to integrate exogenous sequences into STAPLRs are, without limitations, fibroblasts, adipose cells, muscle cells (e.g., skeletal or smooth muscle cells), bone cells, myeloid cells, myeloid progenitor cells (e.g., primitive myeloid progenitor cells).

[0139] The cells may be from established cell lines, or they may be primary cells, where “primary cells”, “primary cell lines”, and “primary cultures” are used interchangeably herein to refer to cells and cells cultures that have been derived from a subject (e.g., a human) and allowed to grow *in vitro* or *ex vivo* for a limited number of passages of the culture. For example, primary cultures include cultures that may have been passaged 0 times, 1 time, 2 times, 4 times, 5 times, 10 times, or 15 times, but not enough times to go through the crisis stage. Primary cell lines can be maintained for fewer than 10 passages *in vitro* or *ex vivo*. In some embodiments, the cells are autologous in the context of cell therapy. In some embodiments, the cells are allogeneic in the context of a cell therapy.

[0140] Primary cells may be harvested from an individual by any suitable method. For example, leukocytes may be suitably harvested by apheresis, leukocytapheresis, density gradient separation, etc., while cells from tissues such as skin, muscle, bone marrow, spleen, liver, pancreas, lung, intestine, stomach, etc. are most suitably harvested by biopsy.

[0141] Any of the foregoing differentiated cell types can be differentiated from PSCs prior to engineering them.

[0142] The present disclosure provides a pharmaceutical composition comprising the engineered cells herein and a pharmaceutically acceptable carrier.

IV. Methods of Identifying STAPLRs

[0143] The present disclosure also provides methods of identifying STAPLRs as sites for safe genomic integration in a mammalian cell (e.g., a human cell). In these methods, the first step is to select a set of cell types for single cell RNA sequencing (“scRNASeq”). Examples of cell types, without limitation, are those referred to herein, including, without limitation, PSCs (e.g., iPSCs), cells in the immune system (e.g., T cells, NK cells, dendritic cells, macrophages/monocytes, or hematopoietic progenitor cells thereof), cells in the cardiovascular system (e.g., ventricular cardiomyocytes, nodal cells, or cardiac progenitor cells), cells in the metabolic system (e.g., hepatocytes and pancreatic beta-cells), cells in the central nervous system (e.g., sensory neurons, motor neurons, interneurons, microglial cells, oligodendrocytes, or progenitor cells thereof), muscle cells (e.g., skeletal muscle cells and smooth muscle cells), adipose cells, and cells in the ocular system (e.g., retinal pigment epithelium cells and photoreceptor cells).

[0144] The second step is to perform an scRNASeq assay wherein the sequencing analysis assigns a unique transcriptome comprising transcribed genes to each cell that passes quality criteria. To pass quality criteria, transcriptomes are filtered to exclude those with high sparsity or missingness and those that are likely derived from more than one cell.

[0145] Next, a Prevalence Score is assigned to each gene. The Prevalence Score is out of “1” and represents the fraction of cells containing at least one transcript of a given gene based on an scRNASeq database of datasets collected. In some embodiments, scRNASeq datasets are obtained from PSCs, dopaminergic neurons and/or their progenitors

(e.g., those at various select differentiation states), microglia and/or their progenitors (e.g., those at various select differentiation states), cardiomyocytes and/or their progenitors (e.g., those at various select differentiation states), oligodendrocyte cell and/or their progenitors (e.g., those at various select differentiation states), or macrophages and/or their progenitors (e.g., those at various select differentiation states).

[0146] After assigning a Prevalence Score, the location of each gene in the mammalian (e.g., human) genome is determined.

[0147] The next step in identifying a STAPLR in the genome of a mammalian cell is to identify neighboring, nonoverlapping genes. By “non-overlapping genes” it is meant that the genes are separated from each other by at least 50 base pairs, at least 75 base pairs, at least 100 base pairs, at least 200 base pairs, at least 300 base pairs, at least 400 base pairs, at least 500 base pairs, at least 1000 base pairs, at least 1500 base pairs, at least 2000 base pairs, at least 2500 base pairs, at least 3000 base pairs, 3500 base pairs, at least 5000 base pairs, at least 10000 base pairs, at least 15000 base pairs, or at least 20000 base pairs on either strand. The transcripts used to calculate genetic distances for identifying non-overlapping genes may be specified by any genomic database, such as NCBI’s RefSeq database and the GENCODE databases.

[0148] In some instances, different genomic databases contain non-consensus gene boundary annotations that may lead to different calculated genetic distances and contrary conclusions as to whether two genes overlap or not. In such instances, two genes are considered non-overlapping if they are determined to be non-overlapping by using at least one genomic database. For example, MLF2 is flanked downstream by its neighboring gene PTMS. As annotated in the NCBI RefSeq database, these genes are non-overlapping, with an intergenic distance of about 13 kb; however, the GENCODE V38 database reports one MLF2 transcript whose transcriptional start site is located within the first intron of PTMS encoded on the opposite strand. In this case, the RefSeq annotations are considered and the GENCODE annotations are not, and this gene pair is classified as non-overlapping.

[0149] Once two or more genes are considered non-overlapping, a Neighbor Score for the pairs of non-overlapping genes or for regions comprising three or more non-overlapping genes is determined. A Neighbor Score is the product of the individual Prevalence Scores and reflects the probability of both genes being transcriptionally active in the aggregate scRNASeq dataset. The Neighbor Score is essentially a ranking of the vicinities of transcriptionally active genes.

[0150] Neighbor Scores are then sorted to obtain a ranking of pairs of non-overlapping genes or a ranking of regions comprising three or more genes. Once the Neighbor Scores are ranked, a pair of genes or a region comprising three or more genes with the best Neighbor Scores is selected and the intergenic region between the genes of the selected pair or region is identified as a potential STAPLR.

[0151] The STAPLR may be targeted for safe genetic integration. Intergenic regions with high-ranking Neighbor Scores are then annotated in order to design homology arms for site-specific integration. In general, sequences to be avoided for integration sites include promoter regions, enhancer regions, CpG islands, epigenetic marks (e.g.,

H3K4Me1, H3K4Me3, and H3K27Ac), DNase I hypersensitivity peaks, conserved regions, and repetitive regions. The UCSC Genome Browser may be used with, but are not limited to, the following gene annotation tracks: GENCODE V32, RefSeq Genes, GTEx RNA-seq, EPDnew Promoters, ENCODE (transcription, H3K4Me1, H3K4Me3, H3K27Ac, and DNase Clusters), GeneHancer, CpG Islands, Conservation 100 vertebrates, and RepeatMasker.

[0152] In selecting a targetable intergenic subregion, known promoter regions and enhancer regions must be avoided. Additionally, conserved regions, repetitive regions, epigenetic marks, and DNase hypersensitivity regions are features that should be minimized in selecting a targetable region. In some embodiments, the targetable intergenic subregion comprises the sequence of an CRISPR endonuclease protospacer adjacent motif (PAM) site. A PAM site is a 2-6 base pair DNA sequence immediately following the DNA sequence targeted by a Cas (e.g., Cas9 or Cpf1) endonuclease. A short oligonucleotide known as a guide RNA (gRNA) is synthesized to perform the function of the tracrRNA-crRNA complex in a CRISPR/Cas gene editing system. A gRNA recognizes gene sequences having a PAM sequence at the 5' or 3' end. Different Cas proteins may recognize different PAMs. For example, Cas9 from *Streptococcus pyogenes* recognizes 5'-NGG-3' ("N": any nucleobase); Cas9 from *Staphylococcus aureus* recognizes 5'-NN-GRR(N)-3'; Cas9 from *Neisseria meningitidis* recognizes 5'-NNNNGATT-3'; Cas9 from *Campylobacter jejuni* recognizes 5'-NNNNRYAC-3' ("Y": a pyrimidine); Cas9 from *Streptococcus thermophilus* recognizes 5'-NNAGAAW-3' ("W": A or T); Cpf1 (Cas12a) from *Lachnospiraceae bacterium* and *Acidaminococcus* sp. recognizes 5'-TTTV-3' ("V": G, A, or C); Cas12b from *Alicyclobacillus acidiphilus* recognizes 5'-TTN-3'; and Cas12b v4 from *Bacillus hisashii* recognizes 5'-ATTN-3', 5'-TTTN-3', and 5'-GTTN-3'.

[0153] Finally, confirmation that the identified intergenic region will safely support an exogenous genetic payload may be carried out by inserting a transgene at a targeted location within the intergenic region using a gene editing system. The gene editing system may be, for example, a CRISPR system (e.g., those using an CRISPR endonuclease disclosed above), a Cre/Lox system, a FLP-FRT system, a TALEN system, a ZFN system, a system that utilizes homing endonucleases, a system that produces homologous recombination, or a system that utilizes non-nuclease dependent viral vectors (e.g., retroviral, AAV, or lentiviral vectors). Constitutive, inducible, tissue-specific, or lineage-specific promoters may be used to direct expression of the inserted transgene.

[0154] In some embodiments, the targeted intergenic region is at least 30, 40, 50, 75, or 100 base pairs in length. In some embodiments, the intergenic region does not comprise a promoter region or an enhancer region. While it may be better for the intergenic region not to comprise conserved regions, repetitive regions, epigenetic marks, and/or DNase hypersensitivity regions, the intergenic region may in fact contain a minimal amount of conserved regions, repetitive regions, epigenetic marks, and/or enzymatic hypersensitivity regions in some embodiments. For example, in some embodiments, the intergenic region will not comprise a CpG Island, an H3K4Me1 epigenetic mark, an H3K4Me3 epigenetic mark, an H3K27Ac epigenetic mark, a DNase I hypersensitivity region, a conserved region, or a repetitive region. However, in some embodiments, the intergenic

region may comprise a CpG Island, an H3K4Me1 epigenetic mark, an H3K4Me3 epigenetic mark, an H3K27Ac epigenetic mark, a DNaseI hypersensitivity region, a conserved region, or a repetitive region. The amount of allowed conserved regions, repetitive regions, epigenetic marks, and/or DNase hypersensitivity regions depends on various factors. These factors include, for example, the size of the intergenic region; the size of the conserved, repetitive, and/or hypersensitivity regions, or epigenetic marks; the presence of gRNA binding sites; or challenges to synthesizing 5' and 3' homology arms for targeting.

[0155] After genomic integration, the transcription level of the integrated transgene is measured and the intergenic region between the selected pair or within the selected region is confirmed to be a STAPLR when the integrated transgene displays sustained transcription (or displays sustained transcription when an inducible promoter regulating the transgene is induced).

[0156] Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention. In case of conflict, the present specification, including definitions, will control. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Throughout this specification and embodiments, the words "have" and "comprise," or variations such as "has," "having," "comprises," or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers. All publications and other references mentioned herein are incorporated by reference in their entirety. Although a number of documents are cited herein, this citation does not constitute an admission that any of these documents forms part of the common general knowledge in the art. As used herein, the term "approximately" or "about" as applied to one or more values of interest refers to a value that is similar to a stated reference value. In some embodiments, the term refers to a range of values that fall within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 10%, or less in either direction (greater than or less than) of the stated reference value unless otherwise stated or otherwise evident from the context.

[0157] According to the present disclosure, back-references in the dependent claims are meant as short-hand writing for a direct and unambiguous disclosure of each and every combination of claims that is indicated by the back-reference. Further, headers herein are created for ease of organization and are not intended to limit the scope of the claimed invention in any manner.

[0158] In order that this invention may be better understood, the following examples are set forth. These examples are for purposes of illustration only and are not to be construed as limiting the scope of the invention in any manner.

EXAMPLES

[0159] In order that this invention may be better understood, the following Examples are set forth. These Examples

are for purposes of illustration only and are not to be construed as limiting the scope of the invention in any manner.

Example 1: Design for STAPLR Targeting

Selection of gRNA

[0160] We used CRISPOR to identify Cas9-based gRNAs that target near the midpoint region where STAPLR construct homology arms flanked the intended site of transgene integration. We excluded gRNAs that had perfect off-targets in the genome. We minimized the use of gRNAs that had a maximum of 3 bp off-target mismatches. Three gRNAs were selected for each STAPLR site as seen in Table 2.

TABLE 2

gRNAs for Targeting Select STAPLR Sites	
STAPLR Site	gRNA Sequence (5'→3')
PRDX1-AKIR1A1	CTTGCAGCACTGCCTAGGCT (SEQ ID NO: 71) TGGTTCTTCAGCAGTGCCT* (SEQ ID NO: 72) CTTGCAGCACTGCCTAGGC (SEQ ID NO: 73)
ACTB-FSCN1	GCCCCACCTTGAAAGATCTGTGA* (SEQ ID NO: 33) TGCCCCACCTTGAAAGATCTGT (SEQ ID NO: 34) CACCTTGAAAGATCTGTAGGG (SEQ ID NO: 35)
RPL34-OSTC	TTATCATGATACTATGTCCC* (SEQ ID NO: 25) AGTTACTGAGCACAGTGCCT (SEQ ID NO: 26) GAGTTACTGAGCACAGTGCC (SEQ ID NO: 27)
AKIRIN1-NDUFPS5	ATTAATGCCTTAAATAGGT (SEQ ID NO: 55) TCTTAAACCTACCTATTAAA* (SEQ ID NO: 56) GCTAATTAAATGCCTTAAAT (SEQ ID NO: 57)

*Preferred embodiments.

[0161] A list of additional Cas9- and Cpf1-based gRNAs for STAPLR targeting is listed in Table 3.

TABLE 3

Additional STAPLR-Targeting gRNAs			
STAPLR Site	Nuclease	gRNA Sequence (5'→3')	
PRDX1-AKIR1A1 Site 1	Cas9	CTTGCAGCACTGCCCTAGGCT (SEQ ID NO: 74)	
		TGGTTCTGCAGCACTGCCCT (SEQ ID NO: 75)	
		TCTTGACAGCACTGCCCTAGGC (SEQ ID NO: 76)	
		TAGGCTGGGCTCAAACCTCCA (SEQ ID NO: 77)	
		CTAGGCTGGGCTCAAACCTC (SEQ ID NO: 78)	
	Cpf1	TGAGAGGATCACTGAGCCC (SEQ ID NO: 79)	
		CTGGAGTTGAGCCCGACCT (SEQ ID NO: 80)	
		TTCTTGTGTTTAGAGAC (SEQ ID NO: 81)	
		AGCCCAGCCTAGGCAGTGCTG (SEQ ID NO: 82)	
		GAGACTGGTCTTGCAAGCACT (SEQ ID NO: 83)	
ACTB-FSCN1	Cas9	TTTAGAGACTGGTTCTGCAG (SEQ ID NO: 84)	
		TTTGTGTTAGAGACTGGTTCTT (SEQ ID NO: 85)	
		TTTGTGTTAGAGACTGGT (SEQ ID NO: 86)	
		GCCCCACCTTGAAGATCTGTA (SEQ ID NO: 38)	
		TGCCCAACCTTGAAGATCTGT (SEQ ID NO: 39)	
		GGGCCAAGAGGCTCAGTCAC (SEQ ID NO: 40)	
		CGATAAAAGAGATCAGGTCGA (SEQ ID NO: 41)	
		CACCTTGAAGATCTGTAGGG (SEQ ID NO: 42)	
		TCCCTACAGATCTTCAAGGT (SEQ ID NO: 43)	
		TTCAAGGTGGCAAGGAAC (SEQ ID NO: 44)	
		TCCCTCCTACAGATCTTCA (SEQ ID NO: 45)	
		CTCCCTACAGATCTTCAAGG (SEQ ID NO: 46)	
		ACCTTGAAAGATCTGTAGGGA (SEQ ID NO: 47)	
		CTTCAAGGTGGCAAGGAAC (SEQ ID NO: 48)	
		ACAGATCTTCAAGGTGGCA (SEQ ID NO: 49)	
		AAGAGATCAGGTCGAAGGCC (SEQ ID NO: 50)	

TABLE 3-continued

Additional STAPLR-Targeting gRNAs		
STAPLR Site	Nuclease	gRNA Sequence (5'→3')
RPL34-OSTC	Cpf1	GGGCAAGGAACCTGGGCCAAG (SEQ ID NO: 51) TAGGGAGGGATAAAAGAGATC (SEQ ID NO: 52) ATCAGGTCCAAGGCCAGGTG (SEQ ID NO: 53) AGGTCCAAGGCCAGGTGCGG (SEQ ID NO: 54) TGAACCACCCACCTGGCCT (SEQ ID NO: 36) TCCCTCCCTACAGATCTCAA (SEQ ID NO: 37)
	Cas9	TTATCATGATACTATGTCCC (SEQ ID NO: 28) AGTTACTGAGCACAGTGCTT (SEQ ID NO: 29) GAGTTACTGAGCACAGTGCC (SEQ ID NO: 30) CAAGCATTATCATGATACTA (SEQ ID NO: 31) TCATGATACTATGTCCCAGGC (SEQ ID NO: 32)
AKIRIN1-NDUF55	Cas9	ATTAATGCCTTTAACATAGGT (SEQ ID NO: 58) TCTTAACCTACCTATTAA (SEQ ID NO: 59) CCACTTCCCTCCTCCATTAA (SEQ ID NO: 60) GCTAATTAAATGCCTTTAACAT (SEQ ID NO: 61) CTTTAATGGAGGAGGGAAAGT (SEQ ID NO: 62) CCTTTAATGGAGGAGGGAAAG (SEQ ID NO: 63) TTTAATGGAGGAGGGAAAGTG (SEQ ID NO: 64)
	Cpf1	ATGGAGGAGGGAAACTGGGGTG (SEQ ID NO: 65) AGAGAAATCTATGTCACCCAC (SEQ ID NO: 66) AAGGCATTAAATTAGCGTTTGC (SEQ ID NO: 67) AATAGGTAGGTTAACAGAGAT (SEQ ID NO: 68) CATGCACTACTTTAAATTTT (SEQ ID NO: 69) AAGTAGTGATGCAAACGCTA (SEQ ID NO: 70)
PRDX1-AKR1A1 Site 2	Cas9	AAGGGCCAAGGGAGTTAGTG (SEQ ID NO: 87) AGCTCCCTCACTAACCCCT (SEQ ID NO: 88) AGGGCCAAGGGAGTTAGTG (SEQ ID NO: 89)
PRDX1-AKIR1A1 Site 3	Cas9	ATGAAAAATAAGCCCGGTAG (SEQ ID NO: 90) CAGGCTGAGTCACCACTACC (SEQ ID NO: 91) ACAGGCTGAGTCACCACTAC (SEQ ID NO: 92)

[0162] For each STAPLR site, human iPSCs were nucleofected with each individual gRNA complexed with Cas9 nuclease in the form of a ribonucleoprotein (RNP). Three days later, the nucleofected cells were harvested, genomic DNA was extracted, and PCR amplification of the genomic region flanking the intended cut site was performed. Purified PCR product was sequenced and the sequencing data were analyzed for overall cutting efficiency through Synthego's ICE Analysis Tool (available at Synthego's website) (FIG. 1). gRNAs were considered to be efficient when showing greater than 50% indel editing.

[0163] The data show that there was at least one efficient gRNA (>50% indel editing) per STAPLR site. The gRNA that had the greatest overall cutting efficiency was selected for use in future experiments to integrate transgenes at STAPLR sites.

Design of STAPLR Homology Arms

[0164] A list of gene neighbors consisting of genes that were both highly expressed was generated. This list was filtered to remove gene pairs that contained at least one gene that is a known tumor suppressor gene or oncogene. Initially, gene pairs with less than 5 kb intergenic distance between them were discounted. However, gene pairs with only about 100 base intergenic distance between flanking genes can also be annotated and tested. Promoter regions, enhancer regions, CpG islands, and regions containing epigenetic markers were avoided in the design. Subregions that avoided regulatory elements and were capable of being synthesized

in a donor plasmid were classified as potential homology arm regions and were used as the basis for a gRNA search (Table 4).

TABLE 4

Parameters for Selecting STAPLRs
STAPLR Site Parameters
Site is flanked by adjacent highly expressed genes
Neither flanking gene is a known tumor suppressor gene or oncogene
>50 base intergenic distance between flanking genes
Site avoids annotated regulatory elements and repetitive regions
Site allows for homology arms to be synthesized and cloned into donor construct
Site has CRISPR gRNAs available that have predicted high efficiencies and low number of off-targets

[0165] After selecting gRNAs with predicted high efficiency, homology arm sequences were finalized to center selected gRNAs within an 800 bp left homology arm and an 800 bp right homology arm that flanked the intended site of transgene integration. Table 5 indicates the intergenic distance in base pair between the two gene neighbors for each exemplary STAPLR site, along with the coordinates for each set of STAPLR left and right homology arms based on the hg38 human reference genome. Gene distances were calculated using NCBI's RefSeq database.

TABLE 5

Intergenic Distance Between STAPLR Gene Neighbors and STAPLR Homology Arm Coordinates					
Gene	Homology Arm Chromosomal Coordinates (Hg38)	Distance from Intergenic Distance (bp)	Distance from Insertion Point to Upstream Gene Neighbor (bp)	Distance from Insertion Point to Downstream Gene Neighbor (bp)	
RPL34-	chr4: 108,638,253-	20,100	8,568	11,532	
OSTC	108,639,852				
ACTB-	chr7: 5,545,025-	62,214	15,223	46,991	
FSCN1	5,546,624				
AKIRIN1-	chr1: 39,012,798-	20,229	7,533	12,696	
NDUFS5	39,014,397				

TABLE 5-continued

Intergenic Distance Between STAPLR Gene Neighbors and STAPLR Homology Arm Coordinates					
Gene Neighbors	Homology Arm Chromosomal Coordinates (Hg38)	Distance from Intergenic Distance (bp)	Distance from Insertion Point to Upstream Gene Neighbor (bp)	Distance from Insertion Point to Downstream Gene Neighbor (bp)	
PRDX1-	chr1: 45,525,540-site 1	27,888	3,449	24,439	
AKR1A1	45,527,139				
PRDX1-	chr1: 45,524,377-site 2	27,888	2,286	25,649	
AKR1A1	45,525,976				
PRDX1-	chr1: 45,528,804-site 3	27,888	6,713	21,222	
AKR1A1	45,530,403				

[0166] Sequences for the left and right homology arms of the targeting constructs based on the hg38 Human Reference Genome are shown in the table below.

TABLE 6

STAPLR Left and Right Homology Arms		
STAPLR Site	Left Homology Arm	Right Homology Arm
AKIRIN1- NDUFS5	TTCTCATTTAGCAAGAGTGGAAAATG TTGTGGATCC CAGGAGAA GCTGAGC TAGTAGTAGGGCTAGATAATAGAGG GCTTGACTAATAGGAGTTGGGCTT ATCTGTGGAAATCTGAAGGGTTGG GAAAATGAATTCAGGGCTAGGTCAAG ACAGAAGGGCAAATCAGACAAGAC ACACAGCCACTAAGAGCCTTTTTT TTTTTGTCCAGTCAGGGATATGAGG CCTATATATTATTAGGAAGGATGATT AACCATGGCTGTAGCTCACCTAACAG GAACGTGGCTTGGGGCGGGTGGGT GGCTCACGCCCTGAAATCCCAGCATT GGGAGGTAGAGGTGGCGGATCACCTG AGGT CAGGGTTCAAGACCAGCCTGAC CAACATGGAGAAACCTCTCTCTACTA AAAATACAAAATTAGCCGGCTGGTGT GTGATGCCGTAACTCCCTGGTACTCG GGAGGCTGAGGCAGGGAGAACCGCTGA ACCTGGGAGGTGGAGGTTGGCTGGTGA TGAGATCACCCATTGCACTCAGCCT GGACAACAAAGAGAAAATCTATCTCA AAAAAAAGAGAGGAATGTGGCTCTTG AAGATCTTCTGACTCTCATCTCCCT GTCCCTTAGACACACCTTCAAAGATC CTTTTCTTTGGTTGGCTGTCTTT CCCCCTCACTTACCTAACCTAACGCCA ATTCCATAGCCACTGCTGGTCTCTTA ATGGAGGAGGAAGTGGGGTGAATAG ATTCTCTTAAACCTTAC (SEQ ID NO: 17)	TATTAAAGGCATTAATTAGCGTTTG ATGCAC TACTTTAAATTCTCTCTCA ATTGCTTGAGTCCAGGAGTTGAGGTT ACGGTGA ACTATGATTTCACCTTGCA GTCTGAGTGGCAACAGAGAGAGACC CTCTTAAAGAAAAAAACTTTTTTCC TCTCAGACTCTGCTCC TAGCATTGACT CACCCATCATTTTTTTTTTCTTCTT GAGATGGAGTTCAAGAAAAAACTCCA GGCTGGAGGCACTGGTATGGTCTCGGC TCACTGCAACCTCCGCCTCCGGGTTC AAGCGATTCTCCCTGCCCTAGCCTCCC AAATAGCTGGGATTACAGGCACCCGCC ACCACACCCAGCTAATT TTTGTATT TTAGCAGAGACGGGGCTCTCCATATT AGCCAGCTGGTCTGGAAACTCTGACT TCAAGT GATCCACCCACTTGGCTCC CAAAGTACTGGGATTACAGGCACGAGC CACCGCGCCGGCCTCACCTATTATA TCTATTCTGTCTTTAAATAGTGA GACTTCACAGCTCAGTCCTTGTCT GAACCTCCATTCTGAGTCGAACCCAG GTACACCTTAGGAAGGGCGGGATCCCG TTATCCTGTCCCTGAACATCAACTACT GAGACAGCTGATGAGCTGGCTCGGT TGTGAGGGCCAGGGCAGTGGGCACATC AGAGCCGAGCGTACAGGCACATCAGG CAGAGCAGACCCAGAGGGCAGGGAGACA CACTGGGGCAGGGCAGTGGGCACATC TGGACCACTCGCAGGG (SEQ ID NO: 18)
RPL34- OSTC	TTATTCCTCACAGGGACAGCTCCCAA CTCTTACCTCTCTCTAACATGCAACA CTCACTCTCACTCTCAGCAGCCTTATA ATATCTGTAAATGTAGAAACTATCAGT TATGAACCTCTGTAACTCCCTTTCTAT CTTGGCACCTCAAACCTTAAATGTAT GTAAAAGCATGGCTTGAATTTGGTA AACTGGAGCTATCAATCTGCGCTG TGTCAACTGAGCATGTTGCCAATT CTATGAGCTCAGTTCTTATCTAAA AAATGAAACTAATGCTACTCCAA AGAATTGTTGGATGGATTAAGTAAGAC AACCTAATATAATTGCTGTGCATAGTA	CACAGGGCACTGTGCTCAGTAACCTAAC CTATTATAACTGTGCTGCTGACAAT ATGTTGTA CTACTGCGCAAAATCAC CAATGTCATATGTCAGCTAAATGTC TTCTTCAATTCAATCTATCTGACAT GCTGAAACTTCCCTTCCAAAGGCTTC TTTGCTATCGCTCCCTGGCTCTTC TTCCCTTGATCTGCTGTATTACT GTAGTATCTCTCTAAATTATGGGGTGA TTCCCTAGGCACTCATCTCAGCTTTC TTCTCTCTCACTCTTTAACTTTACTC TGTGAATTCTGTGAATTATACACTAC ATTGGGTGTGTGTGTGTGTCTGT

TABLE 6-continued

STAPLR Left and Right Homology Arms		
STAPLR Site	Left Homology Arm	Right Homology Arm
	AATCCTAACCTGTTGGTGCCTTATAC TCTAAATCTCTTCTCTCTCTTTCT TTAGAAGATGGCATGCATCTTTCAA GGTAAACATTACCTCTGTGAACCTA TCCCATACTCTTGCCATGCCCTAA TTTGTTCAGGTTACTTAACCTCTCC TTCCCTCTGACATCCTGGCCCTTC TCTCCACAGTTTATTCTCGTCTTA CACAAACATGCCAAGTCATCATCC ATAAAAACCTCTGCAATTACCACTTCA GCTCTTCTCTGTTCATACCCAGCT GTGTTGGGTGCATGAATCATTCACTA CTATTATTTAAATTCATCATCACTA CTTCTCCCTCTGCAATCTGGCTTC CTGCCAACATGCTACTGAAATTATGA ATTAATTCAATCAATTTCACAGCAT TTATCATGATACTATGT (SEQ ID NO: 19)	GTGCACGTGCATTTCTAGGAAGGAG TTCATAGCTTTATCAGCTCTGCAAGG AGGTCTATGACCCTCCAATAACAAACC CTGACCGAGTCCACCTCAGCTGTAGAC ACATTCTCCACTGTGATGACTTCCAT TATCCACATGCTGCTGATTCTGG ATCTCAACCTAAACCTCTTCCCAT GCTCCAGCGACCATCTTCCAACTCTG CCTAGGTCCCTCATGGAGGGTCTTCT AGAATAATAGTTCTCATCAGGCAGGTG GCAGAACCTCTGCTCCACTCTTGC ATGCCCTCTGTGAGGCTTGGGTGGGT TGGGGGTGATGGGAGGGTGCCTGTAT AATCAAGACTCCCTGGGGTGCAGAACT CTGTTCTGACCTTGCCCTTGCTATCC ATGGAATTACTGCTCTATCTGGCCAT TTTCCCATTCTCTTCT (SEQ ID NO: 20)
ACTB-FSCN1	GAAGCCGGGTGGTGGTGGCATGCCCTG TGGTCCAGCCACTTGGGAAACTGAGG TGGAGGATGGCTCAAGCCCAGGAGG CAAGGTCTGTGAGCTATGATTGTGA CACTGCACTCCAGCCTGGCAACAAAG GGAGACCCCTTAACTAACAAATGAAT GAATGAATGAATGAAGGCACACCT GAGCTGCATTCTGGCTGCCAACCT GCCAGCCCCATCCTCAGCCCTCCCT GAGTCTGAGGCCCTCAGGGCTCCACA CAGGGCCAGGCTCATCTGGTCTGC AAATTGACCTTCGCTTCCATTCCCT GCCACACTGCTCACGGGTACCATGGA TGCCTGGATAACTGCCACGGCTCTCTG GAGAAGGCCCTCGTCACTCGCCCT GATGTCAGGTATTCAACCTGCAATCC TCCCTGAGGTGCTGGGCTCTCCAGAG CCCTGGGCTGTGTCAGGGACTCAGG GGAGAACGGGCCCCAACAGCTCTTC CTCACAGACTTCAGAGCTGGGGC CTCCCTGATCCCTCCAGATCCAAG GCCCTGCTGGGACAGCTGGCAGACT GGGACTCCGTGAGGCTGGCTTGACT TGGATATTCTGGTTCACACAGCA GGCACCGTGGCTGTAGTAGGCGTGCAT GGAAAGTCAGGAGGAGCAGACCTGTCA TCTCCCTGCTGGGACAGCCGGT CAGGTGTCTGGGGCTGGGGCAG TGACTGAGCCTTGGGCCAGTCCCTT GCCACCTTGAAAGATCT (SEQ ID NO: 21)	GTAGGGAGGGATAAAGAGATCAGGTCC AAGGCCAGGTGGCGGTGGTCACGCCCTG TAATCCACGACTTGGGAGGCCGAGC CAGGCAGATCGCAGGTCAGGAGTTCG AGACGAGCCCTGCCATCATGGTGAAC CCCGTCTACTAAAAATACAAAATT GCCAGGCATGGTGGCAGGGCCCTGTA ATCCCAGCTGAGGAGGAGAATCGTT GAACCCGGGAGGTGGAGGTTGCAGTGA GCCAGAGGAGGCAACTGCACTCTG CTGGGTAACAAAGCAAGACTCCATCTC AAAAAAAAAAAAAAAAAAATCAA GTCGGCGAGCCCCAGGAAGCAATCT TGGGCTGGCACACCCGATTTCTGTGA CATGAGGCAGCAGACACTCGTCGCTG TCACAGGGTTGCTGGGGTTGGGGCTA GAGGAGAGGAGGAGCCTGTTGAGGG CTCCAAGGAAGGGACAGAGGAGGTTG TCATGGGCTCTGGAGACACCTTTGG TGGTCTGCCCCATCTGGGGCAC TTTGGGGAGGTGAAGGAGGGAAAGCATT AAGGGACAAGACCCCCCGTCCCAATT TCTCTGGAGGCCAGGGTTCTGTAT TCGAAGAACAGGTGTCACACCCAGG AAGTCACACTGATGGCATCTGCCCTGG GCATCAGCATTTAGGGCTGATCACTGA GGTCTGCACCTCCAAAGGGCTGTG CCATTCTGGCGCCCCAAAGGGAAAG AAAAACTCTGAATGTGCACCGGGACA GGACCCATCCCATGGG (SEQ ID NO: 22)
PRDX1-AKR1A1 Site 1	AGAAAATTGCCACCTATTGTATTCA ATATGCCAGGCTGGCTGGGACTT GAGTTTATTCTTAATTCAAGAACAA GGCCAGGACGGCTGGCTCATCGTGA ATCCCAGCACTTGGGAGGCTAAGGCG GGCAGATCAACAGGTCAAGGAGTTGAG ACCAGCTGCCAACATGGGAACATCC CATCTCACTAAAAATCAAAATTAG CTGGGTTCTGGTGTGCACCTGTAGT CCCAGCTACTCAGGGGCTGAGGCGAG AGAATCTTGAACCCAGGGCGAG GTTGCAGTGAGCCAGATGCGCCACT GCACTCCAGCTGAGTGCACAGAGTGAG ACTCCGCTAAAAAAAGAACAAAAG CCGGGCATGGTGGCTCAAGCCTGTAA CCCAGCACTTGGGGAGGCAAGGG CAGATCATGAGGCTAGTAGTCCAGAC CAGCCTGCCAACATGGTAACAACTG TTTCTACTAAAAATCAAAATTAGCT GGCGTGGTGGCAGGCACTGCCAATCC CAGCTACTCAGGAGCTGCCAGCAGGAG (SEQ ID NO: 23)	CTAGGCAGTGCAGCAAGAACAGTCTC TAAACAAACAAAGAAAAAAACAGGA AAAAAAAGAGTTATCCTAGCCTAGAG AAATGTTAACAGTCTATCCCTTAAGA TCATAAAAGTGAATAACTGGACAGAG TACTGAAAAAGAAGAAATCTGTGAA CAAGTCTAAGTGATTGTAGAGTCTCT CCAAAGGACAGTAAAGATTTCAGAAC ATGGGGAGGGCTTCTTGAGGGTC TGATGTATACCTCTAGTTCTATTTTA GATATATTCATTGTGCTCTTCTT TGCTCACCATAGCATGGGATTATTC CCCTTGTTTATTATATTGTAGTAA CAAGTCTAGCACAGTAACCTGGCATAGA AGTACTTAAATAATGTTACAATGTGTT GAAGACTTTGAGATGTAGAATTAAAT TATTGAACTACGTTCTAGGAAGTGT CTCTGGAGCTGTGATGAAATGAATT TAGTGACTAGGAGATTAGGCAAGAGA GCAATTGGGAGGCCACGGCAATGATCC AGACAAGAGAAGTCTCAGAACCTGTAAT

TABLE 6-continued

STAPLR Left and Right Homology Arms		
STAPLR Site	Left Homology Arm	Right Homology Arm
	AATCACTTGAACCCGGAGGCAGAGGT TTTAGTGAAGCTGAGATCCCGCACTGC ACTCCAACCTGGCAACAGAGCAAGAC TCCTGCTCAAAAGAAAAAGAAAAGAAA AAAGAAAAAGAAAAGAAAAGAAAAGAAC TAAAAGAGAGTGGAGCACATTGGCAT GTGACTATAGTCCCAGCTACTGGTAG GCTGAGGTGAGGAGTCACTTGAGCCC TGGAGTTGAGGCCAGC (SEQ ID NO: 23)	ATGACATTAACAGCAGGGTAGAAAGAA GGAAATACAAGGTATATTCTTCAGGC TTTAGGATTAATTGGATTAGGTGAA AGGAAAGAGTAAATTAAATCCAGTTTC TTGGTGACATGAGTGTGATTAACC TGTGACATAAGGAAACACTGAAAGAAAA GCAGGTTGAGGTTAAAATAATGTTA AACAGGCCGGAGCGGTGGCTCACGCC TGTAATCCAACATT (SEQ ID NO: 24)
PRDX1- AKR1A1 Site 2	AATAATTATAATCTTATTGTGATTA TCTATGAGTCTAGTACAGTACTTAGT TCTTATTATAAATTAGAAATGTTAGCT GCTATTAGTATAACTCTTCCATCCTTC ACACTTATAACACCCAGAACACAGTGT AGTGTGTTAACGCGCTCAGACTTTTT TTTAAGATGGAATTGCGTTTCTTG CCCAGGCAGGAGTGCATGGCACGACT TCTGCTCACTGCACCCCTCCACCTCCAG CGTTCAGAACATTCTCTGCTCAGCC TCCAAGTAGCTGGGATTACAGCGCC CACCAACACGCCCGCTAATTTTTGT ATTTTATGAGAGACGGGTTTCTAGTA TGTCAAGGGCTGGTCAAACCTCT GACCTCAGTTGATCTACCTGCCCTCAGC CTCCCAAAGTGTGGGATTACAGGAGT GAGCCACTGAGCCGGCCGCTTCAGA CTTTGTTGAACCTGAAACAACTCTT CCCTGGAGAAAACATGCTTCTAGCAT TACTTTCTCTCCGCTGTTCTTACCT TTACTCATTCCAGGATTACAATTCC CCCTCTTGTGTCGAAACAAATAGCTT TATAGCCAAAATCTTTAAGCAAT TTTGTGTTACATTTCTCTGCTGCCT GCTAGAGTCACTTCTGGAGAGTAGGA TGTGGGTGACATCTGTCCTGTCCTCC CAAAGTCACACAGCTGATATA AAATAGGAGCTCAGTCAGTGAGCATT AATCAATAGTATCATGTTCTGCTCT GTCCCCAGCTCCCTCAC (SEQ ID NO: 93)	TAAACTCCCTGGCCCTTATAGTGGGAT AATGAATGAGGAAGTGCCTCATG TCATTGAGCATCTCAATTTTGTTATT TATCATACTACTTGCATAGACGTGAT TTATTGCTTTTCTTCTTGCTGCTGA GGCTGAATAGTTAGTTGGCATTATTC TTTTTTAATTTTTATGTACTTTAT TCTTTTTATCAGGCAGCCACCCGAGC CAGCAGCCAGGGTAGGCTAGAGAGAC TCCCTCTTCTTGCGAGGCAAGG GCAGCATGAAAGTTAGAGCTTAA TGGATGGTTATTGGCTATGGCAGGCT AAACCAATACACAGTACATACAGTCAT ACTTCAGCCCAAAGAAAATTGCCACCT ATTGATTCACCATATGCCAGGCATTG GGCTGGGACTTGGAGTTATTTC ATTCAAGAACAAAGGCCAGGCACGGTGG CTCATGCTGTAATCCAGCAGTTGG GAGGCTAAGGCCGGCAGATCACAAGGT CAGGAGTTGAGACCAGCCTGGCAAC ATGGAGAAATCCATCTACTAAAAA TACAAAATAGCTGGTTTCGTGGT TGCACCTGAGTCCCAGCTACTCAGGG GGCTGAGGCCAGGAGATCTTGAC CAGGAGGGGAGGTGCAAGTGGCGA GATGCCGCAACTGCACTCCAGCCTGAG TGACAGAGTGAGACTCCGCTAAAAAA AAAGAACAAAAGCCGGCATGGTGGCT CAAGCTGTAATCCAGCAGTTGG GGCCAAGGTGGCAGAT (SEQ ID NO: 94)
PRDX1- AKR1A1 Site 3	TGTTCACGTGTTGCTGTCGACCCCTC TCCCCACAATGCTTGTGACCCCTGAC ACATCCCCCTCCGAGAACACCCAC AAATGATCAATAACTAAGGGAACT CAGAGGTGGGGATCTCCATATGC TGAACGCTGGCTCCCTGGGGCTCTTA TTCTTCTCTATACTTGTCTGTG TCTTTCTTTCCAAGTCTCTGTT CACCTAACAGAACACCCACAGGT GGAGGGCAACCCACCCCTCACCATC TCCACAAAAAAATAAAAATTAGCCAG TGTGTTGTTGTCGCTGTGGCTCAA CTAGTCAGAGGTGAGGTGGGAGGAT TGCTTGAGTCAGAGCTGAGGCTGC AGTGGAGCCGAGATGGCACCCAGTACT CCAGCCTGGGTGACAGTGGAGCAG TGAGACCTGTCCTAAAGATAATAAA ATAAAAATAATGGAAACCCAAATCACAA GACAGCACTCATTTCTTTATT TTTATTCTTGTGAGACAGAGTCTTG TCTGCTGCCAGACTGGAGTGCAGCAG CATGGTTGGGCTCACTGCAACCTCCA CCTCCTGGGTCAAGTGTATTCTGTG CTCAGCCTCCCAAGTGTGAAATTAC AGGTATGTGCAACCATGCTGACTGAT TTTGATTTTAATAGAGACAGGGT ACGCCATGTTGCCAGTGTGTTCGA	CCGGGCTTATTTTACACAGATGATT TAGAATATGTTATGTTAAATT AGATTATCACTAAAGAAAAGAGTTGA ATTCTCAATTACAAGGATTCAATT TTAAAATATATTGAGTGTAGAG CCTACAAATACCAATTTCAGAAACACT GTTATCAAGGATATCTATTGTTAAAA TTGAATGCAACTGAAAGTAACAGATTA ACTGCAATAACAAAGCTAACCAATA GGGATTAAATTTCCTACAAACAG GAGACAGGGCCGGCAGTGGCTCAC GCCTGTAATCCAGCAGTTGGGAGGT CGAGGCGGGTGGATCACTTGAGGTGG GAGCTCAAGACCAGCTGACCAACATG GAGAACCCGGTGTCAAGAAAATAC AAAATTAGCCGAGCTGATGGTACATG CCTGTAATCTCAACTACTCGGGAGGCT GAGGCAAGGAGAATGGCTGAACTGGG AGGTGGAGGTGCACTGAGCCAAGATC GCGCCATTGCACTTCAAGCCTGGGCAA CAAGACGAGCTCGTCTCAAAAAAAA AAAAAAAAAAAGCAAGTAGACAG GGCCAGGGAGTGGCTCATGGTTGA ATCCCAAGCATTGGAGCTCAGGAGTTGG AGAGGATCATTGAGCTCAGGAGTTGG AGACCAGCTGGCGACATAGGGTAC CCATCTCACAAAAATAACAAACAA

TABLE 6-continued

STAPLR Left and Right Homology Arms		
STAPLR Site	Left Homology Arm	Right Homology Arm
	ACTCCTGGTCTCAAGGGATTCAACCAT CTCGGCCTCTCAAAGTACTGGGATTAC AGGCTGAGTCACCACTA (SEQ ID NO: 95)	AACAAAAAAATAGCCAAACATGATGACG CATGCCCTGTAGTCCCAGCTACCCAGGA GGCTGAGATGAGAGAAT (SEQ ID NO: 96)

Example 2: Testing of Inducibility and Transgene Expression at STAPLR in a Pooled Population of Targeted iPSCs

[0167] To test for robustness of inducibility and transgene expression at each of the four annotated STAPLR sites (PRDX1-AKR1A1 (Site 1), ACTB-FSCN1, RPL34-OSTC, AKIRIN1-NDUFS5), the dual component doxycycline-inducible rtTA/TRE system Tet-On 3G from Clontech/Takara (as described in U.S. Pat. No. 9,127,283, incorporated by reference herein in its entirety) was used. For the constitutive component, the Tet-On 3G rtTA (reverse tetracycline transactivator) was expressed biallelically from the GAPDH locus, via the inventors' "sustained transgene expression loci" (STEL) approach (FIG. 2) as described in WO 2021/072329, incorporated by reference herein in its entirety.

[0168] To test inducibility of transgene from a STAPLR site, the TRE3G promoter was used to test expression of an eGFP cargo. A Kozak sequence was included to enable translation initiation and an SV40 PolyA sequence was added to enable transcription termination. In the presence of doxycycline, the rtTA protein binds to and activates the tetracycline-response element (TRE) minimal promoter (FIG. 3). For each STAPLR site, a parental iPSC line with bi-allelic rtTA integration at GAPDH (GAPDH::rtTA iPSCs) was nucleofected with a selected high-efficiency RNP and the corresponding STAPLR targeting construct (STAPLR left homology arm-TRE3G promoter-eGFP-SV40-STAPLR right homology arm). Pools of cells that received both STAPLR RNP and STAPLR targeting construct were fed with media containing 2 µg/ml doxycycline starting at day one post-Nucleofection (FIG. 4) and continuing to day seven post-nucleofection (FIG. 5) in order to induce GFP expression. The parental rtTA iPSC line was also given 2 µg/ml doxycycline media as a control. GFP expression was monitored over the course of a week by fluorescent microscopy. An increase in GFP intensity was observed as cells were treated for longer duration with doxycycline. Preliminary testing of this rtTA/TRE-based transgene expression system at STAPLR indicates robust inducibility and expression of GFP in a pooled population of STAPLR site-targeted iPSCs.

Example 3: Testing of Inducibility and Transgene Expression at STAPLR in a Clonal Population of Targeted iPSCs

[0169] Parental GAPDH::rtTA iPSCs were nucleofected with RNP and a STAPLR targeting construct at each of the four STAPLR sites followed by plating each pooled population of STAPLR-targeted iPSCs at clonal density. Individual clones were picked and screened by PCR across the junctions of the left and right homology arms to confirm accurate integration of the TRE3G-eGFP-SV40 at each of the four STAPLR sites. Targeted iPSC clones were expanded

and treated with media containing doxycycline at a range of 0.1 µg/ml to 5 µg/ml from 0 to 68 hours. Cells were collected at a time course of 0, 3, 8, 24, 48 and 68 hours over this time course of GFP induction and flow cytometric analysis was performed (FIG. 6). The results indicate that maximal GFP induction from all four STAPLR sites can be seen from administration of 0.1 µg/ml doxycycline and after 48 hours of doxycycline administration. STAPLR sites vary in their maximal expression levels of GFP, with the PRDX1-AKR1A1 site demonstrating the highest expression of GFP in doxycycline-induced iPSCs. One clonally derived line from each STAPLR-targeted site and a wildtype unedited iPSC control line was then treated with media containing 2 µg/ml doxycycline for 72 hours (FIG. 7). The AKIRIN1-NDUFS5 STAPLR line showed slightly delayed GFP induction so treatment with media containing 2 µg/ml doxycycline was increased to 6 days (FIG. 7). The results indicate that all four treated STAPLR-targeted iPSC lines could induce high levels of GFP expression, with the PRDX1-AKR1A1 site again demonstrating the highest expression of GFP in doxycycline-induced iPSCs, while the wildtype unedited doxycycline-treated iPSC control line did not express GFP. In all instances, cells that did not receive doxycycline treatment did not express GFP.

Example 4: Testing of Inducibility and Transgene Expression at STAPLR in iPSC-Derived Myeloid Progenitors

[0170] Clonally-derived STAPLR iPSC lines were differentiated into myeloid progenitor cells to demonstrate that transgene integration at STAPLR maintains sustained transgene expression in differentiated iPSCs (Douvaras et al., Stem Cell Reports (2017) 8(6):1516-24). 2 µg/ml doxycycline was added to each STAPLR-targeted clonal line at day 12 of differentiation and doxycycline was replenished daily for three days. Adherent myeloid progenitors were harvested for flow cytometric analysis of GFP induction at day 15 of differentiation. Three of the four TRE-eGFP-SV40 STAPLR lines (PRDX1-AKR1A1, ACTB-FSCN1, RPL34-OSTC) demonstrated efficient GFP induction in heterogeneous adherent myeloid progenitor cells, compared to differentiated cells that did not receive doxycycline (FIG. 8). A wildtype unedited iPSC control line differentiated using the same protocol and similarly treated with doxycycline did not show induction of GFP. One of the TRE-eGFP-SV40 STAPLR lines (AKIRIN1-NDUFS5) demonstrated delayed GFP induction under fluorescent microscopy. This cell line was replenished with doxycycline for an additional three days and adherent myeloid progenitors were harvested for flow cytometric analysis at day 18 of differentiation. FIG. 8 shows the bimodal GFP induction seen from the myeloid

progenitors harvested at day 18 of differentiation. In all instances, cells that did not receive doxycycline treatment did not express GFP.

[0171] STAPLR-targeted lines were further differentiated past 30 days to the point where non-adherent myeloid progenitor cells could be collected in suspension culture. 2 µg/ml doxycycline was added for six days and the non-adherent myeloid progenitor cells were collected for flow cytometric analysis of GFP induction. All four TRE-eGFP-SV40 STAPLR lines cultured past 30 days demonstrated efficient differentiation into triple-positive myeloid progenitors as defined by >80% co-expression of the cell surface markers CD45, CD14 and CX3CR1 (FIG. 9). The doxycycline treated STAPLR lines also demonstrated efficient GFP induction in heterogeneous non-adherent myeloid progenitor cells, compared to a doxycycline treated wildtype unedited control line, with some variability in maximal GFP expression levels (FIG. 10). This data demonstrates that transgene integration at all four STAPLR sites permitted sustained expression of the transgene under external promoter control during and post-differentiation into myeloid progenitor cells.

Example 5: Derivation of Human Induced Pluripotent Stem Cell Line with Inducible Expression of CD19t-IL12 from the PRDX1-AKR1A1 STAPLR Site

[0172] A parental iPSC line with bi-allelic rtTA integration at GAPDH (GAPDH::rtTA iPSCs) was transfected with a selected high-efficiency RNP for the PRDX1-AKR1A1 STAPLR site (Site 1) and a STAPLR targeting construct comprising a doxycycline-inducible promoter (TRE3G)-driven CD19t-IL12 cassette flanked by PRDX1-AKR1A1 left and right homology arms. CD19t was included here as a non-biologically functional cargo; it served as an epitope marker for surrogate detection of IL-12 transgene integration by flow cytometry. Two different gRNAs and their corresponding nucleases were used for targeting at the PRDX1-AKR1A1 STAPLR site. Either a Cpf1-based guide RNA with sequence 5'-GAGACTGGTCTTGCACT-3' (SEQ ID NO: 83) or a Cas9-based guide RNA with sequence 5'-CTTGCAGCACTGCCTAGGCT-3' (SEQ ID NO: 71) were selected to generate clonal lines. The GAPDH::rtTA constitutively expresses the reverse tetracycline transactivator (rtTA) from the GAPDH locus. In the presence of doxycycline, rtTA binds to the TRE3G promoter and induces expression of CD19t and IL-12 driven by the TRE3G promoter (FIG. 11).

[0173] Single cell suspensions of GAPDH::rtTA iPSCs were prepared for transfection with either Cpf1 or Cas9 gRNA RNP complexes and the PRDX1-AKR1A1 targeting pTRE3G-CD19t-IL-12 DNA donor template. Two days post transfection, cells were treated with doxycycline (2 µg/mL) for 48 hours to induce CD19t-IL12 expression that was analyzed using live cell imaging of AF488 conjugated anti-CD19t antibody staining (FIG. 12, Panels A and B). Cells were then dissociated and plated at single cell clonal density. Four days after clonal density plating, growing

colonies were treated with 2 µg/mL doxycycline for 48 hours to induce CD19t-IL-12 expression. Colonies were analyzed with live cell imaging using an AF488-conjugated Ab against CD19t after the 48-hour doxycycline treatment. CD19t positive colonies were identified (FIG. 12, Panels A and B, marked under "Clonal density").

[0174] The data demonstrate that the CD19t-IL-12 expression cassette integration at the PRDX1-AKR1A1 STAPLR site permitted sustained expression of the transgene under external promoter control in both pooled and clonal populations of STAPLR-targeted iPSCs after treatment with doxycycline.

Example 6: Induction of Reporter Transgene Expression at Various Sites Within a STAPLR Intergenic Region in Targeted iPSCs

[0175] To test for robustness of inducibility and transgene expression at two alternate sites within the PRDX1-AKR1A1 intergenic region (PRDX1-AKR1A1 Site 2 and Site 3), we again utilized the dual component doxycycline-inducible rtTA/TRE system. The TRE3G promoter was used to test expression of an EGFP cargo. A Kozak sequence was included to enable translation initiation and an SV40 PolyA sequence was added to enable translation termination, as per the design of the original PRDX1-AKR1A1 targeting construct. In the presence of doxycycline, the rtTA protein binds to and activates the TRE minimal promoter. A parental iPSC line with bi-allelic rtTA integration at GAPDH (GAPDH::rtTA iPSCs) was Nucleofected with a selected high-efficiency RNP and the corresponding PRDX1-AKR1A1 targeting construct (for either Site 2 or Site 3). Three different gRNAs were tested for PRDX1-AKR1A1 Site 2 (SEQ ID NO:87-89) and three different gRNAs were tested for PRDX1-AKR1A1 Site 3 (SEQ ID NO: 90-92). Pools of cells that received both PRDX1-AKR1A1 Site 2 or Site 3 RNP and targeting construct were fed with media containing 2 µg/ml doxycycline starting at day two (Site 2; FIG. 13) or day one (Site 3, FIG. 14) post-Nucleofection and continuing up to day 7 post-Nucleofection (FIG. 15 and FIG. 16) in order to induce GFP expression. GFP expression was monitored over the course of 7 days by fluorescent microscopy or flow cytometry. GFP expression was induced from both PRDX1-AKR1A1 Site 2 and PRDX1-AKR1A1 Site 3. All three gRNAs tested for each site displayed differences in construct targeting efficiencies (different sized peaks seen in flow cytometric histograms), but all were able to induce GFP expression to similarly high intensities (similar log levels of expression) following doxycycline addition. The peak observed around 10⁶ represents edited cells that express high levels of GFP, while the peak observed around 10⁴ represents transient GFP expressed from non-integrated targeting construct. The data demonstrate that multiple sites within the PRDX1-AKR1A1 intergenic region permit robust inducibility and expression of GFP in a pooled population of STAPLR site-targeted iPSCs.

SEQUENCE LISTING

-continued

tgtcaatagg	taagtaagta	aaaaaacacga	tatgaacaat	aatgtctata	caaaaataaa	4140
agcagatgcc	gggcacgggg	gctcatgccc	gtaatcccc	cactttggga	ggcccgaggcg	4200
ggcggtcac	gaggtcagga	gatcgagacc	atccctggct	acacgggtgaa	accccgctc	4260
tactaaaaaa	tacaaaaaaa	taagccagac	atggtggcg	gcccctgttag	tcccaagctac	4320
tcaggaggct	gaggcaggag	aacggcgtga	accaggagg	cgagggttgc	agtgagccga	4380
gatcgcgoca	ctgcactcca	gcctggggga	cajagcggaga	ctccatatca	aaaaaaaaaa	4440
aaaaaaaaaaag	tatatatata	tatagcagat	aaaggggata	gaatgcaggg	tggaagttaga	4500
gggtgcagt	ttgaatatcc	tggtgagggg	agacttcccc	aagaagggga	catcgaagca	4560
aagacctgaa	ggcagtgggg	agggtggcca	taaggatag	tgggaaagag	cttccagca	4620
gagaatagt	ccggggcgg	aggaggctc	gaagctgagg	gaaagcggca	cagaatagag	4680
gttagggagt	tcccgagaca	ccagggtgt	tgggtcettg	ccatcatct	ctaggaccag	4740
tgcgccctc	agacaatgggg	cctccctct	gggtcccat	cttttagggg	cgctgctct	4800
gcttcacct	gccaatatcc	cacccaaaaa	gcaaagagt	caggggacca	aggggcatgc	4860
taaccctaaa	tgtatgtct	ctttctggac	ataatcccc	aatccctggc	cgaggcaccc	4920
aaacccacgc	acatgtctat	ttcatgttag	tcacatgtt	cectctcaa	ggagtattta	4980
tatTTTGT	ttttaattt	ctgtatgt	gttggaaagac	tgacttttc	tgagggaatt	5040
ttaagcagag	gaagctggta	actcaatctg	tttggagtc	aggatcactc	tactcaatgt	5100
aacacaaacgc	taagaggtaa	tgcaaaaggac	aggatcgcc	gtgaagaaaa	cacgggattt	5160
ccatgtcttc	agaaatcgac	atcagaagct	cccatgaggt	ttgttctctg	aaatttctt	5220
ttctgttca	aaatcttag	cttaaacaata	ttgtcagtgg	ttctcacact	tttagattcac	5280
tttaggattct	tctggaaac	ttgtttaaaat	tcagatacc	aagctccatc	aaagatgtcc	5340
ccaaatcgaaa	ctctcgaaag	tcaggccaa	aactttcatg	gtacacagatc	tcctagctga	5400
ttttcaccc	gatggatctt	gaactggcc	ttcacaacga	tgttgttaca	actgtgtacg	5460
gttgggcatt	agacaaacat	gggtttaaat	cataggctca	ccat�aaacca	gttaggtgac	5520
tgagcaattt	tcttacctt	tgtccatctt	agtcaccc	cctgtaaaaa	tgaagataaa	5580
taatcataat	ttcagaggat	ctgagggtt	ctgaaatacc	ttgtgttgc	cacatcgccag	5640
agoacatgtt	tcatgttata	gtatgtcgt	aagccttctt	cctgttgc	gtacaagagc	5700
tgcccttcta	catacttaca	gagaacacaa	aaaaatgc	caagatggcc	aggatgaagt	5760
cttctcattt	agagtattt	aagagcgtg	ataatattgg	aaataaaacca	aaaagccaa	5820
tgatatcttc	ctgtatcat	ttttacccaa	aatgattttt	ttttttctt	ttttaaggaa	5880
ttacacacac	acacccccc	acacacacac	ttggctgatg	tatcagctg	tctttgatta	5940
gtctgggca	aatgtatgc	aggcattctg	aaaacccacaa	ctggagacat	catttctgat	6000
gggcctaata	ctctgggtat	ttcccttcaa	ctgaacttgg	ggaaagttact	tttgatataat	6060
gtacatattgg	aaatattttt	caacatttt	taaaagttag	atacttttt	caggatataat	6120
gtaaaacaca	ctttttctt	caacgtgtat	ggcttaattt	cttgcattt	ctaaggccag	6180
gtatTTTGT	ttgttgggt	taatgtatc	tcatgtgt	agaaaggacc	atattgcacag	6240
ccctggcgac	agtcttttt	ctggcttctt	ggcagacagc	aggaacagct	gctgtttaag	6300
catacccaa	atatTTTacc	tctggaaaga	tcattttgg	aggtgagcag	gcactttgat	6360
ttctcatcc	ctctgaaagg	ttcggtggg	aatggtgggg	gaatagctgg	ccatTTTact	6420
ggggcttgg	ttctggggcc	attgcacat	ttttttttt	ttttttttaa	aacgggatct	6480
cgtcttatca	cacaggctgg	agtgcgtgg	cgcgatctcg	gctcaetgca	agctccgctt	6540
cccggggtca	cgccattctc	ctgcccagc	ctcccgagta	gttggacta	caggccccc	6600
ccactggcc	cagcttattt	ttgttagagac	gagggttccac	cgtgttagcc	6660	
aggatgtgtt	cgtcttctg	acctctgtat	ccacccgcct	cagecctcca	aagtgttggg	6720
attacaggct	tgagccacc	ccccagcca	ccatttgcac	atttccatct	tgcaatccaa	6780
tttgcacact	ggtttctgg	caaattttt	tgaagtgtt	aaagctgtt	tttttaatg	6840
ccatccctaa	tctttaggtt	aaacatttt	ttggaaatgtt	tttccatct	caagaatcta	6900
acagttgtt	tttcagggtt	gaaggatag	gggggggggg	aagatgtataa	ggggaaagggt	6960
acatTTTgtt	tcatgttta	aaatttttt	tactcttgc	attttactt	ccaaacatgt	7020
catcttattt	tttaaaaaaa	ggaatctgt	gtataatgtt	gtttagacc	atTTTgttta	7080
taatTTTctt	taaattttt	gaatgtttt	atgaaatctt	gtaaatgtt	ttttgaaagc	7140
tttagcttata	ttttaggtt	atttttttt	tttaaaggatc	atttccatct	tatactttt	7200
tcttttttag	tttacaagcc	ttgtatgtaa	ttattttgtaa	gacaacttga	tataaaaaat	7260
accctgggtt	tgatattttt	ttcatctgaa	aatgtatgtt	tcacaagatt	tttttttca	7320
tcattaaaga	agtaatcat	tttgcattttt	aaagatatgt	cgatgtgtt	gggatgcac	7380
ttttccctt	tatgtatgt	tttgcattt	ttacttgc	acagcattga	ggcaagttt	7440
ttaaaaataa	gtttgcctt	ttcatatgt	ttcgactata	ttttgtttaat	ctcaatgtat	7500
caagaatgc	ctttgtatg	gataaaat	ttcgaggacta	ttttgtttaat	ctcaatgtat	7560
tcctggcc	ctctggccac	agaaaaaga	tatctgtgg	cgctgttgc	aaaaatgtt	7620
tttgttattt	tttgcatttgc	aaacgtttt	caagtggat	aaaaacgtt	tagatctaca	7680
ttttttttt	tttgcatttgc	tttgcatttgc	tttgcatttgc	tttgcatttgc	tttgcatttgc	7740
acctgggtcc	ctaatgttgc	aaaacaaattt	attttccatct	gaggacagct	cccaacttctt	7800
tacctctt	cttaatgc	cactctactt	cactctcagc	agccttataa	tatctgtaaa	7860
tgttagaaact	atcgttattt	actccctttt	catcttgc	accccttcaaa	ac	7920
ttaatgtat	gtaaaacat	gggttttttt	attttttttt	tttttttttt	tttttttttt	7980
gctgtgtca	cttgcatttgc	tttgcatttgc	tttgcatttgc	tttgcatttgc	tttgcatttgc	8040
atgaaactaa	taatgttgc	tccaaagaat	tttttttttt	tttttttttt	tttttttttt	8100
ataattgttgc	tgcataatgtt	tttgcatttgc	tttgcatttgc	tttgcatttgc	tttgcatttgc	8160
tgtcttctt	tctttaggtt	atggcatat	tttttttttt	tttttttttt	tttttttttt	8220
aactctatcc	catactcttct	tttgcatttgc	tttgcatttgc	tttgcatttgc	tttgcatttgc	8280
tttcccttct	acatcttgc	tttgcatttgc	tttgcatttgc	tttgcatttgc	tttgcatttgc	8340
acatggccaa	gtctcatcat	ccataaaaaac	cttgcatttgc	accacttgc	cttgcatttgc	8400
ctgttcatatc	ccagctgtgt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8460
cctatcaatc	tttgcatttgc	tttgcatttgc	tttgcatttgc	tttgcatttgc	tttgcatttgc	8520
atgaattat	tcatcaatata	tttacaagca	tttgcatttgc	tttgcatttgc	tttgcatttgc	8580
gctcgttac	tcagtcatttgc	ataatactgt	tttttttttt	tttttttttt	tttttttttt	8640

- continued

-continued

-continued

acagctctac	tgcataattat	attgtaaaact	gtgtttctc	caacattttt	ccttcttcat	17820
ctataaatag	aaaaccttg	aatgattgat	tatatgttta	ttaaggcga	tgctcta	17880
cccaactgct	tattcattat	ttcttgggtt	ttcagttaca	gagtttatg	tatgtttaaa	17940
tagaacattt	ttctgttagg	attgtgggtt	aggacatgga	tgataacaga	tatgcatatc	18000
ttactcatga	aaggctctt	ttatacgtgg	cttataagaat	attacaaact	catgtgaaa	18060
aatatttcaa	agcttaagag	gttaaaaaat	aataatgaat	tggttcccg	agcaaaccta	18120
cttttggaa	agtgtatgt	aaatatttac	ataatagatg	atgcagtgt	ttctgactt	18180
gcagttttat	tattatgaa	attaatttt	cattttaaa	ttatataat	ataatttata	18240
tgttttccac	aatgttaggt	gaaattatgt	gtcccttagt	tacacaggaa	gaaggataag	18300
atataaaacg	tatacattt	taaacgtctt	aaatcttatt	acttctacag	caattgcctt	18360
ttaggcttca	tcaggctaca	gactatact	cacttttga	gttttatttc	cttcccagg	18420
gccactttg	tcatatgtaa	agcgttata	ctagttta	cctaacagaa	ttgtcatgaa	18480
gagtatataa	gcacaaatgt	aaatcactt	acatgtacc	tgccaagcag	tggtaatta	18540
atgttagatt	ccttttaag	ctataaaggaa	aattttaaa	cctttaatcc	ctatgtttt	18600
ccaaagcaaa	atgtttttt	tttattttt	tcttcatacc	aaagcaaaat	attttaattt	18660
ttccaaagt	ttcccttactt	ataaaactgt	gggaatttca	gaaacattt	gttttacagt	18720
gcccagagaa	acttctaca	tataaacgt	ggttttgggt	taatgcatt	cattatttgg	18780
gttattatcg	gcctttatcat	aaccttaatgt	tegtgtggct	cggttaagac	ataggcaata	18840
tgactctggg	agagactctc	actaatgaac	agtaaggcga	gcaagectca	actgtactt	18900
gaatttgata	ctacaatgtt	ttgtatgtat	tttcattacc	ttcttcaga	aatttcttaa	18960
caatttgagg	gttagtataat	tagaagtgc	gaaccaatac	atggcatct	gtaaaaataa	19020
ccacgggott	agtaacaagc	aaataaaaaaa	atataaaaac	attggctcta	taatgtgtca	19080
atatgtctcg	gacttattgt	atcgtgact	cttaactt	gtcccaaaaag	gaaggtaaca	19140
tttgttagagt	ctttaaaatgt	tatataaaa	agaaaagca	aacagaaaac	aaaaaaatctt	19200
tttagacctt	tagtccatcc	ctaatactact	attctccatc	atataattt	atctatttgg	19260
attgttaatgt	ttgcccagac	cagagatgg	ggtggttcc	actagggtt	cagtgtgg	19320
atggtaggg	gcagatgtat	ttggatgtat	ttggatgtac	agccaaaggag	acctactaat	19380
attttggat	cagaaggat	gtaaaggagag	aaatccacag	gactttaaga	tttttggcca	19440
gagggat	accttttagc	gacatacaga	tgaccaagg	aagagcaaaag	ttgagaggat	19500
gagtttata	aaatagtatt	atggtcacgc	caagctatca	gcattaaggc	atccctgtag	19560
agatgtctg	ttgacattttgc	ttggatgttgc	ttggatgtttt	aaaaaaatgt	gtAACatatt	19620
acttttagca	cttattttat	tacccacgt	acatccggaa	gtaaaaagct	19680	
agcagctcc	ggctttccct	aacccttga	aaacaccttca	atataaaagg	catttatc	19740
caaaagtcc	cattcatctc	ccgctctgg	aaccgggagc	gccccctgtc	caatcagctt	19800
tcaaatcccc	tgccgtcagcg	agtgccgggg	gtccacccat	attcccttgc	ctccgctt	19860
ccggcttggc	cacgttagaa	gtcccgcc	tccttgcagg	tacacaaatc	accttatat	19920
ccatttagaaa	aatgttagat	acattatc	aatgttaatgt	taatttagatt	aaataattt	19980
caacactta	ttcataatcat	ttattttttt	tattttttt	ttttttttt	tctatgagta	20040
tagtaaccac	ctccgagctt	tacggatctg	cggttccctt	cctcgccgac	gcccactt	20100

SEQ ID NO: 2	moltype = DNA	length = 62214				
FEATURE	Location/Qualifiers					
source	1..62214					
	mol_type = genomic DNA					
	organism = Homo sapiens					
SEQUENCE: 2						
gtgtggcgct	cgccgcgt	ggttttatag	ggggccgcgg	cgggccgtcg	agccataaaa	60
ggcaacttcc	ggAACGGGCG	acgtgttgg	gccccggcgc	gtcactcac	cggttcgccc	120
gcacagtgc	gcattttttt	acccctctc	ccctcccttt	gcgaaaaaaaa	aaaagaggca	180
gagcggagatt	gaggaaagagg	aggagggaga	gttttggcg	tggccgcctt	gggggtgtgg	240
gccccggggc	tggggcgccg	cgccgtggcc	cccgccgc	acgtgtggca	gtgccccgtt	300
cgccccggca	tggccggcc	cgccgtggcc	cccgccgc	ctcgccgtct	ccacccaccc	360
cgggacatcc	taggtgtgg	catcttgg	gcactgaggc	cccaagggtgg	gtggggccagg	420
gtctgcacgg	gtgcggggc	cctgggttct	gtacgcttct	gcagaaggag	cttttggagg	480
gcatggatgt	gcacggcgt	cactccccc	tgccgactt	agacaaactg	ccctgttgc	540
agggccttgc	gaccccttgc	tgtggggctc	agctagctaa	atgtgttgg	ttgggtcacta	600
ggggagagacc	tgggttttgc	ttggatgttgc	ttggatgttgc	ggatgttgcgt	ggcttgcggc	660
cacttagaa	tcgcaggacc	acactccca	ggacaggcga	ggggccaggc	gtccatgttgc	720
tggagggtgc	ccgtgtatgg	ggcttacaaa	ctacccacgc	gcagccctgg	gaaggaaagg	780
gggccttaca	gggcaggggca	ccttttaccc	tggatgttgc	tgttttgcgt	gttacacatc	840
acgccccaggc	aaagacaaggc	ctggggcggtt	atgtgggtgc	ctaggccatc	cgggggccggg	900
gggctgggtc	tacacaggct	gggttgcggc	ccacccgtcc	ttgtatgtt	gttatgttgc	960
gccacagctg	aactgcctc	ccagaccatc	tggaggccgc	tgggggactc	tggggaccaa	1020
gactccatgt	gcacacaggat	tttgggggggg	ggggccgtgt	aggaaacttca	agccagctg	1080
ggaagacccat	gtcccttgc	cccttttcc	ccttgcgttgc	gttcactgtt	tttttttttt	1140
gaccgggtgg	cgagggtccg	ttctgcgtcc	ggaatcacag	actgtgttgc	cccagggttt	1200
gggcacatgc	cgatcgttgc	cgtggggatca	cagagggcga	ttccgttgc	ccttagaggaa	1260
cctgccttcc	ttcttcttccc	ccttcttgg	ggtgggtaca	agaagtgaa	tttttttttt	1320
aaaataggaa	gagtggttgc	gtggggccgc	ctcaggccctg	cgggggctt	tgcaagcgcgt	1380
agttaggtgtt	cactgtgttgc	gtgtcgat	aggcaaaatgt	gagttccat	aaggacatgg	1440
aacctgttgc	caggaaaggat	gtcttgcgtt	ccagttcac	gtacaaggct	gtgttcgc	1500
aatgggactt	gtcccttccct	tgaagggttgc	agaggccaca	gtctgggtgg	aaagatgacc	1560
accacccaggc	acacagtggc	agacacaggat	tcacagtcca	gaagacttgc	tgagccctt	1620
ccatcaccac	cccacgcga	ggcaaaatctc	cccagccctgt	atactccctg	cctcttgggg	1680
atgggactgt	ctgtccctgt	gccttgcata	tttcagctgc	aggcgttgc	ccagctgcct	1740
caggacttgc	caacgttca	ccggcttggcg	ttgtggggcg	atgggggggg	ggcccttctt	1800

-continued

gctgcctgtt	gatgagggct	ggttctgggg	gttcgtggaa	cgttgaactg	ggtgagggtc	1860
tgggggggtg	gggtggggcag	tggggatggg	accggatctg	cagagcagggt	atgagaggtt	1920
gtcagggtgg	cggtccagg	cagaagggtg	gggtgcagag	cagacctgtat	ccccaggccag	1980
gacctgttgtt	gocactgggc	ctccatctag	ctggccaaga	gacaggagct	ccagccccac	2040
cccgccccgg	gctgcagcat	tcctggctgt	tcctctggc	ttgtggttcc	ctctgtctgc	2100
agcggccact	tcctagtgtc	ctccatgca	cccttcaatg	cccateccag	aaagccctc	2160
tttgggaaag	cggttttagcc	tetgcagcca	cgcttcte	cttacccatg	accaaggctt	2220
gcagctggca	tttagaactg	ggccctcage	accaggctgc	gagccctccc	agggtctccg	2280
tcaccccccac	tcctccac	tacacccacc	ccagggtctc	tcgtgttga	atgagtgact	2340
gcctgggtgt	cctcgtgt	cagggtggag	gtggccgagc	agctgggggt	gaggggcagg	2400
gaggaggtg	agagagaaag	caagggaaag	agcaggaaag	gaggaaatgt	agggtgcctgg	2460
ccaggggccg	ggccgagcag	ggggaaatgt	aacaggttct	tttctgtgt	ttggaaacttt	2520
gcctgttaag	ggagctttcg	gagetgtac	tggggcagc	acgcagcccc	cgggagaaccc	2580
tcgggtgaaac	tcagccaaag	ggactccca	tgctgtggc	ggacttttgt	ttttttccct	2640
ccaaggcctt	gtttgggtga	caccccccac	ctgcaaggcc	tccaggcaac	caactcttac	2700
agcgcagacc	cagggtggacc	tcctggccc	tccggcaaaa	caaccaccca	gccacccctgc	2760
cctgggact	gcagcgcct	ccctatctg	ccaggaggga	atccaaaccc	ctttttattta	2820
ttattactat	ttttgataca	agggtttgt	gtgtatgc	ggctggagt	cgatcacagc	2880
tcactgcage	ctcgaccc	tgggtcaaa	gccttcac	gtatgtgg	ctataggcac	2940
acacttccaa	gcctggctaa	ttttttattt	atctgtatag	gtatggcttc	tcaccatgtt	3000
gcccaggctg	gtctccagct	ccgggggggt	caagcgatcc	tcaccaccta	gcctccaaaa	3060
gtctgggat	tacaggcatg	agccaccctg	cccggtccca	agctccattt	tgttagaggaa	3120
ggaactgaga	cgagcttcc	cttcaagat	tgatgttagt	ctaggaggac	agggaccctg	3180
aacacgcttag	tggtctgt	agaggaaat	gtatgtatg	ggggggccaa	cattccagga	3240
gacacaccccc	cattcaca	ctagggaaact	gagtcaaggc	tggtaaagt	ctcaagttcc	3300
caaatacggt	gggaaggact	ctagctctag	ggtccctgag	ggggcaagccc	tgcctccctgc	3360
ctcacactgc	ccctcactgg	taaagggtc	cctggatgt	ccccaaaggcc	tgatccaaaa	3420
gggtgttta	aagtccgaat	ttcttggggcc	tacccctctt	ctcagaattt	atcaccctgag	3480
gccccagag	gacaccatca	gcctatggc	acgtggcagc	aagtgggtc	tctcccaactg	3540
gggtgtatgt	ggttgtac	gtatccca	cattttggg	ggccgaggag	ggaagatcgc	3600
tcagggccat	gtgtgtggaa	caaggtggg	caacacggg	gatcttatct	ctacaaaatc	3660
agttgggtgt	tgtggcacgt	gcctgtactc	ccagcttaat	gggagggtgg	ggcaggagca	3720
ttgcttcagc	gtggaaagtc	aagggttata	ttagctatg	tcaagtact	gcattccagc	3780
ttggatgaca	gagggaaagcc	ctgtctcaaa	aaaaaaaaaa	aaagaaaaaa	aaaatttaat	3840
ttaaaaattt	gggtgtccag	gtggggctgt	gtggctcac	cctgttaatcc	ctgcaactt	3900
gggggtcaag	gggggtggat	caccgtgg	aaaaatttca	agatcagct	ggcaaaatcg	3960
atgaaacccct	gtctctacta	aaaataaaaa	atattagctg	ggcatatgtt	cgctgtgcctg	4020
taatccca	tactcaggag	gctgaggcag	gagaatcg	tgaaccggg	aggcagaggt	4080
tgcagtggcc	cgagattgtc	ccgtgtcact	ccagctgtgg	tgacgaccaa	aactctgtct	4140
cagaaaaaca	acaacaaaaaa	agttgggtc	tccagctttt	attctttaa	aaataaaata	4200
gccccggcaca	gogggtcaca	ccgtatccc	agcattttg	gaggccgagg	caggaggatc	4260
tctcgagttc	aggagttttt	aggagcatgg	ttagaccctg	tctcttacaa	aaaaaaaaaa	4320
aaagacaaaa	ttagccggcc	atgggtgg	tggtctgt	tcccagctac	tggggaggct	4380
aagggtgggg	aactgtgt	gcctgagggg	tcgaggctgc	agtgaaacca	gatcatgc	4440
ccgcacttca	ccgggtgt	tagagacc	gtctccaaa	aaaaaaaaaa	aaagtaaaag	4500
aaaaagaaaa	agtaggtgt	cccggtttat	ttatattaa	tattttttga	gataaggct	4560
ctaaaaattt	tagagacagg	gtctgtct	gtggccagg	ctggagtgc	gtggcatgt	4620
cacagctcac	tcgcaggctt	acccctggg	tccaaaggat	ctctctgtct	cagcctctg	4680
agtagctgg	acggcaggga	tgcaccac	tctgactaat	atttttat	tttggggaga	4740
tggggtctct	ctgcaata	acagggtctg	tccctctgt	ttgcctggc	tgggtctcgaa	4800
ctgtgtacgt	gaagtgtat	tctcacca	gcctccaaa	ttgtgtggat	tagagggtgt	4860
agccactgca	ctggccatca	gtttttgtt	gttgggtt	ttcggtttt	gttgggtt	4920
ttttagatgt	gtgttgc	ctgtggcca	gttggactgt	cagtggcaca	atctcagct	4980
actgcaacct	ctgcctcccg	ggttcagggt	tcaagcgatt	ctccgtcc	agcctccgt	5040
gtagttggaa	ctactgtgt	gcaccaccc	gcccagctag	ttttttttt	ttttgttatt	5100
tttagtagata	tgggttca	ccatgtgt	caggctgt	tcaaaactct	aacttcagct	5160
gatctgcaca	ccttagctc	ccaaagtgt	gatagagac	gggtttact	gtgttagcca	5220
atagtttatt	tatattttta	tttattttcc	ttttttttt	tttttttggc	tggagtctt	5280
ctctgttacc	cagggtggag	tgcagtggca	cgatctccgc	tcactgc	ctccacccccc	5340
ctgggttgc	gccttcttc	tcgttcac	tccctcagtg	ttggggact	agggcccccgc	5400
caccaacggc	ggcttaattt	ttttagttt	agtagagac	gggtttact	gtgttagcca	5460
ggatgtgtct	gatctctgt	ccttgtgtat	cgaccgcctc	ggccctccaa	agtggtggaa	5520
ttacaggcgt	gagccaccgt	ggctggccata	tttttctt	tttttagaca	gagttgtgt	5580
ctgtgtccca	ggctgtgggt	caatggcc	atcttggct	actgcaact	ctgcctcccg	5640
ggttcaagg	atctcttc	ctcagccct	ttagtagct	ggatcacgg	caccacccac	5700
ccccccagg	taattttgt	tttttagtg	agatgggggt	tcaccatgtt	ggccaggctg	5760
gtctcaact	cccgacacta	ggtgatccgt	tcgccttggc	tcctccaaat	ggtgggatta	5820
cagggtgtaa	ccaccacacc	cggtttat	ttatatttt	atttatttt	ttatatttt	5880
atttatttt	ttttagatgg	ctgtcgcc	ggctggatgt	caatggtgcc	5940	
atctcagctc	actgcaacct	ccacccctcg	agttcaagcg	atccctctgc	ctcagccctac	6000
ttagtagctg	ggactacagg	cgctgtccac	cacggccac	taatttttt	tgtatattt	6060
tagagacggg	gtttccat	gttggccagg	atggctcc	tccatctc	ctgacccctgt	6120
gatccgc	cctcagcc	ccaaagtgt	gggattacag	gatgagcc	ccgtgcccag	6180
ccttacttata	ttttttaaat	aaaacagat	ggggtctctc	tcctccctt	ccaccccttc	6240
agcctaagg	gccagttca	ggctgcctt	tccaggaagc	ctgcatgac	ttcctccctc	6300
ttctactcta	acagcagaag	gtttctgt	cagaaagggg	cagtggccca	ggcaggatca	6360

-continued

cccgtaagg	gcacagccag	gagctgaagg	gcctttgcc	tggagcacgt	cctggatcac	6420
actgtccctg	catgcagcac	accgtcactg	accagccgga	catgttcca	gagttactg	6480
gcactgtttt	gagaatagtag	aaaagcaccc	tgaagaaagc	aagcacaatt	cctaagacag	6540
atcttaaatg	agagagagag	caagagaaca	acaagttaa	tggcatcagg	agggaaagca	6600
acaggttcc	ttcagaatga	ggtattaggg	ggtgctgggt	gtagtggctc	atgcctgtaa	6660
teccagcaact	ttgggaggt	gaggoaggca	gatcacgagg	tcaagagtt	aagaccagcc	6720
tggccaacat	aatgaaaccc	cacctctact	aaaagtaca	aaattaaccc	agtatggtgg	6780
tgcacaccta	tagtcccage	tactcggag	gctggggcag	gagaatcgct	tgaacctggg	6840
aggcggaggt	ttcagtgagc	caagatgtg	ccaccgcact	ccagcctgg	ctatagagcg	6900
agactctgtc	tcaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aggagggtgt	ttggccgagt	6960
gtggtggtc	acgcccataa	tcccaaact	ttgggaagac	aaggtaggag	gatcaactg	7020
gtccaggagt	tcaagaccag	cctggcaac	atagcaagac	cttgtctca	tattttaaaa	7080
aaaaaaaaaa	aaaaaggaag	gtgtttttt	ttatttttcg	agttagggc	tcactttgtc	7140
accaggcttgc	atgtcagttg	tggaaatctca	gctcaactca	ggctcaaccc	tttggccctcc	7200
attgatectc	ctacctcagc	ctcctact	actgttacta	cagggtgt	ccacccatgcc	7260
caqctaatagt	ttttagaga	caaaatttt	ccattingcc	aggctggcc	ccaaactctg	7320
ggtaaaagca	atccacactgc	ctcagctcc	caaagtgcgt	ggattacagg	catgagccac	7380
tgacgcgcgc	agagagtgt	gtcttggaa	cagtcagggg	gccaggcgt	gtggctcaag	7440
cctgtatatt	cagcacttgc	ggaggctgt	gggggtggat	cacgggtca	ggagatccag	7500
accatctgg	ctaacacgg	gaaacccctg	ctctactaa	aatataaaa	attagccagg	7560
catggtagcg	ggcacctgt	gtcccagct	cttggggaggc	tgaggcagga	gaatggcatg	7620
aacttgggg	gtggaggtt	cggtgagac	agatcgtgc	actgcactcc	agcctggcg	7680
acagacgcaag	actccatctca	aaaaaaaaaa	agaaaagca	tcaggggagg	ctccctcagag	7740
gagggtacca	gcaagacgctt	ggtccccaca	tacacctgc	ctccagacat	agcctgggt	7800
ggtcacttgc	ctgccttcag	tagetctgc	ctgggcctct	gagctccct	ggcagaggca	7860
ccaggctaa	gggaatccct	gccccagagg	aatggggaa	ttctaaact	tcttgggaga	7920
gtttttaaaa	atctttaaaa	actcccttct	gttccagcgc	cgcagggtac	ccggagaaac	7980
ccttccagcc	ttccctaggga	gttcttggac	ttgttattt	gattgaatta	ttcagaaaa	8040
gaaaatgcgg	tttcttgggc	ttccccctta	cacacacaca	cacacacaca	cacacacac	8100
cacacacaca	ccccacacac	acacacccca	cacacaccac	acacacacac	cacacacac	8160
ccacacacac	acateccacac	cacacacacc	acacacacc	acacacacac	caccccacac	8220
ccacacacaca	cacacacacac	acacacacac	actgaacgct	ttttgtttc	tctggaccca	8280
tctgctgtag	ggccagctct	cccccttctc	caagatgaaaa	agagaccaca	tgtcagaga	8340
ggaactggat	cotttggggc	ctgcaggct	aggctgtgg	cgcttgggg	gggaggctgc	8400
gggcacccctc	atccctcccg	ggactccaa	ccactctcc	agggccaa	gtagcccttg	8460
ggactcagg	cccttctgtt	tgccccggat	ggggccctcag	aatgtgggt	acaagttggg	8520
ccaggagaga	ttgggggtcag	tagcaagtt	aatgtcaagg	caagtgggt	tcccgaggcc	8580
ctggctctga	cacccttctt	gggtgtctca	caggccaa	gtgggtctt	cccatgtctaa	8640
ggactgtccc	aggcagtggc	ttccctggcc	ttggcccttct	gcccctact	gtgaaaagg	8700
gcagggggtt	acctagctcc	cagggtcccc	ctgaggggg	ccagtgggg	gggggtctgca	8760
gttcaggccc	agggtcatgt	cgaggactc	gccccttct	ggctggctgg	gttlatccct	8820
gacgctctcc	tgcctggcc	gggacctcag	aaatgcccct	ttgagctgg	catagtggct	8880
catgcctata	atcccgac	tttgggggg	tgaggcgccc	agatcacctg	aggtcaggag	8940
ttcaagacca	gtctggggaa	catgtgaaa	ccccatctct	accaaaaata	caaaaaattt	9000
gtctgggtgt	gttggggggca	cctgtatcc	caogtactcg	ggaggcggag	gcaggagaat	9060
cagttgaacc	caggagggtgg	agggtgcago	caogtgcata	cgagatccag	ccattgcact	9120
ccagcctgg	cgagaagacg	aaaactca	cattaaagaa	aaagaaagaa	gagacaaga	9180
gagaagagaa	ggctggggac	gggtgtccac	actgttaatc	ccagacttt	gggaggctga	9240
ggcggggcagg	atcgctgtag	gtcaggagtt	cgagaccac	ctggcaaca	gggtgatacc	9300
ccatctctac	taaaaaataca	aaaatttagt	gcgtgtgg	ggggcacct	gtaatccag	9360
ctactcggga	actgtgggg	ggaaatccc	ttgaacccgg	gagggtgggg	ttgcagtgg	9420
ccgagatcgc	gccattgtc	tccagctgg	cgaccaaaa	gcaaaaaattcc	atctcaaaaa	9480
aaaaaaagaga	gagagagaag	aaagggagg	aggcaggac	gcaagcaggc	cccttgggg	9540
gccaggcacc	tctcagcc	agtctccctt	gttctgggg	gggggtgaaga	ttcccacccg	9600
tgtttctgtat	atggggggaca	cgtgttggg	ctggggatgg	gtcacatcc	ttcagccct	9660
cgggggtgcgt	gaagaatcaa	gaacgggg	cgaaacccac	tgggtcacgc	ctgtatccc	9720
agcaagcagg	tagttggca	ataaaaaaa	ttagctggc	gttggggccac	atgcctgtaa	9780
tcccagctac	tcggggggct	aagggtgggg	aattgtctgg	accaggagg	cagagggtgc	9840
agtgagccga	catagccca	ttgcactca	gcctggcaaa	caagagcaaa	actccatctc	9900
aaaaaaaaaa	atggggggaa	agaaaaagaaa	agaaaaaaa	ggatgattgc	ctgaggccag	9960
ctgcagtgt	ccatgtatgt	accccttc	atccgttgc	gagacaa	ggagccctt	10020
ctctaaaaaa	taagttttaa	aataaataaa	atagtctgg	gtactggctc	ctacactgtaa	10080
tcccagcact	ttggggggcc	atgggggggt	gtactctgt	ggtcacag	tcaagaccag	10140
cctggacaac	atgttggaa	tccatctct	ctaaaaat	aaaaaaat	cccaggat	10200
gtggcagggt	cctgtatcc	cagtcactcg	ggatggctgt	gcaggagaa	cgcttgcacc	10260
caggagggtgg	agggttcagt	gaggtgcata	cacaccact	cactccagcc	taggcaacag	10320
agoaagactc	caccacacac	acacacacac	actttttat	ttaaaaattt	aaaaaaatgt	10380
cccttggagg	ttggggcagt	gccccatgg	tgaggcccc	ttgggggggt	ggagaaggag	10440
agaaaaaggaa	gtgttgggg	gggtgcaggcg	gacggggcag	ctgctgact	aggaggccac	10500
tggctcacct	gccttggggc	tcagctgtc	tccttggccca	cccccagttc	agagaggaga	10560
agggaaagaca	gaggccccaa	gggcctactg	tgtgtggcc	gcgtgtcagg	gttggccca	10620
cgtcaccc	tcaaaaagcac	ccccaccc	tgtggccca	gtgggactct	cccatcttc	10680
agatgagtaa	accggggc	ggggaaatgt	gagacttgc	cctatgtt	acacgggtt	10740
gttattacca	ggatgttctg	gaacatgtt	atccagggtcc	tggcccagg	ccaccatcc	10800
tggcgagctt	ccaaggggccc	accaagtcc	tcaagcttca	tgtgttgg	gttctcagcc	10860
ccatgtttca	gtatgggtaaa	ctgagggtt	gaggtggcc	tggggaggg	cagaggccagg	10920

-continued

ctcggtggct	gcagttggcc	tcgtgtccct	gcccgtggggc	gcattcaggg	ttcaacttgggt	10980
gtccctggcgc	ttggcgtggcc	tcctgcgtgc	ccggggccttg	cctgccccaa	ggaaatcccg	11040
atcacagccca	ccgcgcagca	tcaccaggcc	ctccctcagcc	agaatcttc	gcactgtctg	11100
gggttgcggc	gaacctgagc	cgccccgtgg	gggacaggcc	cagctgtctg	ggcttccctga	11160
ggtgtctccc	atacgctgcc	ccgtgtccagg	tgtgtggggac	aggaggggac	ctccctggca	11220
tgaactgtcc	actggggaca	ggggocacaga	gtggccagcc	cctgatccca	tgagacttgg	11280
cctgtgc当地	cccaggggaaa	geccccagacc	cagecctcag	gcagggggag	tcccagaggc	11340
ctctctggct	tccatgtttgc	cggctcagcc	aagaagggtt	ttggggtttc	tcttgggtat	11400
tcggcttcgt	ctggggacaa	ggcacaggca	gggctccctg	acctccctga	ggaagccccg	11460
ccgctcgggt	tcccgatgtt	ctagggggtt	attttttttt	ttagacggag	tctcgctctg	11520
teaccaggc	ttggagtgcag	ttggccgtat	taggtctact	gtaagtcgg	cctccgggt	11580
tcacgcattt	ttccctgcctc	agcctccga	gtagccatata	ggtgcggggc	accacgcgg	11640
gctaactttt	ttttgtat	tttagtagaga	ttgggtttca	ccatgttaac	caggatggc	11700
tcgatcttc	gacccgtgt	tctgcggcc	tcggccctcc	aaagtgttgg	gattacaggc	11760
tttagggccat	ggccccggcc	tttaggggtt	aatttttgc	acttaagtca	ccttgcgttg	11820
cacgggtggc	catgcgtgt	atcttagca	tttggaaaggc	ctaggccggc	agttcacata	11880
aggccagaag	ttcaagacca	gcctggccaa	catggcgaaa	tcctcatctt	actaaaaata	11940
taaaaaatag	cagggtgtgg	ttggcacatgt	ctataatccc	aactactcag	gaggctgagg	12000
catgagaatt	gtttgaatgg	ggggaggtgg	gggtgcgtgt	agctgagatc	acaccactgc	12060
actccagct	gggcacacgt	agagacacccg	tctccaaata	taataataat	aatgtcacgt	12120
gggctctccg	atagccaagg	gcccataagg	ccagagatga	ccaaggccca	atgttgggac	12180
ctccataacct	ttgtccctgg	ggccacactgg	gtgtcacaca	catgtacat	ttgggggtgac	12240
agtggagaaa	ggcacacaga	tgtccctgtt	gtgtgtaccc	gggacatgt	aggctggagg	12300
gtctggccat	gtcagggtgg	gacttagca	tgagccacgt	gtgcacagcc	cagggtgtca	12360
agcggtggc	ctcttagct	cageaaagcc	ggaatgaagc	cgggcgccgt	ggctcacata	12420
tgtatccca	gacacttggg	aggcagggtt	aggcagatca	cccgagatgt	ggagttcaag	12480
accagcctga	ccaaatctgg	gaaacactcg	ctctactaaa	aataaaaaaa	aattagccgg	12540
gtctgggtggc	acatgcgtgt	aatcccgatgt	attcccgagg	ctgaggcaga	agaattgttt	12600
gaacctggaa	ggccggagggtt	tcgggtgaccc	aaaggatgc	cattgacatc	cagccctggc	12660
aacaagagcc	aaactctgtc	tcaaaaaaaaaa	aaaaaaaaaa	agccacagtg	aagagccccc	12720
cagatcccggt	ggggctctac	agaaaaataggc	atagccatct	tcccaacggc	cgagaccccc	12780
agaatagtt	tttggggaaac	ctgtatctt	ctactcgag	ccacttccag	gttttatatt	12840
aagttagggt	cagccatgcc	actcacacac	ttggatgttct	ttggaaatgt	ctcagccctt	12900
tcaagccccc	aaggcctcagc	aaggccctgt	gtgggtcagg	tgcaagccgt	catgcctcta	12960
atccctgtcc	tttggggaggc	ttggggggaa	ttttgggttt	aggccaggaa	tttggagacta	13020
gcctggggaa	catagtgtgaa	ccccatctt	acagaaaaaa	ttttaaaatgt	tagctgttg	13080
tgataacatg	tcctatggc	ccccactact	ctaggaggctg	aggcaggagg	atcgcttgag	13140
gccaggagg	agagacttagc	gtggtaaaca	ttggcaagacc	cctgtctcta	aaaaaacaaa	13200
aaaaagacaaa	aaaatcatgc	cccatgtgt	tgacactgtc	ttggggaggcc	ggggttccag	13260
gcctttcagg	gacttgcac	agaccctcta	gaagctggac	aaaaaatgag	gaaaccaggc	13320
caggcttgg	ggcttttgc	tgttaaaaaa	acactttggg	gggcaaaaggc	agggtgtatc	13380
cttaagggtca	ggagttcaag	accatcttgc	ccaacatgtt	gaaactccat	ctctactgtaa	13440
aatgcaaaaa	gtagctgggc	acatgtgtaa	tcccaatgtt	tttttttttt	tttttttttt	13500
gaggcaggag	aatcgctgt	accaggagg	cgagggttc	agtgtggccaa	gtcaccccca	13560
ctgcacttca	gcttggggca	cagacgtgt	ctgtctcaaa	ttaaaaaaaa	ggggggggcc	13620
ggcgccgggg	cteacgcctg	taatcccc	actttgggg	ggcgaggccg	ggggatcag	13680
aggtcagaag	atcgagaccc	tcctggctta	cacgggtttt	ccccgtttct	actaaaaat	13740
aaaaatttag	ccggggatgg	ttggggggcc	ctatagtccc	agactactgg	ggggctgggg	13800
caggagaatg	gggtgaaatct	ggggagggtt	gtttgcgtgt	agcccgagg	gcccactac	13860
actccaggct	gggcgcacaga	gtgaaatctt	gtctcaaaaa	aaaaaaaaaa	aagggtgggg	13920
aaaccaggct	cataggtat	cgaagtctca	ggccaaatca	cacagcacac	gtggcgagcc	13980
agagccagg	cttggggggc	cggttgggg	gtatgggg	gtatggccct	ggggatattga	14040
ggacaatttt	aatctttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	14100
ctgtcgccca	ggctggagg	cagtggcacc	atctcagctc	actgcacact	ctactcccg	14160
ggttcaactg	attctcttc	tcacgcctcc	cgagtagctc	ggactactaga	ccgcctgcac	14220
catgccttcc	taatttttgc	atttttttgt	ttttttttgt	ggatgttttt	tttccatgt	14280
gtcttgcac	tcctgcaccc	agggtatgt	ccccatctgg	cctccaaatg	ggctgggtat	14340
acagggtgt	gccaccaagg	ccggccgca	tttgaatctt	ctgtggggcc	aggctggct	14400
gtccaggggg	gtatggcagg	ccggccggcc	gttgcgtgt	tgcatgtgt	ttggcccgac	14460
cacttgggg	atcgagggtt	ggggggggcc	caagccagg	agggtcaagg	tgctgtgtcc	14520
atgtattgt	atactgttgc	ccaggctggg	caacaaagggt	agacccttca	actaaactaa	14580
tgaatgtat	aatgtatgaa	tgaaggcaac	ctgagctgt	ttcccttggcc	tgccaaactgt	14640
cccaaggccca	tccctcagcc	ctccctgtgt	ctgagggccc	tgcaaggccc	acacaggccc	14700
aggctccatc	tttttttgc	aaatgttgc	tttttttgc	tttttttgc	tttttttgc	14760
ccgggttccatc	ttggatgttgc	ataactgtca	ccgttcttctt	ggagaaaggcc	ttccgttcatct	14820
caggcccttgc	tgttcaggat	tcaacgtca	aatcttcctt	ttagggtgt	ggcttccctgt	14880
agccctgggc	tgtgtgtcgg	ggacttcagg	gagaacgggg	ccacccccc	ttcttcctca	14940
cagagctttc	agagggtggg	ggccctccctg	atcccctccc	agatcccaag	ggccctgtgt	15000
ccccccaccc	gcagatgggg	actccgtgt	gttgggttcc	gactttggata	tttttttttt	15060
agcacaacage	aggcaccgtt	gttgcgtgt	ccgttccat	ggagaaaggcc	ttccgttcatct	15120
tgtcatctcc	cctgtgttgc	acacacccgg	gtcagctgt	tttggccccc	ttggggccatgt	15180
gactgaccc	tttggggccat	tttgcgtgt	tttgcgtgt	tttgcgtgt	tttgcgtgt	15240
gatcagggtcc	aaaggccagg	ggccgttggcc	acgcctgtt	ttccctgttca	tttggggagcc	15300
gagccaggca	gtatcgcgagg	tcaggatgtt	gagacgtgt	ttggccatcat	tttttttttt	15360
cgtctctact	aaaaatacaa	aaattggcc	ggcatgtgtt	tttttttttt	tttttttttt	15420
tgaggcgagg	gaatcgctt	aaaaaggggag	tttttttttt	tttttttttt	tttttttttt	15480

-continued

actgcactcc	tgcctgggta	acaaaagcaag	actccatctc	aaaaaaaaaa	aaaaaaaaaa	15540
aatatcaagt	tggcgacggc	ccaggaaagca	atcttgggct	gggcacaccc	cattttctgt	15600
acatgaaggc	agcagacacct	cgtcgctgtc	acagggttgc	tgggggttgg	ggcttagagga	15660
gaggaggagc	ctctgttgag	ggctccaaga	agggacacag	aggtgtgtct	catgggtct	15720
ggagacacct	tttgggttgt	tcgtgcctc	catcctgggc	cactttgggg	aggtgaagga	15780
gggaagcatt	aagggacaag	accccccggt	cccaatttct	ctccggagcc	agggtttctc	15840
tgattcgaag	aaacagggtt	cacaacccag	gaagtccact	gatggcat	gcccctgggc	15900
atcagcattt	agggtgtatc	actgggtct	gcacccctca	aggctgtgt	gccccattct	15960
gggcgcceca	aaggggaga	aaaactctg	aatgtgcacc	gggacaggac	ccatccatg	16020
cgttgcacaga	acctcatgt	ageggccggc	gctctggata	ttcaaactgt	ggagcccaac	16080
aaggggacgc	cgtgtgtcg	gaggggtcg	aggggttttg	aggggtgtcg	atgggtgtcg	16140
tggtgaccca	ctggcatgcc	taatacatgc	aagacctgt	gcccaggcc	gagtcgcaga	16200
gcccacccg	ccaagggtac	tgtgtggca	gggtggccac	atccatcagc	tgctggcac	16260
aggattgcga	gcccgtggca	tctctggcca	atgcccaggca	cctctgaagg	gccttccggc	16320
tacagagatc	ctggggatg	ggccgggtgt	tcctctggca	tcatgtgtc	cctgggtct	16380
cattgaact	caacttctgt	tgaccggca	ggcacttctt	aatagatgtc	ctactcttt	16440
tttttttttt	tttggaaacag	aatctcactc	tgtcaccctg	gttggagtgc	agtggatata	16500
tcttggctca	tcgcacccct	tgcctctgg	gttcaggca	ttctctggc	gcagccccc	16560
gagtagctgg	gattacaggc	tcatgtatc	atgcccggct	aattttttgt	attttttagta	16620
gagacggatc	gggggtctgt	tgccaggct	agtgtggcac	tctctggctc	aagcagtcca	16680
cctgccttgg	cctctcaaa	tgctgagatt	acaggcatga	atcaccacac	ccagctaca	16740
ccatccatc	catcgatcca	ttcatccatc	catccatcca	tccatttata	caaagtctgg	16800
ctctctgggt	gtgtggccca	ggctggagcg	cagtgttat	atctcagtc	actgacccctc	16860
tgggttttgc	ctcagcttgc	acctctgggg	ttaaagcaat	tccctctgc	cagcatccca	16920
agtgcgttag	actacaggta	cacaccacta	tgtctgact	tttttatttt	tgtaaagatg	16980
ggatctgtct	gtgtgttcca	ggctggctct	aggtgatact	ccagctggg	cctcccaaa	17040
tgtgtggatt	acagggtgt	gcccactgt	ctgtgtctg	ctagtcctg	gagcaccaaaa	17100
ctccctgcct	ccacaggccc	tttgcactgt	ctgtgcccc	gacataccct	cctgtcaggc	17160
tctccctaa	tttgtctctt	ggtgactcc	tattcagact	tcaaggccca	gccccaaaaga	17220
ccccctccctc	tctggaaaggc	cttgcaggct	ctaccacaca	gagctttttt	agctctttct	17280
tacttatacc	tcagggtgtca	gtgtcagatc	caccccttac	agcaagtctt	ccctgtatccc	17340
tagactgggt	cgggacttct	tgatgtttt	tgtctatact	ggtcagatg	gataactaag	17400
tgtatctgt	tgtatgtgg	tttatgtct	gcttcgtgc	tccacaggca	ggaatgacat	17460
ttgtttttt	tttttttgag	acagggtctc	actcttggcc	caggctggag	tgcaagtgtt	17520
caatcatagc	tcaactgcgc	ctcaaaactc	ggggctcagg	tgatctccc	acccatcgcc	17580
ccggagtagc	taggactaca	agtatacaac	accacacctg	gctaattttt	tttttttttt	17640
tttttttttt	tttttgagat	ggagtcttgc	ctctgtcgcc	aggctggagt	gcagttgtgt	17700
aatcttttgc	tcactgcaac	ctccacccctc	agggttcagg	gcaatttctc	tgcctcagcc	17760
tccctgatag	ctgggactac	aggcgctgtc	cactaggct	ggctaaattt	tgtctcaact	17820
atgagacagg	gtttcaaat	gttggcagg	ctggtctgc	actcctgacc	ttgtgatcca	17880
ccccgcctcg	cctcccaaag	tgctggatt	acaggtgtta	gccaccacgc	ccggcccaat	17940
ttcttgtact	ttttgttagag	atggggcctt	cccttgtgt	ccaggctgt	ctcaactcc	18000
tgggctcaag	caatccccct	gcctcagatc	ccccaaagtct	gggattacag	gtacaaggcca	18060
ccacatctgg	ccttgatacc	tgttttgatc	actttttggat	catggcatgt	agttaggact	18120
cagtaaactc	agctctact	tgcataatagg	aaagaaggaa	ggaaggaaagg	aggaaggaa	18180
ggaaggaagg	gaaggaggga	ggagggagga	gggaaggaa	aaaggaaggaa	aagaaggagg	18240
gagggaggaa	ggaaaggaaa	aaggaaaggaa	ggagagaggg	aaggaaaggaa	gggagggaaag	18300
gaagggaggaa	ggggaggaaa	ggagggaggaa	aaggaaaggaa	ggagggaggaa	ggagggaggaa	18360
ggaaggaagg	aggggaggaa	ggagggaggaa	gagggaggaa	ggagggaggaa	ggagggaggaa	18420
ggggagggaa	gaagggaggaa	gggaggaggaa	ggagggaggaa	ggagggaggaa	gggagggaaag	18480
gaagggaggaa	ggggaggaaa	ggggaggaggaa	ggagggaggaa	ggggaggaggaa	gggagggagg	18540
agggaggaaa	ggggaggaaa	ggggaggaggaa	ggggaggaggaa	ggggaggaggaa	ggggaggaggaa	18600
gagaggaggaa	ggggaggaaa	gagggaggaggaa	ggggaggaggaa	ggggaggaggaa	ggggaggaggaa	18660
ggacagaagt	aacatccgga	agggtgtgt	tatctgtcaa	gtgagtgtc	atctctgtccc	18720
aggagtgtgt	cccgaggagg	tgtggccaca	cccttggcag	gaatgtgtc	ccgggactgc	18780
cagagcttt	gaggaatagc	gttggggagg	aaagggttgc	gecggtttcg	ggctcacacc	18840
ttccaaggat	gtcagaggat	agggtgtgtc	acaggtgtt	cctgggtctc	tgggtgtta	18900
ctgactcacc	gggactctgt	taatattaat	actcagacgc	cagactgcag	cctgttctcc	18960
agaggcccc	agggaaatga	atcaatgtat	gtctgtatc	ctgaatgtgt	tttagaggct	19020
gaccctgggg	aggcagctaa	cccagctgt	ggaacttggca	aggctttct	gaggaagggt	19080
tttttgggct	ggggcttggaa	gtctgtatag	gagtccacca	aagagaaat	tttaggggggg	19140
gttgcacacgg	ggggatgttg	ggggaaactgc	aagtatgtc	tgtgcctgg	ctctttggga	19200
agcgtctgga	gagactttgg	tggggaggatc	tggactggac	accatttcta	tcctggccag	19260
tacccagcaa	agtaaagtga	gaggccgggg	gggggtggctc	acggcgttgc	tccctggat	19320
tttgggaggat	gaggcaggag	aattgtctca	acccgggggg	cgagggttgc	agttagcaca	19380
gttgcacta	ctgcacttca	gcactccac	ctggcgttgc	gagcaagact	ctgttctaaa	19440
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	ctgggtgtcc	tggctcacac	ctgtatccc	19500
accactttgg	gaggccgagg	cgggccgtac	acgggggtac	gggatcgaga	ccatccctggc	19560
taacacatgg	aaaccccgct	tctactaaa	atacaaaaaaa	ttagccgggc	atgttggcgg	19620
gcccctgttag	tcccagatc	tcggcaggat	gaggcaggag	aatggcgtga	acccaggagg	19680
cgagatgtgc	agttagccga	gatcgcgtt	ctgcacttca	gcctgggtga	cagagcgttgc	19740
ctccgactca	aaaaaaaaaa	aaaaaaagggt	tttggggat	ctgggtgtac	aaagtctagt	19800
tggacaaact	ccatttgagaa	agtctatgtt	gatgaggagg	tttgggggggg	catggcgct	19860
gggtgagccc	tcttagcact	gccttccct	ttctctctgc	cctgatccat	cccaagacac	19920
agcgtggggac	aagaagtctag	agggacccct	gggtgtctga	agcaatggac	cggaccatt	19980
gtacagatgg	gaagactgt	gtggggcagg	acagagccca	cgcttacaca	cctaagctca	20040

-continued

gcttgggttc	tggagccat	gaaggcaaga	gacaggcgtc	cctggagtta	ggctccacat	20100
agaggcaagt	tgacttggcc	agcttctggg	cccagcgctg	gggcactgga	agacctgcca	20160
aaaaacttag	cagaagtag	gtactagaad	tctggatca	gcctccaggc	taacagccgc	20220
gggaccttc	ggcttccttc	ccccaaagct	gggggaaggt	gagctttcc	tccccctcag	20280
ccctggccca	gagaagaaaa	cagagctgc	aaggtcacac	agcaagaatg	tggcaactgga	20340
actcagctaa	gagggtggag	agggocttgg	aatctagaga	ttagccggtg	tctggaatgc	20400
gtggggcagg	agtgtctca	gaggaggacc	gcaggggctgg	gcatgggctc	aggagcttcg	20460
getttgtcc	ctaaggccaa	tggggagca	tggagggtat	gagagcagga	gagaagccaa	20520
agatgttca	ggaggtggct	ctacattata	aaggacccaa	gaggagtcca	acacagaccc	20580
tgtgacccgt	gatttgcagat	ttctgtccgc	ctctggccaa	gtaaaaagtt	gtgttagaaag	20640
ttgttaatcat	ggccggggcg	tgggtgcata	cactgttaat	cccagactt	tgggaggcca	20700
aggcatgtgg	atcacgagg	caagagatag	agaccatct	ggccaaactt	gtgaaacccc	20760
gtctctacta	aaaatacaca	aattagccag	gcatgggtgg	gcatacctt	agtcccaagtt	20820
actcaggagg	ctgaggcagg	agaatgttt	gaatccggga	ggttagagg	gcaatgaggt	20880
gagattgcgc	cactgcactc	cagectgcca	atagagcaag	actccgtctc	aaataaaaaaa	20940
aaaaaaaagg	aaaagaaagg	gtataataag	gttgggtcga	tgggctcaag	cctgttaatcc	21000
cggcattttg	ggagggcggag	gaggacagat	cacttggagg	caagagttca	agaccagcct	21060
ggcccaacatc	gaaaaacccc	atcttacta	aaaatacaaa	attagccagg	ggttagggca	21120
catgcctata	atccctagcta	ctccggaggc	tgaggcaggaa	gatcgtctg	aaccggggag	21180
gcagagggtg	tggtaggc	agatggtgc	attgcactcc	agccctggca	acacagtgaa	21240
actccatctc	aaaaaaaaaa	aaaaaaaata	ataataataa	tactaacaac	ctgtgtgtct	21300
gtgttgc	ttccgtttt	tgatgttta	taatttattt	ttggcccccatt	ggcaactccg	21360
agatgttaggc	agggcagtc	tcatacaccc	agtttattt	atatttttattt	tttatttttattt	21420
tttatagaga	tggcttaggg	gatgttctact	atgttggcca	cgctggctt	gaactccctgg	21480
ccccaaagca	tcctccacc	tcagctcc	aaagtcatca	cgtcatttt	tagctgcggg	21540
tcctctggcc	agtagttgc	gtccgtgtgt	gcagctccca	cgcccttccat	cagggcggtt	21600
ccacacgttga	tgaggggagg	tggtaggc	ctgccttaaga	ccccctgtcc	ctctggggcc	21660
agacatctt	caaaggctat	caggagccag	gccctgtctt	gctctgtccc	tgtcaatcc	21720
caacagaggc	acaggcccc	gttttctat	ctgtaaaatg	gggtggggcc	agcgcagttg	21780
ctcatgcctg	taatccccagc	tctttgggg	gctgagggtgg	gtggatca	agagttcagg	21840
agttcaagac	cacccgttgc	acatggtaa	atccctgttc	tactaaaaat	acaaaaaaat	21900
tagccaggca	cgggtggc	cgttgcata	cccagactt	tgggaggctg	agacaggaga	21960
atcgcttga	cccaggaggc	ggagggttca	gtgagccag	atcacgcac	tggactgcag	22020
cctggggcgc	agaatgagac	tctgtctaa	aaaataaaat	aaaataaaat	aaaataaaat	22080
aaaaatgggg	gaaaaacctgc	tgtatggat	tcttggggag	acaacgcocag	cggttccagg	22140
aatcagacca	cgctggcc	cgtccctgt	gttagaaaa	cgtagctgt	cattactgt	22200
tacgccccc	ggggaggagg	aatcaggago	cactggcttc	tctgtgtgtt	ccagggagg	22260
agggcagggg	aggta	agagagag	gagtgttca	ggatcgatg	gaggctgagc	22320
ttctctggca	tgccttagt	cacacccgtc	gacccgttgc	cagcgcagag	gacaaggaggc	22380
aaaaggc	aggcagggtc	tggatctgg	cctctccctgt	tcatccagg	gccactccgg	22440
gcattatcta	gtt	gagacgggt	ctcgctctgg	cgccccagg	ggagtagactgt	22500
ggcgcgatct	cggtcaactg	caagctctcc	ctcccagg	catgcccattc	tcctgcctca	22560
gcctccggag	tggtgggg	tacacggcc	cgccacat	ccccgtact	tttttgtatt	22620
tttagtaggc	acgggttgc	cgatgttgc	cgaggatgtt	cttgcatttcc	tgaccttgcg	22680
atccgcctgt	ctcgccctcc	caaagtgtt	ggatattc	cactgtgc	cgccatattca	22740
ggcttaatca	tgacagccat	cetccctaa	gggtgcacc	ggtgggggtgc	ggggccctcta	22800
tctctgtact	tttgc	ggcccccctt	gccccactt	gcccacttgg	ccatcaccc	22860
cagcttgggg	cgagggg	gatatgttac	tgtccagaaa	ttggggcc	gaggctcc	22920
gaagaggagg	gtgtcaaaat	gcccagg	atcccttcc	ttggggagg	aaccaggagg	22980
cacgccc	cacaccc	acacccgt	actggacatt	tctgggtgg	tcataagaag	23040
taacacgttgc	aggcggggc	cgttgc	ggctgttaat	ccacggactt	cgggaaagg	23100
aggcaggagg	gtca	cgttgc	ccatgttgc	gcaatatttgc	ggagacttgc	23160
cacttgc	aggagg	tttttttttt	gggggggg	tttttttttt	tttttttttt	23220
aaaaaaataca	aaaattagct	gggtgtgg	gcacgtgc	gtatgtcc	ctactcggtt	23280
ggctgagggt	ggagaaacac	ctgaggccc	gaagtttgc	ctgcgttgc	ccatgtatc	23340
accactgcac	tctagctgg	gtgacagac	aaatctgt	ctctaaaaaa	accaaaca	23400
aacaaacaaa	aaaaaaatag	cctgtatgttgc	tgtcagacgc	ctgtatgttgc	agctacttc	23460
gaggcttgc	ttgggaaatc	acttgc	tttttttttt	tttttttttt	tttttttttt	23520
cgagacccca	tcttc	aaaaattat	aaattatc	cacgttgggg	ctcacac	23580
taatccac	aaaaat	aaat	tttttttttt	tttttttttt	tttttttttt	23640
cacccgttgc	cacaa	aaac	tttttttttt	tttttttttt	tttttttttt	23700
cttgggttgc	ttggactc	ctgttgc	tttttttttt	tttttttttt	tttttttttt	23760
atttgc	ggggagg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	23820
gggtggagag	tgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	23880
tcacacgttgc	tcagg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	23940
gggtgtctga	gtgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24000
ctcatccagg	gttag	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24060
aacggcc	ttaggg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24120
atttgc	aggagg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24180
cccagg	agg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24240
cttgggggg	ttgt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24300
gtgtatcc	atgtct	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24360
ggcccttgc	catgtt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24420
ggggggggcc	actctg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24480
ggggctc	cac	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24540
ggggagata	ccagg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24600

-continued

ctcaccggcc	cccacccgc	ccctegggcg	gcccggccca	cttggccga	gttctctcc	24660
ataaaaaggcc	agggtttggga	gccccccggg	agggggcggt	gggtccacgg	gccgcccccg	24720
cccaccgtgt	tccccctccc	cactcgagc	cctcgccgac	cgggggggtgc	agcccagccc	24780
ccttcggccg	accccccacgc	cggggccgt	gaccacgacc	ctgaccggcag	tccccactcc	24840
ggatcggggc	cacagcaccc	cccggcggtc	cccgcgagtc	ccgegetctg	tccccatttt	24900
gegatttgaag	aaatttgagggc	ccgcagcggt	cgtagccgg	ctggggocac	ggggttgggg	24960
actgaaacggc	cggaaaacgcc	acacgtgc	ccaggtggca	agccagctcc	tcatagacgc	25020
acacgaaccc	ttctgtcaagt	gggggtcgat	tgttatattt	cacctatagg	gattggggc	25080
ctcagggtatg	ggggggccctc	ccccgcaett	agactagaat	ccccccctcc	cgcagcgctg	25140
caacttagct	cttacgtgc	cctggcaca	cgatactttt	actcacagcc	tcagttttcc	25200
tgtctgttaa	atggggggag	tttgcat	gccccgtttt	cacaacgggt	tgtgtcaagt	25260
acaacaccca	gocaagggtc	agtgcagggg	aacaggagga	atggggcaag	ggcgggggcc	25320
ggcgccacccc	cagccccgc	cctggcgac	accctctca	gaagaacctg	catttaggg	25380
gagacggatc	ggacactaagc	cggccccgg	gagactcgag	tgagacttag	acataaaggca	25440
caccaatcaag	aaacacccgg	gtgtcgat	aatgcagctg	cgcctccccc	ccccacccca	25500
cttggggagcg	ttagtcaigt	cggtcaggg	gaggctggga	agccqagct	tggtctctat	25560
gtggctccac	ccgcctggct	ctggggccgg	ggcaggtctc	gcctctct	gggcctcaagt	25620
ttcccccttc	ctgtgtcgag	gccaggccgg	ggcctgacc	cctaggggga	ggaggggccgc	25680
gaggagccgg	cggcggggg	gccccgggg	ccctcggggg	agcgccccc	cctccgcucc	25740
cggcccccgg	ccccctgtgc	tcagaggcag	gggtggccct	gagccggcca	cgtccggcc	25800
ccacaggtcc	gccataaaaag	gcaaacgggg	cgggtcgctc	tgccaaacct	agtcatccgg	25860
cccgccggca	gcgcgagcga	cttccctccg	gggcttcccg	tttctggagc	ggggcccccga	25920
gggggttccg	gggtgtggccg	cgacttctga	aatggccggc	tgcttctccg	aaaaaggccgc	25980
ggaggggccgg	ccctcggagc	catgggggg	agggggat	ggccgggggg	ggggggggat	26040
ctccccccgc	ctgccccggc	ctccggat	atcccgccg	tccttaaccc	ctgagctgc	26100
gggagggggg	gaaggggctg	ggctccggcc	ggggctcggg	cgtcctctgg	tccagcaggg	26160
agtggcttcc	cggaagacgc	cggcgagggg	ggcccaagtg	ctgtctcoga	gaaatctcg	26220
cgaggccctcc	ttggcgagaaa	atgtcgaggc	gtgtacgtcc	atgtgttaaa	atgtgaaaaa	26280
cgcggccggc	gaaaccaccc	agaaccctt	cctgcaggga	ggaggggtgg	ttggaccaac	26340
agggggatcg	tcaactgcgc	aaggcaggga	ctaggccat	gttggacaac	gggggtctac	26400
ctcaatgttgc	cccgacacaca	gacgggttcc	tactgtctgg	ccccacttat	gtgaagtct	26460
agaacaacgg	ccacatttc	tggtgacago	aaacagggtt	catttttat	agatgttaggg	26520
gcagggtattt	atgggggggg	gcacagagaa	gagcatgac	tgggggttga	ggggggccatc	26580
caactgattt	gagatccat	atttcttga	gaaaacaact	gaataaaaca	tagcaaataa	26640
ataattgtt	catccgggtt	gttattatgt	accttctgt	ctttctgt	tctttgaat	26700
atttcgggt	ggcgccgtgg	ctccctgctg	taatcccac	actttgggg	gccaagggtg	26760
gcggatcacc	tgagggtcgaa	agttcgagac	caccctgact	gacatggaga	aaccaattag	26820
ccggggcggt	cggcgcatgc	ctgtatcc	acctaactgg	aggctgaggc	aggagaatcg	26880
cttggatccgt	ggaggcgag	gttgggtgt	gcccggaga	ggccatttgc	ctccagctg	26940
gacaacaaga	gcaaaaactcc	gtctcaaaaa	aaaaaaaaaa	agatttcaca	ctttccaaaa	27000
aagggaagaa	agcgaaggg	caaaagctc	gagttggaca	tggtggtcaa	ccactgtgc	27060
cctcagtttcc	tttttata	ctcatgggc	aataatct	ttgggggtaa	ttaaatatgt	27120
atgcgataat	gcaggcaaaag	tgctaaatgc	gaggcttggg	catagatgt	gtgtctcagca	27180
aaggcgcttg	gtgaggggcg	ggcccttctc	ccggggggaa	gggggggg	tgctgggggt	27240
agggtcttgg	atggggaaatgt	tctctggcca	cgacttctgt	caccctgctc	ctgctctggcc	27300
cgacctggcc	ttgttgggg	ggcttgggtt	tctgacaaat	ggggccagcc	aatcccaggc	27360
cctgcctgt	ttgttccatc	ctgtctgt	ttgttccccc	ggccagggtc	actccgggt	27420
agactcgatag	caatgttacc	caacgggaca	caggcaggca	gttccacca	cctccgattc	27480
caggcggccg	gatcttcagg	cagectgtt	gggcagtgag	gacactgggg	ttccagggtc	27540
cagccggta	ggccttagagc	cctggccg	gcactgtgc	gggatcttag	cccagagagc	27600
ccccaaggcc	tgtggccggc	cagagccgg	ctgggaggag	ggcaggagag	ggtgggggtt	27660
ggggaaaggac	gagaaaaagag	aaagttaaaag	gagaacacgt	acttttctgt	atgttatttt	27720
tgaacacccgt	tgggtatgt	ggccgttcc	caggcttct	gagaaaaatgc	ccccatgacc	27780
ctgtatgggg	atattaagacc	ccaaggcaag	gcacagtggc	gtgtgectgt	aatctcagca	27840
cttgggggg	tcaaggccgg	aggatgttgc	gaagctggga	gtttgagacc	agccttggta	27900
atgttagcaag	accccatctc	tacaaaaattt	ttaaaaattt	agccacagat	gttgggtgcac	27960
atctgtatgc	cccgatctacta	ggggggctac	aacaggagga	tcaactgtgt	ccaggagggtc	28020
aaggctcagac	tgagcggtga	ttgttccacc	gcacttccaa	cctgggcaac	acagtggagac	28080
cccgcttc	aaaaaaagaa	agactcccaaa	atacgcaag	gtcaagcagc	ttgaccaagg	28140
tgaccatgt	agaaggcggag	tcagggggaa	tcggggacca	ggggagaaaa	aaagagatatt	28200
gacagacacaa	tggctgggg	gacttgc	ccctcggggg	ggggtagggaa	gcagctccg	28260
ggaatgggt	cccaaaagaag	caggaatgt	gggtgggggg	aagggtctca	tttctgggg	28320
aggcccagg	aaaggccctca	tttctgggg	agggccagg	agggccagg	gtgcagagaa	28380
gccaatgtaca	gttggggatgt	tggggactgt	gcccaggacag	actagacgga	gtctgtcttt	28440
gggttccacc	tagtggggaa	cttcttact	gtcattccctg	gagatgttgc	cctgtgcacgt	28500
ccccacccgg	tgccccggcc	accatggca	tccccatcc	aatcatctt	caccaccccc	28560
aggctcggt	cgcttccccc	caggcttgc	tggacaatgt	gcccattgg	gtatctcg	28620
ctggccccagg	gttctggccat	caccatcca	ctaaatgttgc	caggccaccc	ggggaggatatt	28680
gtttccct	ttacacatgt	gaggaaactgt	gttggcacc	gttggctcat	cctgtatattc	28740
cagcaacttg	gggggggtgaa	gccccggat	cacctgtggat	caggatgttgc	agaccagact	28800
ggctaaacaca	gtgaaactcc	atcttattt	aaatttacaa	aaaggacccgg	gcgtgtggc	28860
aggagctct	aaacccagct	actcttgc	ctgaggcaga	agaactgttt	gaaccacagg	28920
gatgtatgg	gttgcgtgt	gccaatgttgc	caccattgc	ctccagctg	gacgacagag	28980
cgagactctg	tctaaacaca	caacaacaaat	aaaacacaga	tgaggaaact	gagggtctaga	29040
ggggcgccag	cacttgc	gggttgcgttgc	cgacaatgtc	ccaagctgg	gagtttccat	29100
gggatccaag	cctatccccca	agcagccca	gggacacaca	cagcatgtc	gcacacaatc	29160

-continued

gcacaccctg	ccccggcacc	cctgagccctg	ccgcacatcgtc	attcctgtgc	aactaggcac	29220
cctcgagtctc	cacatcgaaa	tggctctggg	gtacacagac	gaagcagaga	gccggcaccc	29280
catecctcggt	ctgagaaatc	cagettcc	tggctggaa	acccgggtat	ggtaacggccc	29340
tgtcctcgat	gttcggctc	tgctgtgacc	tccacccctt	acctaataatgc	agatttcaga	29400
tttcccgagg	tagcctggaa	ggaaaatagct	tatactgtac	agctgaagca	tggcaaagga	29460
agaatctcc	atcaacagga	gtctgtgggc	tgggtgtggg	ggctcacgct	tgtaatacca	29520
gtagcttggg	agggttgatgc	aggactgtc	gaagccagg	ggctgaggct	gcagtgaatgc	29580
atgattgtac	cactgcactc	taeactggaa	aacagagtg	gactgtgtcc	cccaaaaaaca	29640
cataaaattaa	taaaattaaa	ctaaaataaga	gacggggttt	tttttttgt	tttttttttg	29700
agacagaccc	tcactctgtc	acccaggctg	gagtgcagtg	gcacgatata	ggctcacccgc	29760
aaccacact	tcccggat	aaggcattt	cctacccatcg	tctcccaatgt	agctgggact	29820
acaggcgc	gcacccacgc	ctggctattt	ttttggattt	tttagtagaga	tggggtttgc	29880
ccatgttgc	caggctggtc	tcgactctc	aagctcagg	aatccacc	cctcggtctc	29940
ccaaaggct	gggatccatag	gtgtgagcc	ccacggccag	ccatgagaca	gagttttgtc	30000
atgttgc	gactggtctc	agactccca	getcaagca	tecttttacc	tcggcctccc	30060
gaagtgtca	ggttacagg	gtgaggact	gcacccagg	aggaaactt	gattaaggct	30120
ccacacgtt	gctcagctcc	agccctgca	aaatctccag	ccctgcagag	actgagcatc	30180
gagtccctga	ttttctttt	ccagaatgt	gtggaaagg	gtggttataa	ctgactctc	30240
tccaagacgg	aggttctaa	cttcatgat	gcaagactgg	ctgccttca	aaagaaccagg	30300
aaggacccgg	gggtcttca	ccgcgtat	aagaaactgg	gcacccaa	tgacgggcag	30360
ctagatttc	cagaatttct	taatctgatt	ggggcctag	ctatggctt	ccatgactcc	30420
ttctcaagg	ctgtccctc	ccagaagegg	acctgaggac	cccttgggg	tagccttca	30480
acccaccccc	tttttcttca	acctttctgt	catcatctcc	acagccca	catccccctga	30540
gcccagcaca	ccaaaccact	catcgagg	ccacctcg	atggtaat	aacaatgtca	30600
cttttttaaa	acatgaaaaa	aaaattttt	taaaaagaca	ggccagaact	agtggctcat	30660
gcctgtaa	ccagcactt	gggaggccga	ggcggggcga	tcggctggg	ttaggagtt	30720
aaagaccgg	ttggccacat	gggttacat	tgttotctact	aaaatataca	aaatttagctg	30780
agcggttgg	tgacatgtaa	ttcccgat	ttggggagg	gaggcaggag	aatcgcttga	30840
acacgggggg	cgaggctgc	agggagccg	gacagccca	ctgcacttca	gcctggccaa	30900
cagagcaaga	ctccatctcc	aaaaataaa	acaataaaac	aaaatgcca	gagtatgtc	30960
atgtttaagg	ttaatgtt	actctgtt	ttttactcat	gtgtgtaa	caagccccca	31020
agatgtgc	ccaaaataca	gatgttgg	gtggatc	ctgtacttga	gccccctctg	31080
ccacactgac	atgatctca	ttaaagaaag	aaaagagggg	gcaggttgat	cacttgagg	31140
cagcagt	tgaccagaat	ggccaacata	acgaaaacct	cctctactaa	aaatacgaaa	31200
aactagccgg	gtgtgttggc	ctgcacccat	aatcccagct	actcggggg	ctgaggcagg	31260
agaatact	gaaacccggg	ggcagagg	cgactgaa	gagatcacac	cactgcactc	31320
cagcctggc	gacagagg	gactccat	aaaaaaa	aaaaatgtt	acaatgtaa	31380
actttatgtt	atgtatctt	tactacaatt	ttttttaacc	ttgattctga	gaaggggtcc	31440
ctaagcg	tcaatgtca	atcgcttca	cgccccctga	agcttacagc	cctgttact	31500
cctattgtc	cagcgttca	ggaaatgttgc	agccatgtat	tctggaaat	gtatgttgc	31560
gccccatgc	cgaggatc	cgccccctgt	gtatgggtgg	ctcgagctg	gcacacaata	31620
gttctgtgc	cgacatcc	ctcagcagg	aaattgtgaa	aggcacactc	ggtgcctct	31680
ttccccatgc	gcccacgg	agtcgtcg	acggccacac	aqacctggc	gcacaccctg	31740
cccttagtac	catccagct	gtgggtc	cagggtc	gttggatct	ggacaccgtt	31800
gccaatctaa	gtgcaatcc	gagccctgt	cggtatc	tcgcagac	tcgtcgctc	31860
ttccctgac	cagcctc	ggggcagg	gccttcg	tgagggtgg	aggatgtaca	31920
ggttacatgt	gccccccgg	gaaggatgtt	tcaggat	aacacaatgtt	actgtcaaaa	31980
cctcgacgt	ggagccggat	tggccagg	tgtggccgg	ggagacgata	gggttggggt	32040
gggggttggc	acgttgttca	tgaaggat	tcttacaaga	gaccgtggc	aagttccca	32100
ggttacatgt	ggggatctgc	ctagtgtt	gaatgactcg	tattgttcc	aaggttgcag	32160
ctgtgttgc	cgaaatgt	caggatgtt	cctggatgtt	gggggtgggg	cgaggcttac	32220
tccagacggg	ctgggggggg	gggggttgc	cgcgctgt	cccttcc	cactcttc	32280
ggaccgcgg	gtcgcccc	tctcggtt	ccccctcc	ccagacat	acagccacca	32340
gcggcgacg	ggcgggatg	atgttggaa	acttgggtt	ggacggccc	atcctctcg	32400
aacctccgtc	cctcgaccc	attttaggt	ggggatgggg	tgggggttga	gtctggctc	32460
aaaacccggc	tcaatgtct	tgggtggg	caggcagg	ttggccctgtc	cgggccccgg	32520
gagtgtact	gcccgttcc	ggcccttcc	gggacta	cttcttctt	aaataaagg	32580
aatcttcac	ccagcgtt	ggagtgggt	cgcgagata	aaacaggccc	tccagatgt	32640
ttcccttcct	gtgtgttgc	ggggatgt	tgtgttggg	gagagggg	ggcggccct	32700
ccacctcagg	ctgggggggg	ctgggttgc	ggtccaag	ctggacc	cgacgac	32760
cagccccat	ctaaaatgtc	atggatgtt	ccatgtat	caatggccat	aactgtccca	32820
agaatgtat	ttgttgcgt	tacttgc	ctgcgttgc	tccgcctt	tcacctaatt	32880
cagggcaaga	accgtctc	ggcccctgc	cctggggag	gtgggttgc	ggaccggagg	32940
cgaatgtca	gggttgcgt	ggccatccc	aaatatcgct	tcctttagga	agccggacaa	33000
aacgtatcg	cgggggggcc	ctggatgtt	tccttcttata	gagctgg	aggtacagcc	33060
ccatcttgc	ccccacgtac	ccgcgc	tccttcg	catctccct	acctgggttgc	33120
tttgggttgc	gggttccagg	cgacggcc	gtggcg	cgacgttgc	ggccggccaa	33180
gctgggttgc	gtgttgcgt	tccgttgc	ggggcccttgc	ggcagccaga	cacat	33240
tttcagatgt	tcacttgc	ttcccgagg	ggagtgt	ggcgcgtt	cggtgtact	33300
cagcctccgc	ctcccggtgt	caagtgtt	ccctgtact	gcctttcg	tagtgggat	33360
tacaggat	tgccaccac	cccgat	ttttgtttt	tagtagag	gggggttctc	33420
catgttggcc	aggctgggt	cgaacttcc	acctcagg	atccgttgc	ctcgccctcc	33480
caaagtgtct	agattacagg	cgtgttgc	ccgcgc	cagccagac	cattttaaag	33540
tccttggc	cgggcgccgt	ggcttac	gactttggg	aggccgagg	33600	
ggccggatca	cgatgtcagg	agatcg	catcttgc	aacacgttga	aaccctgtct	33660
ctgtctaaaa	tacaaaaat	tagcagg	ttgggttgc	ccgcgttgc	cccacgtact	33720

-continued

caggaggctg	aggcaggaga	atggagtga	ccaggaggc	ggagctgca	gtgagccgag	33780
atcgccccac	tgcacttcag	cctggccgac	agagcaagac	tccgtctcaa	aaaaaaaaaa	33840
aaaaaaaaaa	tgttatccgt	gcatggtggc	gggtgcctgt	agccccagct	actcaggagg	33900
ctgaggcagg	agaatcactt	gaaccggga	ggcagagctt	gcagtgagcc	aagatcacgc	33960
cactgcactg	cagcctggc	gacagagtga	gactccttct	caaaaaaaaaa	aaaaaaaaaa	34020
aaaaaaaaaa	gaaagccctt	tgaagtggag	aaaaggtact	gggttcaaaa	atgaaaaata	34080
agagtcagg	cgacggagt	cagtggctca	cgctgttaat	cccagcac	tgggaggccg	34140
agacgggtgg	atcacgagg	caggagatcg	agaccatcc	gcttaacacg	gtgaaacccc	34200
gtctctacta	aaaatacaca	aaattagccg	ggegtggtgg	cgggcgcc	tagtccagc	34260
tactcagg	gtgaggcag	gagaatggcg	tgaaacccgg	agctggagct	tgcaagtga	34320
cgagatctgt	ccatcgact	gtagccctt	caaaagag	gaaatctgt	ctcaaaaaca	34380
aacaaacaaa	caattagctg	agctggagg	catatgcgt	tagtccca	cactgagag	34440
gctgaggcca	gagggtac	ttaagccag	gaattcgagg	ctgcagtga	ctatgatgc	34500
attactgcac	tccagctgg	gcaataatgt	gagacccccc	tcttagaaag	aaagaaagag	34560
agaaaggaaag	gaaggaagg	agaaaggaaa	aaagaaagg	aaggagggg	gaaggaagga	34620
agggaatgaa	ggaaggaaag	ttccata	caatcaaa	acacaaaac	caacaactc	34680
actaatgtcc	tttctgcaat	caagagtaa	tttttgtt	atcttttgc	tcagataca	34740
atccacaatgt	taacctgtta	acttcactg	gcctcagccc	ttaatcaaa	gcagataga	34800
aagtcaata	aaagtccct	ggaaaggatt	aaataatcc	tagggggct	ccttatgca	34860
agcagccctc	cccttggct	cctgtccat	attcata	tggtttttt	tatttattt	34920
tttactttat	ttatttat	atgacagg	tcgtctgt	ccaggctag	agtgcagtgg	34980
catgatcttg	gtcactgca	acccatccat	ccaggctca	agccatcc	ccacctc	35040
ctcoctgtgt	gtctggacta	taggacaca	ccagcacgc	cggttaattt	ttttttttt	35100
tttgacatgg	agtctcact	tgtcacctag	gtggagtg	atggcaaga	tctccact	35160
ctgcaacactc	ccgcctccgg	gttcaagca	ttatcctgc	tcagcetcc	gagtagctga	35220
gattacagat	ggcaccacc	aagctggct	aattttgt	ttttttttt	agtagagat	35280
gcgttccacc	atgtggc	gactgtt	aaatctgt	cctcgatc	caccacactc	35340
ggcctcc	aaatgtgg	ttataggct	gaacactgt	gcccggctt	ttttttttt	35400
gagacggagt	ctcaactt	ttggccggg	ttggagtg	tgacatgt	acatgatcg	35460
gcctcaccac	aggcctaacc	tctcagg	aagcgagt	cctgcctcc	cctctg	35520
agctgggatt	ataggcgt	gcca	cctggctaa	cctgtt	gttgggtt	35580
ttttttttt	gagacggagt	ttgtctgt	caccagggt	ggagtgc	ggcatgat	35640
gggctca	acttctgc	ctccgggtt	caagtgtt	tcctgc	gcctctg	35700
tagtgggat	ta	cgccactat	ccggc	ttttttttt	ttttttttt	35760
gatggag	tcgtctgtt	ccaggctt	gtgc	cgctt	tttgc	35820
taaaccgat	cttctgc	acccctt	gtgtt	gggtt	tttgc	35880
ccca	ttttgtat	tttagtaga	gggggtt	acatgtt	ggggatgt	35940
tcaatttctt	gac	tctgccc	tcagc	ctcc	aaagtgtt	36000
gtgagccacc	acgcccgg	atttt	tttgc	ggc	gattttt	36060
ttgtcagg	gtactt	ttctgac	gtgtt	ccct	ttttttt	36120
ctgagattac	aggcatg	cacttgc	ggc	ctt	ttttttt	36180
gggttccacc	atgtt	gactgtt	gaact	ccat	ttttttt	36240
tcggcc	aaatgtct	gattt	ctgag	ccact	ttttttt	36300
atttttacta	aaagaca	tttccat	ttggc	agggt	ttttttt	36360
tcaatgtata	ccaccc	cccttgc	tttgc	tttgc	ttttttt	36420
ctgtgtctt	ctgc	cttattt	tttttgc	tttttgc	ttttttt	36480
ctgtcacc	ggcttgc	catgt	atc	atc	ttttttt	36540
gttcaatgt	tttctt	ttttttt	at	at	ttttttt	36600
gagtgc	gtgt	atctc	at	at	ttttttt	36660
cctgc	c	atgtcc	at	at	ttttttt	36720
tttgc	t	atgtt	at	at	ttttttt	36780
cgccc	cc	tttgc	tttgc	tttgc	ttttttt	36840
aagtgtat	ccat	tttgc	tttgc	tttgc	ttttttt	36900
ccagg	tttgc	tttgc	tttgc	tttgc	ttttttt	36960
taacttct	atgtt	tttgc	tttgc	tttgc	ttttttt	37020
catgac	ccat	tttgc	tttgc	tttgc	ttttttt	37080
ttttgtat	cc	tttgc	tttgc	tttgc	ttttttt	37140
tctctatgt	ttgt	tttgc	tttgc	tttgc	ttttttt	37200
atacgtt	tcg	tttgc	tttgc	tttgc	ttttttt	37260
tttgc	tttgc	tttgc	tttgc	tttgc	ttttttt	37320
atgaactt	atgtt	tttgc	tttgc	tttgc	ttttttt	37380
cttttgc	tttgc	tttgc	tttgc	tttgc	ttttttt	37440
aaatttgc	atgtt	tttgc	tttgc	tttgc	ttttttt	37500
agagagt	ccat	tttgc	tttgc	tttgc	ttttttt	37560
gaataat	tttgc	tttgc	tttgc	tttgc	ttttttt	37620
cagggtt	tttgc	tttgc	tttgc	tttgc	ttttttt	37680
atagaaaata	tttgc	tttgc	tttgc	tttgc	ttttttt	37740
ggcaggag	ttacttgc	tttgc	tttgc	tttgc	ttttttt	37800
cgtctccacc	aaaat	tttgc	tttgc	tttgc	ttttttt	37860
gtaatcc	ccat	tttgc	tttgc	tttgc	ttttttt	37920
ccacatctgc	aaacat	tttgc	tttgc	tttgc	ttttttt	37980
gtgtgttgc	gtgc	tttgc	tttgc	tttgc	ttttttt	38040
acccaggagg	ccgaa	tttgc	tttgc	tttgc	ttttttt	38100
cagaatgaga	ctct	tttgc	tttgc	tttgc	ttttttt	38160
gcccggcgc	gtgg	tttgc	tttgc	tttgc	ttttttt	38220
cacaaggat	tttgc	tttgc	tttgc	tttgc	ttttttt	38280

-continued

aatacaaaaa	tgcgtggc	atggtggccc	acacctgtac	tccagactac	tcaggaggt	38340
gagtccagg	aatcttttca	accaggagg	tggaggttgc	agttagccca	gatctcgcca	38400
ctgcactcca	gcttggcaa	gagagcgaga	ctccttctca	aaaaaaaaaa	attacctgag	38460
agcagtggc	caggaggctg	aggtagggagg	atcggtttag	cccaggca	ggaggctgca	38520
gtgagccata	attgcaccac	tgcactaccc	tagcctgggt	gagagggaaa	gaccctgtct	38580
aaaaaaaaaa	aaaaaaaaaa	aaaggataag	aatggaaa	taccagtgc	tcctggctag	38640
aactgtcgta	agggtttctg	taactgtta	gagggtcagg	gacttcagg	atgtatgt	38700
aaggctccag	ctgcataatcc	cttgacttc	cttagcagt	tcttcccat	gggtcaggct	38760
tgtctgtatt	cataggggc	cacatgcctg	cttgaggctc	tggtagactc	agggtttcc	38820
tacagagcag	ggttctggcg	cagcgttcc	aqgaacacgc	cctgagatcg	ctgtgtatgg	38880
ggcattttcg	gttccatcg	ccatccatcg	aggcatgact	ggtgtggaa	agggtggccc	38940
cctaggcacc	agaataga	tctcagctgt	cctgaaatgt	gtggcagtgc	acagcgtggg	39000
ggggaaagag	agatttaca	ggaaatccct	aggcagtca	agagggctg	gttagaggat	39060
tggtctcggg	cggggggtgg	aggggggggg	gtgtcgccc	gcagcggaa	acacaggag	39120
gggtggcagg	gcatggatg	tgggtgtgg	aggagacctt	aggaaaggga	gcctgcaccc	39180
agggattttcg	cagacttcc	aaaggacat	cgggatctag	ggccaaaccc	tctgcacccc	39240
ccatccccca	tcccccaccc	cagagctct	gtcgagcgc	cctctgcctt	tgccgcagg	39300
aaaccagtgg	agggggcgtt	aggggtgtga	tccagattca	ccggagagg	aaacaacccg	39360
gagggggtcg	caggcaggcc	gtggccaa	gtgtcgctc	tccaaaggct	cgtgaacccc	39420
aggggcgc	ctctccatgt	gcgcacca	ttggggctgt	accttcact	caggacccca	39480
ggcccttgc	ccccgggacg	cctctctgg	gtgtcccacc	gcaccactga	ggccgcgcac	39540
gaccctcage	caccgcacgg	cggtcgatcc	cagctgccc	tctcttcga	gcacccctca	39600
ctccccggcc	gcacccccc	cattggctt	ccagacca	gaccctctat	gttgggttca	39660
tttccgcgg	gocacgcctc	ctggtcggca	aaagagccg	gaccgcaca	gccccccccc	39720
actccgcgt	tccctagccc	ctggggcgc	ggatgagccg	eggccgcac	gggggggtcg	39780
tagtagtgac	tgcgaccat	gtgggtgtga	gcaccgaact	cgagcaact	aggaaaggcgc	39840
tggtcttaggg	ctgtgtcg	tgccatagcc	tcggatccca	cctgagagg	aaaaatttcgg	39900
agcctgteta	ccacccccc	agccccgggg	ctctaatccc	tcccaggac	tgacccccc	39960
cccccgttcc	ccagacccctc	cctattttct	ctccctgccc	aggggggggg	gttgggtggac	40020
tttattgtta	acggaaactgt	tacagttaa	aaaactgtag	tcagagagg	gtaaaactacc	40080
ccgagtcata	cagcgtcg	ggggagccca	gaataataac	tgaggatgg	gaaggaaagg	40140
gagaagctgt	ggctgtgt	atcttctaca	tcccaactct	ccccccct	gatacacagt	40200
tatgagctaa	gcacaggagg	gtccctgtag	cacacatca	gcctgcaggc	tgacgatt	40260
tttttttttt	ttttttttta	gacggaaatcc	cttctgtcg	acaggctg	gtgcgttgc	40320
gegatctcg	ctcaactgca	tctccgcctc	ctgggttca	gcgatctcc	tgccttcagcc	40380
tcccgagtag	ctggggactac	aagcccccac	caccacacc	ggctatttt	tatatttaat	40440
agagacgggg	tttccacata	ttggccaggg	tggtctcgaa	ctccctgact	tgtgatctgc	40500
ccgccttc	ggtctccaa	agtgtggg	ttacaggcgt	gagccac	gccccgcctt	40560
tttttttttt	tttttttttt	ttttaaggaa	gtcttcacgt	gtcggccagg	ctggagatca	40620
gtggcgtat	ctcggttca	tgcacatcc	accttcgtgg	tcaaggat	tctctgtct	40680
cacccctctc	agatgttgg	attacagg	ccggccacca	cgccgcgtca	atttttgtat	40740
tttttagtaga	gatggggttt	caccatgtt	gccaggctgg	tctcaaaactc	ctgacccctag	40800
atgatctgc	caccttggcc	tccaaaaat	ctaggattat	aggcgtgact	cactgtgccc	40860
ggccaaattt	tttcaaaatt	ttttaaaaat	ttttaaaaat	caacactta	aaactacaaa	40920
taactgttca	attttttttc	agccagggt	tttggccgt	gtctgcgtc	ccagcttatt	40980
gagaggctg	gatggggagga	ttgttgago	ccaggagg	aaagggtgg	agctatgtac	41040
aatctactcg	actccacac	gggtgacaca	gtcacaaccc	atctctgaa	taaaaacaaa	41100
ataaaaaaaa	aactaatttt	tttcttagaa	aaaaactgaa	aataatgtt	tagctgtat	41160
ccacaatcac	cccttgcatt	ttcttc	agggttattc	tgcttatata	atggccagga	41220
agttataata	ttacttttgg	ggtttttttt	gttttttttt	ttttagatcg	gaatctca	41280
ctgtcatcca	ggctggaaat	cagtccac	atctccgtc	actgcacgt	ccgcctcccg	41340
ggtcatgt	ccatttcct	gcctcagcc	cctgagtag	tgggactaca	cgtgcccacc	41400
accacgcgc	gtataatttt	tgtatttt	tagtggagac	ggggtttac	ctgttagcca	41460
ggatggctc	gatctctga	cctctgtatc	cgccgcctc	ggccttccaa	agtgtggg	41520
ttacaggcgt	gaggccaccc	gcccggcc	ttatgtttt	gagatgtt	ctcaactgtc	41580
cccaaggctgg	atgcgtcg	tgcatactcc	gttcaact	aactctgc	cccggttgca	41640
atcgatccac	ctgcctc	ttcttgat	gtgcgcacca	cagggtc	caaccaagcc	41700
cagctaattt	ttcatat	ttggtagagac	ggggtttgc	catgttgc	aggctgtct	41760
cgaacttcata	actcaac	atcgtcg	ctcagcc	caaagtgtc	gggatccagg	41820
tgtgagccac	tgtgtggcc	aaatcatat	acatca	gtatgtc	catctaaacc	41880
ttttagttct	gaatgttcc	ccagaaat	cttgcag	aaaaatcc	cccatgttt	41940
atttcgttct	tgcatacatt	taatttcc	cggtatggaa	cacttct	tctgtgttg	42000
gtcttcataa	cggtgtact	ttttttttt	tttttagat	gagctgg	ttcactatgt	42060
gcaggatgt	tgcaatgttgc	cgatcttgc	tcacacttc	caacccctgc	cgccgggtt	42120
caagcaatcc	tcccccacca	gccttggc	tagtggat	tacaagc	ccccccacca	42180
actgactaat	ttttgttattt	tcaatagaga	ttgggttca	ccatgttgc	cagggtat	42240
tcaacttgc	gacccat	tccaccc	acagctccc	aaagtgttgc	gattacaggc	42300
gtgagcaact	gcacccaggc	atgttttgc	attttagta	gagctgg	ttcactatgt	42360
tggccaggat	gttctcaaa	tccccaccc	agggtatct	cctgcctc	cctctcaagg	42420
tgctgggatt	acagaaat	gtcaact	ctggactatt	ttatattt	tcatttatt	42480
atttgttgc	cgaggat	tttttttgc	ccaggctgg	gtcaatgg	aaatctcg	42540
cttactgc	tttctac	ctggat	gegatctcc	tgcctcag	ttcccgat	42600
ctggggattac	aggcatcc	caccat	ggcttag	aaaaat	atgtatttag	42660
atgaatttcc	ataccttta	tgtatatt	aagccaa	tttacatt	tgatccctgg	42720
ggtgcactcc	agtttacaac	tcccttct	gccttgg	acctgg	tacgatttgc	42780
atacatctgc	ttctttttt	ttttttttt	gagacag	ttcacttgc	cgcccagg	42840

-continued

tgagtgcagt	ggtgcgatct	tggccggctg	caacctccac	ctcctgggtt	caagcgattc	42900
tccctgcctcc	caaagtgcgt	cgttacagg	tgcggccac	catgcctggc	taacttttgt	42960
atttttttt	ttcttttga	gatggagtct	ccctctgtcg	cccaggctgg	agtgcgtgg	43020
cacgatctcg	gtcactgca	agtcggct	cccagggtca	tgcattctc	ctgcctcagc	43080
ctccccgaga	gtctggacta	caagtgcgg	ccaccatacg	cggtcaattt	tttgcatttt	43140
tagtagagat	gggggttccac	tgtgttagcc	aggatggct	cgatctccctg	acctcggtat	43200
ctgccccatct	cagccctcca	aagtgcgtgt	attacaggcg	tcagccccca	tgcggccct	43260
aattttgtgt	tttttagtgc	aatggggttt	tgccatgtct	tccagggtgg	tctcgaaact	43320
ctgacacctgg	gtgtatgccc	cacctcggtc	tcccaaagt	ctgggattac	aggeggggagc	43380
caccatgcgt	gaccatgtat	gtttttttttt	tttgagttgg	agtctcactc	tgcggccctag	43440
gtcgctgtgc	actgtgtgtga	tctcggtca	ctgcaatctc	agctctccgg	attgaagcg	43500
tttctcgta	ccagcccttc	gaggagctgg	gatcacaggc	acggggccacc	acacgcagct	43560
aatttgcgtgt	attttttagt	gagatgagga	tccaccatgt	tgcccgaggct	gatctcgaa	43620
tccctgcctcc	agggtatccg	ccggccctgg	cctccaaagg	tgctgggatt	acaagcggg	43680
gcccacatgc	ctggccacat	ctgtttttt	tgtcaacccc	tcgggggtga	gcacgtgtct	43740
cctggggact	gtcacccatc	acctttgtaga	cctgtccctc	acccctgtgt	tccttggca	43800
gggatatctc	ctagccccag	gacagcttag	tcacccctcc	caggtcacgg	gccccgccta	43860
attggatctc	tcacccatca	tgagatctca	tcctccccc	ttttgtactgt	ccaccaagca	43920
gcccacatgc	tccttccatca	ccacccccc	gcctctccca	ccccattgtac	gtgtctcaga	43980
gcccacaccc	tcccccaccc	cactggggac	ccccacccgtc	tctgtccacg	caaattggct	44040
cgggactgtat	tattctggaa	ggaaaacggaa	aggggaaggg	actaaggggaa	gagaaggaa	44100
gaggcttctt	ccgggttctt	tttttagtca	aaaatttcatg	ccccgtgtgc	acttcagtt	44160
agaacacact	tccttttgc	ttaacgtat	acagcggttt	caactacaa	tttctggtaa	44220
aatttgcgtcc	ttgggtatccg	gggggtgtgt	ctgggtcaag	ggggggccct	gcaggctggg	44280
aggacacggg	agggggtgggg	ggggggggaa	acgggtgggg	gaaggtggac	gggggtggcc	44340
ggccctctgt	ctgtttttttt	ccctggacaa	tgggtctggc	gtcgcccgag	agataaaggag	44400
tcttgcgtgg	gttgggtatca	cageggccgg	ccgactgtat	gttggggaaag	gtcccttctt	44460
cgcagacatcg	ggccaaaccac	cccccaaaag	ccccggccgc	tcaagtgaga	tctctttca	44520
ttgtacattc	ctgagagcga	gggtcttcaga	gcccccccc	agggtcacca	ggaggctggc	44580
tgtgagctgt	gtgtctgggg	aggcccttc	ccaggccgag	acacctccc	tggctccagg	44640
acacagtcgc	aaaaagatgt	ttgtttgttc	ttttttttttt	tatttttttt	tatttttttt	44700
tcttttgaga	taggtctca	tttcatcacc	ctggctgggg	cgcagtgggt	tgatcatgtc	44760
tcgctgcagc	tttgacttcc	ccagotcaag	tgatctttgt	gcctcagct	gagtagctgg	44820
gactatggat	gtgcaccacc	agggtcggt	gatttttttt	tttcttttga	gagatggagt	44880
ctcaactatgt	tcggccaggct	gggtggaaac	tcctggactt	aagtaatct	ccccccatcc	44940
ggctaaacacg	gtgaaacaccc	gtctctacta	aaaacacaaa	aaaatttagcc	gggcatagt	45000
gcgggacact	gtgtcccaag	ctactcgaga	ggctgaggccc	ggagaatggc	gtgaaccagg	45060
gaggcggggc	ttgcagtgg	ccgagattgc	gccactgcac	tccagctgg	gcgcacagaga	45120
gagactccgt	ttccaaaaaa	aaaaaaaaaa	aaaaaaaaaa	attccctcc	cctgggctc	45180
tgaaaagcgt	gggatttacag	gtgtgacca	ctgtgttca	ttaagat	aatttatgt	45240
aagtaactta	tgttacttcc	cagccggcc	aacatggcga	aatccatct	gtactaaaa	45300
tagaaaaattt	tggctggggc	cgatggctca	cgctgtat	cccagcac	tggggggcca	45360
aggcaggcga	atcacaggat	caggagatcg	agaccatcc	ggcttaacat	gtgaaaccct	45420
gtctctacta	aaaaacacaa	aaatttgcac	gggtgtgggg	cacccgtctc	taatccacgc	45480
tactcaggag	tctgaggcag	gagaactact	tgaacccggg	aggcagagg	tgcaactgag	45540
tgagatctgt	ccattgtact	ccagctgggg	tgacagagcg	agattccatc	tcaaaacaa	45600
aacaacaaaa	ataaaaat	ggggggccac	cttgcgttgc	cagctacttgc	agagggttgc	45660
gcaggagaaat	cgcttgcac	tggggggat	agggttagat	gagccgagat	tgcaact	45720
cgttccacgc	cttgggtgcac	agagcaagac	cctgtctca	aaaaaaaaaa	caaaaaactt	45780
ccacactcaa	agtacataaa	cctgcagta	atctcttgc	atggttatgg	gtgtcttagt	45840
cgaatctcgca	gtctgtggaa	ggccagggggg	acagctccgg	ccccagttgt	tagaaatgt	45900
agccctgcac	atggggat	aacctgtat	cgtatgtat	ggggggagaa	aggacataaa	45960
atttctcata	atgtttatgt	taatttata	tttgcattat	atattttttt	atattttttt	46020
taaataacat	acattaagaa	atttggggcc	ggcgccgtat	ctcatacttgc	taatccacgc	46080
actttgggg	ggcgagggtt	ggggatccac	aggtcaggag	atccgagacc	tccttgcataa	46140
cacgggtgaaa	ccccgtctct	actaaaaata	aaaaaaaaaa	tagtgggttgc	tgggtgggg	46200
cgcctctatgc	ccagatactc	ggggggctga	ggcaggagaa	tgaagtgtac	ccaggaggcg	46260
gagcttgcac	tgagtgcaga	tcgcgcact	gcactccac	ctggggcag	gagcgagact	46320
ccatctcaaa	aaaaaaagaa	agaaaaatttgc	tgtcaacttgc	ttatcttac	ttaaaaattt	46380
ttttttactt	taggtctggcc	ttttgtgtcc	acacctgtta	taacagc	tttggggggcc	46440
gaggcagggtt	catacttgc	ggggaggat	tcggaggacc	cctggcttaac	atgggtgaaac	46500
cctgtctcttca	ttaaaaacac	aaaaatttac	tggatggggt	ttcaccatgt	tggggaggct	46560
gggtctcgac	tcctctcaga	tgatcttcc	tcctcggtct	cccaaagtgt	ttgggactaca	46620
gggtgtggac	accatggcg	gtcttgcac	ttttttttttt	ttttttttttt	ttttttttttt	46680
cccaaggat	atgtgcgtgg	cacgtatcg	gtctactgaa	ggctccac	ccctggatcc	46740
agcaattttt	ttggcttcgc	tcctcgatgt	gttgggttca	caggcatgt	ccaccatgc	46800
cagctaattt	tgttat	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	46860
aaactctcgta	cctcaatgt	tccactcacc	ttggccctcc	aaatgggggg	cccacataa	46920
ccacttgcac	ccggcccaat	taggtgttt	ttttttttttt	ttttttttttt	ttttttttttt	46980
atggagatctc	tttgcgtcc	ggctgtttttt	ttttttttttt	ttttttttttt	ttttttttttt	47040
ctgcctctca	gatccaaatgt	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	47100
gecgctgcac	tcatgcgtgg	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	47160
tatttttttt	tgagacggag	ttttttttttt	ttttttttttt	ttttttttttt	ttttttttttt	47220
tcggctact	gcaagctccg	cctcccaatgt	tcacgcatt	tccttgcctc	agccctccaa	47280
gtagctggga	ctacaggc	ccggccaccac	ggccggctaa	ttttttttttt	ttttttttttt	47340
gacagggtttt	caccgtgtt	ggccaggatgg	tcttgcatt	ctgacccatgt	gttccaccc	47400

-continued

-continued

tttctcaaaa	gtctgtttta	gatagtgcgtg	ggggagatgtg	atgtctggag	ccctggcagc	52020
tgcccttggaga	ttctggaggac	ctgagccctga	aggcagaggtt	caccctgacag	ggaaggcaga	52080
gtggagaatg	ggggaaatgc	cagggtcctt	tccagctctg	ttacagtgtt	aaacacccctgg	52140
aaactccccct	ctttggggct	tcttgtttct	tgtataatg	aaacatcttgc	ctggttagcc	52200
agttgagttt	gagttttctc	tttcttgcat	ttgaagttac	catgactgca	agagggagtt	52260
ccccaaagggg	tctgggaaatg	gcccgaggaa	ataggggtga	ggtgagactc	cactgccccg	52320
acagtgccca	tgtatgaatt	agaaaaaaatgt	ggccaggcgt	gggtggccatc	acctgtatac	52380
ccacgtttt	ggggaggctga	gtggggcagat	catgagggtca	ggagttcgaa	accagccctgg	52440
ccaacatgtt	gaaaccccat	ctctactaaa	aatacaaaaa	ttagctgggc	ttggggcac	52500
acatctgttag	tcccacactc	tcaggaggct	gaggcagaatg	aatcgttga	acctggggag	52560
ttggagggttgc	agtggcggcga	ggtcgtccca	ctgcaetcc	gcctgggtga	cagagtggaga	52620
ctttgttcca	aaaaaaaaaa	aaaaaaaaaa	aagaaaaaaa	agaaagaaaa	agaaaaaaag	52680
tggggccagggt	gtggggctc	atacctgtaa	tcccagact	ttggggaggt	gaggtggggag	52740
gattgttac	agccaggagt	ttgagaccat	gttggacaat	atggtgagac	cttgccttta	52800
aaaaaaaaaaa	aaaaaaaccta	tctggggat	ttgttgcac	cctttagtcc	cagctactta	52860
ggaggggttag	gtgggaaggat	cacttgagtt	aaggggaggag	accacccctc	atattgttct	52920
atgcccattt	tctgccttca	aagaaagaaa	aagtaaaaac	taaaaggcag	aatgaaaaac	52980
cacaggcaga	cagcccaageg	ccacaccctg	ggccctcgat	ttaaagatcg	accctgtatc	53040
taatcgttgc	tgttatctat	agactacaga	cattgtatag	aatgtcactg	tgaaaatcccc	53100
tatctggttt	tttttttttt	taattaccgg	ttgtatgcac	ccccactc	gtaccccttg	53160
cttgctcaat	cacgaccctc	tcacgtgcac	cccccttagag	ttgtgagcc	ttaaaaggga	53220
cagaatttgc	tcactcgggc	agcttggctc	ttgagacag	agtcttgcgt	atgcccctgg	53280
ccaaataaaac	cccttccctt	ttaatcccg	tgtctgat	ttgtctggg	ctcatccctgc	53340
tacagactt	aggaggcaga	gggtgcgtat	agccaaatcc	acggcaactgc	actccagcc	53400
gggtgacaga	gcaagacccc	accaaaaaaa	gaaaagaggc	caggcgcagt	ggctcacgc	53460
caigctaattt	ttgtatctttt	agtagagacg	gggtttcacc	atgttggca	ggctgtatctc	53520
aaactctgtt	cctcaggtga	tccggccaco	tttgccttccc	aaagtgttgc	gataaacagg	53580
gtggagggtc	aactttttaa	agaaaaaaag	gacgaatcag	gacggaggac	attacaagc	53640
tctgttattt	cgaaatttctc	attggcttac	agaaataact	ttggctatgt	attggttata	53700
tgtttagac	agacccacag	ggtggatgg	gtctgtggc	atctggcat	tagctggcc	53760
agggccaggaa	gttccatgtg	caagtagtgc	caggacgtaa	ttatttatgt	caatagaaaag	53820
ttgatgttgc	ttgttgcatt	tttttttttt	ttttgagacg	gagtttact	tttgccttcc	53880
aaagctggagt	gcaatggcgc	gatctcggt	cactgcaacc	tctgttcc	aggttcaagt	53940
gattctctg	cctcagcc	ccgagtagct	gggattacag	gcgtgtgc	ccaaagcccg	54000
ctaatttgc	tatttttatgt	agagacaggac	ttttcaatgt	ttagccaga	ttgttctcgaa	54060
cttctgttgc	tgtcatccac	cgccctcg	ctctcaagg	gtttaggat	caggcatgag	54120
ccaccacggc	ccgccttta	actgtgtt	ttttttttt	tttattatcatt	tttattctgc	54180
tagttccata	aaagcaatgt	aaacaccago	attctataca	tcttgcgtt	gaactcacat	54240
aatgtcttgc	tccctgtatc	tttgacctt	ttgtcttctc	cagtttattt	ctgtttggac	54300
cactggccac	aggaatggag	ttgggggtgg	gtttaggaga	gagaatgttgc	gtttaggggt	54360
atgacttgc	tcatttgc	ttgttacat	ttaaatgtt	ttctgggc	gataattttaa	54420
ggggggcttgc	atttctcaca	tcaaaggc	aagggtttt	ttgtttgtt	ctcactatttgc	54480
ctcaaaaggcc	acatgttgc	aggcagactc	ttgttacccgg	gcagagaatgc	ttatagaa	54540
tttccaggc	ttggcgggtt	tttcatgtt	gggggacaatgc	agcggaaatcc	cattccaaaa	54600
agaaaaaaa	aaaaaaagaa	atttccagg	caggcacat	ggctcacacc	ttgttgc	54660
agcactttgg	gagatttgg	ttggggatgc	acttggggcc	aggatgttgc	gaccagcata	54720
ggcccacatgg	caagaacccca	gttttacaa	aaacattttt	ttttttatgt	agctgggc	54780
gttgggtgcac	atttgcgtt	ccagactact	gggggggtga	gttggggagaa	tcgttgc	54840
cttggggaggc	gagggttgc	tgagttgt	tcgcaccact	gcatttc	cttggcaaca	54900
gaggggagact	ctgttctca	aaaaatcaaa	taaaagaaaaaa	agaacatttc	cattatttgc	54960
gaatgttca	ttggacggcg	cttggccaga	tgcttgc	aaaggcttgc	acacccaaac	55020
cttggtaaaat	tgcccaatgc	ccatctgggg	ttgttgcac	aaaggggccg	gagctgggat	55080
tacaagcg	gaggcaggcg	ccacccca	tttttttttt	tttttttttt	ccagacttttgc	55140
tgggatgttgc	agggttttgc	caggcagttt	tttttttttt	tttttttttt	tttttttttt	55200
atgggcatttgc	agggttttgc	gggttgcgtt	ccggggaga	agaagggggg	gtggcaggag	55260
agatggccat	aggccccccca	ggagccccca	ggcttccagg	tttttttttt	tttttttttt	55320
cttggggatcc	acggggaga	tttttttttt	tttttttttt	tttttttttt	tttttttttt	55380
ggaaaccttacc	ttgggttgc	gggttgggtt	tttttttttt	tttttttttt	tttttttttt	55440
atttgtacttgc	agagegtatgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	55500
agaacacgttgc	tcggggatgg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	55560
agatggccat	aggccccccca	tttttttttt	tttttttttt	tttttttttt	tttttttttt	55620
cccccgttctt	actaaaaacaa	tttttttttt	tttttttttt	tttttttttt	tttttttttt	55680
gctactccat	agggttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	55740
ccggccatcc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	55800
aacccaaaacc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	55860
tcccccgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	55920
agaggttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	55980
aggtggggcc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	56040
gttggccatcc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	56100
aggttgcggcc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	56160
ggggacatctt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	56220
gcacttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	56280
aggaccacgt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	56340
aggttgcggcc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	56400
acagaagagt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	56460
ggttcccccac	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	56520

-continued

ttcaccagg	gttggcatca	cctccctctg	aaggcccac	tggactgac	agttaaactct	56580
ggagccccc	ctgcacagg	cagtcatgt	atggatccac	aggtgagtc	ctgtctttga	56640
agggttacag	tttagtctc	ggagcagtgg	tggggaggte	acaccccccac	ccacateccct	56700
ctttcaaat	ccccatggaa	ctatgagct	gggtccccca	aactgcatgt	tccccactta	56760
aagctggggc	agatctgtt	tgctgtatcc	aggcctca	ttccctgtt	gtaccagaa	56820
gtccatgcag	ccagaattt	tcagggagct	gggcacgggt	gctcatgct	gtatcccgag	56880
cacttcaga	ggctgaggca	ggaggttca	ttgtgatcag	gagtttggga	ccccagctcg	56940
gccaatacaa	tgatgtccct	gccttacca	aaaaattaaa	aacaatttgc	caggccttgg	57000
ggcatacact	tgtgtctca	gtatcaag	ggaggctgag	gtggggaggat	cacttgagtc	57060
caggagttt	tggtgggggt	gagctctgt	tgtaaccact	cactccaggc	tgggtaaagag	57120
aatgagact	tgttttattt	tattttattt	tttagtggag	tctcaact	ctgttttttt	57180
tacccaggc	ttggatgtcag	ttgttca	tcggttact	gcaacccctg	cttcccggt	57240
tcaagagatt	ctcccgctc	agcctccaa	gtagctggg	ttacaggcgt	gtgccaccac	57300
gcctggctaa	ttttttgtt	tttttagtga	gacagggtt	caccatgtt	gccaggctgg	57360
acttgaactc	ctgcacccat	gtatgttgc	cccccttgg	ctcctaaatt	gttgggattt	57420
taggcattag	ccacccggcc	tggccgagac	cctgtttttaa	aaaaattaaat	caggctca	57480
gggggtggctc	acgcctgtaa	tcccgact	ttggggaggcc	gaggaggggc	gatcacctga	57540
gatcgggagt	ttcgagaccag	cctgacca	atgcagaaac	cccgctctca	ctaaaaataac	57600
aaaatttagca	gggcgtgtt	gfcacac	gtatccca	ctacttggg	ggctgaggca	57660
ggagaatcac	ttggacccat	gaa	ggcgttggag	ccgagat	accattccac	57720
tccagcctgg	gaaaaaagag	tgaaaactca	tctcaaaata	taataataat	aataataata	57780
ataataataa	taataatgg	tagatgtt	ggtagaggca	ggacagtggg	agtgacagcc	57840
tggctctgtc	ctcccccgtt	ctgggttgc	atgtgttgc	ttggctca	ggcttgggggg	57900
gcacagctgt	ctgtgtggcc	tgtgttgc	cgggaa	ggccttggcc	gacttctg	57960
ctagaatcag	ctcccccac	tcccccagg	cccatgcagg	gctgtggcc	taccaggcct	58020
cctcgcgggt	gcacagagc	cagcgctg	acagcaac	agacgcatt	cacaggctt	58080
cgcccacag	gcgcgttatt	tacaggggg	tgagccgtt	atctgtgc	gttgcata	58140
tggacagcat	ttctgttccc	atacagc	ctggggcc	ccatatttac	tcccttgc	58200
agaaggaaaa	cgaggctcg	gagggat	gggatggc	tgggttca	caactttaat	58260
gtatgaccctg	ggcaggaa	tcccagg	gcaacagaga	ctcaggcc	gctgggcat	58320
gtggcttca	cttgcattca	cageggat	ggggccat	ggggccat	cactcgagcc	58380
caggagctca	ggggccat	ggggatata	gtgaga	acc	atctctaaa	58440
taggcccgtc	gggggtgact	acgcctgtaa	tcccgact	ttggggagg	gaggccgg	58500
gatcagcagg	tcaggagatt	gagaccat	tggcttac	ggtaa	accc	58560
aaaattacaa	aaaataggcc	ggggctgg	gtctcat	gtatcc	ccacttgg	58620
gggtggggcc	ggcggtat	gagg	ttcgagat	agcct	gacaa	58680
accctgttcc	tactaaaaat	aaaaattt	gccccgg	gtgac	ccgc	58740
cagctactca	ggaggctgag	gcaggaga	cgcttga	acc	atcgagg	58800
gagcccgat	gcactact	cacttc	tggagac	agt	gagact	58860
aaaaaaaaaa	aaattagcc	ggcgtgtt	ggcgttgc	ccctt	tcccttca	58920
gctgaggtag	gagaatggc	tgaacccgg	aggccgg	act	tcgg	58980
ccactgcact	aaagctggg	caacagagc	agacttata	tca	aaaaaaat	59040
ttagccggcc	atggtgttgc	atgtgttgc	ccccagca	ttggga	agggtggg	59100
gattgtttaa	ggccagg	ttgggggg	atgtgttgc	gtatgc	actgcattca	59160
gcctgggtt	cgagtgaga	ccccgtctt	ccctatccc	cgcaaaaaaa	aaaaaacaa	59220
tcaggggtt	cccacccctg	tccttgcaca	gatga	aaac	cgaa	59280
tat	tttgcoc	ccaggcccc	aatctgttgc	tttggact	actgttgc	59340
ctcaggccta	aaaaaaat	gttggagg	gtgttgc	gcaatgttgc	ggtgaga	59400
gggtccccc	ggggggcc	cgaggaa	ccaggaa	gcttggg	gggggtgt	59460
gcagccaa	ggagcaggcc	cgac	ccgtt	gttgc	tttgc	59520
ctgggttgtt	aggacagg	ccgactt	actcagg	ctgat	ccat	59580
gcagccctt	ctgggttgc	cacttccc	tgcagg	actgttgc	gtccat	59640
cagccctgg	ggagctcagg	actggctc	cctctgttca	ccgccttgc	ccaaatgttgc	59700
tcagatggca	cttgggttcc	cgaagg	ccagg	ccat	ccatcc	59760
cccccgggg	cttgggttgg	ggctgttgc	aaggat	cat	tgactgtt	59820
tgtcttgc	ccggacaca	ggtgc	ccat	gtaag	ccat	59880
ggcatgttgc	cgagg	tgttgc	aaatgttact	ctgttgc	ccacgg	59940
tcactctgt	aatcccagta	cttgggagg	cccagg	ccat	cat	60000
taa	ggcc	gggg	gggg	gggg	gggg	60060
ccgggggtt	ttgggtat	ctgttac	atgttca	gggttgc	ggggccat	60120
gttgc	gggg	gggg	gggg	gggg	gggg	60180
ggggccacaga	gaaagactcc	gtctt	aaaaat	gtact	ctgttgc	60240
cccccgcc	gagcttcat	gccat	cttgc	tgc	ccat	60300
ggcagcgt	gtggaccagg	gtctact	atc	acttgc	atc	60360
aacacatctt	acctgtact	gttgc	atc	ccat	gagg	60420
attcaagcca	aggcattt	tatgttca	tat	tat	tat	60480
atctatctat	ctatctat	gtatcttgc	tatgttgc	ccat	ttt	60540
agcgatctt	ctgttgc	ctccaaat	gttgg	ttt	ttt	60600
tggcttgc	gtgttgc	atgttgc	actgttgc	ccat	ttt	60660
aatccctt	cttcgtt	ctgttgc	atgttgc	ccat	ttt	60720
ctaacatctt	attttttca	tgatgtat	tttgc	tttgc	tttgc	60780
tgttgc	gggttgc	ctgttgc	gggg	tttgc	tttgc	60840
ccacaacgtc	agtgttgc	gggggggtt	ccagg	gggg	gggg	60900
ccctgggg	aaacccattt	gaagcaggat	ggca	ggca	ggca	60960
cactgaa	ccctggctgg	gattgc	ccagg	ccagg	ccagg	61020
atcccagttt	atcagc	ggcgttgc	gggg	gggg	gggg	61080

-continued

ccttgtgggc	agccgtgttgt	catgtggca	ttccaggggc	cttttgcac	gtggggaaatg	61140
tccaggaaaa	gcctcagcct	tcggtagggc	gcagaaaagg	gaagtgtccc	tagaggggggt	61200
gggttaggggc	gttggggaggt	gtgtctcgac	ggaaatgtcc	cttgggggg	ggaggatggat	61260
gggttgggat	tctgaggatg	gggggggggg	ctgtaggacc	caccatgtcc	ctccctgtgt	61320
accaggctcg	agtccccatg	aattgggggt	tgggggggg	aggggacactg	gcctggggac	61380
cagagacactg	ggctggctcg	gctcacagtt	gtctggaccc	tgtgatecgt	gtcaaaaaac	61440
gagaacaccca	attctgttcc	tgccttacta	ccaccagggt	ggacctgttac	cttgccatcg	61500
ccaggattgg	aatgttcggc	ctactgtca	accacttcc	cgaggactca	tttttggccac	61560
ccggggccgg	tgccctgggg	acacgggggg	gtctgggttgc	tttttggccac	agccgtgtgt	61620
gggctgaaggc	ctcaaggaca	cacatctcg	cataggaggg	ccaggctgca	gggcctcgga	61680
gacaacactg	ttggggctgt	tggggctctgt	gggtcccaagg	ttctggccctc	accggggctcc	61740
caccggcggt	tcagttccca	gcctttttcc	ctgttgc	tttttttttttgc	tgttttttttttgc	61800
tgtgtcccaa	ggccacccccc	tccaggcage	cttcagaaaa	tttttttttttgc	tttttttttttgc	61860
tctgcacccgt	ggtttttttgc	gtctccatcg	tgacccatcc	tttttttttttgc	tttttttttttgc	61920
agcgaggctt	gggggtgggg	tgtccggaggt	ggggggacata	tttttttttttgc	tttttttttttgc	61980
cgccggggagc	cgggggttgt	agggggtatgt	tttttttttttgc	tttttttttttgc	tttttttttttgc	62040
gaggacacccg	gttgggggttag	agcacccggcg	ggggcagcagc	tttttttttttgc	tttttttttttgc	62100
cttcggccccc	ggggccctcc	cgccggggcc	cagggggggg	tttttttttttgc	tttttttttttgc	62160
ggcgctgacg	tcacccatcg	tataaaatgt	tttttttttttgc	tttttttttttgc	tttttttttttgc	62214

-continued

aacttctgac	ctcggtatcc	actcgcccta	gcctccccaa	gtgctggat	tacagggttg	3060
agccacccga	ctggccacc	aaaccatact	ttaaagtaaa	taatagcac	tttattgtac	3120
tgaagggttg	atattttgac	aatttaaagt	tctggcttc	atgggttctt	gttgcgttt	3180
acttattgtc	gtatctaata	tctcccttt	tttttctt	ttttagacag	agtctcagtc	3240
tgtcaccagg	gctggagtgc	agtggcgaa	tttcagctca	ccgcacccctc	cggcccttgg	3300
gttcaaggac	ttctgctca	gtctctgaa	tagcttgat	tacaggccc	cggccaccacg	3360
cctggctatt	tatattttca	atagagaagg	ggtttcatc	tattggccag	actgggtctca	3420
aacttctgac	ctcggtatcc	acccgctca	acctctgaa	gagctggat	tacagacgtg	3480
agccacccg	ccctgttag	tatctcattt	tatgcctgt	tatctgttag	ctggacagta	3540
tatttggaaa	atttcgggtg	gcatgttgc	tcaggctgg	tgcggtagct	cacgcctgt	3600
atcccaccc	tttcggggc	caagggtgg	agatcatgag	gtcagaagg	tgagaccagc	3660
ctggccaaata	taatgaaacc	ccgttctac	ttaaaatatac	aaaaatcagc	cgggctgggt	3720
ggcgccgc	tgtatccca	gctactaggg	aggctgagcc	aggagaatca	cttgaaccca	3780
ggaggaggag	gttgcgtt	ttcaaggat	cataccattt	cactccagcc	taggaaccaa	3840
gcaaatctcc	atctcaaaaa	aaaaaaaaat	tattatttgc	aggagtgaac	aggacaagaa	3900
tgagggtatc	tatctctaag	aaggattat	gttgcctgt	gtctggatda	ctaccaaatc	3960
atctgtatc	ggtttcagaa	ggctgaaagt	acctcaggctc	atccagat	aatataaccc	4020
ggtgtcaagt	ccacttgaga	actgttttac	tttttttgt	ttgtttttt	ttgtttttt	4080
gagacggagt	ctcaacttgc	tcacccaggc	tggagtgcgg	tggcacgatc	tcggctcaet	4140
acaatctcg	cctccgggt	tcaaaaggat	cccctgcctc	acccctccga	atagcttagga	4200
ttacaggcat	gtgcatcaaa	gcacagctaa	tttttgcattt	tttaatagag	acagggttcc	4260
accatgttgg	ccaggctggc	cttgaactt	tgacgtcagg	tgatctgcct	gcctccggct	4320
ccaaatgtac	ttggatatac	gggatgagcc	accacgcctg	gcccggaaat	ggtttactt	4380
tgattcacct	ttatcttaaa	gggttataat	tttggatcc	caattttagg	tgctgtgg	4440
gagaattccc	accatttagt	agccttaggt	tttgactttt	gtccagtcct	ttgaggctge	4500
tgaaggactt	ctaaaatgtc	agctcagggt	tccaaaggact	tcttccagat	tggcaaatgc	4560
cttaaaggaa	aaaggcgtc	tgactgttag	gtttactct	cacttagact	4620	
tttgggtttagt	tcttccttat	tagctgtta	gtctttttgt	gcctttaaga	gaatgacggg	4680
ctggccatgg	ttggctacgc	ctgtatcc	aaacatttgg	gaggctgagg	tggggaggat	4740
atttgagacc	agccctggca	acatcagaag	aacctgcctt	tgcacaaaat	aaactcaaaa	4800
aatcaggcc	gtgtgttgcg	acacacctgt	ggtctcagat	tcttctgggg	ctgagggtgg	4860
agaatcacaa	gctctgggg	tttggatgtc	tttggatgtc	tttggatgtc	tagacttca	4920
gcttgggtga	cagagcaaga	cttgggtc	aaaaaaactat	atataatata	ataattttt	4980
ttaactgt	ttttttttt	agggtggat	tttgcgttgc	cgcccaggct	ggagtgcagt	5040
ggcatgatct	ctgcgtact	caacccccc	ctcccccgg	caatgttgc	ttctgcctca	5100
gcttccccaa	gtatgtgg	ttaacggc	ctggccacac	acctggctaa	tttttgcatt	5160
ttaatagag	aggacattt	accatgttg	ccaggctgtt	ctcgaaactt	tgacctgagg	5220
cgatccaccc	gccttggc	cccaaagtgc	ttggattaca	gggtgtggcc	atcgcccca	5280
ccatcttca	tttgcgttta	gttgcgtt	tcagtaggtt	ttttctattt	tttttagag	5340
atgttctcgc	tcttgcgtcc	tttgcgtt	tttgcgttgc	aatcatagct	cactgcagct	5400
caagcgaact	ccccaccc	ggccctccaa	agtgctgaga	ttacagggtt	gatctccag	5460
attttatttt	ttaaaatata	gggtgaccgt	cacccatggc	atggccctca	tgaggctat	5520
tgcctcagg	ttggatcc	ttttttttt	ttttttttt	tttgcgttgc	tttgcgttgc	5580
agtctcgtc	ttgtcacc	gttggatgt	agtggcgaa	tctcgttca	ctgcaagctc	5640
cactcttgc	gttcacgc	tttccccc	tcagcctccc	caatgttgc	gactatagat	5700
gccggccacc	acgcctggc	aatttttttt	tgtatttttt	gttagagacgg	ggtttccacgg	5760
tgtatgcac	gtatgttgc	atctgcctt	gtgtatccgc	ttccctccgg	tcccaaaatgt	5820
ctaggattac	aggcgtgac	cacccggccc	ggccaaatgt	atccagat	tctaatgtc	5880
catttctagg	agtgtaaat	cttataattt	ttgtctcatc	agggttttca	ttgttgcatt	5940
aaggctgtt	acctgc	acac	ctgtttgt	acccaaatgt	ctttttttt	6000
aagttagaaa	agacatgtt	atgttgcgtt	ttgtgtttaa	tcatttcgac	agttttactt	6060
gtgttcttgc	ttttttttt	tttgcgttgc	tttgcgttgc	atgtatattt	ttttttttt	6120
tctgttaat	ggccctagcc	tttgcgtt	tttgcgttgc	ataacgttac	tacccaaat	6180
ataaataaaat	agtcaact	atctactgt	ttttttttt	tttgcgttgc	tttgcgttgc	6240
aggcaaggta	aaagtggca	agtgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	6300
gcgattctga	tcaggagg	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	6360
atgaggccgg	gtgtgttgc	tcacacccgt	aatccacca	tttgcgttgc	tttgcgttgc	6420
cagatcacaa	ggtcaggagg	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	6480
actaaaaata	aaaaatata	ggccaggcc	tttgcgttgc	tttgcgttgc	tttgcgttgc	6540
ggaggctgt	cgaggaggat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	6600
cacggccact	tcactcc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	6660
aaaaaaaagaat	gggtcaggat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	6720
gcagagagga	cattctcatt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	6780
tggagctagg	ttatgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	6840
atctctgttgc	aatctgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	6900
ggggcaatac	cagactaaga	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	6960
tcagaggat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	7020
acctaagagg	aacgtggctc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	7080
tttggaggt	agagggtggc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	7140
catggagaaa	cctttctct	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	7200
tgtatccct	gttacttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	7260
gttgcgttgc	gctgagat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	7320
atctcaaaaa	aaaagagagg	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	7380
gtcccccttgc	acacaccc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	7440
cactttac	aatccagcc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	7500
agtgggttgc	catagattt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	7560

-continued

tgcactactt	taaaattttc	ctctcaattt	cttgagtcc	ggagttcgag	gttacggtga	7620
actatgattt	caccatgc	gtccagctga	ggcaacacag	agagacccctc	ttaaagaaaa	7680
aaaaaaactt	ttcccttc	actctgtccc	tagcattgac	tcaccccatc	atttttttt	7740
tttttcttga	gatggagttt	caaaaaaaaa	ctccaggctg	gaggcagttg	tatggtctcg	7800
gctcaactga	acctccgcct	ccccgggttca	agcgattctc	cctgcctcag	cctcccaaata	7860
agotgggatt	acaggcaccc	gccaccacac	ccagctaattt	ttttgtattt	tttagcagaga	7920
cggggcttct	ccatattagc	caagctggc	ttgaaactct	gacttcaga	gttccaccca	7980
ccttgcgcct	ccaaagttact	gggttacagc	gcacgagcca	ccgcgeccgg	ccttcaccta	8040
ttatatatctat	tcttgtcttt	taaatagtga	gtgacttcac	agctcagttt	ctttgttctg	8100
aacttccattt	cttgatcgca	acccaggatc	accttaggaa	gggcgggatc	ccgttatcc	8160
gtcccttctt	atcaactact	gagacatgtt	ctcggtgtt	gacggaggcc	8220	
ccttcagagt	cagcagagcc	gagctacag	cgcacatcag	cgacagcaga	ccagaggcag	8280
ggaagacaca	ctggggcagg	gccagtgccc	acatctggac	cacctcgcc	ggctctcgag	8340
ggaaaccgtga	ctacaggc	gttggtgatg	gtccccctgc	gaggctgc	acccaaggtt	8400
ttgttgggtt	ggaaacgtat	gacagcacat	acactaagaa	ggtggaaacc	gggggctggg	8460
tgcagtggc	catccagca	ttaggqagcc	caaggcaggt	ggatcacct	8520	
aggtcaggag	ttcgagacca	gcctggccaa	tatcgtaaaa	ccccgtctct	actttaaaatt	8580
gaattttaaa	aaactggggg	ccaggccccgg	ttggctcatgc	ctgtatccc	agcactttgg	8640
gaggccaaagg	cggtggatc	atgggttgc	gagatcgaga	ccatcttgc	caacatggg	8700
aaatcccgct	tctactaaaa	atccaaaaaa	tagccagac	ttgttgggg	cacctgtatg	8760
ccagctactc	aggaggctaa	ggcaggagaa	tcgcttgcac	ccgggaggcc	gagatcacgg	8820
tgagcagaga	tagccactt	gtatccago	ctggcgacag	agcgagactg	tatctcaaaa	8880
acaacaaacaa	agcaaaacac	tgggaatgg	tttgcaccc	accttgcagg	gtctgtgg	8940
agctgggcag	gccccaaatgt	gttttgcac	gttaaggaaa	aatccatc	tgcgtccgg	9000
gtcacagagt	ggggggccag	ggggagttcc	tgctgaggcc	agggcettgc	caggettccg	9060
tccctatcca	gtgccaagg	ttggagctcc	caggctgtca	gcatctcc	atgaggccag	9120
aaggggagg	cttaaaatca	agtcacatc	ccagcgatgg	agtcctatgg	ccacaccgtc	9180
ttggaaatgt	ggagcgcctc	tcggccgt	ctgtgcac	ctccaaatgt	gaagtgcac	9240
ccttgcgtgt	gatcttctg	ccctcccaa	ttttgcattt	tgcacattaa	agtttactt	9300
ttaattatgt	tttaatttgg	agaattttaa	aaatgtactt	aatgtat	ttggctgtc	9360
ttctcttcc	gtatccatcc	gttattctc	tgaaacttcc	tcctttaac	tcccgatca	9420
aataatttt	aggcttctgt	agtcgttcc	tgcgcagcc	aatcagcga	ttcagcagtc	9480
ctcaacctag	agaagtcac	aatctgtgg	gcaggagact	ccttacttgc	gtaaatgttc	9540
tttgggtctg	gacagagaga	gggcagggtg	tttcttggg	aaaggatattc	agctgcctg	9600
gcagggtccc	tccaggatgt	gtatattcc	tcatctgt	tcaggaaatc	gactctacta	9660
taaagctata	ttgtatctata	aaaccagact	cctggccagg	cgccgtgg	cacgcctgt	9720
atccccagac	tttgggaggc	cgagacgggt	gatcacgagg	tcaggatgtc	aagaccagcc	9780
tggccaaat	ggtggaaac	catctctact	aaaaatacaa	aaaaatttgc	ggggcgtgg	9840
ggcacgcggc	tgtaatccca	gctatccag	aggctgaggc	agagaattgc	ttaaaacctgg	9900
agaggccag	gttgcagtga	gctgagatc	caccactgc	ctccagoc	ggcgcacagag	9960
ttagacttcg	tctcaaaaa	taataataa	aaaataaaaa	aataaaaaaa	caaactccag	10020
tgtgttaatt	tatgtatgaa	atatatgtca	ggttagactgg	tcctttggcc	ttacttagtt	10080
attttcttca	tgtaatccata	ctgttgcgg	agtttgcatt	aatgtttaa	gcaatgtctat	10140
aaatccaaatg	gagaggatgt	agtttgcata	gggatggat	tttttaccaa	atcaagtttac	10200
ccctccacc	ataagcttgc	ctcgtctt	ccccaaatca	tttttgcata	gtgattctgt	10260
attgtttcc	ctgttcaact	tgatatttcc	tccagcccc	aatggcaca	tcttagatgt	10320
cacttagat	atcgctctgg	tttccattat	atccagcc	tttagatcc	tttgcgtttt	10380
acattactt	gttcttc	tctgttgc	ctgttgcata	gcccattac	caggtaggaa	10440
gtaggggag	tctccaaaca	ccatcttcc	ccaaatgtgc	acccagotta	ttcaaatctg	10500
tgtggcgtg	gttccaggat	acagatctaa	ataaaatgtc	caatgtgc	aattataagat	10560
ggccctatgt	tagaagggtt	ctgttgcgg	tttttgcatt	ttatgtacta	gagggtgtgt	10620
aacatgggg	aaccttacag	gtcaaggaca	gaaatcccg	gaatgcagac	ccaggaaacc	10680
ttagacggc	tctgcctaca	accttgcaga	caccctgtc	atgcagcagc	tgaggctgt	10740
tcaggggg	gggactcagg	gccttgcatt	agcacttcc	gccccgtt	tttaccctgg	10800
gagcaaggct	atgttgggg	gaatgtat	tccttccca	ctgggtgtcg	gggcctccgg	10860
atccatgtaa	cgaaaaacac	caatagctgt	ggcttgcct	tetacatcat	aaggccct	10920
ttgattttcg	tataaaaaac	ccatgtaa	aatggatat	catatccagg	caaagaaaa	10980
gtaaaggtag	ggaagggggg	aggggacago	aacattaccc	cgcttgcac	atagggagca	11040
acaaaatgc	ggcacgcggc	agttggca	aggctcagt	ctagcagg	gtactctgg	11100
atttttagt	tctcgcctt	tgccagcc	gaagactgc	acctatcc	ttgcacccct	11160
atcatgcata	cagcaagact	gtatgcgg	cttgggtgg	gaggctggg	aatcatcatt	11220
taacaatata	tataatagac	atttacagg	ctgttggat	ctcaaaaagg	cctgttaat	11280
cctggagaaat	ttaggaat	ttcaccagaa	aatgtatgt	taaaactggat	ctgaatggat	11340
taggcattc	ccaggagac	agttggaa	tagcagca	ttgggtggat	cattccaggc	11400
gaagaaatgt	ctccaaac	cttggagcc	gaaagcaatt	tttggatgg	gagaacacga	11460
tttttgcgtt	gggcgtatgt	caatggatc	tgaggat	gaaatgggt	aatgtatcc	11520
aatgcacaa	ccaagaattt	ttacccat	tttgcgtt	gacacacagc	cagggcagaa	11580
tgaatggc	ccttcttgc	tttttgcata	tctttctgt	acagaagatc	actggctaca	11640
acattttgt	tttttcttca	tccaaaga	ccaaggccca	tttgcgtt	tttgcgtt	11700
taagttata	gggttgcacgt	tttgcgtt	ttatgtat	tttttgcatt	tttgcgtt	11760
ttagatttgc	atgcacccat	cacccaa	gtatgcac	taccctt	tttgcgtt	11820
atccctcgcc	ccccccccc	ccttcttcc	agttcccaaa	aatccactgt	atcatcc	11880
agccctcgcc	ttctcatagc	tttgcgtt	catatcgt	agaacata	atgtttgatt	11940
ttccatcc	gagttactt	acttagaata	tttgcgtt	atctcatcc	tttgcgtt	12000
aatgcattt	atttccat	ttttatgg	tttgcgtt	tttgcgtt	tttgcgtt	12060
ccacagatc	ttttttttt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	12120

- continued

-continued

```
SEQ ID NO: 4          moltype = DNA    length = 27888
FEATURE              Location/Qualifiers
source               1..27888
                     mol_type = genomic DNA
                     organism = Homo sapiens
```

```

SEQUENCE: 4
gcaccacccac gccccagctaa tttttgtat ttttagtaga gacgagggttt catcatgttg 60
gctaggctgg tctccaaacgc ctgacacctag gtgatcccccc cgcctttggcc tcccaaagtgg 120
cttagaaatgt aagcgcgagc cccgcacatt ggcccttatt ccacaatatt cataaaatgtt 180
cactactaga ttatccctttt actagactac ttcttacccct tcttagttagta cctttactact 240
agattaccct gagtccacca cataatttgtt ttcttcctcc atccctttt tcttcaaaca 300
ttcaccttat cttatgtaaa aatgttagatt taggccacccctt aggaggggccgg atcataaagg 360
caggagatcg agaccatctt ggcaccaatgtt gttaaaccctt gtctctgtca aaataaaaaaa 420
atttagccga gtgtgggtggcc ttccctgttag tcttagactat ttggggaggctt gaggccggg 480
aatcgcttga accccggggc agagggttgcgtt gtagccggagc agcaccatctg cactccggcc 540

```

-continued

tggcaacaga	gcaagactcc	gtctcaaaaa	acaaaacgtt	gatttcctga	gcatttagcta	600
aagtcttgc	ctcaatcttt	cccacccca	ctttcaagg	aaaatgtata	aatactaaaa	660
acctcctgag	aaccttttg	gaagaaacag	ccacaaaagt	gtctgtaca	tggtttagtc	720
agtttagtca	ctcatttcaa	ttaacactat	ctcttatcca	ccccctccac	tgcaagctgaa	780
ggaaatctt	ctgaagtcg	atcatgtcac	tgttattcg	ggtttaaaag	ctggggaaaga	840
aacagccaca	aaagtgtctg	tgacatgg	tagtcagtt	agtcaactcat	ttcaattaac	900
actatcttt	atccatccct	tecaactcg	ctgaaggaa	tcttctgaa	gtctgatcat	960
gtcaactgtta	ttcgtgttt	aaaagctgg	gctaaagaaa	gaaagaaagt	aaaattttaaa	1020
aagaaaagaa	aaataaaaaaa	taaaaataaa	gctgaggctc	tgtatcaat	ttactgagtt	1080
aaaactctc	actgtcagg	ccggccgcgg	tggttcatgt	ctgtaatccc	agcactttgg	1140
gaggctgagg	aggggcagat	acctggagg	aggagttcg	gaccgcctt	acgtggagaa	1200
acccatctc	tactaaaaat	acaaaatagg	ccacgtgtt	ttggcgatgt	ctgtatccc	1260
agctactctg	gaggctgagg	caggtgaatc	gcttgaatca	gggaggcaga	gggttgcagt	1320
agctgatata	ggccatccat	actccagct	gggcaacaag	agcgaacatc	tgtctcaaaa	1380
aaaacacacac	acacacacac	acaaaactc	ctcaactgt	tttttaggtt	gaacttggc	1440
aagtctaaa	ctcatttac	cttactttt	tcaactgt	aatggaaata	atattaatac	1500
tttattgtga	ttatctatgt	agttcagtt	agtactttagt	tcttattata	aatttagaaat	1560
gttagctgt	atagtataaa	ctcttccat	cttcacactt	ataacaccca	gaacacagt	1620
tagttgttca	aaggcctttag	actttttttt	ttaagatgg	atttcggett	tcttgcccg	1680
gcaggagtg	aatggcacga	cttcgtctca	ctgcacccct	caccccttgc	gttcaagcaa	1740
ttctcctgcc	tcagcctccc	aagtagctgg	gattacagg	gcccaccc	acgcccggct	1800
aatttttgtt	attttttagt	gagacgggt	ttcagatgt	cagccagg	ggtctcaac	1860
tcctgacactc	agttgtatca	cctgcctcg	cctccaaat	tgctggatt	acaggagtg	1920
gcaactgagc	ccggccgcct	tcagactt	gtttgaacct	ggaaaaactt	ttccctggag	1980
aaaacatgtc	tttcagcatt	actttctct	cggtgtttc	ttacctttac	tcattccag	2040
gatttacaat	cccccttctt	tggttccaa	caatagttt	tatagccaa	aaaatcttt	2100
aagcaatttt	agttttat	tttcttctt	gctgttgc	gtgactttt	ggagagtagg	2160
atgtgggtga	catctgtct	tggttccca	aagtccaaac	aaagcatctg	atataaaaata	2220
ggagctca	caatgttgc	gtatcatgt	tctgttctg	tcccccagct	2280	
cctcaactaac	cccccttgc	tttatagtgg	gataatgaat	gaggaaatgc	tgtctcatgc	2340
tcattgttgc	tctcaat	ttttttat	catactt	tgcatagacg	tgatttttt	2400
gttttttct	tctttgc	agggtata	gttagttgc	attttctt	ttttttat	2460
ttatgttaact	tttattctt	tttacaggg	agccacccca	gccagcagcc	agcgttaggt	2520
cagagagact	cccccttctt	tcttgcagg	gcaagggcg	catgaaagg	cagagctaa	2580
taatgttgc	tttttttttgc	ttggggaggct	aaaccaatac	acgtacata	cagtcaatct	2640
tcagccccaa	gaaaatggcc	accttattgt	ttcaccat	ggcaggatt	gggctggca	2700
cttgagtttta	ttttttcaat	tcaaaacaa	ggccaggac	ggtggctcat	gcctgtatc	2760
ccagcactt	ggggggctaa	ggccggcaga	tcacaagg	aggagttt	gaccagctg	2820
gccaacatgg	agaaatccca	tcttactaa	aaatacaaa	attagctgg	tttcgtgtg	2880
tgcacccgt	gtccctgt	ctcagggg	tgaggcagg	gaatcttctg	aaccaggag	2940
ggcgagggt	caatgttgc	agatcgcc	actgcactt	agccctgat	acagagtg	3000
actccgtct	aaaaaaaaaa	acaaaagccg	ggcatgg	ctcaagctg	taatccagc	3060
actttgggg	gcaaaatgtt	gcatgtat	tttgcgtat	ttccagacca	gcctgtatc	3120
catgttgaaa	acetttttct	actaaaaata	caaaaatag	ttggggctgg	ttggcaggcac	3180
tttcaatccc	agctactctg	gagggttgc	caggagaatc	acttgaaccc	gggaggcaga	3240
ggtttttagt	agcttagatc	ccggccact	actccaaat	ggcaacaga	gcaagactt	3300
gttcaaaaaa	aaaaaaaaaa	gaaaatggaa	aaaaagaaaa	gaaaatggaa	ctaaaagaga	3360
gttggagcaca	tttggcatgt	gataatgt	ccagctactt	gttggctgt	ggtgagagga	3420
tcattgttgc	cttggagttt	gagcccgcc	ttggcgtgc	tgcagaacc	agtctctaa	3480
caaacaataga	aaaaaaaaaa	ggaaaaaaa	agagttatcc	ttgccttag	aatatgttac	3540
agtctatccc	ttttagatca	taaaatgtt	ataactggac	agagtagtca	aaaaagaga	3600
aatctgttgc	acaatgttca	gttggatgt	gatgttctc	aaaggacat	aaagatttca	3660
gaaccatgg	gggggggtt	tcttgcagg	tctgtat	acctcttgc	cttattttat	3720
atataattcca	tttgcgtctc	tttattgt	caccatttc	atgggat	ttccccttgc	3780
tttattttata	tttgcgtat	caatgttgc	acatgtactt	gcatgtat	acttaataaa	3840
ttgttataat	tttgcgtat	tttttttttgc	tttttttttgc	tttttttttgc	tacgttttgc	3900
gaagttgact	tttgcgtat	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	3960
gagcaattgg	ggggccacgg	caatgttca	gacaagagaa	gtctcagaac	tgtatatgt	4020
tcattgttgc	ggggggatgg	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4080
attggatata	ggatgttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4140
tgttcaatcc	tttgcgtat	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4200
tgtttaacac	ggccggggat	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4260
gcaggaggat	catctgtat	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4320
gttccacta	aaaacacaaa	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4380
acttgggg	cttggggat	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4440
aagatggcac	cactgttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4500
aaaaaaaaat	ggccgggtgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4560
gcaggtat	tgttgcgtat	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4620
gttctgtat	aaatataaaaa	aaaaaaaaaa	aaatataac	gggttgcgtat	tttttttttgc	4680
taatccca	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4740
gcaatgttgc	aaatataaaaa	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4800
aaaaaaaaaa	aaaaagatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4860
ctgactatcc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4920
attcaatgtt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	4980
cttaacatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	5040
gaaaatataat	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	5100

-continued

gcatgggtgt	gcatgcctgt	aaagccagca	cttgggaga	ccgagggtgg	cagatcactt	5160
gagcccaagg	atttgagata	aacctggaca	acatggcgaa	actccatctg	tgggggaaaa	5220
gcaagagaga	ttagatgttt	actgtgtctg	tgttagaaa	agtagacata	ggagactcca	5280
ttttgtctg	tactaagaaa	aattttctgt	ccttggattt	ctgttaactt	atgaccctac	5340
ccctaacc	gtgtctctg	gaacatgtgc	tgtgtcaaac	tcagggtaa	atggattaag	5400
ggctgtgcaa	gatgtgtttt	gttaaacaga	tgtttgaagg	cagcatgtctc	cttaagagtc	5460
atcaccaactc	cctaattctca	agtacccagg	gacacaaacaa	ctgcggaaagg	ctgcaggggac	5520
cttcgtctag	gaaaggccagg	tattgtccaa	ggtttctccc	catgtatag	tctgaaatatt	5580
ggcctgtgg	gaaggggaaag	acctgaccgt	cccccagccc	gacacccgta	aagggtctgt	5640
gtctgaggagg	attatataaa	gaggaaaggca	tgctcttctg	agttgagaca	agaggaaaggc	5700
atctgtctcc	tcggcgatccc	ttggcgatgtt	aatgttctgg	tataaaaccc	gattgtacgt	5760
tccatctact	gagataggga	aaaacccctt	taaggcttga	ggtgggacat	gccccccagca	5820
atgtctgtttt	gttaaagcatt	gagatgttta	tgtgtatgca	tatctaaaag	cacagcacatt	5880
gattcttac	cttgcgtat	atgcaaaagaa	cattttttcac	gtgttttgtt	gctgaccctc	5940
tcccccaaat	tgtctgtga	ccctgacaca	tccccctctc	cgagaaacac	ccacaaatgaa	6000
taaataataa	ctaaggaaac	tcagaggctg	gccccggatct	ccatatgtct	aacgtgtgtc	6060
ccctgggtcc	ccttattttt	ttctctatac	tttgttctctg	tgttttttctt	tttttccaagt	6120
ctctcggttcc	acctaacaac	aaacccccc	agggtgtggag	gggcaaccca	cccccttccacc	6180
atctccacaa	aaaataaaaaa	attagccca	tgtgtgtggg	tgtgtgtgt	gtcttaacta	6240
gtccagggc	tgagggtggg	ggatggctgt	agtccccagag	ctcgaggctg	cagtggccg	6300
agatggcacc	actgtactcc	agcctgggtt	acagagttag	cagagttaga	ccctgtctca	6360
aaaataataa	aaataaaaata	atgggaaacc	aaatcacaag	acagactca	ttttttctttt	6420
atttatttttta	tttttttttta	gacagagtgt	tgtctgtctg	cccgactgg	agtgcaggag	6480
catgggttcg	gtctactgca	accccttccat	cttgggttca	agtgttctt	gtgtctcagc	6540
ctcccaagta	gctggattta	caggatgtgt	ccaccatgccc	tgactgattt	ttgttatttt	6600
aatagagaca	gggttacgccc	atgttttca	gtctgggtt	gaactcttgg	tctcaaggga	6660
ttcacccatc	toggcctctc	aaagtagctgg	gattacaggc	tgagtccaca	ctaccgggtt	6720
tatTTTTCAT	acagatgatt	tagatgtt	ttatgtttaa	ttttttttaga	ttatctactaa	6780
aaagaaaaacag	ttttaattctt	caatttacaa	ggattcaatt	tttaaaatttta	tatttgaatg	6840
tacagaggcc	tacaaaatacc	aattttcgaaa	acactgtttt	caaggatatac	tattttgttaa	6900
aatttgaatgc	aactgtttttt	aacagattaa	ctgcaaaataa	caagcttaa	caaataagggg	6960
atttattttt	cccatatc	aaagagacag	ggccgggcac	agttgttccac	gcctgttaatc	7020
ccagcac	gggggggtcg	ggccgggtgg	tcacttgggg	ttggggac	aagaccagcc	7080
tgaccaacat	ggagaaaccc	cgtgtctacg	aaaatataca	aatttagccgg	acgtgtatgtt	7140
acatgcgtgt	aatcttcaact	actcgccgg	ctgaggaggc	agaatggctt	gaaccttggga	7200
gggtgggggtt	cgagtggcc	aagatcgccg	catttgccat	ccagcctggg	caacaagacg	7260
agtcgtct	aaaaaaaaaa	aaaaaaaaaa	aaaaagcaag	tagacgggc	cagggtggagt	7320
ggctcatgtt	tgtaatccca	gcattttggg	aggccggat	gagaggatca	tttgagctca	7380
ggaggttggg	accaggcttc	gcccgttgg	ggccatagg	gttccatcc	tctacaaaaaa	7440
caaacaaaaaa	atagccaaac	atgtacgac	atgcctgtat	tccctgtatc	ccaggaggct	7500
gagatgagag	aattgtattt	atctaggggg	ttttagggcc	agtggccat	gaatgcacca	7560
ctgcacttca	gcctgagca	cagagcaaga	ccctgcctca	acaaaaaaaa	agcaaaacaa	7620
aatgtataaa	agcaaaatgg	gcaggggcaca	gtgggttccac	ctgttaatcc	tagactttt	7680
ggaggccgg	gcaggccgg	tacttgcgtt	caggatgtt	aaacccagg	ggccaaatctg	7740
gtctaaacttc	gtctacta	aaaaatataca	aaaaatatacg	aggcggtgt	gtggacaccc	7800
gtatatccca	ctactcttca	ggctgaggca	ggataatcg	ttaaaccggg	gaggcgatca	7860
tgcctactca	ccccaaacttc	ggcttccat	caaggctctg	tctcaaaaaa	aaaaaaaaaa	7920
aaaagaagata	gcacatgttt	aactgtatgtt	aacttccata	gagacagc	caggccaaat	7980
gagagcttga	caggcacagg	gttacttgc	gccttactga	ggaaaaataa	ctagacatga	8040
gcaaaaggaga	tccttcagaaa	cctagaggca	aaatgttcatc	atatgcattt	ttcgttcaaa	8100
acttgcggg	aggaacatca	gcacccagg	gttttccat	gttttccat	tttttttttt	8160
ttttttttttt	atgtggatc	ttgttccat	cgccggatgt	gaatgtccat	ggcgccatct	8220
tggctacttca	ccaccccttc	ctccccgggtt	caagtgtttt	tcttgcctca	gccttcttgc	8280
tagctggat	tacagggtgt	cgccaccat	cctggctaat	ttttgttattt	tttagtagaga	8340
cagggtttttt	gocatgttgc	ccaggctgtt	ctcgtacttc	tgacccctt	tgatccaccc	8400
gactcagect	cccaacttc	taggtttttt	ggcggtggat	accacgc	gccagcattt	8460
tctaagatgt	tcaggaggg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8520
acatggcaaa	ggtgacaaag	ggccgttata	aaaaagaaat	aaaaacat	gtccccctgt	8580
aaaggggat	aaaaagaaat	tcaaaatgtt	cagtgttcc	aaaaggcc	ctgttagctt	8640
ttttgttcttc	ttttgttgcgt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8700
tgtatgttca	aaaaaaatgg	gagatgttgc	tttttttttt	tttttttttt	tttttttttt	8760
ctcacgectt	taatcccaat	actttttttt	gttttttttt	gttttttttt	gttttttttt	8820
tttcgagatca	tcctggccaa	catgttgc	tttttttttt	tttttttttt	tttttttttt	8880
ccactcggtt	ttgtgttcc	ctgttatccc	aaactactcg	ggggctgggg	caggagatcc	8940
gttttgcaccc	gagaggccgg	gggttgcgtt	agcaggatgt	gcacccat	actccaggct	9000
ggggcaataga	gtggatctt	gtctcaaaaa	aaatgttca	tttttttttt	tttttttttt	9060
gcaagtagcc	ttatgtaaa	aagggttgc	tttttttttt	tttttttttt	tttttttttt	9120
tttgcataatca	agctaaatgg	aagcttgcata	tttttttttt	tttttttttt	tttttttttt	9180
aaacgaaaat	gaaagaaatgg	tttttttttt	tttttttttt	tttttttttt	tttttttttt	9240
gatgttataat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	9300
ctgttataacaa	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	9360
gtctgtatcc	ccaaactttt	ggggggccaa	tttttttttt	tttttttttt	tttttttttt	9420
aagaccagcc	ttggccaaat	ggcggaaaccc	tttttttttt	tttttttttt	tttttttttt	9480
gcatgggtgt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	9540
gaatccgggaa	ttggccaaat	tttttttttt	tttttttttt	tttttttttt	tttttttttt	9600
gacagagcgca	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	9660

-continued

tgtgttagat ttgttttaa	actatacagt	gtctttttg	tatagttaa	acactactaa	9720
atgtgtctt acacatttc	tttttagttt	tctagagacg	gagtcttgc	ctgttgccca	9780
ggctggagtg cagtggacc	atcatagctc	actgcaccc	ggaattcttgc	ggctcaagca	9840
atctcttac ctcaggctcc	tgagtagctg	ggactataga	tggtgtcac	cacgccttgc	9900
taatttggaa aaaaattttt	tttttttga	gatggagttt	tgctcttgc	gcccggctg	9960
gagtccagt tcaccatctc	tgctcaactgc	aacctccgc	tccagagtc	aagaatctct	10020
cctgtctcaa cttctgaga	aattggggat	acaggagctc	gcaccaacac	caggctattt	10080
ttttgtatct ttagtagaga	tggggttca	ccatattggc	caggctgtc	tcgaactct	10140
gacctcgat ggtccgcctg	ccttggctc	ccaaagtgc	ggcaccacg	ccgcaccta	10200
aacattttt ttgttgccg	aggaggctg	gctcacac	gtatccccag	cactttggg	10260
ggtctggatg ggtggatct	gaggctcagg	gatggagacc	atctggctc	acacagtgaa	10320
accccgcttc tactaaaaat	acaaaaattt	agtttggct	gatggcaggc	gcctgtatgc	10380
ccagctactc gggaggctg	ggcaggagaa	tggcgtgaa	ccgggaggcg	gatctgcag	10440
tgagctgaga ttgcgcact	gcactccgc	ctgggtcaga	gatgtgaa	ctgtctcaaa	10500
aaaaaaaaaa aaattttttt	ttgttgttgc	gacagagact	ctttatgttgc	cccgacacttgc	10560
tcttcaattc ctgggtctaa	gcagtccctc	tccttgc	tcccaaaatgc	ctgtgactat	10620
agtttttga ttagaaaaaa	aaaaaaataca	agttaaaaac	tggttatgatc	actccacact	10680
gtatcaggaa ctca	gatccctc	aactccctgg	ttcaagcgt	tctccggcc	10740
cagcctccca agtagctggg	actacaggc	cgcaactgc	tgagcagct	attttgc	10800
atttagata gatgggtt	ccccatgttgc	gtcaggatgt	tctcgatctc	ctgacccat	10860
gatctgcggc ctcggcctc	ccaaaggctc	gggattacag	cgctgagtc	ggcctactt	10920
tattcttata ataaccctat	ggtcacaaaa	tgtctgtct	acttccaggc	tggacaacat	10980
agtaagaccc agtctccaaa	aaaaaaattt	ttaattatc	ttaaaaaat	aaaaataaaat	11040
gtatttaagt tgcttttac	ctgtttttt	ttaaagaga	tgaggcttca	ctatattg	11100
caggctggc ttgaaatcc	gggctcaaga	gattctccca	ccttggctc	ccaaatgtct	11160
gggattacag gtgtactg	ctactccag	cctaccta	ctttatttgc	acttttctt	11220
tctttttttt tgtagatggc	atcttgc	ttggaccagg	ccagactg	gtggcgaat	11280
ctca	tcgacttct	gcctgcagg	ttcaagcaat	tctccgtc	11340
agtagctggg atcacaggc	tgccacca	cgctggct	attttgc	tttttagtga	11400
gacagggtt ccgcattgttgc	gccaggctg	tctcttac	ctgacccat	gtggggac	11460
cttggcttcc agatgtctca	ggatgtctca	catgaggcc	cacgcctgc	cttttttttgc	11520
agacggatc ttgctgtc	ccccaggctg	gatgtc	gtcaatctt	ggttca	11580
aaacctctgc ctccccagg	caagcaatc	tccaaccc	gcctcccg	tagctgaa	11640
gataggtgc tgccgcact	cctggctat	gtat	tttttag	tagagac	11700
actggctagg ctgtctcg	actcttgc	tcaggatgt	cacccgc	gaccctcc	11760
agtgctggg ttataagc	ccgcacccgg	ctttttttt	tttttttgc	agatga	11820
gggtgtgatgt gttgtctat	gcctatgttgc	gcacgtt	cggggtgt	aggcgt	11880
atca	cttggcaaca	ttgtta	agccat	ccatct	11940
aaaaattttaa aattatgtat	gtgtgtgt	aggta	cttgc	acttgggagg	12000
ctgaggtggg aggtgtt	gaggtgttgc	gggtgagg	ttgtgt	gtgatcat	12060
ca	ctgcgtggc	aaacagat	gaccctgt	ccaaaaacaa	12120
gagagatggg gtctgtct	gtcatcttgc	ctggagg	gtggctgt	atgggtc	12180
ggccgccttaa actccccagc	ccaaatgtat	tctccgc	aaggc	cttgc	12240
attacaggag tatgtcc	tgccagg	gaat	tttgc	tttgc	12300
actcattttt tttttttaaa	tctgc	attttttt	tttttttgc	tttttttgc	12360
aggcagaatt tgctgttgc	gccagg	gtatgc	cacaat	tttgc	12420
acccgcctc ctgggttca	agcgttctc	ctgc	cccg	cttgc	12480
caggcatgtg tccatctc	cgatgtt	ttgtat	ttgtgt	tttttgc	12540
atgttggc	agatgtctc	gatcttgc	tttgc	tttgc	12600
agtgcttaga ttacaggcgt	gagccact	acc	ccgc	tttgc	12660
gatatacc	ttaatttca	gtatgttgc	ataatttgc	tttgc	12720
atatactatgt	gttttgc	at	aaat	tttgc	12780
tttagtgc	tttttttca	tttgc	tttgc	tttgc	12840
ttttttttt ctccctt	tttcc	caat	tttgc	tttgc	12900
gatatagtcc agacccaggc	agtaaaagg	acatgt	tttgc	tttgc	12960
aacttggaa actaaatgttgc	tttttat	ctatgt	tttgc	tttgc	13020
attctcagat tttttttt	tttgc	tttgc	tttgc	tttgc	13080
tcttaatctc aacactatttgc	gtat	tttgc	tttgc	tttgc	13140
cttatgttgc ataggatgttgc	tgatgttgc	tttgc	tttgc	tttgc	13200
ccccccacac acgggtgg	ctgttcc	tttgc	tttgc	tttgc	13260
aaaaaaatgtc	ctatgttgc	tttgc	tttgc	tttgc	13320
cctggctggc aggggtc	cttgc	tttgc	tttgc	tttgc	13380
tattcatacgt ttaacttgc	ctatccat	tttgc	tttgc	tttgc	13440
gcttgc	aaacatgtc	tttgc	tttgc	tttgc	13500
cactgtatc ttttttgc	tttgc	tttgc	tttgc	tttgc	13560
aca	tttgc	tttgc	tttgc	tttgc	13620
caagaatgtc tca	tttgc	tttgc	tttgc	tttgc	13680
atcttgc	tttgc	tttgc	tttgc	tttgc	13740
caactcaga	tttgc	tttgc	tttgc	tttgc	13800
cagatggacc ctatgttgc	tttgc	tttgc	tttgc	tttgc	13860
ttacgggttgc	tttgc	tttgc	tttgc	tttgc	13920
ccctaa	tttgc	tttgc	tttgc	tttgc	13980
tataaaataa aacaatttca	tttgc	tttgc	tttgc	tttgc	14040
tttcttctt aactttaaca	tttgc	tttgc	tttgc	tttgc	14100
aaacccctgc	tttgc	tttgc	tttgc	tttgc	14160
tagcttgaca	tttgc	tttgc	tttgc	tttgc	14220

-continued

ctccgtccgc	tctagccgta	aagcaactatt	tgcaatcccc	ggaggtgtatg	ttgcagatct	14280
cagtgccttc	cccaattctc	tccacttact	cttgccttaa	atccgtgttgc	aaaaaatgtgg	14340
gettaagtgt	taccaggctca	gtggtagctt	tccgtacaac	ctcccttattc	ccagtgctgat	14400
tttcagtgcc	ccttcctctg	tatcccaco	ataccctgt	catatctca	tcatatgttt	14460
acatagtcc	ataattgttt	ataattttatg	tttcttttcc	ccctttggac	catgtatctt	14520
gtgagtcaga	gttatgggtgg	ttgtaaatgt	ctcagagcat	tgaatagaaa	ccgatacaac	14580
cattgcattg	aactatgcat	tccttgagggg	caggccata	tcttcaccc	tgcataccaa	14640
catctagcat	ataccaggta	aatagtgggt	gttcagtgaa	tgaagaatgt	atacatgaat	14700
gcttcaggca	gaagggtgct	ctcatgtgg	tagtacaca	tttgggtgatc	ttttttgtt	14760
tgttggtttgc	cttgtttttt	gttgcgtgt	ttttttgtt	ttgagacaga	gtcttgctt	14820
gtcccacagg	atggagtcga	gttgcgtgt	ctcggtccat	tgcatactcc	gcctgccccg	14880
ttaaagctat	tccctgcct	cagectcgga	gtagctggga	ctacagggtt	gcccacccac	14940
acacagctaa	ttattgtatt	tttagtagag	acagtgtttc	accatgttgg	ccaggatgtt	15000
ctcaatctcc	tgacccctgt	atccacccaa	cttggccctt	caaagtgtca	ggattacagg	15060
catgagccac	cgcacccggc	ctttttttat	tataaagaa	agagtttgc	tcttgcacc	15120
caagtttggag	tgcaatgtgt	tgattttgc	tcactgcaac	ctctgcctcc	ttgggtcaag	15180
tgattctct	gcctcagcct	cccgagtago	tgggactaca	gacaccacc	accacaccca	15240
gtaaattttt	gtatTTTTAG	tagagacago	attttacat	attgaccagg	ctgggtctcg	15300
actccctgacc	tcagggtatc	cacccgttcc	ggccctccaa	agtgtggga	ttacagggtgt	15360
gagccacccgc	acctggccct	tttttttttgc	ttcagtgta	atcttttca	ttcagcttgc	15420
tgcccaaggct	ggagtgcaat	ggcgtgtatc	caatttactg	caacccocac	ctccgggtt	15480
caagcacttc	tcctgcctta	gtccctgtag	tagctggac	tatgggctgt	caccaccaag	15540
cccggtatg	ttttgttattt	tctgttagaaa	cgggggtttcg	ccatgttgg	cagactgtgc	15600
tcgaactct	gacccctgt	gacccacccgg	cctcagcttcc	ccaaagtgtt	gggattacag	15660
gaataagcc	ccatgtcg	ccatattttgg	tgcattttat	tttgcgtgt	atgcacgtgc	15720
aaagaggaca	tgttggagcc	aggagactga	tggaaatcttgc	cccttaccca	atttatttgg	15780
atctgttata	atgttacaa	tcccatat	atgttacagc	agaaaatgtc	gaacaggggc	15840
gggcgcgggt	gttcacccgt	atggatcccg	cacttttgg	ggccgaggcg	ggccgatcac	15900
aagggtcgaa	aatcaagacc	atccgtctaa	catgttgc	ccccgttctt	actaaaaaaat	15960
acaaaaaaaaat	agccggggct	cgtggcggg	gctgtgttc	ccagctactc	gggaggctga	16020
gcccggggaa	tggcggtgaa	cgaggaggcg	gaggttgcg	tgagccccaa	tcaccccaat	16080
gcactccagc	ctggccgg	gagcgagact	ccgcctcaaa	aaaaaaaaaa	aaaagaaaaaag	16140
aaaaagaaaaaag	aaaagaaaaat	gtcttacaga	tatttgcgt	attaatca	ctcaatttcc	16200
ccatTTTCA	agatctttagt	acctgttcc	ttcttacata	tttgcgtgt	tttgcgtgt	16260
ataaaaattaa	tataattttt	actatTTTTC	tttttttttttgc	tttgcgtgt	tttgcgtgt	16320
tcagccctc	tccaaacccgt	gccaacacaa	tttttttttgc	tttgcgtgt	tttgcgtgt	16380
acaggatgt	gtttttactt	tggaaacta	tttttttttgc	tttgcgtgt	tttgcgtgt	16440
tatcccttc	tttcctggag	gggggggggg	ggaatatacg	gtatTTTTCT	tttttttttt	16500
cttgcgtat	gttgcgttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	16560
cactgtcaacc	tccgccttc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	16620
gggattacag	gttgcgttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	16680
cgctctgtcg	cccaagactgg	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	16740
cctgggttca	ttccatcttc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	16800
caactatgtac	cggtcaattt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	16860
aggatgttct	cgaaacttctg	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	16920
attacaggca	ttagccacca	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	16980
tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17040
gttgcgttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17100
tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17160
tcaggctgtt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17220
tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17280
tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17340
ctgcaacc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17400
agctaattttt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17460
ggatctcttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17520
tgagccacca	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17580
ctgcctttag	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17640
tataaaactgg	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17700
actccagatgt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17760
aatgtcttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17820
gttccagggtt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17880
ctgaaatctt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	17940
cacaggtcaag	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	18000
aaagggttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	18060
tgtcttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	18120
tgaatgtacta	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	18180
ttgttgcgttca	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	18240
tccaggagtt	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	18300
aaaaattttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	18360
ctgagggtgg	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	18420
cattgttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	18480
aaaaaaaaaaaa	ggaaagcaaa	aaacccaaacc	aaaacaaacaa	aaaagataac	ataaaaaacaa	18540
aacaaaaggaa	agccaaaagc	caggcatgtt	ggctcacatc	tgtatccca	gcacttttgc	18600
aagccaaaggc	aagcagatca	cctgagggtca	gcagttcgag	accagccaga	ccaacatgg	18660
gaaaccctgt	ttataactaaa	aaaaatata	tatataatata	tacaaatata	aaaatttagcc	18720
aggcgtgggt	gtgcgttgc	gtatccctag	ctattctgg	ggcttgggg	ggagaatcgc	18780

-continued

ttgaaccgaa	ggggcagagg	ttgttgttag	ccaagatcg	actattgcac	tccagcctgg	18840
gcaacaagag	caaaaactccg	tctcaaaaag	aaaaaaaaaa	aaaaaaggaaag	caaaggaaac	18900
tcaagagacta	aaaaatggag	gtctcaggga	cgggcctggc	tgagggtggg	agatcacttg	18960
aggcgggaaag	ttcaagacca	gcctgggaa	catagagaga	cactgttctt	acaaaaaaagt	19020
aaaaaaaatta	gcaacgcata	gtggtgcgtg	cctatagccc	tagttactca	gaagactgag	19080
gttaagaggat	cccttgaggc	ctggagtttg	agattacat	gagctatgtat	tataccactg	19140
cactccagcc	tgcaaaatag	agcaagaata	aacttttggc	tgggcattgt	agctcatgcc	19200
tgttaatccca	gcaactttggg	aggctgagg	gagtagatca	tgagggtcagg	agttcaagac	19260
cagcctgtatc	aatatggtga	aaccccgctt	ctaataaaaa	tacaaaaaaa	attagctgag	19320
ccttgggtca	ggcaggctgt	atccccagta	ctcaggagggc	tgaggtagga	gaattgttgg	19380
aacctggggag	cgagaggat	cagtggagca	agatcgcgc	attgcactt	gggtgtgcaa	19440
aggagagact	cttggagcaag	actccggago	aagactccat	gttaaaaaaaaa	aaaaaaaaaaa	19500
aaagaataaa	cctttaaaga	aaaaaaagaa	aaatcatata	gtggcttata	atctcccgac	19560
ccagaatcc	accaggatgt	gccccttgg	ctttgcata	ctctaaatata	tgtatcacct	19620
ggaattaaac	caactgtt	aaaaaaagaa	gaagaaaaaa	agaaggaaat	tcaagttctt	19680
aaagcccaagc	actaqttagag	ctcaqacaq	ttacccctt	atccataatga	agtagqaagg	19740
gcaccactat	gtaccttagca	cccaggaaat	ttaacccctt	actcacgttc	tgcctggaga	19800
aggaaggccag	gtgcataatgc	agggatctgg	ctccctgaact	ctgtcaaggc	tgaggcagg	19860
cttggataaga	aatcttacc	cccattgtcg	gcttgcgtac	agccctcc	ctcatctgc	19920
acaaccccttga	ttcaggggca	ggaactgt	cactggctt	tggaaaggct	aggacttga	19980
acttcgttga	acttcggagc	tgaagggtgt	aacagagctg	gtgccaacc	cacttttagt	20040
aaccccatat	ttttttttcc	cccccaactt	acaacatata	aaagacattt	tttccacaac	20100
aataacttaca	tattttat	atccctgtt	tgctgttat	tatccatgg	atgcacaaca	20160
atttattttt	gatattccct	ttccataat	attacgtta	tatgttttta	tttttttata	20220
ttaaaaatat	atttttaaa	tagagacagg	gtctcgat	gttgcgggg	ctggctcaa	20280
actgtgtggc	ccaaacatac	cttcaatctt	ggccctccaa	agtgtgtgg	ttacaggcgt	20340
gagccaccac	gtcagctg	ttttttttat	ctaaacacaa	cagtaaacag	cttgcctcc	20400
tttcttattat	ttccctaaac	taagttctt	acaattgtcg	aatcagaagg	gttgcatt	20460
taacattttc	atatatata	tcaaaattgc	ttatataat	gtttttttt	tctttcttct	20520
acttgtat	ttaagttat	atattttaa	ttgtataact	ttttttttt	ttttttttt	20580
tttttttagag	acagggtctt	gcttgcata	ccaggctgg	gtgcgtgg	acaatcatag	20640
ctaactgtt	ctccaaatct	ctgggttca	gtgtccctt	cacccatgg	tccattgtag	20700
ctaggactac	aggtgcacaa	caccccttatt	tttttagaga	cagggttca	gtatgtgc	20760
ctggcctcat	gtgatccccc	cacccagct	atgtccctt	ttatcaac	cacatgtat	20820
tccctggaaat	taacatccctg	ttccatttc	ccacaactt	gcacatata	attctctgc	20880
tcagactacc	attctccagc	ttccctttaac	ttcccttccca	acaacagtct	attttccaa	20940
aaaactcttgc	gcacatggta	gatatcata	aatattggc	agtcgaatga	ataaatcaat	21000
cacttctacc	cacttttagc	acagagcago	ctctgtgggt	gacatgtgc	acagaagacc	21060
gtaaaaattag	gaaatgtcgat	tgacgggggg	aatccgcacc	gcccataatgc	aggctgtctg	21120
cttccctgtcg	cggttattcc	cttgcgtcc	tgccctagcc	cacaccattt	cctcccttcc	21180
ccccccacccc	cgactttcca	gtccccagcc	atgggttca	gttgcgtca	gcctccccc	21240
caccacttca	ctgcagcatt	gattttgc	actgcataaa	tagaactgt	tcctaccctg	21300
ataataactat	agcttctatg	ccaaacagct	tagctctaca	taatcatgt	tgtgcattgt	21360
cagggttgaag	tgtgtttgc	ttttttttat	tggtgtat	tgccatgtca	tgagagagaa	21420
agagactcgca	accctgtcg	gagttcttct	cttctattct	ggagttctgt	ctctgtcagg	21480
gatgggtggc	agttttagt	agtttactt	gaaggggtag	gagaagaggg	aacagtttc	21540
ctatttcaac	ctagctgtca	gcagcttata	ttgaatataa	ataataatca	ttaccacta	21600
ttaaatcttct	atgttaccacc	aggcacttta	catccattgt	ttcattttat	ccccaaacaa	21660
ctaaaaaaat	taagcatggc	cagttttgt	gctcatgtt	ataactcgat	cactttttgg	21720
aggccaagg	ggggaggatcg	tttgagccca	ggagttttag	atcacctgg	caaataatagt	21780
aagacactgt	ctctcaaaa	aataacaaa	attagctgat	cacccgtgg	gacagttgt	21840
gtcccaatgtt	ctggggggc	tgagggtgg	ggatgtttt	agccatgggg	gtcaacactg	21900
cagtgcgc	tgatcatgtc	actgcactt	agccatgtat	atagacta	actctgttc	21960
aaaaaaaataaa	taacatataat	ttaagaaattt	aagaattttt	cttttgtca	taatcaggaa	22020
aaaagcaata	aatgggtctg	aggccaaat	cactctgtt	ggatggggg	tttggatctca	22080
gatccccaaa	gatgggttag	ctgactgtca	aatgcgttca	ccaggccct	gttgcaggct	22140
tgttgttct	cagccagccac	ttgtatggta	ccacaggat	tgaggggctt	gagggggaga	22200
ataatgcctc	atgtcgcacgg	cccaggggca	tttccagtag	ctacctgtt	agggacttcc	22260
actacttggg	ggaaaatgtat	caggcttctt	gaagttgtt	tcaagatgtt	ccagatgtt	22320
ggaatgtat	ggccctataa	ggggatgtat	aggaaactttaa	atatgttctt	taaatgtgtcg	22380
ctcattttacaa	tgtctactaga	aggcccaatgc	ataaagtcaat	gatctgttt	aggccccctt	22440
atcgccctgc	ttcgatataat	ggtctcttgc	gaacctgtac	tctgttctt	attcttattc	22500
agacagccat	gcctacatgt	agcaggaaat	agtcacttgg	ggtagcttgc	22560	
ggacacaaatc	agagacttgg	acaggcaggca	ttttttttat	catctcttct	accttgcac	22620
attctcttta	gggttctatgaa	accctgtca	aatgtctaaa	cggtcattgt	gggttctgt	22680
gaacttttaccc	tttcccttccc	tcctgtctat	ttgtccatata	agggtatgtc	ccttcttctt	22740
cctagacagc	cattctact	ttcttagtct	tcatctgtt	caccccttct	tgtgcacaca	22800
caaggccaa	cctgtcaca	cacataggta	tgcatctgtc	ctttcttata	tttgccttcc	22860
tggactgtt	accttgcagg	cagagaacaa	ggcttttttt	tttttttttt	tttttttttt	22920
gtgtatgtat	ctcaactctgt	cactcaggact	gggtgtcagg	gaacatctt	ggcttcatgc	22980
aacctgcacc	ttctgggttc	aacacttct	cctgtatgtc	ttgggttata	gggtgcctgc	23040
atcacacccg	gtaatttttt	gtatttttt	tttttttttt	ttgtttttt	tttttttttt	23100
ggattttattt	cctcacatgtt	cagcataactt	ttcatatgtat	cctacacata	tcagatgtat	23160
actctgttgc	gggtgagat	gtgacactga	cccagcacag	ctgtgagggt	gtgacacggc	23220
aggttagcag	gtgcgcacag	aggggcttcc	ttctaaatgg	gggttgcagg	tagatgtttc	23280
agctccgcag	agcttatttca	gtgtcttgc	tttgcgttgc	ttcccttgc	ttcttaataaa	23340

-continued

agctgcagat	ggagtcaactc	agtccacat	ggctggggaa	gcctcgccaa	acttacaatc	23400
aagggtggaa	gcacccttc	acagggtgg	aggaaagaga	atgagtgct	aatttttcta	23460
ttttttagtag	agatggggtt	ccaccatgtt	ggccagtc	gtctcaaact	cctaacctca	23520
agtgatctc	ccacccgtgc	tcccaaagt	ctgggattgc	aggtgtgagc	caccgtgcc	23580
ggccaaggct	ggtccat	tgtgcccagt	gaaaggacta	ggacatagta	tatgcta	23640
aatatgttagc	taaatgtat	atccatgatc	ccccccacaa	ttatccatt	attttttttc	23700
tttttttttt	tctttttttt	tttagacaag	agtctcg	tgteacccag	gctggagtg	23760
agtgccatga	tcttgacc	ctgcaac	caccatcaa	gttcaaggaa	ttctcatg	23820
tcagcctca	aagtagctg	gattacag	gtgcacc	atgeccagg	aatttttgt	23880
atttttacta	gagatgggg	ttcacatgt	tgcccaagg	gatcttaac	tcctaatctc	23940
aagtgtatca	cccgcttgg	cctccaa	tgtgggat	acaagttg	gccaccgc	24000
ccagctccc	tccatcat	tctccactt	catatctcc	gtcttcca	ctggacc	24060
ccccgctgt	tagcacgtg	tcatgctc	atttccaca	tagtggaaa	cttccctcc	24120
atctgcatca	agtca	aaccata	ttttctc	cttactt	ttttttttt	24180
tgagacagg	tctactc	ttgtact	tttgtatgg	ttgggt	cgat c	24240
tgccgcctt	gctccagg	ctca	ctccac	cagtc	cttca	24300
aacacagg	catgctac	tacca	attttgtat	ttttttaga	gatgggat	24360
caccatgt	cccggtgt	tcttgagaa	ccctgggt	aagtgt	cctgc	24420
cctccca	agtgatca	tttttctt	tttctt	tttctt	tttctt	24480
gagaagtc	tctctgatt	ctccatgt	tcatctc	tttctt	atttgc	24540
atctggctc	ggattaac	tttccactg	ggagctc	acattgg	tcaatgat	24600
tcatgtcc	caatccaa	agtttctgt	ttggat	acaat	atataaagac	24660
tatggta	agtggatgt	caataa	ttgttctt	tatttctt	tgctca	24720
tgtttttt	ttttttttt	atttattt	tttgtg	ttggat	tttgc	24780
cagactgg	tgca	ccatctcg	tcactg	ctctgc	cagg	24840
caattctct	gctc	ctca	ttggatt	ggcac	accatg	24900
gtaat	ttttttt	ttttagatgg	tttccac	tttggcc	tttggcc	24960
actcc	cc	tttggat	ggcc	tttggat	tttggat	25020
gagcc	cc	tttggat	tttggat	tttggat	tttggat	25080
ggaaaa	aa	tttggat	tttggat	tttggat	tttggat	25140
tcat	cc	tttggat	tttggat	tttggat	tttggat	25200
aaacac	cc	tttggat	tttggat	tttggat	tttggat	25260
agatgagg	ttttagat	tttggat	tttggat	tttggat	tttggat	25320
ccttctt	ttttagat	tttggat	tttggat	tttggat	tttggat	25380
ttttagat	ttttagat	tttggat	tttggat	tttggat	tttggat	25440
gagtgtat	ttttagat	tttggat	tttggat	tttggat	tttggat	25500
taaac	ttttagat	tttggat	tttggat	tttggat	tttggat	25560
tttttagt	aa	tttggat	tttggat	tttggat	tttggat	25620
tgatct	cc	tttggat	tttggat	tttggat	tttggat	25680
ttttagat	ttttagat	tttggat	tttggat	tttggat	tttggat	25740
acc	cc	tttggat	tttggat	tttggat	tttggat	25800
gattac	gg	tttggat	tttggat	tttggat	tttggat	25860
ctccatgt	gt	tttggat	tttggat	tttggat	tttggat	25920
tccca	at	tttggat	tttggat	tttggat	tttggat	25980
gaata	cc	tttggat	tttggat	tttggat	tttggat	26040
tcaag	gt	tttggat	tttggat	tttggat	tttggat	26100
atcc	at	tttggat	tttggat	tttggat	tttggat	26160
tgta	ag	tttggat	tttggat	tttggat	tttggat	26220
tcaag	cc	tttggat	tttggat	tttggat	tttggat	26280
cgct	at	tttggat	tttggat	tttggat	tttggat	26340
gtcg	gg	tttggat	tttggat	tttggat	tttggat	26400
ttct	ct	tttggat	tttggat	tttggat	tttggat	26460
aattt	tttggat	tttggat	tttggat	tttggat	tttggat	26520
ctgac	cc	tttggat	tttggat	tttggat	tttggat	26580
cacc	cc	tttggat	tttggat	tttggat	tttggat	26640
gttc	cc	tttggat	tttggat	tttggat	tttggat	26700
gga	gg	tttggat	tttggat	tttggat	tttggat	26760
ctgt	gt	tttggat	tttggat	tttggat	tttggat	26820
acac	ca	tttggat	tttggat	tttggat	tttggat	26880
gaaa	aa	tttggat	tttggat	tttggat	tttggat	26940
tact	at	tttggat	tttggat	tttggat	tttggat	27000
gcaaaa	ata	tttggat	tttggat	tttggat	tttggat	27060
cctggat	gt	tttggat	tttggat	tttggat	tttggat	27120
atata	tg	tttggat	tttggat	tttggat	tttggat	27180
ggcc	agg	tttggat	tttggat	tttggat	tttggat	27240
tcac	gg	tttggat	tttggat	tttggat	tttggat	27300
aaaata	aaa	tttggat	tttggat	tttggat	tttggat	27360
ggct	gg	tttggat	tttggat	tttggat	tttggat	27420
ccact	gt	tttggat	tttggat	tttggat	tttggat	27480
aatatt	tttggat	tttggat	tttggat	tttggat	tttggat	27540
gcac	at	tttggat	tttggat	tttggat	tttggat	27600
cactt	cc	tttggat	tttggat	tttggat	tttggat	27660
ttgc	at	tttggat	tttggat	tttggat	tttggat	27720
gaaactt	tttggat	tttggat	tttggat	tttggat	tttggat	27780
tgaat	cc	tttggat	tttggat	tttggat	tttggat	27840
ccggaaa	aa	tttggat	tttggat	tttggat	tttggat	27888

-continued

SEQ ID NO: 5 moltype = DNA length = 24142
 FEATURE Location/Qualifiers
 source 1..24142
 mol_type = genomic DNA
 organism = Homo sapiens
 SEQUENCE: 5
 gccggggcgc ggctgggtgac gcagcaaaat gcccgggcct tgccggcccg cgtgcgcagt 60
 ctttgttttgggtccggc gcggtttct agagatatac acggaaaagg cggggcctcc 120
 cagcttccttc gcgatccctt ctggcccgcc cccggccctt cgtccgcac cccaccccca 180
 ctccggcccc tgcgtatccc tacgegtcc cgectcteet ccegtctgt cttctegegt 240
 tagaaaaaccg tttagtgccca aactgcctg ccatgcgcct teccggattct ctgcgaaaa 300
 cgggtcgctt gacatggga ggcaggctg agggggaggc acacggatc gggatcttgg 360
 aaactgaaa tggcgcgtt gttactctt agggagataa tcagtcact gatgcattcat 420
 tccaccacaa aaacctgtaa aaagectgt ctgtggaaaa gggctcaatt cagaatcaag 480
 ggcctgggtc cacctcgcca tttttctgaa ggttggaggc tctgtccagg gacagcggaa 540
 attcgaggat ctttgoggag gaggagatt tgccaaatgg ggccggctt tgccttgc 600
 tttagacccctt agtgcgtat agtgcacccgaa aaggatgtt gggggagag tcttagtttc 660
 tgggaatggg gagagtagga cattgtgtt ttccacgcac atcggttggg ggagaagggg 720
 tccccctcg ctttgcaca catattagta gggtgttagta ggggttagcaa ccaagagacc 780
 caaaacacgtc ctgcctaaga aagaagcgtc taacgggttggg ttggggaaaa atactctgt 840
 ggcgtcattt tgaaacaga aggaggatc acaaaatgtc taaaacctt tctctgaaagaa 900
 cggggcttca taataacgcac taaggatattt agtgcgtgt tggtgcaggc caagtagtgg 960
 gagctacgaa acctggaccc tagctttaat gccattaaaa agggaaattt ggcggggcgc 1020
 cgggtcgctt cgcctataat cccagttact tggaaatgtc ggccggccgc tcaactgttgc 1080
 tcaggagttc aggacccgc accaaacat gtgttgcattt catctctact aaaaataacta 1140
 attagcaggc cttgtggca cttgcgttgc atgccttataa ctactgttgc tgaggcgggg 1200
 ggtatagtcg aaccccgaaac ccggggaggcg gagggttgcag tgagcacttc aacttggcg 1260
 acagatgtcg accctgtgtt aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa 1320
 tcaaggctgtt aatccacca cttcgggagg cagggggggg catggcactc ggtcaggagg 1380
 tttagcccat cttggcttac atgttgcattt cccgtcttca taaaataaa aaaaaaaaaatca 1440
 gecgggtgtt gtggcacacg cctctgttgc cagctactcg ggagagcggc acaggagaat 1500
 cgttgcaccc cggggaggccg aggttgcgtt gggccggat cgcgcactgtt cacttcggcc 1560
 tggggcagac agtgcgttgc ttatccaaaa aaaaaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa 1620
 ggccgtactgg tttcaggccctt taatcccaactt acctttgggg gcggaggccca tggttactt 1680
 ggagatcagg agttcaagac cagttgtgtt aacatgttgc aacccctgtc cttactaaaa 1740
 atacaaaataa ttggcccaagg agcgtatgtt gacgcctgtt atccaaatccat ttggggggc 1800
 cggggggggc ggatcgcgg agtgcaggat cttggcccaaca ttggccaaaacc 1860
 cccatcttac taaaataaca aaaaatgttgc ttggccgttgc tgccataat cccaggatact 1920
 caggagtctt aggtggggaa atctcttgc cccggaaagac gcagggttgc gcgagttctag 1980
 attgttaccac tgcgttccat cttgtgttgc gtggacttgc ttgttaaaaaaa aaaaaaaaaaaa 2040
 aaaaaaaaaaaa gaaaaagaaaa aatagggtcactt aaccaatataa ttggggcact gaatgtcgc 2100
 ctaatttttttttt atgttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc 2160
 tgccgtatgtt gtatcacagg tcacttgcgtt ctgtacttctt gggtccaaattt gatcttcgg 2220
 taacgcactga atttaaatat atgttccaggc tcacaatattt caaaagagaa ctttatagac 2280
 tgatgttgcg aatttttttttgc atttttttttgc ttgtttaattt gaaagggttgcg aacatccatctt ttatagaa 2340
 ggctttcttca taagataaca actttagatgtt ttgtttaattt gttttagggaa gtcacccggaa 2400
 aagggttgcac aatgttgcattt atgttcccttgc gggccggggc gccggggctc acggctgttgc 2460
 tcccagcatt ttggggggcc gaggccggcg aatcaccttgc ggttggggat tccggaccc 2520
 cctggaccaacc atggagaaaccc cccatcttca ctaaaaaaaaccaaaatgtc cggggctgtgtt 2580
 ggcgcgtccgc tggaaatccca gttaccaggc aggttgcggc aggaaatgttgc ttgttacccca 2640
 ggaggttgcgg gttcaggtaa ggcggatccat tggccatgttgc ttttttttttgc ctttttttttgc 2700
 gcaaaactctt gtctcaaaaaa aagaaaatgtc tcttgggtat agggataaaa aattttggcg 2760
 ggcgcgtgtt ctcatccgtt taatcccaacttgc acctttgggg gccaaggccat gttgttgcacc 2820
 tgatgttgcg agtgcgttgc cgccttgcgc aacatgttgc aacccatctt ctatcaaaaaa 2880
 tcaaaaaatgttgcgtt ggttgcggatccat ggttgcgttgc ggttgcgttgc ggttgcgttgc 2940
 ggcggatccat ttttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc 3000
 cttccagccgtt ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc 3060
 gacggggccgc ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc 3120
 atcatgttgcg ctttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc 3180
 aaaatatacaaa aatttagccat ctttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc 3240
 ctgttgcgttgc accttgcgttgc tggggggggcc ggttgcgttgc ggttgcgttgc ggttgcgttgc 3300
 gcttgggttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc 3360
 gatattatgtt cccatgttgc ttttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc 3420
 gaggggttgcg ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc 3480
 agggagatccat ttttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc 3540
 ccaggatgttgc atgcgttgc acgttgcgttgc ttttttttttgc ctttttttttgc ctttttttttgc 3600
 ccaggatgttgc ttttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc 3660
 tggggatccat ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc 3720
 gtttttttttgc aaggatgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc 3780
 ggaggatgttgc ttttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc 3840
 cacctgttgc ttttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc 3900
 aaggacccgc tggccatccat ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc 3960
 ggttgcgttgc ttttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc ctttttttttgc 4020
 tccacccggccat ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc 4080
 ctccagccgtt ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc ggttgcgttgc 4140

-continued

gctcatgcct	gtaagcccag	cactttggga	ggccgaggcg	ggtggtcac	gaggtcagga	4200
gatttagacc	atccctggcta	acatggtcaa	accccatctc	tactaaaaat	acaaaaaaatt	4260
agccaggcgt	ggtggcttcg	cctgtgtcc	caactactcg	ggagggtcg	gcaggagaat	4320
ggtgttagacc	caggagacgg	agcttgcagt	gagocgagat	cacgcactc	cacttcagcc	4380
tgggcgcacg	agcgagactc	cgtctcaaaa	caaacaacaa	aacagaaaca	aaaacaacaa	4440
cctggccggg	tgtgggtgc	cctgctgtta	atccacgcac	tttgggaggc	caaggtgggc	4500
agatcacctg	aagtccaggag	ttcgagacca	gcctggggca	catggtaaaa	ccttgcgtct	4560
actaaaaata	caaaaatag	ctggggctgg	ttggcaggcgc	ctgttaattcc	agctacttgg	4620
gaggctgagg	caggagaatc	gettgaacct	ggggaggtgg	ggttgcagt	agccaagatg	4680
gcaccactgc	actccgcct	ggggcgcage	gagactccgt	ctcaaaaaaa	aaaaaaaaaa	4740
aaaaaaaaaa	statttacc	gtggggcgt	gtatccca	cactttggga	gaccggggc	4800
ggtggtcac	aaggtcagga	gttcaagacc	agctggcca	atatggtaaa	accccatctc	4860
tactaaaaat	acaaaaatta	gccaggcgt	gtggcgcgt	cctgtgtcc	cagctactca	4920
ggaggcagag	gcaggagaac	cacttcaga	cgggaggcag	agttgcagt	gagccgaat	4980
agcaccactc	cactccagcc	ttggcaacag	agtggagatc	tgtctcaaa	aaaaaaaggtc	5040
agaaggattt	tacgttaaa	aatgacatgc	attgaatgaa	gtattttaa	aatggtgcc	5100
gggcgcagt	gctgaagcct	gtaatcccag	cactttggga	ggctgaggcg	ggtggtcac	5160
aaggtcagga	gatcggaccc	atccctggta	acatggtcaa	accccatctc	tactaaaaat	5220
acaagaaatt	agctggcg	cgtggcccat	gcctgtgtcc	ccagctactc	gggaggctga	5280
ggcaggagag	tcacttgcac	ccgggagggt	gatgtcagt	gagccaaat	tgcaccactg	5340
cactccagcc	ttggcaacag	agcgagactc	tgtctcaaa	aataagtaaa	taatataaaa	5400
aataaaaaata	taatggccct	tctgtgtgg	acgggtggct	atgcctgtaa	acccgggact	5460
tttgggaggag	gagggtgggt	gatcacttgc	ggtcaggag	tttgcagac	ctggcaaca	5520
tggcaaaacc	tgttttcc	aaaaatac	aaatggcc	cggcgcagt	gctcacac	5580
gtatccca	cacttaggg	ggcgaaggca	ggttgatcat	gaggtcagga	gttcaagacc	5640
agccctgcgc	agatgtgaa	acccatcc	tactaaaaat	acaaaaaaat	gctgggtgca	5700
gtggcagggt	cctggatcc	cagetacta	ggagggtcg	acaggagat	tgcttggacc	5760
caggcaggcag	agggtcagt	gagccgagat	cacacgtt	cactccagcc	taggtgatag	5820
aggagtaaaa	ctccatctc	aaaaaaaat	taaaaaat	agctggcat	ggtgatata	5880
gcctgtgtt	ccagctactt	ggggggctaa	ggcaggagaa	ttacttgcac	ctgggagggt	5940
taggttgcag	tgcgttgc	tcgtgtgt	gactgtcgt	tctggcaac	agaggcaggac	6000
tctgtctc	aaaaaaaat	aaagatctc	tacccacgt	ctccaaac	caggctgtgg	6060
caatccccag	gctgtgtata	ggtaactggc	catggccat	taggaaagca	gccacac	6120
aagagatgaa	cgctccat	gccagcatta	ctgcctgac	tctac	tcgtacat	6180
tgtatggatt	agatcttcat	aggttgtcga	acccat	taactgcaca	ttcaagggt	6240
ctagggttgc	ctcgccat	tgaatctt	ttttttttt	tttttttgc	tacaggatct	6300
ctgtcacc	gggtggagt	ccgtggcgc	atctggctc	actgcac	ccgcctctg	6360
ggttcaagcg	attctctgc	ctcagectc	tgagttagt	ggattacagg	tgcttgcac	6420
catgccccgc	taatttttt	tttat	tttagacgg	tgcgtct	tcacca	6480
ggagtgcagt	ggtgcgtat	cagetact	caagcttcgc	ctccca	gacgcatt	6540
tccttcctca	gcctccctag	tagtggac	tacaggtgc	tgccac	tctggctat	6600
ttttgtattt	ttagtagaga	ttgggttca	ctgtgttgc	caggatgtc	tcgatcac	6660
ctacttgcgt	atctgcctc	ctcgccctc	caaagtgc	ggattacgg	cgtgaccc	6720
tgcccttc	tgtggccgc	taatttttt	ttat	tttagat	gagatgggg	6780
ttggccaggct	ggtcttgc	tcccaac	ttgtgtat	ccagcc	tctccaa	6840
tgctaaatg	acaggcgt	gcccactgc	ctggccact	cgagaa	atgttgc	6900
gaacggttt	atccaaat	tatccctca	ccatggaa	acttgc	tgcac	6960
atttctgtt	aaaaaaactt	tgaggacc	tgttctacc	caa	atctc	7020
gtttacat	gaa	ttat	taagacc	ttat	ggatggc	7080
tcctgtgtc	ccagtact	ggggccaa	atgggagg	ttgttgc	caggat	7140
agaccagct	ggactacata	ggggaggc	gtcact	aaa	at	7200
tgtgtgtat	gtgtat	ggat	tttgc	tttgc	tttgc	7260
catgttgc	gtgagacc	tcgcgt	tttttcc	tat	tttgc	7320
tagtagatc	tcaaaaat	ttgtt	taata	ttat	tttgc	7380
ctatggacaa	accagtcat	gtat	tttgc	tttgc	tttgc	7440
gcagtaggg	ttggacata	caagat	tttgc	tttgc	tttgc	7500
ccagtagaa	aaagaat	tttgc	tttgc	tttgc	tttgc	7560
caccatttgc	tgagactgaa	agg	tttgc	tttgc	tttgc	7620
tatatttgc	atgtggaa	agg	tttgc	tttgc	tttgc	7680
tcatgttgc	tgggtgaa	aaa	tttgc	tttgc	tttgc	7740
gataacagg	agaacaagg	tttgc	tttgc	tttgc	tttgc	7800
agagaaggc	taggaaga	ggagg	tttgc	tttgc	tttgc	7860
tttctgtct	ctgtctct	tttgc	tttgc	tttgc	tttgc	7920
aatataata	catacata	caca	tttgc	tttgc	tttgc	7980
catecc	tgtctgt	ggagtgc	tttgc	tttgc	tttgc	8040
ctccccaggc	ttaggtgc	tccca	tttgc	tttgc	tttgc	8100
gcactacc	acccagct	tttgc	tttgc	tttgc	tttgc	8160
gctcaggctg	gtctccact	cctgggtt	tttgc	tttgc	tttgc	8220
gtctccat	ccactgttgc	tttgc	tttgc	tttgc	tttgc	8280
tgaatttagt	ccactgttgc	tttgc	tttgc	tttgc	tttgc	8340
taagcttgc	tttgc	tttgc	tttgc	tttgc	tttgc	8400
atgtgtatc	taacttgc	tttgc	tttgc	tttgc	tttgc	8460
gagggttgc	ggagccgaga	tttgc	tttgc	tttgc	tttgc	8520
ttgtctcaat	taaaaaata	ataata	tttgc	tttgc	tttgc	8580
caggatgacc	tttgc	tttgc	tttgc	tttgc	tttgc	8640
caggcgttgg	ttcgttgc	tttgc	tttgc	tttgc	tttgc	8700

-continued

cagttcttct	gccttagcct	ccgcata	gggattata	ggcatgcacc	accacactg	8760
gctaattttt	tgtatTTTA	gtagagacag	ggttttgaca	tgttggcag	gctggctcg	8820
aactctgac	ctcagggtat	ctgcacact	cagectcca	aatgtctgg	attacagtgt	8880
gagccacccgt	goctggccat	ctgcgtgt	tcttaatgtc	attgtctcg	gcctttcca	8940
ggagtatgaa	tcttcagttt	atcccattaa	tctttccatt	ttgtctacaa	atatgcttag	9000
gtgttccatt	ctaaaaagca	agacaaaaca	aaccttccct	caggccct	ttttttttt	9060
tttgagatgg	agtcgc	tgtcgcc	gctagatgt	agtg	ggcgtca	9120
ctgcaagtc	tgcccccgg	gttcaaggca	tttcgtc	tcagccccc	gagtagctgg	9180
gactacagg	gcccgcaca	atgcggca	aactttgt	acttttagta	gagacggggt	9240
ttcaactgtgt	tagccaggat	ggtcata	tcctcac	atgatccc	cacccgtc	9300
ctaaaatgtc	gggattacag	gtgtgagca	ccggcctgg	ccacccagg	cctcc	9360
atgctgtgt	ctgtttaact	ttttccat	tcttaagat	actccata	ctatctc	9420
taaaggta	tctttgtgt	ccattcatt	gttcccact	ccccagct	ccaaagtgg	9480
tacactgaa	gocatcaat	tgtcaattt	aaaggccct	tttcc	ggc	9540
ggeccatgcc	tataatcc	ttat	tttgggg	taggcagat	gectgagcc	9600
gatcagecgt	ggcaacatgg	tggaaac	tcttccacaa	acccccc	ccccaaat	9660
agccatgtgg	cttgggtgtgg	tggeta	ctgtatccc	agca	acttgg	9720
ttggcagatc	acctgtgg	aggatgtt	gaccagctt	accaat	atgg	9780
tctctacta	aaaaat	attagctgg	cgtgg	ggc	ctgtgt	9840
cttggggaggc	tgagacagaa	gaat	ccctgggg	ccag	gggtgt	9900
agatcatgcc	actgcactcc	agca	cactgg	cc	agatgt	9960
aaaaaaaaaa	aaaat	tgggatgtt	gactctc	tgtat	ccca	actacttgg
aggctgaggc	aggagaatag	cttggaa	ggagg	gttgc	agatcg	10020
caccactgca	ctccagctg	ggca	acacag	ttagact	tttcc	10080
aaaaaacagt	tagcgtgt	tgg	tcag	ctactt	aggagg	10140
tcacttgagc	ccaggagg	aagg	tttgggg	ttagt	gggtgg	10200
ttgggtgacag	agacactgt	tcaaaaaaa	aaaaaa	aaagg	ccgtt	10260
ctgttctttt	ttttttttt	tttgg	gagaca	gtttagt	ttttcc	10320
agtggttac	tgcac	ccct	ctgggtt	ttctgggtt	aa	gttgc
cctccctgtgt	agatggatt	acagg	tttgg	tttgg	tttgc	10440
tagtagagac	gggttac	catgt	gggtt	gggtt	gggtt	10500
atccacccgc	cttgg	tttgg	tttgg	tttgg	tttgg	10560
ctgttcttgg	tctttt	tttgg	tttgg	tttgg	tttgg	10620
ggttctttag	cctca	ggaa	ggaa	ggac	ggac	10680
tttattttttag	aaataa	aaaaaa	aaaaaa	ggct	actt	10740
gctactgttt	gactactt	tttattt	tttattt	tttattt	tttattt	10800
atttatgagc	tttcttggaa	aggat	ggatgg	tttgg	ccac	10860
tagacttatt	gggttaactt	ctgactt	tttgg	tttgg	tttgg	10920
ttgggtgtgtc	ttttgg	tttgg	tttgg	tttgg	tttgg	10980
cgaggggggtc	atttt	tttgg	tttgg	tttgg	tttgg	11040
acatgtgt	tttct	tttgg	tttgg	tttgg	tttgg	11100
atccctgtgc	atagaat	tttgg	tttgg	tttgg	tttgg	11160
tttacccgtc	tttgg	tttgg	tttgg	tttgg	tttgg	11220
tttgacatagt	tttgg	tttgg	tttgg	tttgg	tttgg	11280
tttattttttttttag	tttgg	tttgg	tttgg	tttgg	tttgg	11340
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	11400
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	11460
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	11520
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	11580
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	11640
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	11700
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	11760
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	11820
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	11880
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	11940
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12000
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12060
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12120
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12180
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12240
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12300
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12360
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12420
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12480
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12540
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12600
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12660
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12720
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12780
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12840
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12900
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	12960
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	13020
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	13080
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	13140
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	13200
tttgggttact	tttgg	tttgg	tttgg	tttgg	tttgg	13260

-continued

ttaattcaga	aatcaaaaat	ccagcaatct	cacttctggg	tgtttagaact	gtagaaaaata	13320
aaatttttga	ccccggacgg	tggctcacgc	ctgtcatccc	aggccgaggc	gggtggatca	13380
cgaggtcagg	agatcgagac	catctggcc	aacatgggtg	aaccccgct	ttactaaaaa	13440
tacaaaaaatt	accctgggt	ggtggcggtg	gcttgtatgc	ccagctactc	cgaggcgctg	13500
gacaagagaa	tcacttgaac	ccgagaggca	gagggttggg	tgagccgaga	ttggggcatt	13560
gaactccagg	cttggcaaca	agagcggaaa	tctgtctaaa	taaataaata	aataaataaa	13620
aatataaaaat	tttaagcccc	ccaaactgact	gaatggatcc	ccatctggcc	aaggaggacc	13680
ccaaactct	gaaaagtag	ttcaggcaat	gtggggaaatg	ggaggcatg	aatatgcct	13740
attataactct	cttcccttt	gagttctggc	acaactgacc	agcattaaca	ttaaaacaga	13800
gatcttgcgt	gggcactatg	gctcatgtt	gtaatcccac	cacttggga	ggcccgagt	13860
gttggatcac	tctaggctcag	gagttcgaga	ccagcctggc	caacatgtat	aaaccccgct	13920
tctactaaaa	atacaaaaaa	atacaaaaaa	ataaataaaa	aattagctgg	gtgttgtatg	13980
gcatgtctgt	aatcccagct	acttggggagg	ctgagacaga	agaatcgctt	gaatctggga	14040
ggccggagggtt	gcagtgaaaca	gagatcacac	taccgcacgc	cagcctggc	aacaagagcg	14100
aaactcccgtc	tcaaaaaaac	agaaacaaca	acaacaaaaaa	gagttcgaga	ctagectggc	14160
caacatagca	aaaccccgct	tctactaaaa	atacaatgt	tagcggcgtc	tggtggcgctg	14220
tgtctgtat	cccagctact	caggaggctg	aggcaggaga	attgctgaa	ccaggaggat	14280
ggagggttca	gttgatcag	atccctggcc	tgcaactccag	cccaaggggac	agagtgagag	14340
tctgtctcaa	aaaataaaaa	taaataaaaat	gaataaaaata	aaacaagaga	tttaaagact	14400
gacaaacagag	gggggggtcg	gtggctcaag	cctgtatcc	tagctcttgc	ggagggtcaag	14460
gcaggccgat	caggggttt	agaccagct	ggccaacatg	gtgaaaccct	gtctctacta	14520
aaaataaaaa	aatcagctgg	gctgtgtatg	gcacacccgt	aatcccccgt	actcaggaga	14580
ctgaggccagg	agaatatctt	gaaccgggg	gaccaggatg	gtctgtatg	gaaatcgaaac	14640
cactgcactt	tagcttggc	aaacagatgt	gactctctca	aaaaaaaaaa	ggccgggggg	14700
aaggacaaaa	gagactagca	aaattccaa	atgatgtatg	tataacatca	cataaaagat	14760
aaaaaaaaaa	aaatcaaaaat	attttaccc	aaaataatgt	tctttgtat	attttgaat	14820
gttgggcaca	gttggctttt	cctgtatgg	cagtaactgg	aggcagaggc	aggagtattt	14880
cttgaagtca	ggagtttagag	accaggctg	gcaacaaga	cagactctgt	ctctataaaa	14940
ataatttaat	taaatataaa	taaataggcc	aggtgccgt	gctcaegct	gtatccac	15000
cactttggg	ggcccgaggcg	ggcagatcac	gagggtcagg	gatcggagacc	atccctgtca	15060
acacgggtgaa	acacgggtt	tactaaaaat	acaaaaaaa	aaaaaaa	aaatttagcca	15120
ggcgttgggg	caggcgcctg	tagtccagg	tactcaggag	gtctggggcag	gagaatggcg	15180
tgaacccagg	aggcagagct	tgcaataago	caagattgc	ccactgcact	ccagcctggg	15240
caagaatgag	actgtctcaa	aaaataaaaa	taaataaata	aggcttggg	tggtggctca	15300
cgcctgtat	ccacggactt	tggggggccg	agggtggccg	atccctgagg	caggagctcg	15360
agaccagct	aacatgtga	aaaccccgct	ctactaaaaat	tacaaaaat	agccaggatg	15420
ggttgggtgt	gctataatc	ccagactact	aggaggctg	ggcagaaaggaa	tcacttgaac	15480
ccaggaagcg	gagggttccg	tgagctgaga	tcgcgccact	gcactccagc	ctgggcaaca	15540
gagagagact	tcctctaaaa	aaaaaaaat	taaataat	taaactgtcc	gccaaggccat	15600
cttttgggg	ggggaaattt	catctgtaaa	taaatactgt	gaaatctat	agatctttt	15660
cctaccagg	ccacccaaat	ctgaagagat	taactgacag	tctagtttt	tttgggttgg	15720
tgtttgtttt	tgagatggat	tcttgcctg	ttgcccaggc	tggagtgcag	tgccaggatc	15780
tcggctgact	gcaatctctg	cctcccccgt	tcgaacgtt	tttctgctc	agccctcccta	15840
gtagccagg	cgacaggcgc	acgcacaa	gtctggctaa	ttttttttt	tttttttttt	15900
gaggagtctc	gttttgc	ccagggttgg	tgcaatgtt	gcaatctcg	ctcaactgca	15960
cctctgtctc	ctgggttca	gcaatttctt	tgctttagcc	tccagatgt	ctgggactac	16020
aggcatgtat	caccaacgcct	ggctatattt	ttttttttt	ttttgtat	tttagtagaga	16080
cggggtttct	catgttgg	cagggttgc	tcgaactctt	gacctcagg	gatccactg	16140
cctccagccct	ccaaatgtct	gggttacag	gcatgaggca	ccatgcctgg	cctgtttgtc	16200
tgtttttttt	agatggaggt	cccaactctgt	cacccaggct	ggagtgcatt	ggtgtatct	16260
caigctactt	caacccctcc	ctccctgggtt	caagaggatc	tcccctatca	gcctctcgag	16320
tagctggat	tacaggcacc	tgccatctat	cctggctaat	tgtattttt	tgagaaaggat	16380
atttcacccat	gttggtcagg	cttggggag	tgtagcagg	aggaggacca	gaagtcacte	16440
tggtggccat	tttgggtttt	ttgggattt	ggcggtctt	ttactgcaac	ctgtttcttt	16500
tgtttgtttt	ttttttggaga	ttgggttca	ctctgtcttc	caggctggag	tgccaggccgg	16560
tgatctcgcc	tcattgcac	ctcagcttcc	tgggttccag	cgatcttc	gcctcaggct	16620
cccgaggatgc	tgggattaca	ggctgtgtct	accacaactg	gctaattttt	gtatttttat	16680
tagagacagg	gttttgcct	tttggccagg	ctgtctctca	acttttggcc	tcaggatgtc	16740
cattcgcct	gacccctccaa	agtgttgggg	tttacaggat	gaaccatgt	caagcatgag	16800
ccaatgtgtc	ttggcgttgg	tttacacttgc	tttacattttt	ccatccccc	ttataagaga	16860
accccttaatc	tttttttttt	tttggatcgg	agtttcttac	tttccccc	gctggatgtc	16920
agtggcacaat	tctccgttca	ctgcaaccc	cgctctgtt	tttggaa	tttccctgtcc	16980
tcagccttca	gagtatctgg	gattacagg	gcccccaacc	atgcctagct	atttttttt	17040
tatttttat	agacacgggc	tttccactat	tttggccagg	ttgttcttgc	aaactctgtac	17100
tgtgtatccac	ctgccttggc	ctccccaaat	gtcttttca	caggctgtat	ccaccgtgccc	17160
tggcctttttt	tttttttttt	ttatattttaa	gtcaggatctt	gtctgttgc	ccaggcttgc	17220
gtccagtggc	acgatcttgg	ctcaacttca	cctccgcctc	ctgaggatca	tgattctcc	17280
tgcctcaacc	ttcccgatgt	ctgggattac	aagcatgcgc	cactacaccc	ggctcatattt	17340
tatattttta	gttagatcg	ggatcttca	ttttggccagg	gggtggcttca	aaactctgtac	17400
ctcagggtat	ccaccacact	cggttttca	aagtgggtat	tttggccagg	tttggccagg	17460
cgccccagtt	tctgtactt	cttcaggat	aaaaaggccg	atgatattcc	tgccctacta	17520
tttaggatctc	ttgtgttgc	tttggatagg	tttggatagg	tttggatagg	tttggatagg	17580
ggccatttat	aaactctggag	tttcaacaaa	tttggatgtc	tttggatagg	tttggatagg	17640
atttaaagaa	acattcagg	gttggccat	tttggctatcc	cagcacttgc	tttggatagg	17700
ggaggccaaag	ggggccggat	cacccatgt	tttggatagg	tttggatagg	tttggatagg	17760
gtgaagcccc	atctctacta	aaaatgc	tttggatgtc	tttggatagg	tttggatagg	17820

-continued

agtcccagct	actcgggagg	tgaatcgctt	gaacctggga	ggcagagggtt	gcagttagct	17880
gagattgcgc	cactactc	cagectgagt	gacagaggga	gactccatct	caaaaaaaaa	17940
aaaaaaaaaaa	aaaagaggaa	tcataacaata	tttgcctt	tgtggatggc	tcatttcatt	18000
tggcataatg	ccaacaat	atgttcago	atgtgacaga	atttcttcc	tttttaaggc	18060
tgaataataat	atataatatac	acatataata	aaaataataat	tttatacacac	cacacatact	18120
gttttttttt	gagatggagt	tttgagggtt	tgtcttgc	atccaggctg	gtgttcaatg	18180
gcatgatctt	ggctcaetgc	aaacctcagec	tccggggatc	aaaggactct	cctgcctcag	18240
cctccgaaatg	agctggggat	aggggcatag	gcaaccacac	ctggcttaatt	ttgtttttt	18300
tttttgagac	agagtcgc	tctgtcgcc	aggctggagt	gcagtgggtc	tgtgtcagct	18360
caactgcaacc	ttccatcccc	agggttgaatg	gattcttgc	cctcagccac	ctggcttaatt	18420
tttgttgcgt	ttttttgtaga	gacagggtt	cacatgttgc	gccaggctgg	tcttgaattc	18480
ctgacccatcg	gtgatcgcc	tgccctggcc	tgccagagggt	ttgggattac	aggcgtgagc	18540
caccatgc	agtcttat	attttttga	cacaggatct	cactttgtca	cccagggtgg	18600
agtgttagtgg	tgtgatcatg	gctcaactca	acctctgcct	cccagggtca	agggatctc	18660
ccacccatcg	ctccctagta	gttggggatc	ccacccatgc	ctggcttaatt	ttgttatttt	18720
gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtttttaga	tggagtttgc	18780
ctcttattgc	ccaggctgga	gtgeaatgg	atgatcttgc	ctcaactgca	cctctgcctc	18840
ctgggtttaa	tggtatctt	tgccctcag	tcctaagtag	cgggattata	gacgtccaca	18900
accatgcgt	gtctat	gtat	tttagtggact	gggttcatcg	tgtggccag	18960
gtctggcttg	aactctgc	ctcaagggt	ccatccgc	tgtgtccca	aagtgttgg	19020
attataaggcg	tgagccccag	tgccctggcg	gtgtggttt	ttttgttagag	acagtgttt	19080
gcottttttt	ttttttgggg	ggggatggag	tctaacttgc	tctccctggc	tggagtgtgt	19140
agtacagtgg	tgtgttgc	gctcaactga	acctctgc	ccagggtca	agtgtattct	19200
ctacccatcg	ctccctagta	gttggacta	caggcatgt	ccacccact	tggcttaatt	19260
ttgttatttt	agtagagaag	gggtgtca	aefttggcca	ggcagggttc	gaactctga	19320
cetcaggcga	tctggccggc	tcggcctcc	aaagtgtca	gatcgocat	tttgcctcagg	19380
ctagtc	aaagtctcg	ctcaaacat	ccttataact	ccaggoccaa	gcgatccctgg	19440
cttcccaaaag	tgctgggatt	acaggcatg	gccactgc	ccggccgaat	aatattctc	19500
tgttgcata	taccacatct	tatthaacca	ttcaactgt	gatggact	tgggtttctg	19560
ctgccttgc	ctattgtgaa	taatgtctt	atagatgttgc	ttgttaacact	gagttccctga	19620
ttgttaattct	tttgggttta	aaacccaga	tggaaatttgc	ggatcatat	gtaaatctat	19680
ttaattttt	ttgaggatct	gcccataat	tccatagttgc	ctgcaccatt	taacattccc	19740
accaacagt	cacaagtgtt	ccaaatctt	cacatcttgc	ctaacactt	tttttttga	19800
taagttagtca	gcctaataatgg	ttgagggtac	atctcattt	ttgttttaat	tacattttt	19860
tctttttttt	ttttttttaa	gggtca	cgatctcg	gccaggctg	gagtgcata	19920
gtgcgatctc	gactcatc	aacctccat	tccctgggtt	aagcgattct	cctgcacac	19980
cctccctgt	agctggggat	acagacatgc	gccaccacac	ccagcttaatt	ttttttttt	20040
agtagagatg	ggctttctcc	atgttggtca	ggctgggttc	gaactccaa	cctcagggtc	20100
tccggccacc	ttggccctcc	aaagtgtctgg	gattataaggc	atgagccacc	gtgcctggcc	20160
ttttctctt	ttttttgtt	ggatgttgc	tcttgggtt	ttggctggag	tgcaatgggt	20220
tgtatctcg	taactgcac	ctctgcctcc	cagggtca	ccattctct	gcctcaggct	20280
cctgagtagc	ttgggattata	ggcataggcc	accatggccc	gtaatttttgc	tatttttgt	20340
agagatgggg	tttcttcat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	20400
gcctgcctca	gcctccaaa	gtgtgtaaat	tacagggtgt	agccactgc	ccggcccta	20460
gcaccccttc	atgtgttca	tgatcttgc	ctatcttgc	tggataat	tcttcatggc	20520
cctttccat	tttgaatc	agtttttgc	gatgtgttgc	agtttttgc	atagtcttgc	20580
tattatcccc	tttacatgcata	catcttgc	aaatatttgc	tcttatttca	ttggctgtct	20640
tttttactgt	tttgcgtt	tttacacaa	ttttaatttgc	ttgtgttgc	tttgcgttgc	20700
ttttttttt	ttgtgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	20760
ctgtcacc	ggctacatgc	cagtgggtgc	atcttggctc	actgttac	ccgcctccca	20820
gggtca	atcttgc	ctcagccctc	caagtagtgc	ggactacagg	tgtgcgc	20880
catgttcc	taattttgc	atcttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	20940
tcttagcc	gtatgagcc	ctctgcgt	ctccaqaa	aaaaatttca	atcattctag	21000
taatagagga	agttatctgc	atttcaatttgc	ttgtgttgc	aaagacttgc	aggaaggccc	21060
gttgcgttag	ctcaggccta	taatctc	actttggggat	gtcgaggcc	gaggatcatt	21120
tgagggtcagg	agtttgc	ttgtgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21180
tacaaaaatgg	tgctggggat	ttgtgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21240
gcagaattgc	ttgaacatgg	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21300
tccaggcgt	ggcgacagac	aaatctgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21360
agaggtagga	attaatgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21420
ggaatgttgc	aacactactg	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21480
aaaaaaatgg	ggaatgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21540
tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21600
aatatggat	aaaggcttc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21660
gtacacgttca	ctacttc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21720
gagaaatgttca	taaacttc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21780
ttttttttt	ttttttttt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21840
atcatggctc	actgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21900
ctcaaggtat	cctccac	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	21960
accttagttag	tatttttat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	22020
tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	22080
gcagaagcc	cctccat	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	22140
tttgcgttgc	agggttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	22200
gggttca	gatcttc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	22260
ccatgc	taatttttt	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	22320
cgccca	gggttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	22380

-continued

caagcaattc	tcctgccta	gcctccctgag	tagctggat	tataggcgcc	cgccaccacg	22440
cccagctaat	ttttctgttt	tttagtagaga	cggggtttca	ccatgttgtt	caggctgtgc	22500
tcaaaacccot	gaccctgtga	tccacccccc	tcggccccc	aaactgttgt	gattacaggc	22560
atagccccac	gcgcgtccct	tttttttgtt	gttattttt	gtagatgg	ggtttactg	22620
tgtagccatc	aatggtttgg	atctccctgac	cttgcaccc	gcctgcctg	ctgtgcctag	22680
cctcccaagt	ttgggggatt	acagggtgtga	gcctactgtac	ccggccggcc	tcatttttt	22740
tggagaccaa	aagactgtaa	tctataaaaa	tgtcttaact	ccttttagttc	taaaggccca	22800
tttttagatgt	tagccacatt	ttttccat	ctaacatatt	cttcttaggt	gtatccatc	22860
atgttcattt	caggattctt	gcctctatcat	ttacttgtaa	ctttccattt	tctaaattgt	22920
agcccttta	tttcctccaa	accattctt	gacaacccaa	actgagttt	tcttggaaatt	22980
cctatgttgc	tatctgttaa	ctaactgtta	tgtttttagt	agcctgggg	agtcttagtt	23040
tttttttttt	gtccataat	gtatattatgg	ctgccttaat	cagaagctac	ttttgttagatt	23100
aaaatgtctg	tagagaatat	gtttaagagcc	agaaaaaggtt	caatatactc	aaaatttcc	23160
ctttcaggag	gtagctttt	catgtaccc	caacttttc	ttcaacccctg	taccaagctc	23220
tttacatgttc	atatgtaaacc	aatgttaact	ccttcagggc	cacaaaaggc	ccatgtgtca	23280
tttcaagaaa	tttggggc	gttggggc	tgtagattac	atacagaaga	atcggtatc	23340
gttaacagg	aacatctgg	aaagaaagg	gaaaatggaa	actattttt	tgagacatgt	23400
tctcaactctg	tcgcccaggg	tggagtgcag	tgggtgtatc	tcaggattt	ctcagtgtatc	23460
ctattttttt	tagaggccgg	gttttgcata	gtttcccaagg	ctgggtgtca	attcttggcc	23520
taaagagatc	cacccctgc	ggccctccaa	agtgttggg	ttacagggt	gagccgtgc	23580
ggctgttcag	aaatggaaaa	ctttatgtc	aatgttgtat	tccttaatgtt	aaagaaggttc	23640
taacagccct	gtatagtaac	tttttttttt	tttttttttt	tttgagatg	gagttttgtct	23700
ctgtctgtcg	cccaggctgg	agtgcagtgg	cgtgtatcc	gtctactgc	agctctgtcc	23760
cccaaggttt	caccatctc	ctgcctcage	ctcccaagca	gttggacta	cagggtcccg	23820
ccacccacgc	cggtttat	ttttgtat	ttttgtat	gggggttcaa	cgtgtttagcc	23880
aggatggct	caatctcc	acccctgtat	ctggccccc	tttggcttcc	caaagtgc	23940
ggattacagg	cgtgagccac	caacgcctgg	caatgtttcc	ttaaaagagg	tataaagtct	24000
gagatctaa	gtggggaaatc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	24060
taattaaata	attactggaa	ccttttttaag	tttttttttt	tttttttttt	tttttttttt	24120
ggggaaagaga	ccagtcaccc	ac				24142

SEQ ID NO: 6 moltype = DNA length = 13221
 FEATURE Location/Qualifiers
 source 1..13221
 mol_type = genomic DNA
 organism = Homo sapiens
 SEQUENCE: 6
 gacgtcacga tagaccggcc caccggggc cgggccgtgg gagccccccc ccctggcgcc 60
 ggtggatgaga tggctggggc cgagtcggcagg gggccgcggg gccatggggc aagctcagg 120
 gacgagacgc tgaggccggag aggggtcccc tgggcacgaa gaagggggagc tcatttagtaa 180
 cctcgatttg ggttctgtgg actggaaagag gctccataaa ggcaggggag gagaaggaga 240
 agaggaaatc attcttcggg agggaccggg ttggggatgg aaccttggggg cgcactgcgg 300
 cacgttctct ctcccttcctc aaggggaaac tggaggacgaa gggccagatg ccccttgggg 360
 agaggccggc gctctgtgg cgacgaggccg gggccggatcc ctggggggag cttaaagccg 420
 actctactgg agaagcagtc tacccttagag ggaggatggg cagtttctt gcgggggagg 480
 cgcttggggc tgccccccgc tgaggggagc cgttcctggc agggccggat ccccccggatg 540
 acgcgaggaga ggtggaaactc tccggccccc atctaggatcg ggcttcggaa atcggtatgg 600
 ggcagaggat ccagggtggg gggttggggc ccaqatggcg ctctccggcc aggctggaaac 660
 cggaggccct tgagaactctg cggtggagag ggacctggcc cttcccccgt taccatctg 720
 tgaagggttg gactggggaa aggtctacatg tctgaatcta ttatgtttaa gccacttctc 780
 ctttcctttt actctaattt ttgtgtctc aagagctccc agtcttcac tgactttct 840
 ttgttgttctt cttaggtcaa aggtttaaaatg tttttggggaa caaaggatgtg gagagcttcg 900
 ttaacctgga tcacccaggg gaggagctaa aacggggccc agaagacatcg cttccaaggc 960
 cagagactga tactttgaca tactaaatgt ctgtcataaa atctgggttctt gctggagaga 1020
 agggcccttc cctttgggggt cgggcttca gggactgttca ggtatggactt taaccacact 1080
 catccaaact ccagctttcc ctgggttctt actctgggtt ttacccatcat tagtgtgtg 1140
 tggagttctga cccaatgaga aagacttggg cccagtaatt ccctggggga gaaaggagggg 1200
 gtccctgggtg actcaggaaa aggtgtctagg aacagggttcc ctacgtttc tttggacttg 1260
 atcattatta aagggaaaaa aacaataata gagacaggga cgcagatgtg ataattaaga 1320
 tataagctga cagactggca agtcaaaaggaa aacacagat aatacaagta aatttggttt 1380
 tcattgtat ttctggggga agagacttag taaaatctt tgccttcac atcttcac 1440
 tccccaccat gcaacttccc cttaggttcc tggcccaaat gaaaggatgt tgacatgtga 1500
 tcaagaggat tccccatcaa ggtttctgtgg ctgaaacta actaaagaga tcctatccat 1560
 tttttttctt ttgtgtcttc tttttttcttcaaa agcaaggatct ttatgttccca gccaatcaac 1620
 aaatgtttaa ggccctggcgc agtggcttcgc gctgtatcaatc ccaacactt gggaggctga 1680
 ggcggccaca tcacaagatc aggagttcga gaccagctg accaacaag tgaatccatcg 1740
 tctctactaa aaataaaaaaa attagctggg cgtgggtggca catgtctgtaa atcccaacta 1800
 ctcaggaaacg tgaggccaggaa gaatccgttgc aaccaggag atggagatgg cagtggccaa 1860
 agatgtgtcc actgcacttc agccctggccgc acagtggatcc tccgtctcaaa aacacaacta 1920
 aacacaaaaaa cgaaatattag cggggcatgg tagcaagcac tctgtatccc aacagtggcc 1980
 tgaggccaggaa gaatctacttgc agcctggggaa gtggagggttgc cagtggccgc agatcacgc 2040
 attcgcactcc agcctgggggg atggagccggactctgttcc aaaaaaaaaaaaataaaaata 2100
 taaaaaacaa atgtttaatg aggccctggaa ctgtcttccatgg tacatgttatgtaaagccatca 2160
 tacaaaaatata aagacacatgttgc ctctggccatgg tagaaggccata cagtctgttgc gccacatccc 2220
 tttttttttat ttacccatgttgc tccgttttttgc ttatgtttaatgatgttgc gatgtggatgt 2280
 gggagtttca tcatgttgc cagggtgttgc tcaaaactctt gacgtcaagt gattcgcggca 2340

-continued

aacttggcct	ccaaaagtgc	tgggattaca	gacatgaacc	actgtgetca	gccacctgct	2400
atcctttacg	taccgcctt	agggtgcctag	agtccctttct	cttccttcctt	ccaaaactta	2460
agaatctgt	ggccaggcgc	ggttagtcad	gcctgtatac	ccagcaett	gggaggcgcga	2520
ggtggttgg	tcacttggg	tcaggagtt	gagactagcc	tggcaactt	ggccaacatg	2580
gcaaaacct	gtctctacta	aaaatacataa	aaaattagcc	aggcatagtg	gcacgtgcct	2640
gtactcccag	ctacttggg	ggctgaggca	tgagaatcgc	ttgaacccag	cagcagagg	2700
tgcaatgagc	cgagatctt	ccactgcact	ccagectggg	cgacagagca	aggctccgtc	2760
tcaataaaaa	aaagaaaaaa	aaaaaaaggaa	attgttggcc	aggcacatg	actcacacct	2820
gtaattccaa	cacttggg	agccaaggca	ggaggctcg	ttgagccctg	gagttggaga	2880
ccagcctcgag	caacatagt	agacgttgc	tcttcgtat	aaaaaattag	ccaggtgtgg	2940
tgtgtgtgc	ctgtatccc	agactatgg	gaggctcg	ttggaggatt	gtttggcca	3000
taatcacacc	actgcactcc	agcttggag	acagagcaag	gccctgtctc	taaaaaagaa	3060
aaaaagaaaa	aaaaaaaacc	tcttgaacag	gtgaggatga	aagagcaat	gcaagagccc	3120
attttcagag	ttcttggg	cataccat	aggcagtgt	ctactccagg	aattaaaggg	3180
atgtatattt	gcaagtaacc	agagttgtt	tctgttttcc	tttccgggt	cttcctct	3240
ccttcatttc	cttctctc	cttctctca	ctcttcttct	ctcttcttct	ctccatccct	3300
cttgtctttt	ccctctctcc	cttctctttt	ttttttttt	ttttttttag	acggagtctc	3360
actctgtgc	ccaggctgg	gtcagtcgg	acgatctcg	ctcaactgca	gctccgcctc	3420
ccgggttca	cccttctcc	tacccatcage	tcccgagat	ctgggactac	agggtccgc	3480
acacctgcgt	aatttttta	tatttttagt	agagacagg	tttcacccgt	ttagccagga	3540
tggcttgc	ctccgtac	cgtgtatgc	ccgcctcag	ctcccaaatg	gctggattta	3600
caggcatgag	ccacagtgc	cggeccctt	tttctttatc	tctctcttc	tctctcttc	3660
tttttcttc	ctccctctcc	tctctctt	ttttttttt	ttttttttt	ttttttttt	3720
ttttcttc	ttccctttt	ttttttttt	agagttttgc	ttttggcc	caggtgcag	3780
tgoaaaggca	tgatctggc	gcaccacaa	ctccgcctcc	cagttgggt	tcaagtgatt	3840
ctctctgcctc	accctccgg	gtacgttgg	ttacaggcat	gtgccaaccac	gcctgactaa	3900
ttttgcattt	tttagttaga	caggattttt	ccatgttgc	caggctggc	tcgaaactcc	3960
taacttcagg	gtatctggc	ccctggcctc	ccaaagtgtt	gggattacag	gcatgagcca	4020
ctgcacccgg	cttgcctcc	ctcccttct	tttttctt	tttctttt	tttctttt	4080
ctttcttc	tccttcctt	ctctttttt	ttttttttt	ttttttttt	ttttttttt	4140
actgtgtgt	acagtgtcc	ataggtttaga	tgttagaagg	agagcagg	ctggcagggg	4200
tgaggctcaa	tcataaaggca	gataaggatt	gttggcaagg	aagagaaca	aagtccagc	4260
attcaactgtt	ctatctgt	gagccaaagt	catgaagtcc	ttgagagg	acagccggc	4320
cagaaaagg	gtcgctctt	cagatgttag	gtcatgcctc	ttggagcaatc	ccaggctgt	4380
gagacatagg	atgggggg	gggtgggg	gtgatgggg	aatgggggg	gactgagg	4440
tttgcctgag	cgccatgagc	gggtgtgg	gagtgtgt	ggatgtgt	tgcatttgc	4500
tttgggggagg	gagataaact	ataaaaaagg	ggcctaagg	cccccaagg	gagagagggg	4560
ggaatgc	ctcagtgaga	gagaaaagt	acttgagagc	atttttttaa	agtccac	4620
tcagctttaa	aaaataataat	aaataatataa	ataatataat	atataatatag	ccaaatata	4680
actgtatgc	tactctgt	caagtactgt	tcatggttt	ggagatacag	taatgtat	4740
aaataaaagc	aaggtagaa	ggggatggag	aacaggcagg	caagctgg	ggcaggcag	4800
tgttgc	gtatgtat	actaggaaag	acccactgt	tgaagtgaca	tttaagcaga	4860
gacactgaa	agggtgaga	aatagctgg	atgagagcat	tctggggaga	agtacgca	4920
gcaggccagg	tcgcgttgt	caagcgtgt	atcccagc	tttggggaggc	tgagggtgg	4980
ggatcacttg	aatccaggag	ttcagagacc	gcctggacc	catgggtaaa	ccctgtct	5040
gctaaaaat	acaaaaattt	ggctgggg	gggtggctac	gcctgtatac	ccagca	5100
ggggggccaa	ggcgggtgg	tcatgggt	aggagatgg	gaccatcc	gttaaacacag	5160
tgaacccctg	tctctactaa	aaaaataaaa	aaaaataaaa	attagttgg	tgtgtgg	5220
ggcacctgca	atccca	ctcaggagg	tgaggcagg	aaattgtt	aacccgggag	5280
gcagagctt	cgatgtgg	agatcacacc	actgcactcc	acgctgggg	acagagcag	5340
actctgttc	aaaaatataa	caaaaaaaa	gggtgacaa	cacaaagg	ctcaagat	5400
caggaaggac	aaagggatgt	tagtgg	aaggacagg	aagaagg	tgtgct	5460
cacgtgtgaa	aaagggaga	gcagtgtg	gagaccagg	cacagg	gagggag	5520
agtcatgtgg	gactttctaa	gccttgc	aacctgc	aggatcact	tggctgtgt	5580
atggaggaca	gactgtgg	gggggg	cgacggaaa	agaaagg	ccct	5640
tgttagccata	atccatgt	gaccttgc	ttggctttag	caggtagaa	gcagttag	5700
caatgagaag	gggtcggt	tattttgt	gtatgtt	gatagacta	cgaaattt	5760
ctgtatggat	gtatgtat	gagagagg	tcaaggat	ttccagg	tctggct	5820
gaactgttta	cactgttca	taactgtt	aaaatgtaca	aatagtgtt	tgagaat	5880
ttaaagegg	gactgtgg	aaactcg	ctgtgtgg	atccggg	actgac	5940
acacttattt	cttgcacta	gccttgc	ggggatgg	ttggacacat	gaatgggg	6000
cttccacacc	tagtttact	ataagatatt	ggatattt	ccgttgc	tagatat	6060
aatggatct	gttggat	gcactgg	ttcaattt	tctttaat	tttcaag	6120
acatccgtgc	ggggccagg	tgtaatccca	acacttgg	aggctgg	ggcagg	6180
agtggtatca	cttgcacta	gggttgc	actagctgg	ccaacatgg	gaaacc	6240
tttctactaa	aaatacataa	tagtggca	tagtggca	tgctgtat	ctcagact	6300
tggggggcc	agacagggt	atcacgg	caggat	tttgc	ggccaacat	6360
gtgaaaccc	gtttctatta	aaggcgg	ttgcgtgt	ctgagat	ggccat	6420
tccagctgg	atgatagg	gagatcc	ctcaaaaa	aaaaaaat	tgctgt	6480
aaacttttgc	gggtgtgt	tgtggtt	ggctacag	gtttaaagg	accagac	6540
gaagggccgt	atttgc	agaatttgc	ctgtgg	aatttaca	gtctgtt	6600
attttagtta	gaaaggcgt	ttatgcatt	caggaaat	tttgc	tttgc	6660
ccaggtatgg	aaatagg	tgaagat	aaaaaaa	aatttca	ttcatgt	6720
aaaccctaag	gagcttaagc	ttaaagt	cgagtgc	tttgagc	ggcaggcc	6780
ggatttagaca	gcagaggagg	cactcaag	atgaggat	tgggggg	aagaaaag	6840
ggctgacccc	ttggcgcgaa	tgtatgg	gaggacag	ccatcaact	aaagag	6900

-continued

cccaaggaa	ttcagtggg	aacaaggtag	gttgaggag	ctgtgtgca	tccgggaga	6960
actgaacaga	acagaactgg	aatcatca	ctgaaactcc	ggagagaagc	tagtgccaga	7020
gagtttgaag	tgtgtttgag	gtgtgttgc	gcagtggac	ttaaccat	tgcagtacaa	7080
catggatag	tttagggagt	ctgtgaagag	agaacaggac	ctctgagatt	gaagtcacct	7140
gggagggttc	agtggacact	gtcatcgct	ccttctggc	atgaatctt	cccttcctca	7200
gttttctgg	aaaaacaagt	cttttagctgt	tactccctgt	cctatttcc	cttgccctt	7260
cattttact	ggtccccctgg	atcagccaca	cagtctgtt	aatctatgaa	aatgtgaaca	7320
cttgaacaa	acacacacct	atgttaaagt	atgtgccaac	agtggccctg	atgacctgaa	7380
acacagcgc	caggtgtct	ccctgactt	ctacttaagg	attactgtt	atggagggtt	7440
aaaagggtga	aaaaggaggca	gtttagggcc	aaggcttaaa	gtgttgctat	cccatatggg	7500
aaaatagcc	ctagccacat	gtggctattt	aatggccgtt	agcgtggct	7560	
cacacctgta	atcccgac	tttggggggc	cgaatgggc	ggatcaggag	gtcaagagat	7620
cgagaccatt	cctggcaac	atggtaaac	ctcatctcta	ctaaaaatac	aaaattagc	7680
caggccatgt	ggcgegtacc	tgttagtccc	getactcggg	aggctgaggc	aggataattt	7740
cttgaatct	ggagatggag	gttgcata	gcccagatcg	tgccactgca	ctccagcctg	7800
ccaaatagc	gagactccgt	ctcaataaa	aaataataaa	atagtttaa	ttaaaaggcc	7860
gtgcgcgtt	gctcatgcct	gtaatcttag	cactttggg	ggccggagggt	ggcagatgc	7920
ctgagctcg	gagttcgaga	ccagcttggg	caacatgggt	aaaccccttc	tctactaaaa	7980
tacaaaagaa	attagcaagg	ttgtgttgc	tgtgcctgt	gttgcctgt	cttggggaggc	8040
tgaggccagg	gaatttctt	aactttgggg	gcaggggtt	cagtgcgcg	agatcatgcc	8100
actgaattcc	agcactccag	cctgggtgac	agagtggac	tctgtctcta	aaaaataata	8160
ataataataa	taattaat	aaagtaaaaa	aattcagtcc	ctcagttggc	ctagccacat	8220
tttcaagtgt	tcaatcacca	cacgactatg	ggctactatg	tttggaaagt	cagataaaga	8280
acattttccat	catcacagaa	agttttgtt	gacagactgt	gaaggccccc	ataaaatgtt	8340
ccttccttct	ggacttttgtt	gaaggcccat	gaagacatgt	tgaagatgt	atgtggggga	8400
gaaatattaa	gcaggacaga	gagagtcaag	caatagcaaa	ctattctagg	tgtaaaagca	8460
aggcgcaggc	cggttgcgggt	ggcttacacd	tgtatccca	gcactttggg	aggccaaacac	8520
aggtgtatca	cgagggtcagg	aatttttagac	tagccggggc	aagatgttgc	aaccctgtct	8580
ctactaaaa	tacaaaattt	aacttggccgc	agttggccgg	gcctgtatc	ccagctactc	8640
aggaggctga	ggcaggagga	gaacttgcac	ccgggagaca	gagggttgcag	tgagccgaga	8700
ttgcgcact	gcaactccagc	ctggggata	gagcaagact	ctctctcgga	aaaaaaaaaaa	8760
aaaaggccaa	aggcattccc	agggagggg	agggggataa	agaaaaagca	atcaagattt	8820
tgaagaagg	cccacgtgg	ggttcaccc	tataatggg	gctgggggg	gtgtatggct	8880
tgagccccc	gggtccagac	cagtcgago	aacatgttgc	aaaccgcatt	tatgtaaaaa	8940
taaaaaaaat	aggccatgtt	gggtggggcc	cttcagcttct	agctactgt	gttgggggtt	9000
gaggatcat	tgagccctgt	ggggcagagg	ttgcagcgcg	ctaagatgt	gttactgtc	9060
tccagccctgg	gtgacagaggt	gagaccctgt	cccaaaaaaa	aaaaaaaaga	atctgtaaagga	9120
gaggatcaaa	ggtggaaaccc	attcttttt	tactggggag	aaatttcatg	gagaaaagta	9180
ggaaaaaccc	aaaaactttt	ttgtgtatgt	tcaaaacata	aaagagactt	ccaaaccagg	9240
catctttccca	ttagtttgc	ggaatgttca	aagtccagac	gaggggagcc	agtttctgg	9300
caagcatgtt	tgttttccca	gcttcctgt	agaagggttgc	gagactgttgc	agatgtgtc	9360
aaaaatggt	ccacttggct	aggaaataga	ttgtttattt	agctttcttg	tttgcctctg	9420
gtgtgttgtt	agatactatg	ccacaggaga	cattatgttgc	ggtggaaagac	caagggaggc	9480
gattttgtgt	tgagaacaa	atatggggaa	aagggtatgc	gttctgcata	gtcaggagga	9540
atcaggagg	cactcgagaa	aagggtatcc	cacggagac	ggacgcattgc	ttggatgtgc	9600
ttaatgtgc	ctggaaatgc	atcttcctt	atttttaa	cccatgggg	taacacactc	9660
aatcatgtatc	ccggccatgt	cgtatgtca	tgccctgtatc	cccagactt	ttggggccgc	9720
aggtggccag	atcacatgt	gtcaagatgt	cgacaccagg	ttggccaaaca	tagtgaacc	9780
cgctcttcac	tggaaaataca	attttatcca	gggtgttgc	cacatgcctg	taatccca	9840
tgcttgggg	gtctggatgt	gaggttgc	tgaacccagg	aggcagaggt	tgttagtgc	9900
ggagatgttgc	ccactgtcact	gcacgttggg	tgacagatgt	agactctgtc	tcaagacaa	9960
aataataaaa	aaacatgtatc	gaagtttaggc	taatttatgt	cttctctgt	gactccat	10020
cctctcaact	ctccactgt	tttgcgtttt	aaaatgtttt	gattgtaaaat	aaaaaaaagca	10080
aggaaaaatg	gggttaatgg	gggatagaaa	atggggatcc	ttggggatgga	atggggataa	10140
agtaaaacca	ggagttggaga	tagcccccc	aactttcagg	tctctgttcc	ccaaggcctt	10200
gaatgataaa	ttggccatgc	tttccccatt	caactatgttgc	gagaattttt	caatgtggact	10260
atccaggaa	ctggagaggg	ggaaaattt	accaacacgg	gaaataggca	cactgaaaaa	10320
gaacacttcc	cacacatgtt	atcaaggtaa	atttcttagga	tgtcaggcca	gaagtcatcg	10380
tttccagggg	gttcagaccc	ttctttgttc	cacctggc	cctttttccc	gttctcgaaa	10440
aatgtgttgc	aaacatgtatc	ttccccccatt	caactatgttgc	gagaattttt	caatgtggact	10500
gttagatgttgc	gttattttgc	taatggccaa	gcctccacaa	agcctcgaaa	atgttccac	10560
agagaaaaagc	aaacacacggc	cattttgcac	cggtggccccc	agaggtctg	aggcttacacc	10620
ctaggactgt	cggttgcgtt	tcagatccca	ctcactccac	attcatccca	agtcttactt	10680
ccagcttcc	tctcatgtaa	ccaaacacttc	agaagacttgc	tttgcgttcc	aaccagagag	10740
agatggctaa	aggatgttgc	atccatgggg	aaaggaaatcc	gatecaggccc	tccagtttcc	10800
tagatctttt	ccctggggtc	ccttcggatcc	ttgtgttca	gtctttatgt	tctcaaggca	10860
ctcccttggg	ctttggggca	ggattttccat	tcttcgttcc	tccatcttc	tggattttcc	10920
caactgtatc	gtctgtgtc	ccagacaccc	ctccctggcc	aaaatcccg	gtccctcccg	10980
cctcccttc	cctaatcaca	ccttcattcc	accccaacggcc	ctgaggccca	gcccacatgc	11040
cggccccacag	cccacgttac	ccctgggttgc	cccaatgttgc	gaagttgggg	gccagggtccc	11100
ttatggaa	gttggagggttgc	cttcattttgc	ggccgggggg	agggacactg	11160	
gaggggttgc	ttccctgtgg	ggcttggatgt	ggggggatgttgc	cccttcctac	atccccat	11220
aaggcaggcc	aaggcaggat	gttttgcgtt	tggctgttgc	aatatgtatc	tccacccat	11280
aacagggttt	ctattttgtatc	tcaggaaaaaa	taaggaggaa	gtggatgttgc	tttgcgttgc	11340
gcccgttgc	gagggggggg	aggaggatgttgc	gtgacccggag	tggggggccgg	ttaccaggc	11400
aacagggttac	gagcccccac	gggtgtccgc	gacagcaggcc	gggggttgcgg	gttggatgttgc	11460

-continued

-continued

ctgggctaac	cgcagttcc	gcctccctgg	ttcaaggcat	tctcctgcct	cagcccgcg	2400
atgtatcgaa	attacagaca	cccgactac	cccccgtctaa	ttttgtattt	tttagtagaga	2460
cgagcttct	ccatgttgg	caggctgctc	tgcgaactccc	gacccatgg	gatccaccca	2520
cctctgcctc	ccaaaggccct	gggattacag	gctgtggcca	ctgcgcctaa	cctttgttt	2580
tacttttatt	ttattttttg	agacaaagtt	ttgctcgta	cccaggcaga	agtcaatgg	2640
ctcaatctct	ggtcaactgca	acccctcgct	cttgggttca	agcgattctc	cagccctcage	2700
ttccccgagta	gtgggatta	caggcgcgt	ccacgacgccc	cagctaaattt	tttttttttt	2760
tttttttttt	tttagtcgaga	cggggttca	ccatgttggc	cacgctggtc	tcgaactccc	2820
gacgtcactgt	gatccgccc	cctcagecctc	ccaaagtgt	gggattacag	gcctgagcca	2880
cccgccacgg	caaaaaaaaaa	attattata	tttatttttg	gacggagtt	2940	
cgtctgtcg	cccaaggctgc	agtgcaegg	cgccatctcc	gctcaactgca	agctccgcct	3000
cccggttca	cgccattctc	ctgttcagc	ctcccgagta	gtctggacta	caggccccc	3060
ccaccgcge	cggtcaattt	tttgtatttt	tagtagaggg	ggggtttca	cgtgttagcc	3120
aggatggct	cgatctctca	acctcggt	ccacccgcct	cgccctccca	aagtgttggg	3180
attacagggc	tgagccacgg	cgccccggca	agaaaaaaaatt	atcagggga	gtctctctcg	3240
agccattttt	ggctgcctat	aaaaaaattaa	aggccggacg	cggtgttca	cgcctgtatt	3300
ctcagcactt	ttggagaccg	aggcagggtgg	atcacgtgag	gtctggagtt	caataccagc	3360
ttggccaaaca	ttgttggaaacc	ccgttctctac	taaaaataca	aaattagccg	ggcatgttgc	3420
cgcattgcgt	taatccgc	tactcgggag	gctggggcag	gagaatcgct	tgaacgcggg	3480
aggccggagg	tgtgttgagc	cgatgtggc	ccatcactac	acagctggg	caacaagagc	3540
aaaactgtgt	ctaaaaaaaa	aataaaaaata	aaaataaaaa	aaaagaaggc	atcttcact	3600
tatccccatc	accttcgtt	tccatgtctt	tccttctca	agctaaacta	ttcaatctct	3660
ttttgttccat	tcttcacat	ctcattatct	cctttagtct	tccttatttt	tttctactga	3720
cacaaccat	tttccctgt	tttattttt	attaagacaa	gagtctttt	ctgtcaccca	3780
ggctggagt	gatctggct	cactgtactt	ccgcctccca	gttctatgct	attctctcg	3840
ctcagcctcc	agagtagctg	ggatattacgg	cacacgcac	cacgcctggc	taatttctgt	3900
atttttagta	gagacgggg	ttcaactatg	tggccaggct	gttcttgaa	tgtgtaccc	3960
aggtgtatcca	cccgccctgg	cctccctaa	tgctgggat	acaggcatga	gccactgcgc	4020
ccggcctaa	ctgttaattt	aaaccaatag	aaaaagagt	aagaatata	gtattnaaaa	4080
aataggacgg	aagtcttgg	tatggctt	tctccctgtt	gttttctca	acctttctta	4140
tttatttaggc	tccaactgtc	aaggatgtgg	acaaaaattt	aaaagtgtcc	atttatccct	4200
ggatagtgat	caaaccgtg	cctcaattaa	gcaaggtaac	cctctcaaca	tggtgacgca	4260
tatgtgtcac	aactaacagt	taaaagaccc	ccqaaaaatgg	cagctccctt	gtctcccccac	4320
tgccggcggat	caaagtaaga	cgttccaaac	atggtggcac	aagatcttg	cgtcatttcc	4380
tgtatgttgc	tctatataag	gggcaggatt	tccgctttcg	ct		4422

SEQ ID NO: 8 moltype = DNA length = 3975
 FEATURE Location/Qualifiers
 source 1..3975
 mol_type = genomic DNA
 organism = Homo sapiens

SEQUENCE: 8

ggccgcctcgc	ctgctgtctca	ctgggatata	gtctggggc	ccgttgcatt	gtggtcttgg	60
tagtctttgc	gagaacggca	aatggcggtt	cgagaggcgc	gtcggaaattc	tcggctcgct	120
gtctccctggc	ttggcgctgt	ttgttcattc	tgggcttagt	gagcaggggac	aatacgtcgc	180
cctctcaggaa	ctttgtctc	tcctttaaaac	ccgaggattc	accgtcaagc	ttctggtttc	240
tctgacatcc	tcccccaggc	gcccccttcat	cgccacccgt	gatccagttg	agaagctgtg	300
ttggaggctca	ggggatggag	gcttttttgt	cccttgcggaa	accggaaat	at	360
ccggagacaca	tgacgacacat	gaaaattttt	cccttgcgttct	tggctcccac	tccgtacac	420
agatgaacgg	tgttacctt	cctcatgtt	cttgcgtt	ccatactcc	agagctgtc	480
cttgcataa	cctcttcaaa	tgtcttcaac	tctccctgt	tccttgcataa	tgggtata	540
ataagtgc	tggaaaaagc	atgagattt	aaagtccata	atgtttgtt	ctagggttctg	600
gtgttattta	atccatgtat	cttggaaag	tatctgtt	tggttctgtt	aaaacggaaa	660
tgctaatgt	tatttaccag	gactttttt	tgctggggc	agatgtttat	tatctccgtt	720
tcatgtat	gttcagagaa	tataatgtt	ttgttccaa	tcagcaatgt	aatgtgtgg	780
aaagctgtgg	tttgagccct	gactgttcaac	accggggca	aaatgtttt	gtctgtt	840
catcttatgg	atgttacacat	tcatgttcaac	tcatgttgc	atgtttgtt	ccagcactgt	900
gggttcatgc	tttactgtct	gataacgtat	tttcattttt	ttttttttt	tttcttagac	960
agggtcttgc	tctgttccat	aaaccttgc	agggtcttgc	ccgtttttt		1020
tttggttttgtt	tttgggtttt	aacggatgtt	actctggctt	tgttgcacca	ctggggatgt	1080
gttgcggat	acaggaggcc	tcatgttca	tttacgttgc	gaaatcttc	tcttcaggga	1140
tcctcccaag	tagctggat	ttcaggatgt	tgcacccata	ccttgcgttca	attttgcatt	1200
tgtttttttt	aattttttt	tgtttttgtt	gaaatgtgt	gtgttatttt	tttgcgttgc	1260
tgtgtgtat	gttatgtttt	acataaaatgtt	gggttgcggg	ttttttccat	taatgtttat	1320
ctatattat	atgtttggcc	gaaatctgtt	atgtttttt	actcttcacc	aaatcttacc	1380
ttatctcaca	acaaactgtt	atgtttttgt	atatttactt	tccatataacc	ttaccaggca	1440
tcagagtggg	gtctctactc	tggaaataaa	tatctgcatt	aaggatgttt	gaagacgtt	1500
gtctctttcc	caggaggcac	aaagtgc	tcttttctc	atcgcttgc	caaattaaac	1560
ctaccggaca	gtttactaca	gaacttcaaa	acagctggc	gcagtggtc	acgcctgtaa	1620
tcccagaaat	ttggggggcc	gagggtggcg	gatcaccgtt	ggtcggggat	tcaagaccag	1680
cctgaccac	atggagaaac	ccctgttcc	ctaaaaat	agaatttgc	gggcgttgc	1740
ccatatgtct	gtatcccag	gtactgtt	ggctggggca	ggagaatcac	ttgaccgg	1800
gaggtagagg	ttgcgggttgc	ccgagatcac	accattgtac	tccagcttgc	gcaacaagag	1860
caagaattca	tctcaaaaaga	aaaaaaaag	aacttaagca	taagccacaa	ataaggtt	1920
tcattacctc	atcaatgtt	ggataacttac	atttatcaag	tttgcgttgc	taatgtt	1980
aaaaatattat	tgtatgttgc	aaggggata	tatatatatt	atgttattaa	atatgtt	2040

-continued

SEQ ID NO: 9 moltype = DNA length = 13991
 FEATURE Location/Qualifiers
 source 1..13991
 mol_type = genomic DNA
 organism = Homo sapiens
 SEQUENCE: 9
 tattgttgc ttgttatttt tgccctggtaa ggccctgagat ggagggtcttt tccttcagggg 60
 atgcacagac tcggggctgtt ggatgtgacc tgccctgtt gccccaaatat tcatatcctg 120
 ttgttgcggg aacctgttag gaggctgagt aattctgtat ctttgcatca ggcacgttgg 180
 gaagttagcgcc catctacttgg cacatottaa gcactgcatg ctcccttggaa gtcttctcaa 240
 gaatgtctca aattggccac ctcaggagtg aegcatcagg gcaatgttgg tccgttggtt 300
 tcaaggcgtt gctggacccctt ggcttaatgt ttccatccccctt ctgtctgtc cagtgcct 360
 gtttagtaat ttgaagatc caagcgactt acatattaaga gttaggaaat cctatgttgc 420
 aaaatcacaa aagggtggga caggggctttt ctccatcaga cactatagtc agtgtgcata 480
 ggtgacccgtc aggaggcaac agtccctggcc agaccttgcc actcgccccca taaggcccata 540
 tctttatgtac tttttattttgg tggccacccag agacccatgtg taaaacagggtt tttaggaagg 600
 ttatggccaa tctcttcatac ctcaggatc acgaaaaactt ttacttccccca aaaccatgtc 660
 tctttgttca agaagtattt attgggttact tatgggttag gcatgtatgt aagccctggg 720
 gatatacgta taaccaagac acataaggcc ctacaggctt ctggaggaga cagaactaagt 780
 aattgtatcaa tttagtttaat tacagtccgg aagaatgttta tgaaatgtc agggcgatgt 840
 gcttattat ttaatcgat aagatctgaa agatgttgg ggtttaattttt gttaaaggagg 900
 ttaaaggatggg gaacagccattt ccataacaat tggacatgtt agggaaaatgtt ctggaaactt 960
 gaagacgctt ctggaaaagc ctaaaagaaga ctgtttaggac agcaactaaa agtaagggggg 1020
 acctgttgggtt agaggcgatg gcataaaaaca atgttgcattt tggttaggaaat tgatgtatttt 1080
 aaggccctaaaa agactgttga gatttttattt attttttttt tgaatgttgg tctctgttc 1140
 cgaggctgtt gatatcaaaa attttttttttt tttatagaga ttatgttccctt ttatgttgc 1200
 caggctgtttt ttgtttttttt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 1260
 tgccatgttgc ccatctcgcc tcactgttcc accttgccttcc ttgttgcctt ccgttgcgg 1320
 gcttccatgtt ccgttgcgtt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 1380
 tgtaggtttttt tggtttttttt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 1440
 ctcaagtgtat ctggccgcctt cagccctccca aagtacttggg attataggcg tcaggccctcc 1500
 caggctgtttt ttgtttttttt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 1560
 ggattacacg tttttttttt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 1620
 acaaaatcgat tttttttttt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 1680
 cagccatgttccatgttgcgtt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 1740
 aatgttcccttccatgttgcgtt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 1800
 ctttttttttccatgttgcgtt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 1860
 aacttgggggtt tggtttttttt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 1920
 tggttgcgttccatgttgcgtt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 1980
 attcaaatgttccatgttgcgtt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 2040
 ctttttttttccatgttgcgtt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 2100
 attcttgcgttccatgttgcgtt tggtttttttt tggtttttttt tggtttttttt tggtttttttt 2160

-continued

-continued

taacagcaat	gttcaggaa	caagagagat	aaccttaaac	tctgactgcc	agttagctgg	6780
gcagaacaga	gocacattt	tcttttttca	aaagccaaatg	ggagaaaat	cactgaattt	6840
tttttctcg	caaggaacat	ccctgagaaaa	gagaatgcgt	ccctgagggt	aggccctctga	6900
aatggccact	ttggggggca	gctgtcttt	tttgtcgaag	ctgttagggat	gaaataagcc	6960
ccagtcctcc	gtagcgctcc	caggcttattt	aggacaaggaa	aattccacc	taataaattt	7020
tggtcagatc	ggttgtctgt	tctcaaacc	tgttcccttg	taagatgtt	tcaatgacaa	7080
tgtgtgc	aaacttcattt	geatattttt	ttttgcctt	gtcctgtgt	cttgcgtat	7140
cgtccctgc	ccatggcctt	tgtatattt	tattacc	tgaagcacgt	gatctctgt	7200
acccacaccc	tattcgtaca	ctccctcccc	tttggaaat	cactataaa	aacttgcgt	7260
tttacgcgt	ttggggggcat	cacggaaac	gcccacatgt	gtatgtccc	ccgaacaccc	7320
aggtttaaa	tttctctttt	ttgtatctt	teccctttat	tctcagactg	ggcgacactt	7380
agggaaaata	gaaaagaact	tacatgacta	tcgggggcag	gttccccca	taagecttgg	7440
ggacaagaat	gagacttcat	ctaaaaaaac	aaacaaacaa	acacaaaagc	aaaaaagaat	7500
acttcacagat	agtttccaa	ttcttaggaa	acaaggcaga	gagaaacaaa	aaacatgtc	7560
caggccggc	gcgggtgtc	acgectgtaa	teccagcact	ttgggaggcc	gaggcgccg	7620
gatcacgagg	tcaggagatt	gagaccatcc	cggtctaaac	gggtaaaccc	cgtctctact	7680
aaaaatcacaa	aaaattagcc	ggggttagt	gccccgcct	gtagtcctag	ctacttggg	7740
ggctgaggca	ggagaatggc	gtgaccccg	gaggccgagc	ttgcagtgg	ccgagatccc	7800
gccaactgc	tccggctcg	gcgacagago	gagacttgcgt	ctcaaaaaaca	aaacaaaaaca	7860
aaacaaaaaaac	aaaaaaaac	atgtccaga	ttttgttca	aagagtatac	cttactctc	7920
atgcctgtaa	tccacact	ttgggaggcc	gaggctggc	gatcacaagg	tcaggagatt	7980
gagaccatcc	ttggctaacac	agtggaaaccc	cgtctctact	gaaaatacaa	aaaaaatagc	8040
caggccgtt	ggcgggtgtc	tgttagtcc	gctactcggg	aggctgtgg	aggagaacag	8100
tgtgaacct	ggaggccgg	tttcgtat	gctgagatgt	ttccagctca	ctccagctg	8160
ggagacagag	cgagacttca	tctcaaaaa	aaaaaaaaaa	agaggatacc	ttactcaatt	8220
tattaaagga	cataaaatgc	tcaaagtaa	tttccttgc	tctgaaaaac	aaaataagga	8280
tcagcaat	tccaaagcaaa	agttttaaa	tttgcgttgc	tttgcgttgc	acatttgcgt	8340
aactcttgc	tttacttgc	tttgcgttgc	ttttgcgttgc	tcatgtttt	tgtacgttt	8400
tcctttatc	caatgtcaca	atctccaa	ttatcgttgc	cttgcgttgc	agagcactg	8460
tcaaagtct	atagctttt	ataaaacctt	ttttggaaat	gattaaacgg	aaacaacaaat	8520
tgtctgtga	ttacaaaatgc	ttcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	8580
tttgattatc	ttcggtgtt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	8640
ctcagacatt	agaattttgc	aaatccata	caattttgg	acatgtatta	ttcacaaaaa	8700
tgtaacctaa	agaagattga	acacgttac	aagtaccta	atgtttcaaa	taatcctgtt	8760
tagctcttc	ttcttgcac	tcttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	8820
ggaaagacaa	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	8880
ttgaaacact	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	8940
actcagaaac	ataaaaagaa	aaaaaaaaaa	cttttataa	gcctttacaa	attttgcacaa	9000
agagcaggat	atgccttac	gagttttttt	tttttttttt	tttttttttt	tttttttttt	9060
aaaaatcat	acaatact	tttgcgttgc	tttttttttt	tttttttttt	tttttttttt	9120
caagattat	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	9180
agtgccgtgg	catgatttca	gcttacttca	acccctgtt	cacgggttca	gcaatttttgc	9240
tgtctcagag	ttcccggtat	ctgggactac	agggtgcgttgc	cacaatgcac	agcttatttt	9300
tgtattttttgc	tttgcgttgc	tttttttttt	tttttttttt	tttttttttt	tttttttttt	9360
atctgaccc	tataattttgc	ttcaatttttgc	tttttttttt	tttgcgttgc	tttgcgttgc	9420
gttgcggcagg	atgcgttgc	gttgcgttgc	tttgcgttgc	tttgcgttgc	tttgcgttgc	9480
ttcaagcaat	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	9540
tgcccgatca	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	9600
tcttgcactt	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	9660
aggcgtgagc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	9720
tttaccaata	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	9780
agttatccat	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	9840
ccaggcttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	9900
cttgatcttcc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	9960
ggctatatttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10020
actctcttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10080
gagccaccac	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10140
ttttaggc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10200
cacggacagg	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10260
tggaccatttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10320
catgggtgtt	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10380
aggtcgccgg	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10440
caaaaaatttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10500
caggataatc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10560
actccagctt	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10620
ttaaaaatatttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10680
tctgttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10740
aggttcaagc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10800
ccatgccttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10860
tggctcttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10920
tacaggcgttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	10980
gagacagatc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	11040
caaccccttcc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	11100
tacaggcccc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	11160
ccatgttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	11220
aaagtgttgc	tttttttttt	tttgcgttgc	tttttttttt	tttgcgttgc	tttgcgttgc	11280

-continued

-continued

aactcctgat	gtcatgataa	caccccgcc	tcccaacatg	ctgggattac	aggcgttata	1440
acagtggcat	attagacta	gttaactttt	ttttttcttt	ttaaaattttt	tttttatttt	1500
aaatggatt	tcgctctgtt	tgccaggctg	gagtgcata	gcatacttc	agctcaactgc	1560
aacctccatc	tcccaagttc	aagtgattct	ctcacctctg	cctcccgagt	agctggact	1620
acagtccagg	gccaccacgc	ccagctaatt	ttcatatttt	tagtacagac	agggtttcac	1680
catgttgacc	aggctggctc	cgaactccctg	acctcaagtg	atccgtcccc	cctcccccgc	1740
ccggcctccc	aaagtgtctgg	gattacagg	gtgagccac	atgcggcc	tattttattc	1800
ttaatattta	attttttag	acagggtgtc	actgagac	aagtgcagtg	gcatactat	1860
ggctcaactgc	agccctgcact	tccgggctg	caatgatct	gcacatcg	cctcccaagt	1920
ttctgggacc	acaggtatat	accggccatgc	gtggcttaat	tttgatattt	ttgttagagac	1980
ggggtatttc	atgttgtcc	aggctgttgc	caaaactgtgt	tgctcaagca	atccacccac	2040
ctcggccctcc	caaagtgtc	ggattacagg	tctgagccat	agtttaacatg	ttttagaagct	2100
tctgggtttt	gtttttgaga	tggagttctca	cccaggctgg	atgtgcagtg	acaccatctt	2160
ggctcaactgc	aacctccacc	tcccaaggta	aaggatctt	cttgccctcac	ttccctgagta	2220
gctgggatta	caggcatgca	ccacccgtgt	caacttaggtt	ttgtatattt	atgttagagaca	2280
gggtttacc	atactgatg	ttggccaggc	tggctctgaa	tttgtgactt	gaaatgatct	2340
tcccacccatc	gcctcccgaa	gtgctggat	atcaggcatg	agccacccgtg	cctggctaga	2400
aacatgtttt	taaaagaaac	aggccagggt	tagtggctaa	cgcttacaat	ccccaggact	2460
tgggaggccg	aggagggtgg	atcaacttgc	gtttaggttgc	gttagggcagc	ctgggcaaca	2520
tgtatgaaacc	atgtctctac	taaaaatgc	aaaaatgc	atgtgtgttgc	gccccgttgc	2580
gtaatccca	ctactcagg	ggctgagaca	gtagaatc	ttgaacccgg	gaggcagagg	2640
tttcagcaac	ccgagatgg	gcccgtcg	tccagccctgg	acaacaaagg	tgaaaactcca	2700
tctcaagaag	aaaaaaaaaa	gacccagata	tagaaatgt	gaaatgcct	gggcatgttg	2760
gttcatgcct	gtatactc	cacttggta	ggctgaa	gggtggatcac	ttaagctc	2820
gagttttaaga	cccgccctgg	gaatgtggcg	aaaccccg	tctacaaaa	ttacttagt	2880
ttgttgtgtc	cacccgtgt	ccggactact	cggtgggttgc	aggtgggggg	attgttgt	2940
cccaaggagc	agagggtgtc	gtgagctgt	atgtgtccac	tgcactocag	cttgggcaac	3000
agagccagac	cttgcataa	aaaaaaaat	tatatgggtt	tataaggaaac	ccagaatatt	3060
aaaaaaat	cttgcataa	acgagcatat	ttagaggact	caacactt	caattttt	3120
agaacttact	acaaaggact	aatagtcaag	actgtgtgttgc	tagcataagg	atagacat	3180
agatagaatt	gcaaggatcg	aaataaacac	atctatgttgc	aattgtattt	tttttattaa	3240
aaatcacaat	aggcaagcat	atgttcaatt	gtttttttgt	tttaattttt	tttttttca	3300
gatggagtct	cactctgtt	cccaggctgg	agttgtacaa	tcttggttca	ctacaactc	3360
caccctgtgg	gttcaagt	ttctctgtcc	tcagccccc	tgtagctgg	attacagg	3420
catgcca	cgcctgtgtc	atttttgtat	tttttagtgc	gatgggggtt	caccatgtt	3480
ggccaggatgg	tctcaatctc	ctgacccatgc	cctcccaaa	tgtgtggatt	acaggcgt	3540
ggccactgac	cttactgtgtc	agttgttttt	tgacaagagt	gccaagacta	ttgaatgtc	3600
aacaagtgtt	gctggataa	ccaaaatagct	acatgc	aatgttgc	ggacccctct	3660
gttataactgt	atataatgt	aaaggccctt	tcttagaccc	agccataat	tggataaaa	3720
caatttctca	taaacttca	gagaatgttgc	tgccagaa	aaaaccaaca	ctttccagta	3780
aataacaa	gaagccagag	tctacaactt	ttgttcatca	caatatttgc	tataacacag	3840
acgtgtat	tgtaatgg	aatatggaca	caagatgttgc	gagaaaagca	gtcaacagaa	3900
acagagaaac	agacccaa	atgttgc	gcaagagaa	actttaaaat	3960	
atctatcaaa	atataatgttgc	ggccgtgtc	ggtgttgc	gcctgttac	ccagcactt	4020
caagactgt	gggtgggttgc	tcacccgttgc	tcaggatgtt	aaagaccaggc	tggccaaat	4080
ggtgaaacc	cgctctact	aataacacaa	aaagttagcc	ggtgtgttgc	catgtgtct	4140
taatccacgc	tacttgc	gtcaaggcgt	gagaatctact	tgaacccagg	aggttgaggt	4200
tgcgtgt	tgatgttgc	ccatgttgc	ccggccctgg	taacccgtgt	agactatctc	4260
aaaaaaaaaa	aaaatgttca	agacttaata	gtcataatgc	gtgaacagac	cggaatctt	4320
agaagagaaa	cagaaacaca	aaaagac	aatgttgc	ttacaaagaa	cacccgt	4380
atctatcaat	ataatatttgc	tacccgttgc	taacaacaa	tttagggcc	ggcacccgt	4440
ctcatgtgt	ataatccacgc	actttggat	ggccggacgg	gttgcgttgc	aggtttagg	4500
atccggatca	tcctgtgttgc	cacccgttgc	cccccccttc	actttaaaa	caaaaaat	4560
ggccggccatg	gtgggtggcg	cctgtgttgc	cagctactcg	ggaggctgt	gcaaggaaat	4620
ggcgttaccc	tgggggccat	agctgtgt	gagccgttgc	cggtccact	cactccagcc	4680
tggggggccat	agcaagactc	catctcaaa	aaaaaaaaaa	aaaaaaaaaa	aagaggggg	4740
ggccagggtca	gtgggttgc	cctgttata	cagccacttgc	ggaggccgg	gtgggttgc	4800
tgcgtgt	caggatgttgc	agaccaggct	ggctaaatgc	gttgcaccc	gttctacta	4860
aaaaatacaa	aatcagggttgc	gctgtgttgc	gcacacccgt	agttccacgt	acttgggggg	4920
ctgaggcagg	agaaatgttgc	gaatctgttgc	gtcaggatgttgc	tcgtgttgc	aatgttgc	4980
cactgtc	cgtccgttgc	cagacgttgc	tcacccatca	aaaaaaaaaa	aaaagaaaaa	5040
aaaaaggaaa	cgttcgtgt	agtttgc	agatcaa	aggat	ataacaaagag	5100
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaagcc	aatacaggaa	atctcaaaat	5160
caggctgtt	taaaatgttgc	aaaatgttgc	tatataatgc	accctgtat	tgtgtatgc	5220
caataa	agataatgc	cacccatata	tcacccatca	aaaaatgttgc	atttgc	5280
caggca	aatctgttgc	gtatgttgc	aaaacaaac	ttgcctaa	aggggatagg	5340
gattgtat	ggaggat	aaggactt	ctagggtgt	ggaatgttgc	tatatccgt	5400
atataatgttgc	gattaaatgttgc	atataatgc	tcatgttgc	tcatgttgc	taccc	5460
tttttttttt	tttttgc	gatgttgc	tttttgc	aggctgttgc	aggctgttgc	5520
catctgggt	cactgttgc	tcacccatca	gggttcaat	gatttgc	tttcaccc	5580
ccgtgt	gttattacat	gtatgttgc	ccatgttgc	cgatgttgc	atttttgt	5640
gagacgggg	ttctccatgt	ttgttgc	gggttgc	gttgc	ccaccc	5700
cctgcctc	cctcccaaa	tggttgc	ataggatgt	gcaactgt	ctggccctcc	5760
aaagaggatt	ttagaagaga	ggaaacccgt	caaaggat	ttttaagct	gttatgt	5820
tccttaggt	agccctgttgc	gaccat	tcaaaatgttgc	tagaattaa	aacaacaa	5880
aactc	cgttgc	tttagatata	taaaaatgttgc	atgttgc	ggccgttgc	5940

-continued

-continued

tagctctggc	taaaaactag	agcaaccctg	gcctgccgt	gctcctgtg	cccagacaac	10560
tccctccaat	tgaaaaggat	gaggggaact	caaagtata	atgtctact	tgaggcagta	10620
agttcgttag	acatatcact	tgccctatta	acatcaagca	tccaaaacc	cagtctgggt	10680
cagtttgcc	cagagtggg	ttttagaaad	acgggttctc	ctgggatct	atacttagcc	10740
cagaatcagt	tgcaaaagcc	aggccatago	gaattgtcc	gcagecaga	tagcagagaa	10800
tctgacggca	gcaggcagaa	ggagccgcto	cattgcagta	agccaagatc	gcgcacattg	10860
cctcattaca	tcaagcatcc	caaaaacc	tctgggtca	tttgc	agtgagg	10920
gtagaacacg	ggttctctg	ggatctatac	ctagcccc	atacg	ttcaaa	10980
ttttggactt	tctctctctc	tctctctata	tatataat	tgtgtgtgt	tgtgtgtgt	11040
tgtgtgtgt	tatacacaca	catacata	atgtgtata	tattgtat	gtgtgtgt	11100
gtatataat	atataat	atttttttt	ttttttttt	ttgagacaga	gtttcactt	11160
tgttgacaaa	gttggagtgc	aatggcacga	tttgcgtca	ctgcaac	tcgcctctgg	11220
gttcaagtga	ttctctgtc	tcagecctc	gagtagtgg	gagtagcagg	gtgcacacc	11280
gccccccgc	aattttgtca	tttttagat	agatgggg	tcacatgtt	agccaggat	11340
gtctcgatct	cgtacatcc	tgatctgc	acctcggc	cccaaa	tggttata	11400
ggcatgaccc	accatacc	gccttcatt	ttatattt	agtagagaca	ggatttca	11460
atgtggcc	ggctggctc	gaactctg	cetcagg	tgcccaact	tcagcag	11520
ccaaagtgc	ggaatttac	gcgtgagca	ccgcctgc	aattttttt	attttagt	11580
gagacagggt	ttcaccatgt	tgccagg	ggtgttga	ac	atgtgat	11640
tcacccatcg	cctccaaa	tgctgaaat	acatgag	accatgcca	gcctctg	11700
caaagttct	ttaattgca	agtagagat	attgtttt	tttgcgtt	tttgcgtt	11760
gatggagtct	cgctgtgtca	cccaggctg	agtg	tgatctca	gtca	11820
acctccac	cccaaggat	agtgttctc	ctgc	ctc	ctgc	11880
ca	gatgtgc	ccacatgc	ccgtt	ttttgtt	tagtagagat	11940
cgtgttagcc	aggatgg	cgatctt	ac	ctcgtgtat	ctgc	12000
aagtgcgtgg	attacaggca	caagccac	tgcccag	gagat	tttca	12060
gatatttco	gataatta	tgaacaca	cacacac	atcta	acata	12120
ta	acacatgt	atacata	aaagat	atgg	ctgtttt	12180
tcctgc	ccttcc	catcttta	tctatgg	agcc	ctct	12240
ttgaatcg	ttac	tttgcac	gttgc	tttgc	gtcatt	12300
gctagtca	ttggactgt	gccaagg	aggat	gtat	ttcagg	12360
aaatgtgt	agctgtgtc	ttgtat	aaatcaat	tttgcgtt	tttgcgtt	12420
aggcctg	agctcag	agccaagg	cttgc	ctcgt	gtcc	12480
ttgggcttca	tttatca	atacgca	gag	ctat	ctcgt	12540
gcaacatgt	ctgtgttgc	acacagat	gca	gat	ggat	12600
gagaagatgg	tcgttccc	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	12660
atcaaaagg	cttgcagg	tttgcgtt	aa	agcgttca	gttgc	12720
aagtgtaaa	gcctctgtt	tgagatgtt	tctgt	gttgc	tttgcgtt	12780
gaaatgtaca	ttggactgt	gaccagg	actat	tttgcgtt	tttgcgtt	12840
ttgaaacatgc	cacattact	agatc	tttgcgtt	tttgcgtt	tttgcgtt	12900
tgtgttttta	gataaaa	ctacc	tttgcgtt	tttgcgtt	tttgcgtt	12960
ccaagg	gtctaaatt	acgttccat	gg	tttgcgtt	tttgcgtt	13020
acaagatgg	gttggag	taactgt	actgt	gttgc	tttgcgtt	13080
gagaaatgg	ctatgtt	ctta	gggggg	ttttttt	aaactgt	13140
tgtcgtac	actatgt	aggat	tttgcgtt	tttgcgtt	tttgcgtt	13200
gtgtgtgt	agaatc	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13260
caatcttc	ttggc	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13320
aagtccatc	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13380
gaacgggttgc	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13440
cccaactgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13500
catttatgt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13560
gttctactt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13620
cttccac	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13680
ttttttttt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13740
cgatctcg	ttactgt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13800
tcccaatgt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13860
gttagataaa	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13920
tttccgc	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	13980
catgtccat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14040
accctatgt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14100
gcctggccat	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14160
acttctgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14220
ctaaaa	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14280
aagaaatgt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14340
aagttttctt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14400
ctgcca	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14460
tgaagatgt	tcttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14520
gcagccttgc	tcttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14580
ctaatca	cgttgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14640
gagtgc	actgt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14700
tccctctgt	cac	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14760
ttttgtatt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14820
tcctgtatct	agg	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14880
gcctcgtgc	ccggc	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	14940
gatgggggtc	tcactgtt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	15000
gcctcgtgc	cccaaaatgt	tttgcgtt	tttgcgtt	tttgcgtt	tttgcgtt	15060

-continued

tttctgagtc	aaaggacaag	aaaataaaca	gcctagaaaag	acagacagcc	ttgatTTTT	15120
tatTTTtagt	tttcattgg	tttgattttt	cttgagacag	agtctcac	tgtcgcccag	15180
gctggagtgc	agtggcacga	tctgactca	ctgcaacc	tgcetctgg	gttcaagcag	15240
tttcctcgcc	tgcctcagcc	tttcgttgg	cctcaatga	tccgcctgg	tcagcctccc	15300
agagtgtgg	gattacagcc	atgagccact	gcatccagcc	gacagacagc	cttttagacta	15360
tagccccctt	actccagcc	caceccctaco	cccatcagca	aaggccaagt	ggaaaagct	15420
gatctccaaa	cotttgttag	ctgtaatgaa	gtactcctt	ccttccccac	cagggtgatg	15480
tagaagaggc	ctagttatcc	actcaggct	ttcatccata	atcaaggcca	ccttccccaa	15540
caatgtcggt	ggaggcccac	tecctctcc	agecgggtgt	gtatcagtga	atgcctaata	15600
gagaaactgg	actccacac	ctgcgcagca	gtaaagaagga	gcccgtccaa	caccgggtgc	15660
aatggaggcc	aagtggaaaa	cctggactt	cadacccctc	tggtaaaaat	aaggtagcat	15720
ccgcacttcc	cccacagaag	caatgttga	ggaggccttc	tgaacacaa	gctataaagt	15780
aagatccaga	ggctgggcgc	agtggctcat	gcctgtatc	cctgcactt	gggaggctga	15840
gggtggagga	ttgtcttaagc	ccaggagct	aagaccagct	tgggcaacat	acggagac	15900
tgctgtcaca	aaaaaaacaa	aaacaaaaaa	attagctggg	catgtatg	cctgcctgt	15960
gtctctagtc	cttgggaggc	tgatgtgg	ggatgttgg	aggccaaag	gttggactg	16020
cagtgtgtca	tgacccact	gcactccago	ctgagtgaca	aggtgagacc	ctatccaaa	16080
acattaaata	agatccagag	tcttataaca	tgatgtgca	aatgtccagg	tcttaaagt	16140
atttgacccg	gaaaccagaa	attttgaat	agaaaggata	atctatatac	actaacacca	16200
aaatgacaca	gatattagaa	ttatgtaaa	aaatgttga	atgtccatc	ataaaaatgc	16260
ttcaatgtat	ataacgtgtt	caaatacaaa	catacttgg	acaaatgaaa	aatggaaag	16320
tctggaaat	ataatgtctc	aacagatca	tacataaaag	agaaccaa	atggaaattt	16380
ggaataaaaa	atatacatgt	tgaattttaa	aactcaatg	ataagctcg	caacagaatg	16440
gagaggacag	agggaaagaa	gagttgttgg	acagaaaaac	acaaatttac	caatctgaac	16500
actagggaaa	aaaaatagaa	tggaaaaaa	aaaaaaacaga	gcctcagg	catgtgagac	16560
aataacataa	gatctaatac	tcatgtcatt	ggactatagac	aaggagaaga	aatattaaaa	16620
gaacaaatag	ccgaaatatttgc	cccaaaat	gcaaaagaca	aactgacaga	ttcacatagc	16680
tgagcaacc	tcaaaacag	tatggccat	agaaaaacaca	tcataaaac	ctcaacat	16740
actttttttt	tttttttttt	tttgagatgg	agtctccctc	tgctgtccag	gctggagcac	16800
agtggcacaa	tctcagttcg	ctgcaacc	cgccctctgg	gttcaagcga	ttctcttgcc	16860
tcagccttcc	aagttagctg	gactatagac	gtgttaccac	atgcctgg	aattttttgt	16920
atttttagta	gagatgggg	tttccatgt	tgcccaggct	ggtctcgaaa	tcctgactc	16980
aggtgtctg	ttcgccctcg	actccaaag	ttgtggatt	acaggtgtga	gccaccgc	17040
cccagccaca	aattctgaa	agcagccaga	gagaaatgat	acattaccta	tagggggaaa	17100
acaaatcaa	tgtatgtgaa	tttctcgtaa	ggcaccatgg	aggccaggag	gaagtggaaac	17160
aacatttttt	tttttagaca	agagtcttac	tctttcacc	agggtggatg	gcagtggcgc	17220
gatcttggct	cactgcaacc	tccgctccc	aggttcaac	aattctctg	cctcagcctc	17280
ccaaatgtct	ggggctacag	gcgcgcacc	atgcctgg	aattttgtt	tttttagtaa	17340
agacaggagt	ttcaccat	tggccaggat	gttctcgaa	tcatgtactc	gttatctg	17400
caccttggcc	tcccaaaatgt	ctgggattac	aggcctgagc	cactgc	agctggaaaca	17460
tttttagat	gotgaaagaa	aagaactatc	aacttggaa	cccatactg	gtgaacacat	17520
ccttcaggaa	ttggggggaa	gggtgttggg	ggaggaaatc	atgacattt	tacatgaggg	17580
aaaaactaa	ggaatttgtc	actggcagat	ttactttata	gggattacta	aaggcaattt	17640
tctaaacaaa	ggggatgtt	acaaaaggat	tcttggaa	tcaggcagg	aaaaaggaca	17700
aaaagagaca	acatacaggt	caataata	gattgttctt	ttctcttgc	gtatctaa	17760
ttatgttca	caaggtaagc	aaaaattac	tttgtctaa	ttgtgttgc	tatatgtaga	17820
ggaatatactt	aagatgttga	cattataat	ggagatgtaa	aggacacaga	aaaaggttt	17880
tttcttttatt	tgtatgttgg	tagaatttgc	cagaggat	gaggatagt	tttctac	17940
cactcaact	ggtaaaatgt	caacacc	agactgtat	agtttact	actccag	18000
ggggcagaca	gcgagac	gcctaaaa	ggaaagaaa	aaaagaaaa	aagaaaagaa	18060
aaaggaacac	atacaaact	taatgtgt	cgttaccaaa	aacagactt	aaaaataatg	18120
gaacactggc	aggcgcgtgt	gtctatgc	gtatcacag	agtttggg	ggcaaggca	18180
gggtggatcac	ctgagatc	gagttcgaga	ccacgc	caacatgtt	aaacccctc	18240
tctactaaaa	ataaaaaat	taggtgg	gtgttgg	acggctgtt	acccagcact	18300
atgggaggcc	gagggtgg	gatcat	aaagat	gagaccat	tggccaa	18360
ggtgaaaccc	ccgccttact	aaaaaaat	tttttttttgc	tttttttttgc	tttttttttgc	18420
tgtaaaacca	gcacttgg	ggcttgg	agagatc	tttgc	tttgc	18480
tttgc	tttgc	tttgc	tttgc	tttgc	tttgc	18540
tctcaaaaa	aaaacaaaa	ccccaaaaaa	caaaacaa	aaaacttata	gacagaaat	18600
caacaggcgt	gagggtgtt	ctcacact	taatggata	ttagattaa	cactacc	18660
agccaaacagg	atgtatgt	cacttata	acatcaaa	aaaaaaac	acattttt	18720
taggcacca	tgaaacatgt	atcaagat	accatat	tttttttataa	acaaacatca	18780
acaaatataa	aagaattgaa	atcatacaga	gtgtatgac	aaaatggact	caaacttata	18840
attaataaca	caaaacaa	ctccaggaa	atcttcaaa	acttggaaac	taaataacac	18900
atttctaaat	aatctat	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	18960
ctgtatgt	aaaatataa	tttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	19020
ggaagggaaat	taatagact	aatgttca	tttttttttgc	tttttttttgc	tttttttttgc	19080
ataataacaag	cttccac	aggaaattt	aaaagaacaa	aataaaac	aaacaaacag	19140
aaagaaggaa	aagaagagaa	atcttataa	tttgc	tttgc	tttgc	19200
acaaaacaga	aaaactgg	tttgc	tttgc	tttgc	tttgc	19260
ttaatttgc	acagcttctc	tttgc	tttgc	tttgc	tttgc	19320
agoctcaacc	tgccggg	tttgc	tttgc	tttgc	tttgc	19380
acagggtcat	ccat	tttgc	tttgc	tttgc	tttgc	19440
catgttgc	agactgg	tttgc	tttgc	tttgc	tttgc	19500
caaaatgtc	agattacaga	tttgc	tttgc	tttgc	tttgc	19560
aaggctgacc	aaaaaagaga	tttgc	tttgc	tttgc	tttgc	19620

-continued

cactttggga	ggccgaggca	gatggatcac	tttagctca	gagttcaaga	ccagcttggg	19680
caacatcatca	aaaaaccac	cctgcctcta	caaaaataac	aaaattaac	caggcacgg	19740
ggcacacgtg	tagtccggc	tacttgggg	gctgagggtgg	gaggatcca	taaacctggg	19800
aggccggagg	tgcagtggc	cgagatcatg	ccactgact	ccagcttggg	cagcagagta	19860
agacttcata	tcaaaaaaaaaa	aaaaggattat	tttatagag	acagggtctt	gctatgtgc	19920
catacgatc	tttgaactcc	tggggccaa	tgatccctca	gcctcagctc	tcccaaagt	19980
ctgggattac	aggcatgaac	cactgtggc	accccataca	tttctatcat	caggaataaaa	20040
acaagagota	tcaactacaga	ctctcgat	aacgatagga	taataaggga	atttttatgt	20100
attgtttatt	tatgttattt	tttttgaga	cagagtctcg	ctctgtcacc	aggctggagt	20160
atgggtgggt	gatctcggt	cactacaat	tctgacttcc	ttgttcaagt	gatactctg	20220
cetcagcc	ccaaatgt	gggattacag	ccacggccca	ccacggccca	ctaactttt	20280
tatTTTATG	agagacgggt	tttccatgt	tggccaggat	ggtcttgaac	tcctgactc	20340
gtgatctggc	tgcttggact	ccccaaagtac	tgagattaca	ggcatgagcc	accgtgcctg	20400
gccctggaaa	ttttataaaac	acacatttt	tttatatgtt	actaacacac	atttgacaac	20460
tttagatgaaa	ttgactaaat	cctttaaaaa	taaaaactat	tacaacatac	ccgatatgaa	20520
gtatataatt	tcaatgtt	tataatcaat	aaggacattt	aatttgtat	ttaaaactt	20580
acccccaaga	aatctttact	ccagatgatt	tcagtgaga	attctatcaa	acgtctaaag	20640
aattaacacc	gatttacaa	atctgtatca	gaaaatagaa	gagatgaac	atttcctaag	20700
tcattttatg	aggccatct	tacccat	aaaaccagac	aaggaaaaata	caaaaaaagat	20760
gactacagac	caatgaccc	taataagaa	acagacacaa	aaatcccaa	caaaacatca	20820
acaatagggc	caggcacgg	gcctatgc	tgtatccca	gcactttggg	aggctgaagt	20880
gggtggatca	cctcaagg	ggagtggag	accagctga	ccaaatctgt	gaaaccctgt	20940
ctctactaaa	aaaaaaaat	aatgttgg	gctgttgggt	gctgttgcgt	aatcccaagt	21000
actccggagg	ctgaggcagg	aaatcaact	agaactcggt	agggtggat	tgcagagac	21060
caagatcg	ccactgact	ccagctgac	tctgttctca	taaataata	aataaataaa	21120
taaataacaa	atagaggcc	ggcacgg	ctcacgcct	taatccca	actttggag	21180
gccaaggcag	ggggatcacc	tgaggtgg	agttcgagac	cagctgacc	aatatggaga	21240
aaccccatcc	ctactaaaa	tacaaaattt	ggcaggcaca	gtggctat	cctgtatcc	21300
cagcacttt	ggggcctag	gtgggggat	cagggtca	gaagttcaaa	accagctgg	21360
ccaaacatggc	aaaacccctgt	ctctactaaa	aataaaaaaa	ttaaccagc	atggtggca	21420
acacctgtaa	tccctcgat	acaggaggt	gaggcaggag	aattgttga	acccggggagg	21480
ggggagggtt	atggggccaa	gttcacca	tttgcactca	gcttgggtt	cagagcgaga	21540
cttcatctca	aaaataaaat	aaaataaaa	atacaatata	aaaatttggc	gggcattgt	21600
gtgatctgc	gtatccca	ctactcggt	gactgaggca	ggagaattgc	ttgaacctgg	21660
ggggcactagg	ttggcgatgg	cggagattt	gttattgcac	tctagcttgg	gcaacaagag	21720
ggaaactccg	tctccaaaga	aaaaaaat	ataagaattt	agcaatata	ttaaaagaaatt	21780
atataactatg	accagggtgg	gattattca	gggatgtca	gctaagtca	tatccaaaa	21840
tcaatcaatg	aaaaaaaatc	aatcaatgt	ttccaccctg	ttaacaggct	aaagaaaaat	21900
caattcagac	agagaaaaac	cttgcacaaa	tttgcattacc	atttcataata	aaaactctca	21960
ggaaactaga	aaaaaaagg	aatttctca	atttgcata	gtatctaca	aaaacatca	22020
actagcatca	tacttcacag	tgagagact	aatgtttcc	ccctgactga	acacaaggca	22080
agcatgtca	ctctcacc	tcttattca	cacagtctg	gaaattgtag	tctgttaat	22140
aaqaatagga	aaagaatttt	aaaacatca	atattggaa	gaaagatata	aggcggtct	22200
tataatggc	tgatggctt	tatagaaaa	tttgcataaa	ctaaaaaaa	aaaaaaagct	22260
agaaaccatg	agtgtatca	gcaaggtac	aatacaat	caacatata	atatacaat	22320
atatatattt	gcaatgaaa	cacggacatc	aaatttaa	acataatatt	tataatcaca	22380
aaaaaaaggaa	atattttatg	atgtatcc	taaaacatgt	acaagactca	tatgtgaaa	22440
actagaaaaat	gtgttggaa	gaaatcaag	aaatctca	tagacagata	tactgtttc	22500
atgacttca	agactcatca	cgttgcatt	gtcaatttctc	cctaaattaa	tatacaatgt	22560
tagtgcatt	ctatccaaa	tcccgat	attttgtt	ataaaagattt	tttataaaatg	22620
ataggaaaa	ggcaaaaaaa	cttagatgt	tgaacataat	ttttttttt	tttttttttga	22680
ggccagatct	cactctgt	cccagactgg	accacagtgg	cgtgtatcaca	gctcatttgc	22740
gccttgatcat	cttgggctca	gttatacc	ctacctcggt	ctcccaatgt	actggacca	22800
cagggttgt	ccaccacg	tagttaattt	cctttttttt	ttttttggaa	cgaggctt	22860
ctctgtggcc	cagggttgt	tgcaatggca	tgacccctgc	tcactgca	ctccgcctcc	22920
cagggttcacg	tcttgcct	gtttcagct	cccgagtt	tgggactaca	ggcactcgcc	22980
accacgcctg	gtatatttt	ttttgtt	tagtagac	gagggttac	cgtgttgc	23040
aggatgttct	ccatctctg	accttgc	ccggccatct	cgccctccca	aagtgttgg	23100
attacaggcg	tgagccac	agccccggcc	aggcccgat	aattttaa	aattttttt	23160
tatacagaca	aggccctct	gtgttggccca	gggtgtttt	atttctggat	tcaagcaatc	23220
ctcccccac	aggcccttca	agtgttgc	ttacaggat	gagccactgt	gcccagatca	23280
aaataatattt	cattaaaaaa	actttttt	agacaagg	ttgtctgtc	acccaggctg	23340
gagtgcacgt	gcacatctt	ggcccaact	agccctccga	gtagttca	ctgttagacat	23400
gtggccacac	acccacat	ttttcaat	ttttgttgc	acgggggtt	gcccattgt	23460
ccaggctgt	tttgcactt	ttgtatca	caatttgc	aacttggct	cccaatgt	23520
tggactaca	ggcgtgcact	accacacgt	gcctaaaaat	tatttttaa	aagaaggat	23580
agccccggcc	agtgttgcac	gcctgtat	ccagactt	gggagggtt	ggccgggtt	23640
tcacttcagg	attttgc	cagttggcc	aacatgttgc	accacatct	ctactaaaa	23700
tacaaaaat	tagccagg	tgatggc	tgccctgtat	tccagat	cggttaggt	23760
aggcaggaga	gttgcatt	tctgggg	agaggttgc	gtgagttgc	gtcgctgc	23820
tgcactccat	gactccat	ctgggcaaca	agagcgat	tccatctca	aaataaaaaa	23880
ataaaaaagg	ctgggcatt	tggctac	ctgtatccc	agcacttgg	gagggttgc	23940
caggcagatc	acgaggatc	gagttcaaga	ccagccgtac	caacatgttgc	aaacccatc	24000
tctactaaa	ataaaaaat	tagccggca	tggtggcgt	cacccgtat	cccagctact	24060
taggaggctg	aggcaggaga	atagcttgc	ccggggagg	ggagggttgc	gtgaaacc	24120
actgcacat	tgcacacc	cctgggctca	agagcgac	tccatctca	aaaaaaa	24180

-continued

aaaaagaagat	taaagtggga	ggaatcactg	cctaattca	agaatacagc	tatagtaatt	24240
atgactgtgg	ggttatgttg	gatggaaaga	cacatcaatc	aatggaaagac	aatagaggac	24300
ccagaaatac	cccatgcaat	acagtggagg	aaggatagt	ttgtcaacaa	atggtactga	24360
agcaaaatgg	tgttcatagg	aaaaacaaa	aaacaactcc	acagaaaatc	gctttctgtg	24420
atctgataca	gcctgcctag	aagcagaagt	tctggctgcc	catcacactg	tttttagtaa	24480
acctagagtt	ttatgtgaa	tatttgagt	ggcccaggcg	eggtggotca	cgccctgtat	24540
cccagcaactt	tgggaggccg	aggcagggt	atcatgcgg	caggagatcg	agaccatcc	24600
ggotaacacg	gtgaaacacc	ctctctacta	aaaaatacaaa	aaattagcca	agtgtggtg	24660
cggggectg	tagtcccage	tacttgggg	gctgaggcg	gagaatca	tgaacccggg	24720
aggccggagg	tgcagtgagc	caagatcgt	ccactgca	ccagctgga	taacagagca	24780
agatcttgtt	aaaaaaaaaa	aattataaa	gatggaaaat	atggagagcc	accaaggaa	24840
caataattct	gcatctagac	tgccctcaga	cttggcaga	cttgaactgc	aacatcaat	24900
cttccctggg	tcttttagctt	gctggccagt	tctgcagatt	cagatttgc	agccccacca	24960
ataagacaat	tcccttattt	gccttagt	tgatgaaatc	agagatctag	gctgggtgca	25020
gccccatcg	cttgcatacc	tagacttgg	ggaggctgag	acaggcggt	cacttggggc	25080
caggagtc	agaccacgc	gaccaacatg	acaataccc	atctctgct	aaaatacaaa	25140
aaaatttagcc	agatatggt	gcatgcaact	gtaatcccg	ctattttga	ggctgaggta	25200
caagaatcat	ttgaacccag	gaggcagagg	ttgcatgaa	gtctgctgaa	aaaaaaaaaa	25260
gtaaaataaa	aatgttaaa	tgaccattaa	aaaaaaaagg	aaccaaccta	aacccatgc	25320
cttatacata	aaatgttac	aatggaaaca	ggtttaatg	taaactgtaa	agtataaga	25380
tgttttagaa	acaagataga	agaaaaatcat	caggacatg	gagtcttcc	ccagggtctca	25440
ggggaaaagg	cttcagacat	atgcacaaa	caagaacccat	aaaagaaaaa	atataatacat	25500
aattggactt	catccaaattt	acaaactttt	gctctgtgaa	agacccccc	aaggaggatg	25560
aaagacaaac	tacaaagtga	gagaaaaat	ttgcaaaacca	tatatctgac	aaaggactta	25620
tattttagat	atataaaag	aacactcaag	gcccggcatt	gtggctact	cctgtataac	25680
cagcacatgg	ggaggcacaag	gcaggcagat	cacttgaggc	cgggagttt	agaccacgc	25740
gcoocatatt	gtaaaaaccc	atctctgtca	aaaatacaaa	atttagotgg	gcacgggtt	25800
gcatgcctgt	aatcccaatg	attcaagagg	ctgaggcagc	agaatcac	gaacccagg	25860
ggcagaggtt	gcagggagcc	gagattgcac	cactgca	cagcctggc	tacagagct	25920
gactctgtct	aaaaaaaaaa	aaaaaaaaaa	aagggtggc	gtgggtggc	aggcctggaa	25980
tcccgacatt	ttggggggc	gaggggggc	gatcagcgg	tcaggagatt	gagaccatcc	26040
tggctaacac	ggtgaaaccc	catctctact	aaaatacaaa	aaaataggc	ttgggtgtgg	26100
ggcagggtct	tgtgttccca	gctactcagg	aggctgaggc	gggagaatga	catgaatctg	26160
ggaggcggag	cttcgcgcga	gagccactgc	attccagtc	agcctgggt	gcagagca	26220
actctgtctc	aaaaaaaaaa	caacaaaaac	aaacactcaa	atctcaacag	26280	
taagaaaaac	accagataat	cccttagaa	aatgtc	aaaaaaaactt	acttaagata	26340
tttccccaaa	gagaatatac	agaaggc	taaacacgt	aaagaatgtt	aacataacta	26400
tccatttaga	aaattaagac	catgtgaga	taccattaca	tacttggag	aatagctaaa	26460
ataaaaaat	aggggccagg	gttgcgttgc	gtgcctgtt	tcccgact	ttggggaggct	26520
gaggcaggca	gatcactga	gtccaggagt	ttggagacc	tctggccaa	atagtga	26580
cctgtctct	ctaaaaatac	aaaaatagg	ttggcatgat	ggtgtgtc	acctgtata	26640
ccagctactt	ggggggctga	gataggagaa	ttgcttga	ctggggagg	gaggctgc	26700
tgagccgaga	tcatgcact	gtactgc	cggggtgaca	gagtgaact	ctgtctca	26760
aaataataat	aaaaataaaa	aaataataaa	acagaatacc	acatgtgaa	ggatgtggac	26820
aattgaatct	ttttttttt	ttttttttt	ctgagataga	gtcttgc	gtcaccac	26880
ctggaggtca	gtggcacat	cttggctca	tgcaac	gcctctgg	ttgaagca	26940
tccctctct	cagccctccca	actactgtgg	attacaggc	tgccacca	cacttggc	27000
attttgcatt	tttttagaa	gtatgggtt	accatattgg	ccaggctgt	ctcgaaact	27060
tgacctcg	atctccgc	ctcagcc	caaagtctg	gtattacagg	cgtgagcc	27120
catgccgc	caacaattga	atcttcata	cattactgt	ggaaatgt	aatggaaat	27180
caactctgt	actaatttt	gtatttat	tttatttt	ttaagagaca	aaaacttgg	27240
gtgcactgtt	gcccata	cttactgc	cctccaa	ctgggtc	gcaatcc	27300
caacctccac	caccgagtag	ctgggactac	aggtgtgc	cccaaaatgt	ggctaa	27360
ttaattttta	atttttgtac	acatagggc	taattacgt	gcccaggct	gtcttgc	27420
cctgagctca	aatgtatcc	ctgccttgc	ctcccagt	gttggat	caggcatg	27480
tcaccatgc	cggccctact	atttgttta	aaacactaa	atcccttac	catatgaccc	27540
agaaatttt	ctacttggca	tttattccag	ataaatgaa	atttatgtt	acacgaa	27600
ctgtatcga	tttttataa	cacatttatt	tgtaatagcc	aaaacctgg	aaacaacaaa	27660
atgaccatca	aatatgtaa	tggcttaaa	aaccatgt	tatccatata	acatgtatcg	27720
agttaaatgg	gagcccccca	aaagacatgt	acatata	actctgt	gtatgtatgt	27780
gacgttgc	aatagcgtt	ttgaagccgg	ggccgggtgc	tcacgttgc	aatcccgac	27840
cttggggagg	ccttaggtgg	cagatcacct	gaggctggg	gttcgagacc	agcctgacca	27900
acttggagaa	accccgctc	cactaaaaat	acaaaattag	cccggtgtt	tggcacatgc	27960
ctgtatccc	actactcg	gtctggcagg	agaactact	gaacccagg	ggcaaaagg	28020
gtgtatgagcc	gagatcg	cattgc	cagcctgt	aacaagag	aaactccatc	28080
tcgaaaaaaa	aaaaataaa	aatagggtt	ttgcagtg	gattaatgt	aggatctcg	28140
gatgatatac	acttagatta	tctgggtgg	ccctaaatcc	aatgaca	gtctttttt	28200
cccccaagca	taatccct	tggtgtgg	accacagg	tctggcc	cacctgg	28260
aattttttt	tttttttgc	acaaaatgc	cctgcacat	aggctgg	ccagtggt	28320
tatctgggt	cattgc	teccgttcc	aggttca	aattctc	cctcagc	28380
cccgatgt	gggat	ccatggcc	ctat	ttttt	tat	28440
agagacgggg	ttttgc	ttggccat	ttggctt	gaaattat	ttc	28500
gctgccc	gtctcc	gtgctgg	tacagg	ggccaccat	cccac	28560
atttgcatt	tttttttgc	acaaaatgc	ctcactatgt	agggacagg	ctcactatgt	28620
gttctca	acac	tcctgg	tttgc	tttgc	tttgc	28680
acaaagggt	gtccacgt	tcacccata	agtgtcc	taagagac	gagagg	28740

-continued

-continued

cgccaccatg	ccggcataat	tttttgtatt	tttagtagag	acggggtttc	accatgttag	33360
ccaggatggt	ttcaatctcc	tgacccgtg	atccacccgc	cttgcctcc	caaagtgtcg	33420
ggattacagg	cgtgagccac	cgtgoctgg	ctcttttact	tttttaaatg	tgaatttttag	33480
aaaattttaa	attacacatg	tggccgggg	acagtggctc	atgcctgtaa	tccttagact	33540
ttgggaggtc	aagggtggtg	gatcaactg	agtcagggtt	ttgaaagcag	cgtggccaac	33600
atgacaaaac	cccgctcta	ctgaaaatac	aaaaatttgc	tgggtgtgg	agcgggtgtc	33660
tgtatccca	gtaactcggg	agactgaggg	aggagaatcg	cttgaaccca	ggagacggag	33720
gttgcgtga	gocgagattt	cttcaactgc	cttcagttctg	gatgcagag	agagactcca	33780
tgtcaaata	aataaaataa	taaataaata	aataaaataa	taaataaata	aaatagtgc	33840
tctgggttag	ggcgtggct	aatgtctgt	atcccagcac	tttgagggc	caagacagga	33900
gcatcggtt	ggccggaggg	ttcaagacca	gtttggggca	cataggaga	ccccatctt	33960
acaaaataaa	aaaatttgc	caagtgtgt	ggcacatgtc	catggttcca	actcttggg	34020
aggtgggggg	atcaatttgg	cctgggggtt	cgaggctgt	gtgtgtcatg	attgcacatc	34080
tgtatccca	cctggggcaac	agatgtggat	cctgtctca	aaaataaaat	aaaatagtta	34140
ctcccaatgt	catgtttttc	tattttttac	ttcttcttaa	cacttattaa	taactgcatt	34200
agacagaaaa	caaacaggta	aacaatgt	ctacatgtat	acaaatgtat	caaagttag	34260
aagaaaataa	agcaaggctt	aaaggagggg	gctactttgg	atgggggttc	aaaaagggcc	34320
tctttgagaa	ggtgacttct	aatgtgttat	ttaggtgtat	aaaatggcac	caatcatgca	34380
atgattttag	ggaaatttccaa	gcagccagaa	cagcttgc	gagtccaaa	gtgggacggg	34440
aattcactta	ttagaggaaac	agaaaggaa	tttgatttgc	tagatccag	tgagcaaggg	34500
ggtcctggg	cacgagtc	atcatttga	cattatgttc	cattgaaaag	agtgtatatt	34560
ttagtgtaa	tgcatattgg	aactgtgt	gtgacatgtt	ttatgttttt	tttttttttt	34620
tttttttgg	atggatgtt	ccttgggtt	cccaggctgg	agtgcaatgg	cttgatctcg	34680
gctcaccgc	acctctgcct	ccttgggttca	agtgtatctc	ctgccttcg	cttccaagta	34740
gctgggat	ttggcatgca	tcaccactc	cagctaattt	tgtattttta	gttaggtgg	34800
ggtttctca	ttgtgttgc	gctgggtcg	aactctgtgc	tttaggtgtat	ccgcctgcct	34860
cagccctccca	aaatgtggg	attacaggaa	tgagccacgg	cgccccggcc	tacccatgg	34920
aatccgcggc	cctcgtctc	ccaaaaatgt	gggatttacaa	gtgtgagcc	ccgcacccag	34980
ccactgtatct	atgttttaa	gaatttact	tttgctgtcg	tatggatgt	aatgggggg	35040
ggagaaaaaa	ttggaaatagg	ccatccagaa	ggctgggtca	tttgccttgg	tgatggatgg	35100
caigtgttgg	accaagatgg	tgacagcaag	atagaagaaa	ttgggtgtatt	caggatacat	35160
atggggggca	ttttgttgc	actggccgt	agggataaga	ggaatgagag	aaagaaaaat	35220
caagaatgtac	ttctggctgg	gtgcgggtt	tcacgcgtat	aatccagca	ctttgggggg	35280
ctgaagcagg	ttggatcacct	gaggttgc	gtttgagacc	agcctgtacca	atatgtatgg	35340
acctctatcc	tagttaataa	acaaaaatgt	gtctggcatg	ttggcaagca	cctgtatcc	35400
cagctactcg	ggagggtcg	acaggagaat	cgcttgc	cagggggtgg	agggtgtat	35460
gagccggatgt	tgcgccattt	cactccagct	ttggcaacaa	gagcaaaact	ccatctcaaa	35520
aaaaagaaaaaa	aaaaagact	tctaaatctt	gtacttgcgt	aactgggtag	gtggtcaagc	35580
cattttctga	gatggccaaa	actgggggg	caigttctgt	ttgtgttgg	ttgtttagaa	35640
ccagatata	aaatgtgtgt	ggacatgtat	actttgtat	gtctttagaa	caagcagggt	35700
gggaaggagg	cagttgata	tgagtctgg	gtctggggaa	gagatgtat	agctgtatgt	35760
acggatagaa	gtcaatctg	gatgttgc	ggcgccggcc	ctcacccctg	tagtcccac	35820
actttgggg	gocgagggc	gccaatttgc	tgaggtcg	agtttgcgc	cagccgtgtc	35880
aatatgtgt	aaactctgt	ctactaaaaat	tacaaaaataa	tagccaggca	ttgtggcagg	35940
cacttgcgt	cccgactact	ttgggggtgt	agacacaaga	atcgcttgc	cccgaaaggc	36000
agagggttgc	gtgagctgc	atcgccca	tgacttgc	cctgggtgc	agagtgtat	36060
tccatctca	aaaataaaaa	aataataaa	aataaaaaaa	aatctggatg	tcatctgt	36120
atggatgttgg	gggatcaact	agggaaatgg	gggaggaaga	atgttcttgc	cagagcttgc	36180
agatcttgc	agtccctccc	aatttgc	ttctgtctt	tttcttgc	ttgtacatgt	36240
tacactgcaca	gagttttatgt	ctctactcg	tttcatattt	cacccatca	ctgtcacatc	36300
tcttgcgtcc	agaggatcc	atggacaaac	tttgcgtat	tttgcgtat	tttgcgtat	36360
ggcagtatttgc	ttttgttgc	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	36420
tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	tttttttttt	36480
agttgggtat	ctcagtcac	tgcaacctcc	aactccgt	tttgcgtat	tttgcgtat	36540
gggggttccat	catgttggcc	ttggatgtt	tttgcgtat	tttgcgtat	tttgcgtat	36600
ccggctccca	aaatgtgtgg	attacaggcg	tgagccacgg	ccgcctggct	agaataat	36660
tttttttaca	taggttttac	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	36720
ataagacttt	ttttaaagct	gatactggcc	atggatttgc	tttgcgtat	tttgcgtat	36780
tgggtgaaat	ttctctcccc	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	36840
gaaaatgtgaca	tttttttttt	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	36900
aatatgtggc	cagttttcc	aacggcttta	tttgcgtat	tttgcgtat	tttgcgtat	36960
aggaaaagca	gaccatttca	gtcaaaggct	tttgcgtat	tttgcgtat	tttgcgtat	37020
tccgttata	aatggaaaca	gatttttatt	tttgcgtat	tttgcgtat	tttgcgtat	37080
tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	37140
tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	37200
aaaaaaatgtt	taagactctg	aaaaacaaaa	caaaggatct	gcaacat	tttgcgtat	37260
tcaaaaatgtt	agttcgttcc	atgcgttta	tttgcgtat	tttgcgtat	tttgcgtat	37320
tttagtctcc	atgagtctcc	aaagtttttc	tttgcgtat	tttgcgtat	tttgcgtat	37380
atcagaaggcc	ttccatccagg	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	37440
aagagaacca	aaacaagaca	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	37500
aaaaaacaca	attggaaagg	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	37560
cccaaggact	ttggggaggcc	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	37620
ctggacaaca	ttgtgttacc	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	37680
gctcacatct	gtatccca	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	37740
gatcagectt	ttcagcatgg	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	37800
gggtgtcgtat	gtcataccac	tttgcgtat	tttgcgtat	tttgcgtat	tttgcgtat	37860

-continued

-continued

-continued

tgccgcggcgg	cggtctgggaa	gccccgttgc	atcgccccccg	gccccccgcgc	cccagctccc	47040
ggcaggccgc	ggcccgccgac	aggcggcttg	acaggccggc	gcacagccgc	agtgcgggcc	47100
tcgcacactgc	ageccccccgc	gcccggccgc	gggagccggc	aagactcgga	gctggggttt	47160
cccccctcgca	ggcgctgcca	gggttgcagg	gaggcgttgt	ccccggggaaa	atgggacacg	47220
gagccctgcg	ggggccgggc	gcccggaccc	agggagccggg	caggcgtggg	cgggggaaagg	47280
cgttcccccgg	ccccagggtct	cgggccgtgg	ccccagccgc	ctgtcggtcc	tcggtcgcgc	47340
gtccccggcgc	tttgcacggc	acgttgggc	ggcgcgcggg	cccacccggc	ccgccccccagg	47400
ctcgccggcc	cgccccctct	ccggccgtggg	gaaaacgggtt	ggggcccaagg	ggggagggtc	47460
gcagggggttgc	agagcacccgt	gtgagggtcc	tgtcgccgc	cgccctgcgg	cctagggggt	47520
gggctggggc	ttggggggcc	actctctgt	gtctcgggga	ttatgtggat	gaccagggtt	47580
gcccgggggtc	ggggggggcc	cgccgggttc	aacaagtgt	tgtgaggggc	cgagatgt	47640
cggggtagatgc	ccccaggagg	acgtttctc	cgggaggagg	ccttggggaa	gcggggcttt	47700
gtatgttaaat	gggcgtcttgc	gagaggtat	gggggttctg	ggatgtaa	ggtgttctgt	47760
gaggaggggc	attttttctgc	ccccccggat	gagggttaggg	atggaaacctt	ggcctgtagg	47820
tcccaagatgt	tgaggctgtc	acttcagcac	ctggccctggc	actagaaggt	aatgggagga	47880
ggggtaaagtgt	ctaagataaa	ggggcaaaagg	gtgcccaggac	ctgggtctt	tcctttctgc	47940
tttctgtggc	tagatccctt	ctagatgtt	tttgggaaaga	gggctcacac	gctgtggca	48000
caagggggtgc	cctggcagggg	aggctctgt	gcccggggct	gcctcacaggc	ctccctacacc	48060
atttttatgt	ttttttttgtt	atctcttgc	atctctgtgt	tatgtggggg	atctgcctat	48120
gacccaggag	gtgtgtctgc	ccccctgtgt	cctcttaggt	cagattcgag	aatctggag	48180
gagctgtggc	cacagaccca	tttgttccag	ccatgttaggc	tacctcgctg	tctctacaca	48240
gcctgtgtgc	cggttattcc	ctataaggaa	aaatcaacag	ctcttgccct	ccctggaccc	48300
taacggccaa	gtcaaaaggct	ggtacccat	ttggctttgc	ggcctgttagg	aaggctgtta	48360
gtctgaaggc	cataccatgt	gtgtctgt	ggagcccttg	gggtgttggg	ccatgtgtag	48420
cagatgttgc	ttggaaagtga	gtctgaccc	gtctgaggaa	ggctgcocag	tctggctgt	48480
ggagcccaagc	ttggccttaga	gcaatgttat	tttcaactaa	ctctctccac	tagctaaact	48540
gaccagaaaa	aatgggctt	ctgtcccat	ccccaccccc	atccttcatgc	tgtgcacact	48600
atccccacaa	cagccccatgg	ataaaaggaa	tgtacatttgc	atctaatggc	tgctgggtgc	48660
catgcatggc	atcaaccagg	actacatttgc	catttgacaa	tgttcttaaa	acatctctgt	48720
aggtgaaata	ttaccccccatt	tttgcagat	aaagaaactga	gtctcagaga	ggtaaagtgt	48780
cttggcccaag	atcacagcaa	tttgtaaagt	ccaaaccaggaa	atataatacc	aggctctgtct	48840
tacttcaatgt	ctggatcatt	tttcaataata	tctgtgtggg	acagagtggg	agaaaccaggaa	48900
gaatttctgc	ctatttctgc	ccatgtgttg	cttcttgccct	gttattgtgt	tcatgtctta	48960
gtcttttggaa	aatacttccc	atctgttgc	ccactttatc	ctttagatcat	tcctggggagg	49020
taatgttgcac	aggatgttgc	taatagataaa	taatgtatgt	tgctgttgc	ggcaatgtggc	49080
actatataatgc	acttttacatgc	caggatgttgc	cactttatcat	gcagaacccca	accctatagg	49140
gtctgggttgc	actgtgtcgtc	atccccatttgc	tgttagatgt	gggtgttgcgg	ctcaagagaga	49200
ttaaatggct	tctccagggt	cacacagcta	gtaaatggag	tagaattttaa	acccacagca	49260
ttctgtatgc	ggatctgtgc	gtcttacat	ccaaatgtgc	caaataatggaa	taataatagt	49320
tccttaggtat	aggagtttacat	ttttatgtat	tgctttatgtat	ttttagataca	tcatcataca	49380
taatataccccc	caaaccatcat	gaattttatgtat	ttggatgtat	tgtcttcatc	ccatcttccat	49440
ggtaaaggaaa	ctgaggatca	gataggcagg	acttagaaacctt	ctgtttccca	agtcccagg	49500
ttcctttaggt	gacccttttgc	gtctgttgcac	attcacccttgc	tttcccttc	tttcccttc	49551
SEQ ID NO: 11		moltype = DNA	length = 93			
FEATURE		Location/Qualifiers				
source		1..93				
		mol_type = genomic DNA				
		organism = Homo sapiens				
SEQUENCE: 11						
gtccgcggag	gaaaaggag	tcctttaaa	gagaacgcgg	ctgcgaccat	tgcgcgtgc	60
cgcaggccgg	cgccccctgg	agggtgcgaag	ggc			93
SEQ ID NO: 12		moltype = DNA	length = 376			
FEATURE		Location/Qualifiers				
source		1..376				
		mol_type = genomic DNA				
		organism = Homo sapiens				
SEQUENCE: 12						
tttcagactgtat	ctatTTTTT	catggttggg	ggagcagccgg	gagacggggaa	gggtgtctagg	60
aggagaggggc	ggggaaaggag	ggctttctcc	tcttccaaatc	ctctgtatctt	cctgtgttt	120
ggaggtgtcgat	atttattgt	ccccagccca	ctgtccccccca	cgaccccccgc	ccccccaccc	180
ccccccacgc	cccccccccac	ctccgagccg	gttctgtgc	cagggtggcgc	tggagaaagg	240
aacagcggac	cccgccccaa	tcattactaa	tgttagcaatc	gtctgcataaa	acccaggccaa	300
cggttccat	ttggctgtat	ctctgttgc	gacgtcatcg	ggacgtacta	agactagggt	360
ttggccgaga	gtcgga					376
SEQ ID NO: 13		moltype = DNA	length = 791			
FEATURE		Location/Qualifiers				
source		1..791				
		mol_type = genomic DNA				
		organism = Homo sapiens				
SEQUENCE: 13						
aactggccgg	cgccgggtca	cgtggcaca	gcagttttcc	gggcctgagg	gtcccgctc	60
atccggcccc	cgccccatcc	ggagggtctcg	gcccctctcc	agccaggaaa	aggctggaga	120
gtccgcggc	cgccggacgc	tgcttccctt	aaggccaaagg	gagcagcagg	catgttgcgt	180

-continued

ctggccccagg	caaccaggcg	gcacgggtcc	aggttacctt	tccaaagggtt	ggggccctcg	240
agggctccgt	tttttctca	aggttccctga	gttagatgt	ggggtttggc	ataggaggt	300
gggtggccac	ggcagactg	gcaggagtc	ccctgtcaca	aacacccctc	cttcctctgt	360
ccatgttcc	ggcagtgcag	gggacagcca	gagggtatct	ggcttcctgg	cttcctcttc	420
ccttgttcca	aaggccgacc	gcctctgtgc	gggtttttca	accggggcaga	gtcttggaa	480
ctctggccca	gcctctttaa	gttgaggccc	gtatgtcgcc	gaggccaaagg	aaaacgtgcc	540
ggagtccctc	ctccctggcc	tcctcaggt	gcattcaccc	taccccaaaa	tctgaaccag	600
aatgcgttac	agggttaatca	ttttcatgtt	gttggggagg	accagattag	ctgtctgggt	660
ggtggggat	ggggcacaag	ctagtcctgt	gactcaaca	aaccttcca	gggtgtgacta	720
ggccaggaa	tccttgcatc	tccttacccc	cagcccttgg	gagacaaca	caaacaaacag	780
gatacaggaa	g					791

```
SEQ ID NO: 14          moltype = DNA    length = 696
FEATURE                Location/Qualifiers
source                 1..696
mol_type = genomic DNA
organism = Homo sapiens
```

SEQUENCE: 14

SEQUENCE. 14		
ggcgcttagag	ccgctgcgcgc cgctttcta gaaccttccc ccccaactaac ggcgttccg	60
ctacgtcagg	ccgtcgcgta aacgccttat ccgcgcgcaa tggcgaaaag gcttacgcc	12
ccacccattacg	ccaaatgcgt actcttccca cccttgccgc cagagacagt accgcacgtt	18
acttcgttaa	atgcgtccaa tgaattgcgg aaggcttagag tcctctgtacttactctt	24
tggaaataggg	tcccgcccccc tgccttggcg aaggccagggt gagaaaacgtt cgccgcaggtt	30
gaaaataacg	ccgacgggag gggcttataa cgcgcgttccgcgatccggcccttacacc	36
gggagggttgt	ccctgcaggta acgccaatga taaccccccgc cagaaaaatc tttagtagcct	42
tcccttttttgc	ttttccgtgc cccaaacttgc cggattgtact cggcccttc cggaaacacc	48
cgtaaatcaact	tctatgtccaa ttatgttca cgcgcgcgttccgcgcgttccgcgcgtc	54
tgtggaaactg	accccttgggt tacaggagag ttgcgtctgt aaatgggtt caaagggttgc	60
ctatgtttta	agctcccaac tcccccttccca ccagcgctgtt gaggatttca caccctcgca	66
ccgcaggggc	gaggaaagtgg gcggagttccg gttttg	69

```
SEQ ID NO: 15          moltype = DNA    length = 362
FEATURE
source           Location/Qualifiers
                  1..362
mol_type = genomic DNA
organism = Homo sapiens
```

SEQUENCE: 15

```

gtgtccctctt cttaaaggaa tatttttc tcagccattt atccaggttt gccaaggatc 60
ttctgcacta aagagatgga ctaacgtaat gactagggtt agtatctt tcacaaggcc 120
tgccctctttt aatccaaaat tggaaaacctt gggagtgctt agtatttcc atagatacta 180
ttcaactaat tactcgatc aactctgtgc ttgttactac ggtcaccatt ttaaaggaga 240
aaaactactt caaaaaatgg atttataago cggaaatttcc tcaagggtcac acattgcatt 300
atttccaaag caatgttcat cctgagggtt gttgtttctt tggtcgcccc ggttggagtgc 360
ca 362

```

```
SEQ ID NO: 16          moltype = DNA  length = 2722
FEATURE                Location/Qualifiers
source                 1..2722
                      mol_type = genomic DNA
                      organism = Homo sapiens
```

SEQUENCE: 16

SEQUENCE: 16						
tgtgcttgc	atgctgctt	tggttgatgg	agggtggggag	aatgtcagtt	gaagtggtat	60
tggaaactgt	cctccacagt	gccatgctct	accttcccaa	aacgaggaaat	gtcttgactt	120
gcagaagaca	caaattttca	aatttttgt	gatataatgg	tctgtcttggag	ttgggttggtg	180
taatttttca	agtagaaaca	gtcttcactg	ctaagatgg	tttggggctca	ggacacatgg	240
tctactcgaa	aaacactgact	gcaagtttgc	atctttgttca	caagaaggcca	ggaaaggatgg	300
taactatgtt	ttcaggggcca	gactaagggtc	agatgggtta	tttagtttac	tttactatgt	360
tcaggaaatgg	ctaaactaga	gatacttacaa	gtgaacagaa	ctgaaacacag	gtcaaatgtt	420
taatgtatgt	tatcttccat	taaaatagag	tagtttttgt	ggccctaaaa	tcctctgtgt	480
tttgcgttgtt	catcttccttc	tccttgctta	cgcattttgc	caaaggccaa	gaatgtacaa	540
tgccaaatggc	aagagtgaac	ccttagtgtt	actaaatctt	ggatgggggg	tgataatgtt	600
atatacaacgt	agggttctt	gttacacaaat	gttacactt	ggcagggtgg	agtgtatgtt	660
ataaagggggg	agggtatata	tggtggaaaat	gtcttacactt	tttctcaattt	tgctgtgtt	720
ctaaaaactgt	tctaaaaat	taatgttctt	aaaaaaaaaa	aaaaaacggcc	aaagtccgggt	780
gttttacactt	gttacccat	cactttgggg	ggccggacggca	ggcagatct	tcggggatgt	840
gagatcaaga	ccagccctggc	caacatgggt	aaacccccc	tctactaaaa	atataaaaag	900
ccaggcgttag	ttggccggccgc	ctgtatcccc	agctacttgg	ggggctgggg	caggagaatc	960
gtttgttgcacc	aggaggccgg	gggttgcagt	agcccaagatc	ggcccaactgc	actttccacgt	1020
gggttgcacgg	agtgatgtt	attttccgtt	aaaaaaacaaag	caacaaaaat	ttttccctcaaa	1080
gacgttataaa	acttagtggc	aatcttcatgt	tgttttttttgc	ggggacttgc	tacaaaaataa	1140
tttgcgtttag	attttttttgc	tttcaaaaaat	atgttcatat	atatgggggt	ggtgacataa	1200
ctacaactac	tttagaggaa	cctaatttttca	aaaacgtttt	tcagggttgc	aaagtgttgc	1260
gtttatcttgc	aatataat	aggggaaatcca	caaaaaaaaga	gggttttttttgc	caatcttgcata	1320
caaaaatgttca	cagtatttttca	aaaactgttgc	atgttttttttgc	tatggggaaa	gggttttttttgc	1380
tctatccccca	ttaatcttcaaa	atgttttttttgc	agcccaatgtt	taggggttgc	gttgggttttgc	1440
cagtgtttctt	cacccgttgc	tacaaatcaaa	ctggggatgtt	ttttatcttgc	acccaaatgtt	1500

-continued

tcataacttccctagatatta	ggtaattttc	tcccttatgtc	cgttggttcg	ctcttcaccc	1560
cttagagaatc	actgcattagg	tcactgcago	ctcgacttcc	tggtgctcgag	1620
ctcagccctcc	caaagcgctg	ggattacagg	catgagccac	cgtgcccagc	1680
attttctgaa	caaatgatgc	atttagcca	ggaatctcta	gtatatttt	1740
tcgtggggaa	agaccataat	cagctatTTT	attcttggaaa	tttaggaaaa	1800
tatcttaaac	agtggaaaaaa	tatTTTTGAA	agtatggca	cttccccgaa	1860
atatcttccct	atgttctgaa	atcttaggac	cttcgatgtt	tccggaaatcg	1920
cagcaattta	ttttagaaaaaa	taaggcaaac	cggctggggc	cggtggttca	1980
cccagcactt	tgagaggccg	aggcggggcg	atcgcttgg	ctcaggagtt	2040
ctggggcaaca	cagcggaaacc	tcgtttctac	taaaaataca	aaaattggcc	2100
cgcctgttgt	cccagctgt	cgggaaagctg	aggcggggaga	atcgcttga	2160
ggaaatttgcg	gtgagccgag	atcgctc	tgcacttcc	cctgggtgac	2220
cctgtctcaa	aaaagaaaaaa	aagaaagaaa	gaaagaaaaa	aaagaaagaa	2280
aaagaaaaaa	aaagaaaaaa	aaagaaaaaa	aaagaaaaaa	aaagaaaaaa	2340
agaatttccgc	tcacgtccgc	gccttccat	agcAAAAATGG	aggacttctc	2400
caactcccaa	gaagcaatgc	gaggdggaa	ggcggtaact	atTTTTCCC	2460
tctaggttca	atattggaaa	taggtctga	cgtttaataa	acacagcgtc	2520
atctgggggg	gaactggtaa	ctcgaaaacc	aaataactcg	tcttccgaaa	2580
aaccttccct	tctacaagag	ggtccggaaa	ccactgttac	gcccatttgg	2640
cttgggggggg	gcaaaaggcgc	tgaaggcga	agtgcgaca	cccgccgc	2700
ccgttagacgg	aacttgcct	tt			2722

SEQ ID NO: 17 moltype = DNA length = 800
 FEATURE Location/Qualifiers
 source 1..800
 mol_type = genomic DNA
 organism = Homo sapiens

SEQUENCE: 17

ttttcattta	gcaagagtgg	aaaactgttg	tggatccca	gagaagtctg	gagctaggt	60
gtaggggcta	gataatagag	ggctttgtact	aataggat	tgggccttat	cctgtggaaa	120
tctgaagggt	tttggggaaa	tgaattcagg	gcttaggtcga	agacagaagg	gcaaatacca	180
gactaagaca	cacagccact	aagagctct	ttttttttt	ttgtccagtc	agaggtatg	240
aggcctatat	atttatttagg	aaggatgatt	aaccatggct	gtagcttac	ctaagaggaa	300
cgtggctt	ggccgggtgt	cgggtgctca	cgcctgtat	cccagcatt	tgggaggtag	360
agggtgggggg	atcacatcg	gtcaggatgt	caagaccacg	ctgaccaaca	tgggaaaacc	420
tcttcttctac	taaaaataca	aaatttagccg	ggcgtgggtt	tgcatgcctg	taatccctgt	480
tactcgggg	gtgaggcag	gagaatcgct	tgaaccctgg	aggtggaggt	tgccgtgagc	540
tgagatcaca	ccattgca	ccagcttgg	caacaagac	aaaactctat	ctcaaaaaaa	600
aagagggaa	tgtggctt	gaaaatctt	tctagtc	tcttccctgt	ccccttagac	660
acaccttcaa	agatccctt	tctttggt	tgcgttgc	tccccctca	ctttaccaa	720
tccagcccaa	ttccatagcc	actgtgg	cctttatgg	aggagggaa	tggggtgaca	780
tagattctct	taaacatacc					800

SEQ ID NO: 18 moltype = DNA length = 800
 FEATURE Location/Qualifiers
 source 1..800
 mol_type = genomic DNA
 organism = Homo sapiens

SEQUENCE: 18

tatTTAAAGG	cattaatag	cgtttgc	actacttta	aaatTTTCC	ctcaattgt	60
tgagtccagg	aggtcaggat	tacggtaac	tatgattca	ccatgc	ccagctgagg	120
caacagagag	agacccttt	aaaaaaaaaa	aaaactttt	cctctc	agac	180
gcattqactc	acccatcat	ttttttttt	tttcttggaa	tgggttca	agaaaaaaact	240
ccaggcttga	ggcagtggta	tggtctgg	tcactgc	ctccgc	ccgggttcaag	300
cgatttcc	tgccttgc	tcccaatag	ctggattac	aggcaccc	caccacaccc	360
agctaatttt	ttgttatttt	agcagagac	gggtttctc	atattagc	agctggctg	420
gaactcttgc	cttcaatgt	tccaccacc	ttggccccc	aaatgtactgg	gattacaggc	480
acgagccacc	gcgcggggcc	ttcacctatt	atatctt	ttgttttta	aatagtgt	540
gacttccac	ctcagtctt	ttgttctt	cttccatt	gatgtc	ccaggta	600
cttaggaagg	cgggatccc	tttattctt	ccctgaat	caactactg	gacagctgt	660
gaagctggct	cggtgtgt	tttggggcc	ctcagatgt	gcagagccg	gcgttacag	720
cacatcaggc	agagcagacc	agaggcagg	aagacacact	ggggcagg	cagtggcac	780
atctggacca	cctgcgagg					800

SEQ ID NO: 19 moltype = DNA length = 800
 FEATURE Location/Qualifiers
 source 1..800
 mol_type = genomic DNA
 organism = Homo sapiens

SEQUENCE: 19

ttattcttca	cagaggacag	ctcccaact	tttacctctc	tccctaatgc	aacactcact	60
ctcaacttca	gcagectt	aatatctgt	aatgtgaaa	ctatcgat	tgaacttct	120
taactccctt	ttcatctt	gcacatc	aaaactt	aatgtaaa	atgggttt	180
aaattggtaa	actggagct	tatcaatct	tggctgtgt	aacttgagca	tgttgccaa	240
tttctatgt	cctcgtttt	tttatctaa	aatgaaaact	aataatgc	actccaaaga	300
attgttggat	ggattaagta	agacaacc	ataataattgc	tgtgcata	aaatccta	360

-continued

tgttggttgc	cttataccctc	taatotttctt	tctgctctct	tttctttaga	agatggcatg	420
cacattttcc	aaggtaaatt	ttactcttgt	tgaactctat	cccaatactct	cttggccatgg	480
ccctaatttt	gttcagttac	ttaactcttt	cccttcctcc	tgacatctt	ggtccttc	540
ttttccacagt	tttattctct	cgtttttacac	aaacatggcc	aagtcttacat	attccataaaa	600
accttgttca	ttaccacttc	agctcttctt	ttctgttcat	accaggctgt	gttgggtgcc	660
atgaatcat	cactattttt	atttttaaaa	ttccttatcac	tactttcccc	ctctgttcaatc	720
tggcttctcc	tgccaaacatg	ctactgaat	taatgaatta	attcatcaaa	tatttacaag	780
cattttatcat	qatactatqt					800

SEQ ID NO: 20 moltype = DNA length = 800

FEATURE Location/Qualifiers

source 1..800

```
mol_type = genomic DNA
```

organism = Homo sapiens

SEQUENCE : 2C

cacaggcact	gtgctca	actcaagtca	ttataatact	gtgtgttga	caatatgttg	60
tacttactgg	ccaaaaatca	ccaatgtcta	attgtcaagc	taaatgtctt	cttccaaattt	120
caatctatct	gacatgtcgg	aaaccttccct	ttccaaaggct	tctttgtcat	cgctcccttg	180
gtcttcgttgc	tcctttgtatc	tttgttgcatt	tacttttgcgt	tatcccttca	attatgggggg	240
tgtatccctta	ggcgcattcata	ttcaagctttc	ttctcttca	ctcttttaac	tttactctgt	300
gaattttcctg	taatttataca	ctacatttgg	tgtgtgttgc	tgtgtgtctg	tgtgcacgtg	360
catttttctta	ggaaaggaggat	catacgcttt	atcagctctg	caaggaggct	tatgaccctt	420
ccaaataacaca	ccctgtaccac	ttccacatctca	cgtgttagaca	catttttcca	ctgtgtatgc	480
ttccattttatc	ctccacatgc	tgctgttgc	ttgtatcttca	acccttaaaaacc	tctttttccat	540
gttcccgatgt	ccatcttccc	aactctggct	agggttttcc	tggagggtcc	tttcttgcata	600
atagtttctca	tcaggcagggt	ggcagaaacct	ctgtcttcca	ctcccttgcatt	gccccctctgt	660
gaggcttgggg	tgggttgggg	gtgtatgggg	gggtgtcgctgt	ataatcagag	tcccttgggg	720
gcagaagctt	tgttcttgcac	tttgccttttgc	ctatccatgg	aattactgt	ctatcttgcgc	780
catttttcca	tttcttgcattt					800

SEQ ID NO: 21 moltype = DNA length = 800

FEATURE Location/Qualifiers

source 1,800

mol_type = genomic DNA

```

SEQUENCE: 21
gaaggccgggt gtgggtggtc atgctgttgc tcccaggcac ttggggaaact gaggtggggag 60
gattgtcata gcccaggagg tcaaggctgc tggtagatgtt gattgtgaca ctgcactcca 120
gccttggggaa caaaaggggaa ccccttaact aactaaatgtaa atgaatgtaa gaatgtatgtaa 180
aggcaacacctg agctgcattt cttggcttc caaacctgtcc agccccatcc ctcagccctc 240
cttggatgttc aggggcccttc aggtttccaca caggccaggcccttgc ttttcgttccaaa 300
tttgacactt ccgttccattt ccctgcacca tctgttcacagg gtcaccatgg atgtggata 360
actgtccaccc tcttccttggaa gaaggccctcc gtcatactgg ccctctgtgtt caggatattca 420
acctgtcaaat cttccctgtt ggtgtgtggg ttccttcacaggcc cttgggtgtt gtgcaggggaa 480
ctcaggggggaa aacggggccccc accacgttcc ttcttcacaggcc acgtttccaga ggctgggggg 540
cttccctgtatcc cccctcccaaa tcccaagggtt cctgtctggcc ccacccgtca gactggact 600
ccgttggggctt gggctcttgcac ttggatattt tggttccaggcc acacagcagg caccctgggtt 660
gtatgttggcgcc tgcattggaa gtcaggaggaa gcagacgttccatctccctt gcttggggatca 720
ggccctggggcc acggctgttcc ggcggcttgg ggccagtgtac tgagcttccggccagggtt 780
cttggccaccatc ttgttggatdt 800

```

SEQ ID NO: 22 moltype = DNA length = 800

FEATURE Location/Qualifiers

source 1,800

mol type = genomic DNA

organism = Homo sapiens

SEQUENCE : 22

gtaggggggg	ataaaagat	caggccaag	gccaggtcg	gtggttcag	cctgtaatcc	60
cacgacttgg	ggagggccgg	ccaggcagat	cgcgcgatca	ggagtctcg	acgaggctgg	120
ccatcatgtt	gaaaccggct	ctctactaaa	aataaaaaaa	ttggccaggc	atggggcgc	180
ggccctgtaa	tccccatgtt	ggcaggagaa	tcgcttgaa	ccggggatgt	gggttgcag	240
tgagccgaga	gcgaggccat	gcactctgc	ctgggttaaca	aagcaagact	ccatctcaaa	300
aaaaaaaaaa	aaaaaaaaaa	tcaagtctgg	cgcagccccca	ggaagaaatc	ttgggtctggg	360
cacacccgtt	ttctgtgtca	tgaaggcagc	gacaccttgt	cgctgtcaca	gggttgcgtt	420
ggttggggcc	tagaggagag	gaggagcctt	tgttgaggggc	tccaaagaagg	cacagagagg	480
ttttgttcat	gggtcttgg	gacacctttt	gggtgttctg	tgcccttccat	cctggggccac	540
ttttggggagg	tgaaggaggg	aagcattaa	ggacaagacc	ccccgttccc	aattttcttc	600
cgaggccagg	gtttctctga	ttcgaagaaa	cagggtgtac	aaccaggaa	gttccactgt	660
ggcatactgtt	ctggggccat	agcattttag	gctgtatact	gagggtctgc	cctcccaagg	720
ctctgtgtcc	catttcgtgg	cgccccaaag	gggaagaaaa	actcctgaat	gtgcacccggg	780
acaggaccca	tccccatgcgg					800

SEQ ID NO: 23 moltype = DNA length = 800

FEATURE	Location/Qualifiers
---------	---------------------

source 1..800

mol type = genomic DNA

-continued

```

SEQUENCE: 23          organism = Homo sapiens
agaaaattgc cacctattgt attcaccata tgccaggcat tgggctggc acttgagttt 60
attttctaa ttcaagaaca aggccaggca cggtggctca tgccctgtaat cccagcacct 120
tgggaggcta aggcgggcag atcacaaggt caggagttt agaccagcct ggccaacatg 180
gaaaaatccc atctctacta aaaatacaa aatttagctgg gtttcgtgt gtgcacctgt 240
agtcccagct atcagggggg ctgaggcagg agaatctctt gaacccagga ggccggagggt 300
gcagtgagcc gagatcgccg cactgcactc cagcctgagt gacagagta gactccgtct 360
aaaaaaaaaaag aacaaaagcc gggcatgggt gctcaagact gtaateccag cactttggga 420
ggccaagggtt ggcagatcat gaggctcgtt gttccagaccc acatggtaa 480
aacctgttcc tactaaaaat aaaaaaaaaaactt gctggggctg gtggcaggca cttgcaatcc 540
caqctactca ggaggctgca gcaggagaat cacttgaacc cggggaggcag aggttttagt 600
gagctgagat cccgccactg cactccaacc tgggcaacag agcaagactc tgtctcaaaa 660
aaaaaaaaaaag aaaaaaaaaaaa 660
atttggcatg tgactatagtt cccagctact ttggtaggctt aggtgagagg atcacttgag 720
ccctggagtt tgagccccagc 780
800

SEQ ID NO: 24          moltype = DNA length = 800
FEATURE
source
1..800
mol_type = genomic DNA
organism = Homo sapiens

SEQUENCE: 24
ctaggcagt ctgcaagaac cagtctctaa acaaacaaga aaaaaaaaaaaac agaaaaaaa 60
aaaggttac ctggcctaga gaaatgttaa cagtctatcc ccttaagatc ataaaagtga 120
aataactgga cagactactg aaaaaaaaaaaag aatctgtt aacaagtctt agtGattttgt 180
agagtcttc caaaggacac taaaagttt aqaaccatgg ggaggaggct ttcttgcaagg 240
gtctgtatgt tacctcttagt tctattttata gatatattcc atttgcctt cttcatttgc 300
tcaccattag catggaaattt atccccctt ttttattttt atttgttagt acaagtcttag 360
cacagtaact ggcatalogt tacttataaa atgttacaat gtgttgaaga ctttttggat 420
gttagaattttt atattttgtt ctaegtttca ggaagtttgc tctggaaact gtgtgtatgaa 480
tgaattttgtt gacttagggaa ttagggaaac agagcaattt ggaggccacg gcaatgtatcc 540
agacaagaga agtctcagaa ctgtatattt acatttagca cagggttagaa agaaggaaat 600
acaaggatata ttcttcagg cttttaggatt aatttggattt aggttgaag gaaagagttaa 660
attaaatccca gttttttgtt gacatgttgc atgttcattaa ccttgttgcacat agggaaact 720
ggaaagaaaaag cagggttgcg gttttttttttaa atgtttaaaca ggcggggagc ggtggctcac 780
gcctgttaatc ccaacatttt 800

SEQ ID NO: 25          moltype = DNA length = 20
FEATURE
misc_feature
1..20
note = Oligonucleotides
source
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 25
ttatcatgtt actatgtccc 20

SEQ ID NO: 26          moltype = DNA length = 20
FEATURE
misc_feature
1..20
note = Oligonucleotides
source
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 26
agttactgag cacagtgcct 20

SEQ ID NO: 27          moltype = DNA length = 20
FEATURE
misc_feature
1..20
note = Oligonucleotides
source
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 27
gagttactgaa gcacagtgcct 20

SEQ ID NO: 28          moltype = DNA length = 20
FEATURE
misc_feature
1..20
note = Oligonucleotides
source
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 28

```

-continued

ttatcatgat actatgtccc	20
SEQ ID NO: 29	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 29	
agttactgag cacagtgccct	20
SEQ ID NO: 30	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 30	
gagttactga gcacagtgcc	20
SEQ ID NO: 31	moltype = DNA length = 21
FEATURE	Location/Qualifiers
misc_feature	1..21
	note = Oligonucleotides
source	1..21
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 31	
caagcattta tcatgatact a	21
SEQ ID NO: 32	moltype = DNA length = 21
FEATURE	Location/Qualifiers
misc_feature	1..21
	note = Oligonucleotides
source	1..21
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 32	
tcatgatact atgtcccagg c	21
SEQ ID NO: 33	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 33	
gcccacctt aagatctgt	20
SEQ ID NO: 34	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 34	
tgcacacattt gaagatctgt	20
SEQ ID NO: 35	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 35	
caccttgaag atctgttaggg	20
SEQ ID NO: 36	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides

-continued

source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 36		
tgaaccacgg caccctggcct		20
SEQ ID NO: 37	moltype = DNA length = 21	
FEATURE	Location/Qualifiers	
misc_feature	1..21	
	note = Oligonucleotides	
source	1..21	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 37		
tccctcccta cagatcttca a		21
SEQ ID NO: 38	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 38		
gccccacattg aagatctgtta		20
SEQ ID NO: 39	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 39		
tgcaccaccc tt gaagatctgt		20
SEQ ID NO: 40	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 40		
ggcccaagag gtcagtacac		20
SEQ ID NO: 41	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 41		
ggataaaagag atcagggtcca		20
SEQ ID NO: 42	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 42		
caccttgaag atctgttaggg		20
SEQ ID NO: 43	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 43		
tccctacaga tcttcaaggt		20

-continued

SEQ ID NO: 44	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 44	
ttcaagggtgg gcaaggaact	20
SEQ ID NO: 45	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 45	
tccctcccta cagatcttca	20
SEQ ID NO: 46	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 46	
ctccctacag atcttcaagg	20
SEQ ID NO: 47	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 47	
accttgaa tctgttaggga	20
SEQ ID NO: 48	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 48	
cttcaagggtg ggcaaggaac	20
SEQ ID NO: 49	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 49	
acagatcttc aagggtggca	20
SEQ ID NO: 50	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 50	
aagagatcag gtccaaaggcc	20
SEQ ID NO: 51	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = Oligonucleotides
source	1..20
	mol_type = other DNA

-continued

SEQUENCE: 51	organism = synthetic construct	
gggcaaggaa ctggggcaag		20
SEQ ID NO: 52	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
source	note = Oligonucleotides	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 52		
tagggagggg taaagagatc		20
SEQ ID NO: 53	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
source	note = Oligonucleotides	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 53		
atcagggtcca aggccaggtg		20
SEQ ID NO: 54	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
source	note = Oligonucleotides	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 54		
agggtccaagg ccaggtgcgg		20
SEQ ID NO: 55	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
source	note = Oligonucleotides	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 55		
attaatgcct ttaaataggt		20
SEQ ID NO: 56	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
source	note = Oligonucleotides	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 56		
tcttaaacct acctattttaa		20
SEQ ID NO: 57	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
source	note = Oligonucleotides	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 57		
gctaattaat gcctttaaat		20
SEQ ID NO: 58	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
source	note = Oligonucleotides	
	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 58		
attaatgcct ttaaataggt		20
SEQ ID NO: 59	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	

-continued

misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 59		
tcttaaacct acctatcaa		20
SEQ ID NO: 60	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 60		
ccacttccct cctccattaa		20
SEQ ID NO: 61	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 61		
gtctaattaat gcctttaaat		20
SEQ ID NO: 62	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 62		
ctttaatggc ggagggaaat		20
SEQ ID NO: 63	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 63		
ccttaatggc aggaggaaag		20
SEQ ID NO: 64	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 64		
tttaatggcg gagggaaatg		20
SEQ ID NO: 65	moltype = DNA length = 21	
FEATURE	Location/Qualifiers	
misc_feature	1..21	
	note = Oligonucleotides	
source	1..21	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 65		
atggaggagg gaagtgggggt g		21
SEQ ID NO: 66	moltype = DNA length = 21	
FEATURE	Location/Qualifiers	
misc_feature	1..21	
	note = Oligonucleotides	
source	1..21	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 66		

-continued

```

agagaatcta tgtcacccca c 21
SEQ ID NO: 67      moltype = DNA  length = 21
FEATURE           Location/Qualifiers
misc_feature      1..21
note = Oligonucleotides
source            1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 67
aaggcattaa ttagcgtttgc 21
SEQ ID NO: 68      moltype = DNA  length = 21
FEATURE           Location/Qualifiers
misc_feature      1..21
note = Oligonucleotides
source            1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 68
aataggttgg tttaagagaa t 21
SEQ ID NO: 69      moltype = DNA  length = 21
FEATURE           Location/Qualifiers
misc_feature      1..21
note = Oligonucleotides
source            1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 69
catgcactac tttaaaattt t 21
SEQ ID NO: 70      moltype = DNA  length = 21
FEATURE           Location/Qualifiers
misc_feature      1..21
note = Oligonucleotides
source            1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 70
aagtatgtca tgcaaacgct a 21
SEQ ID NO: 71      moltype = DNA  length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = Oligonucleotides
source            1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 71
cttgcagcac tgccttaggt 20
SEQ ID NO: 72      moltype = DNA  length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = Oligonucleotides
source            1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 72
tggttcttgc agcactgcct 20
SEQ ID NO: 73      moltype = DNA  length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = Oligonucleotides
source            1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 73
tcttgagca ctgcctaggc 20
SEQ ID NO: 74      moltype = DNA  length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = Oligonucleotides

```

-continued

source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 74		
cttgcagcac tgcctaggct		20
SEQ ID NO: 75	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 75		
tggttcttgc agcaactgcct		20
SEQ ID NO: 76	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 76		
tcttgccaga ctgccttaggc		20
SEQ ID NO: 77	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 77		
taggctggc tc当地actcca		20
SEQ ID NO: 78	moltype = DNA length = 19	
FEATURE	Location/Qualifiers	
misc_feature	1..19	
	note = Oligonucleotides	
source	1..19	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 78		
ctaggctggg ctc当地actc		19
SEQ ID NO: 79	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 79		
tgagaggatc acttgagccc		20
SEQ ID NO: 80	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 80		
ctggagtttg agcccagcct		20
SEQ ID NO: 81	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Oligonucleotides	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 81		
ttctttgttt gtttagagac		20

-continued

```

SEQ ID NO: 82      moltype = DNA length = 21
FEATURE           Location/Qualifiers
misc_feature      1..21
note = Oligonucleotides
source            1..21
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 82
agcccagcct aggcagtgtc g                                21

SEQ ID NO: 83      moltype = DNA length = 21
FEATURE           Location/Qualifiers
misc_feature      1..21
note = Oligonucleotides
source            1..21
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 83
gagactgggtt cttgcagcac t                                21

SEQ ID NO: 84      moltype = DNA length = 21
FEATURE           Location/Qualifiers
misc_feature      1..21
note = Oligonucleotides
source            1..21
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 84
tttagagact gggttctgca g                                21

SEQ ID NO: 85      moltype = DNA length = 21
FEATURE           Location/Qualifiers
misc_feature      1..21
note = Oligonucleotides
source            1..21
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 85
tttgtttaga gactgggttct t                                21

SEQ ID NO: 86      moltype = DNA length = 21
FEATURE           Location/Qualifiers
misc_feature      1..21
note = Oligonucleotides
source            1..21
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 86
tttgtttgtt tagagactgg t                                21

SEQ ID NO: 87      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = Synthetic
source            1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 87
aaggggccaag ggagtttagtg                                20

SEQ ID NO: 88      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = Synthetic
source            1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 88
agctccctca ctaactccct                                20

SEQ ID NO: 89      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = Synthetic
source            1..20
mol_type = other DNA

```

-continued

-continued

-continued

QEIIISAAGKE	LSEAFKQKTS	EILSHAHAA	DQPLPTTLKK	QEEKEILKSQ	LDSLLGLYHL	480
LDWFAVDESN	EVDPEFSARL	TGIKLEMEPS	LSFYNKARNY	ATKKPYSVEK	FKLNFQMPML	540
ASGWDVNKEK	NNGAILFVKN	GLYYLGIMPK	QKGRYKALSF	EPTEKTSEG	DKMYYDYFPD	600
AAKMPIPKST	QLKAVTAHFQ	THTTPILLSN	NFIEPLEITK	EIYDLNNPEK	EPKKFQATA	660
KKTDQDKGYR	EALCKWIDFT	RDFLSKYTKT	TSIDLSSLRP	SSQYKDLGEY	YAELNPLLYH	720
ISFQRIAEKE	IMDAVETGKL	YLFQIYNKDF	AKGHHGKPML	HTLYWTGLFS	PENLAKTSIK	780
LNGQAELFYR	PKSRMKRMAH	RLGEKMLNKK	LKDQKTPIPD	TLYQELYDYV	NHRLSHDLS	840
EARALLPNVI	TKEVSHEIIK	DRRFTSDKFF	FHVPITLNYQ	AANSPSKFNQ	RVNAYLKEHP	900
ETPIIGIDRG	ERNLIYITVI	DSTGKILEQR	SLNTIQQFDY	QKKLDNRKE	RVAARQAWSV	960
VGTIQLDKQG	YLSQVIHEIV	DLMHQYQAVV	VLENLNFGFK	SKRTGIAEKA	VYQOFEKMLI	1020
DKLNCLVLKD	YPAEKVGVL	NPYQLTDQFT	SFAKMGQTSG	FLFVYPAPYT	SKIDPLTGTV	1080
DPFWWKTIN	HESRKHFLEG	FDFLHYDVKT	GDFILHFKMN	RNLNSFQRGLP	GFMPAWDIVF	1140
EKNETQFDK	GTPFIAGKRI	VPVIENHRFT	GRYRDLYPAN	ELIALLEEKG	IVFRDGSNIL	1200
PKLLENDSSH	AIDTMVALIR	SVLQMRNSNA	ATGEDYINSP	VRDLNGVCFD	SRFQNPEWPM	1260
DADANGAYHI	ALKGQLLLH	LKESKDLKLQ	NGISNQDWLA	YIQELRN	1307	

SEQ ID NO: 99 moltype = AA length = 1331
 FEATURE Location/Qualifiers
 REGION 1..1331
 note = Synthetic
 source 1..1331
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 99

MTQFEGFTNL	YQVSKTLRFE	LIPQGKTLKH	IQEQGFIEED	KARNDHYKEL	KPIIDRIYKT	60
YADQCLQLVQ	LDWENLSAAI	DSYRKEKTEE	TRNALIEEQAA	TYRNAIHDF	IGRTDNLTDA	120
INKRHABEIYK	GLFKAELFNG	KVLKQLGTVT	TTEHENALLR	SFDKFTTYFS	GFYENRKNVF	180
SABDISTAIP	HRIVQDNFPK	FKENCHIFTR	LITAVPSLRE	HFENVKKAIG	IFVSTSIEEV	240
FSFPFYNQOLL	TQTQIDLYNQ	LLGGISREAG	TEKIKGLNEV	LNLAIQKNDE	TAHIIASLPH	300
RFIPLFKQIL	SDRNTLFSIL	EEFKSDEEVI	QSPCKYKTL	RNENVLETAE	ALFNELNISID	360
LTHIFISHKK	LETISSALCD	HWTDLRNALY	ERRISELTGK	ITKSAKEKVQ	RSLKHEDINL	420
QEIIISAAGKE	LSEAFKQKTS	EILSHAHAA	DQPLPTTLKK	QEEKEILKSQ	LDSLLGLYHL	480
LDWFAVDESN	EVDPEFSARL	TGIKLEMEPS	LSFYNKARNY	ATKKPYSVEK	FKLNFQMPML	540
ASGWDVNKEK	NNGAILFVKN	GLYYLGIMPK	QKGRYKALSF	EPTEKTSEG	DKMYYDYFPD	600
AAKMPIPKST	QLKAVTAHFQ	THTTPILLSN	NFIEPLEITK	EIYDLNNPEK	EPKKFQATA	660
KKTDQDKGYR	EALCKWIDFT	RDFLSKYTKT	TSIDLSSLRP	SSQYKDLGEY	YAELNPLLYH	720
ISFQRIAEKE	IMDAVETGKL	YLFQIYNKDF	AKGHHGKPML	HTLYWTGLFS	PENLAKTSIK	780
LNGQAELFYR	PKSRMKRMAA	RLGEKMLNKK	LKDQKTPIPD	TLYQELYDYV	NHRLSHDLS	840
EARALLPNVI	TKEVSHEIIK	DRRFTSDKPL	FHVPITLNYQ	AANSPSKFNQ	RVNAYLKEHP	900
ETPIIGIDRG	ERNLIYITVI	DSTGKILEQR	SLNTIQQFDY	QKKLDNRKE	RVAARQAWSV	960
VGTIQLDKQG	YLSQVIHEIV	DLMHQYQAVV	VLENLNFGFK	SKRTGIAEKA	VYQOFEKMLI	1020
DKLNCLVLKD	YPAEKVGVL	NPYQLTDQFT	SFAKMGQTSG	FLFVYPAPYT	SKIDPLTGTV	1080
DPFWWKTIN	HESRKHFLEG	FDFLHYDVKT	GDFILHFKMN	RNLNSFQRGLP	GFMPAWDIVF	1140
EKNETQFDK	GTPFIAGKRI	VPVIENHRFT	GRYRDLYPAN	ELIALLEEKG	IVFRDGSNIL	1200
PKLLENDSSH	AIDTMVALIR	SVLQMRNSNA	ATGEDYINSP	VRDLNGVCFD	SRFQNPEWPM	1260
DADANGAYHI	ALKGQLLLH	LKESKDLKLQ	NGISNQDWLA	YIQELRN	1320	
HAAPPKKKRK V						1331

SEQ ID NO: 100 moltype = AA length = 1103
 FEATURE Location/Qualifiers
 REGION 1..1103
 note = Synthetic
 source 1..1103
 mol_type = protein
 organism = synthetic construct

SEQUENCE: 100

MTIRSMKLKL	KIYSGRSAPQ	LROGLWRLHR	LLNEGTYAAM	DWLVHMRQEA	LPGKSKKEIR	60
AELERRVRQQ	QEKGNGVQNDQ	VPMDEVLSAL	RQLYELLVPS	AVNNSGDAQT	LSRKFLSPLV	120
DPNSEGGKGT	SNAGAKPGW	KKQEAQGDPW	EKDYERWLKR	KQADPTAEIL	GKLETAGLKP	180
LPFLYLTNEVK	DWRMPLTSK	QYVNRWDRDM	FQQAIHELLS	WETWNRKVNE	ERAKLKETVR	240
RFFEQHLang	KDWLSPQAY	EANRQALRD	MAISPSPDRF	ITRRQIKGWS	ELYERWNKLA	300
PTASVEAMYQ	EVRVHQVKLKG	GTFGDADLYR	FLAKPENVH	WRDHQERLHY	YAAYNDLHKR	360
LMSAKEQAQF	TLPDPVAHPL	WVRFDARDGN	LFTYILQADS	SKQRSRRYVN	FSRFLWPVED	420
GYFEETENVK	VELALSKQFY	RQVIVHDNPT	GKQKITFQDY	SSKEILEGHL	GGAKLQLDRN	480
FLRKSGRDFE	VELDFGPFLN	VVLDDLPKQK	VKNGRQLQSP	GQALLVKSRP	NDIPKVYGYK	540
PDALAALEQ	ASGEESLGSE	SLRQGFRVMS	IDLGVRSAAA	ISVFSVKGEK	TREGDKVCY	600
VGETGLFAVH	DRSFLRLPG	ESSEKRVNVE	RDKRKTERMQ	IRYHIRTLLAR	VLRLANKATP	660
MDRIKAVQDV	LNDIESTRFM	NDHDHHVYNH	ALETLRITYAP	DHQGIWEQV	IAAHRLQLEHH	720
VGVIVGEWRK	NWGDKDRRTV	GLSMNDNIEEL	DEMRRLLISW	SRRARYPREA	KPFQVNESNP	780
VHLLRHLQNL	KEDRLKQLAN	LIVMTALGYV	YDSKEKKWKA	AYPACQLILF	EDLQRYRFHL	840
DRSARENSQL	MKWHAHSIPI	YVWMQGEPYG	LQIGDVWAGF	TSRYHAKTGA	PGIRCKALTE	900
KDFQQGRLL	SLVAEGMFTL	QEVTGTLKPGD	IVPAEGGELF	VTLADDSGDR	IVITHADINA	960
AQNVQKREW	ANSERFRVAC	RSVQIASQEC	FIPSSESVAK	KMGKGVFVRD	FSFHKDMEVY	1020
HWNNQVKLTA	KNVPTDHSD	LQDLQDYQAI	LEEARESSSS	YKTLFRDPSG	FFFPPDDVVWP	1080
QNIYWRREVKK	TITALLRKRI	MST				1103

-continued

SEQ ID NO: 101	moltype = AA length = 1091
FEATURE	Location/Qualifiers
REGION	1..1091
	note = Synthetic
source	1..1091
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 101	
MPIRSFKLKL VTHNGDSTYM DKLRRGLWKT HVIINRGIAY YMNTLALMRQ EPYGSKSREE	60
VRLDLLSTLR EQQRNNWSE QTGTDELLS LSRRVYELLV PSAIGEKGDA QMLSRKFSLSP	120
LVDPNSEGGR CTAKSGRKPR WKKMMEEGHP DWEKEKEKDA AKKAEDPTAS ILADLEAVGL	180
LPLFPLFSDE QKEIRWLPPK KRFQVRTWDR DMFQQALERI LSWESWNRRV AEEYKLQAQ	240
RDEVYAKYLE DAGSWLNDLQ TFEKQREEL AEVSFEPNTE YLITRQIRG WKEVYEWWSK	300
TSENASQEL WRMVADVQTA MAGAGFDPKV YQFLSQPKHH HIWREHPNRL FYYSKYNEVR	360
EKLNRRAKQAA FTLPDPVHE PLWTRFDARG GNIHDYEISK VGKQYHVTFS SLILPEAQSW	420
VEIENVTVGI CNSQLQKQI RLDGYADKKQ KVKYDDYSSR FELTGVLGGA KIQFDRKHLK	480
KA AHLRAEAE TGPIFLNVVV DVEPFLEVKN GRLRTPLGQV LQVNTRDWPK VVDYKAKELS	540
VIMENTQIGN ENGVSTIEAG MRIMSIDLGQ RTAAAVSIFE VISKKPDEKE TKLFYPIADT	600
DLYAVHRRSL LLRLPGEEIS SKKMICRKE RARISSLVRY QIRLLSEVLR LHTQGTAEQR	660
RFKLDELLVS IQRKLELDQS EWISELEKLF DYIDESABKW KEALVVAHRT LEPIVVEAVR	720
NWKSLSKEN KDRRIAGIS IWSIEELEET RKLLIAWSKH SREP GIPKRL EKEETFAPEH	780
LQHQVNQKDD RLKQMANLFB MTALGYKYDE GNKRWVEAAPP ACQVILFEDL SRYRFALDRP	840
RRENRRLMKW AHRSIPRLTY MQAELFGIQV GDVYSAYTSR FHAKTGAPGI RCHALTEADL	900
QSNSYVNVQL IKDKFIQDNQ TEILKAGQIV PWQGGELFVT FADRSGASLA VIHADINAAQ	960
NLQKRFWQHN SEVFRVPCVK VKGGLVPVYE KMRKLFKGKL FVNIDDPESK EVYRWEHSTK	1020
MKSKTTPVDL ESEDIDHEEL SDEWEDMQEG YKTLRDPNG FFWSSDSWIP QKDFWIRVKS	1080
RIGKSLREQI R	1091
SEQ ID NO: 102	moltype = AA length = 1091
FEATURE	Location/Qualifiers
REGION	1..1091
	note = Synthetic
source	1..1091
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 102	
MPIRSFKLKL VTHNGDSTYM DKLRRGLWKT HVIINRGIAY YMNTLALMRQ EPYGSKSREE	60
VRLDLLSTLR EQQRNNWSE QTGTDELLS LSRRVYELLV PSAIGEKGDA QMLSRKFSLSP	120
LVDPNSEGGR CTAKSGRKPR WKKMMEEGHP DWEKEKEKDA AKKAEDPTAS ILADLEAVGL	180
LPLFPLFSDE QKEIRWLPPK KRFQVRTWDR DMFQQALERI LSWESWNRRV AEEYKLQAQ	240
RDEVYAKYLE DAGSWLNDLQ TFEKQREEL AEVSFEPNTE YLITRQIRG WKEVYEWWSK	300
TSENASQEL WRMVADVQTA MAGAGFDPKV YQFLSQPKHH HIWREHPNRL FYYSKYNEVR	360
EKLNRRAKQAA FTLPDPVHE PLWTRFDARG GNIHDYEISK VGKQYHVTFS SLILPEAQSW	420
VEIENVTVGI CNSQLQKQI RLDGYADKKQ KVKYDDYSSR FELTGVLGGA KIQFDRKHLK	480
KA AHLRAEAE TGPIFLNVVV DVEPFLEVKN GRLRTPLGQV LQVNTRDWPK VVDYKAKELS	540
VIMENTQIGN ENGVSTIEAG MRIMSIDLGQ RTAAAVSIFE VISKKPDEKE TKLFYPIADT	600
DLYAVHRRSL LLRLPGEEIS SKKMICRKE RARISSLVRY QIRLLSEVLR LHTQGTAEQR	660
RFKLDELLVS IQRKLELDQS EWISELEKLF DYIDESABKW KEALVVAHRT LEPIVVEAVR	720
NWKSLSKEN KDRRIAGIS IWSIEELEET RKLLIAWSKH SREP GIPKRL EKEETFAPEH	780
LQHQVNQKDD RLKQMANLFB MTALGYKYDE GNKRWVEAAPP ACQVILFEDL SRYRFALDRP	840
RRENRRLMKW AHRSIPRLTY MQAELFGIQV GDVYSAYTSR FHAKTGAPGI RCHALTEADL	900
QSNSYVNVQL IKDKFIQDNQ TEILKAGQIV PWQGGELFVT FADRSGASLA VIHADINAAQ	960
NLQKRFWQHN SEVFRVPCVK VKGGLVPVYE KMRKLFKGKL FVNIDDPESK EVYRWEHSTK	1020
MKSKTTPVDL ESEDIEHEEL SDEWEDMQEG YKTLRDPNG FFWSSDSWIP QKDFWIRVKS	1080
RIGKSLREQI R	1091

1. A genetically modified mammalian cell, comprising an exogenous nucleotide sequence integrated in a sustained transcriptionally active payload region (STAPLR) in the genome of the cell, wherein the STAPLR is selected from the group consisting of the intergenic region between the RPL34 gene and the OSTC gene;

the intergenic region between the ACTB gene and the FSCN1 gene;

the intergenic region between the AKIRIN1 gene and the NDUFS5 gene;

the intergenic region between the PRDX1 gene and the AKR1A1 gene;

the intergenic region between the PTGES3 gene and the NACA gene;

the intergenic region between the MLF2 gene and the PTMS gene;

the intergenic region between the RAB13 gene and the RPS27 gene;

the intergenic region between the JTB gene and the RAB13 gene;

the intergenic region between the AKR1A1 gene and the NASP gene;

the intergenic region between the NDUFS5 gene and the MACF1 gene;

the intergenic region between the SRSF9 gene and the DYNLL1 gene;

the intergenic region between the MYL6B gene and the MYL6 gene;

the intergenic region between the GPX1 gene and the RHOA gene;

the intergenic region between the HNRNPA2B1 gene and the CBX3 gene;

the intergenic region between the ROMO gene and the RBM39 gene;
 the intergenic region between the PA2G4 gene and the RPL41 gene; and
 the intergenic region between the NDUFB10 and the RPS2 gene.

2. A method for modifying a mammalian cell, comprising integrating an exogenous nucleotide sequence in a sustained transcriptionally active payload region (STAPLR) in the genome of the cell, wherein the STAPLR is selected from the group consisting of:

the intergenic region between the RPL34 gene and the OSTC gene;
 the intergenic region between the ACTB gene and the FSCN1 gene;
 the intergenic region between the AKIRIN1 gene and the NDUFS5 gene;
 the intergenic region between the PRDX1 gene and the AKR1A1 gene;
 the intergenic region between the PTGES3 gene and the NACA gene;
 the intergenic region between the MLF2 gene and the PTMS gene;
 the intergenic region between the RAB13 gene and the RPS27 gene;
 the intergenic region between the JTB gene and the RAB13 gene;
 the intergenic region between the AKR1A1 gene and the NASP gene;
 the intergenic region between the NDUFS5 gene and the MACF1 gene;
 the intergenic region between the SRSF9 gene and the DYNLL1 gene;
 the intergenic region between the MYL6B gene and the MYL6 gene;
 the intergenic region between the GPX1 gene and the RHOA gene;
 the intergenic region between the HNRNPA2B1 gene and the CBX3 gene;
 the intergenic region between the ROMO gene and the RBM39 gene;
 the intergenic region between the PA2G4 gene and the RPL41 gene; and
 the intergenic region between the NDUFB10 and the RPS2 gene.

3. The method of claim 2, wherein the integrating step is performed by using a CRISPR/Cas system; a Cre/Lox system; a FLP-FRT system; a TALEN system; a ZFN system; homing endonucleases; random integration; homologous recombination; a transposase; or a non-nuclease-dependent viral vector.

4. The method of claim 2, wherein the integrating step is performed by using a CRISPR/Cas system comprising a guide RNA, and wherein

the STAPLR is the intergenic region between the RPL34 gene and the OSTC gene and the gRNA is selected from SEQ ID NOs: 25-32,
 the STAPLR is the intergenic region between the ACTB gene and the FSCN1 gene and the gRNA is selected from SEQ ID NOs: 33-54,
 the STAPLR is the intergenic region between the AKIRIN1 gene and the NDUFS5 gene and the gRNA is selected from SEQ ID NOs: 55-70, or

the STAPLR is the intergenic region between the PRDX1 gene and the AKR1A1 gene and the gRNA is selected from SEQ ID NOs: 71-92.

5. The method of claim 3, wherein the CRISPR/Cas system comprises a gRNA-dependent nuclease of type I, type II, type III, type IV, type V, or a variant thereof.

6. The method of claim 3, wherein the CRISPR/Cas system comprises a gRNA-dependent nuclease selected from the group consisting of Cas9, Cpf1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas12, Cas13, Cas100, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, CasX, CasY, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, CasPhi, MAD7, and Csf4.

7. A DNA molecule comprising a nucleotide sequence of interest flanked by a 5' homologous region (HR) and a 3' HR, wherein the 5' and 3' HRs are at least 95% homologous to a first genomic region (GR) and a second GR, respectively, in a sustained transcriptionally active payload region (STAPLR) in the genome of a mammalian cell, wherein the STAPLR is selected from the group consisting of:

the intergenic region between the RPL34 gene and the OSTC gene;
 the intergenic region between the ACTB gene and the FSCN1 gene;
 the intergenic region between the AKIRIN1 gene and the NDUFS5 gene;
 the intergenic region between the PRDX1 gene and the AKR1A1 gene;
 the intergenic region between the PTGES3 gene and the NACA gene;
 the intergenic region between the MLF2 gene and the PTMS gene;
 the intergenic region between the RAB13 gene and the RPS27 gene;
 the intergenic region between the JTB gene and the RAB13 gene;
 the intergenic region between the AKR1A1 gene and the NASP gene;
 the intergenic region between the NDUFS5 gene and the MACF1 gene;
 the intergenic region between the SRSF9 gene and the DYNLL1 gene;
 the intergenic region between the MYL6B gene and the MYL6 gene;
 the intergenic region between the GPX1 gene and the RHOA gene;
 the intergenic region between the HNRNPA2B1 gene and the CBX3 gene;
 the intergenic region between the ROMO gene and the RBM39 gene;
 the intergenic region between the PA2G4 gene and the RPL41 gene; and
 the intergenic region between the NDUFB10 and the RPS2 gene.

8. The DNA molecule of claim 7, wherein each of the 5' and 3' HRs is independently about at least 50, at least 100, at least 150, at least 200, at least 250, at least 300, at least 350, at least 400, at least 450, at least 500, at least 550, at least 600, at least 650, at least 700, at least 750, at least 800, at least 850, at least 900, at least 950, at least 1000, at least 1100, at least 1200, at least 1300, at least 1400, at least 1500,

at least 1600, at least 1700, at least 1800, at least 1900, or at least 2000 base pairs long; or between 50 to 1500 base pairs long.

9. The DNA molecule of claim 7-Gr-8, wherein the 5' and 3' HRs are at least 95% homologous to SEQ ID NOs: 17 and 18, SEQ ID NOs: 19 and 20, SEQ ID NOs: 21 and 22, SEQ ID NOs: 23 and 24, SEQ ID NOs: 93 and 94, or SEQ ID NOs: 95 and 96, respectively.

10. The genetically modified mammalian cell of claim 1, wherein:

- the intergenic region between the RPL34 gene and the OSTM gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 1;
- the intergenic region between the ACTB gene and the FSCN1 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 2;
- the intergenic region between the AKIRIN1 gene and the NDUFS5 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 3;
- the intergenic region between the PRDX1 gene and the AKR1A1 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 4;
- the intergenic region between the PTGES3 gene and the NACA gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 5;
- the intergenic region between the MLF2 gene and the PTMS gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 6;
- the intergenic region between the RAB13 gene and the RPS27 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 7;
- the intergenic region between the JTB gene and the RAB13 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 8;
- the intergenic region between the AKR1A1 gene and the NASP gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 9;
- the intergenic region between the NDUFS5 gene and the MACF1 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 10;
- the intergenic region between the SRSF9 gene and the DYNLL1 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 11;
- the intergenic region between the MYL6B gene and the MYL6 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 12;
- the intergenic region between the GPX1 gene and the RHOA gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 13;

the intergenic region between the HNRNPA2B1 gene and the CBX3 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 14;

the intergenic region between the ROMO gene and the RBM39 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 15;

the intergenic region between the PA2G4 gene and the RPL41 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 16; and/or

the intergenic region between the NDUFB10 and the RPS2 gene comprises a nucleotide sequence at least 95% identical to SEQ ID NO: 97.

11. The genetically modified mammalian cell of claim 1, wherein the exogenous nucleotide sequence comprises a transgene.

12. The genetically modified mammalian cell of claim 11, wherein the transgene encodes a therapeutic protein; a cellular marker; or a protein that regulates the differentiation state or activity of the cell.

13. The genetically modified mammalian cell of claim 1, wherein the cell is a human cell.

14. The genetically modified mammalian cell of claim 1, wherein the cell is a pluripotent stem cell (PSC).

15. The genetically modified mammalian cell of claim 1, wherein the cell is:

- a) a cell in the immune system;
- b) a cell in the cardiovascular system;
- c) a cell in the metabolic system;
- d) a cell in the central nervous system;
- e) a muscle cell;
- f) an adipose cell; or
- g) a cell in the ocular system.

16. The method of claim 3, wherein the non-nuclease-dependent viral vector is selected from the group consisting of: a retroviral vector, an adeno-associated viral (AAV) vector, and a lentiviral vector.

17. The genetically modified mammalian cell of claim 11, wherein the transgene comprises a constitutive or inducible promoter.

18. The genetically modified mammalian cell of claim 12, wherein the therapeutic protein is selected from the group consisting of: a protein deficient or defective in a genetic disease, a cytokine, and a recombinant antigen receptor.

19. The genetically modified mammalian cell of claim 12, wherein the transgene encodes SOX10, IL-10, IL-12, CD19t, or ThPOK.

20. The genetically modified mammalian cell of claim 14, wherein the pluripotent stem cell (PSC) is an induced PSC (iPSC).

* * * * *