



US 2016033323A1

(19) **United States**

(12) **Patent Application Publication**
MIZUGUCHI et al.

(10) **Pub. No.: US 2016/033323 A1**

(43) **Pub. Date: Nov. 17, 2016**

(54) **CONDITIONALLY REPLICATING
ADENOVIRUS**

G01N 33/574 (2006.01)

C12Q 1/68 (2006.01)

C12N 9/12 (2006.01)

C12N 15/113 (2006.01)

(71) Applicant: **NATIONAL INSTITUTE OF
BIOMEDICAL INNOVATION**, Osaka
(JP)

(52) **U.S. Cl.**

CPC *C12N 7/00* (2013.01); *C12N 9/1276*

(2013.01); *C12Y 207/07049* (2013.01); *C12N*

15/113 (2013.01); *G01N 33/57492* (2013.01);

C12Q 1/6897 (2013.01); *C07K 14/005*

(2013.01); *C12N 2710/10021* (2013.01); *C12N*

2840/203 (2013.01); *C12N 2310/141*

(2013.01); *G01N 2333/075* (2013.01); *G01N*

2333/705 (2013.01)

(72) Inventors: **Hiroyuki MIZUGUCHI**, Ibaraki-shi
(JP); **Fuminori SAKURAI**, Ibaraki-shi
(JP)

(73) Assignee: **NATIONAL INSTITUTE OF
BIOMEDICAL INNOVATION**, Osaka
(JP)

(21) Appl. No.: **15/183,661**

(22) Filed: **Jun. 15, 2016**

(57)

ABSTRACT

Related U.S. Application Data

(63) Continuation-in-part of application No. 14/240,216,
filed on Mar. 12, 2014, filed as application No.
PCT/JP2012/053814 on Feb. 17, 2012.

The object of the present invention is to provide a novel
conditionally replicating adenovirus and a reagent compris-
ing the same for cancer cell detection or for cancer diagno-
sis.

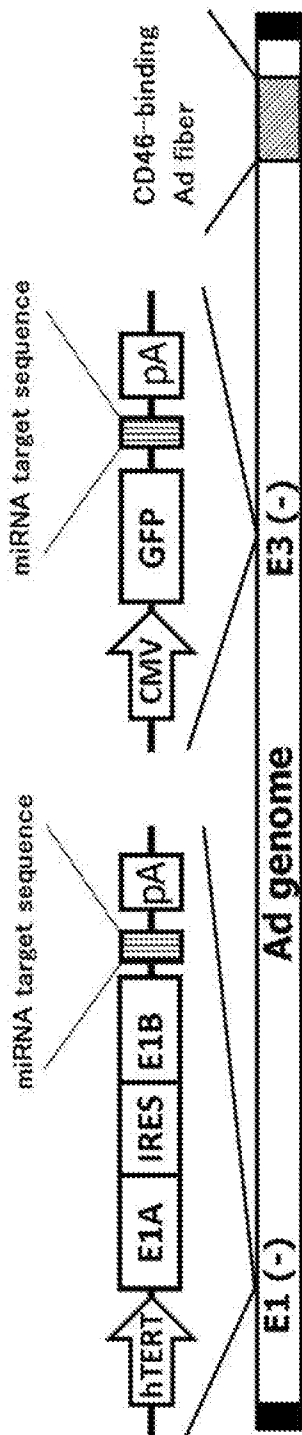
(30) **Foreign Application Priority Data**

Aug. 23, 2011 (JP) 2011-181414

The present invention provides a polynucleotide, which
comprises human telomerase reverse transcriptase (hTERT)
promoter, E1A gene, IRES sequence and E1B gene in this
order and which comprises a target sequence of a first
miRNA. The present invention also provides a recombinant
adenovirus, which comprises a replication cassette compris-
ing the above polynucleotide, wherein the replication cas-
sette is integrated into the E1 region of the adenovirus
genome.

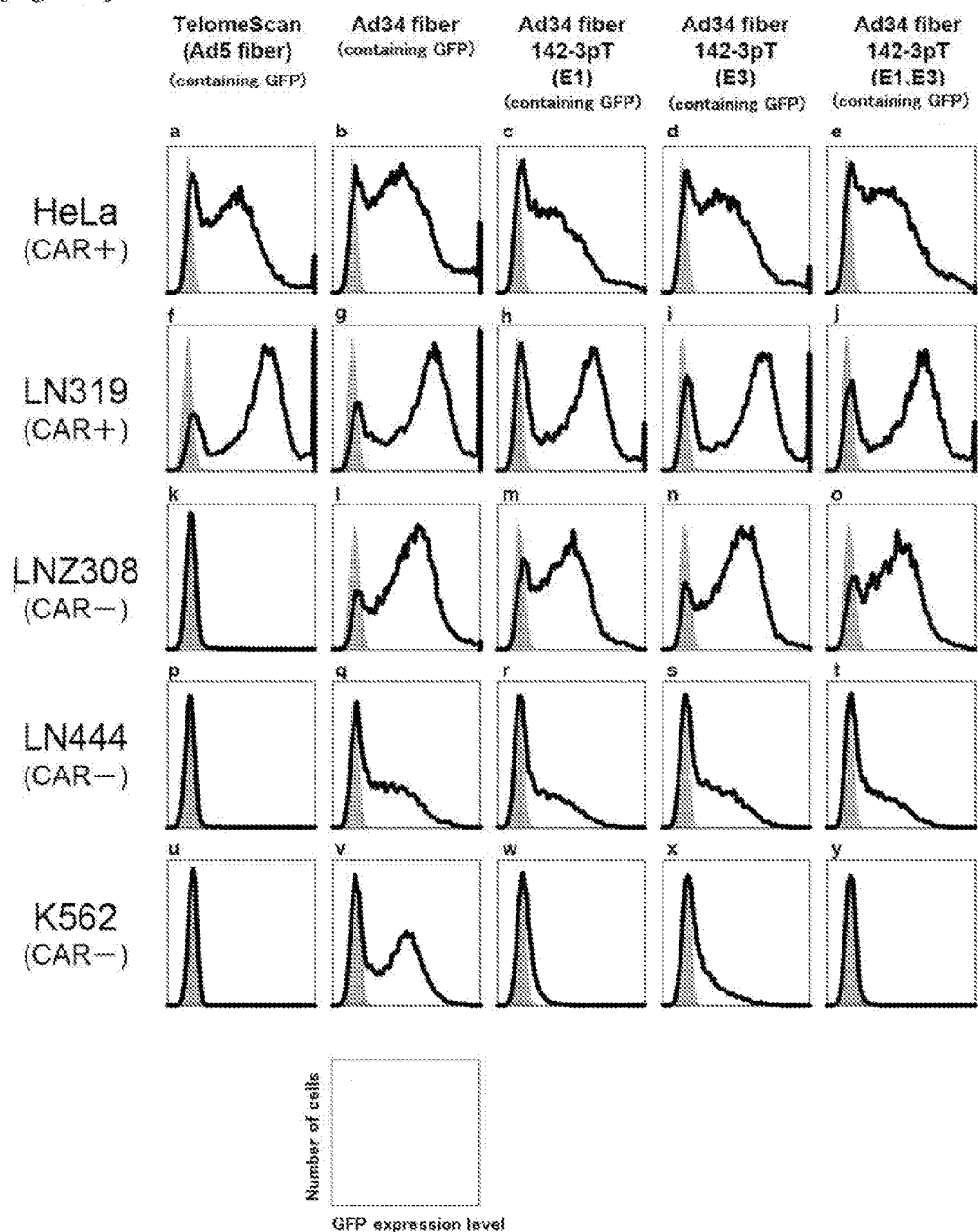
Publication Classification

(51) **Int. Cl.**
C12N 7/00 (2006.01)
C07K 14/005 (2006.01)



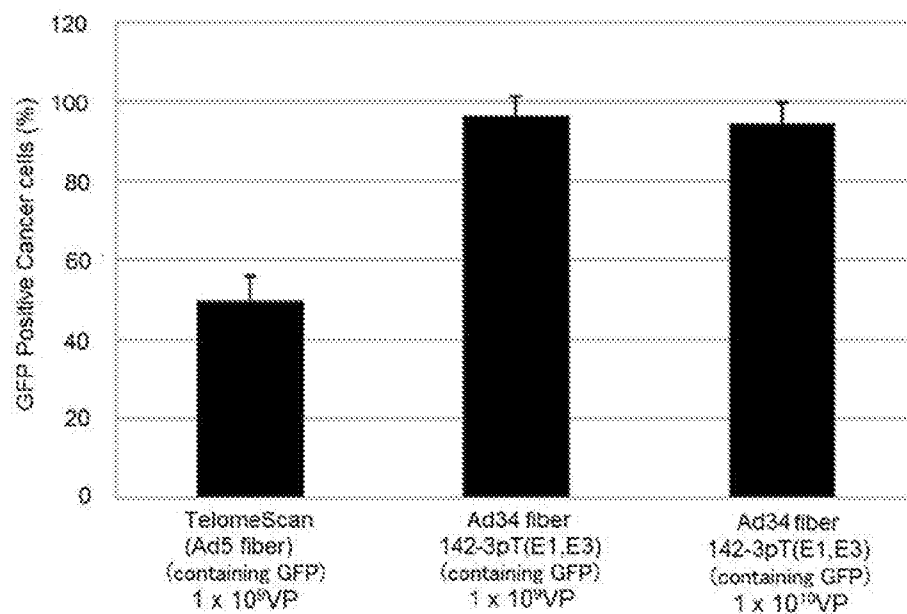
[Figure 1]

[Figure 2]



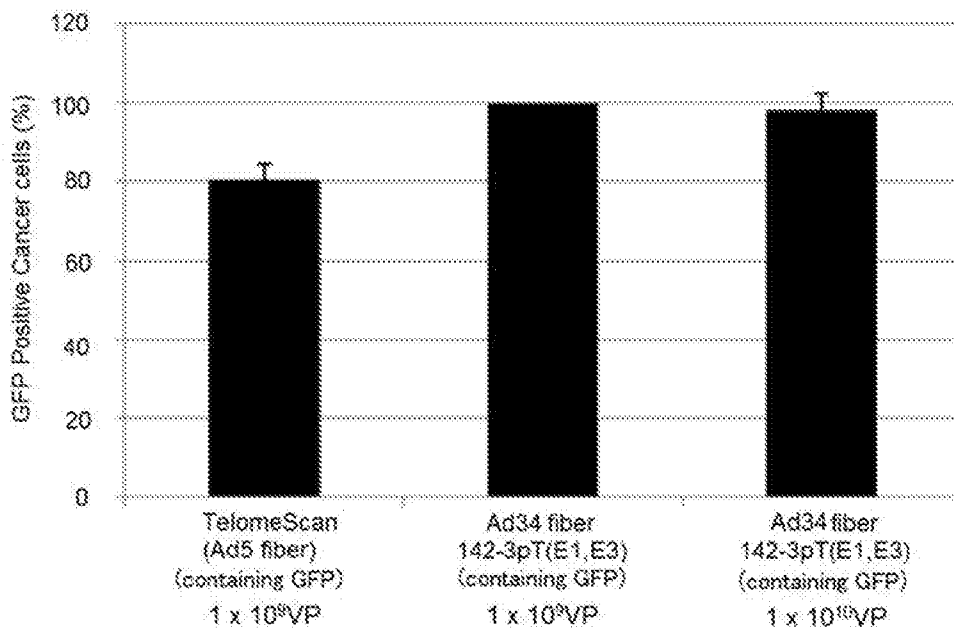
[Figure 3]

H1299 cells

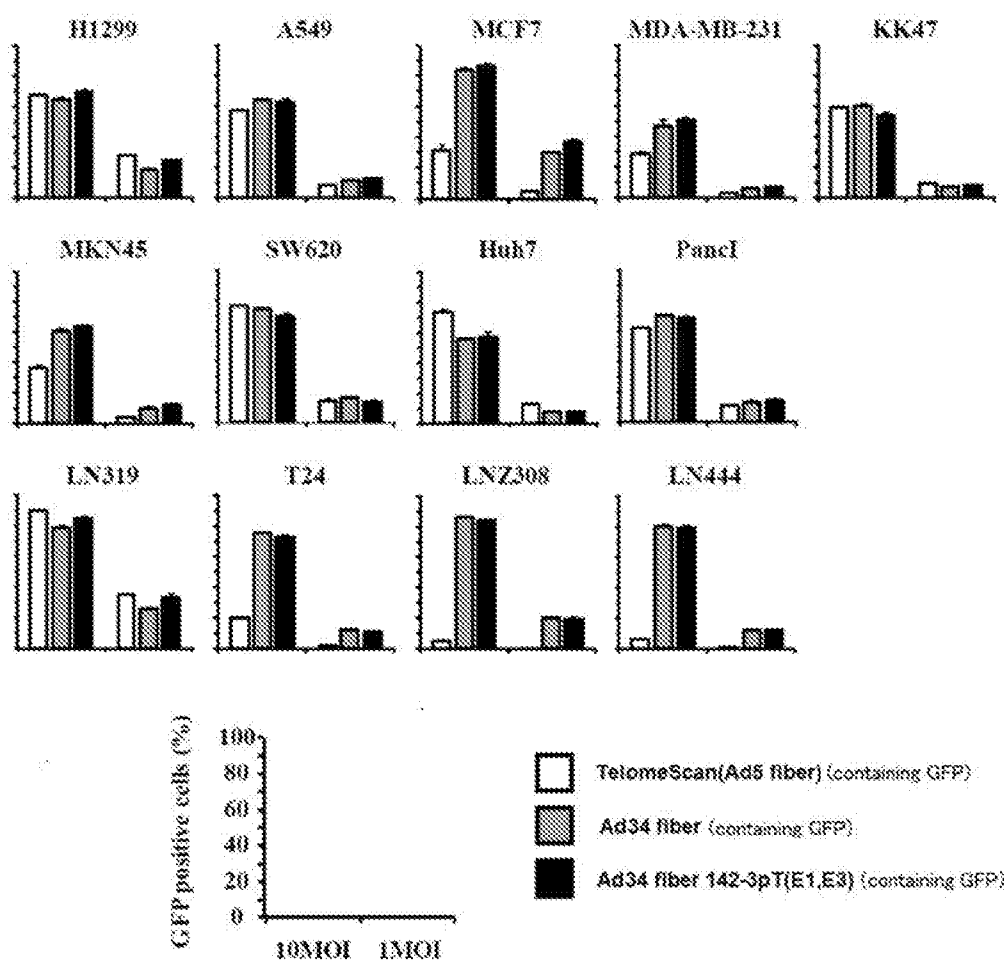


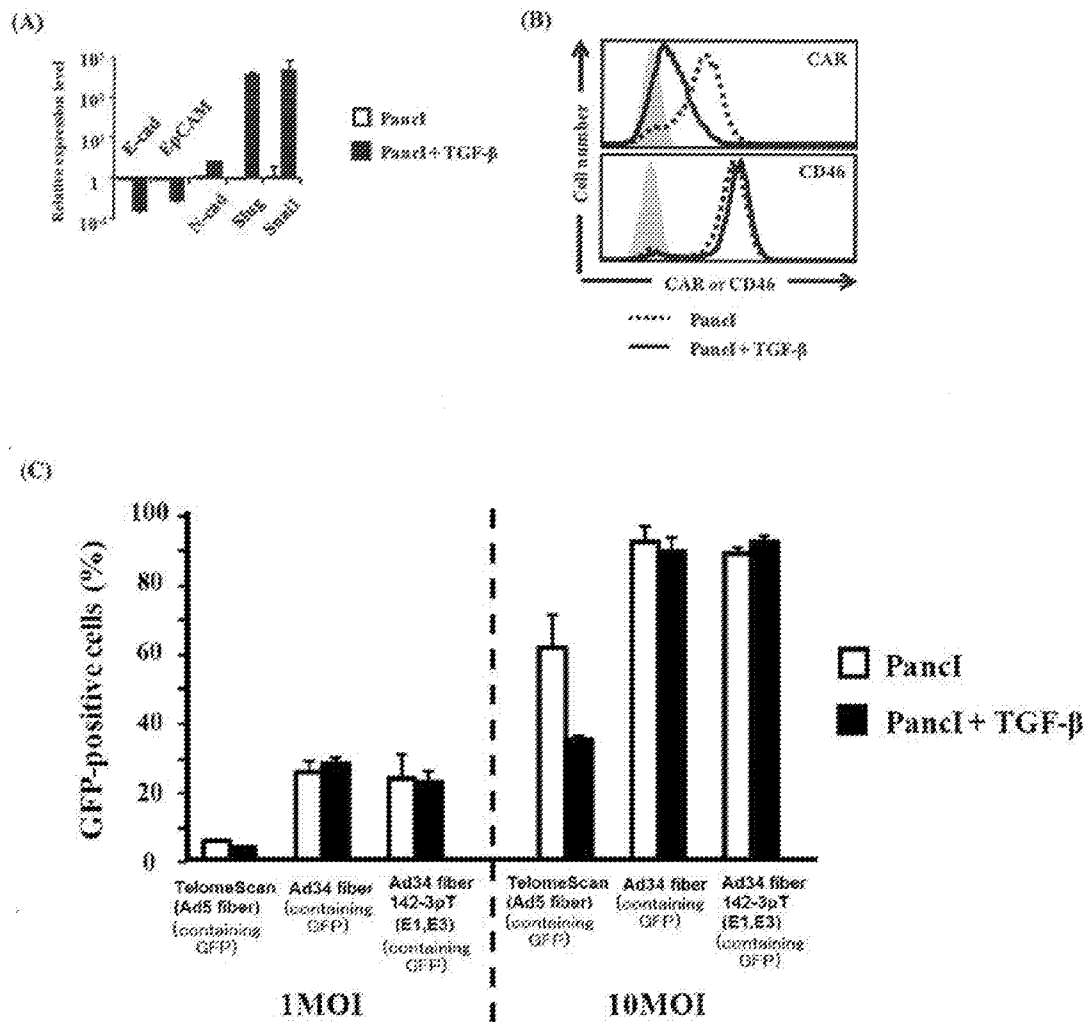
[Figure 4]

A549 cells



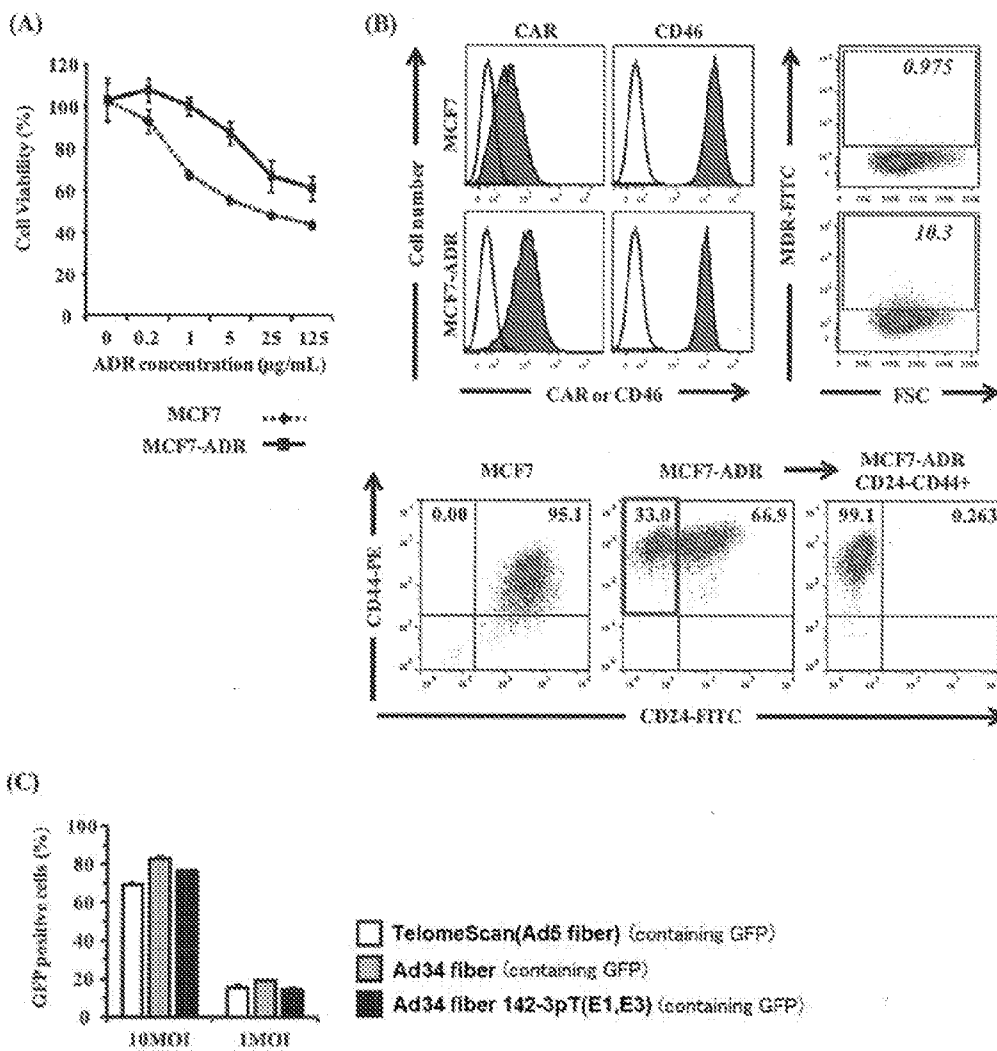
[Figure 5]



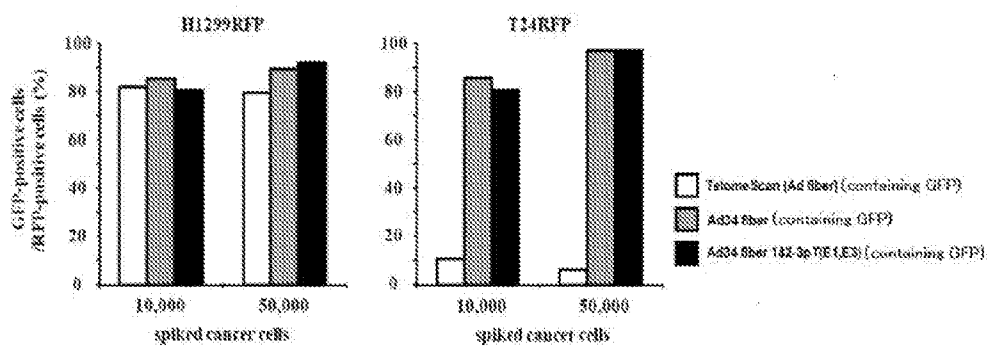


[Figure 6]

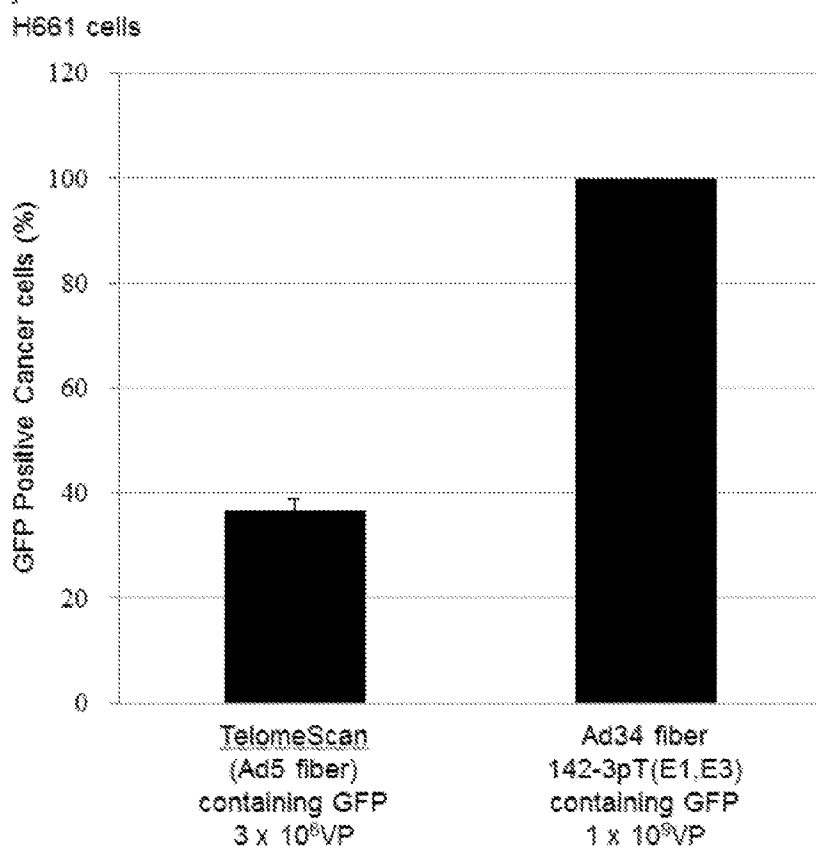
[Figure 7]



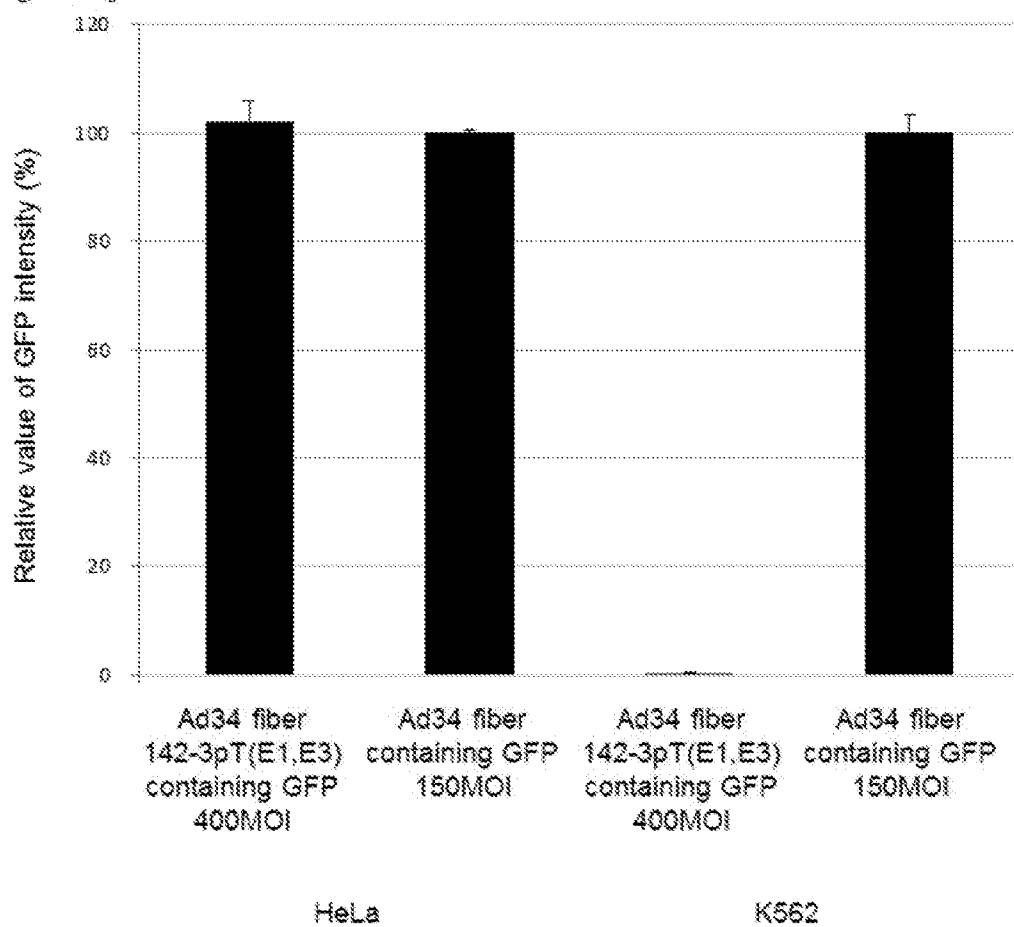
[Figure 8]



[Figure 9]



[Figure 10]



CONDITIONALLY REPLICATING ADENOVIRUS

[0001] This application is a continuation-in-part application of U.S. application Ser. No. 14/240,216 filed on Mar. 12, 2014, which is the U.S. National Phase of International Patent Application No. PCT/JP2012/053814, filed on Feb. 17, 2012, and claims the benefit of the filing date of JP application 2011-181414 filed on Aug. 23, 2011. These documents are hereby incorporated by reference for all purposes.

REFERENCE TO SEQUENCE LISTING SUBMITTED VIA EFS-WEB

[0002] This application includes an electronically submitted sequence listing in .txt format. The .txt file contains a sequence listing entitled "2016-06-15_SequenceListing_4456-0189PUS2.txt" created on Jun. 15, 2016 and is 137,556 bytes in size. The sequence listing contained in this .txt file is part of the specification and is hereby incorporated by reference herein in its entirety.

TECHNICAL FIELD

[0003] The present invention relates to a novel conditionally replicating adenovirus and a reagent comprising the same for cancer cell detection or for cancer diagnosis.

BACKGROUND ART

[0004] Techniques currently used for cancer diagnosis mainly include (i) those using large-sized testing instruments (e.g., MRI) and (ii) those for measuring tumor markers or the like in blood, and expectations are now focused on (ii) which are simple techniques with less burden on patients. In particular, cancer cells circulating in the peripheral blood of cancer patients (i.e., circulating tumor cells (CTCs)) show a close relationship with clinical symptoms because these cells increase the risk of systemic metastasis and because the prognosis of patients with CTCs is significantly poor. Thus, it has been expected to develop a technique for simple and highly sensitive detection of CTCs as a predictive factor or surrogate marker for prognosis.

[0005] Techniques used for CTC detection include detection with a cancer-related antigen such as EpCAM (epithelial cell adhesion molecule) or cytokeratin-8 (e.g., Cell-Search system) and detection by means of RT-PCR, etc. However, these cancer-related antigens are also expressed on normal epithelial cells and hence are highly likely to cause false positive detection, while cell morphology characteristic of cancer cells cannot be observed at the same time in the case of PCR detection. For these reasons, there has been a demand for a new technique in terms of sensitivity, simplicity, accuracy and costs.

[0006] On the other hand, the inventors of the present invention have already developed a conditionally replicating adenovirus which grows specifically in cancer cells and expresses GFP (GFP-expressing conditionally replicating adenovirus: GFP-CRAd) (which is referred to as TelomeScan®; OBP-401 or Telomelysin-GFP) (Patent Document 1: WO2006/036004). Moreover, the inventors of the present invention have also developed a simple technique for CTC detection using this TelomeScan (Non-patent Document 1: Kojima T, et al, J. Clin. Invest., 119: 3172, 2009).

[0007] However, since TelomeScan has the fiber protein of adenovirus type 5 and infects via coxsackievirus and adenovirus receptor (CAR) in target cells, TelomeScan may not infect cells which do not express CAR. In particular, it is known that CAR expression is reduced in highly malignant cancer cells which are highly invasive, metastatic and proliferative (Non-patent Document 2: Okegawa T., et al, Cancer Res., 61: 6592-6600, 2001); and hence TelomeScan may not detect these highly malignant cancer cells. Moreover, although less likely, TelomeScan may give false positive results by infecting and growing in normal blood cells (e.g., leukocytes) to cause GFP expression.

[0008] For these reasons, there has been a demand for a reagent for cancer cell detection and a reagent for cancer diagnosis, each of which detects almost all cancer cells including CAR-negative ones and does not give any false positive results in normal blood cells.

PRIOR ART DOCUMENTS

Patent Documents

- [0009] Patent Document 1: WO2006/036004
 [0010] Non-patent Document 1: Kojima T., et al, J. Clin. Invest., 119: 3172, 2009
 [0011] Non-patent Document 2: Okegawa T., et al, Cancer Res., 61: 6592-6600, 2001

SUMMARY OF THE INVENTION

Problem to be Solved by the Invention

[0012] The present invention has been made under these circumstances, and the problem to be solved by the present invention is to provide a reagent for cancer cell detection and a reagent for cancer diagnosis, each of which detects almost all cancer cells including CAR-negative ones and does not give any false positive results in blood cells, as well as to provide a conditionally replicating recombinant adenovirus which is useful as such a reagent.

Means to Solve the Problem

[0013] As a result of extensive and intensive efforts made to solve the above problem, the inventors of the present invention have found that not only CAR-positive cells, but also CAR-negative cells can be detected when the fiber of adenovirus type 5 in TelomeScan is replaced with another adenovirus fiber binding to CD46, which is highly expressed on almost all human cells, particularly cancer cells in general. Moreover, the inventors of the present invention have succeeded in avoiding any false positive results in blood cells by integration of a microRNA (miRNA)-mediated gene regulatory system into TelomeScan, which led to the completion of the present invention.

[0014] Namely, the present invention is as follows.

[0015] (1) A polynucleotide, which comprises human telomerase reverse transcriptase promoter, E1A gene, IRES sequence and E1B gene in this order and which comprises a target sequence of a first microRNA.

[0016] (2) The polynucleotide according to (1) above, wherein the first microRNA is expressed in non-cancer cells.

[0017] (3) The polynucleotide according to (1) or (2) above, wherein the first microRNA is at least one selected from the group consisting of miR-142, miR-15, miR-16,

miR-21, miR-126, miR-181, miR-223, miR-296, miR-125, miR-143, miR-145, miR-199 and let-7.

[0018] (4) A recombinant adenovirus, which comprises a replication cassette comprising the polynucleotide according to any one of (1) to (3) above, wherein the replication cassette is integrated into the E1 region of the adenovirus genome.

[0019] (5) The recombinant adenovirus according to (4) above, which further comprises a labeling cassette comprising a reporter gene and a promoter capable of regulating the expression of the gene, wherein the labeling cassette is integrated into the E3 region of the adenovirus genome.

[0020] (6) The recombinant adenovirus according to (5) above, wherein the labeling cassette further comprises a target sequence of a second microRNA.

[0021] (7) The recombinant adenovirus according to (4) above, wherein a cell death-inducing cassette comprising a gene encoding a cell death induction-related protein and a promoter capable of regulating the expression of the gene is further integrated into the E3 region of the adenovirus genome.

[0022] (8) The recombinant adenovirus according to (7) above, wherein the cell death-inducing cassette further comprises a target sequence of a second microRNA.

[0023] (9) The recombinant adenovirus according to (6) or (8) above, wherein the second microRNA is expressed in non-cancer cells.

[0024] (10) The recombinant adenovirus according to (9) above, wherein the second microRNA is at least one selected from the group consisting of miR-142, miR-15, miR-16, miR-21, miR-126, miR-181, miR-223, miR-296, miR-125, miR-143, miR-145, miR-199 and let-7.

[0025] (11) The recombinant adenovirus according to (5) or (6) above, wherein the reporter gene is a gene encoding a protein which emits fluorescence or a gene encoding an enzyme protein which generates a luminophore or a chromophore upon enzymatic reaction.

[0026] (12) The recombinant adenovirus according to any one of (5) to (10) above, wherein the promoter is human telomerase reverse transcriptase promoter or cytomegalovirus promoter.

[0027] (13) The recombinant adenovirus according to any one of (4) to (12) above, which further comprises a gene encoding a CD46-binding fiber protein.

[0028] (14) The recombinant adenovirus according to (13) above, wherein the CD46-binding fiber protein comprises at least the fiber knob region in the fiber protein of adenovirus type 34 or 35.

[0029] (15) A reagent for cancer cell detection, which comprises the recombinant adenovirus according to any one of (4) to (14) above.

[0030] (16) A reagent for cancer diagnosis, which comprises the recombinant adenovirus according to any one of (4) to (14) above.

[0031] (17) The reagent according to (15) above, wherein the cancer cells are derived from a biological sample taken from a subject.

[0032] (18) The reagent according to (17) above, wherein the biological sample is blood.

[0033] (19) The reagent according to (15) or (18) above, wherein the cancer cells are circulating tumor cells.

[0034] (20) The reagent according to any one of (15) and (17) to (19) above, wherein the cancer cells are drug-resistant cancer cells.

[0035] (21) The reagent according to any one of (15) and (17) to (20) above, wherein the cancer cells are cancer stem cells.

[0036] (22) The reagent according to any one of (15) and (17) to (21) above, wherein the cancer cells are cancer cells having undergone epithelial-mesenchymal transition or mesenchymal-epithelial transition.

[0037] (23) A method for cancer cell detection, which comprises contacting cancer cells with the recombinant adenovirus according to (11) above and detecting the fluorescence or color produced by the cancer cells.

[0038] (24) The method according to (23) above, wherein the cancer cells are derived from a biological sample taken from a subject.

[0039] (25) The method according to (24) above, wherein the biological sample is blood.

[0040] (26) The method according to (25) above, wherein the cancer cells are circulating tumor cells.

[0041] (27) A recombinant adenovirus, which comprises:

[0042] a replication cassette comprising a polynucleotide comprising a human telomerase reverse transcriptase promoter, E1A gene, IRES sequence and E1B gene in this order, and a target sequence of a first microRNA, wherein the first microRNA is miR-142 and the target sequence comprises a nucleotide sequence having at least 98% identity to the nucleotide sequence consisting of SEQ ID NO: 52, and wherein the replication cassette is integrated into the E1 region of the adenovirus genome;

[0043] a labeling cassette comprising a reporter gene, a promoter capable of regulating the expression of the reporter gene, and a target sequence of a second microRNA, wherein the second microRNA is miR-142 and the target sequence comprises a nucleotide sequence having at least 98% identity to the nucleotide sequence consisting of SEQ ID NO: 52, and wherein the labeling cassette is integrated into the E3 region of the adenovirus genome; and

[0044] a gene encoding a CD46-binding fiber protein comprising at least the fiber knob region in the fiber protein of adenovirus type 34 or 35, which comprises a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 50 and is integrated into the adenovirus genome.

[0045] (28) The recombinant adenovirus according to (27) above, wherein the reporter gene is a gene encoding a protein which emits fluorescence or a gene encoding an enzyme protein which generates a luminophore or a chromophore upon enzymatic reaction.

[0046] (29) The recombinant adenovirus according to (27) above, wherein the promoter capable of regulating the expression of the reporter gene is a human telomerase reverse transcriptase promoter or cytomegalovirus promoter.

[0047] (30) The recombinant adenovirus according to (27) above, wherein the replication cassette comprises a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 53.

[0048] (31) The recombinant adenovirus according to (27) above, wherein the labeling cassette comprises a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 54.

[0049] (32) A recombinant adenovirus, which comprises:
[0050] a replication cassette comprising a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 53 and is integrated into the E1 region of the adenovirus genome;

[0051] a labeling cassette comprising a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 54 and is integrated into the E3 region of the adenovirus genome; and

[0052] a gene encoding a CD46-binding fiber protein comprising a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 50 and is integrated into the adenovirus genome.

[0053] (33) A recombinant adenovirus, which comprises a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 51.

Effects of the Invention

[0054] The present invention enables simple and highly sensitive detection of CAR-negative cancer cells without detection of normal blood cells (e.g., leukocytes).

BRIEF DESCRIPTION OF THE DRAWINGS

[0055] FIG. 1 is a schematic view showing an example of the structure of the recombinant adenovirus of the present invention.

[0056] FIG. 2 shows the results measured for activity of recombinant adenoviruses by flow cytometry.

[0057] FIG. 3 shows the results detected for H1299 cells contained in blood samples.

[0058] FIG. 4 shows the results detected for A549 cells contained in blood samples.

[0059] FIG. 5 shows the results measured for activity of the recombinant adenovirus of the present invention in various types of cancer cells.

[0060] FIG. 6 shows the results detected for cancer cells having undergone epithelial-mesenchymal transition (EMT).

[0061] FIG. 7 shows the results detected for cancer stem cells.

[0062] FIG. 8 shows the results detected for H1299 and T24 cells contained in blood samples by using a red fluorescent protein.

[0063] FIG. 9 shows the results detected for H661 cells contained in blood samples.

[0064] FIG. 10 shows the results measured for activity of the recombinant adenovirus of the present invention in HeLa and K562 cells.

DESCRIPTION OF EMBODIMENTS

[0065] The present invention will be described in more detail below. The following embodiments are illustrated to describe the present invention, and it is not intended to limit the present invention only to these embodiments. The present invention can be implemented in various modes, without departing from the spirit of the present invention. Moreover, this specification incorporates the contents disclosed in the specification and drawings of Japanese Patent Application No. 2011-181414 (filed on Aug. 23, 2011), based on which the present application claims priority.

1. SUMMARY

[0066] TelomeScan (i.e., a conditionally replicating adenovirus comprising hTERT promoter, E1A gene, IRES sequence and E1B gene integrated in this order into the E1-deficient region of adenovirus type 5 and comprising cytomegalovirus (CMV) promoter and GFP integrated in this order into the E3-deficient region of adenovirus type 5),

which has been previously developed by the inventors of the present invention, has problems in that: (i) TelomeScan may not detect highly malignant cancer cells where CAR expression is reduced; and (ii) TelomeScan may detect normal blood cells as false positive. As a result of extensive and intensive efforts made to solve these problems, the inventors of the present invention have found that highly malignant CAR-negative cancer cells can be detected when the fiber of adenovirus type 5 in TelomeScan is replaced with another adenovirus fiber binding to CD46, which is highly expressed on almost all human cells, particularly cancer cells in general. Moreover, the inventors of the present invention have also found that when a target sequence of miR-142-3p, which is miRNA, is integrated into each of the replication and labeling cassettes in TelomeScan, virus growth and labeling protein expression can be prevented in normal blood cells to thereby prevent the occurrence of false positive results in normal blood cells.

[0067] Namely, in a preferred embodiment of the present invention, the recombinant adenovirus of the present invention is a recombinant adenovirus, in which a replication cassette comprising hTERT promoter, E1A gene, IRES sequence, E1B gene and a target sequence of microRNA is integrated into the E1 region of the adenovirus genome and a labeling cassette comprising a reporter gene, a promoter capable of regulating the expression of the gene and a target sequence of microRNA is integrated into the E3 region of the adenovirus genome, and which comprises a gene encoding a CD46-binding adenovirus fiber protein (FIG. 1). This recombinant adenovirus has the following features.

(i) Because of comprising a gene encoding a CD46-binding adenovirus fiber protein, this recombinant adenovirus is able to infect almost all cells including CAR-negative cells.

(ii) Because of comprising hTERT promoter, this recombinant adenovirus grows specifically in hTERT-expressing cancer cells and also increases reporter gene expression upon growth, whereby the production of a labeling protein, a chromophore or the like can be increased to detectable levels.

(iii) Because of comprising a target sequence of miRNA, this recombinant adenovirus can prevent the occurrence of false positive results even when the virus infects normal cells having hTERT promoter activity, because expression of this miRNA prevents not only growth of the virus but also expression of the reporter gene. In particular, because of comprising a target sequence of miRNA which is expressed specifically in blood cells, this recombinant adenovirus can prevent the occurrence of false positive results even when the virus infects normal blood cells having hTERT promoter activity, because expression of this miRNA prevents not only growth of the virus in blood cells but also expression of the reporter gene.

[0068] The present invention has been completed on the basis of these findings.

2. RECOMBINANT ADENOVIRUS

(1) Replication Cassette

[0069] The present invention relates to a polynucleotide, which comprises human telomerase reverse transcriptase (hTERT) promoter, E1A gene, IRES sequence and E1B gene in this order and which comprises a target sequence of microRNA. In addition, the present invention relates to a recombinant adenovirus, which comprises a replication cas-

sette comprising the above polynucleotide, wherein the replication cassette is integrated into the E1 region of the adenovirus genome.

[0070] By the action of the above polynucleotide (or a replication cassette comprising the same), the recombinant adenovirus of the present invention can grow specifically in cancer cells and can also be prevented from growing in cells which express the desired miRNA. For example, if the target sequence of miRNA contained in the replication cassette of the present invention is a target sequence of miRNA which is expressed specifically in blood cells, the recombinant adenovirus of the present invention grows specifically in hTERT-expressing cancer cells and is prevented from growing in blood cells.

[0071] Human telomerase reverse transcriptase (hTERT) promoter is a promoter for reverse transcriptase which is an element of human telomerase. Although human telomerase activity will be increased by splicing of hTERT mRNA, post-translational modification of hTERT protein and other events, enhanced hTERT gene expression, i.e., increased hTERT promoter activity is thought to be the most important molecular mechanism. Human telomerase has been confirmed to show increased activity in 85% or more of human cancers, whereas it shows no activity in most normal cells. Thus, the use of hTERT promoter allows a gene downstream thereof to be expressed specifically in cancer cells. In the present invention, the hTERT promoter is located upstream of E1A gene, IRES sequence and E1B gene, whereby the virus can grow specifically in hTERT-expressing cancer cells.

[0072] hTERT has been confirmed to have many transcription factor binding sequences in a 1.4 kbp region upstream of its 5'-terminal end, and this region is regarded as hTERT promoter. In particular, a 181 bp sequence upstream of the translation initiation site is a core region important for expression of its downstream genes. In the present invention, although any sequence may be used as long as it includes this core region, an upstream sequence of approximately 378 bp which covers this core region in its entirety is preferred for use as the hTERT promoter. This sequence of approximately 378 bp has been confirmed to have the same efficiency of gene expression as the 181 bp core region alone. The nucleotide sequence of a 455 bp long hTERT promoter is shown in SEQ ID NO: 1.

[0073] In addition to the sequence shown in SEQ ID NO: 1, the nucleotide sequence of hTERT promoter includes the nucleotide sequences of polynucleotides which are hybridizable under stringent conditions with DNA consisting of a nucleotide sequence complementary to DNA consisting of SEQ ID NO: 1 and which have hTERT promoter activity. Such polynucleotides may be obtained from cDNA and genomic libraries by known hybridization techniques (e.g., colony hybridization, plaque hybridization, Southern blotting) using a polynucleotide which consists of the nucleotide sequence shown in SEQ ID NO: 1 or a fragment thereof as a probe.

[0074] For preparation of cDNA libraries, reference may be made to "Molecular Cloning, A Laboratory Manual 2nd ed." (Cold Spring Harbor Press (1989)). Alternatively, commercially available cDNA and genomic libraries may also be used for this purpose.

[0075] Stringent conditions in the above hybridization include, for example, conditions of 1×SSC to 2×SSC, 0.1% to 0.5% SDS and 42° C. to 68° C., more specifically

prehybridization at 60° C. to 68° C. for 30 minutes or longer and the subsequent 4 to 6 washings in 2×SSC, 0.1% SDS at room temperature for 5 to 15 minutes.

[0076] As to detailed procedures for hybridization, reference may be made to "Molecular Cloning, A Laboratory Manual 2nd ed." (Cold Spring Harbor Press (1989); particularly Section 9.47-9.58), etc.

[0077] In a preferred embodiment of the invention, the hTERT promoter comprises the nucleotide sequence of SEQ ID NO: 1 or a variant thereof. The variant of the hTERT promoter comprises the nucleotide sequence having at least 90%, preferably at least 95%, more preferably at least 96%, and even more preferably at least 97%, 98%, or 99% identity to that of SEQ ID NO: 1, wherein the nucleotide sequence has an hTERT promoter activity.

[0078] E1A and E1B genes are both included in the E1 gene of adenovirus. This E1 gene refers to one of the early genes among the virus early (E) and late (L) genes related to DNA replication, and it encodes a protein related to the regulation of viral genome transcription. E1A protein encoded by the E1A gene of adenovirus activates the transcription of a group of genes (e.g., E1B, E2, E4) required for infectious virus production. E1B protein encoded by the E1B gene of adenovirus assists late gene (L gene) mRNAs to accumulate into the cytoplasm of infected host cells and inhibits protein synthesis in the host cells, thereby facilitating virus replication. The nucleotide sequences of the E1A and E1B genes are shown in SEQ ID NO: 2 and SEQ ID NO: 3, respectively. In addition to the sequences shown in SEQ ID NO: 2 and SEQ ID NO: 3, the nucleotide sequences of the E1A and E1B genes include nucleotide sequences which are hybridizable under stringent conditions with DNA consisting of a nucleotide sequence complementary to DNA consisting of SEQ ID NO: 2 or SEQ ID NO: 3 and which encode a protein having E1A or E1B activity. Procedures and stringent conditions for hybridization are the same as those described above for the hTERT promoter.

[0079] In a preferred embodiment of the invention, the E1A gene comprises the nucleotide sequence of SEQ ID NO: 2 or a variant thereof. The variant of E1A gene comprises the nucleotide sequence having at least 90%, preferably at least 95%, more preferably at least 96%, and even more preferably at least 97%, 98%, or 99% identity to that of SEQ ID NO: 2, wherein the nucleotide sequence encodes a polypeptide having an E1A activity.

[0080] In a preferred embodiment of the invention, the E1B gene comprises the nucleotide sequence of SEQ ID NO: 3 or a variant thereof. The variant of E1B gene comprises the nucleotide sequence having at least 90%, preferably at least 95%, more preferably at least 96%, and even more preferably at least 97%, 98%, or 99% identity to that of SEQ ID NO: 3, wherein the nucleotide sequence encodes a polypeptide having an E1B activity.

[0081] Identity of any two nucleotide sequences can be determined by those skilled in the art based, for example, on a BLAST 2.0 computer alignment. BLAST 2.0 searching is known in the art and is publicly available, for example, at ncbi.nlm.nih.gov/BLAST/.

[0082] IRES (internal ribosome entry site) sequence is a protein synthesis initiation signal specific to the picomavirus family and is considered to serve as a ribosomal binding site because of having a sequence complementary to the 3'-terminal end of 18S ribosomal RNA. It is known that translation of mRNAs derived from viruses of the picomavirus

family is mediated by this sequence. The efficiency of translation from the IRES sequence is high and protein synthesis occurs even from the middle of mRNA in a manner not dependent on the cap structure. Thus, in the virus of the present invention, the E1A gene and the E1B gene, which is located downstream of the IRES sequence, are both translated independently by the action of hTERT promoter. With the use of the IRES sequence, hTERT promoter-mediated expression regulation occurs independently in both the E1A gene and the E1B gene, and hence virus growth can be more strictly limited to cells having telomerase activity when compared to the case where any one of the E1A gene or the E1B gene is regulated by the hTERT promoter. Moreover, the IRES sequence inserted between the E1A gene and the E1B gene can increase the growth capacity of the virus in host cells. The nucleotide sequence of the IRES sequence is shown in SEQ ID NO: 4. In addition to the sequence shown in SEQ ID NO: 4, the nucleotide sequence of the IRES sequence includes nucleotide sequences which are hybridizable under stringent conditions with DNA consisting of a nucleotide sequence complementary to DNA consisting of SEQ ID NO: 4 and which encode a protein having IRES activity. Procedures and stringent conditions for hybridization are the same as those described above for the hTERT promoter.

[0083] In a preferred embodiment of the invention, the IRES sequence comprises the nucleotide sequence of SEQ ID NO: 4 or a variant thereof. The variant of the IRES sequence comprises the nucleotide sequence having at least 90%, preferably at least 95%, more preferably at least 96%, and even more preferably at least 97%, 98%, or 99% identity to that of SEQ ID NO: 4, wherein the nucleotide sequence has an IRES activity.

[0084] miRNA generally refers to short single-stranded RNA of approximately 15 to 25 nucleotides and is considered to regulate the translation of various genes upon binding to its target sequence present in mRNA. Thus, for example, when miRNA-expressing cells are infected with a recombinant adenovirus comprising a desired gene and a target sequence of the miRNA, the desired gene is prevented from being expressed in these cells. Such a target sequence of miRNA may be inserted into any site as long as a desired gene is prevented from being expressed, but it preferably inserted into an untranslated region of the desired gene, more preferably downstream of the desired gene.

[0085] The target sequence of miRNA to be used in the present invention includes target sequences of miRNAs which are expressed in non-cancer cells. Non-cancer cells are intended to mean cells that are not malignant tumor cells, and examples include normal cells, benign tumor cells and so on. Normal cells include, for example, normal blood cells, normal endothelial cells, normal fibroblasts, normal stem cells and so on. On the other hand, circulating tumor cells are regarded as cells originating from malignant tumors, and hence they fall within malignant tumor cells in the present invention.

[0086] The target sequence of miRNA to be used in the present invention also includes target sequences of miRNAs which are expressed specifically in blood cells. In the present invention, "blood cells" may include not only normal blood cells, but also cancerous blood cells. Namely, in the present invention, "miRNA which is expressed specifically in blood cells" may be expressed specifically in normal blood cells or may be expressed specifically in both normal

blood cells and cancerous blood cells. Even when expressed specifically in both normal blood cells and cancerous blood cells, miRNA can also reduce false positive cases of normal blood cells during detection of circulating tumor cells and thereby ensures accurate detection of circulating tumor cells released from solid cancers. In the present invention, "miRNA which is expressed specifically in blood cells" is more preferably miRNA which is expressed in normal blood cells but is not expressed in cancerous blood cells.

[0087] In the present invention, blood cells include, but are not limited to, leukocytes (i.e., neutrophils, eosinophils, basophils, lymphocytes (T cells and B cells), monocytes, dendritic cells), CD34-positive cells, hematopoietic cells, hematopoietic stem cells, hematopoietic progenitor cells, peripheral blood mononuclear cells (PBMCs) and so on. Likewise, cancerous blood cells include leukemia cells, lymphoma cells and so on. In the present invention, being "expressed specifically" in certain cells is intended to mean not only that expression is limited only to the intended cells, but also that expression levels are higher in the intended cells than in other cells. For example, being "expressed specifically in blood cells" is intended to mean not only that expression is limited only to blood cells, but also that expression levels are higher in blood cells than in any cells other than blood cells.

[0088] miRNA which is expressed specifically in blood cells includes, for example, miR-142, miR-15, miR-16, miR-21, miR-126, miR-181, miR-223, miR-296 and so on, with miR-142, miR-15 and miR-16 being preferred.

[0089] Although miRNA is single-stranded RNA, it is possible to use a target sequence of either strand of premature double-stranded RNA as long as a desired gene can be prevented from being expressed. For example, there are miR-142-3p and miR-142-5p for miR-142, and a target sequence of either miRNA may be used in the present invention. Namely, in the present invention, "miR-142" includes both miR-142-3p and miR-142-5p, with miR-142-3p being preferred. Likewise, in the present invention, "miR-15" includes the sense strand (referred to as "miR-15S") and antisense strand (referred to as "miR-15AS") of premature double-stranded RNA. The same applies to other miRNAs.

[0090] miR-142-3p gene is located at a site where translocation occurs in B cell leukemia (aggressive B cell leukemia), and is known to be expressed in hematopoietic tissues (e.g., bone marrow, spleen, thymus), but not expressed in other tissues. Moreover, miR-142-3p has been observed to be expressed in mouse fetal liver (fetal hematopoietic tissue) and hence is considered to be involved in differentiation of the hematopoietic system (Chang-Zheng Chen, et al., Science, 2004).

[0091] In this embodiment, gene expression is regulated in two stages in a selective manner, because specific gene expression is caused in cancer cells by the action of hTERT promoter and gene expression in blood cells is regulated by the action of miRNA.

[0092] In another embodiment, the target sequence of miRNA to be used in the present invention includes a target sequence of miRNA whose expression is suppressed in cancer cells. miRNA whose expression is suppressed in cancer cells includes, for example, miR-125, miR-143, miR-145, miR-199, let-7 and so on. In this embodiment, specific gene expression in cancer cells is doubly regulated by the action of hTERT promoter and miRNA.

[0093] Although miRNA molecules have been initially found in nematodes, yeast and other organisms, there are currently found several hundreds of miRNAs in humans and mice. The sequences of these miRNAs are known, and sequence information and so on can be obtained by access to public DBs (e.g., miR Base: the microRNA database (world wide web.mirbase.org), microRNA.org—Targets and Expression (world wide web.microrna.org/microrna/home.do)).

[0094] The sequences of miR-142, miRNA-15, miRNA-16, miR-21, miR-126, miR-181, miR-223, miR-296, miR-125, miR-143, miR-145, miR-199 and let-7 are shown below.

miR-142-3p: (SEQ ID NO: 5)
5' - UGUAGUGUUUCCUACUUUAUGGA

miR-142-5p: (SEQ ID NO: 6)
5' - CAUAAAGUAGAAAGCACUACU

miR-15S: (SEQ ID NO: 7)
5' - UAGCAGCACAUAAUGGUUUGUG

miR-15AS: (SEQ ID NO: 8)
5' - CAGGCCAUUUGUGCUGCCUCA

miR-16S: (SEQ ID NO: 9)
5' - UAGCAGCACGUAUUUUGGCG

miR-16AS: (SEQ ID NO: 10)
5' - CCAGUAUUUACUGUGCUGCUGA

miR-21S: (SEQ ID NO: 11)
5' - UAGCUUUAUCAGACUGAUGUUGA

miR-21AS: (SEQ ID NO: 12)
5' - CAACACCAGUCGAUGGGCUGU

miR-126S: (SEQ ID NO: 13)
5' - UCGUACCUGAGUAUUUUGGCG

miR-126AS: (SEQ ID NO: 14)
5' - CAUUUUUACUUUUGGUACGCG

miR-181: (SEQ ID NO: 15)
5' - AACAUUCAACGCUGUCGGUGAGU

miR-223S: (SEQ ID NO: 16)
5' - UGUCAGUUUGUCAAAUACCCCA

rniR-223AS: (SEQ ID NO: 17)
5' - CGUGUAUUUGACAAGCUGAGUU

miR-296-3p: (SEQ ID NO: 18)
5' - GAGGGUUGGGUGGAGGCUCUCC

miR-296-5p: (SEQ ID NO: 19)
5' - AGGGCCCCCUCUAAUCCUGU

-continued

miR-125: (SEQ ID NO: 20)
5' - UCCCUGAGACCCUUUAACCGUGA

miR-143S: (SEQ ID NO: 21)
5' - UGAGAUGAAGCACUGUAGCUC

miR-143AS: (SEQ ID NO: 22)
5' - GGUGCAGUGCUGCAUCUCUGGU

miR-145S: (SEQ ID NO: 23)
5' - GUCCAGUUUCCAGGAUCCCU

miR-145AS: (SEQ ID NO: 24)
5' - GGAUCCUGGAAAACUGUUCU

miR-199: (SEQ ID NO: 25)
5' - CCCAGUGUUCAGACUACCUGUUC

let-7: (SEQ ID NO: 26)
5' - UGAGGUAGUAGGUUGUAUAGUU

[0095] In the present invention, a single unit of a target sequence of miRNA is composed of a sequence complementary to the whole or part of the miRNA, and has a nucleotide length of 7 to 30 nucleotides, preferably 19 to 25 nucleotides, more preferably 21 to 23 nucleotides. In the present invention, a single unit of a target sequence of miRNA is intended to mean a nucleotide sequence having the minimum length required for serving as a target of certain miRNA. More specifically, it is intended to mean an oligonucleotide of at least 7 nucleotides in length selected from complementary sequences of the nucleotide sequences shown in SEQ ID NOs: 5 to 26, and such an oligonucleotide may comprise substitution, deletion, addition or removal of one or several nucleotides at any site(s).

[0096] The target sequence as a whole to be integrated into the polynucleotide or recombinant adenovirus of the present invention may comprise several copies of a single unit of target sequence in order to ensure effective interaction between miRNA and the target sequence. The target sequence as a whole to be integrated into the recombinant adenovirus may be of any length as long as it can be integrated into the viral genome. For example, it may comprise 1 to 10 copies, preferably 2 to 6 copies, and more preferably 2 or 4 copies of a single unit of target sequence (John G. Doench, et al., Genes Dev. 2003 17:438-442). An oligonucleotide of appropriate length may be inserted between single units of target sequence contained in the target sequence as a whole. The length of such an oligonucleotide of appropriate length is not limited in any way as long as the target sequence as a whole can be integrated into the recombinant adenovirus genome. For example, such an oligonucleotide may be of 0 to 8 nucleotides in length. Moreover, in the case of comprising several units of a target sequence of miRNA, the target sequences in the respective units may be those toward the same miRNA or those toward different miRNAs. Furthermore, in the case of comprising target sequences toward the same miRNA, the target sequences in the respective units may have different lengths and/or different nucleotide sequences.

[0097] The target sequence of miRNA to be contained in the polynucleotide of the present invention (or a replication

cassette comprising the same) can also be referred to as a “target sequence of a first microRNA” in order that the polynucleotide, when integrated into the recombinant adenovirus, should be distinguished from other miRNA target sequences present in the recombinant adenovirus.

[0098] When miR-142-3p is used as miRNA in the present invention, a target sequence thereof may be exemplified by sequences comprising the following sequences, by way of example.

(i) Sequence comprising two units of a target sequence of miR-142-3p:

(SEQ ID NO: 27)
5'-gcgccctccataaaagtaggaaacactacacagctccataaaagtagga
aacactacattataagcggtac,

each underline represents a single unit of a target sequence of miR-142-3p)

(ii)-1. Sequence comprising four units of a target sequence of miR-142-3p:

(SEQ ID NO: 28)
5'-ggcctccataaaagtaggaaacactacacagctccataaaagtagga
caactacattaatccataaaagtaggaaacactacaccactccataaaagt
ggaaacactacagctac,

each underline represents a single unit of a target sequence of miR-142-3p)

(ii)-2. Sequence comprising four units of a target sequence of miR-142-3p:

(SEQ ID NO: 55)
5'-ccataaaagtaggaaacactacacagctccataaaagtagga
acattaatccataaaagtaggaaacactacaccactccataaaagt
acactaca

each underline represents a single unit of a target sequence of miR-142-3p)

[0099] The nucleotide sequence of SEQ ID NO: 55 corresponds to the sequence from position 5 to 109 in the nucleotide sequence of SEQ ID NO: 28.

(ii)-3. Sequence comprising four units of a target sequence of miR-142-3p:

(SEQ ID NO: 52)
5'-ccataaagtaggaaacactacacagctccataaagtacgaaacactac
attaatccataaaagtaggaaacactacaggactccataaagt
aggaacactaca-3',

each underline represents a single unit of a target sequence of miR-142-3p)

[0100] In a preferred embodiment of the invention, the sequence comprising four units of a target sequence of miR-142-3p comprises the nucleotide sequence of SEQ ID NO: 52 or 55, or a variant thereof. The variant of the sequence comprising four units of a target sequence of miR-142-3p comprises the nucleotide sequence having at least 90%, preferably at least 95%, more preferably at least 96%, and even more preferably at least 97%, 98%, or 99% identity to that of SEQ ID NO: 52 or 55.

[0101] The nucleotide sequence of SEQ ID NO: 55 has 98.1% (103/105) identity to the nucleotide sequence of SEQ ID NO: 52. The nucleotide sequence of SEQ ID NO: 52 has 98.1% identity to the nucleotide sequence of SEQ ID NO:

55. That is, the nucleotide sequences of SEQ ID NOs: 52 and 55 have 98.1% (103/105) nucleotide identity each other.

[0102] In the present invention, a target sequence of miRNA is placed downstream of the construct of hTERT promoter-E1A gene-IRES sequence-E1B gene, and the resulting polynucleotide comprising the hTERT promoter, the E1A gene, the IRES sequence, the E1B gene and the target sequence of miRNA in this order (which polynucleotide is referred to as a replication cassette) is integrated into the adenovirus genome, whereby E1 gene expression and virus growth can be prevented in cells expressing the miRNA.

[0103] In the present invention, a target sequence of miRNA is integrated downstream of the E1B gene or the reporter gene described later, whereby a gene located upstream thereof is prevented from being expressed. Although the details of this mechanism are not clear, a possible mechanism is as follows. First, miRNA-RISC (RNA-induced silencing complex) cleaves a target sequence on mRNA to thereby remove polyA from the mRNA. This would reduce the stability of the mRNA to cause degradation of the mRNA and hence prevention of gene expression. Alternatively, miRNA-RISC would recruit polyA ribonuclease, as in the case of normal miRNA, to cause polyA degradation, as a result of which the stability of mRNA would be reduced and gene expression would be prevented.

[0104] It should be noted that there are previous reports showing that the miRNA-induced inhibitory effect against gene expression was not obtained for the expression (translation) of a gene inserted downstream of the IRES sequence (Ramesh S. Pillai et al., Science 309, 1573(2005); Geraldine Mathonnet, et al., Science 317, 1764 (2007)). However, when the inventors of the present invention confirmed gene expression for the recombinant adenovirus of the present invention comprising hTERT promoter, E1A gene, IRES sequence, E1B gene and a target sequence of miRNA in this order, the miRNA was found to sufficiently prevent the expression of the E1B gene inserted downstream of the IRES sequence. This is a new finding in the present invention.

[0105] The genes to be contained in the replication cassette of the present invention can be obtained by standard genetic engineering techniques. For example, it is possible to use nucleic acid synthesis with a DNA synthesizer, which is commonly used as a genetic engineering technique. Alternatively, it is also possible to use PCR techniques in which gene sequences serving as templates are isolated or synthesized, and primers specific to each gene are then designed to amplify the gene sequence with a PCR system (Current Protocols in Molecular Biology, John Wiley & Sons (1987) Section 6.1-6.4) or gene amplification techniques using a cloning vector. The above techniques can be easily accomplished by those skilled in the art in accordance with Molecular cloning 2nd Edt. Cold Spring Harbor Laboratory Press (1989), etc. For purification of the resulting PCR product, known techniques can be used. If necessary, conventionally used sequencing techniques may be used to confirm whether the intended gene has been obtained, as expected. For example, dideoxynucleotide chain termination sequencing (Sanger et al. (1977) Proc. Natl. Acad. Sci. USA 74: 5463) or the like may be used for this purpose. Alternatively, an appropriate DNA sequencer (e.g., ABI PRISM (Applied Biosystems)) may also be used for sequence analysis.

[0106] In the present invention, the target sequence of miRNA can be obtained by being designed and synthesized such that each single unit of target sequence is complementary to the whole or part of the nucleotide sequence of the miRNA. For example, a target sequence of miR-142-3p can be obtained by synthesizing DNA such that it is complementary to the nucleotide sequence of miR-142-3p.

[0107] Then, the respective genes obtained as above are ligated in a given order. First, the above genes are each cleaved with known restriction enzymes or the like, and the cleaved DNA fragment of each gene is inserted into and ligated to a known vector in accordance with known procedures. As a known vector, pIRES vector may be used, by way of example. The pIRES vector comprises the IRES (internal ribosome entry site) sequence of encephalomyocarditis virus (ECMV) and is capable of translating two open reading frames (ORFs) from one mRNA. With the use of the pIRES vector, it is possible to prepare a "polynucleotide which comprises hTERT promoter, E1A gene, IRES sequence and E1B gene in this order and which comprises a target sequence of microRNA" by sequentially inserting the required genes into a multicloning site. Such a target sequence of miRNA may be inserted into any site, but it is preferably inserted downstream of the hTERT promoter-E1A-IRES-E1B construct. For DNA ligation, DNA ligase may be used. Alternatively, CMV promoter contained in a known vector (e.g., pShuttle) may be removed with known restriction enzymes and a sequence cleaved from the hTERT promoter-E1A-IRES-E1B-miRNA target sequence with appropriate restriction enzymes may then be inserted into this site, if necessary. Once the E1 gene required for adenovirus growth is allowed to be expressed under the control of the hTERT promoter, the virus can be grown specifically in cancer cells.

[0108] In a preferred embodiment of the invention, hTERT promoter and E1A gene, E1A and IRES sequence, IRES sequence and E1B gene, and E1B gene and miRNA target sequence may be linked by spacer sequences.

[0109] In another preferred embodiment of the invention, the replication cassette comprises the nucleotide sequence of SEQ ID NO: 53 or a variant thereof. The variant of the polynucleotide cassette comprises the nucleotide sequence having at least 90%, preferably at least 95%, more preferably at least 96%, and even more preferably at least 97%, 98%, or 99% identity to that of SEQ ID NO: 53, wherein the nucleotide sequence encodes a polypeptide having an activity to enable an adenovirus to replicate selectively in tumor cells. The activity can be determined by infecting the adenovirus comprising the replication cassette to tumor cells and measuring viral titer of the adenovirus.

[0110] (2) Labeling Cassette

[0111] In yet another embodiment, the present invention relates to a recombinant adenovirus in which the above replication cassette is integrated into the E1 region of the adenovirus genome and a labeling cassette is further integrated into the E3 region of the adenovirus genome. Such a labeling cassette comprises a reporter gene and a promoter capable of regulating the expression of the gene, and may further comprise a target sequence of miRNA.

[0112] The adenovirus E3 region contains 11.6 kDa ADP (adenovirus death protein), and ADP has the function of promoting cell damage and virus diffusion. The recombinant adenovirus of the present invention is designed to eliminate any viral genome region like the E3 region containing ADP,

which encodes a protein having the function of promoting cell damage and virus diffusion, so that the timing of cell death is delayed to facilitate identification of cancer tissues by production (emission, expression) of fluorescence (e.g., GFP). This is also effective in that circulating tumor cells (CTCs) described later can be detected alive over a long period of time.

[0113] The reporter gene to be contained in the labeling cassette in the recombinant adenovirus of the present invention is not limited in any way, and examples include a gene encoding a protein which emits fluorescence, a gene encoding an enzyme protein which generates a luminophore or a chromophore upon enzymatic reaction, a gene encoding an antibiotic, a gene encoding a tag-fused protein, a gene encoding a protein which is expressed on the cell surface and binds to a specific antibody, a gene encoding a membrane transport protein, and so on. Examples of a protein which emits fluorescence (i.e., a labeling protein) include a green fluorescent protein (GFP) derived from luminous jellyfish such as *Aequorea victoria*, its variants EGFP (enhanced-humanized GFP) and rsGFP (red-shift GFP), a yellow fluorescent protein (YFP), a cyan fluorescent protein (CFP), a blue fluorescent protein (BFP). GFP derived from *Renilla reniformis* and so on, and genes encoding these proteins can be used in the present invention. The above protein which emits fluorescence is preferably GFP or EGFP.

[0114] Likewise, examples of an enzyme protein which generates a luminophore or a chromophore upon enzymatic reaction include β -galactosidase, luciferase and so on. β -Galactosidase generates a blue chromophore from 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal) upon enzymatic reaction. On the other hand, luciferase generates a luminophore upon enzymatic reaction with luciferin. Firefly luciferase, bacterial luciferase, *Renilla* luciferase and so on are known as members of luciferase, and those skilled in the art would be able to select an appropriate enzyme from known luciferase members.

[0115] Moreover, the promoter capable of regulating the expression of the above gene is not limited in any way as long as it is a suitable promoter compatible with the virus used for the expression of the above desired gene. Examples include, but are not limited to, CMV promoter, hTERT promoter, SV40 late promoter, MMTV LTR promoter, RSV LTR promoter, SR α promoter, β -actin promoter, PGK promoter, EF-1a promoter and so on. Preferably, CMV promoter or hTERT promoter can be used for this purpose.

[0116] The target sequence of miRNA to be integrated into the labeling cassette may be either the same or different from the target sequence of miRNA to be integrated into the replication cassette.

[0117] In the present invention, the target sequence of miRNA is placed within the untranslated region of the reporter gene, preferably downstream of this gene, whereby the reporter gene can be prevented from being expressed. Namely, in the present invention, the labeling cassette preferably comprises a promoter capable of regulating the reporter gene, the reporter gene and the target sequence of microRNA in this order. The target sequence of miRNA to be integrated into the labeling cassette is referred to as a "target sequence of a second microRNA" in order that it should be distinguished from the target sequence of miRNA to be contained in the replication cassette. Other explanations on miRNA are the same as described above.

[0118] In a preferred embodiment of the invention, the labeling cassette preferably comprises the sequence comprising four units of a target sequence of miR-142-3p. The sequence comprises the nucleotide sequence of SEQ ID NO: 52 or 55, or a variant thereof. The variant of the sequence comprising four units of a target sequence of miR-142-3p comprises the nucleotide sequence having at least 90%, preferably at least 95%, more preferably at least 96%, and even more preferably at least 97%, 98%, or 99% identity to that of SEQ ID NO: 52 or 55.

[0119] In a preferred embodiment of the invention, examples of a replication cassette include a replication cassette comprising CMV promoter, EGFP gene and a target sequence of miR-142p. In a preferred embodiment of the invention, the replication cassette comprising CMV promoter, EGFP gene and a target sequence of miR-142p comprises the nucleotide sequence of SEQ ID NO: 54 or a variant thereof. The variant of the polynucleotide cassette comprises the nucleotide sequence having at least 90%, preferably at least 95%, more preferably at least 96%, and even more preferably at least 97%, 98%, or 99% identity to that of SEQ ID NO: 54.

[0120] Details on how to obtain, purify and sequence the recombinant genes to be contained in the labeling cassette of the present invention are the same as described above for the replication cassette.

[0121] (3) Cell Death-Inducing Cassette

[0122] In yet another embodiment, the present invention relates to a recombinant adenovirus in which the above replication cassette is integrated into the E1 region of the adenovirus genome and a cell death-inducing cassette is integrated into the E3 region of the adenovirus genome. Such a cell death-inducing cassette comprises a gene encoding a cell death induction-related protein and a promoter capable of regulating the expression of the gene, and may further comprise a target sequence of microRNA.

[0123] The cell death-inducing cassette used in the recombinant adenovirus of the present invention comprises a gene encoding a cell death induction-related protein and a promoter capable of regulating the expression of the gene. Thus, for example, when the recombinant adenovirus of the present invention is infected into cancer cells, the virus grows specifically in the cancer cells to thereby increase the intracellular expression level of the cell death induction-related protein and induce cell death only in the cancer cells without damaging other normal cells.

[0124] Such a gene encoding a cell death induction-related protein is intended to mean a gene encoding a protein related to the induction of cell death in specific cells. Examples of a cell death induction-related protein include immunological proteins such as PA28. PA28 is a protein which activates intracellular proteasomes and which elicits immune reactions and also induces cell death when overexpressed. Moreover, TRAIL can also be exemplified as an apoptosis-inducing protein. TRAIL refers to a molecule which induces apoptotic cell death upon binding to its receptor on the cell surface.

[0125] Moreover, another example of the gene encoding a cell death induction-related protein is a tumor suppressor gene, which has the function of suppressing the growth of cancer cells. Examples of such a tumor suppressor gene include the following genes used in conventional gene therapy. SEQ ID NO (nucleotide sequence) and GenBank Accession No. are shown below for each gene.

[0126] p53 (SEQ ID NO: 29; Accession No. M14694): multiple types of cancer

[0127] p15 (SEQ ID NO: 30; Accession No. L36844): multiple types of cancer

[0128] p16 (SEQ ID NO: 31; Accession No. L27211): multiple types of cancer

[0129] APC (SEQ ID NO: 32; Accession No. M74088): colorectal cancer, gastric cancer, pancreatic cancer

[0130] BRCA-1 (SEQ ID NO: 33; Accession No. U14680): ovarian cancer, breast cancer

[0131] DPC-4 (SEQ ID NO: 34; Accession No. U44378): colorectal cancer, pancreatic cancer

[0132] FHIT (SEQ ID NO: 35; Accession No. NM 112012): gastric cancer, lung cancer, uterine cancer

[0133] p73 (SEQ ID NO: 36; Accession No. Y11416): neuroblastoma

[0134] PATCHED (SEQ ID NO: 37; Accession No. U59464): basal cell carcinoma

[0135] Rbp1110 (SEQ ID NO: 38; Accession No. M15400): lung cancer, osteosarcoma

[0136] DCC (SEQ ID NO: 39; Accession No. X76132): colorectal cancer

[0137] NF1 (SEQ ID NO: 40; Accession No. NM 000267): neurofibroma type 1

[0138] NF2 (SEQ ID NO: 41; Accession No. L11353): neurofibroma type 2

[0139] WT-1 (SEQ ID NO: 42; Accession No. NM 000378): Wilms tumor

[0140] The target sequence of miRNA to be contained in the cell death-inducing cassette may be either the same or different from the target sequence of miRNA to be integrated into the replication cassette. In the present invention, the target sequence of miRNA is placed within the untranslated region of the gene encoding a cell death induction-related protein, preferably downstream of this gene, whereby the cell death induction-related protein can be prevented from being expressed. Namely, in the present invention, the cell death-inducing cassette preferably comprises a promoter capable of regulating the gene encoding a cell death induction-related protein, the gene encoding a cell death induction-related protein and the target sequence of microRNA in this order. Other explanations on miRNA are the same as described above.

[0141] Details on how to obtain, purify and sequence the recombinant genes to be contained in the cell death-inducing cassette of the present invention are the same as described above for the replication cassette.

[0142] To determine whether or not cell death has been induced, morphological observation described below may be conducted for this purpose. Namely, once cells adhered onto the bottom surface of a culture vessel have been infected with the recombinant virus of the present invention and incubated for a given period, the cells will be rounded and detached from the bottom surface and then will float as shiny cells in the culture solution, as observed under an inverted microscope. At this stage, the cells have lost their vital mechanism and hence a determination can be made that cell death has been induced. Alternatively, cell death can also be confirmed with a commercially available kit for living cell assay which uses a tetrazolium salt (e.g., MTT, XTT).

[0143] (4) CD46-Binding Fiber Protein

[0144] In yet another embodiment, the recombinant adenovirus of the present invention may comprise a gene encoding a CD46-binding adenovirus fiber protein.

[0145] Adenovirus vectors which are now commonly used are prepared structurally based on adenovirus type 5 (or type 2) belonging to Subgroup C among 51 serotypes of human adenovirus. Although adenovirus type 5 is widely used because of its excellent gene transfer properties, adenovirus of this type has a problem of being difficult to infect cells with low expression of coxsackievirus and adenovirus receptor (CAR) because its infection is mediated by binding to CAR on target cells. In particular, CAR expression is reduced in highly malignant cancer cells which are highly invasive, metastatic and proliferative, and hence an adenovirus having the fiber protein of adenovirus type 5 may not infect such highly malignant cancer cells.

[0146] In contrast, CD46 is expressed on almost all cells except for erythrocytes in humans and is also expressed on highly malignant cancer cells. Thus, a recombinant adenovirus comprising a gene encoding a CD46-binding adenovirus fiber protein can also infect CAR-negative and highly malignant cancer cells. For example, adenovirus types 34 and 35 bind to CD46 as their receptor and thereby infect cells (Marko Marttila, et al., *J. Virol.* 2005, 79(22):14429-36). As described above, CD46 is expressed on almost all cells except for erythrocytes in humans, and hence adenovirus types 34 and 35 are able to infect a wide range of cells including CAR-negative cells. Moreover, the fiber of adenovirus consists of a knob region, a shaft region and a tail region, and adenovirus infects cells through binding of its fiber knob region to the receptor. Thus, at least the fiber knob region in the fiber protein is replaced from adenovirus type 5 origin to adenovirus type 34 or 35 origin, % hereby the virus will be able to infect CAR-negative cells via CD46.

[0147] Because of comprising a gene encoding a CD46-binding adenovirus fiber protein, the recombinant adenovirus of the present invention is able to infect almost all cells except for erythrocytes and thus able to infect highly malignant CAR-negative cancer cells which are highly invasive, metastatic and proliferative. In the present invention, "CAR-negative" cells are intended to mean cells where CAR expression is low or cells where CAR is not expressed at all.

[0148] 57 serotypes have now been identified for human adenovirus, and these serotypes are classified into six groups, i.e., Groups A to F. Among them, adenovirus types belonging to Group B have been reported to bind to CD46. Adenovirus types belonging to Group B include adenovirus types 34 and 35, as well as adenovirus types 3, 7, 11, 16, 21 and 50, by way of example.

[0149] For use as a CD46-binding adenovirus fiber protein in the present invention, preferred is the fiber protein of adenovirus belonging to Group B, more preferred is the fiber protein of adenovirus type 3, 7, 34, 35, 11, 16, 21 or 50, and even more preferred is the fiber protein of adenovirus type 34 or 35.

[0150] The nucleotide sequence of a gene encoding the fiber protein of adenovirus type 34, 35, 3, 7, 11, 16, 21 or 50 is available from a known gene information database. e.g., the GenBank of NCBI (The National Center for Biotechnology Information). Moreover, in the present invention, the nucleotide sequence of a gene encoding the fiber protein of adenovirus type 34, 35, 3, 7, 11, 16, 21 or 50 includes not only the nucleotide sequence of each gene available from a

database as described above, but also nucleotide sequences which are hybridizable under stringent conditions with DNA consisting of a nucleotide sequence complementary to DNA consisting of each nucleotide sequence available from a database and which encode a protein with binding activity to CD46.

[0151] The binding activity to CD46 can be evaluated when a recombinant adenovirus having DNA comprising the nucleotide sequence is measured for its infectivity to CD46-expressing cells. The infectivity of such a recombinant adenovirus may be measured in a known manner, for example, by detecting GFP expressed by the virus infected into CD46-expressing cells under a fluorescence microscope or by flow cytometry, etc. Procedures and stringent conditions for hybridization are the same as described above.

[0152] The recombinant adenovirus of the present invention may comprise the entire or partial region of a CD46-binding adenovirus fiber protein, such that at least the fiber knob region in the fiber protein binds to CD46. Namely, in the present invention, the CD46-binding adenovirus fiber protein may comprise at least the fiber knob region in the fiber protein of adenovirus belonging to Group B, more preferably at least the fiber knob region in the fiber protein of adenovirus of any type selected from the group consisting of type 34, type 35, type 3, type 7, type 11, type 16, type 21 and type 50, and even more preferably at least the fiber knob region in the fiber protein of adenovirus type 34 or 35. Moreover, the technical idea of the present invention is not limited to these fiber proteins as long as the intended protein binds to CD46, and it also covers various proteins capable of binding to CD46 as well as proteins having a motif capable of binding to CD46.

[0153] Alternatively, in the present invention, the CD46-binding fiber protein may comprise a region consisting of the fiber knob region and the fiber shaft region in the fiber protein of adenovirus belonging to Group B, more preferably a region consisting of the fiber knob region and the fiber shaft region in the fiber protein of adenovirus of any type selected from the group consisting of type 34, type 35, type 3, type 7, type 11, type 16, type 21 and type 50, and even more preferably a region consisting of the fiber knob region and the fiber shaft region in the fiber protein of adenovirus type 34 or 35.

[0154] In the present invention, the CD46-binding fiber protein may comprise the fiber shaft region or the fiber tail region in the fiber protein of adenovirus of any type (e.g., type 2, type 5) other than the above types, as long as it comprises at least the fiber knob region in the fiber protein of adenovirus belonging to Group B.

[0155] Examples of such a fiber protein include, but are not limited to, fiber proteins which comprise a region consisting of not only the fiber knob region and the fiber shaft region in the fiber protein of adenovirus of any type selected from the group consisting of type 34, type 35, type 3, type 7, type 11, type 16, type 21 and type 50, but also the fiber tail region in the fiber protein of adenovirus type 5.

[0156] The nucleotide sequences of a gene encoding the fiber knob region in the fiber protein of adenovirus type 34, a gene encoding the fiber shaft region in the fiber protein of adenovirus type 34 and a gene encoding a region consisting of the fiber knob region and the fiber shaft region in the fiber protein of adenovirus type 34 are shown in SEQ ID NOs: 47, 48 and 49, respectively.

[0157] Likewise, the nucleotide sequence of a gene encoding a region consisting of not only the fiber knob region and the fiber shaft region in the fiber protein of adenovirus type 34, but also the fiber tail region in the fiber protein of adenovirus type 5 is shown in SEQ ID NO: 50. In the present invention, the nucleotide sequence of such a gene includes not only the nucleotide sequence shown in SEQ ID NO: 50, but also nucleotide sequences which are hybridizable under stringent conditions with DNA consisting of a nucleotide sequence complementary to DNA consisting of the nucleotide sequence shown in SEQ ID NO: 50 and which encode a protein with binding activity to CD46. Procedures for evaluation of the binding activity to CD46, procedures and stringent conditions for hybridization are the same as described above.

[0158] In a preferred embodiment of the invention, the gene encoding a CD46-binding adenovirus fiber protein comprises the nucleotide sequence of SEQ ID NO: 50 or a variant thereof. The variant of the gene encoding a CD46-binding adenovirus fiber protein comprises the nucleotide sequence having at least 90%, preferably at least 95%, more preferably at least 96%, and even more preferably at least 97%, 98%, or 99% identity to the nucleotide sequence of SEQ ID NO: 50.

[0159] To prepare the recombinant adenovirus of the present invention, a polynucleotide comprising the replication cassette, the labeling cassette and/or the cell death-inducing cassette may be excised with appropriate restriction enzymes and inserted into an appropriate virus expression vector. A preferred virus expression vector is an adenovirus vector, more preferably an adenovirus type 5 vector, and particularly preferably an adenovirus type 5 vector which comprises a gene encoding a CD46-binding adenovirus fiber protein (e.g., the fiber protein of adenovirus type 34 or 35).

[0160] As shown in Example 2 described later, GFP expression in blood cells was sufficiently suppressed in both cases where a miRNA target sequence was inserted downstream of the replication cassette and where a miRNA target sequence was inserted downstream of the labeling cassette, whereas GFP expression in blood cells was unexpectedly significantly suppressed in a case where miRNA target sequences were simultaneously inserted downstream of the replication cassette and downstream of the labeling cassette, respectively. This is a new finding in the present invention.

[0161] In the present invention, the recombinant adenovirus may be obtained in the following manner, by way of example.

[0162] First, pHMCMV5 (Mizuguchi H. et al., Human Gene Therapy, 10: 2013-2017, 1999) is treated with restriction enzymes and a target sequence of miRNA is inserted to prepare a vector having the target sequence of miRNA. Next, pSh-hAIB comprising a construct of hTERT promoter-E1A-IRES-E1B (WO2006/036004) is treated with restriction enzymes and the resulting fragment comprising the hTERT promoter-E1A-IRES-E1B construct is inserted into the above vector having the target sequence of miRNA to obtain a vector comprising hTERT promoter-E1A-IRES-E1B-miRNA target sequence. On the other hand, pHMCMV-GFP-1 (pHMCMV5 comprising EGFP gene) is treated with restriction enzymes to obtain a fragment comprising CMV promoter and EGFP gene, and this fragment is inserted into the above vector having the target sequence of miRNA to obtain a vector comprising a construct of CMV-EGFP-miRNA target sequence. Then, the vector comprising

hTERT promoter-E1A-IRES-E1B-miRNA target sequence and the vector comprising CMV-EGFP-miRNA target sequence are each treated with restriction enzymes and ligated together to obtain a vector in which hTERT promoter-E1A-IRES-E1B-miRNA target sequence is integrated into the E1-deficient region of the adenovirus genome and CMV-EGFP-miRNA target sequence is integrated into the E3-deficient region of the adenovirus genome. Alternatively, when a vector comprising a gene encoding a CD46-binding adenovirus fiber protein is used as a vector to be inserted with the DNA fragments comprising the respective constructs, it is possible to obtain a vector in which hTERT promoter-E1A-IRES-E1B-miRNA target sequence is integrated into the E1-deficient region of the adenovirus genome and CMV-EGFP-miRNA target sequence is integrated into the E3-deficient region of the adenovirus genome and which comprises a gene encoding a CD46-binding adenovirus fiber protein. Moreover, this vector may be linearized with a known restriction enzyme and then transfected into cultured cells (e.g., 293 cells) to thereby prepare an infectious recombinant adenovirus. It should be noted that those skilled in the art would be able to easily prepare all viruses falling within the present invention by making minor modifications to the above preparation procedures.

[0163] In a preferred embodiment of the invention, the recombinant adenovirus comprises a nucleotide sequence comprising:

[0164] a replication cassette comprising a polynucleotide comprising a human telomerase reverse transcriptase promoter, E1A gene, IRES sequence and E1B gene in this order, and a target sequence of a first microRNA, wherein the first microRNA is miR-142, and wherein the replication cassette is integrated into the E1 region of the adenovirus genome;

[0165] a labeling cassette comprising a reporter gene, a promoter capable of regulating the expression of the reporter gene, and a target sequence of a second microRNA, wherein the second microRNA is miR-142, and wherein the labeling cassette is integrated into the E3 region of the adenovirus genome; and

[0166] a gene encoding a CD46-binding fiber protein comprising at least the fiber knob region in the fiber protein of adenovirus type 34 or 35, which is integrated into the adenovirus genome;

[0167] wherein the recombinant adenovirus comprises the nucleotide sequence of SEQ ID NO: 51 or a variant thereof. The variant comprises the nucleotide sequence having at least 90%, preferably at least 95%, more preferably at least 96%, and even more preferably at least 97%, 98%, or 99% identity to the nucleotide sequence of SEQ ID NO: 51.

3. REAGENT FOR CANCER CELL DETECTION OR REAGENT FOR CANCER DIAGNOSIS

[0168] As described above, the recombinant adenovirus of the present invention has the following features.

(i) This recombinant adenovirus infects almost all cells except for erythrocytes, and is also able to infect highly malignant CAR-negative cancer cells.

(ii) This recombinant adenovirus grows specifically in hTERT-expressing cancer cells and also increases the expression level of a reporter gene upon growth, whereby the production of a labeling protein, a chromophore or the like can be increased to detectable levels.

(iii) This recombinant adenovirus can prevent the occurrence of false positive results even when the virus infects

normal cells having hTERT promoter activity, because miRNA expression prevents not only growth of the virus, but also expression of a reporter gene. In particular, because of comprising a target sequence of miRNA which is expressed specifically in blood cells, this recombinant adenovirus can prevent the occurrence of false positive results even when the virus infects normal blood cells having hTERT promoter activity, because expression of this miRNA prevents not only growth of the virus in blood cells but also expression of a reporter gene.

[0169] Thus, the recombinant adenovirus of the present invention can be used as a reagent for cancer cell detection or as a reagent for cancer diagnosis. In particular, because of having the above features, the recombinant virus of the present invention is extremely effective for detection of circulating tumor cells (CTCs) present in blood.

[0170] On the other hand, since 2004 when CTCs, which are cancer cells present in blood, were reported to serve as a prognostic factor for post-operative breast cancer patients in the *New England Journal of Medicine* (Cristofanilli M. et al., *The New England Journal of Medicine*, 2004, 781-791), CTCs have been measured as a biomarker in many clinical trials conducted in Europe and North America. Particularly in breast cancer, prostate cancer and skin cancer, CTCs have been proven to be an independent factor which determines the prognosis of these cancers. Moreover, in Europe, in the clinical trial in adjuvant setting of prostate cancer (SUCCESS), the number of CTCs counted is added to the inclusion criteria and only patients in whom one or more cells have been detected are included. This trial is a large-scale clinical trial including 2000 cases or more, and attention is being given to the results. Moreover, there is also a clinical trial in which an increase or decrease per se in CTCs is one of the clinical endpoints (MDV3100).

[0171] In recent years, the FDA in the United States has issued guidelines for approval and authorization of molecular-targeted anticancer agents, and hence the CTC test has become more important in cancer diagnosis. The guidelines issued by the FDA define that genetic changes in molecular targets in tumors should be tested before selection of molecular-targeted anticancer agents. When attempting to achieve the guidelines by conventional techniques, there arises a need for surgical biopsy from tumor tissues in patients to conduct genetic testing, which will impose a very strong burden on the patients. To solve this problem, efforts are now made to conduct genetic testing on CTCs collected from blood, and this strategy is referred to as "liquid biopsy" in contrast to the conventional "biopsy." Once this strategy has been achieved, genetic testing of tumor tissues can be conducted simply by blood collection and the burden on patients can be reduced greatly. For these reasons, the CTC test is receiving great attention as a highly useful testing technique in the clinical setting.

[0172] The CellSearch System of Veridex LLC is the only CTC detection device currently approved by the FDA, and most of the CTC detection methods used in clinical trials are accomplished by this CellSearch System. The CellSearch System is based on techniques to detect cancer cells with EpCAM antibody and cytokeratin antibody.

[0173] However, CTC detection techniques are designed to detect several to several tens of cells from among a billion of blood cells, and it is therefore very difficult to improve their sensitivity and accuracy. Thus, some problems are also pointed out in CTC detection methods based on the Cell-

Search System. For example, it is pointed out that cancer cells which are negative in the CTC test based on the CellSearch System are detected as being positive in another test, and that there are great differences in sensitivity and accuracy, depending on the cancer type (Allard W. J. et al., *Clinical Cancer Research*, 2004, 6897-6904). Moreover, the CellSearch System is also pointed out to have a problem of low CTC detection rate for lung cancer in the clinical setting (*ibid*).

[0174] Likewise, the CellSearch System is also pointed out to have a problem of reduced CTC detection rate because the expression of cell surface antigens including EpCAM is reduced in cancer cells having undergone epithelial-mesenchymal transition (EMT) (Anieta M. et al., *J Natl Cancer Inst*, 101, 2009, 61-66, Janice Lu et al., *Int J Cancer*, 126(3), 2010, 669-683).

[0175] Further, to conduct the above "liquid biopsy," additional steps are required for concentration and phenotyping or genotyping of CTCs, which require more sensitive and more accurate CTC detection techniques than simply counting the number of CTCs.

[0176] In contrast to this, because of having the above features (i) to (iii), the recombinant adenovirus of the present invention allows simple, highly sensitive and highly accurate detection of CTCs in blood without detection of leukocytes and other normal blood cells. Further, the reagent of the present invention allows detection of CTCs alive, so that the source organ of the detected CTCs can be identified upon analyzing surface antigens or the like present on the cell surface of the CTCs. Thus, the recombinant adenovirus of the present invention is useful for CTC detection and cancer diagnosis.

[0177] Moreover, the recombinant adenovirus or reagent for cancer cell detection of the present invention can be used to detect cancer cells having undergone EMT or mesenchymal-epithelial transition (MET). EMT is a phenomenon in which cancer cells lose their properties as epithelium and acquire features as mesenchymal lineage cells tending to migrate into surrounding tissues, and EMT is also involved in invasion and/or metastasis of cancer cells. On the other hand, mesenchymal-epithelial transition (MET) is a phenomenon in which mesenchymally derived cells acquire features as epithelium. As described above, it is difficult to detect cancer cells having undergone EMT by known techniques including the CellSearch System. In contrast, the present invention allows detection of cancer cells having undergone EMT or MET. The recombinant adenovirus of the present invention is therefore useful for cancer cell detection and for cancer diagnosis.

[0178] Further, the recombinant adenovirus of the present invention can also be used to detect drug-resistant cancer cells. Drugs intended in the present invention are those used for cancer chemotherapy. Examples of such drugs include, but are not limited to, adriamycin, carboplatin, cisplatin, 5-fluorouracil, mitomycin, bleomycin, doxorubicin, daunorubicin, methotrexate, paclitaxel, docetaxel and actinomycin D, etc. Moreover, the recombinant virus of the present invention can also be used to detect cancer stem cells. In the present invention, cancer stem cells refer to cells (stem cells) serving as the origin of cancer cells. Cancer stem cells also include those having drug resistance.

[0179] In the present invention, the type of cancer or tumor to be detected or diagnosed is not limited in any way, and cells of all cancer types can be used. Examples include

solid cancers or blood tumors, more specifically brain tumor, cervical cancer, esophageal cancer, tongue cancer, lung cancer, breast cancer, pancreatic cancer, gastric cancer, small intestinal cancer, duodenal cancer, colorectal cancer, bladder cancer, kidney cancer, liver cancer, prostate cancer, uterine cancer, uterine cervical cancer, ovarian cancer, thyroid cancer, gallbladder cancer, pharyngeal cancer, sarcoma, melanoma, leukemia, lymphoma and multiple myeloma (MM). Most (85% or more) of the cancer cells derived from human tissues show increased telomerase activity, and the present invention allows detection of such telomerase-expressing cancer cells in general.

[0180] Moreover, in the present invention, CTCs are not limited in any way as long as they are cancer cells present in blood, and they include not only cancer cells released from solid cancers, but also blood tumor cells such as leukemia cells and lymphoma cells as mentioned above. However, in cases where CTCs are blood tumor cells, the miRNA target sequence contained in the adenovirus of the present invention is preferably a target sequence of miRNA which is expressed specifically in normal blood cells.

[0181] To prepare the reagent of the present invention, the recombinant adenovirus may be treated, e.g., by freezing for easy handling and then used directly or mixed with known pharmaceutically acceptable carriers (e.g., excipients, extenders, binders, lubricants) and/or known additives (including buffering agents, isotonicizing agents, chelating agents, coloring agents, preservatives, aromatics, flavorings, sweeteners).

4. METHOD FOR CANCER CELL DETECTION OR METHOD FOR CANCER DIAGNOSIS

[0182] Furthermore, the recombinant adenovirus of the present invention can be used for cancer cell detection or cancer diagnosis by contacting the same with cancer cells and detecting the fluorescence or color produced by the cancer cells.

[0183] In the present invention, the term “contact(ing)” is intended to mean that cancer cells and the recombinant adenovirus of the present invention are allowed to exist in the same reaction system, for example, by adding the recombinant adenovirus of the present invention to a sample containing cancer cells, by mixing cancer cells with the recombinant adenovirus, by culturing cancer cells in the presence of the recombinant adenovirus, or by infecting the recombinant adenovirus into cancer cells. Moreover, in the present invention, “fluorescence or color” is not limited in any way as long as it is light or color produced from a protein expressed from a reporter gene, and examples include fluorescence emitted from a labeling protein (e.g., GFP), light emitted from a luminophore generated by luciferase-mediated enzymatic reaction, blue color produced from a chromophore generated by enzymatic reaction between β -galactosidase and X-gal, etc.

[0184] Cancer cells for use in the method for cancer cell detection or in the method for cancer diagnosis may be derived from a biological sample taken from a subject. Such a biological sample taken from a subject is not limited in any way as long as it is a tissue suspected to contain cancer cells, and examples include blood, tumor tissue, lymphoid tissue and so on. Alternatively, cancer cells may be circulating tumor cells (CTCs) in blood, and explanations on CTCs are the same as described above.

[0185] Cancer cell detection and cancer diagnosis using the reagent of the present invention may be accomplished as follows, by way of example.

[0186] In cases where the biological sample taken from a subject is blood, the blood sample is treated by addition of an erythrocyte lysis reagent to remove erythrocytes and the remaining cell suspension is mixed in a test tube with the reagent of the present invention at a given ratio (0.01 to 1000 MOI (multiplicity of infection), preferably 0.1 to 100 MOI, more preferably 1 to 10 MOI). The test tube is allowed to stand or rotated for culture at room temperature or 37° C. for a given period of time (e.g., 4 to 96 hours, preferably 12 to 72 hours, more preferably 18 to 36 hours) to facilitate virus infection into cancer cells and virus growth. GFP fluorescence production in the cell fraction is quantitatively analyzed by flow cytometry. Alternatively, GFP-expressing cells are morphologically analyzed by being observed under a fluorescence microscope. This system allows highly sensitive detection of CTCs present in peripheral blood. This method can be used for detection of CTCs which are present in trace amounts in peripheral blood.

[0187] In cases where flow cytometry is used for CTC detection, CTCs may be detected by determining whether each cell is GFP-positive or GFP-negative, e.g., in accordance with the following criteria.

[0188] First, groups of cells in a sample which is not infected with any virus are analyzed to obtain a background fluorescence value. A threshold is set to the maximum fluorescence value. Subsequently, groups of cells in samples which have been infected with the virus of the present invention are analyzed and groups of cells in a sample showing a fluorescence value equal to or greater than the threshold are determined to be GFP-positive. In the case of using a blood sample taken from a subject, GFP-positive cells can be detected as CTCs. Further, these GFP-positive cells (CTCs) may be concentrated for phenotyping or genotyping.

[0189] In the present invention, examples of a subject include mammals such as humans, rabbits, guinea pigs, rats, mice, hamsters, cats, dogs, goats, pigs, sheep, cows, horses, monkeys and so on.

[0190] The amount of the reagent of the present invention to be used is selected as appropriate, depending on the state and amount of a biological sample to be used for detection and the type of detection method to be used, etc. For example, in the case of a blood sample, the reagent of the present invention can be used in an amount ranging from about 0.01 to 1000 MOI, preferably 0.1 to 100 MOI, and more preferably 1 to 10 MOI per 1 to 50 ml, preferably 3 to 25 ml, and more preferably 5 to 15 ml of the blood sample. MOI refers to the ratio between the amount of virus (infectious unit) and the number of cells when a given amount of cultured cells are infected with a given amount of virus particles, and is used as an index when viruses are infected into cells.

[0191] To infect the recombinant virus into cells, the following procedures may be used for this purpose. First, cells are seeded in a culture plate containing an appropriate culture medium and cultured at 37° C. in the presence of carbon dioxide gas. The culture medium is selected from DMEM, MEM, RPMI-1640 and others commonly used for animal cell culture, and may be supplemented with serum,

antibiotics, vitamins and so on, if necessary. The cultured cells are inoculated with a given amount of the virus, for example, at 0.1 to 10 MOI.

[0192] For confirmation of virus growth, the virus-infected cells are collected and treated to extract their DNA, followed by real-time PCR with primers targeting an appropriate gene possessed by the virus of the present invention, whereby virus growth can be quantitatively analyzed.

[0193] In cases where GFP gene is used as a reporter gene, labeled cells may be detected as follows: cells showing virus growth will emit a given fluorescence (e.g., a green fluorescence for GFP) upon irradiation with an excitation light, so that cancer cells can be visualized by the fluorescence. For example, when the virus-infected cells are observed under a fluorescence microscope, GFP fluorescence production can be seen in the cells. Moreover, to observe the virus-infected cells over time, GFP fluorescence production can be monitored over time with a CCD camera.

[0194] Moreover, the reagent of the present invention also allows real-time detection of cancer cells present in vivo. To label and detect cells in vivo in a real-time manner, the recombinant adenovirus of the present invention may be administered in vivo.

[0195] The reagent of the present invention may be applied directly to the affected area or may be introduced in vivo (into target cells or organs) in any known manner, e.g., by injection into vein, muscle, peritoneal cavity or subcutaneous tissue, inhalation from nasal cavity, oral cavity or lungs, oral administration, catheter-mediated intravascular administration and so on, as preferably exemplified by local injection into muscle, peritoneal cavity or elsewhere, injection into vein, etc.

[0196] When the reagent of the present invention is administered to a subject, the dose may be selected as appropriate, depending on the type of active ingredient, the route of administration, a target to be administered, the age, body weight, sex and/or symptoms of a patient, and other conditions. As a daily dose, the amount of the virus of the present invention serving as an active ingredient may usually be set to around 10^6 to 10^{11} PFU (plaque forming units), preferably around 10^9 to 10^{11} PFU, given once a day or in divided doses.

[0197] Real-time in vivo monitoring of fluorescence from cancer cells has the advantage of being used for in vivo diagnostic agents. This is useful for so-called navigation surgery and so on. Details on navigation surgery can be found in WO2006/036004.

[0198] Further, the reagent of the present invention is useful for detection of CTCs as a biomarker, and hence the reagent of the present invention can be used to determine prognosis.

[0199] For example, in cases where GFP is used as a labeling protein in the virus of the present invention, a biological sample taken from a cancer patient before being treated by any cancer therapy (e.g., chemotherapy, radiation therapy, surgical operation) and a biological sample taken at a time point after a certain period (e.g., 1 to 90 days) has passed from the treatment are each infected with the virus of the present invention. Next, GFP-positive cells contained in the sample taken before the treatment and GFP-positive cells contained in the sample taken at a certain time point after the treatment are compared for their number under the same conditions. As a result, if the number of GFP-positive cells after the treatment becomes smaller than the number of

GFP-positive cells before the treatment, a determination can be made that prognosis has been improved.

[0200] The present invention will be further described in more detail by way of the following illustrative examples, which are not intended to limit the scope of the invention.

Example 1

Preparation of Ad34 Fiber 142-3pT

[0201] (1) Preparation of pHMCMV5-miR-142-3pT pHMCMV5 (Mizuguchi H. et al., Human Gene Therapy, 10: 2013-2017, 1999) was treated with NotI/KpnI and the resulting fragment was ligated to a double-stranded oligo, which had been prepared by annealing the following synthetic oligo DNAs, to thereby prepare pHMCMV5-miR-142-3pT (pre).

miR-142-3pT-S1:

(SEQ ID NO: 43)
5' - GGCCTCCATAAAGTAGGAAACACTACACAGCTCCATAAAGTAGGAA
ACACTACATTAATTAAGCGGTAC - 3' ,

each underline represents a miR-142-3p target sequence)

miR-142-3pT-AS1:

(SEQ ID NO: 44)
5' - CGCTTAATTAATGTAGTGTTCCTACTTTATGGAGCTGTGTAGTGT
TCTACTTTATGGA - 3' .

each underline represents a miR-142-3p target sequence)

[0202] Then, pHMCMV5-miR-142-3pT(pre) was treated with PacI/KpnI and the resulting fragment was ligated to a double-stranded oligo, which had been prepared by annealing the following synthetic oligo DNAs, to thereby obtain pHMCMV5-miR-142-3pT having 4 repeats of a miR-142-3p target sequence.

miR-142-3pT-S2:

(SEQ ID NO: 45)
5' - TCCATAAAGTAGGAAACACTACAGGACTCCATAAAGTAGGAAACA
CTACAGTAC - 3' ,

each underline represents a miR-142-3p target sequence)

(SEQ ID NO: 46)
5' - TGTAGTGTTCCTACTTTATGGAGTCTGTAGTGTTCCTACTTTAT
GGAAT - 3' ,

miR-142-3pT-AS2:

each underline represents a miR-142-3p target sequence)

[0203] (2) Preparation of E1 Shuttle Plasmid pHM5-hAIB-miR-142-3pT

[0204] pSh-hAIB (WO2006/036004) was digested with I-CeuI/PmeI and the digested product was electrophoresed on an agarose gel. A band of approximately 4.5 kbp (hAIB cassette) was excised from the gel and treated with GENECLEAN II (Q-Biogene) to purify and collect a DNA fragment. The purified DNA fragment (hAIB cassette) was ligated to a fragment which had been obtained from pHMCMV5-miR-142-3pT by being digested with NheI, treated with Klenow Fragment and further digested with I-CeuI, thereby obtaining pHM5-hAIB-miR-142-3pT having

hTERT promoter, E1A gene, IRES (internal ribosomal entry site) sequence, E1B gene and a miR-142-3pT target sequence.

[0205] (3) Preparation of E3 Shuttle Plasmid pHM13CMV-EGFP-miR-142-3pT

[0206] pEGFP-NI (Clontech) was digested with *Apal* and *NotI*, and the resulting digested product was inserted into the *Apal/NotI* site of pHMCMV5 to obtain pHMCMVGFP-1. pHMCMVGFP-1 was digested with *PmeI/HindIII*, and the digested product was electrophoresed on an agarose gel. A band of approximately 750 bp (EGFP) was excised from the gel and treated with GENECLAN II to purify and collect a DNA fragment. The purified DNA fragment (EGFP) was ligated to a fragment which had been obtained from pBlue-scriptII KS+ by being digested with *HindII/HindII*, thereby preparing pBSKS-EGFP. pBSKS-EGFP was digested with *Apal/XbaI*, and the digested product was electrophoresed on an agarose gel. A band of approximately 750 bp (EGFP) was excised from the gel and treated with GENECLAN II to purify and collect a DNA fragment. The purified DNA fragment (EGFP) was ligated to a fragment which had been obtained from pHMCMV5-miR-142-3pT by being digested with *Apal/XbaI*, thereby obtaining pHMCMV5-EGFP-miR-142-3pT. pHMCMV5-EGFP-miR-142-3pT was digested with *BglIII*, and the digested product was electrophoresed on an agarose gel. A band of approximately 2 kbp (CMV-EGFP-miR-142-3pT) was excised from the gel and treated with GENECLAN II to purify and collect a DNA fragment. The purified DNA fragment (CMV-EGFP-miR-142-3pT) was ligated to a fragment which had been obtained from pHM13 (Mizuguchi et al., *Biotechniques*, 30: 1112-1116, 2001) by being digested with *BamHI* and treated with CIP (Alkaline Phosphatase, Calf Intest), thereby obtaining pHM13CMV-EGFP-miR-142-3pT.

[0207] (4) Preparation of pAdHM49-hAIB142-3pT-CG142-3pT

[0208] pAdHM49 (Mizuguchi et al, *J. Controlled Release* 110; 202-211, 2005) was treated with *I-CeuI/PI-SceI* and the resulting fragment was ligated to pHM5-hAIB-miR-142-3pT which had also been treated with *I-CeuI/PI-SceI*, thereby preparing pAdHM49-hAIB142-3pT in which hTERT promoter, E1A gene, IRES sequence, E1B gene and a miR-142-3pT target sequence were integrated into the E1-deficient region of the Ad vector. pAdHM49 is a recombinant adenovirus in which a region covering genes encoding the fiber knob and fiber shaft of the adenovirus type 5 fiber is replaced with a region covering genes encoding the fiber knob and fiber shaft of the adenovirus type 34 fiber, and hence pAdHM49 comprises the nucleotide sequence (SEQ ID NO: 49) of a gene encoding a region consisting of the fiber knob region and the fiber shaft region in the fiber protein of adenovirus type 34. The nucleotide sequence of a gene encoding the pAdHM49 fiber protein (i.e., the fiber knob region and fiber shaft region of the adenovirus type 34 fiber and the fiber tail region of the adenovirus type 5 fiber) is shown in SEQ ID NO: 50. In the nucleotide sequence shown in SEQ ID NO: 50, the nucleotide sequence of a gene encoding the fiber tail region of the adenovirus type 5 fiber is located at nucleotides 1 to 132, the nucleotide sequence of a gene encoding the fiber shaft region of the adenovirus type 34 fiber is located at nucleotides 133 to 402, and the nucleotide sequence of a gene encoding the fiber knob region of the adenovirus type 34 fiber is located at nucleotides 403 to 975. Namely, in the nucleotide sequence shown

in SEQ ID NO: 50, the nucleotide sequence of a region derived from the adenovirus type 5 fiber is located at nucleotides 1 to 132, while the nucleotide sequence of a region derived from the adenovirus type 34 fiber is located at nucleotides 133 to 975.

[0209] Then, pAdHM49-hAIB142-3pT was digested with *Csp45I* and the resulting fragment was ligated to a fragment which had been obtained from pHM13CMV-EGFP-miR-142-3pT by being digested with *Clal*, thereby obtaining pAdHM49-hAIB142-3pT-CG142-3pT in which hTERT promoter, E1A gene, IRES sequence, E1B gene and a miR-142-3pT target sequence were integrated into the E1-deficient region of the adenovirus vector and CMV promoter. EGFP and a miR-142-3pT target sequence were integrated into the E3-deficient region of the adenovirus vector, and which further comprised a gene encoding the fiber protein of adenovirus type 34.

[0210] (5) Preparation of Ad34 Fiber 142-3pT(E1,E3)

[0211] pAdHM49-hAIB142-3pT-CG142-3pT was linearized by being cleaved with a restriction enzyme *PacI* whose recognition site was present at each end of the adenovirus genome therein, and the linearized product was transfected into 293 cells seeded in a 60 mm culture dish by using Lipofectamine 2000 (Invitrogen). After about 2 weeks, a recombinant adenovirus Ad34 fiber 142-3pT(E1,E3) was obtained (FIG. 1).

[0212] The Ad34 fiber 142-3pT(E1, E3) corresponds to the Adenovirus type 5; TelomeScan F35 strain, which was deposited on Feb. 6, 2014, with the American Type Culture Collection (ATCC®) (10801 University Boulevard, Manassas, Va., 20110), and was provided with the deposit Accession No. PTA-120968. Please note that the depository ATCC® is an International Depository Authority (IDA) under the Budapest Treaty located in the USA.

Example 2

Activity Measurement of Ad34 Fiber 142-3pT(E1,E3)

(1) Cells

[0213] HeLa (derived from human uterine cancer cells) and LN319 (derived from human glioma cells) were used as CAR-positive cells, while LN308 (derived from human glioma cells), LN444 (derived from human glioma cells) and K562 (derived from human myelogenous leukemia cells) were used as CAR-negative cells. K562 cells are expressing miR-142-3p. DMEM (10% FCS, supplemented with antibiotics) was used for HeLa, LN319. LN308 and LN444 cells, while RPMI-1640 medium (10% FCS, supplemented with antibiotics) was used for K562 cells. These cells were cultured at 37° C. under saturated vapor pressure in the presence of 5% CO₂.

[0214] (2) Activity Measurement of Ad34 Fiber 142-3pT (E1,E3) by Flow Cytometry

[0215] Cells of each line were seeded in a 24-well plate at 5×10^4 cells/500 ul/well and treated with Ad34 fiber 142-3pT (E1,E3) at an MOI of 10. As a control, TelomeScan (i.e., a conditionally replicating adenovirus comprising hTERT promoter, E1A gene, IRES sequence and E1B gene integrated in this order into the E1-deficient site of adenovirus type 5 and comprising CMV promoter and GFP integrated in this order into the E3-deficient site of adenovirus type 5) was used. After culture for 24 hours, the cells were collected and

the number of GFP-positive cells was measured using a flow cytometer MACSQuant (Miltenyi Biotec).

[0216] The results obtained are shown in FIG. 2. In the specification and FIG. 2, “TelomeScan (Ad5 fiber)” represents TelomeScan, while “Ad34 fiber” represents a recombinant adenovirus which comprises hTERT promoter, E1A gene, IRES sequence and E1B gene integrated in this order into the E1-deficient site of the adenovirus genome and also comprises CMV promoter and GFP integrated in this order into the E3-deficient site of the adenovirus genome and which comprises a gene encoding a fiber protein derived from adenovirus type 34. Likewise, “Ad34 fiber 142-3pT (E1)” represents a recombinant adenovirus which further comprises a target sequence of miR-142-3p integrated into the E1-deficient region (downstream of the E1B gene) in the above Ad34 fiber, while “Ad34 fiber 142-3pT(E3)” represents a recombinant adenovirus which further comprises a target sequence of miR-142-3p integrated into the E3-deficient region (downstream of the GFP gene) in the above Ad34 fiber. Likewise, “Ad34 fiber 142-3pT(E1,E3)” represents a recombinant adenovirus which further comprises a target sequence of miR-142-3p integrated into each of the E1- and E3-deficient regions (downstream of the E1B gene and downstream of the GFP gene, respectively) in the above Ad34 fiber. Moreover, in FIG. 2 and the subsequent figures. “(containing GFP)” is intended to mean that the GFP gene is inserted into each viral genome.

[0217] As a result of activity measurement, when LNZ308, LN444 and K562, which are CAR-negative cells, were infected with TelomeScan (Ad5 fiber), no GFP-positive cell was detected (FIG. 2, panels k, p and u). In contrast, when these cells were infected with Ad34 fiber, GFP-positive cells were detected (85.5% positive in LNZ308, 58.4% positive in LN444, and 63.7% positive in K562) (panels l, q and v).

[0218] This result indicated that the recombinant adenovirus of the present invention having a gene encoding the fiber protein of adenovirus type 34 allowed significant detection of CAR-negative cells.

[0219] Further, in the case of K562 cells which are CAR-negative and are expressing miR-142-3p, GFP-positive cells were 63.7% upon infection with Ad34 fiber (panel v), whereas GFP-positive cells were 12.2% upon infection with Ad34 fiber 142-3pT(E1) and 34.8% upon infection with Ad34 fiber 142-3pT(E3), and no GFP-positive cell was detected upon infection with Ad34 fiber 142-3pT(E1,E3) (panels w, x and y). Namely, the detection rate of K562 cells was significantly reduced when using an adenovirus comprising a target sequence of miR-142-3p integrated into either the E1- or E3-deficient region of the adenovirus genome, and K562 cells were no longer detected when using an adenovirus comprising a target sequence of miR-142-3p integrated into each of the E1- and E3-deficient regions.

[0220] This result indicated that the recombinant virus of the present invention comprising a target sequence of miR-142-3p did not detect highly miR-142-3p-expressing cells, such as normal blood cells.

[0221] Moreover, the above result was further investigated. When the case of infecting K562 cells with Ad34 fiber (panel v) was compared with the case of infecting the same cells with Ad34 fiber 142-3pT(E1) (panel w), the rate (%) of GFP-positive cells in the case of the infection with Ad34 fiber 142-3pT(E1) was decreased by 19% (12.2%/63.7%). This result shows an effect based on miR-142-3pT (target

sequence of miR-142-3p) integrated into the E1-deficient region. On the other hand, when the case of infecting K562 cells with Ad34 fiber (panel v) was compared with the case of infecting the same cells with Ad34 fiber 142-3pT(E3) (panel x), the rate (%) of GFP-positive cells in the case of the infection with Ad34 fiber 142-3pT (E3) was decreased by 54% (34.8%/63.7%). This result shows an effect based on miR-142-3pT integrated into the E3-deficient region. These results indicate that, in the case of using an adenovirus wherein miR-142-3pT has been integrated into both E1- and E3-deficient regions in this experiment, if the result obtained shows an additive effect based on two sequences of miR-142-3pT integrated into both E1- and E3-deficient regions, the rate (%) of GFP-positive cells would be expected to be about 10% (19% \times 54%).

[0222] However, contrary to the above expectation, in this Example, in the case (panel y) of infecting K562 cells with Ad34 fiber 142-3pT (E1, E3) wherein miR-142-3pT had been integrated into both E1- and E3-deficient regions, almost no GFP-positive cells were detected.

[0223] This fact indicates that, in the effects exerted by the present invention, the effect based on two sequences of miR-142-3pT integrated was not an additive effect but a synergistic effect.

[0224] As described herein above, CTC detection techniques are required to detect several to several tens of cells among a billion of blood cells, and therefore desired to have a high accuracy in the detection of CTC. In addition, in order to increase the CTC detection accuracy, the CTC detection techniques are desired to bring the number of cells detected as false positive boundlessly close to zero. In general, 1 μ L of blood is known to contain thousands of white blood cells, and these white blood cells are desired not to be detected as false positive.

[0225] In the CTC detection techniques, even the presence of several false positive cells has a great influence on the determination of whether a subject has CTC.

[0226] As described above, in this Example, contrary to the expectation, in the case of infecting K562 cells with Ad34 fiber 142-3pT (E1, E3) wherein miR-142-3pT had been integrated into both E1- and E3-deficient regions, almost no GFP-positive cells were detected. This result shows unexpectedly remarkable effect according to the present invention.

Example 3

Detection of Cancer Cells in Blood Samples Using Ad34 Fiber 142-3pT(E1,E3)

[0227] 5×10^4 H1299 cells (CAR-positive) were suspended in 5 mL blood and erythrocytes were lysed to collect PBMCs. To these PBMCs, a virus was added in an amount of 1×10^9 , 1×10^{10} or 1×10^{11} VPs (virus particles) and infected at 37° C. for 24 hours while rotating with a rotator. The cells were collected and immunostained with anti-CD45 antibody, and GFP-positive cells were observed under a fluorescence microscope. CD45 is known to be a surface antigen of blood cell lineage cells except for erythrocytes and platelets. “GFP Positive Cancer cells (%)” found in the vertical axis of FIGS. 3 and 4 represents the “number of GFP-positive and CD45-negative cells (%) among GFP-positive cells.”

[0228] As a result, many false positive cells (GFP-positive and CD45-positive cells) were observed upon infection with

TelomeScan (Ad5 fiber), whereas false positive cells were very few upon infection with Ad34 fiber 142-3pT(E1,E3), so that cancer cells were able to be specifically detected.

[0229] Moreover, as a result of quantitative analysis on the detection specificity of H1299 cells, many false positive cells were detected in the case of TelomeScan (Ad5 fiber) upon virus infection at 1×10^9 VPs, whereas the detection specificity was 90% or higher and some samples showed 100% detection specificity in the case of Ad34 fiber 142-3pT(E1,E3) even when the amount of virus infection was increased (FIG. 3). Likewise, quantitative analysis was also performed on A549 cells (CAR-positive cells) in the same manner, indicating that the detection specificity was 100% upon virus infection at 1×10^9 VPs (FIG. 4). These results indicated that the recombinant virus of the present invention allowed specific detection of cancer cells contained in the PBMC fraction.

[0230] In view of the foregoing, the detection reagent and diagnostic reagent of the present invention were demonstrated to allow detection of highly malignant CAR-negative cancer cells and, on the other hand, to ensure no false positive detection of highly miR-142-3p-expressing normal blood cells (e.g., leukocytes), etc.; and hence they were shown to be very effective for detection of circulating tumor cells (CTCs) in blood.

Example 4

Activity Measurement of Ad34 Fiber 142-3pT(E1,E3) in Various Human Cancer Cell Lines

(1) Cells

[0231] The cancer cells used in this example were human non-small cell lung cancer-derived H1299 cells, human lung cancer-derived A549 cells, human breast cancer-derived MCF7 cells, human breast cancer-derived MDA-MB-231 cells, human bladder cancer-derived KK47 cells, human gastric cancer-derived MKN45 cells, human colorectal cancer-derived SW620, human liver cancer-derived Huh7 cells, human pancreatic cancer-derived Panel cells, human glioma-derived LN319 cells, human bladder cancer-derived T24 cells, human glioma-derived LN308 cells, and human glioma-derived LN444 cells.

(2) Activity Measurement of Ad34 Fiber 142-3pT(E1,E3) by Flow Cytometry

[0232] 5×10^4 cancer cells of each line were suspended in 500 μ l medium, to which 100 μ l of a conditionally replicating Ad suspension prepared at 5×10^5 or 5×10^6 pfu/ml was then added. The resulting mixture of the cells and the conditionally replicating Ad was seeded in a 24-well plate and cultured at 37° C. for 24 hours. The cells were collected and centrifuged at 1500 rpm for 5 minutes. After removal of the medium, the cells were suspended in 300 μ l of 2% FCS-containing PBS and measured for GFP-positive rate using a flow cytometer (MACS Quant Analyzer, Miltenyi Biotec). The data obtained were analyzed by FCS multi-color data analysis software (Flowjo).

[0233] As a result, Ad34 fiber 142-3pT(E1,E3) was found to efficiently infect almost all cancer cells, and 60% or more of the cancer cells were GFP-positive. Particularly in the case of CAR-negative cells (T24, LN308, LN444), their GFP-positive rate was significantly improved when compared to conventionally used TelomeScan (FIG. 5).

[0234] This result indicated that the recombinant virus of the present invention allowed efficient detection of not only CAR-positive cells but also CAR-negative cells.

Example 5

Detection of Cancer Cells Having Undergone Epithelial-Mesenchymal Transition (EMT)

[0235] Human pancreatic cancer PancI cells were cultured for 6 days in the presence of 10 ng/mL recombinant TGF- β 1 to thereby induce epithelial-mesenchymal transition (EMT). After induction of EMT, relative expression of mRNAs encoding E-cadherin, EpCAM, hTERT, N-cadherin, Slug and Snail was measured by real-time RT-PCR. In addition, CAR and CD46 expression in the Panc I cells was analyzed by flow cytometry. The virus of the present invention was infected into the cells in the same manner as shown in Example 4.

[0236] As a result, upon culture in a TGF-3-containing medium, the expression of EMT marker genes Slug, Snail and N-cadherin were increased, while the expression of epithelial markers E-cadherin and EpCAM was reduced, thus indicating that EMT has been induced (FIG. 6A). Moreover, upon EMT induction, CAR expression was reduced whereas CD46 expression was not reduced at all (FIG. 6B). Further, when conventionally used TelomeScan was used for PancI cells having undergone EMT only about 35% of these cells were GFP-positive, whereas almost 90% or more of the cells were GFP-positive in the case of Ad34 fiber 142-3pT(E1,E3) (FIG. 6C).

[0237] These results indicated that the recombinant virus of the present invention allowed highly sensitive detection of cancer cells having undergone epithelial-mesenchymal transition (EMT).

Example 6

Detection of Cancer Stem Cells

[0238] MCF7 cells and MCF7-ADR cells (cancer cells resistant to the anticancer agent adriamycin) were each seeded in a 96-well plate at 1×10^3 cells/well, and on the following day, adriamycin was added thereto at 0.2, 1, 5, 25 or 125 μ g/mL. After 24 hours from the addition of adriamycin, an alamarBlue® cell viability reagent was used to measure cell viability (value: mean \pm S.D. (n=6)).

[0239] MCF7 cells and MCF7-ADR cells were also analyzed by flow cytometry for expression of CAR, CD46, P-glycoprotein (MDR), CD24 and CD44. 5×10^5 MCF7-ADR cells were suspended in 100 μ l of 2% FCS-containing PBS, and FITC-labeled mouse anti-human CD24 antibody and PE-labeled mouse anti-human CD44 antibody were each added thereto in a volume of 1 μ l, followed by reaction for 1 hour on ice under light-shielded conditions. After washing with 4 ml of 2% FCS-containing PBS, the suspension was centrifuged at 1500 rpm for 5 minutes to remove the supernatant by aspiration. The cells were suspended again in 100 μ l of 2% FCS-containing PBS and subjected to a cell sorter (FACS Aria II cell sorter; BD Biosciences) to sort a CD24-negative and CD44-positive cell fraction. The data obtained were analyzed by FCS multi-color data analysis software (Flowjo). In human breast cancer cells, a fraction having the characteristics of CD24-negative and CD44-positive cells is known to be cancer stem cells (Al-Hajj M., et al., Proc Natl Acad Sci USA, 100; 3983-3988,

(2003)). The virus of the present invention was infected into the cells in the same manner as shown in Example 4.

[0240] As a result, MCF7-ADR cells showed significantly high viability even in the presence of adriamycin when compared to MCF7 cells and hence were found to have drug resistance ability (FIG. 7A). MCF7-ADR cells were also found to highly express CAR and CD46 as in the case of MCF7 cells. Moreover, MCF7-ADR cells were also found to highly express MDR, which is a membrane protein responsible for drug elimination ability (FIG. 7B). Further, when Ad34 fiber 142-3pT(E1,E3) was infected into CD24-negative and CD44-positive cells among MCF-ADR cells, 80% or more of the cells were GFP-positive. In contrast, about 70% of the cells were GFP-positive in the case of conventionally used TelomeScan (FIG. 7C).

[0241] These results indicated that the recombinant virus of the present invention allowed detection of drug-resistant cancer cells. Moreover, it was also indicated that the recombinant virus of the present invention allowed detection of cancer stem cells.

Example 7

Detection of Cancer Cells in Blood Samples Using Ad34 Fiber 142-3pT(E1,E3)

[0242] H1299 cells or T24 cells were infected with a lentivirus vector expressing a red fluorescent protein (monomeric red fluorescent protein; RFP) at an MOI of 100 and cultured. To obtain cell clones, the cells were then seeded in a 96-well plate at 0.1 cells/well and cultured until colonies were formed. RFP-expressing cells were selected under a fluorescence microscope and subjected to extended culture, followed by flow cytometry to measure the intensity of RFP expression. Then, cells showing high intensity of RFP expression were identified as RFP-expressing cells.

[0243] Human peripheral blood mononuclear cells (hPBMCs) obtained from 1.0 mL of human peripheral blood were suspended in 800 μ L of RPMI-1640 medium (10% FCS, supplemented with antibiotics). To the hPBMC suspension, cancer cells prepared at 1.0×10^5 or 5.0×10^{-5} cells/mL were added in a volume of 100 μ L (in FIG. 8, "spiked cancer cells" represents the number of cancer cells added to the hPBMC suspension). Further, a conditionally replicating Ad suspension prepared at 2×10^8 pfu/mL was added in a volume of 100 μ L to give a total volume of 1 mL, followed by culture at 37° C. for 24 hours while slowly rotating with a rotator.

[0244] The cell suspension cultured for 24 hours after virus infection was centrifuged at 300 \times g for 5 minutes to remove the supernatant. A cell fixative was added in a volume of 200 μ L and reacted at 4° C. under light-shielded conditions for 15 minutes. After addition of 1 mL PBS, the suspension was centrifuged at 300 \times g for 5 minutes to remove the supernatant. The cells were suspended in 2% FCS-containing PBS and measured for GFP-positive rate using a flow cytometer (MACS Quant Analyzer: Miltenyi Biotec). The data obtained were analyzed by FCS multi-color data analysis software (Flowjo).

[0245] In this study, cancer cells labeled with RFP (red fluorescent protein) were mixed into hPBMCs to examine whether the cancer cells in hPBMCs were able to be detected. As a result, in the case of CAR-positive cancer cells (H1299), TelomeScan (Ad5 fiber) and Ad34 fiber 142-3pT(E1,E3) were both able to detect 80% or more of the

cancer cells. On the other hand, in the case of CAR-negative cancer cells (T24), TelomeScan (Ad5 fiber) achieved very low detection efficiency (about 10% of the cells were detected as being GFP-positive), whereas Ad34 fiber 142-3pT(E1,E3) was able to detect 80% or more of the cancer cells (FIG. 8).

[0246] This result indicated that the recombinant adenovirus of the present invention allowed efficient detection of not only CAR-positive cancer cells but also CAR-negative cancer cells.

Example 8

Full Sequencing of Ad34 Fiber 142-3pT (E1,E3) Genome

[0247] In the present example, a full sequencing of Ad34 Fiber 142-3pT (E1,E3) genome was acquired by constructing a shotgun library using the Nextera XT DNA Library Preparation Kit and the Nextera XT Index kit (illumina, Inc.), followed by sequencing the clones using MiSeq System (illumina, Inc.) to provide 2,412,121 base pairs sequencing data.

Method

(1) Test Samples

[0248] In this Example, adenoviral genome DNA was prepared from the test article using a Qiagen viral DNA extraction method and final concentration was 8.1 ng/ μ L and the total amount was 10 μ L.

(2) Shotgun Library Construction and DNA Sequencing

(2-1) Tagment Genomic DNA and Amplify Libraries

[0249] One ng of purified Ad34 Fiber 142-3pT (E1,E3) was used for generating the library by applying the Nextera XT DNA Library Preparation Kit according to the manufacturer's instructions with the exception of using the primers of the Nextera XT Index kit. In brief, DNA sample was fragmented and tagged with adapter sequences by Nextera XT transposase and tagged with Index sequence.

(2-2) Clean Up Libraries

[0250] The resulting fragments were purified by using Agencourt AMPure XP beads (Beckman Coulter), separated by agarose gel electrophoresis and extracted from the gel in the range of 350-600 bases. The fragments were size-selected by a High Sensitivity DNA Chip on the Bioanalyzer 2100 (Agilent Technologies, Inc.) and quantified by Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Inc.) before loading on the sequencing chip. The final concentration of the resulted library was 1.3 ng/ μ L and total amount was 10 μ L.

(2-3) Sequencing and Assembly

[0251] The resulted clean up library was denatured by 0.2N NaOH for 5 min. After denaturation, it was diluted into 12 pM by hybridization buffer before loading on the MiSeq chip. After 300-bp paired-end read sequencing on the MiSeq platform (Illumina), the data were base called and reads with the same barcode were collected and assigned to a sample on the sequencing instrument. The uncallable ends of the

MiSeq reads (B in input file) were automatically trimmed by the Sickle: sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) software (Joshi N A, Fass J N (2011)) and Fastx tool kit (Ver 0.0.13). De novo assembly was performed as follows: the failed reads (Y in header information for the quality score) were removed by using Bayes hammer implemented, the resulted sequences were assembled at parameter k=21, 33, 55, 77, 99, 127 and repaired mismatches by SPAdes Genome Assembler (Ver 3.6.0) (J. Comp. Biol. 19(5) (2012): 455-77.). The assemble data was aligned to reference sequence by Bowtie2 (Ver 2.2.4).

Result

[0252] The assembly resulted in a contig length of 35,324 bp (SEQ ID NO: 51), in the range of 20,000-2000,000 reads. The sequences of genes or regulatory elements were aligned for the resulting full sequence of Ad34 Fiber 142-3pT (E1,E3).

[0253] The location thereof at SEQ ID NO: 51 as follows:

[0254] The human telomerase reverse transcriptase promoter starts at position 554 and ends at position 1008, therefore consisting of 455 nucleotides.

The E1A gene starts at position 1041 and ends at position 1939, therefore consisting of 899 nucleotides.

[0255] The E1B gene starts at position 2616 and ends at position 4438, therefore consisting of 1823 nucleotides.

[0256] The IRES sequence starts at position 1961 and ends at position 2581, therefore consisting of 621 nucleotides.

[0257] The first microRNA sequence (SEQ ID NO: 52) starts at position 4508 and ends at position 4612, therefore consisting of 105 nucleotides.

[0258] The second microRNA sequence (SEQ ID NO: 52) starts at position 30715 and ends at position 30819, therefore consisting of 105 nucleotides.

[0259] The cytomegalovirus (CMV) promoter starts at position 29,381 and ends at position 29,966, therefore consisting of 586 nucleotides.

[0260] The GFP gene starts at position 29,988 and ends at position 30,704, therefore consisting of 717 nucleotides.

[0261] The replication cassette (SEQ ID NO: 53) starts at position 554 and ends at position 4612, therefore consisting of 4059 nucleotides.

[0262] The labeling cassette (SEQ ID NO: 54) starts at position 29381 and ends at position 30819, therefore consisting of 1439 nucleotides.

[0263] The gene encoding a CD46-binding fiber protein (SEQ ID NO: 50) starts at position 31192 and ends at position 32166, therefore consisting of 975 nucleotides.

[0264] The sequence comprising four units of a target sequence of miR-142-3p was determined as follow:

(SEQ ID NO: 52)

5' - tccataaagttaggaacactacacagctccataaagttaggaacact
acataatccataaagttaggaacactacaggactccataaagttagga
acactaca-3'

each underline represents a single unit of a target sequence of miR-142-3p).

Example 9

Detection of Cancer Cells in Blood Samples Using Ad34 Fiber 142-3pT(E1,E3)

(1) Cells

[0265] H661 (derived from human lung cancer cells) was used as a CAR-positive cell without expression of miR-142-3p.

(2) Detection of Cancer Cells Infected with the Ad34 Fiber 142-3pT(E1,E3) Sequenced at Example 8 by Fluorescence Microscope Observation

[0266] One hundred of H661 cells were suspended in 7.5 mL blood and erythrocytes were lysed to collect PBMCs. To these PBMCs, a virus was added in an amount of 1×10^9 VPs (virus particles) for Ad34 fiber 142-3pT(E1,E3) or 3×10^6 for TelomeScan (Ad5 fiber) and infected at 37° C. for 24 hours while rotating with a rotator. The cells were collected and immunostained with anti-CD45 antibody, and GFP-positive cells were observed under a fluorescence microscope. CD45 is known to be a surface antigen of blood cell lineage cells except for erythrocytes and platelets. "GFP Positive Cancer cells (%)" found in the vertical axis of FIG. 9 represents the "number of GFP-positive and CD45-negative cells (%) among GFP-positive cells."

[0267] As a result, many false positive cells (GFP-positive and CD45-positive cells) were observed upon infection with TelomeScan (Ad5 fiber), whereas false positive cells were not observed upon infection with Ad34 fiber 142-3pT(E1, E3), so that cancer cells were able to be specifically detected.

[0268] Moreover, as a result of quantitative analysis on the detection specificity of H661 cells, many false positive cells were detected in the case of TelomeScan (Ad5 fiber) upon virus infection at 3×10^6 VPs, whereas the detection specificity was showed 100% detection specificity in the case of Ad34 fiber 142-3pT(E1,E3) even when the amount of virus infection was increased (1×10^9 VPs) (FIG. 9). These results indicated that the recombinant virus of the present invention allowed specific detection of cancer cells contained in the PBMC fraction.

Example 10

Activity Measurement of Ad34 Fiber 142-3pT(E1,E3)

(1) Cells

[0269] HeLa (derived from human uterine cancer cells) was used as a CAR-positive cell without expression of miR-142-3p, while K562 (derived from human myelogenous leukemia cells) was used as a CAR-negative cell expressing miR-142-3p. CD293 medium without phenol containing DMEM (10% FCS, supplemented with antibiotics) was used for HeLa cell, while RPMI-1640 medium (10% FCS, supplemented with antibiotics) was used for K562 cell. These cells were cultured at 37° C. under saturated vapor pressure in the presence of 5% CO₂.

(2) Activity Measurement of Ad34 Fiber 142-3pT(E1,E3) Sequenced at Example 8 by Multi-Plate Reader

[0270] Cells of each line were seeded in a 12-well plate at 8×10^5 cells/well and treated with Ad34 fiber 142-3pT(E1, E3) at an MOI of 400. As a control, Ad34 fiber was used at an MOI of 150, which can produce the same intensity of GFP fluorescence using Ad34 fiber 142-3pT(E1,E3) at an

MOI of 400 in HeLa cells. After culture for 24 hours, the cells were collected and lysed by GLO lysis buffer (Promega Corporation).

[0271] The fluorescence intensity of the lysate was measured using a multi-plate reader, PowerScan HT (DS Pharma Biomedical Co., Ltd.).

[0272] The results obtained are shown in FIG. 10. In FIG. 10, the fluorescence intensity of each cell infected with Ad34 fiber 142-3pT(E1,E3) was shown as a relative value against those of the cell infected with Ad34 fiber, assuming the latter was 100%.

[0273] As a result of activity measurement, when K1562 (miR-142-3p-negative cells) were infected with Ad34 fiber 142-3pT(E1,E3), the detected intensity of GFP fluorescence was very little (0.45%) as compared to those of Ad34 fiber (FIG. 10).

[0274] This result indicated that the recombinant virus of the present invention comprising a target sequence of miR-142-3p did not detect highly miR-142-3p-expressing cells, such as normal blood cells.

INDUSTRIAL APPLICABILITY

[0275] Reagents comprising the recombinant adenovirus of the present invention enable simple and highly sensitive detection of CAR-negative cancer cells without detection of normal blood cells (e.g., leukocytes).

SEQUENCE LISTING FREE TEXT

[0276] SEQ ID NO: 4: synthetic DNA

[0277] SEQ ID NOS: 5 to 26: synthetic RNA

[0278] SEQ ID NOS: 27 to 28: synthetic DNA

[0279] SEQ ID NOS: 43 to 46: synthetic DNA

[0280] SEQ ID NOS: 50 to 55: synthetic DNA

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 55

<210> SEQ ID NO 1

<211> LENGTH: 455

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

```
tggccctcc ctcgggttac cccacagcct aggcggattc gacctctctc cgctggggcc      60
ctcgtggcg tccttgacc ctgggagcgc gagcggcgcg cgggcgggga agcgcggccc      120
agacccccgg gtccgcccgg agcagctgcg ctgtcggggc caggcccggc tcccagtgga      180
ttcgcgggca cagacgccc ggaccgcgct cccacagtgg cggagggact ggggacccgg      240
gcacccgtcc tgcccctca ccttccagct ccgctctctc cgcgcggacc ccgccccttc      300
ccgacccctc ccgggtcccc ggcccagccc cctccggggc ctcccagccc ctccccttcc      360
tttccggggc cccgcccctc cctcggggcg cgagtttcag gcagcgtgc gtctgtgtgc      420
gcacgtggga agccctggcc ccggccaccc ccgcg                                     455
```

<210> SEQ ID NO 2

<211> LENGTH: 899

<212> TYPE: DNA

<213> ORGANISM: Adenovirus

<400> SEQUENCE: 2

```
acaccgggac tgaaaatgag acatattatc tgcccaggag gtgttattac cgaagaaatg      60
gccgcagatc ttttggacca gctgatcgaa gaggtactgg ctgataatct tccacctcct      120
agccattttg aaccacctac ccttcacgaa ctgtatgatt tagacgtgac ggcccccgaa      180
gatcccaacg aggaggcggg ttccgagatt tttcccgact ctgtaatggt ggcgggtgcag      240
gaagggattg acttactcac ttttccgccc gcgcccgggt ctccggagcc gcctcacctt      300
tcccggcagc ccgagcagcc ggagcagaga gccttgggtc cggtttctat gccaaacctt      360
gtaccggagg tgatgatct tacctgccac gaggtggct ttcacccag tgacgacgag      420
gatgaagagg gtgaggagtt tgtgttagat tatgtggagc accccgggca cggttgcagg      480
tcttgtcatt atcaccggag gaatacgggg gaccagata ttatgtgttc gctttgctat      540
atgaggacct gtggcatggt tgtctacagt cctgtgtctg aacctgagcc tgagcccag      600
```

-continued

ccagaaccgg agcctgcaag acctaccgc cgtcctaaaa tggcgctgc taccctgaga	660
cgcccgacat cacctgtgtc tagagaatgc aatagtagta cggatagctg tgactccggt	720
ccttctaaca cacctcctga gatacaccgc gtggctccgc tgtgccccat taaaccagtt	780
gccgtgagag ttggtgggcg tcgccaggct gtggaatgta tcgaggactt gcttaacgag	840
cctgggcaac ctttgactt gagctgtaaa cgccccaggc cataagggtg aaacctgtg	899

<210> SEQ ID NO 3
 <211> LENGTH: 1823
 <212> TYPE: DNA
 <213> ORGANISM: Adenovirus

<400> SEQUENCE: 3

ctgacctcat ggaggcttgg gagtgtttgg aagatttttc tgctgtgctg aacttgctgg	60
aacagagctc taacagtacc tcttggtttt ggaggtttct gtggggctca tcccaggcaa	120
agttagtctg cagaattaag gaggattaca agtgggaatt tgaagagctt ttgaaatcct	180
gtggtgagct gtttgattct ttgaatctgg gtcaccaggc gcttttcaa gagaaggta	240
tcaagacttt ggatttttcc acaccggggc gcgctgcggc tgctgttct ttttgagtt	300
ttataaagga taaatggagc gaagaaacc atctgagcgg ggggtacctg ctggattttc	360
tggccatgca tctgtggaga gcggtgtga gacacaagaa tcgcctgcta ctgtgtctt	420
ccgtccgccc ggcgataata ccgacggagg agcagcagca gcagcaggag gaagccaggc	480
ggcgccggca ggagcagagc ccatggaacc cgagagccgg cctggacct cggaatgaa	540
tgtgtacag gtggctgaac tgtatccaga actgagacgc attttgaaa ttacagagga	600
tgggcagggg ctaaaggggg taaagagga gcgggggct tgtgaggcta cagaggaggc	660
taggaatcta gcttttagct taatgaccag acaccgtcct gagtgatta cttttcaaca	720
gatcaaggat aattgcgcta atgagcttga tctgctggcg cagaagtatt ccatagagca	780
gctgaccact tactggctgc agccagggga tgattttgag gagctatta gggatatgc	840
aaaggtggca cttagccag attgcaagta caagatcagc aaacttgtaa atacagga	900
ttgttgctac atttctggga acggggccga ggtggagata gatacggagg atagggtggc	960
ctttagatgt agcatgataa atatgtggcc ggggtgctt ggcatggacg ggggtgttat	1020
tatgaatgta aggttactg gcccattt tagcggtagc gttttcctgg ccaataccaa	1080
ccttatccta caggtgtaa gcttctatgg gtttaacaat acctgtgtgg aagcctggac	1140
cgatgtaagg gttcgggct gtgcctttta ctgctgctgg aaggggtgg tgtgtcgccc	1200
caaaagcagg gcttcaatta agaaatgcct ctttgaaagg tgtacctgg gtatcctgctc	1260
tgagggtaac tccagggtgc gccacaatgt ggcctccgac tgtggttct tcatgctagt	1320
gaaaagcgtg gctgtgatta agcataacat ggtatgtggc aactgcgagg acagggcctc	1380
tcagatgctg acctgctcgg acggcaactg tcacctgctg aagaccattc acgtagccag	1440
ccactctgc aaggcctggc cagtgtttga gcataacata ctgaccgct gttccttga	1500
tttgggtaac agggggggg tgttcctacc ttaccaatgc aatttgagtc actaagat	1560
attgcttgag cccgagagca tgtccaaggt gaacctgaac ggggtgtttg acatgaccat	1620
gaagatctgg aaggtgctga ggtacgatga gaccgcacc aggtgcagac cctgcgagtg	1680
tggcggtaaa catattagga accagcctgt gatgctggat gtgaccgagg agctgaggcc	1740

-continued

cgatcacttg gtgctggcct gcaccocgc tgagttggc tctagcgatg aagatacaga 1800
 ttgaggtact gaaatgtgtg ggc 1823

<210> SEQ ID NO 4
 <211> LENGTH: 605
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic DNA

<400> SEQUENCE: 4

tgcatctagg gggccaatt cggccctct ccctccccc ccctaactg tactggcga 60
 agccgcttg aataaggccg gtgtgctgtt gtctatatgt gattttccac catattgccg 120
 tcttttgga atgtgagggc ccggaacct ggcctgtct tcttgacgag cattcctagg 180
 ggtctttccc ctctcgcaa aggaatgcaa ggtctgtga atgctgtgaa ggaagcagtt 240
 cctctggaag cttcttgaag acaaacacg tctgtagcga ccctttgcag gcagcggaac 300
 cccccactg gcgacaggtg cctctgccc caaaagccac gtgtataaga tacacctgca 360
 aaggcggcac aacccagtg ccacgttgtg agttggatag ttgtgaaag agtcaaatgg 420
 ctctctcaa gcgtattcaa caaggggctg aaggatgcc agaaggtacc ccattgtatg 480
 ggatctgatc tggggcctcg gtgcacatgc tttacatgtg tttagtcgag gttaaaaaaa 540
 cgtctaggcc ccccgaaaca cggggacgtg gtttccttt gaaaaacacg atgataagct 600
 tgcca 605

<210> SEQ ID NO 5
 <211> LENGTH: 23
 <212> TYPE: RNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 5

uguaguguuu ccuacuuuau gga 23

<210> SEQ ID NO 6
 <211> LENGTH: 21
 <212> TYPE: RNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 6

cauaaaguag aaagcacuac u 21

<210> SEQ ID NO 7
 <211> LENGTH: 22
 <212> TYPE: RNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 7

uagcagcaca uaaugguuug ug 22

<210> SEQ ID NO 8
 <211> LENGTH: 22
 <212> TYPE: RNA

-continued

<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 8

caggccauau ugugcugccu ca 22

<210> SEQ ID NO 9
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 9

uagcagcacg uaaauauugg cg 22

<210> SEQ ID NO 10
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 10

ccaguauuaa cugugcugcu ga 22

<210> SEQ ID NO 11
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 11

uagcuuauca gacugauguu ga 22

<210> SEQ ID NO 12
<211> LENGTH: 21
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 12

caacaccagu cgaugggcug u 21

<210> SEQ ID NO 13
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 13

ucguaccgug aguaauaaug cg 22

<210> SEQ ID NO 14
<211> LENGTH: 21
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

-continued

<400> SEQUENCE: 14
cauuuuuacu uuugguacgc g 21

<210> SEQ ID NO 15
<211> LENGTH: 23
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 15
aacauucaac gcugucggug agu 23

<210> SEQ ID NO 16
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 16
ugucaguug ucaaauaccc ca 22

<210> SEQ ID NO 17
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 17
cguguauuug acaagcugag uu 22

<210> SEQ ID NO 18
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 18
gaggguuugg uggaggcucu cc 22

<210> SEQ ID NO 19
<211> LENGTH: 21
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 19
agggccccc cucaaucug u 21

<210> SEQ ID NO 20
<211> LENGTH: 24
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 20
ucccugagac ccuuuaaccu guga 24

-continued

<210> SEQ ID NO 21
<211> LENGTH: 21
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 21
ugagaugaag cacuguagcu c 21

<210> SEQ ID NO 22
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 22
ggugcagugc ugcaucucug gu 22

<210> SEQ ID NO 23
<211> LENGTH: 23
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 23
guccaguuuu cccaggauc ccu 23

<210> SEQ ID NO 24
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 24
ggauccugg aaauacuguu cu 22

<210> SEQ ID NO 25
<211> LENGTH: 23
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 25
cccaguguuc agacuaccug uuc 23

<210> SEQ ID NO 26
<211> LENGTH: 22
<212> TYPE: RNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic RNA

<400> SEQUENCE: 26
ugagguagua gguuguauag uu 22

<210> SEQ ID NO 27
<211> LENGTH: 69
<212> TYPE: DNA

-continued

<213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic DNA

<400> SEQUENCE: 27

gcggcctcca taaagtagga aactacac agctccataa agtaggaac actacattat 60
 aagcgtac 69

<210> SEQ ID NO 28
 <211> LENGTH: 113
 <212> TYPE: DNA
 <213> ORGANISM: Artificial
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic DNA

<400> SEQUENCE: 28

ggcctccata aagtaggaaa cactacacag ctccataaag taggaaacac tacattaatt 60
 ccataaagta ggaaacacta caccactcca taaagtagga aactacacag tac 113

<210> SEQ ID NO 29
 <211> LENGTH: 1307
 <212> TYPE: DNA
 <213> ORGANISM: Homo sapiens

<400> SEQUENCE: 29

accgtccagg gagcaggtag ctgctgggct cgggggacac tttcgttcg ggctgggagc 60
 gtgctttcca cgacggtag acgcttccct ggattggcag ccagactgcc ttccgggtca 120
 ctgccatgga ggagcccgag tcagatccta gcgtcgagcc ccctctgagt caggaaacat 180
 tttcagacct atggaaacta ctctctgaaa acaacgttct gtcccccttg ccgtcccaag 240
 caatggatga tttgatgctg tccccggacg atattgaaca atggttcaact gaagaccag 300
 gtccagatga agctcccaga atgccagagg ctgctccccg cgtggcccct gcaccagcga 360
 ctctacacc ggcgcccct gcaccagccc cctctggccc cctgtcatct tctgtccctt 420
 cccagaaaac ctaccagggc agctacggtt tccgtctggg cttcttgcatt tctgggacag 480
 ccaagtctgt gacttgacg tactcccctg ccctcaaca gatgttttgc caactggcca 540
 agacctgccc tgtgcagctg tgggttgatt ccacacccc gcccggcacc cgcgtccgctg 600
 ccatggccat ctacaagcag tcacagcaca tgacggaggt tgtgaggcgc tgccccacc 660
 atgagcgtg ctccagatag gatggtctgg cccctcctca gcattctatc cgagtgggaag 720
 gaaatttgct tgtggagtat ttggatgaca gaaacacttt tcgacatagt gtggtggtgc 780
 cctatgagcc gcctgaggtt ggctctgact gtaccacct ccaactacaac tacatgtgta 840
 acagttcctg catggggcgc atgaaccgga ggcccatcct caccatcacc aactgggaag 900
 actccagtgg taatctactg ggacggaaaca gctttgaggt gcgtgtttgt gcctgtcctg 960
 ggagagaccg gcgcacagag gaagagaatc tccgcaagaa aggggagcct caccacgagc 1020
 tgccccagg gagcactaag cgagcactgc ccaacaacac cagctcctct ccccagccaa 1080
 agaagaaacc actggatgga gaattttca cccttcagat ccgtgggctg gacgcttcg 1140
 agatgttccg agagctgaat gaggccttgg aactcaagga tgcccaggct ggaaggagc 1200
 caggggggag cagggctcac tccagccacc tgaagtcaa aaagggtcag tctacctcc 1260
 gccataaaaa actcatgttc aagacagaag ggctgactc agactga 1307

-continued

```

<210> SEQ ID NO 30
<211> LENGTH: 837
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 30
gaggactccg cgacggctccg caccctgcgg ccagagcggc ttgagctcg gctgcttccg      60
cgctaggcgc tttttcccag aagcaatcca ggcgcgcccg ctggttcttg agcgcaggga     120
aaagcccgga gctaacgacc ggccgctcgg cactgcacgg ggccccaagc cgcagaagaa     180
ggacgacggg agggtaatga agctgagccc aggtctccta ggaaggagag agtgcgccgg     240
agcagcgtgg gaaagaaggg aagagtgtcg ttaagtttac ggccaacggt ggattatccg     300
ggccgctgcg cgtctggggg ctgcggaatg cgcgaggaga acaagggcat gccagtgagg     360
ggcggcagcg atgagggtct ggccaacgccc gcgcggggac tagtgagaga ggtgcgacac     420
tctctgggag ccggcgcgga tcccacgga gtcaaccggt tcggggaggcg cgcgatccag     480
gtcatgatga tgggcagcgc ccgctggcgg gagctgctgc tgctccaagg cgcggagccc     540
aactgcgcag accctgccac tctcaccgga ccggtgcatg atgctgcccg ggagggcttc     600
ctggacacgc tgggtgtgct gcaccggggc ggggcgcggc tggacgtgcg cgatgcctgg     660
ggtcgtctgc ccgtggactt ggccgaggag cggggccacc gcgacgtgac agggtaactg     720
cgcacagcca cgggggactg acgccagggt ccccagccgc ccacaacgac tttattttct     780
tacccaattt cccaccccca cccacctaata tcgatgaagg ctgccaacgg ggagcgg      837

```

```

<210> SEQ ID NO 31
<211> LENGTH: 987
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 31
cgagaggggg gagaacagac aacgggcggc ggggagcagc atggagccgg cggcggggag     60
cagcatggag ccttcggctg actggctggc cacggcccgcg gccccggggtc gggtagagga     120
ggtgcggggc ctgctggagg cggggggcgt gcccaacgca ccgaatagtt acggtcggag     180
gccgatccag gtcatgatga tgggcagcgc ccgagtggcg gagctgctgc tgctccaagg     240
cgcggagccc aactgcgcgg accccgccac tctcaccgga cccgtgcacg acgctgcccg     300
ggagggcttc ctggacacgc tgggtgtgct gcaccggggc ggggcgcggc tggacgtgcg     360
cgatgcctgg ggccgtctgc ccgtggacct ggctgaggag ctggggccatc gcgatgtcgc     420
acggtaacctg cgcgcggctg cggggggcac cagaggcagt aacctgccc gcatagatgc     480
cgcggaaggt ccctcagaca tccccgattg aaagaaccag agaggctctg agaaacctcg     540
ggaaacttag atcatcagtc accgaaggtc ctacagggcc acaactgccc ccgccacaac     600
ccaccccgct ttcgtagttt tcatttagaa aatagagctt ttaaaaatgt cctgcctttt     660
aacgtagata taagccttcc cccactaccg taaatgtcca tttatatcat tttttatata     720
ttcttataaa aatgtaaaaa agaaaaaacac cgcttctgcc ttttactgt gttggagttt     780
tctggagtga gcaactcacgc cctaagcgca cattcatgtg ggcatttctt gcgagcctcg     840
cagcctccgg aagctgtcga cttcatgaca agcattttgt gaactagggg agctcagggg     900
ggttactggc ttctcttgag tcacactgct agcaaatggc agaaccaaaag ctcaaataaa     960

```

-continued

```

aataaaataa ttttcattca ttcactc 987

<210> SEQ ID NO 32
<211> LENGTH: 8972
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 32

gtccaagggg agccaaggat ggctgcagct tcatatgac agttgttaa gcaagttgag 60
gcaactgaaga tggagaactc aaatcttcga caagagctag aagataattc caatcatctt 120
acaaaactgg aaactgaggc atctaataatg aaggaagtac ttaaacact acaaggaagt 180
attgaagatg aagctatggc ttctcttgga cagattgatt tattagagcg tcttaagag 240
cttaacttag atagcagtaa tttccctgga gtaaaactgc ggtcaaaaat gtcctccgt 300
tcttatggaa gccgggaagg atctgtatca agccgttctg gagagtgcag tcctgttcct 360
atgggttcat ttccaagaag aggggttgta aatggaagca gagaaagtac tggatattha 420
gaagaacttg agaaagagag gtcattgctt cttgctgac tgacaaaaga agaaaaggaa 480
aaagactggt attacgctca acttcagaat ctactaaaa gaatagatag tcttccttta 540
actgaaaatt tttccttaca aacagatatg accagaaggc aatggaata tgaagcaagg 600
caaatcagag ttgcatgga agaacaacta ggtacctgcc aggatatgga aaaacgagca 660
cagcgaagaa tagccagaat tcagcaaatc gaaaaggaca tacttctgat acgacagctt 720
ttacagtccc aagcaacaga agcagagagg tcactctcaga acaagcatga aaccggctca 780
catgatgctg agcggcagaa tgaaggtcaa ggagtgggag aaatcaacat ggcaacttct 840
ggtaatggtc aggggtcaac tacacgaatg gaccatgaaa cagccagtgt tttgagttct 900
agtagcacac actctgcacc tcgaaggctg acaagtcatc tgggaacca ggtggaatg 960
gtgtattcat tgttgcaat gcttggctact catgataagg atgatatgc gcgaactttg 1020
ctagctatgt ctagctccca agacagctgt atatccatgc gacagtctgg atgtcttct 1080
ctctcatcc agcttttaca tggcaatgac aaagactctg tattgttggg aaattcccgg 1140
ggcagttaaag aggctcgggc cagggccagt gcagcactcc acaacatcat tcaactcacag 1200
cctgatgaca agagaggcag gcgtgaaatc cgagtccttc atcttttga acagatacgc 1260
gcttactgtg aaacctgttg ggagtggcag gaagctcatg aaccaggcat ggaccaggac 1320
aaaaatccaa tgccagctcc tgttgaacat cagatctgtc ctgctgtgtg tgttctaatg 1380
aaactttcat ttgatgaaga gcatagacat gcaatgaatg aactaggggg actacaggcc 1440
attgcagaat tattgcaagt ggactgtgaa atgtacgggc ttactaatga ccactacagt 1500
attacactaa gacgatatgc tggaaatggct ttgacaaact tgacttttgg agatgtagcc 1560
aacaaggcta cgctatgctc tatgaaaggc tgcattgagag cacttgtggc ccaactaaaa 1620
tctgaaagtg aagacttaca gcaggttatt gcaagtgttt tgaggaatth gtcttggcga 1680
gcagatgtaa atagtaaaaa gacgttgcga gaagtggaa gtgtgaaagc attgatggaa 1740
tgtgtcttag aagttaaaaa ggaatcaacc ctcaaaagcg tattgagtgc cttatggaat 1800
ttgtcagcac attgcactga gaataaagct gatatatgtg ctgtagatgg tgcacttgca 1860
tttttggttg gcaactctac ttaccggagc cagacaaaaca ctttagccat tattgaaagt 1920
ggaggtggga tattacggaa tgtgtccagc ttgatagcta caaatgagga ccacaggcaa 1980

```

-continued

atcctaagag	agaacaactg	tctacaaact	ttattacaac	acttaaaatc	tcatagtttg	2040
acaatagtca	gtaatgcatg	tggaaacttg	tggaatctct	cagcaagaaa	tcctaaagac	2100
caggaagcat	tatgggacat	gggggcagtt	agcatgctca	agaacctcat	tcattcaaag	2160
cacaaaaatga	ttgctatggg	aagtgctgca	gctttaagga	atctcatggc	aaataggcct	2220
gcgaagtaca	aggatgccaa	tattatgtct	cctggctcaa	gcttgccatc	tcttcatggt	2280
aggaaacaaa	aagccotaga	agcagaatta	gatgctcagc	acttatcaga	aacttttgac	2340
aatatagaca	atttaagtcc	caaggcatct	catcgtagta	agcagagaca	caagcaaagt	2400
ctctatggty	attatgtttt	tgacaccaat	cgacatgatg	ataataggtc	agacaatttt	2460
aatactggca	acatgactgt	cctttcacca	tatttgaata	ctacagtgtt	accagctcc	2520
tcttcatcaa	gaggaagctt	agatagtctt	cgttctgaaa	aagatagaag	tttgagagaga	2580
gaacgcggaa	ttggctatag	caactacat	ccagcaacag	aaaatccagg	aacttcttca	2640
aagcgaggtt	tcagatctc	caccactgca	gcccagattg	ccaaagtcac	ggaagaagtg	2700
tcagccattc	atacctctca	ggaagacaga	agttctgggt	ctaccactga	attacattgt	2760
gtgacagatg	agagaaatgc	acttagaaga	agctctgctg	cccatacaca	ttcaaacact	2820
tacaatttca	ctaagtcgga	aaattcaaat	aggacatggt	ctatgcctta	tgccaaatta	2880
gaatacaaga	gatcttcaaa	tgatagttta	aatagtgcca	gtagttagta	tggttatggt	2940
aaaagaggtc	aatgaaacc	ctcgattgaa	tcctattctg	aagatgatga	aagtaagttt	3000
tgagttatg	gtcaatacc	agccgacct	gcccataaaa	tacatagtgc	aatcatatg	3060
gatgataatg	atggagaact	agatacacca	ataaattata	gtcttaata	ttcagatgag	3120
cagttgaaat	ctggaaggca	aagtccttca	cagaatgaaa	gatgggcaag	acccaaacac	3180
ataatagaag	atgaaataaa	acaaagttag	caaagacaat	caaggaatca	aagtacaact	3240
tatcctgttt	atactgagag	cactgatgat	aaacacctca	agttccaacc	acattttga	3300
cagcaggaat	gtgtttctcc	atacaggcca	cggggagcca	atggttcaga	aacaaatcga	3360
gtgggttcta	atcatggaat	taatcaaaat	gtaagccagt	ctttgtgtca	agaagatgac	3420
tatgaagatg	ataagcctac	caattatagt	gaacgttact	ctgaagaaga	acagcatgaa	3480
gaagaagaga	gaccaacaaa	ttatagcata	aaatataatg	aagagaaacg	tcagtgggat	3540
cagcctattg	attatagttt	aaaatagccc	acagatattc	cttcatcaca	gaaacagtca	3600
ttttcattct	caaagagttc	atctggacaa	agcagtaaaa	ccgaacatat	gtcttcaagc	3660
agtgagaata	cgtccacacc	ttcatcta	gccaagaggc	agaatcagct	ccatccaagt	3720
tctgcacaga	gtagaagtgg	tcagcctcaa	aaggctgcca	cttgcaaagt	ttcttctatt	3780
aaccaagaaa	caatacagac	ttattgtgta	gaagatactc	caatatgttt	ttcaagatgt	3840
agttcattat	catctttgtc	atcagctgaa	gatgaaatag	gatgtaatca	gacgacacag	3900
gaagcagatt	ctgctaatac	cctgcaaata	gcagaaataa	aagaaaagat	tggaactagg	3960
tcagctgaag	atcctgtgag	cgaagttcca	gcagtgctac	agcaccctag	aaccaaatec	4020
agcagactgc	agggttctag	tttatcttca	gaatcagcca	ggcacaaaagc	tgttgaattt	4080
tcttcaggag	cgaaatctcc	ctccaaaagt	gggtctcaga	cacccaaaag	tccacctgaa	4140
cactatgttc	aggagacccc	actcatgttt	agcagatgta	cttctgtcag	ttcacttgat	4200
agttttgaga	gtcgttcgat	tgccagctcc	gttcagagtg	aacatgcag	tggaatggta	4260

-continued

agtggcatta taagccccag tgatcttcca gatagccctg gacaaacat gccaccaagc 4320
agaagtaaaa cacctccacc acctcctcaa acagctcaa ccaagcgaga agtacctaaa 4380
aataaagcac ctactgctga aaagagagag agtggaccta agcaagctgc agtaaatgct 4440
gcagttcaga gggctcaggt tcttccagat gctgatactt tattacattt tgccacggaa 4500
agtactccag atggattttc ttgttcatcc agcctgagtg ctctgagcct cgatgagcca 4560
tttatacaga aagatgtgga attaagaata atgcctccag ttcaggaaaa tgacaatggg 4620
aatgaaacag aatcagagca gcctaagaa tcaaatgaaa accaagagaa agaggcagaa 4680
aaaactattg attctgaaaa ggacctatta gatgattcag atgatgatga tattgaaata 4740
ctagaagaat gtattatttc tgccatgcca acaaagtcac cacgtaaagc aaaaaagcca 4800
gcccagactg cttaaaaatt acctccacct gtggcaagga aaccaagtca gctgcctgtg 4860
tacaaaacttc taccatcaca aaacaggttg caaccccaaa agcatgttag ttttacaccg 4920
ggggatgata tgccacgggt gtattgtgtt gaaggacac ctataaaactt tccacagct 4980
acatctctaa gtgatctaac aatcgaatcc cctccaaatg agttagctgc tggagaagga 5040
gttagaggag gagcacagtc aggtgaattt gaaaaacgag ataccattcc tacagaaggc 5100
agaagtacag atgaggtcca aggagggaaa acctcatctg taaccatacc tgaattggat 5160
gacaataaag cagaggaagg tgatattctt gcagaatgca ttaattctgc tatgccaaa 5220
gggaaaagtc acaagccttt cegtgtgaaa aagataatgg accaggtcca gcaagcatct 5280
gcgtctgtct ctgcacccaa caaaaatcag ttatagtgta agaaaaagaa accaacttca 5340
ccagtaaac ctataccaca aaatactgaa tataggacac gtgtaagaaa aaatgcagac 5400
tcaaaaaata atttaaatgc tgagagagtt ttctcagaca acaagattc aaagaaacag 5460
aatttgaaaa ataattccaa ggacttcaat gataagctcc caaataatga agatagagtc 5520
agaggaagtt ttgcttttga ttcacctcat cattacacgc ctattgaagg aactccttac 5580
tgtttttcac gaaatgatc tttgagttct ctatgatttg atgatgatga tgttgacctt 5640
tccagggaaa aggtgaatt aagaaaggca aaagaaaata aggaatcaga ggctaaagtt 5700
accagccaca cagaactaac ctccaaccaa caatcagcta ataagacaca agctattgca 5760
aagcagccaa taaatcgagg tcagcctaaa cccatacttc agaacaatc cacttttccc 5820
cagtcaccca aagacatacc agacagaggg gcagcaactg atgaaaagtt acagaatfff 5880
gctattgaaa atactccagt ttgcttttct cataattcct ctctgagttc tctcagtgac 5940
attgaccaag aaaacaacaa taaagaaaat gaacctatca aagagactga gccccctgac 6000
tcacagggag aaccaagtaa acctcaagca tcaggctatg ctctaaatc atttcatggt 6060
gaagatacc cagtttgttt ctcaagaaac agttctctca gttctcttag tattgactct 6120
gaagatgacc tgttgagga atgtataagc tccgcaatgc caaaaaagaa aaagccttca 6180
agactcaagg gtgataatga aaaacatagt cccagaaata tgggtggcat attaggtgaa 6240
gatctgacac ttgatttgaa agatatacag agaccagatt cagaacatgg tctatcccct 6300
gattcagaaa attttgattg gaaagctatt caggaagtg caaattccat agtaagtagt 6360
ttacatcaag ctgctgctgc tgcattgta tctagacaag ctctgctga ttcagattcc 6420
atcctttccc tgaatcag gaaatctctg ggatcaccat tcatcttac acctgatcaa 6480
gaagaaaaac cctttacaag taataaagc ccaagaaatc taaaaccagg ggagaaaagt 6540

-continued

acattggaaa	ctaaaaagat	agaatctgaa	agtaaaggaa	tcaaaggagg	aaaaaaagt	6600
tataaaagt	tgattactgg	aaaagttcga	tctaattcag	aaatttcagg	ccaaatgaaa	6660
cagccccctt	aagcaaacat	gccttcaatc	tctcgaggca	ggacaatgat	tcatattcca	6720
ggagttcgaa	atagctcctc	aagtacaagt	cctgtttcta	aaaaaggccc	acccttaag	6780
actccagcct	ccaaaagccc	tagtgaaggt	caaacagcca	ccacttctcc	tagaggagcc	6840
aagccatctg	tgaaatcaga	attaagccct	gttgccaggc	agacatccca	aatagggtgg	6900
tcaagtaaag	cacttcttag	atcaggatct	agagattcga	ccccttcaag	acctgcccag	6960
caaccattaa	gtagacctat	acagtctcct	ggccgaaact	caatttcccc	tggtagaat	7020
ggaaataagt	ctcctaacaa	attatctcaa	cttccaagga	catcatcccc	tagtactgct	7080
tcaactaagt	cctcaggttc	tggaaaaatg	tcatatacat	ctccaggtag	acagatgagc	7140
caacagaacc	ttaccaaaca	aacaggttta	tccaagaatg	ccagtagtat	tccaagaagt	7200
gagtctgcct	ccaaaaggact	aaatcagatg	aataatggta	atggagccaa	taaaaaggta	7260
gaactttcta	gaatgtcttc	aactaaatca	agtggaagtg	aatctgatag	atcagaaga	7320
cctgtattag	tacgccagtc	aactttcatc	aaagaagctc	caagcccaac	cttaagaaga	7380
aaatggagg	aatctgcttc	atgtgaatct	ctttctccat	catctagacc	agcttctccc	7440
actaggtccc	aggcacaaa	tccagtttta	agtccttccc	ttcctgatat	gtctctatcc	7500
acacattcgt	ctgttcaggc	tggtggatgg	cgaaaactcc	cacctaatct	cagtcccact	7560
atagagtata	atgatggaag	accagcaaa	cgccatgata	ttgcacggtc	tcattctgaa	7620
agtccttcta	gacttccaat	caataggta	ggaacctgga	aacgtgagca	cagcaaacat	7680
tcatcatccc	ttcctcgagt	aagcacttgg	agaagaactg	gaagtctatc	ttcaattctt	7740
tctgttctat	cagaatccag	tgaaaaagca	aaaagtgagg	atgaaaaaca	tgtgaactct	7800
atctcaggaa	ccaaacaaag	taaagaaaac	caagtatccg	caaaaggaac	atggagaaaa	7860
ataaaagaaa	atgaatttcc	tcccacaaat	agtacttctc	agaccgtttc	ctcaggtgct	7920
acaaatgggt	ctgaatcaaa	gactctaatt	tatcaaatgg	cacctgctgt	ttctaaaaaca	7980
gaggatgttt	gggtgagaat	tgaggactgt	cccatcaaca	atcctagatc	tggaagatct	8040
cccacaggta	atactcccc	ggtgattgac	agtgtttcag	aaaaggcaaa	tccaaacatt	8100
aaagattcaa	aagataatca	ggcaaaaaca	aatgtgggta	atggcagtg	tcccattgct	8160
accgtgggtt	tggaaaatcg	cctgaactcc	tttattcagg	tggatgcccc	tgaccaaaaa	8220
ggaaactgaga	taaaaccagg	acaaaataat	cctgtccctg	tatcagagac	taatgaaagt	8280
tctatagtg	aacgtacccc	atcagttct	agcagctcaa	gcaaacacag	ttcacctagt	8340
gggactgttg	ctgccagagt	gactcctttt	aattacaacc	caagccctag	gaaaagcagc	8400
gcagatagca	cttcagctcg	gccatctcag	atcccaactc	cagtgaataa	caacacaaag	8460
aagcgagatt	ccaaaactga	cagcacagaa	tccagtggaa	cccaaagtcc	taagcgccat	8520
tctgggtctt	accttgtgac	atctgtttta	aagagaggaa	gaatgaaact	aagaaaattc	8580
tatgttaatt	acaactgcta	tatagacatt	tgtttcaaaa	tgaaacttta	aaagactgaa	8640
aaatthtga	aataggtttg	attcttgta	gaggtttttt	gttctggaag	ccatatttga	8700
tagtatactt	tgtcttcact	ggtcttattt	tgggaggcac	tcttgatggt	taggaaaaaa	8760
atagtaaagc	caagtatggt	tgtacagtat	gttttacatg	tatttaaagt	agcaccatc	8820

-continued

```

ccaacttcct ttaattattg cttgtcttaa aataatgaac actacagata gaaaatatga 8880
tatattgctg ttatcaatca tttctagatt ataaactgac taaacttaca tcagggaaaa 8940
attggtattt atgcaaaaaa aatggttttt gt 8972

```

```

<210> SEQ ID NO 33
<211> LENGTH: 5711
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

```

```

<400> SEQUENCE: 33

```

```

agctcgctga gacttctctg accccgcacc aggctgtggg gtttctcaga taactgggcc 60
ctctcgctca ggaggccttc accctctgct ctgggtaaaag ttcattggaa cagaaaagaaa 120
tggatttata tgctctctgc gttgaagaag tacaanaatgt cattaatgct atgcagaaaa 180
tcttagagtg tcccactctg ctggagttga tcaaggaacc tgtctccaca aagtgtgacc 240
acatattttg caaattttgc atgctgaaac ttctcaacca gaagaaaggg ccttcacagt 300
gtcctttatg taagaatgat ataaccaaaa ggagcctaca agaaagtacg agatttagtc 360
aactgtttga agagctattg aaaatcattt gtgcttttca gcttgacaca ggtttggagt 420
atgcaaacag ctataatttt gcaaaaaagg aaaataactc tcctgaacat ctaaaagatg 480
aagtttctat catccaaagt atgggctaca gaaaccgtgc caaagactt ctacagagtg 540
aaccgaaaa tccttctctg caggaacca gtctcagtgt ccaactctct aaccttgaa 600
ctgtgagaac tctgaggaca aagcagcgga tacaacctca aaagacgtct gtctacattg 660
aattgggatc tgattcttct gaagataccg ttaataaggc aacttatgct agtgtgggag 720
atcaagaatt gttacaaatc acccctcaag gaaccagga tgaatcagt ttggattctg 780
caaaaaaggc tgcttgtaa ttttctgaga cggatgtaac aaatactgaa catcatcaac 840
ccagtaataa tgatttgaac accactgaga agcgtgcagc tgagaggcat ccagaaaagt 900
atcagggtag ttctgtttca aactgcatg tggagccatg tggcacaat actcatgcca 960
gctcattaca gcatgagaac agcagtttat tactcactaa agacagaatg aatgtagaaa 1020
aggctgaatt ctgtaataaa agcaaacagc ctggcttagc aaggagccaa cataacagat 1080
gggctggaag taaggaaaca tgtaatgata ggcggactcc cagcacagaa aaaaaggtag 1140
atctgaatgc tgatcccctg tgtgagagaa aagaatggaa taagcagaaa ctgccatgct 1200
cagagaatcc tagagatact gaagatgttc cttggataac actaaatagc agcattcaga 1260
aagttaatga gtgggtttcc agaagtgatg aactgttagg ttctgatgac tcacatgatg 1320
gggagtctga atcaaatgcc aaagtagctg atgtattgga cgttctaata gaggtagatg 1380
aatattctgg ttcttcagag aaaatagact tactggccag tgatcctcat gaggctttaa 1440
tatgtaaaag tgaagagatt cactccaaat cagtagagag taatattgaa gacaaaatat 1500
ttgggaaaac ctatcggaag aaggcaagcc tcccactt aagccatgta actgaaaaatc 1560
taattatagg agcatttgtt actgagccac agataatata agagcgtccc ctcaaaaata 1620
aattaaagcg taaaaggaga cctacatcag gccttcatcc tgaggatttt atcaagaaag 1680
cagatttggc agttcaaaag actcctgaaa tgataaatca gggaaactaac caaacggagc 1740
agaatggtca agtgatgaat attactaata gtggtcatga gaataaaaca aaaggtgatt 1800
ctattcagaa tgagaaaaat cctaaccctaa tagaatcact cgaaaaagaa tctgctttca 1860

```

-continued

aaacgaaagc tgaacctata agcagcagta taagcaatat ggaactcgaa ttaaatatcc	1920
acaattcaaa agcacctaaa aagaataggc tgaggaggaa gtctttctacc aggcattatc	1980
atgcgcttga actagtagtc agtagaaatc taagcccacc taattgtact gaattgcaaa	2040
ttgatagttg ttctagcagt gaagagataa agaaaaaaaa gtacaaccaa atgccagtca	2100
ggcacagcag aaacctacaa ctcatggaag gtaaagaacc tgcaactgga gccaaagaaga	2160
gtaacaagcc aatgaacag acaagtaaaa gacatgacag cgatactttc ccagagctga	2220
agttaacaaa tgcacctggt tcttttacta agtgttcaaa taccagtga cttaaagaat	2280
ttgtcaatcc tagccttcca agagaagaaa aagaagagaa actagaaaca gttaaagtgt	2340
ctaataatgc tgaagacccc aaagatctca tgttaagtgg agaagggtt ttgcaactg	2400
aaagatctgt agagagtagc agtatctcat tggctacctg tactgattat ggcactcagg	2460
aaagtatctc gttactggaa gtttagcactc tagggaaggc aaaaacagaa ccaataaat	2520
gtgtgagtc gtgtgcagca tttgaaaacc ccaagggact aatcatggt tgttccaaag	2580
ataatagaaa tgacacagaa ggctttaagt atccattggg acatgaagtt aaccacagtc	2640
gggaaacaag catagaaatg gaagaagtg aacttgatgc tcagtatttg cagaatacat	2700
tcaaggtttc aaagcggcag tcatttgctc cgttttcaaa tccaggaaat gcagaagagg	2760
aatgtgcaac atttctctgc cactctgggt ccttaaagaa acaagtcca aaagtcactt	2820
ttgaatgtga acaaaaggaa gaaaatcaag gaaagaatga gtctaatac aagcctgtac	2880
agacagttaa tatcactgca ggctttcctg tggttggtca gaaagataag ccagttgata	2940
atgccaaatg tagtatcaaa ggaggtccta ggttttgtct atcatctcag ttcagaggca	3000
acgaaactgg actcattact ccaataaac atggactttt acaaaacca tctcgtatac	3060
caccactttt tcccataag tcatttgta aactaaatg taagaaaaat ctgctagagg	3120
aaaactttga ggaacattca atgtcacctg aaagagaaat gggaaatgag aacattccaa	3180
gtacagttag cacaattagc cgtaataaca ttagagaaaa tgtttttaa gaagccagct	3240
caagcaatat taatgaagta ggttccagta ctaatgaagt gggctccagt attaataaaa	3300
taggttccag tgatgaaac attcaagcag aactaggtag aacagaggg ccaaaattga	3360
atgctatgct tagattaggg gttttgcaac ctgaggtcta taaacaaagt cttcctggaa	3420
gtaattgtaa gcatcctgaa ataaaaaagc aagaatatga agaagtagtt cagactgtta	3480
atacagattt ctctccatat ctgatttcag ataacttaga acagcctatg ggaagtagtc	3540
atgcatctca ggtttgttct gagacacctg atgacctgtt agatgatggt gaaataaagg	3600
aaagatactag ttttctgtaa aatgacatta aggaaagttc tgctgttttt agcaaaagcg	3660
tccagaaagg agagcttagc aggagtccta gccctttcac ccatacacat ttggctcagg	3720
gttaccgaag aggggccaag aaattagagt cctcagaaga gaacttatct agtgaggatg	3780
aaagacttcc ctgcttccaa cacttggtat ttggtaaagt aaacaatata ctttctcagt	3840
ctactaggca tagcaccggt gctaccgagt gtctgtctaa gaacacagag gagaatttat	3900
tatcattgaa gaatagctta aatgactgca gtaaccaggt aatattggca aaggcatctc	3960
aggaacatca ccttagtgag gaaacaaat gttctgctag cttgttttct tcacagtgca	4020
gtgaattgga agacttgact gcaaatatac acaccagga tcctttcttg attggttctt	4080
ccaacaaat gaggcatcag tctgaaagcc agggagttgg tctgagtgac aaggaattgg	4140

-continued

```

tttcagatga tgaagaaaga ggaacgggct tggaaagaaa taatcaagaa gagcaaagca 4200
tggattcaaa cttagggtgaa gcagcatctg ggtgtgagag tgaacaagc gtctctgaag 4260
actgctcagg gctatcctct cagagtgaca ttttaaccac tcagcagagg gataccatgc 4320
aacataacct gataaagctc cagcaggaaa tggctgaact agaagctgtg ttagaacagc 4380
atgggagcca gccttctaac agctaccctt ccatcataag tgactcttct gcccttgagg 4440
acctgcgaaa tccagaacaa agcacatcag aaaaagcagt attaacttca cagaaaagta 4500
gtgaatacce tataagccag aatccagaag gcctttctgc tgacaagttt gaggtgtctg 4560
cagatagttc taccagtaaa aataaagaac caggagtgga aaggatcatcc ccttctaaat 4620
gcccattcatt agatgatagg tggatcatgc acagttgctc tgggagtctt cagaatagaa 4680
actaccatc tcaagaggag ctcatlaagg ttgttgatgt ggaggagcaa cagctggaag 4740
agtctgggcc acacgattg acggaacat cttacttgc aaggcaagat ctgaggggaa 4800
ccccttacct ggaatctgga atcagcctct tctctgatga ccctgaatct gatccttctg 4860
aagacagagc cccagagtca gctcgtgttg gcaacatacc atcttcaacc tctgattga 4920
aagttcccca attgaaagt gcagaatctg cccagagtcc agctgctgct catactactg 4980
atactgctgg gtataatgca atggaagaaa gtgtgagcag ggagaagcca gaattgacag 5040
cttcaacaga aagggtcaac aaaagaatgt ccatggtggt gtctggcctg accccagaag 5100
aatttatgct cgtgtacaag tttgccagaa aacaccacat cactttaact aatctaatta 5160
ctgaagagac tactcatggt gttatgaaaa cagatgctga gtttgtgtgt gaacggacac 5220
tgaatatatt tctaggaatt gcgggaggaa aatggtagt tagctatttc tgggtgacct 5280
agtctattaa agaaagaaaa atgctgaatg agcatgattt tgaagtcaga ggagatgtgg 5340
tcaatggaag aaaccaccaa ggtccaaagc gagcaagaga atcccaggac agaaagatct 5400
tcagggggct agaaatctgt tgctatgggc ccttcaccaa catgccaca gatcaactgg 5460
aatggatggt acagctgtgt ggtgcttctg tggtaagga gctttcatca ttcaccctg 5520
gcacagggtt ccaccaatt gtggttgtgc agccagatgc ctggacagag gacaatggct 5580
tccatgcaat tgggcagatg tgtgaggcac ctgtggtgac ccgagagtgg gtgttgaca 5640
gtgtagcact ctaccagtgc caggagctgg acacctact gataccccag atccccaca 5700
gccactactg a 5711

```

<210> SEQ ID NO 34

<211> LENGTH: 2680

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 34

```

ggttatcctg aatacatgtc taacaathtt ccttgcaacg ttagctgttg ttttctactg 60
tttccaaag atcaaaattg cttcagaaat tggagacata tttgatttaa aaggaaaaac 120
ttgaacaaat ggacaatatg tctattacga atacaccaac aagtaatgat gcctgtctga 180
gcattgtgca tagtttgatg tgccatagac aagggtggaga gagtgaacaa tttgcaaaaa 240
gagcaattga aagtttggtg aagaagctga aggagaaaaa agatgaattg gattctttaa 300
taacagctat aactacaaat ggagctcatc ctagttaaag tgttaccata cagagaacat 360
tggatgggag gcttcagggt gctggtcgga aaggatttcc tcatgtgatc tatgccctc 420

```

-continued

tctggaggtg	gcctgatcct	cacaaaaatg	aactaaaaca	tgtaaataat	tgccagtatg	480
cgtttgactt	aaaatgtgat	agtgtctgtg	tgaatccata	tcactacgaa	cgagttgtat	540
cacctggaat	tgatctctca	ggattaacac	tcagagtaa	tgctccatca	agtatgatgg	600
tgaaggatga	atatgtgcat	gactttgagg	gacagccatc	gttgtccact	gaaggacatt	660
caattcaaac	catccagcat	ccaccaagta	atcgtgcatc	gacagagaca	tacagcacc	720
cagctctggt	agccccatct	gagtctaata	ctaccagcac	tgccaacttt	cccaacattc	780
ctgtggcttc	cacaagtctg	cctgccagta	tactgggggg	cagccatagt	gaaggactgt	840
tgcatagtag	atcagggcct	cagccaggac	agcagcagaa	tggatttact	ggtcagccag	900
ctacttacca	tcataacagc	actaccacct	ggactggaag	taggactgca	ccatacacac	960
ctaatttgcc	tcaccaccaa	aacggccatc	ttcagcacca	cccgcctatg	ccgccccatc	1020
ccggacatta	ctggcctggt	cacaatgagc	ttgcattcca	gcctcccatt	tccaatcatc	1080
ctgctcctga	gtattggtgt	tccattgctt	actttgaaat	ggatgttcag	gtaggagaga	1140
catttaaggt	tccttcaagc	tgccctattg	ttactgttga	tggatagctg	gaccttctg	1200
gaggagatcg	cttttgtttg	ggtcaactct	ccaatgtcca	caggacagaa	gccattgaga	1260
gagcaaggtt	gcacataggc	aaaggtgtgc	agttggaatg	taaaggtgaa	ggtgatgttt	1320
gggtcaggtg	ccttagtgac	cacgcggtct	ttgtacagag	ttactactta	gacagagaag	1380
ctgggcgtgc	acctggagat	gctgttcata	agatctacc	aagtgcata	ataaaggtct	1440
ttgatttgcg	tcagtgctcat	cgacagatgc	agcagcaggc	ggctactgca	caagctgcag	1500
cagctgcccc	ggcagcagcc	gtggcaggaa	acatccctgg	cccaggatca	gtaggtggaa	1560
tagctccagc	tatcagtctg	tcagctgctg	ctggaattgg	tgttgatgac	cttcgtcgct	1620
tatgcatact	caggatgagt	tttgtgaaag	gctggggacc	ggattacca	agacagagca	1680
tcaaagaaac	accttctgctg	attgaaatc	acttacaccg	ggccctccag	ctcctagacg	1740
aagtacttca	taccatgccg	attgcagacc	cacaaccttt	agactgaggt	cttttaccgt	1800
tggggccctt	aaccttatca	ggatggtgga	ctacaaaata	caatcctggt	tataatctga	1860
agatatattt	cacttttctt	ctgctttatc	ttttcataaa	gggttgaaaa	tgtgtttgct	1920
gccttgctcc	tagcagacag	aaactggatt	aaaacaattt	tttttctctc	ttcagaactt	1980
gtcaggcatg	gctcagagct	tgaagattag	gagaaacaca	ttcttattaa	ttcttcacct	2040
gttatgtatg	aaggaatcat	tccagtgcct	gaaaatttag	cccttataaa	cgtcttagag	2100
ccttttatct	gcagaacatc	gatatgtata	tcattctaca	gaataatcca	gtattgctga	2160
ttttaaggc	agagaagttc	tcaaagttaa	ttcacctatg	ttattttgtg	tacaagttgt	2220
tattgttgaa	catacttcaa	aaataatgtg	ccatgtgggt	gagtaattt	taccaagagt	2280
aactttactc	tgtgtttaa	aatgaagtta	ataatgtatt	gtaatcttct	atccaaaata	2340
tttttgcaa	gttatattag	tgaagatggt	ttcaattcag	attgtcttgc	aacttcagtt	2400
ttatttttgc	caaggcaaaa	aactcttaat	ctgtgtgtat	attgagaatc	ccttaaaatt	2460
accagacaaa	aaaattttaa	attacgtttg	ttattcctag	tggatgactg	ttgatgaagt	2520
atacttttcc	cctgttaaac	agtagttgta	ttcttctgta	ttctaggca	caaggttggt	2580
tgctaagaag	cctataagag	gaatttcttt	tccttcattc	atagggaag	gttttgtatt	2640
ttttaaaaca	ctaaaagcag	cgtcactcta	cctaattgtct			2680

-continued

```

<210> SEQ ID NO 35
<211> LENGTH: 1095
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 35
tccccgctct gctctgtccg gtcacaggac tttttgccct ctgttcccgg gtcacctcagg 60
cggccaccca gtgggcacac tcccaggcgg cgctccggcc ccgcgctccc tccctctgcc 120
tttcattccc agctgtcaac atcctggaag ctttgaagct caggaaagaa gagaaatcca 180
ctgagaacag tctgtaaagg tccgtagtgc tatctacatc cagacggtgg aaggagagaga 240
aagagaaaga aggtatccta ggaatacctg cctgcttaga ccctctataa aagctctgtg 300
catcctgccca ctgaggactc cgaagaggtg gcagtcttct gaaagacttc aactgtgagg 360
acatgtcgtt cagatttggc caacatctca tcaagccctc tgtagtgttt ctcaaaacag 420
aactgtcctt cgctcttgtg aataggaaac ctgtggtacc aggacatgct cttgtgtgcc 480
cgctgcggcc agtggagcgc ttccatgacc tgcgtcctga tgaagtggcc gatttgtttc 540
agacgaccca gagagtcggg acagtgggtg aaaaacattt ccatgggacc tctctcacct 600
tttccatgca ggatggcccc gaagccggac agactgtgaa gcacggtcac gtccatgttc 660
ttcccaggaa ggctggagac tttcacagga atgacagcat ctatgaggag ctccagaaac 720
atgacaagga ggactttcct gcctcttggg gatcagagga ggaatggca gcagaagccg 780
cagctctcgc ggtctacttt cagtgcacac gatgttttcc agatcctgaa ttccagcaaa 840
agagctattg ccaaccagtt tgaagaccgc cccccgcct ctcccaaga ggaactgaat 900
cagcatgaaa atgcagtttc ttcattctac catcctgtat tcttcaacca gtgatcccc 960
acctcgggtc ctccaactcc cttaaaatac ctagacctaa acggctcaga caggcagatt 1020
tgaggtttcc ccctgtctcc ttattcggca gccttatgat taaacttct tctctgctgc 1080
aaaaaaaaaa aaaaa 1095

```

```

<210> SEQ ID NO 36
<211> LENGTH: 2234
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 36
aggggacgca gcgaaaccgg ggcccgcgcc aggccagccg ggacggacgc cgatgcccgg 60
ggctgcgacg gctgcagagc gagctgcctt cggaggccgg cgtggggaag atggcccagt 120
ccaccgccac ctcccctgat gggggcacca cgtttgagca cctctggagc tctctggaac 180
cagacagcac ctacttcgac cttccccagt caagccgggg gaataatgag gtggtgggcg 240
gaacggattc cagcatggac gtcttccacc tggagggcat gactacatct gtcattggccc 300
agttcaatct gctgagcagc accatggacc agatgagcag ccgcgcggcc tcggccagcc 360
cctacacccc agagcacgcc gccagcgtgc ccaccactc gccctacgca caaccagct 420
ccaccttga caccatgtgc ccggcgctgt tcatccccct caacaccgac taccctggac 480
cccaccactt tgaggtcact ttccagcagt ccagcacggc caagtcagcc acctggacgt 540
actccccgct cttgaagaaa ctctactgcc agatcgccaa gacatgcccc atccagatca 600
aggtgtccac ccgcccacc ccaggcactg ccatccgggc catgcctgtt tacaagaaag 660

```

-continued

```

cggagcacgt gaccgacgtc gtgaaacgct gcccacaacca cgagctcggg agggacttca 720
acgaaggaca gtctgtctcca gccagccacc tcataccgct ggaaggcaat aatctctcgc 780
agtatgtgga tgacctgtc accggcaggc agagcgtcgt ggtgccctat gagccaccac 840
aggtggggac ggaattcacc accatcctgt acaacttcat gtgtaacagc agctgtgtag 900
ggggcatgaa cgggcggccc atcctcatca tcataccct ggagatgcgg gatgggcagg 960
tgctgggccc cgggtccttt gagggccgca tctgcgctg tcctggccgc gaccgaaaag 1020
ctgatgagga ccaactaccg gagcagcagg cctgaacga gagctccgcc aagaacgggg 1080
ccgccagcaa gcgtgccttc aagcagagcc cccctgccgt ccccgccctt ggtgccggtg 1140
tgaagaagcg gcggcatgga gacgaggaca cgtactacct tcagggtcga ggcgggaga 1200
actttgagat cctgatgaag ctgaaagaga gcctggagct gatggagtgt gtgccgcagc 1260
cactggtgga ctctatcgg cagcagcagc agctcctaca gaggccgagt cacctacagc 1320
ccccgtccta cgggcgggtc ctctcgccta tgaacaaggt gcacgggggc atgaacaagc 1380
tgccctcctg caaccagctg gtgggccagc ctccccgca cagttcggca gctacacca 1440
acctggggcc cgtgggcccc gggatgctca acaaccatgg ccacgcagtg ccagccaacg 1500
gcgagatgag cagcagccac agcggccagt ccatggtctc ggggtccca tgcactccgc 1560
caccccccta ccacgccgac cccagcctcg tcagttttt aacaggattg ggggttccaa 1620
actgcatcga gtatttcacc tcccagggt tacagagcat ttaccactg cagaacctga 1680
ccattgagga cctgggggccc ctgaagatcc ccgagcagta ccgcatgacc atctggcggg 1740
gcctgcagga cctgaagcag ggccaagact acagcaccgc gcagcagctg ctccgctcta 1800
gcaacgcggc caccatctcc atcggcggct caggggaact gcagcgcag cgggtcatgg 1860
aggcctgca ctcccgctg cggcacacca tcaccatccc caaccgccc gggccaggcg 1920
gcggccctga cgagtgggcg gacttcggct tcgacctgcc cgactgcaag gcccgcaagc 1980
agcccatcaa ggaggagttc acggaggccc agatccactg agggcctcgc ctggctgcag 2040
cctgcgccac cggccagaga cccaagctgc ctccccctc cttcctgtgt gtccaaaact 2100
gcctcaggag gcaggacctt cgggctgtgc cgggggaaag gcaaggtccg gcccatcccc 2160
aggcacctca caggccccag gaaaggccca gccaccgaag ccgctgtgg acagcctgag 2220
tcacctgcag aacc 2234

```

<210> SEQ ID NO 37

<211> LENGTH: 4344

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 37

```

atggcctcgg ctggtaacgc cgcgagccc caggaccgcg gcggcggcgg cagcggctgt 60
atcgggtccc cgggacggcc ggctggagcc gggaggcgca gacggacggg ggggctgcgc 120
cgtgctgcgg cgcggaccg ggaactatctg caccggccca gctactgca cgcgccttc 180
gctctggagc agatttcaa ggggaaggct actggccgga aagcggcact gtggctgaga 240
gcgaagtffc agagactctt atttaaactg ggttgttaca tcaaaaaa ctgcggcaag 300
ttcttggtg tgggcctcct catatttggg gccttcgccc tgggattaaa agcagcgaac 360
ctcgagacca acgtggagga gctgtgggtg gaagttggag gacgagtaag tcgtgaatta 420

```

-continued

aattatactc gccagaagat tggagaagag gctatgttta atcctcaact catgatacag	480
accocctaaag aagaagggtgc taatgtcctg accacagaag cgctcctaca acacctggac	540
tggcactcc aggccagccg tgtccatgta tacatgtaca acaggcagtg gaaattggaa	600
catttgtgtt acaaatcagg agagcttacc acagaaacag gttacatgga tcagataata	660
gaatatcttt acccttgttt gattattaca cctttggact gcttctggga aggggcgaaa	720
ttacagtctg ggacagcata cctcctaggt aaacctcctt tgcggtggac aaacttcgac	780
cctttggaat tcctggaaga gttaaagaaa ataaactatc aagtggacag ctgggaggaa	840
atgctgaata aggctgaggt tggtcatggt tacatggacc gccctgcct caatccggcc	900
gatccagact gccccccac agcccccaac aaaaattcaa ccaaacctct tgatatggcc	960
cttgttttga atggtggatg tcatggctta tccagaaagt atatgcaactg gcaggaggag	1020
ttgattgtgg gtggcacagt caagaacagc actggaaaac tcgtcagcgc ccatgcctctg	1080
cagaccatgt tccagtaat gactcccaag caaatgtacg agcacttcaa ggggtacgag	1140
tatgtctcac acatcaactg gaacgaggac aaagcggcag ccatcctgga ggcctggcag	1200
aggacatag tggaggtggt tcatcagagt gtcgcacaga actccactca aaagtgctt	1260
tccttcacca ccacgacctt ggacgacatc ctgaaatcct tctctgacgt cagtgtcatc	1320
cgcgaggcca gcggctactt actcatgctc gcctatgcct gtctaaccat gctgcgctgg	1380
gactgtccca agtcccaggg tgccgtgggg ctggtggtggc tcctgctggt tgcactgtca	1440
gtggctgcag gactgggctt gtgctcattg atcggaatct cctttaacgc tgcaacaact	1500
caggttttgc cttttctcgc tcttgggtgtt ggtgtggatg atgtttttct tctggcccac	1560
gccttcagtg aacagggaca gaataaaaga atcccttttg aggacaggac cggggagtgc	1620
ctgaagcgca caggagccag cgtggccctc acgtccatca gcaatgtcac agccttcttc	1680
atggccgctg taatcccaat tcccgcctg cgggcgttct ccctccaggc agcggtagta	1740
gtggtgttca attttgccat ggttctgctc atttttctg caattctcag catggattta	1800
tatcgacgag aggacaggag actggatatt ttctgctggt ttacaagccc ctgcgtcagc	1860
agagtgattc aggttgaacc tcaggcctac accgacacac acgacaatac ccgctacagc	1920
ccccacctc cctacagcag ccacagcttt gcccatgaaa cgcagattac catgcagtcc	1980
actgtccage tccgcacgga gtacgacccc cacacgcagc tgtactacac caccgctgag	2040
ccgcgctccg agatctctgt gcagcccctc accgtgacac aggacacctt cagctgccag	2100
agcccagaga gcaccagctc cacaagggac ctgctctccc agttctccga ctccagcctc	2160
cactgcctcg agccccctg tacgaagtgg acaactctcat cttttgctga gaagcactat	2220
gctcctttcc tcttgaacc aaaagccaag gtagtgggtga tcttcctttt tctgggcttg	2280
ctgggggtca gcctttatgg caccacccga gtgagagacg ggtggacct tacggacatt	2340
gtacctcggg aaaccagaga atatgacttt attgctgcac aattcaaata cttttctttc	2400
tacaacatgt atatagtcac ccagaaagca gactaccgga atatccagca cttactttac	2460
gacctacaca ggagtctcag taactgtgaag tatgtcatgt tggaaagaaa caaacagctt	2520
ccaaaaatgt ggctgcaacta cttcagagac tggcttcagg gacttcagga tgcatttgac	2580
agtgactggg aaaccgggaa aatcatgccca aacaattaca agaatggatc agacgatgga	2640
gtccttgctc acaaacctcct ggtgcaaac ccgagccgag ataagcccat cgacatcagc	2700

-continued

```

cagttgacta aacagcgtct ggtggatgca gatggcatca ttaatcccag cgctttctac 2760
atctacctga cggcttgggt cagcaacgac cccgtcgcgt atgctgcctc ccaggccaac 2820
atccggccac accgaccaga atgggtccac gacaaagccg actacatgcc tgaacaagg 2880
ctgagaatcc cggcagcaga gcccatcgag tatgcccagt tccctttcta cctcaacggg 2940
ttgcgggaca cctcagactt tgtggaggca attgaaaaag taaggacat ctgcagcaac 3000
tatacgagcc tggggctgtc cagttacccc aacggctacc ccttcctctt ctgggagcag 3060
tacatcggcc tccgccactg gctgctgctg ttcacacagc tgggtgtggc ctgcacattc 3120
ctcgtgtgcy ctgtcttctc tctgaacccc tggacggccg ggatcattgt gatggtcctg 3180
gcgctgatga cggctgagct gttcggcatg atgggcctca tcggaatcaa gctcagtgcc 3240
gtgcccgtgg tcatcctgat cgcttctgtt ggcataggag tggagtccac cgttcacggt 3300
gctttggcct ttctgacgcy catcggcgac aagaaccgca gggctgtgct tgccctggag 3360
cacatgtttg caccgcctct ggatggcgcc gtgtccactc tgetgggagt getgatgctg 3420
gcgggatctg agttcgactt cattgtcagg tatttctttg ctgtgctggc gatcctcacc 3480
atcctcggcy ttctcaatgg gctggttttg cttcccgtgc tttgtcttt ctttgacca 3540
tactcagagg tgtctccagc caacggcttg aaccgcctgc ccacaccctc cctgagcca 3600
ccccccagcy tggctccgct cgccatgccc cccggccaca cgcacagcgg gctgattcc 3660
tccgactcgy agtatagttc ccagacgaca gtgtcaggcc tcagcgagga gcttcggcac 3720
tacgaggccc agcaggcgcg gggaggccct gccaccaag tgatcgtgga agccacagaa 3780
aaccctgctc tcgcccactc cactgtggtc catcccgaat ccaggcatca cccaccctcg 3840
aacccgagac agcagcccca cctggactca gggtccttgc ctcccggacg gcaaggccag 3900
cagccccgca gggaccccc cagagaaggc ttgtggccac ccctctacag accgcgcaga 3960
gacgcttttg aaatttctac tgaagggcat tctggcccta gcaatagggc ccgctggggc 4020
cctcgcgggg ccggttctca caaccctcgy aaccagcgt cactgccat gggcagctcc 4080
gtgcccggct actgccagcc catcaccact gtgacggctt ctgctcctgt gactgtcgcc 4140
gtgcacccgc cgctgtccc tgggcctggg cggaaacccc gagggggact ctgccagggc 4200
taccctgaga ctgaccacgg cctgtttgag gacccccacg tgctttcca cgtccggtgt 4260
gagaggaggg attcgaaggt ggaagtcatt gagctgcagg acgtggaatg cgaggagagg 4320
ccccgggaa gcagctcca ctga 4344

```

<210> SEQ ID NO 38

<211> LENGTH: 4740

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 38

```

ttcggtttt tctcagggga cgttgaaatt atttttgtaa cgggagtcgg gagaggacgg 60
ggcgtgcccc gcgtgcgcgc gcgtgctct ccccggcgt cctccacagc tcgctggctc 120
ccgcgcgga aaggcgtcat gccgcccaca accccccgaa aaacggccgc caccgcccgc 180
gctgcccgcg cggaaacccc ggcaccgccc ccgcccctc ctctgagga ggaccagag 240
caggacagcy gcccggagga cctgcctctc gtcaggcttg agtttgaaga aacagaagaa 300
cctgatttta ctgcattatg tcagaaatta aagataccag atcatgtcag agagagagct 360

```


-continued

tggtaaactt	gggagaaagt	ttcatctgtg	gatggagtat	tgggaggtta	tattcaaaag	420
aaaaaggaac	tgtggggaat	ctgtatcttt	attgcacgag	ttgacctaga	tgagatgtcg	480
ttcactttac	tgagctacag	aaaaacatac	gaaatcagtg	tccataaatt	ctttaactta	540
ctaaaagaaa	ttgataccag	taccaaagtt	gataatgcta	tgtcaagact	gttgaagaag	600
tatgatgtat	tgtttgcact	cttcagcaaa	ttggaaagga	catgtgaact	tatatatttg	660
acacaaccca	gcagttcgat	atctactgaa	ataaattctg	cattggtgct	aaaagtttct	720
tggatcacat	ttttattagc	taaaggggaa	gtattacaaa	tggagatga	tctggtgatt	780
tcatttcagt	taatgctatg	tgtccttgac	tattttatta	aactctcacc	tcccatgttg	840
ctcaaagaac	catataaaac	agctgttata	cccattaatg	gttcacctcg	aacaccagg	900
cgaggtcaga	acaggagtgc	acggatagca	aaacaactag	aaaatgatac	aagaattatt	960
gaagttctct	gtaaagaaca	tgaatgtaat	atagatgagg	tgaaaaatgt	ttatttcaaa	1020
aattttatac	cttttatgaa	ttctcttgga	cttgtaacat	ctaattggact	tccagagggt	1080
gaaaatcttt	ctaaacgata	cgaagaaatt	tatcttaaaa	ataagatct	agatcgaaga	1140
ttatttttgg	atcatgataa	aactcttcag	actgattcta	tagacagttt	tgaaacacag	1200
agaacaccac	gaaaaagtaa	ccttgatgaa	gaggtgaata	taattcctcc	acacactcca	1260
gtaggactg	ttatgaacac	tatccaacaa	ttaatgatga	ttttaaattc	tgcaagtgat	1320
caaccttcag	aaaatctgat	ttcctatttt	aacaactgca	cagtgaatcc	aaaagaaagt	1380
atactgaaaa	gagtgaagga	tataggatac	atctttaaag	agaaatttgc	taaagctgtg	1440
ggacaggggt	gtgtcgaat	tggatcacag	cgatacaaac	tggagttcg	cttgatttac	1500
cgagtaatgg	aatccatgct	taaatcagaa	gaagaacgat	tatccattca	aaattttagc	1560
aaactctcga	atgacaacat	ttttcatatg	tctttattgg	cgtgcgctct	tgaggttgta	1620
atggccacat	atagcagaag	tacatctcag	aatcttgatt	ctggaacaga	tttgtctttc	1680
ccatggattc	tgaatgtgct	taatttaaaa	gcctttgatt	tttacaagt	gatcgaaagt	1740
tttatcaaa	cagaaggcaa	cttgacaaga	gaaatgataa	aacatttaga	acgatgtgaa	1800
catcgaatca	tggaatccct	tgcattgctc	tcagattcac	ctttatttga	tcttattaaa	1860
caatcaaagg	accgagaagg	accaactgat	caccttgaat	ctgcttgctc	tcttaatctt	1920
cctctccaga	ataatcacac	tgacagcagat	atgtatcttt	ctcctgtaag	atctccaaag	1980
aaaaaggtt	caactacgcg	tgtaaattct	actgcaaatg	cagagacaca	agcaacctca	2040
gccttcacga	cccagaagcc	attgaaatct	acctctcttt	cactgtttta	taaaaaagtg	2100
tatcggctag	cctatctccg	gctaaataca	ctttgtgaac	gccttctgtc	tgagcaccca	2160
gaattagaac	atatcatctg	gacccttttc	cagcacaccc	tgcagaatga	gtatgaactc	2220
atgagagaca	ggcatttgga	ccaaattatg	atgtgttcca	tgtatggcat	atgcaaagtg	2280
aagaatatag	accttaaaat	caaaatcatt	gtaacagcat	acaaggatct	tcctcatgct	2340
gttcaggaga	cattcaaacg	tgttttgatc	aaagaagagg	agtatgattc	tatttatagta	2400
ttctataact	cggctctcat	gcagagactg	aaaacaaata	ttttgcagta	tgcttccacc	2460
aggcccccta	ccttgccacc	aatacctcac	attcctcgaa	gcccttacia	gtttcctagt	2520
tcacccttac	ggattcctgg	agggaacatc	tatatttcac	ccctgaagag	tccatataaa	2580
atctcagaag	gtctgccaac	accaacaaaa	atgactccaa	gatcaagaat	cttagtatca	2640

-continued

attggtgaat cattcgggac ttctgagaag ttccagaaaa taaatcagat ggtatgtaac 2700
agcgaccgtg tgctcaaaag aagtgtgaa ggaagcaacc ctccataacc actgaaaaaa 2760
ctacgctttg atattgaagg atcagatgaa gcagatggaa gtaaacatct cccaggagag 2820
tccaaatttc agcagaaact ggcagaaatg acttctactc gaacacgaat gcaaaagcag 2880
aaaatgaatg atagcatgga tacctcaaac aaggaagaga aatgaggatc tcaggacctt 2940
ggtggacact gtgtacacct ctggattcat tgtctctcac agatgtgact gtataacttt 3000
cccagttct gtttatggcc acatttaata tcttcagctc tttttgtgga tataaaatgt 3060
gcagatgcaa ttgtttgggt gagtcctaag ccacttgaaa tgttagtcat tgttatttat 3120
acaagattga aaatctgtg taaatctctc catttaaaaa gttgtagcag attgtttcct 3180
cttccaaagt aaaattgctg tgctttatgg atagtaagaa tggccctaga gtgggagtcc 3240
tgataacca ggctgtctg actactttgc cttcttttgt agcatatagg tgatgtttgc 3300
tctgtttttt attaatttat atgtatattt ttttaattta acatgaacac ccttagaaaa 3360
tgtgtcctat ctatcttcca aatgcaattt gattgactgc ccattcacca aaattatcct 3420
gaactcttct gcaaaaaatg atattattag aaattagaaa aaaattacta attttacaca 3480
ttagatttta ttttactatt ggaatctgat aactgtgtg cttgttttat aaaattttgc 3540
tttaattaa ataaaagctg gaagcaaagt ataaccatat gatactatca tactactgaa 3600
acagatttca tacctcagaa tgtaaaagaa cttactgatt attttcttca tccaacttat 3660
gtttttaaat gaggattatt gatagtactc ttggttttta taccattcag atcactgaat 3720
ttataaagta cccatctagt acttgaaaa gtaaagtgtt ctgccagatc ttaggtatag 3780
aggaccctaa cacagtatat cccaagtgca ctttctaag tttctgggtc ctgaagaatt 3840
aagatacaaa ttaattttac tccataaaca gactgttaat tataggagcc ttaatttttt 3900
ttcatagag atttgtctaa ttgcactcga aaattattct gccctcctta atttgggaag 3960
gtttgtgttt tctctggaat ggtacatgct tccatgtat cttttgaaact ggcaattgct 4020
tatttatctt ttattttttt aagtcagtat ggtctaacac tggcatgttc aaagccacat 4080
tatttctagt ccaaaattac aagtaatcaa gggtcattat gggtaggca ttaatgtttc 4140
tatctgattt tgtgcaaaag cttcaaatga aaacagctgc attagaaaa gaggcgcttc 4200
tccccctccc tacacctaaa ggtgtattta aactatcttg tgtgattaac ttatttagag 4260
atgctgtaac ttaaaatagg ggatatttaa ggtagcttca gctagctttt aggaaaatca 4320
ctttgtctaa ctcaaatga tttttaaaaa gaaatctggg cttgttagaa acaaaaattt 4380
tattttgtgc tcatttaagt ttcaaaacta ctattttgac agttattttg ataacaatga 4440
cactagaaaa cttgactcca tttcatcatt gtttctgcat gaatatcata caaatcagtt 4500
agtttttagg tcaagggtct actatttctg ggtcttttgc tactaagttc acattagaat 4560
tagtgccaga attttaggaa cttcagagat cgtgtattga gatttcttaa ataatgcttc 4620
agatattatt gctttattgc ttttttgat tggttaaaac tgtacattta aaattgctat 4680
gttactattt tctacaatta atagtttctg tattttaaaa taaattagtt gtttaagagtc 4740

<210> SEQ ID NO 39

<211> LENGTH: 4608

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

-continued

<400> SEQUENCE: 39

atggagaata gtcttagatg tgtttgggta cccaagctgg cttttgtact cttcggagct 60
tccttgctca gcgcgcactc tcaagtaacc ggttttcaaa ttaaagcttt cacagcactg 120
cgcttcctct cagaaccttc tgatgccgtc acaatgcggg gaggaaatgt cctcctcgac 180
tgctccgcgg agtccgaccg aggagttcca gtgatcaagt ggaagaaaga tggcattcat 240
ctggccttgg gaatggatga aaggaagcag caactttcaa atgggtctct gctgatacaa 300
aacatacttc attccagaca ccacaagcca gatgagggac tttaccaatg tgaggcatct 360
ttaggagatt ctgggctaact tattagtccg acagcaaaag tgcagtagc aggaccactg 420
aggttccttt cacagacaga atctgtcaca gccttcctgg gagacacagt gctactcaag 480
tgtgaagtca ttggggagcc catgccaaca atccactggc agaagaacca acaagacctg 540
actccaatcc caggtgactc ccgagtgggt gtcttgccct ctggagcatt gcagatcagc 600
cgactccaac cgggggacat tggaaattac cgatgctcag ctcgaaatcc agccagctca 660
agaacaggaa atgaagcaga agtcagaatt ttatcagatc caggactgca tagacagctg 720
tattttctgc aaagaccatc caatgtagta gccattgaag gaaaagatgc tgcctggaa 780
tgttgtgttt ctggctatcc tccaccaagt tttacctggt tacgaggcga ggaagtcatc 840
caactcaggt ctaaaaagta ttctttattg ggtggaagca acttgcttat ctccaatgtg 900
acagatgatg acagtggaat gtatacctgt gttgtccat ataaaaatga gaatattagt 960
gcctctgcag agctcacagt cttggttccg ccattggttt taaatcatcc ttccaacctg 1020
tatgcctatg aaagcatgga tattgagttt gaatgtacag tctctggaaa gcctgtgccc 1080
actgtgaatt ggatgaagaa tggagatgtg gtcattccta gtgattattt tcagatagtg 1140
ggaggaagca acttacggat acttgggggtg gtgaagtcat atgaaggctt ttatcaatgt 1200
gtggctgaaa atgaggctgg aaatgcccag accagtgcac agctcattgt ccctaagcct 1260
gcaatcccaa gctccagtgt cctcccttcg gctcccagag atgtggctcc tgtcttggtt 1320
tccagccgat ttgtccgtct cagctggcgc ccacctgcag aagcgaaagg gaacattcaa 1380
actttcacgg tctttttctc cagagaaggt gacaacaggg aacgagcatt gaatacaaca 1440
cagcctgggt ccttcagct cactgtggga aacctgaagc cagaagccat gtacaccttt 1500
cgagttgtgg cttacaatga atggggaccg ggagagagtt ctcaaccat caaggtggcc 1560
acacagcctg agttgcaagt tccagggcca gtagaaaacc tgcaagctgt atctacctca 1620
cctacctcaa ttcttattac ctgggaaccc cctgcctatg caaacggtcc agtccaaggt 1680
tacagattgt tctgcactga ggtgtccaca ggaaaagaac agaatataga ggttgatgga 1740
ctatcttata aactggaagg cctgaaaaaa ttcaccgaat atagtcttcg attcttagct 1800
tataatcgct atggctccgg cgtctctact gatgatataa cagtggttac actttctgac 1860
gtgccaagtg ccccgcctca gaacgtctcc ctggaagtgg tcaattcaag aagtatcaaa 1920
gttagctggt tgctcctcc atcaggaaca caaaatgat ttattaccgg ctataaaatt 1980
cgacacagaa agacgaccgg caggggtgag atggaaacac tggagccaaa caacctctgg 2040
tacctattca caggactgga gaaaggaagt cagtacagtt tccaggtgtc agccatgaca 2100
gtcaatggta ctggaccacc ttccaactgg tatactgcag agactccaga gaatgatcta 2160
gatgaatctc aagtctctga tcaaccaagc tctctctatg tgaggcccca gactaactgc 2220

-continued

atcatcatga gttggactcc tcccttgaac ccaaaccatcg tgggtgcgagg ttatattatc	2280
ggttatggcg ttgggagccc ttacgctgag acagtgcgtg tggacagcaa gcagcgatat	2340
tattccattg agaggttaga gtcaagttcc cattatgtaa tctccctaaa agcttttaac	2400
aatgccggag aaggagtacc tctttatgaa agtgccacca ccaggtctat aaccgatccc	2460
actgacccag ttgattatta tcctttgctt gatgatttcc ccacctcggg cccagatctc	2520
tccaccccca tgcctccacc agtaggtgta caggctgtgg ctcttaccba tgatgctgtg	2580
agggtcagct gggcagacaa ctctgtccct aagaacaaa agacgtctga ggtgcgactt	2640
tacaccgtcc ggtggagaac cagcttttct gcaagtgcaa aatacaagtc agaagacaca	2700
acatctctaa gttacacagc aacaggcctc aaaccaaaca caatgatga attctcggtc	2760
atggtacaaa aaaacagaag gtccagtact tggagcatga ctgcacatgc caccacgtat	2820
gaagcagccc ccacctctgc tcccgaagac tttacagtca ttactagga agggaagcct	2880
cgtgccgtca ttgtgagttg gcagcctccc ttggaagcca atgggaaaat tactgcttac	2940
atcttatttt ataccttggc caagaacatc ccaattgatg actggattat ggaacaatc	3000
agtggtgata ggcttactca tcaaatcatg gatctcaacc ttgatactat gtattacttt	3060
cgaattcaag cacgaaatcc aaaaggagtg gggccactct ctgatcccat cctcttcagg	3120
actctgaaag tggaacaccc tgacaaaatg gctaataacc aaggtcgtca tggagatgga	3180
ggttatggcg cagttgatac taatttgatt gatagaagca ccctaaatga gccgccaatt	3240
ggacaaaatg acccccgcga tggcagtgct actcctcaga agaacagcaa cctgcttggtg	3300
atcattgtgg tcaccgttgg tgtcatcaca gtgctggtag tggatcatcg ggctgtgatt	3360
tgaccccgac gctcttcagc ccagcagaga aagaaacggg ccaccacag tgcctggcaaa	3420
aggaagggca gccagaagga cctccgaccc cctgatcttt ggatccatca tgaagaaatg	3480
gagatgaaaa atattgaaaa gccatctggc actgaccctg cagggaaggga ctctcccatc	3540
caaagtgtcc aagacctcac accagtcaag cacagccagt cagaacacca actggggaagc	3600
aaaagcacct ctcatctagg tcaagacact gaggaagcag ggagctctat gtccactctg	3660
gagaggtcgc tggctgcacg ccgagccccc cgggccaagc tcatgattcc catggatgcc	3720
cagtccaaca atcctgctgt cgtgagcgcc atcccgggtgc caacgctaga aagtgccag	3780
taccagga tccctccgctc tcccacctgt ggatacccc acccgcagtt cactctccgg	3840
cctgtgccat tcccacact ctcaagtggc cgaggtttcg gagcaggaag aagtcagtca	3900
gtgagtgaag gaccaactac ccaacaacca cctatgctgc ccccatctca gcctgagcat	3960
tctagcagcg aggagggcacc aagcagaacc atccccacag cttgtgttcg accaactcac	4020
ccactccgca gctttgctaa tcctttgcta cctccaccaa tgagtgcaat agaaccgaaa	4080
gtcccttaca caccactttt gtctcagcca gggcccactc ttcctaagac ccatgtgaaa	4140
acagcctccc ttgggttggc tggaaaaagc agatcccctt tgcttctgt gtctgtgcca	4200
acagccctg aagtgtctga ggagagccac aaaccaacag aggattcagc caatgtgtat	4260
gaacaggatg atctgagtga acaaatggca agtttgaag gactcatgaa gcagcttaat	4320
gccatcacag gctcagcctt ttaacatgta tttctgaatg gatgaggtga atttccggg	4380
aactttgcag cataccaatt acccataaac agcacacctg tgtccaagaa ctctaaccag	4440
tgtacaggtc acccatcagg accactcagt taaggaagat cctgaagcag ttcagaagga	4500

-continued

ataagcattc cttctttcac aggcacagc aattgtcaaa tgatgattat gagttcccta	4560
aacaaaagca aagatgcatt ttcactgcaa tgtcaaagtt tagctgct	4608

<210> SEQ ID NO 40

<211> LENGTH: 8959

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 40

ccccagcctc cttgccaacg ccccttttcc ctctccccct cccgctcggc gctgaccccc	60
catccccacc cccgtgggaa cactggggagc ctgcactcca cagaccctct ccttgccctc	120
tccctcacct cagcctccgc tccccgcct cttcccgcc cagggcgccg gccaccctt	180
ccctccgccc cccccggcc gcggggagga catggcccg cacaggccgg tggaatgggt	240
ccaggccgtg gtcagccgct tcgacgagca gcttccaata aaaacaggac agcagaacac	300
acataccaaa gtcagtactg agcacaacaa ggaatgtcta atcaatattt ccaaatataa	360
gttttctttg gttataagcg gcctcactac tattttaag aatgttaaca atatgagaat	420
atitggagaa gctgctgaaa aaaatttata tctctctcag ttgattatat tggatacact	480
ggaaaaatgt cttgctgggc aaccaaggga cacaatgaga ttagatgaaa cgatgctggt	540
caaacagtgt ctgccagaaa tctgccattt tcttcacacc tgtcgtgaag gaaaccagca	600
tgacagtgaa cttcggaatt ctgcctctgg ggttttattt tctctcagct gcaacaactt	660
caatgcagtc tttagtgcga tttctaccag gttacaggaa ttaactgttt gttcagaaga	720
caatgttgat gttcatgata tagaattgtt acagtatata aatgtggatt gtgcaaaatt	780
aaaacgactc ctgaaggaaa cagcatttaa atttaaagcc ctaaagaagg ttgctgagtt	840
agcagttata aatagcctgg aaaaggcatt ttggaactgg gtgaaaaatt atccagatga	900
atitcaaaa ctgtaccaga tcccacagac tgatatggct gaatgtgcag aaaagctatt	960
tgacttgggt gatggttttg ctgaaagcac caaacgtaaa gcagcagttt ggccactaca	1020
aatcattctc cttatcttgt gtccagaaat aatccaggat atatccaaag acgtggttga	1080
tgaaaacaac atgaataaga agttatttct ggacagtcta cgaagagctc ttgctggcca	1140
tgaggaaagt aggcagctga cagaaagtgc tgcaattgcc tgtgtcaaac tgtgtaaagc	1200
aagtacttac atcaattggg aagataactc tgtcattttc ctactgttcc agtccatggt	1260
ggttgatctt aagaacctgc tttttaatcc aagtaagcca ttctcaagag gcagtcagcc	1320
tgacagatgt gatctaatga ttgactgctc tgtttcttgc tttcgtataa gccctcacia	1380
caaccaacac tttaaagatct gcctggctca gaattcaact tctacatttc actatgtgct	1440
ggtaaatcca ctccatcgaa tcatcaccia ttccgcattg gattggtggc ctaagattga	1500
tgctgtgtat tgtcactcgg ttgaaacttc aaatatgttt ggtgaaacac ttcataaagc	1560
agtgcagggt tgtggagcac acccagcaat acgaatggca ccgagcttta catttaaaga	1620
aaaagtaaca agccttaaat ttaaagaaaa acctacagac ctggagacia gaagctataa	1680
gtatcttctc ttgtccatgg tgaaactaat tcatgcagat ccaaagctct tgctttgtaa	1740
tccaagaaaa caggggcccg aaaccaagc cagtacagca gaattaatta cagggctcgt	1800
ccaactggtc cctcagtcac acatgccaga gattgctcag gaagcaatgg aggctctgct	1860
ggttcttcat cagttagata gcattgatt gtggaactcc gatgctcctg tagaaaactt	1920

-continued

ttgggagatt agctcacaaa tgcttttita catctgcaag aaattaacta gtcacaaat	1980
gcttagtagc acagaaatc tcaagtgggt gcgggaaata ttgatctgca ggaataaatt	2040
tcttcttaaa aataagcagg cagatagaag ttctgtcac tttctcttt tttacgggt	2100
aggatgtgat attccttcta gtggaatac cagtcaaag tccatggatc atgaagaatt	2160
actacgtact cctggagcct ctctccgaa gggaaaagg aactcctcta tggatagtgc	2220
agcaggatgc agcggaaacc cccaatttg ccgacaagcc cagacaaac tagaagtggc	2280
cctgtacatg tttctgtgga accctgacac tgaagctgtt ctggttgcca tgtcctgtt	2340
ccgccacctc tgtgaggaag cagatatccg gtgtgggtg gatgaagtgt cagtgcataa	2400
cctcttgccc aactataaca cattcatgga gtttgcctc gtcagcaata tgatgtcaac	2460
aggaagagca gcacttcaga aaagagtgat ggactgctg aggcgcattg agcatccac	2520
tgaggaaac actgaggcct gggaagatc acatgcaaaa tgggaacaag caacaaagct	2580
aatccttaac tatccaaaag ccaaaatgga agatggccag gctgctgaaa gccttcacaa	2640
gaccattgtt aagaggcga tgtcccatgt gagtggagga ggatccatag atttgtctga	2700
cacagactcc ctacaggaat ggatcaacat gactggcttc ctttgtgccc ttggaggagt	2760
gtgctccag cagagaagca attctggcct ggcaacctat agcccaccca tgggtccagt	2820
cagtgaacgt aagggttcta tgatttcagt gatgtctca gagggaaacg cagatacacc	2880
tgtcagcaaa tttatggatc ggctgttgc cttaatggtg tgaacctatg agaaagtggg	2940
acttcaata cggaccaatg ttaaggatct ggtgggtcta gaattgagtc ctgctctgta	3000
tccaatgcta ttaacaaat tgaagaatac catcagcaag tttttgact cccaaggaca	3060
ggttttattg actgatacca atactcaatt tgtagaacaa accatagcta taatgaagaa	3120
cttgctagat aatcatactg aaggcagctc tgaacatcta gggcaagcta gcattgaaac	3180
aatgatgta aatctggta ggtatgttcg tgtgcttggg aatatggtcc atgcaattca	3240
aataaaaaac aaactgtgtc aattagttga agtaatgatg gcaaggagag atgaccttc	3300
atthtgccea gagatgaaat ttaggaataa gatggtagaa tacctgacag actgggttat	3360
gggaacatca aaccaagcag cagatgatga tgtaaatgt cttacaagag atttgacca	3420
ggcaagcatg gaagcagtag tttcacttct agctggcttc cctctgcagc ctgaagaagg	3480
agatggtgtg gaattgatgg aagccaaatc acagttattt cttaaatact tcacattatt	3540
tatgaacctt ttgaatgact gcagtgaagt tgaagatgaa agtgcgcaaa cagggtggcag	3600
gaaacgtggc atgtctcgga ggctggcatc actgaggcac tgtacggtec ttgcaatgtc	3660
aaacttactc aatgccaaag tagacagtgg tctcatgcac tccataggct taggttacca	3720
caaggatctc cagacaagag ctacatttat ggaagttctg acaaaaatcc ttcaacaagg	3780
cacagaattt gacacacttg cagaaacagt attggctgat cggtttgaga gattggtgga	3840
actggtcaca atgatgggtg atcaaggaga actccctata gcgatggctc tggccaatgt	3900
ggttccttgt tctcagtggt atgaactagc tcgagttctg gttactctgt ttgattctcg	3960
gcatttactc taccactgct tctggaacat gttttctaaa gaagtagaat tggcagactc	4020
catgcagact ctctccag gcaacagctt ggccagtaaa ataatgacat tctgtttcaa	4080
ggtatatggt gctacctatc tacaaaaact cctggatcct ttattacgaa ttgtgatcac	4140
atcctctgat tggcaacatg ttagctttga agtggatcct accaggttag aaccatcaga	4200

-continued

gagccttgag gaaaaccage ggaacctcct tcagatgact gaaaagtctt tccatgccat	4260
catcagttcc tctcagaat tccccctca acttcgaagt gtgtgccact gtttatacca	4320
ggtggttagc cagcgtttcc ctcaaacag catcgggtgca gttaggaagt ccatgttctc	4380
cagatttate aatcctgcca ttgtctcacc gtatgaagca gggatttttag ataaaaagcc	4440
accacctaga atcgaaagg gcttgaagtt aatgtcaaag atacttcaga gtattgccaa	4500
tcatgttctc ttcacaaaag aagaacatat gcggccttcc aatgattttg tgaaaagcaa	4560
ctttgatgca gcacgcaggt ttttccttga tatagcatct gattgtccta caagtgatgc	4620
agtaaactcat agtctttcct tcataagtga cggcaatgtg cttgctttac atcgtctact	4680
ctggaacaat caggagaaaa ttgggcagta tctttccagc aacagggatc ataaagctgt	4740
tggaagacga ccttttgata agatggcaac acttcttgca tacctgggtc ctccagagca	4800
caaacctgtg gcagatacac actggtcag ccttaacctt accagttcaa agtttgagga	4860
atztatgact aggcacaggt tacatgaaaa agaagaattc aaggctttga aaacgttaag	4920
tattttctac caagctggga cttccaaagc tgggaatcct attttttatt atgttgcacg	4980
gagggtcaaa actgggtcaaa tcaatgggtga tttgctgata taccatgtct tactgacttt	5040
aaagccatat tatgcaaagc catatgaaat tgtagtggac cttaccata cgggacctag	5100
caatcgcttt aaaacagact ttctctctaa gtggtttgtt gtttttctg gctttgctta	5160
cgacaacgtc tccgcagtct atatctataa ctgtaactcc tgggtcaggg agtacacca	5220
gtatcatgag cggctgctga ctggcctcaa aggtagcaaa aggcttgttt tcatagactg	5280
tcttgggaaa ctggctgagc acatagagca tgaacaacag aaactacctg ctgccacctt	5340
ggctttagaa gaggacctga aggtattcca caatgctctc aagctagctc acaaagacac	5400
caaaagtctt attaaagttg gttctactgc tgtccaagta acttcagcag agcgaacaaa	5460
agtcctaggg caatcagctt ttctaaatga catttattat gcttcggaaa tgaagaaat	5520
ctgcctagta gatgagaacc agttcacctt aaccattgca aaccagggca cgcgctcac	5580
cttcatgca caggagtgtg aagccattgt ccagctctac attcatatcc ggaccctgtg	5640
ggaactgtca cagcccagct ctatcccca acacaccaag attcggccaa aagatgtccc	5700
tgggacactg ctcaatatcg cattaactaa tttaggcagt tctgaccoga gtttaaggtc	5760
agctgcctat aatcttctgt gtgccttaac ttgtacctt aatttaaaaa ttaggggcca	5820
gttactagag acatcaggtt tatgtatccc tgccaacaac accctcttta ttgtctctat	5880
tagtaagaca ctggcagcca atgagccaca cctcacgtta gaatttttgg aagagtgtat	5940
ttctggattt agcaaatcta gtattgaatt gaaaacacct tgtttggaat acatgactcc	6000
atggctgtca aatctagttc gtttttgcaa gcataatgat gatgccaac gacaaagagt	6060
tactgtctat cttgacaagc tgataacaat gaccatcaat gaaaaacaga tgtacccatc	6120
tattcaagca aaaatatggg gaagccttgg gcagattaca gatctgcttg atgttgtact	6180
agacagtttc atcaaaaacca gtgcaacaggt tggtttggga tcaataaaaag ctgaggtgat	6240
ggcagatact gctgtagctt tggtctctgg aaatgtgaaa ttggtttcaa gcaaggttat	6300
tggaaggatg tgcaaaataa ttgacaagac atgcttatct ccaactccta ctttagaaca	6360
acatcttatg tgggatgata ttgctatctt agcacgctac atgctgatgc tgtccttcaa	6420
caattccctt gatgtggcag ctcatcttcc ctacctcttc cagttgtta ctttcttagt	6480

-continued

agccacaggt	ccgctctccc	ttagagcttc	cacacatgga	ctggtcatta	atatcattca	6540
ctctctgtgt	acttggtcac	agcttcattt	tagtgaagag	accaagcaag	ttttgagact	6600
cagctctgaca	gagttctcat	tacccaaatt	ttacttctgt	tttggcatta	gcaaagtcaa	6660
gtcagctgct	gtcattgcct	tccgttccag	ttaccgggac	aggtcattct	ctcctggctc	6720
ctatgagaga	gagacttttg	ctttgacatc	cttggaaca	gtcacagaag	ctttgttggg	6780
gatcatggag	gcatgcata	gagatattcc	aacgtgcaag	tggtgggacc	agtggacaga	6840
actagctcaa	agatttgcct	tccaatataa	tccatccctg	caaccaagag	ctcttgttgt	6900
ctttgggtgt	attagcaaac	gagtgctctc	tgggcagata	aagcagataa	tccgtattct	6960
tagcaaggca	cttgagagtt	gcttaaaagg	acctgacact	tacaacagtc	aagttctgat	7020
agaagctaca	gtaatagcac	taaccaaatt	acagccactt	cttaataagg	actcgcctct	7080
gcacaaagcc	ctcttttggg	tagctgtggc	tgtgctgcag	cttgatgagg	tcaacttgta	7140
ttcagcaggt	accgcacttc	ttgaacaaaa	cctgcatact	ttagatagtc	tccgtatatt	7200
caatgacaag	agtcagaggt	aagtatttat	ggcaatccgg	aatcctctgg	agtggcactg	7260
caagcaaatg	gatcattttg	ttggactcaa	tttcaactct	aactttaact	ttgcattggt	7320
tggcacactt	ttaaaaggtt	acaggcatcc	ttcacctgct	attgttgcaa	gaacagtcag	7380
aattttacat	acactactaa	ctctggttaa	caaacacaga	aattgtgaca	aatttgaagt	7440
gaatacacag	agcgtggcct	acttagcagc	tttacttaca	gtgtctgaag	aagttcgaag	7500
tcgctgcagc	ctaaaacata	gaaagtcact	tcttcttact	gatatttcaa	tggaaaatgt	7560
tcctatggat	acatatccca	ttcatcatgg	tgacccttcc	tataggacac	taaaggagac	7620
tcagccatgg	tctctcccac	aaggttctga	aggatacctt	gcagccacct	atccaactgt	7680
cggccagacc	agtccccgag	ccaggaaaac	catgagcctg	gacatggggc	aaccttctca	7740
ggccaacact	aagaagttgc	ttggaacaag	gaaaagtttt	gatcacttga	tatcagacac	7800
aaaggctcct	aaaagccaag	aaatggaatc	agggatcaca	acacccccca	aaatgaggag	7860
agtagcagaa	actgattatg	aaatggaaac	tcagaggatt	tcctcatcac	aacagcacc	7920
acatttacgt	aaagtttcag	tgtctgaatc	aaatgttctc	ttggatgaag	aagtacttac	7980
tgatccgaag	atccaggcgc	tgcttcttac	tgttctagct	acactggtaa	aatataccac	8040
agatgagttt	gatcaacgaa	ttctttatga	atacttagca	gaggccagtg	ttgtgtttcc	8100
caaagtcttt	cctggttgge	ataatttgtt	ggactctaag	atcaacaccc	tgttatcatt	8160
gtgccaaagat	ccaaatttgt	taaatccaat	ccatggaatt	gtgcagagtg	tggtgtacca	8220
tgaagaatcc	ccaccacaat	accaaacatc	ttacctgcaa	agttttgggt	ttaatggctt	8280
gtggcggttt	gcaggaccgt	tttcaaagca	aacacaaatt	ccagactatg	ctgagcttat	8340
tgtaagttt	cttgatgcct	tgattgacac	gtacctgcct	ggaattgatg	aagaaaccag	8400
tgaagaatcc	ctcctgactc	ccacatctcc	ttaccctect	gcaactgcaga	gccagcttag	8460
tatcactgce	aaccttaacc	tttctaattc	catgacctca	cttgcaactt	cccagcattc	8520
cccaggaatc	gacaaggaga	acgttgaact	ctccccctacc	actggccact	gtaacagtgg	8580
acgaactcgc	cacggatccg	caagccaagt	gcagaagcaa	agaagcctg	gcagtttcaa	8640
acgtaatagc	attaagaaga	tcgtgtgaag	cttgcttgcct	ttctttttta	aatcaactt	8700
aacatgggct	cttcaactag	gaccccttcc	ctgtccttgc	cctttccccc	catggtgtaa	8760

-continued

tgctgcactt cctgttttat aatgaacca tccggtttgc catgttgcca gatgatcaac	8820
tcttcgaagc cttgcctaaa tttaatgtcg ccttttcttt aacttttttt cttctacttt	8880
tggcgtgtat ctggatatg taagtgttca gaacaactgc aaagaaagtg ggaggtcagg	8940
aaacttttaa ctgagaaat	8959

<210> SEQ ID NO 41

<211> LENGTH: 2257

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 41

acggcagccg tcagggaccg tcccccaact ccccttccg ctcaggcagg gtectcgccg	60
cccatgctgg ccgctgggga cccgcgcagc ccagaccgtt cccgggcccg ccagccggca	120
ccatggtggc cctgaggcct gtgcagcaac tccagggggg ctaaagggtc cagagtgcag	180
gcccgtgggg gcgaggggtc cgggcctgag ccccgcccca tggccggggc catcgcttcc	240
cgcatgagct tcagctctct caagaggaag caaccaaga cgttcaccgt gaggatcgtc	300
accatggacg ccgagatgga gttcaattgc gagatgaagt ggaaggga ggacctttt	360
gatttggtgt gccgactct ggggtccga gaaacctggt tctttggact gcagtacaca	420
atcaaggaca cagtggcctg gctcaaatg gacaagaagg tactggatca tgatgttca	480
aaggaagaac cagtcacctt tcacttcttg gccaaatatt atcctgagaa tgctgaagag	540
gagctgggtc aggagatcac acaacattta ttcttcttac aggtaaagaa gcagatttta	600
gatgaaaaga tctactgccc tcctgaggct tctgtgctcc tggcttctta cgcctccag	660
gccaaagtat gtgactacga cccagtggt cacaagccgg gatttttggc ccaagaggaa	720
ttgcttccaa aaaggtaaat aaatctgtat cagatgactc cggaaatgtg ggaggagaga	780
attactgctt ggtacgcaga gcaccgaggc cgagccaggg atgaagctga aatggaatat	840
ctgaagatag ctcaggacct ggagatgtac ggtgtgaact actttgcaat ccggaataaa	900
aagggcacag agctgctgct tggagtggat gccctggggc ttcacattta tgaccctgag	960
aacagactga cccccagat ctccttccc tggaaatgaaa tccgaaacat ctcgtacagt	1020
gacaaggagt ttactattaa accactggat aagaaaattg atgtcttcaa gtttaactcc	1080
tcaaagcttc gtgtaataa gctgattctc cagctatgta tcgggaacca tgatctattt	1140
atgaggagaa gaaagccga ttctttgaa gttcagcaga tgaagccca ggccaggag	1200
gagaaggcta gaaagcagat ggagcggcag cgcctcgctc gagagaagca gatgaggag	1260
gaggctgaac gcacgaggga tgagtggag aggaggetgc tgcagatgaa agaagaagca	1320
acaatggcca acgaagcact gatgctgtct gaggagacag ctgacctgtt ggctgaaaag	1380
gcccagatca ccgaggagga ggcaaaactt ctggcccaga agccgcaga ggctgagcag	1440
gaaatgcagc gcatcaaggc cacagcgatt cgcacggagg aggagaagcg cctgatggag	1500
cagaagggtc tggaaagccga ggtgctggca ctgaagatgg ctgaggagtc agagaggag	1560
gccaaagagg cagatcagct gaagcaggac ctgcaggaag cacgcgaggc ggagcgaaga	1620
gccaaagcaga agctcctgga gattgccacc aagcccact acccgcccat gaaccaatt	1680
ccagcaccgt tgctcctgca cataccaagc ttcaacctca ttggtgacag cctgtctttc	1740
gacttcaag atactgacat gaagcggctt tccatggaga tagagaaaga aaaagtggaa	1800

-continued

tacatggaaa agagcaagca tctgcaggag cagctcaatg aactcaagac agaaatcgag	1860
gccttgaaac tgaagagag ggagacagct ctggatattc tgcacaatga gaactccgac	1920
aggggtggca gcagcaagca caataaccatt aaaaagctca ccttgcagag cgccaagtec	1980
cgagtggcct tctttgaaga gctctagcag gtgacccagc caccacagga cctgccactt	2040
ctctgctac cgggaccgcg ggatggacca gatatcaaga gagccatcca tagggagctg	2100
gctggggggt tccgtgggag ctccagaact tccccagct gagtgaagag cccagccct	2160
cttatgtgca attgccttga actacgacc tgtagagatt tctctcatgg cgttctagtt	2220
ctctgacctg agtctttggt ttaagaagta tttgtct	2257

<210> SEQ ID NO 42

<211> LENGTH: 2969

<212> TYPE: DNA

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 42

ccaggcagct ggggtaagga gttcaaggca ggcacacac ccgggggctc tccgcaacct	60
gaccgcctgt ccgctcccc acttcccgcc ctccctcca cctactcatt caccaccca	120
cccaccaga gcgggagcgg cagcccaggc gcccgggccc cgccgtctcc tgcggegat	180
cctggacttc ctcttgctgc aggacccggc ttccacgtgt gtcccgagc cgccgtctca	240
gcacacgctc cgctccgggc ctgggtgctt acagcagcca gagcagcagg gagtccggga	300
cccggggggc atctgggcca agttaggcgc cgccgaggcc agcgtgaac gtctccaggg	360
ccggaggagc cgccgggctt ccgggtctga gccgcagcaa atgggctccg acgtgcggga	420
cctgaaacgc ctgctgcccc ccgtccccct cctgggtggt ggccggggct gtgccctgcc	480
tgtgagcggc gcggcgagct gggcgccggt gctggacttt gcgccccgg gcgcttcggc	540
ttacgggtcg ttgggcggcc ccgcccggcc accggctccg ccgccacccc cgccgcccgc	600
gcctcactcc ttcataaac aggagccgag ctggggcggc gcggagccgc acgaggagca	660
gtgctgagc gccttactg tccacttttc cggccagttc actggcacag ccggagcctg	720
tcgctacggg cccttcggtc ctccctccgc cagccaggcg tcatccggcc aggccaggat	780
gtttcctaac gcgccatacc tgcccagctg cctcgagagc cagcccgcta ttcgcaatca	840
gggttacage acggteacct tcgacgggac gccagctac ggtcacacgc cctcgacca	900
tgccggcagc tcccccaacc actcattcaa gcatgaggat cccatgggccc agcagggctc	960
gctgggtgag cagcagctact cgggtcccgc cccggtctat ggctgccaca cccccaccga	1020
cagctgcacc ggcagccagg ctttgctgct gaggacgccc tacagcagtg acaatttata	1080
ccaaatgaca tcccagcttg aatgcatgac ctggaatcag atgaacttag gagccacctt	1140
aaagggccac agcacagggt acgagagcga taaccacaca acgcccaccc tctgcccaggc	1200
ccaatacaga atacacacgc acgggtgtctt cagaggcatt caggatgtgc gacgtgtgccc	1260
tggagtagcc ccgactcttg tacggtcggc atctgagacc agtgagaaac gcccttcat	1320
gtgtgcttac ccaggctgca ataagagata ttttaagctg tcccacttac agatgcacag	1380
caggaagcac actggtgaga aaccatacca gtgtgacttc aaggactgtg aacgaagggt	1440
ttctcgttca gaccagctca aaagacacca aaggagacat acagggtgta aaccattcca	1500
gtgtaaaact tgtcagcga agttctccc gtcggaccac ctgaagaccc acaccaggac	1560

-continued

```

tcatacaggt gaaaagccct tcagctgtcg gtggccaagt tgtcagaaaa agtttgcccg 1620
gtcagatgaa ttagtccgcc atcacaacat gcatcagaga aacatgacca aactccagct 1680
ggcgcttga ggggtctccc tgggggaccg ttcagtgtcc caggcagcac agtgtgtgaa 1740
ctgctttcaa gtctgactct ccaactctcc tcaactaaaa ggaaacttca gttgatcttc 1800
ttcatccaac ttccaagaca agataccggg gcttctggaa actaccaggt gtgcctggaa 1860
gagttggtct ctgcctgcc tacttttagt tgactcacag gcctggaga agcagctaac 1920
aatgtctggt tagttaaag cccattgcca tttggtgtgg atttctact gtaagaagag 1980
ccatagctga tcatgtcccc ctgacccttc ccttcttttt ttatgctcgt tttcgctggg 2040
gatggaatta ttgtaccatt ttctatcatg gaatatttat aggccagggc atgtgtatgt 2100
gtctgcta at gtaaactttg tcatggttcc catttactaa cagcaacagc aagaaataaa 2160
tcagagagca aggcacggg ggtgaatctt gtctaacatt cccgaggtca gccaggtctc 2220
taacctggaa agcaggatgt agttctgcca ggcaactttt aaagctcatg catttcaagc 2280
agctgaagaa aaaatcagaa ctaaccagta cctctgtata gaaatctaaa agaattttac 2340
cattcagtta attcaatgtg aacactggca cactgctctt aagaaactat gaagatctga 2400
gatttttttg tgtatgtttt tgactctttt gagtggtaat catatgtgtc tttatagatg 2460
tacatactc cttgcacaaa tggaggggaa ttcattttca tcaactggag tgccttagt 2520
gtataaaaac catgctggta tatggcttca agttgtaaaa atgaaagtga ctttaaaaga 2580
aaatagggga tggctcagga tctccactga taagactgtt ttaagtaac ttaaggacct 2640
ttgggtctac aagtatatgt gaaaaaatg agacttactg ggtgaggaaa tccattgttt 2700
aaagatggtc gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgttttgtt 2760
ttttaagga ggaatttat tatttaccgt tgcttgaat tactgtgtaa atatatgtct 2820
gataatgatt tgctctttga caactaaat taggactgta taagtactag atgcatcact 2880
gggtgttgat cttacaagat attgatgata acacttaaaa ttgtaacctg catttttcac 2940
tttgctctca attaaagtct attcaaaag 2969

```

```

<210> SEQ ID NO 43
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic DNA

```

```

<400> SEQUENCE: 43

```

```

ggcctccata aagtaggaaa cactacacag ctccataaag taggaaacac tacattaatt 60
aagcgggtac 69

```

```

<210> SEQ ID NO 44
<211> LENGTH: 61
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic DNA

```

```

<400> SEQUENCE: 44

```

```

cgcttaatta atgtagtgtt tctacttta tggagctgtg tagtgtttcc tactttatgg 60

```

```

a

```

```

61

```

-continued

<210> SEQ ID NO 45
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic DNA

<400> SEQUENCE: 45

tccataaagt aggaaacact acaggactcc ataaagtagg aaacactaca gtac 54

<210> SEQ ID NO 46
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: synthetic DNA

<400> SEQUENCE: 46

tgtagtggtt cctactttat ggagtctctgt agtgtttect actttatgga at 52

<210> SEQ ID NO 47
<211> LENGTH: 573
<212> TYPE: DNA
<213> ORGANISM: Adenovirus

<400> SEQUENCE: 47

ttatggactg gaataaaccc tccacctaac tgtcaaattg tggaaaacac taatacaaat 60
gatggcaaac ttacttttagt attagtaaaa aacggagggc ttgttaatgg ctacgtgtct 120
ctagttggtg tatcagacac tgtgaaccaa atgttcacac aaaagacagc aaacatccaa 180
ttaagattat attttgactc ttctggaaat ctattaactg atgaatcaga cttaaaaatt 240
ccacttaaaa ataaatcttc tacagcgacc agtgaaactg tagccagcag caaagccttt 300
atgccaagta ctacagctta tcccttcaac accactacta gggatagtga aaactacatt 360
catggaatat gtactacat gactagttat gatagaagtc tatttccctt gaacatttct 420
ataatgctaa acagccgtat gatttcttcc aatgttgect atgccataca atttgaatgg 480
aatctaattg caagtgaatc tccagaaagc aacatagcta cgctgaccac atcccccttt 540
ttcttttctt acattacaga agcagacaac taa 573

<210> SEQ ID NO 48
<211> LENGTH: 270
<212> TYPE: DNA
<213> ORGANISM: Adenovirus

<400> SEQUENCE: 48

ggagttctta ctcttaagtg tttaacccca ctaacaacca caggcggatc tctacagcta 60
aaagtgggag ggggacttac agtggatgac actgatggta ccttacaaga aaacatacgt 120
gctacagcac ccattactaa aaataatcac tctgtagaac tatccattgg aaatggatta 180
gaaactcaaa acaataaact atgtgocaaa ttgggaaatg gggtaaaatt taacaacggt 240
gacatttgta taaaggatag tattaacacc 270

<210> SEQ ID NO 49
<211> LENGTH: 843
<212> TYPE: DNA
<213> ORGANISM: Adenovirus

-continued

<210> SEQ ID NO 51
<211> LENGTH: 35324
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic DNA

<400> SEQUENCE: 51

tatattattg atgatgcctt attttggatt gaagccaata tgataatgag ggggtggagt 60
ttgtgacgtg gcgcggggcg tgggaacggg gcggtgacg tagtagtggt gcggaagtgt 120
gatgttgcaa gtgtggcgga acacatgtaa gcgacggatg tggcaaaagt gacgtttttg 180
gtgtgcgccg gtgtacacag gaagtgacaa ttttcgcgcg gttttaggcg gatgtttag 240
taaatttggg cgtaaccgag taagatttgg ccattttcgc gggaaaactg aataagagga 300
agtgaaatct gaataathtt gtgttactca tagcgcgtaa tctctagcat cgtaactata 360
acggtcctaa ggtagegaaa gctcagatct cccgatcccc tatggtgcac tctcagtaca 420
atctgctctg atgccgcata gttaagccag tatctgctcc ctgcttgtgt gttggaggtc 480
gctgagtagt gcgcgagcaa aatttaagct acaacaaggc aaggcctgac cgacaattct 540
agcctcgacg cgttgcccc tccctcgggt taccccacag cctaggccga ttcgacctct 600
ctccgctggg gccctcgtg gcgtccctgc accctgggag cgcgagcggc gcgcgggcgg 660
ggaagcgcgg cccagacccc cgggtccgcc cggagcagct gcgctgtcgg ggcagggccg 720
ggctcccagt ggattcgcgg gcacagacgc ccaggaccgc gctccccacg tggcggaggg 780
actggggacc cgggcacccg tcctgcccct tcaccttcca gctccgcctc ctccgcgagg 840
acccgcccc gtcccagacc ctcccgggtc cccggcccag ccccctcggg gccctcccag 900
ccccctccct tcctttcgcg ggccccgccc tctcctcgcg gcgagagttt caggcagcgc 960
tgcgctcctg tgcgcacgtg ggaagccctg gccccggcca ccccccgat agatctcgag 1020
aattcacgcg aattcggcct acaccgggac tgaaaatgag acatattatc tgccacggag 1080
gtgttattac cgaagaaatg gccgcccagc ttttggacca gctgatcgaa gaggtactgg 1140
ctgataatct tccacctcct agccattttg aaccacctac ccttcacgaa ctgtatgatt 1200
tagacgtgac ggcccccgaa gatccccacg aggaggcggg ttcgcagatt tttcccagct 1260
ctgtaatggt ggcggtgcag gaagggattg acttactcac ttttcgcgcg gcgcccgggt 1320
ctccggagcc gcctcacctt tcccggcagc cgcgagcagc ggagcagaga gccttgggtc 1380
cggtttctat gccaaacctt gtaccggagg tgatcgatct tacctgccac gaggtgggt 1440
ttccaccagc tgacgacgag gatgaagagg gtgaggagtt tgtgttagat tatgtggagc 1500
accccgggca cggttgacag tcttgtcatt atcaccggag gaatacgggg gacccagata 1560
ttatgtgttc gctttgctat atgaggacct gtggcatggt tgtctacagt cctgtgtctg 1620
aacctgagcc tgagcccagc ccagaaccgg agcctgcaag acctaccgc cgtcctaaaa 1680
tgggcctcgc tatcctgaga cgcgccacat cacctgtgtc tagagaatgc aatagtagta 1740
cggatagctg tgactccggt ccttctaaca cacctcctga gatacaccgc gtggtcccgc 1800
tgtgccccat taaaccagtt gccgtgagag ttggtgggog tccccaggct gtggaatgta 1860
tcgaggactt gcttaacgag cctgggcaac ctttggactt gagctgtaaa cgcgccaggc 1920
cataaggtgt aaacctgtga agccgaatc gcgctcagca tgcatctagg gcggccaatt 1980

-continued

ccgccctct cccccccc cctctccctc cccccccc taacgttact ggccgaagcc	2040
gcttgaata aggccggtgt gcgtttgtct atatgttatt ttccaccata ttgccgtctt	2100
ttggcaatgt gaggggcccg aaacctggcc ctgtcttctt gacgagcatt cctaggggtc	2160
ttccctct cgcceaagga atgcaaggtc tgttgaatgt cgtgaaggaa gcagttcctc	2220
tggaagcttc ttgaagacaa acaacgtctg tagcgacct ttgcaggcag cggaaacccc	2280
cacctggcga caggtgcctc tgcggccaaa agccacgtgt ataagataca cctgcaaagg	2340
cggcacaacc ccagtgccac gttgtgagtt ggatagttgt ggaagagtc aatggctct	2400
cctcaagcgt attcaacaag gggctgaagg atgcccagaa ggtaccccat tgtatgggat	2460
ctgatctggg gcctcggtc acatgcttta catgtgttta gtcgaggta aaaaaactc	2520
taggcccc gaaccacggg gacgtggtt tcccttgaaa aacacgatga taagcttgc	2580
acaaccggg atcctctaga tgcgaaatc ggcttctgac ctcatggagg cttgggagtg	2640
tttgaagat tttctgctg tgcgtaactt gctggaacag agctctaaca gtacctctg	2700
gttttgagg tttctgtggg gctcatccca ggcaaagta gtctgcagaa ttaaggagga	2760
ttacaagtgg gaattgaag agcttttgaa atcctgtggt gagctgtttg attccttgaa	2820
tctgggtcac caggcgcttt tccaagagaa ggtcatcaag actttggatt ttccacacc	2880
ggggcgctc gcggctgctg ttgctttttt gagttttata aaggataaat ggagcgaaga	2940
aaccatctg agcggggggt acctgctgga tttctggcc atgcatctgt ggagagcgg	3000
tgtgagacac aagaatcgcc tgcactgtt gtcttccgtc cgccggcga taataccgac	3060
ggaggagcag cagcagcagc aggaggaagc caggcggcgg cggcaggagc agagcccatg	3120
gaacccgaga gccggcctgg acctcgga atgaatgtt tacagggtgc tgaactgtat	3180
ccagaactga gacgcatttt gacaattaca gaggatgggc aggggctaaa ggggtaaaag	3240
agggagcggg gggcttgtga ggctacagag gaggctagga atctagctt tagcttaatg	3300
accagacacc gtccctgagtg tattactttt caacagatca aggataattg cgctaatgag	3360
cttgatctgc tggcgcagaa gtattccata gacgagctga ccacttactg gctgcagcca	3420
gggatgatt ttgaggagc tattagggtat tatgcaaagg tggcacttag gccagattgc	3480
aagtacaaga tcagcaaact tgtaaatac aggaattgtt gctacatctc tgggaacggg	3540
gccgaggtg agatagatac ggaggatagg gtggccttta gatgtagcat gataaatatg	3600
tggccggggg tgcctggcat ggacggggtg gttattatga atgtaagggt tactggcccc	3660
aattttagcg gtacggtttt cctggccaat accaacctta tctacacgg tgtaagcttc	3720
tatgggttta acaataacct tgtggaagcc tggaccgatg taagggttcg gggctgtgcc	3780
tttactgct gctggaaggg ggtggtgtgt cgccccaaa gcagggttc aattaagaaa	3840
tgctctttg aaaggtgtac cttgggtatc ctgtctgagg gtaactccag ggtgcgccac	3900
aatgtggcct ccgactgtgg ttgcttcagt ctagtgaaaa gcgtggctgt gattaagcat	3960
aacatgggat gtggcaactg cgaggacagg gcctctcaga tgctgacctg ctggacggc	4020
aactgtcacc tgctgaagac cattcacgta gccagccact ctgcaaggc ctggccagtg	4080
tttgacata acatactgac ccgctgttcc ttgcatttgg gtaacaggag gggggtgttc	4140
ctacctacc aatgcaattt gactcacact aagatattgc ttgagccga gagcatgtcc	4200
aaggtaaac tgaacggggt gtttgacatg accatgaaga tctggaagg gctgaggtac	4260

-continued

gatgagacce gcaccaggtg cagaccctgc gagtgtggcg gtaaacatat taggaaccag 4320
cctgtgatgc tggatgtgac cgaggagctg aggcccgatc acttgggtgct ggccctgcacc 4380
cgcgctgagt ttggctctag cgatgaagat acagattgag gtactgaaat atgtgggcaa 4440
gccgaatttc gacccgggcy gcctagcgtt tctagcgttt aaacgggccc tctagactcg 4500
agcggcctcc ataaagtagg aaacactaca cagctccata aagtaggaaa cactacatta 4560
attccataaa gtaggaaaca ctacaggact ccataaagta ggaaacacta cagtaccaag 4620
cttaagttaa aaccgctgat tagcctcgac tgtgccttct agttgccagc catctgttgt 4680
ttgcccctcc cccgtgcctt ccttgaccct ggaaggtgcc actcccactg tcctttccta 4740
ataaaatgag gaaattgcat cgcattgtct gagtaggtgt cattctattc tggggggtgg 4800
ggtggggcag gacagcaagg gggaggattg ggaagacaat agcaggcatg ctggggatgc 4860
ggtgggctct atggcttctg aggcggaaaag aaccagcaga tctgcagatc tgaattcatc 4920
tatgtcgggt gcggagaaaag aggtaatgaa atggcatcga ctccaagatc tgggcgtggt 4980
taagggtggg aaagaatata taagggtggg gtcttatgta gttttgtatc tgttttgca 5040
cagccgcccgc cgcctagagc accaactcgt ttgatggaag cattgtgagc tcatatttga 5100
caacgcgcac gccccatgg gccgggggtgc gtcagaatgt gatgggetcc agcattgatg 5160
gtcgcccctg cctgcccga aactctacta ccttgaccta cgagaccgtg tctggaacgc 5220
cgttgagagc tgcagcctcc gccgcgctt cagccgctgc agccaccgcc cgcgggattg 5280
tgactgactt tgccttctg agcccgttg caagcagtgc agcttcccgt tcatccgcc 5340
gcatgacaa gttgacggct cttttggcac aattggattc tttgaccgg gaacttaatg 5400
tcgtttctca gcagctgtg gatctgcgcc agcaggtttc tgccctgaag gcttcctccc 5460
ctcccaatgc ggtttaaacc ataaataaaa aaccagactc tgtttggatt tggatcaagc 5520
aagtgtcttg ctgtctttat ttagggggtt tgcgcgcgog gtagggccgg gaccagcgg 5580
ctcggctcgt gagggctcgt tgtattttt ccaggacgtg gtaaaggatga ctctggatgt 5640
tcagatacat gggcataaag ccgtctctgg ggtggaggta gcaccactgc agagcttcat 5700
gctgcggggt ggtgtttag atgatccagt cgtagcagga gcgctgggcy tggtgccata 5760
aaatgtcttt cagtacaag ctgattgcca ggggcaggcc cttggtgtaa gtgtttaca 5820
agcggttaag ctgggatggg tgcatacgtg gggatagatg atgcattctg gactgtattt 5880
ttaggttggc tatgttccca gccatatacc tccggggatt catgttgtgc agaaccacca 5940
gcacagtgta tccggtgcac ttgggaaatt tgcattgtag cttagaagga aatgcgtgga 6000
agaacttga gacgccttg tgacctcaa gattttccat gcattcgtcc ataatgatgg 6060
caatgggccc acgggcggcg gcctgggcy agatatttct gggatcacta acgtcatagt 6120
tgtgttccag gatgagatcg tcataggcca tttttacaaa gcgcgggcyg aggggtgccag 6180
actgcggtat aatggttcca tccggcccag gggcgtagtt acctcacag atttgcattt 6240
cccacgcttt gagttcagat ggggggatca tgtctacctg cggggcgatg aagaaaacgg 6300
tttccggggt aggggagatc agctgggaag aaagcaggtt cctgagcagc tgcgacttac 6360
cgcagccggt gggcccgtaa atcacaccta ttaccggctg caactggtag ttaagagagc 6420
tgcagctgcc gtcatecctg agcagggggg ccaactcgtt aagcatgtcc ctgactcgca 6480
tgttttccct gaccaaaatc gccagaaggc gctcgccgcc cagcgatagc agttcttgca 6540

-continued

aggaagcaaa gtttttcaac ggtttgagac cgtccgccgt aggcattgctt ttgagcgttt 6600
gaccaagcag ttccaggcgg tcccacagct cggtcacctg ctctacggca tctcgatcca 6660
gcatatctcc tcgtttcgcg ggtttggggcg gctttcgtg tacggcagta gtcggtgctc 6720
gtccagacgg gccagggtca tgtctttcca cgggcgcagg gtccctcgta gcgtagtctg 6780
ggtcacggtg aaggggtgcg ctccgggctg cgcgctggcc aggggtgcgct tgaggctggt 6840
cctgctggtg ctgaagcgtt gccggtcttc gccctgcgcg tcggccagggt agcatttgac 6900
catggtgtca tagtccagcc cctccgcgcc gtggcccttg gcgcgcagct tgcccttggg 6960
ggaggcgcgg cacgaggggc agtgcagact tttgagggcg tagagcttgg gcgcgagaaa 7020
taccgattcc ggggagtagg catccgcgcc gcaggccccc cagacggtct cgcattccac 7080
gagccagggt agctctggcc gttcggggtc aaaaaccagg tttccccat gctttttgat 7140
gcgtttctta cctctggttt ccatgagccg gtgtccacgc tcggtgacga aaaggctgtc 7200
cgtgtccccg tatacagact tgagaggcct gtccctcgagc ggtgttccgc ggtccctctc 7260
gtatagaaac tcggaccact ctgagacaaa ggctcgcgtc caggccagca cgaaggaggc 7320
taagtgggag gggtagcggc cgttctccac tagggggtcc actcgcctcca ggggtggaag 7380
acacatgtcg ccctcttcgg catcaaggaa ggtgattggt ttgtagggtt aggccacgtg 7440
accgggtggt cctgaagggg ggctataaaa ggggggtggg gcgcgctcgt cctcactctc 7500
ttccgcatcg ctgtctgcga gggccagctg ttgggggtgag tactccctct gaaaagcggg 7560
catgacttct gcgctaagat tgtcagtttc caaaaacgag gaggattga tattcacctg 7620
gcccgcggtg atgcctttga ggggtggcgc atccatctgg tcagaaaaga caatctttt 7680
gttgtcaagc ttggtggcaa acgacccgta gagggcgttg gacagcaact tggcgatgga 7740
gcgcagggtt tggttttgtg cgcgatcgcc gcgctccttg gcccgcatgt ttagctgcac 7800
gtattcgcgc gcaacgcacc gccattcggg aaagacggtg gtgcgctcgt cgggcaccag 7860
gtgcacgcgc caaccgcggt tgtgcagggt gacaaggtea acgctggtgg ctacctctcc 7920
gcgtaggcgc tcgttggtcc agcagaggcg gccgcccttg cgcgagcaga atggcggtag 7980
ggggtctagc tgcgtctcgt ccggggggtc tgcgtccacg gtaaagaacc cgggcagcag 8040
gcgcgcgctg aagtagtcta tcttgcatcc ttgcaagtct agcgcctgct gccatgcgcg 8100
ggcggaage gcgcgctcgt atgggttgag tgggggaccc catggcatgg ggtgggtgag 8160
cgcggaggcg tacatgccgc aaatgtcgta aacgtagagg ggctctctga gtattccaag 8220
atatgtaggg tagcatcttc caccgcggat gctggcgcgc acgtaatcgt atagttcgtg 8280
cgagggagcg aggaggtcgg gaccgaggtt gctacgggcg ggctgctctg ctcggaagac 8340
tatctgcctg aagatggcat gtgagttgga tgatatggtt ggacgctgga agacggtgaa 8400
gctggcgtct gtgagaccta ccgctcacg cacgaaggag gcgtaggagt cgcgcagctt 8460
gttgaccagc tcggcggtga cctgcacgtc tagggcgcag tagtccaggg tttccttgat 8520
gatgtcatac ttatcctgct cctttttttt ccacagctcg cggttgagga caaactcttc 8580
gcggtctttc cagtactctt ggatcggaaa cccgtcggcc tccgaacggt aagagcctag 8640
catgtagaac tggttgacgg cctggtaggc gcagcatccc ttttctaagg gtagcgcgta 8700
tgctcgcgcg gccttcggga gcgaggtgtg ggtgagcgc aaggtgtccc tgaccatgac 8760
tttgaggtac tggatattga agtcagtgtc gtcgcatccg ccctgctccc agagcaaaaa 8820

-continued

gtccgtgcgc	tttttgaac	gcggatttgg	cagggcgaag	gtgacatcgt	tgaagagtat	8880
ctttcccgcg	cgaggcataa	agttgcgtgt	gatgcggaag	ggccccgca	cctcggaaacg	8940
gttgttaatt	acctggggcg	cgagcacgat	ctcgtcaaag	ccgttgatgt	tgtggcccac	9000
aatgtaaagt	tccaagaagc	gcgggatgcc	cttgatggaa	ggcaattttt	taagttcctc	9060
gtaggtgagc	tcttcagggg	agctgagccc	gtgctctgaa	agggcccagt	ctgcaagatg	9120
agggttggaa	gcgacgaatg	agctccacag	gtcacgggcc	attagcattt	gcaggtggtc	9180
gcgaaaggtc	ctaaactggc	gacctatggc	cattttttct	ggggtgatgc	agtagaaggt	9240
aagcgggtct	tgttcccagc	ggtcccaccc	aaggttcgcg	gctaggtctc	gcgcggcagt	9300
cactagaggc	tcattctccg	cgaacttcat	gaccagcatg	aagggcacga	gctgcttccc	9360
aaagggcccc	atccaagtat	aggtctctac	atcgtaggtg	acaagagac	gctcgggtgcg	9420
aggatgcgag	ccgatcggga	agaactggat	ctcccgcac	caattggagg	agtggctatt	9480
gatgtgggta	aagtagaagt	ccctgcgacg	ggccgaacac	tcgtgctggc	ttttgtaaaa	9540
acgtgcgcag	tactggcagc	ggtgcacggg	ctgtacatcc	tgacagaggt	tgaacctgacg	9600
accgcgcaca	aggaagcaga	gtgggaattt	gagcccctcg	cctggcgggt	ttggctgggtg	9660
gtcttctact	tcggctgctt	gtccttgacc	gtctggtgc	tcgaggggag	ttacggtgga	9720
tcggaccacc	acgccgcgcg	agcccaaagt	ccagatgtcc	gcgcgcggcg	gctcggagctt	9780
gatgacaaca	tcgcgcagat	gggagctgtc	catggtctgg	agctcccgcg	gcgtcaggtc	9840
aggggggagc	tcctgcagggt	ttacctcgca	tagacgggtc	agggcgcggg	ctagatccag	9900
gtgataccta	atttccaggg	gctggttggg	ggcggcgtcg	atggcttgca	agaggccgca	9960
tccccgcggc	gcgactacgg	taccgcgcgg	cgggcggtgg	gccgcggggg	tgtccttgga	10020
tgatgcacat	aaaagcgggt	acgcggggca	gcccccgag	gtaggggggg	ctccggaccc	10080
gccgggagag	ggggcagggg	cacgtcggcg	ccgcgcgcgg	gcaggagctg	gtgctgcgcg	10140
cgtaggttgc	tggcgaacgc	gacgacgcgg	cggttgatct	cctgaatctg	gcgcctctgc	10200
gtgaagacga	cgggcccggg	gagcttgaac	ctgaaagaga	gttcgacaga	atcaatttgc	10260
gtgtcgttga	cggcggcctg	gcgcaaaatc	tcctgcacgt	ctcctgagtt	gtcttgatag	10320
gcgatctcgg	ccatgaactg	ctcgatctct	tcctcctgga	gatctccgcg	tccggctcgc	10380
tcccagggtg	cggcgaggtc	gttggaatg	cgggccatga	gctcgcgaga	ggcggtgagg	10440
cctccctcgt	tccagacgcg	gctgtagacc	acgccccctt	cggcatcgcg	ggcgcgcagt	10500
accacctcgc	cgagattgag	ctccacgtgc	cgggcgaaga	cggcgtagtt	tcgcaggcgc	10560
tgaagagagt	agttgagggg	ggtggcgggtg	tgttctgcca	cgaagaagta	cataaccag	10620
cgtcgcaacg	tggattcgtt	gatatcccc	aaggcctcaa	ggcgtccat	ggcctcgtag	10680
aagtccacgg	cgaagtgtaa	aaactgggag	ttgcgcgcgg	acacggtaa	ctcctcctcc	10740
agaagacgga	tgagctcggc	gacagtgtcg	cgcacctcgc	gctcaaaggc	tacaggggccc	10800
tcttctctct	cttcaatctc	ctcttcata	agggcctccc	cttctctctc	ttctggcggc	10860
ggtgggggag	gggggacacg	gcggcgacga	cggcgaccgc	ggaggcggtc	gacaaaagcgc	10920
tcgatcatct	ccccgcggcg	acggcgcagt	gtctcgggtg	cggcgcggcc	gttctcgcgg	10980
gggcgcagtt	ggaagacgcc	gcccgtcatg	tcccggttat	gggttgccgg	ggggctgcca	11040
tgcggcaggg	atacggcgtc	aacgatgcat	ctcaacaatt	gttgtgtagg	tactccgcgc	11100

-continued

ccgagggacc	tgagcgagtc	cgcacgcacc	ggatcggaaa	acctctcgag	aaagcgctct	11160
aaccagtcac	agtcgcaagg	taggctgagc	accgtggcgg	gcggcagcgg	gcggcggtcg	11220
gggttgtttc	tggcggaggt	gctgctgatg	atgtaattaa	agtaggcggg	cttgagacgg	11280
cggatggtcg	acagaagcac	catgtccttg	ggtccggcct	gctgaatgcg	caggcggtcg	11340
gccatgcccc	aggcttcggt	ttgacatcgg	cgcaggtctt	tgtagtagtc	ttgcatgagc	11400
ctttctaccg	gcacttcttc	ttctccttcc	tcttgctctg	catctcttgc	atctatcgct	11460
gcggcggcgg	cggagtttgg	ccgtaggtgg	cgcctcttcc	ctcccatgcg	tgtgaccccc	11520
aagccccctca	tgggtgaag	cagggctagg	tggcgcgaaa	cgcgctcggc	taatatggcc	11580
tgctgcacct	gcgtgaggtg	agactggaag	tcacccatgt	ccacaaagcg	gtggtatgcg	11640
cccgtgttga	tgggtgaagt	gcagttggcc	ataacggacc	agttaacggt	ctggtgacct	11700
ggctgcgaga	gctcggtgta	cctgagacgc	gagtaagccc	tgcagtcaaa	tacgtagtcg	11760
ttgcaagtcc	gcaccaggtg	ctggtatccc	acaaaaaagt	gcggcggcgg	ctggcggtag	11820
aggggccagc	gtaggggtgg	cggggctccg	ggggcgagat	cttccaacat	aaggcgatga	11880
tatccgtaga	tgtacctgga	catccaggtg	atgccggcgg	cgggtggtgga	ggcgcgcgga	11940
aagtgcgga	cgcggttcca	gatgttgcgc	agcggcaaaa	agtgtcccat	ggtcgggacg	12000
ctctggcccg	tcaggcgcgc	gcaatcgttg	acgctctagc	gtgcaaaagg	agagcctgta	12060
agcgggcact	cttccgtggt	ctggtggata	aattcgcaag	ggtatcatgg	cggacgacct	12120
gggttcgagc	cccgtatccg	gcccgtccgc	gtgatccatg	cggttaccgc	ccgctgtctg	12180
aaccaggtg	tgcgacgtca	gacaacgggg	gagtgtctct	tttggtctcc	ttccaggcgc	12240
ggcggctgct	gcgctagctt	ttttggccac	tggccgcgcg	cagcgtaacg	ggttaggctg	12300
gaaagcga	gcattaagtg	gctcgcctcc	tgtagccgga	gggttatctt	ccaagggttg	12360
agtcgcggga	cccccggttc	gagtcctcga	ccggccggac	tgcggcgaac	gggggtttgc	12420
ctccccgtca	tgcaagacct	cgttgcaaaa	ttcctccgga	aacagggacg	agcccccttt	12480
ttgcttttcc	cagatgcatc	cgggtgctcg	gcagatgcgc	ccccctctcc	agcagcggca	12540
agagcaagag	cagcggcaga	catgcagggc	accctcccct	cctcctaacc	cgtcaggagg	12600
ggcgacatcc	gcggttgacg	cggcagcaga	tggtgattac	gaacccccgc	ggcgcggggc	12660
ccggcactac	ctggacttgg	aggagggcga	gggcctggcg	cggctaggag	cgcctctctc	12720
tgagcggcac	ccaaggtgtc	agctgaagcg	tgatacgcgt	gaggcgtacg	tgcgcgggca	12780
gaacctgttt	cgcgaccgcg	agggagagga	gcccaggag	atgcgggatc	gaaagtcca	12840
cgcagggcgc	gagctgcggc	atggcctgaa	tcgcgagcgg	ttgctgcgcg	aggaggactt	12900
tgagcccagc	gcgcaaccg	ggattagtcc	cgcgcgcgca	cacgtggcgg	ccgccacct	12960
ggtaaccgca	tacgagcaga	cggtgaacca	ggagattaac	tttcaaaaa	gctttaacaa	13020
ccacgtgcgt	acgcttgggg	cgcgcgagga	ggtggctata	ggactgatgc	atctgtggga	13080
ctttgtaagc	gcgctggagc	aaaacccaaa	tagcaagccg	ctcatggcgc	agctgttctt	13140
tatagtcgag	cacagcaggg	acaacgaggg	attcagggat	gcgctgctaa	acatagtaga	13200
gcccaggggc	cgtcggctgc	tcgatttgat	aaacatcctg	cagagcatag	tggcgcagga	13260
gcgcagcttg	agcctggctg	acaaggtggc	cgcctcaaac	tattccatgc	ttagcctggg	13320
caagttttac	gcccgaaga	tataccatcc	cccttacggt	cccatagaca	aggaggtaaa	13380

-continued

gatcgagggg	ttctacatgc	gcatggcgct	gaaggtgctt	accttgagcg	acgacctggg	13440
cgtttatcgc	aacgagcgc	tccacaaggc	cgtgagcgtg	agccggcggc	gcgagctcag	13500
cgaccgag	ctgatgcaca	gcctgcaaag	ggcctggct	ggcacgggca	gcgcgatag	13560
agaggccgag	tcctactttg	acgcgggccc	tgacctgcgc	tgggccccaa	gccgacgcgc	13620
cctggaggca	gctggggccc	gacctgggct	ggcggtgcca	cccgcgcgcg	ctggcaacgt	13680
cggcggcgtg	gaggaatatg	acgaggacga	tgagtacgag	ccagaggacg	gcgagtacta	13740
agcggatgatg	tttctgatca	gatgatgcaa	gacgcaacgg	acccggcggg	gcgggcggcg	13800
ctgcagagcc	agccgtccc	ccttaactcc	acggacgact	ggcgccaggt	catggaccgc	13860
atcatgtcgc	tgactcgcgc	caatcctgac	gcgttccggc	agcagccgca	ggccaaccgg	13920
ctctccgcaa	ttctggaagc	ggtggtccc	gcgcgcgcaa	acccacgca	cgagaaggtg	13980
ctggcgatcg	taaaccgcgt	ggccgaaaac	agggccatcc	ggcccacgca	ggccggcctg	14040
gtctacgagc	cgctgcttca	gcgcgtggct	cgttacaaca	gcgcaacgt	gcagaccaac	14100
ctggaccggc	tggtggggga	tgtgcgcgag	gccgtggcgc	agcgtgagcg	cgcgacgag	14160
cagggaacc	tgggtccat	ggttgcaacta	aacgccttcc	tgagtacaca	gcccggccaac	14220
gtcccgggg	gacagggagga	ctacaccaac	tttgtgagcg	cactgcggct	aatggtgact	14280
gagacaccgc	aaagtgaggt	gtaccagtct	ggccagact	atTTTTTcca	gaccagtaga	14340
caaggcctgc	agaccgtaaa	cctgagccag	gctttcaaaa	acttgacggg	gctgtggggg	14400
gtcggggctc	ccacaggcga	ccgcgcgacc	gtgtctagct	tgctgacgcc	caactcgcgc	14460
ctgttgctgc	tgctaatagc	gcccttcacg	gacagtggca	gcgtgtccc	ggacacatac	14520
ctaggtcact	tgctgacact	gtaccgcgag	gccataggtc	aggcgcagct	ggacgagcat	14580
actttccagg	agattacaag	tgtagccgc	gcgctggggc	aggaggacac	gggcagcctg	14640
gaggcaacc	taaactacct	gctgaccaac	cggcggcaga	agatcccctc	ggtgcacagt	14700
ttaaaccagc	aggaggagcg	cattttgcgc	tacgtgcagc	agagcgtgag	ccttaacctg	14760
atgcgcgagc	gggtaaccgc	cagcgtggcg	ctggacatga	ccgcgcgcaa	catggaaccg	14820
ggcatgtatg	cctcaaaccg	gccgtttatc	aaccgcctaa	tggactactt	gcatcgcgcg	14880
gcccgcgtga	accccgagta	tttaccacat	gccatcttga	acccgcactg	gctaccgccc	14940
cctggtttct	acaccggggg	attcgaggtg	cccgagggtg	acgatggatt	cctctggggc	15000
gacatagacg	acagcgtggt	ttccccgcaa	ccgcagaccc	tgctagagtt	gcaacagcgc	15060
gagcaggcag	agggcggcgt	gcgaaaggaa	agcttccgca	ggccaagcag	cttgctccgat	15120
ctaggcgcgtg	cggccccgcg	gtcagatgct	agtagcccat	ttccaagctt	gatagggtct	15180
cttaccagca	ctcgcaccac	ccgcccgcgc	ctgctgggcg	aggaggagta	cctaaacaac	15240
tcgctgctgc	agccgcagcg	cgaaaaaac	ctgcctccgg	catttcccaa	caacgggata	15300
gagagcctag	tggacaagat	gagtagatgg	aagacgtacg	cgcaggagca	cagggacgtg	15360
ccaggccccg	gcccccccac	ccgtcgtcaa	aggcacgacc	gtcagcgggg	tctggtgtgg	15420
gaggacgatg	actcgcgaga	cgacagcagc	gtcctggatt	tgggaggggg	tggcaaccgg	15480
tttgcgcacc	ttcggcccag	gctggggaga	atgttttaaa	aaaaaaaaaa	gcatgatgca	15540
aaataaaaaa	ctcaccacag	ccatggcacc	gagcgttggg	tttctgtat	tccccttagt	15600
atgcggcgcg	cggcgatgta	tgaggaaggt	cctcctcct	cctacgagag	tgtggtgagc	15660

-continued

gcggcgccag	tggcgcgggc	gctgggttct	cccttcgatg	ctccccgga	cccgcggtt	15720
gtgcctccgc	ggtacctgcg	gcctaccggg	gggagaaaca	gcatccgtta	ctctgagttg	15780
gcacccctat	tcgacaccac	ccgtgtgtac	ctgggtggaca	acaagtcaac	ggatgtggca	15840
tccttgaact	accagaacga	ccacagcaac	tttctgacca	cggtcattca	aaacaatgac	15900
tacagcccgg	gggaggaag	cacacagacc	atcaatcttg	acgaccggtc	gcactggggc	15960
ggcgacctga	aaaccatcct	gcataccaac	atgccaaatg	tgaacgagtt	catgtttacc	16020
aataagttta	aggcgcgggg	gatgggtgctg	cgcttgcccta	ctaaggacaa	tcaggtggag	16080
ctgaaatacg	agtgggtgga	gttcacgctg	cccaggggca	actactccga	gaccatgacc	16140
atagacctta	tgaacaacgc	gatcgtggag	cactacttga	aagtgggcag	acagaacggg	16200
gttctggaaa	gcgacatcgg	ggtaaagttt	gacaccgcga	acttcagact	ggggtttgac	16260
cccgtaactg	gtcttctcat	gcctggggta	tatacaaacg	aagccttcca	tccagacatc	16320
atthttgtgc	caggatgcgg	ggtggacttc	acccacagcc	gcctgagcaa	cttgttgggc	16380
atccgcaagc	ggcaaccctt	ccaggagggc	tttaggatca	cctacgatga	tctggagggt	16440
ggtaacattc	ccgcaactgt	ggatgtggac	gcctaccagg	cgagcctgaa	agatgacacc	16500
gaacaggggc	ggggtggcgc	aggcggcagc	aacagcagtg	gcagcggcgc	ggaagagaac	16560
tccaacgcgg	cagccgcggc	aatgcagccg	gtggaggaca	tgaacgatca	tgccattcgc	16620
ggcgacacct	ttgccacaag	ggctgaggag	aagcgcgctg	aggccgaagc	agcggccgaa	16680
gctgcggccc	ccgctgcgca	acccgaggtc	gagaagcctc	agaagaaacc	ggtgatcaaa	16740
cccctgacag	aggacagcaa	gaaacgcagt	tacaacctaa	taagcaatga	cagcaccttc	16800
acccagtacc	gcagctggta	ccttgcatac	aactacggcg	accctcagac	cggaatccgc	16860
tcattggacc	tgctttgcac	tcctgacgta	acctcgggct	cggagcaggt	ctactggctg	16920
ttgccagaca	tgatgcaaga	ccccgtgacc	ttccgctcca	cgcgccagat	cagcaacttt	16980
ccggtgggtg	gcgccagact	gttgcctcgt	cactccaaga	gcttctacaa	cgaccaggcc	17040
gtctactccc	aactcatccg	ccagtttacc	tctctgaccc	acgtgttcaa	tcgctttccc	17100
gagaaccaga	ttttggcgcg	cccgccagcc	cccaccatca	ccaccgtcag	tgaaaacggt	17160
cctgctctca	cagatcacgg	gacgtacccg	ctgcgcaaca	gcatcggagg	agtccagcga	17220
gtgaccatta	ctgacgccag	acgccgcacc	tgcccctacg	tttacaaggc	cctgggcata	17280
gtctcggcgc	gcgtcctatc	gagccgcact	ttttgagcaa	gcatgtccat	ccttatatcg	17340
cccagcaata	acacaggctg	gggcctgcgc	ttcccaagca	agatgtttgg	cggggccaag	17400
aagcgtccg	accaaacacc	agtgcgcgtg	cgcgggcact	accgcgcgcc	ctggggcgcg	17460
cacaaaacgc	gcccactcgg	gcgcaccacc	gtcgatgacg	ccatcgacgc	ggtgggtggag	17520
gaggcgcgca	actacacgcc	cacgcgcgca	ccagtgtcca	cagtggacgc	ggccattcag	17580
accgtgggtg	gcggagcccc	gcgctatgct	aaaatgaaga	gacggcgagg	gcgcgtagca	17640
cgctgccacc	gcccgcgacc	cggcactgcc	gcccacacgc	cggcgggcgc	cctgcttaac	17700
cgcgcacgct	gcaccggcgc	acgggcccgc	atgcgggccc	ctcgaaggct	ggccgcgggt	17760
attgtcactg	tgccccccag	gtccaggcga	cgagcggcgc	ccgcagcagc	cgcggccatt	17820
agtgtatga	ctcagggtcg	caggggcaac	gtgtattggg	tgccgcgactc	ggttagcggc	17880
ctgcgcgtgc	ccgtgcgcac	ccgccccccg	cgcaactaga	ttgcaagaaa	aaactactta	17940

-continued

gactcgtact gttgtatgta tccagcggcg gcggcgcgca acgaagctat gtccaagcgc 18000
aaaatcaaag aagagatgct ccaggtcacg gcgcccggaga tctatggccc cccgaagaag 18060
gaagagcagg attacaagcc ccgaaagcta aagcgggtca aaaagaaaa gaaagatgat 18120
gatgatgaac ttgacgacga ggtggaactg ctgcacgcta ccgcccagcag gcgacgggta 18180
cagtggaaaag gtcgacgcgt aaaacgtggt ttgacgaccg gcaccaccgt agtctttacg 18240
cccgtgagc gctccaccgg cacctacaag cgcgtgtatg atgaggtgta cggcgacgag 18300
gacctgcttg agcaggccaa cgagcgcctc ggggagtttg cctacggaaa gcggcataag 18360
gacatgctgg cgttgccgct ggacgagggc aaccaaacac ctacgctaaa gcccgtaaca 18420
ctgcagcagg tgctgcccgc gcttgccaccg tccgaagaaa agcgcggcct aaagcgcgag 18480
tctggtgact tggcaccac cgtgcagctg atggtacca agcgcagcg actggaagat 18540
gtcttgaaa aatgaccgt ggaacctggg ctggagcccg aggtccgctg gcggccaatc 18600
aagcaggtgg cgcgggact gggcgtgcag accgtggagc ttcagatacc cactaccagt 18660
agcaccagta ttgccaccgc cacagagggc atggagacac aaacgtccc ggttgctca 18720
gcggtggcgg atgcccggt gcaggcggtc gctgcggcgc cgtccaagac ctctacggag 18780
gtgcaaacgg acccgtggat gtttcgctt tcagccccc ggcgcccgcg cgttcgagg 18840
aagtacggcg ccgccagcgc gctactgccc gaatatgccc tacatccttc cattgcccct 18900
accccccgt atcgtggcta cacctaccgc cccagaagac gagcaactac ccgacgccga 18960
accaccactg gaaccgcgcg ccgcccgcgc cgtgcaccgc ccgtgctggc cccgatttcc 19020
gtgcccaggg tggtcgcga aggagcagc accctggtgc tgccaacagc gcgctaccac 19080
cccagcatcg tttaaaagcc ggtctttgtg gttcttgtag atatggcctt caactgccc 19140
ctccgtttcc cggtgccggg attccagga agaatgcacc gtaggagggg catggccggc 19200
cacggcctga cggcggcat gcgtcgtgcg caccaccggc ggcggcgcgc gtcgcaccgt 19260
cgcatgcgcg gcggtatcct gcccctcctt attccactga tcgcccggc gattggcgcc 19320
gtgcccggaa ttgcatccgt ggccttgtag gcgagagac actgattaaa aacaagtgc 19380
atgtgaaaa atcaaaataa aaagtctgga ctctcacgct cgcttggtcc tgtaactatt 19440
ttgtagaatg gaagacatca actttgcgtc tctggcccgc cgacacggct cgcgcccgtt 19500
catgggaaac tggcaagata tcggcaccag caatatgagc ggtggcgcct tcagctgggg 19560
ctcgtgtgg agcggcatta aaaatttcgg ttccaccgtt aagaactatg gcagcaaggc 19620
ctggaacagc agcacaggcc agatgctgag ggataagtgg aaagagcaaa atttccaaca 19680
aaagtggtga gatggcctgg cctctggcat tagcggggtg gtggacctgg ccaaccaggc 19740
agtgcaaaa aagattaaca gtaagcttga tccccgcctt cccgtagagg agcctccacc 19800
ggcgtggag acagtgtctc cagaggggcg tggcgaagc cgtccgcgcc ccgacaggga 19860
agaaactctg gtgacgcaaa tagacgagcc tccctgtac gaggaggcac taaagcaagg 19920
cctgcccacc acccgtccca tcgcccctat ggctaccgga gtgctgggccc agcacacacc 19980
cgtaacgctg gacctgcctc cccccgccga caccagcag aaacctgtgc tgccaggccc 20040
gaccgcccgt gttgtaaccg gtcctagccg cgcgtccctg cgcgcgcgcc ccagcggctc 20100
gcgatcgttg cggcccgtag ccagtggcaa ctggcaaacg aactgaaca gcatcgtggg 20160
tctgggggtg caatccctga agcgcgcagc atgcttctga tagctaacgt gtcgtatgtg 20220

-continued

tgatcatgtat gcgtccatgt cgcgccaga ggagctgctg agccgcgcg cgcgccgtt 20280
ccaagatggc taccctctcg atgatgccgc agtggcttta catgcacatc tcgggccagg 20340
acgcctcgga gtacctgagc cccgggctgg tgcagtttgc ccgcgccacc gagacgtact 20400
tcagcctgaa taacaagttt agaaaccca cgggtggcgc tacgcacgac gtgaccacag 20460
accggcccc a gcgtttgagc ctgcggttca tocctgtgga ccgtgaggat actgcgtact 20520
cgtacaaggc gcggttcacc cttagctgtgg gtgataaccg tgtgctggac atggcttcca 20580
cgtactttga catccgcggc gtgctggaca ggggccctac ttttaagccc tactctggca 20640
ctgcctacaa cgcctgggt cccaagggtg ccccaaatcc ttgcgaatgg gatgaagctg 20700
ctactgctct tgaataaac ctagaagaag aggacgatga caacgaagac gaagtagacg 20760
agcaagctga gcagcaaaaa actcaogtat ttgggcaggc gccttatctt ggtataaata 20820
ttacaaagga gggatttcaa atagggtctg aaggtcaaac acctaaatat gccgataaaa 20880
catttcaacc tgaacctcaa ataggagaat ctgagtggtg cgaacagaa attaatcatg 20940
cagctgggag agtcctaaaa aagactacc caatgaaacc atgttacggt tcatatgcaa 21000
aaccacaaa tgaatatgga gggcaaggca ttcttgtaa gcaacaaat ggaagctag 21060
aaagtcaagt ggaatgcaa tttttctcaa ctactgaggc agccgcaggc aatgggtgata 21120
acttgactcc taaagtggta ttgtacagt aagatgtaga tatagaaacc ccagacactc 21180
atattctta catgccact attaaggaag gtaactcacg agaactaatg ggccaacaat 21240
ctatgcccc aaggcctaat tacattgctt ttagggacaa ttttattggt ctaatgtatt 21300
acaacagcac gggtaatatg ggtgttctgg cgggccaaag atcgcagttg aatgctggtg 21360
tagatttgca agacagaaac acagagcttt cataccagct tttgcttgat tccattgggtg 21420
atagaaccag gtacttttct atgtggaatc aggtgttga cagctatgat ccagatgtta 21480
gaattattga aatcatgga actgaagatg aacttccaaa ttactgcttt ccaactgggag 21540
gtgtgattaa tacagagact cttaccaagg taaaacctaa aacaggtcag gaaaatggat 21600
gggaaaaaga tgctacagaa ttttcagata aaaatgaaat aagagttgga aataattttg 21660
ccatggaaat caatctaaat gccaacctgt ggagaaattt cctgtactcc aacatagcgc 21720
tgtatttgcc cgacaagcta aagtacagtc cttccaactg aaaaatttct gataaaccaa 21780
acacctacga ctacatgaac aagcgagtgg tggctcccgg gctagtggac tgetacatta 21840
accttgagc acgctggctc cttgactata tggacaactg caaccattt aaccaccacc 21900
gcaatgctgg cctgcgctac cgctcaatgt tgctgggcaa tggctgctat gtgcccttc 21960
acatccaggt gcctcagaag ttctttgcca ttaaaaacct ccttctctg ccgggctcat 22020
acacctacga gtggaacttc aggaaggatg ttaacatggt tctgcagagc tccctaggaa 22080
atgacctaa ggttgacgga gccagcatta agtttgatag catttgctt tacgccacct 22140
tttccccat ggcccacaac accgcctcca cgcttgaggc catgcttaga aacgacacca 22200
acgaccagtc ctttaacgac tatctctccg ccgccaacat gctctacct ataccgcca 22260
acgtaccaa cgtgcccata tccatcccct ccgcaactg ggcggcttc ccgggctggg 22320
ccttcaocg ccttaagact aaggaaacc catcactggg ctcgggctac gaccttatt 22380
acacctactc tggctctata cctacactag atggaacctt taacctaac cacacctta 22440
agaaggtggc cattacctt gactctctg tcagctggcc tggcaatgac cgctgctta 22500

-continued

```

cccccaacga gtttgaatt aagcgctcag ttgacgggga gggttacaac gttgcccagt 22560
gtaacatgac caaagactgg ttccctggac aaatgctagc taactataac attggctacc 22620
agggttctta tatcccagag agctacaagg accgcatgta ctcttctttt agaaacttec 22680
agcccatgag ccgtcagggtg gtggatgata ctaaatacaa ggactacca caggtgggca 22740
tcctacacca acacaacaac tctggatttg ttggtacct tgccccacc atgcgcggaag 22800
gacaggccta ccctgctaac ttcccctatc cgcttatagg caagaccgca gttgacagca 22860
ttaoccagaa aaagtttctt tgcgatcgca ccctttggcg catcccattc tccagtaact 22920
ttatgtccat gggcgccactc acagacctgg gccaaaacct tctctacgcc aactccgccc 22980
acgcgctaga catgactttt gagggtggatc ccatggacga gccaccctt ctttatgttt 23040
tgtttgaagt ctttgacgtg gtccgtgtgc accagccgca ccgcgcgctc atcgaaaccg 23100
tgtacctgcy cacgcccctc tcggccggca acgcccacaac ataagaagc aagcaacatc 23160
aacaacagct gcgcgccatg gctccagtga gcaggaactg aaagccattg tcaaagatct 23220
tggttggtgg ccatatttt tgggcacctg tgacaagcgc tttccaggct ttgtttctcc 23280
acacaagctc gcctgcgccca tagtcaatc ggccggctgc gagactgggg gcgtacactg 23340
gatggccttt gcctggaacc cgcaactcaa aacatgctac ctctttgagc cctttggctt 23400
ttctgaccag cgactcaagc aggtttacca gtttgagtac gagtcaactc tgcgcccgtg 23460
cgccattgct tcttcccccg accgctgtat aacgctggaa aagtccccc aaagcgtaca 23520
ggggcccaac tcggccgctt gtggactatt ctgctgcatg tttctccacg cctttgcca 23580
ctggcccaaa actcccatgg atcacaacct caccatgaac cttattaccg ggtacccaa 23640
ctccatgctc aacagtcccc aggtacagcc cacctgcgt cgcaaccagg aacagctcta 23700
cagcttctcg gagcgccact cgccctactt ccgagccac agtgcgcgaga ttaggagcgc 23760
cacttctttt tgtcacttga aaaacatgta aaaataatgt actagagaca cttcaataa 23820
aggcaaatgc ttttatttgt aactctcgg gtgattatct acccccacc ttgccgtctg 23880
cgccgtttaa aatcaaaagg ggttctgccc cgcatcgcta tgcgccactg gcagggacac 23940
gttgcgatac tgggtgttag tgctccactt aaactcaggc acaaccatcc gcggcagctc 24000
ggtgaagttt tcaactcaca ggctgcgcac catcaccaac gcgtttagca ggtcgggcgc 24060
cgatatcttg aagtgcagct tggggcctcc gccctgcgcg ccgaggttgc gatcacagc 24120
gttgcagcac tggaaacta tcagcgccgg gtggtgcacg ctggccagca cgctcttgc 24180
ggagatcaga tccgctcca ggtcctccgc gttgctcagg gcgaacggag tcaactttg 24240
tagctgcctt cccaaaaagg gcgctgccc aggttttgag ttgcaactgc accgtagtgg 24300
catcaaaagg tgaccgtgcc cggctcgggc gttaggatc agcgcctgca taaaagcctt 24360
gatctgctta aaagccacct gagcctttgc gccttcagag aagaacatgc cgcaagactt 24420
gccgaaaaac tgattggcgc gacaggccgc gtcgtgcacg cagcaccttg cgtcggtgtt 24480
ggagatctgc accacatttc ggccccacc gttcttcacg atcttggcct tgetagactg 24540
ctccttcage gcgcgctgcc cgttttgcct cgtcacatcc atttcaatca cgtgctcctt 24600
atztatcata atgcttccgt gtagacactt aagctcgct tcgatctcag cgcagcggtg 24660
cagccacaac gcgcagccc tgggctcgtg atgctttag gtcacctcgc caaacgactg 24720
caggtaacgc tgcaggaatc gcccacatc cgtcacaag gtcttgttgc tggtaaggt 24780

```

-continued

cagctgcaac ccgcggtgct cctcgttcag ccaggtcttg catacggccg ccagagcttc 24840
cacttggtca ggcagtagtt tgaagttcgc ctttagatcg ttatccacgt ggtacttgtc 24900
catcagcgcg cgcgcagcct ccatgccctt ctcccacgca gacacgatcg gcacactcag 24960
cgggttcate accgtaattt cactttccgc ttcgctgggc tcttctctt cctcttgctg 25020
ccgcatacca cgcgccactg ggtcgtcttc attcagccgc cgcactgtgc gcttacctcc 25080
tttgccatgc ttgattagca ccggtggggt gctgaaaccc accatttgta gcgccacatc 25140
ttctctttct tctcgtctgt ccacgattac ctctggtgat ggcgggcgct cgggcttggg 25200
agaagggcgc ttctttttct tcttgggcgc aatggccaaa tccgccgccg aggtcgatgg 25260
ccgcgggctg ggtgtgcgcg gcaccagcgc gtcttgatgat gagtcttct cgtcctcgga 25320
ctcgatacgc cgctcatcc gcttttttgg gggcgcccgg ggaggcggcg gcgacgggga 25380
cggggacgac acgtcctcca tggttggggg acgtcgcgcc gcaccgcgct cgcgctcggg 25440
ggtggtttcg cgtctctctt ctcccgact ggccatttcc ttctctata ggcagaaaaa 25500
gatcatggag tcagtcgaga agaaggacag cctaaccgcc ccctctgagt tcgccaccac 25560
cgctccacc gatgcgccca acgcgcctac caccttcccc gtcgaggcac ccccgcttga 25620
ggaggaggaa gtgattatcg agcaggaccc aggttttgta agcgaagacg acgaggaccg 25680
ctcagtacca acagaggata aaaagcaaga ccaggacaac gcagaggcaa acgaggaaca 25740
agtggggcgg ggggacgaaa ggcattggca ctacctagat gtgggagacg acgtgctggt 25800
gaagcatctg cagcgcagct gcgccattat ctgcgacgcg ttgcaagagc gcagcgatgt 25860
gcccctcgcc atagcggatg tcagccttgc ctacgaacgc cacctattct caccgcgctg 25920
acccccaaa cgccaagaaa acggcacatg cgagcccaac ccgcgcctca acttctaccc 25980
cgtatttgcc gtgccagagg tgcttgccac ctatcacatc tttttccaaa actgcaagat 26040
accctatcc tgccgtgcc aaccgagccg agcggacaag cagctggcct tgcggcaggg 26100
cgctgtcata cctgatatcg cctcgtcaa cgaagtgcc aaaaatcttg agggctcttg 26160
acgcgacgag aagcgcgcgg caaacgctct gcaacaggaa aacagcgaaa atgaaagtca 26220
ctctggagtg ttggtggaac tcgaggggta caacgcgcgc ctagccgtac taaaacgcag 26280
catcgaggtc acccacttgg cctaccgggc acttaaccta ccccccaagg tcatgagcac 26340
agtcagatg gatctgatcg tgcgcgctgc gcagccctg gagagggatg caaatttgca 26400
agaacaaaca gaggagggcc taccgcagct tggcgacgag cagctagcgc gctggcttca 26460
aacgcgcgag cctgccgact tggaggagcg acgcaaaacta atgatggccg cagtgtcctg 26520
taccgtggag cttgagtgca tgcagcgggt ctttgctgac ccggagatgc agcgaagct 26580
agaggaaaca ttgcaactaca cctttcgaca gggctacgta cgccaggcct gcaagatctc 26640
caacgtggag ctctgcaaac tggctccta ccttgggaatt ttgcacgaaa accgccttgg 26700
gcaaaaactg cttcattcca cgctcaaggg cgaggcgcgc cgcgactacg tccgcgactg 26760
cgtttactta tttctatgct acacctggca gacggccatg ggcgtttggc agcagtgtt 26820
ggaggagtgc aacctcaagg agctgcagaa actgctaaag caaaaactga aggacctatg 26880
gacggccttc aacgagcgt ccgtggcgcg gcacctggcg gacatcattt tccccgaacg 26940
cctgcttaa accctgcaac aggtctgccc agacttcacc agtcaaagca tgttgcaaaa 27000
ctttaggaac ttatctctag agcgtcagg aatcttcccc gccacctgct gtgcaacttc 27060

-continued

tagcgacttt gtgccatta agtacgcga atgccctccg ccgcttggg gccactgcta 27120
ccttctgcag ctagccaact accttgccta ccactctgac ataatggaag acgtgagcgg 27180
tgacggtcta ctggagtgtc actgtogctg caacctatgc accccgcacc gctccctggt 27240
ttgcaattcg cagctgctta acgaaagtca aattatcggg acctttgagc tgcagggtcc 27300
ctcgcctgac gaaaagtccg cggctccggg gttgaaactc actccggggc tgtggacgtc 27360
ggcttacctt cgcaaatctg tacctgagga ctaccacgcc cagcagatta ggttctacga 27420
agaccaatcc cggccgcta atgcggagct taccgcctgc gtcattacc agggccacat 27480
tcttgccaa ttgcaagcca tcaacaaagc ccgccaagag tttctgctac gaaagggacg 27540
gggggttac ttggaccccc agtccggcga ggagctcaac ccaatcccc cgccgcccga 27600
gccctatcag cagcagccgc gggcccttgc tcccaggat ggcacccaaa aagaagctgc 27660
agctgccgcc gccaccacgc gacgaggagg aatactggga cagtcaggca gaggaggttt 27720
tggcagagga ggaggaggac atgatggaag actgggagag cctagacgag gaagcttccg 27780
aggtcgaaga ggtgtcagac gaaacaccgt caccctcggg cgcattcccc tcgccggcgc 27840
cccagaaatc ggcaaccggt tccagcatgg ctacaacctc cgctccctcag gcgccgccgg 27900
cactgccctg tcgccgaccc aaccgtagat gggacaccac tggaaaccagg gccggtaagt 27960
ccaagcagcc gccgccgcta gcccaagagc aacaacagcg ccaaggctac cgctcatggc 28020
gcyggcaca gaacgcata gttgcttctg tgcaagactg tgggggcaac atctccttcg 28080
cccgccgctt tcttctctac catcagggcg tggccttccc ccgtaacatc ctgcattact 28140
accgtcatct ctacagccca tactgcaccg gcggcagcgg cagcaacagc agcggccaca 28200
cagaagcaaa ggcgaccgga tagcaagact ctgacaaagc ccaagaaatc cacagcggcg 28260
gcagcagcag gaggaggagc gctgcgtctg gcgcccacg aaccctatc gaccgcgag 28320
cttagaaaca ggatttttcc cactctgtat gctatattc aacagagcag gggccaagaa 28380
caagagctga aaataaaaa caggtctctg cgatccctca cccgcagctg cctgtatcac 28440
aaaaagcgaag atcagcttcc gcgcaocgctg gaagacgagg aggtctctct cagtaatac 28500
tgcgcgctga ctcttaagga ctagtctcgc gccctttctc aaatttaagc gcgaaaacta 28560
cgtcatctcc agcggccaca ccggcgcca gcacctgttg tcagcgccat tatgagcaag 28620
gaaattccca cgccctacat gtggagttac cagccacaaa tgggacttgc ggtggagct 28680
gcccagact actcaaccgc aataaactac atgagcgcgg gacccacat gatatcccg 28740
gtcaacggaa tacgcgcccc ccgaaaccga attctcctgg aacaggcggc tattaccacc 28800
acacctcgtg ataaccttaa tccccgtagt tggcccgctg ccctgggtga ccaggaaagt 28860
cccgtccca ccaactgtgtt acttcccaga gacgccagg ccgaagtca gatgactaac 28920
tcaggggccc agcttccggg cggcttctgt cacagggtgc ggtcgcccgg gcagggata 28980
actcacctga caatcagagg gcgaggtatt cagctcaacg acgagtcggg gagctcctcg 29040
cttggctccc gtccggacgg gacatttcag atcggggggc ccggccgctc ttcattcacg 29100
cctcgtcagg caatcctaac tctgcagacc tcgtcctctg agccgcgctc tggaggcatt 29160
ggaactctgc aatttattga ggagtttgtg ccatcggtct actttaacc cttctcggga 29220
cctcccggcc actatccgga tcaatttatt cctaactttg acgcggtaaa ggactcggcg 29280
gacggctacg actgaatggt aatatgactc tottaaggta gccaaatagg gataattcga 29340

-continued

tttcgcgaag	ggccccgagct	cggtaccocgg	ggatctgcat	tagttattaa	tagtaatcaa	29400
ttacggggtc	attagttcat	agcccatata	tggagtcccg	cgttacataa	cttacggtaa	29460
atggccccgc	tggctgaccg	cccaacgacc	cccgccatt	gacgtcaata	atgacgtatg	29520
ttcccatagt	aacgccaata	gggactttcc	attgacgtca	atgggtggag	tatttacggt	29580
aaactgccca	cttggcagta	catcaagtgt	atcatatgcc	aagtacgccc	cctattgacg	29640
tcaatgacgg	taaatggccc	gcctggcatt	atgccagta	catgacctta	tgggactttc	29700
ctacttgga	gtacatctac	gtattagtca	tcgctattac	catggtgatg	cggttttggc	29760
agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	attccaagt	ctccaccca	29820
ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	ggactttcca	aaatgtcgta	29880
acaactccgc	cccattgacg	caaatggcgg	gtaggcgtgt	acggtgggag	gtctatataa	29940
gcagagctgg	tttagtgaac	cgtcagatcc	gctagccggg	cgccaccatg	gtgagcaagg	30000
gcgaggagct	gttcaccggg	gtggtgccca	tcctggtega	gctggacggc	gacgtaaacg	30060
gccacaagtt	cagcgtgtcc	ggcagggcgg	agggcgatgc	cacctacggc	aagctgaccc	30120
tgaagtcat	ctgcaccacc	ggcaagctgc	cogtgccctg	gcccaccctc	gtgaccacc	30180
tgaactacgg	cgtgcagtgc	ttcagccgct	accccgacca	catgaagcag	cacgacttct	30240
tcaagtccgc	catgcccga	ggctaactcc	aggagcgcac	catcttcttc	aaggacgacg	30300
gcaactacaa	gaccccgccc	gaggtgaagt	tcgagggcga	caccctgggt	aaccgcatcg	30360
agctgaaggg	catcgacttc	aaggaggacg	gcaacatcct	ggggcacaag	ctggagtaca	30420
actacaacag	ccacaacgtc	tatatcatgg	ccgacaagca	gaagaacggc	atcaaggtga	30480
acttcaagat	ccgccacaac	atcgaggacg	gcagcgtgca	gctcgccgac	cactaccagc	30540
agaacacccc	catcgggcag	ggccccgtgc	tgctgcccga	caaccactac	ctgagcacc	30600
agtccgccct	gagcaaagac	cccaacgaga	agcgcgatca	catggtcctg	ctggagtctg	30660
tgaccgcccg	cgggatcact	ctcggcatgg	acgagctgta	caagtaaagc	ggcctccata	30720
aagtaggaaa	cactacacag	ctccataaag	taggaaacac	tacattaatt	ccataaagta	30780
ggaaacacta	caggactcca	taaagtagga	aacactacat	ctagatcata	atcagccata	30840
ccacatttgt	agaggtttta	cttgttttaa	aaaacctccc	acacctcccc	ctgaacctga	30900
aacataaaat	gaatgcaatt	gttgttggtta	acttgtttat	tgcagcttat	aatggttaca	30960
aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	ttttcactg	cattctagtt	31020
gtggtttgtc	caaacctac	aatgtatcct	aagcttggtta	ccgagctcgg	atcatccagc	31080
acagtggcgg	ccgctcgacc	tgcaggcatg	gcggccgcat	cgaatcgcg	atataacagg	31140
gtaatattaa	gttcctgtcc	atccgcaccc	actatcttca	tgttggtgca	gatgaagcgc	31200
gcaagaccgt	ctgaagatag	cttcaacccc	gtgtatccat	atgacacgga	aaccggtcct	31260
ccaactgtgc	ctttttctac	tcctcccttt	gtatccccca	atgggtttca	agagagtccc	31320
cctggagttc	ttactcttaa	gtgtttaacc	ccactaaca	ccacaggcgg	atctctacag	31380
ctaaaagtgg	gagggggact	tacagtggat	gacactgatg	gtaccttaca	agaaaacata	31440
cgtgctacag	caccattac	taaaaataat	cactctgtag	aactatccat	tggaaatgga	31500
ttagaaactc	aaaacaataa	actatgtgcc	aaattgggaa	atgggttaaa	atthaacaac	31560
ggtgacattt	gtataaagga	tagtattaac	accttatgga	ctggaataaa	ccctccacct	31620

-continued

aactgtcaaa ttgtggaaaa cactaataca aatgatggca aacttacttt agtattagta 31680
aaaaacggag ggcttgtaa tggctacgtg tctctagttg gtgtatcaga cactgtgaac 31740
caaatgttca cacaaaagac agcaaacatc caattaagat tatattttga ctcttctgga 31800
aatctattaa ctgatgaatc agacttaaaa attccactta aaaataaatc ttctacagcg 31860
accagtgaia ctgtagccag cagcaaaagcc tttatgcca gtactacagc ttatcccttc 31920
aacaccacta ctagggatag tgaaaactac attcatggaa tatgttacta catgactagt 31980
tatgatagaa gtctatttcc cttgaacatt tctataatgc taaacagccg tatgatttct 32040
tccaatgttg cctatgccat acaatttgaa tggaaatctaa atgcaagtga atctccagaa 32100
agcaacatag ctacgctgac cacatcccc ttttctttt cttacattac agaagacgac 32160
aactaaaatg aattcagaat cgtttgtgtt atgtttcaac gtgtttattt tcaattgca 32220
gaaaatttca agtcattttt cattcagtag tatagcccca ccaccacata gcttatacag 32280
atcaccgtac cttaatcaaa ctcacagaac cctagtattc aacctgccac ctccctccca 32340
acacacagag tacacagtcc tttctcccc gctggcctta aaaagcatca tatcatgggt 32400
aacagacata ttcttaggtg ttatattcca cacggtttcc tgcgagcca aacgctcatc 32460
agtgatatta ataaactccc cgggcagctc acttaagttc atgtcgtgtt ccagctgctg 32520
agccacaggc tgctgtccaa cttgcggttg cttaacgggc gggaaggag aagtcacgc 32580
ctacatgggg gtagagtcac aatcgtgcat caggataggg cggtggtgct gcagcagcg 32640
gcgaataaac tgctgcccgc gccctccgt cctgcaggaa tacaacatgg cagtggctc 32700
ctcagcgatg attcgcaccg cccgcagcat aaggcgcctt gtcctccggg cacagcagcg 32760
caccctgatc tcaactaaat cagcagatg actgcagcac agcaccacaa tattgttcaa 32820
aatcccacag tgcaaggcgc tgtatocaaa gctcatggcg gggaccacag aaccoacgtg 32880
gccatcatac cacaagcgca ggtagattaa gtggcgacc ctcataaaca cgctggacat 32940
aaacattacc tcttttgga tgttgtaatt caccacctcc cgggtaccata taaacctctg 33000
attaaacatg gcgcatcca ccaccatcct aaaccagctg gccaaaacct gcccgccggc 33060
tataactgc aggaacccg gactggaaca atgacagtgg agagcccagg actcgttaacc 33120
atggatcacc atgctcgtca tgatatcaat gttggcaca cacaggcaca cgtgcataca 33180
cttctcagg attacaagct cctcccgcgt tagaaccata tcccaggga caaccatc 33240
ctgaatcagc gtaaatccca cactgcaggg aagacctcgc acgtaactca cgttgtgcat 33300
tgtcaaatg ttacattcgg gcagcagcgg atgatcctcc agtatgtag cgcgggtttc 33360
tgtctcaaaa ggaggtagac gatccctact gtacggagtg cgcagagaca accgagatcg 33420
tgttggtcgt agtgtcatgc caaatggaac gccggacgta gtcataattc ctgaagcaaa 33480
accaggtgcg gcggtgacaa acagatctgc gtctccggtc tcgccgctta gatcgtctg 33540
tgtagtagtt gtagtatatc cactctctca aagcatccag gcgccccctg gcttcgggtt 33600
ctatgtaaac tccttcacgc gccgctgccc tgataacatc caccaccgca gaataagcca 33660
caccagcca acctacacat cgttctgcy agtcacacac gggaggagcg ggaagagctg 33720
gaagaacct gttttttttt ttattccaaa agattatcca aaacctcaaa atgaagatct 33780
attaagtga cgcgctcccc tccggtggcg tggtaaaact ctacagccaa agaacagata 33840
atggcatttg taagatggtg cacaatggct tccaaaaggc aaacggcctt cactccaag 33900

-continued

```

tggacgtaaa ggctaaaccc ttcaggggtga atctcctcta taaacattcc agcaccttca 33960
accatgceca aataattctc atctcgccac cttctcaata tatctctaag caaatcccga 34020
atattaagtc cggccattgt aaaaatctgc tccagagcgc cctccacctt cagcctcaag 34080
cagcgaatca tgattgcaaa aattcaggtt cctcacagac ctgtataaga ttcaaaagcg 34140
gaacattaac aaaaataccg cgatcccgtg ggtcccttcg cagggccagc tgaacataat 34200
cgtgcaggtc tgcacggacc agcgcggcca cttccccgcc aggaacctag acaaaagaac 34260
ccacactgat tatgacacgc atactcggag ctatgctaac cagcgtagcc ccgatgtaag 34320
cttgttgcat gggcggcgat ataaaatgca aggtgctgct caaaaaatca ggcaaaagcct 34380
cgcgcaaaaa agaaagcaca tcgtagtcat gctcatgcag ataaaggcag gtaagctccg 34440
gaaccaccac agaaaagac accatttttc tctcaaacat gtctgcgggt ttctgcataa 34500
acacaaaata aaataacaaa aaaacattta aacattagaa gcctgtctta caacaggaaa 34560
aacaaccctt ataagcataa gacggactac ggccatgccg gcgtgaccgt aaaaaaactg 34620
gtcacctgta ttaaaaagca ccaccgacag ctctcgggtc atgtccggag tcataatgta 34680
agactcggta aacacatcag gttgattcac atcggtcagt gctaaaaagc gaccgaaata 34740
gccccgggga atacataccc gcaggcgtag agacaacatt acagccccc a taggaggtat 34800
aacaaaatta ataggagaga aaaacacata aacacctgaa aaaccctcct gcctaggcaa 34860
aatagcacc tcccgtcca gaacaacata cagcgcttcc acagcggcag ccataacagt 34920
cagccttacc agtaaaaaag aaaacctatt aaaaaaacac cactcgacac ggcaccagct 34980
caatcagtca cagtgtaaaa aagggccaag tgcagagcga gtatatatag gactaaaaaa 35040
tgacgtaacy gttaaagtcc acaaaaaaca cccagaaaac cgcacgcgaa cctacgcca 35100
gaaacgaaag ccaaaaaacc cacaaacttc tcaaatcgtc acttccgttt tcccacgtta 35160
cgtcacttcc cattttaaga aaactacaat tcccacacc tctagagaca agttactccg 35220
ccctaaaacc tacgtacccc gccccgttcc cagccccgc gccacgtcac aaactccacc 35280
ccctcattat catattggtc tcaatccaaa ataaggtata ttat 35324

```

```

<210> SEQ ID NO 52
<211> LENGTH: 105
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic DNA

```

```

<400> SEQUENCE: 52

```

```

tccataaagt aggaaacact acacagctcc ataaagtagg aaacactaca ttaattccat 60
aaagtaggaa acactacagg actccataaa gtaggaaaca ctaca 105

```

```

<210> SEQ ID NO 53
<211> LENGTH: 4059
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: synthetic DNA

```

```

<400> SEQUENCE: 53

```

```

tggccccctc ctcggttac cccacagcct aggcggattc gacctctctc cgtgggggcc 60
ctcgtggcg tcctgcacc ctgggagcgc gagcggcgcg cgggcgggga agcgcggccc 120

```

-continued

agacccccgg gtcgcgccgg agcagctgcg ctgtcggggc caggccgggc tcccagtgga	180
ttcgcgggca cagacgcccc ggaccgcgct cccacagtgg cggagggact ggggacccgg	240
gcacccgtcc tgccccttca ccttcagct cgcctctc cgcgaggacc cgcgccgtc	300
ccgacccctc ccgggtcccc ggcccagccc cctccgggccc ctcccagccc ctccccttcc	360
tttcgcggcg ccgcacctct cctcgcggcg cgagtctcag gcagcgctgc gtctgctgc	420
gcacgtggga agccctggcc ccggccacc cgcgataga tctcgagaat tcacgcgaat	480
tcggcttaca ccgggactga aaatgagaca tattatctgc cacggagggtg ttattaccga	540
agaaatggcc gccagctctt tggaccagct gatcgaagag gtactggctg ataactctcc	600
acctcctagc cattttgaac cacctacct tcacgaactg tatgatttag acgtgacggc	660
ccccgaagat cccaacgagg aggcggtttc gcagatttt cccgactctg taatgttggc	720
ggtgcaggaa gggattgact tactcacttt tccgcggcg cccggttctc cggagccgcc	780
tcacctttcc cggcagcccg agcagccgga gcagagagcc ttgggtccgg tttctatgcc	840
aaaccttgta ccggagggtga tcgatcttac ctgccacgag gctggctttc caccagtgga	900
cgacgaggat gaagagggtg aggagtgtg gttagattat gtggagcacc ccgggcacgg	960
ttgcaggctc tgtcattatc accggaggaa tacgggggac ccagatatta tgtgtctgct	1020
ttgctatatg aggacctgtg gcattgtttg ctacagctct gtgtctgaac ctgagcctga	1080
gcccagacca gaaccggagc ctgcaagacc taccgcctgt cctaaaatgg cgctgctat	1140
cctgagagcc ccgacatcac ctgtgtctag agaatgcaat agtagtacgg atagctgtga	1200
ctccggtcct tctaacacac ctctgagat acaccgggtg gtcccgtgtg gcccattaa	1260
accagttgcc gtgagagttg gtgggcgtcg ccaggctgtg gaatgtatcg aggacttgc	1320
taacgagcct gggcaacctt tggacttgag ctgtaaacgc cccaggccat aagggtgtaa	1380
cctgtgaagc cgaattcgcg tcgagcatgc atctagggcg gccaatccg cccctctccc	1440
ccccccccct ctcccctccc cccccctaa cgttactggc cgaagccgct tggaaataagg	1500
ccggtgtgcg tttgtctata tgttattttc caccatattg ccgtcttttg gcaatgtgag	1560
ggcccggaaa cctggccctg tcttcttgac gagcattcct aggggtcttt cccctctcgc	1620
caaaggaaat caaggtctgt tgaatgtctg gaaggaagca gttcctctgg aagcttcttg	1680
aagacaaaac acgtctgtag cgacccttg caggcagcgg aacccccac ctggcgacag	1740
gtgcctctgc ggccaaaagc cacgtgtata agatacacct gcaaaggcgg cacaaacca	1800
gtgccacggt gtgagttgga tagttgtgga aagagtcaaa tggctctcct caagcgtatt	1860
caacaagggg ctgaaggatg ccagaagggt acccattgt atgggatctg atctggggcc	1920
tcggtgcaca tgctttacat gtgtttagtc gaggttaaaa aaactctag gcccccgaa	1980
ccacggggac gtggttttcc tttgaaaaac acgatgataa gcttgcaca acccgggatc	2040
ctctagagtc gaaattcggc ttctgacctc atggaggctt gggagtgttt ggaagatfff	2100
tctgctgtgc gtaactgtct ggaacagagc tctaacagta cctcttggtt ttggaggttt	2160
ctgtggggct catcccagcc aaagttagtc tgcagaatta aggaggatta caagtgggaa	2220
tttgaagagc ttttgaatc ctgtgggtgag ctgtttgatt ctttgaatct gggtcaccag	2280
gcgcttttcc aagagaaggc catcaagact ttggatffff ccacaccggg gcgctgtgcg	2340
gctgctgttg cttttttgag ttttataaag gataaatgga gcgaagaaac ccatctgagc	2400

-continued

```

ggggggtacc tgctggattt tctggccatg catctgtgga gagcggttgt gagacacaag 2460
aatcgctgc tactgttgc ttcgctccg cggcgataa taccgacgga ggagcagcag 2520
cagcagcagg aggaagccag gcggcgccg caggagcaga gcccatggaa cccgagagcc 2580
ggcctggacc ctcggaatg aatgtgttac aggtggctga actgtatcca gaactgagac 2640
gcattttgac aattacagag gatgggcagg ggctaaaggg ggtaagagg gagcgggggg 2700
cttgtgagge tacagaggag gctaggaatc tagcttttag cttaatgacc agacaccgtc 2760
ctgagtgtat tacttttcaa cagatcaagg ataattgcgc taatgagctt gatctgctgg 2820
cgcagaagta ttccatagag cagctgacca cttactggct gcagccaggg gatgattttg 2880
aggaggctat tagggtatat gcaaaggtg cacttaggcc agattgcaag tacaagatca 2940
gcaaacttgt aaatatcagg aattgttgc acatctctgg gaacggggcc gaggtggaga 3000
tagatacggg gataggggtg gcctttagat gtagcatgat aaatatgtgg ccgggggtgc 3060
ttggcatgga cgggggtggtt attatgaatg taaggtttac tgcccccaat tttagcggta 3120
cggttttcct ggccaatacc aaccttatcc tacacggtgt aagcttctat gggtttaaca 3180
atacctgtgt ggaagcctgg accgatgtaa gggttcgggg ctgtgccttt tactgctgct 3240
ggaagggggg ggtgtgctgc cccaaaagca gggcttcaat taagaaatgc ctctttgaaa 3300
ggtgtacctt gggatccctg tctgagggta actccagggt gcgccacaat gtggcctccg 3360
actgtgggtt cttcatgcta gtgaaaagcg tggctgtgat taagcataac atgggatgtg 3420
gcaactcga ggacagggcc tctcagatgc tgacctgctc ggacggcaac tgtcacctgc 3480
tgaagaccat tcacgtagcc agccactctc gcaaggcctg gccagtgttt gagcataaca 3540
tactgacccg ctgttctctg catttgggta acaggagggg ggtgttctta ccttaccaat 3600
gcaatttgag tcacactaag atattgcttg agcccagag catgtccaag gtgaacctga 3660
acggggtggt tgacatgacc atgaagatct ggaaggtgct gaggtacgat gagaccgca 3720
ccaggtgcag accctgcgag tgtggcggtg aacatattag gaaccagcct gtgatgctgg 3780
atgtgaccga ggagctgagg cccgatcact tgggtgctggc ctgcaccgcg gctgagttt 3840
gctctagcga tgaagataca gattgaggtg ctgaaatag tgggcaagcc gaatttcgac 3900
ccggcgccgc tagcgtttct agcgtttaa cgggcccctc agactcgagc ggccctcata 3960
aagtaggaaa cactacacag ctccataaag taggaaacac tacattaatt ccataaagta 4020
gaaacacta caggactcca taaagtagga aacactaca 4059

```

<210> SEQ ID NO 54

<211> LENGTH: 1439

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: synthetic DNA

<400> SEQUENCE: 54

```

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60
cgttacataa cttacggtaa atggcccgcg tggctgaccg cccaacgacc cccgccatt 120
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180
atgggtggag tatttacggt aaactgcccc cttggcagta catcaagtgt atcatatgcc 240
aagtaegccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta 300

```

-continued

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac	360
catggtgatg cggtttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg	420
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg	480
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt	540
acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagccggt	600
cgccaccatg gtgagcaagg gcgaggagct gttcaccggg gtggtgcccc tcttggtcga	660
gctggacggc gacgtaaacg gccacaagtt cagcgtgtcc ggcgagggcg agggcgatgc	720
caactacggc aagctgaccc tgaagtcat ctgcaccacc ggcaagctgc ccgtgccttg	780
gcccaccctc gtgaccaccc tgacctacgg cgtgcagtgc ttcagccgct accccgacca	840
catgaagcag cagcacttct tcaagtccgc catgcccga ggctacgtcc aggagcgcac	900
catcttcttc aaggacgacg gcaactacaa gaccgcgcc gaggtgaagt tgcaggcgca	960
caacctgggt aaccgcatcg agctgaaggg catcgacttc aaggaggacg gcaacatcct	1020
ggggcacaag ctggagtaca actacaacag ccacaacgtc tatatcatgg ccgacaagca	1080
gaagaacggc atcaagtgta acttcaagat ccgccacaac atcgaggacg gcagcgtgca	1140
gctcgcgcac cactaccagc agaacacccc catcggcgac ggccccgtgc tgetgcccga	1200
caaccactac ctgagcacc agtccgcct gagcaaagac cccaacgaga agcgcgatca	1260
catggtcctg ctggagtctg tgaccgcgcg cgggatcact ctcggcatgg acgagctgta	1320
caagtaaagc ggctccata aagtaggaaa cactacacag ctccataaag taggaaacac	1380
tacattaatt ccataaagta ggaaacacta caggactcca taaagtagga aacactaca	1439

<210> SEQ ID NO 55
 <211> LENGTH: 105
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic DNA

<400> SEQUENCE: 55

tccataaagt aggaaacact acacagctcc ataaagtagg aaacactaca ttaattccat	60
aaagtaggaa acactacacc actccataaa gtaggaaaca ctaca	105

1. A recombinant adenovirus, which comprises:
 a replication cassette comprising a polynucleotide comprising a human telomerase reverse transcriptase promoter, E1A gene, IRES sequence and E1B gene in this order, and a target sequence of a first microRNA, wherein the first microRNA is miR-142 and the target sequence comprises a nucleotide sequence having at least 98% identity to the nucleotide sequence consisting of SEQ ID NO: 52, and wherein the replication cassette is integrated into the E1 region of the adenovirus genome;
 a labeling cassette comprising a reporter gene, a promoter capable of regulating the expression of the reporter gene, and a target sequence of a second microRNA, wherein the second microRNA is miR-142 and the target sequence comprises a nucleotide sequence having at least 98% identity to the nucleotide sequence consisting of SEQ ID NO: 52, and wherein the labeling cassette is integrated into the E3 region of the adenovirus genome; and

a gene encoding a CD46-binding fiber protein comprising at least the fiber knob region in the fiber protein of adenovirus type 34 or 35, which comprises a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 50 and is integrated into the adenovirus genome.

2. The recombinant adenovirus according to claim 1, wherein the reporter gene is a gene encoding a protein which emits fluorescence or a gene encoding an enzyme protein which generates a luminophore or a chromophore upon enzymatic reaction.

3. The recombinant adenovirus according to claim 1, wherein the promoter capable of regulating the expression of the reporter gene is a human telomerase reverse transcriptase promoter or cytomegalovirus promoter.

4. The recombinant adenovirus according to claim 1, wherein the replication cassette comprises a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 53.

5. The recombinant adenovirus according to claim 1, wherein the labeling cassette comprises a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 54.

6. A recombinant adenovirus, which comprises:

a replication cassette comprising a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 53 and is integrated into the E1 region of the adenovirus genome;

a labeling cassette comprising a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 54 and is integrated into the E3 region of the adenovirus genome; and

a gene encoding a CD46-binding fiber protein comprising a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 50 and is integrated into the adenovirus genome.

7. A recombinant adenovirus, which comprises a nucleotide sequence having at least 99% identity to the nucleotide sequence consisting of SEQ ID NO: 51.

* * * * *